genes

MDPI

Article

A High-Quality, Long-Read *De Novo* Genome Assembly to Aid Conservation of Hawaii's Last Remaining Crow Species

Jolene T. Sutton [1,*], Martin Helmkampf [1], Cynthia C. Steiner [2], M. Renee Bellinger [1], Jonas Korlach [3], Richard Hall [3], Primo Baybayan [3], Jill Muehling [3], Jenny Gu [3], Sarah Kingan [3], Bryce M. Masuda [4] and Oliver A. Ryder [2]

1 Department of Biology, University of Hawaii at Hilo, Hilo, HI 96720, USA; mhelmkam@hawaii.edu (M.H.); belling@hawaii.edu (M.R.B.)
2 Institute for Conservation Research, San Diego Zoo, Escondido, CA 92027, USA; CSteiner@sandiegozoo.org (C.C.S.); ORyder@sandiegozoo.org (O.A.R.)
3 Pacific Biosciences, Menlo Park, CA 94025, USA; jkorlach@pacificbiosciences.com (J.K.); rhall@pacificbiosciences.com (R.H.); pbaybayan@pacificbiosciences.com (P.B.); jmuehling@pacificbiosciences.com (J.M.); jgu@pacificbiosciences.com (J.G.); skingan@pacificbiosciences.com (S.K.)
4 Institute for Conservation Research, San Diego Zoo Global, Volcano, HI 96785, USA; bmasuda@sandiegozoo.org
* Correspondence: jtsutton@hawaii.edu; Tel.: +1-808-932-7183

Received: 6 June 2018; Accepted: 27 July 2018; Published: 1 August 2018

Abstract: Genome-level data can provide researchers with unprecedented precision to examine the causes and genetic consequences of population declines, which can inform conservation management. Here, we present a high-quality, long-read, *de novo* genome assembly for one of the world's most endangered bird species, the ʻAlalā (*Corvus hawaiiensis*; Hawaiian crow). As the only remaining native crow species in Hawai ʻi, the ʻAlalā survived solely in a captive-breeding program from 2002 until 2016, at which point a long-term reintroduction program was initiated. The high-quality genome assembly was generated to lay the foundation for both comparative genomics studies and the development of population-level genomic tools that will aid conservation and recovery efforts. We illustrate how the quality of this assembly places it amongst the very best avian genomes assembled to date, comparable to intensively studied model systems. We describe the genome architecture in terms of repetitive elements and runs of homozygosity, and we show that compared with more outbred species, the ʻAlalā genome is substantially more homozygous. We also provide annotations for a subset of immunity genes that are likely to be important in conservation management, and we discuss how this genome is currently being used as a roadmap for downstream conservation applications.

Keywords: runs of homozygosity (ROH); inbreeding depression; major histocompatibility complex; toll-like receptors; behavior; SMRT sequencing

1. Introduction

Whole-genome sequencing of threatened and endangered taxa enable conservation geneticists to transition from a reliance on limited numbers of genetic markers toward increased resolution of genome-wide genetic variation [1,2]. Such genome-level data offer unprecedented precision to examine the causes and genetic consequences of population declines and to apply these results to conservation management (reviewed in [3,4]). Moreover, continued decreases in the costs of genomic sequencing technologies make this information increasingly available for non-model organisms,

including those with large genomes (e.g., [5,6]; also see [7–9] for cost comparisons across platforms). Although challenges remain for bridging the gap between generating genomic data and applying this information to species management, this gap continues to close (for detailed discussions, see [3,10–13]).

Here, we describe long-read, whole-genome sequencing and *de novo* assembly for the critically endangered ʻAlalā. With 142 birds alive as of March 2018, this species is one of the most endangered avian species to have its genome assembled. The genome now provides valuable resources for conservation efforts, such as positional information and sequence data for candidate genes that are likely to have important fitness consequences (e.g., genes associated with immunity, mate choice, learning, and behavior). A high-quality genome also provides a tool for developing and mapping large numbers of genome-wide markers (e.g., single nucleotide polymorphisms; SNPs), which will improve estimates of relatedness and individual inbreeding coefficients (e.g., [14–17]). Improved relatedness estimates will be important for choosing mating pairs in the conservation-breeding (i.e., captive-breeding) program, where inbreeding depression (i.e., loss of fitness due to inbreeding) has been observed during pedigree analysis [18]. The genome will also be valuable for comparative studies aimed at understanding the evolution of tool use and other behaviors (e.g., [19]). Such comparative genome analyses could be especially important for conservation purposes, as they offer the potential to identify the genetic basis of fitness-related traits, both within and across species.

Study Species and Aims

Historically widespread within mesic and dry forest habitats on the Island of Hawai ʻi, the ʻAlalā population declined rapidly during the twentieth century. Likely causes for the decline include habitat destruction and introductions of avian diseases and ungulates [20,21]. By 1970, the population was estimated at fewer than 100 individuals, at which point a small-scale conservation-breeding program was initiated. During the 1990s the total number of birds declined to less than 20, and in 1996 the last wild individual to supplement the newly modernized, more intensive breeding program was collected [18]. In 2002 the species became extinct-in-the-wild, but by then the number of birds in the breeding program was increasing (reviewed in [22]). Today, the ʻAlalā is one of the most endangered endemic bird species in Hawai ʻi, having existed entirely in captivity from 2002–2016 [18]. All extant individuals are descended from nine genetic founders that established the conservation-breeding program [21,22]. In 2016, a long-term reintroduction program was initiated in an attempt to establish a self-sustaining population in the wild. Although a detailed pedigree has been established and utilized for captive management, including choosing breeding pairs, the current population exhibits signs of inbreeding depression [18]. For example, the species suffers from low hatching success [22]. Until establishment of the long-read genome assembly described here, molecular genetic studies were limited to small numbers of traditional genetic markers (e.g., microsatellite loci, amplified fragment length polymorphism; AFLP, and mitochondrial DNA markers [23,24]). These studies identified extremely low genetic diversity, which suggested that conservation efforts would benefit from a whole-genome approach that could generate resources for assessing the remaining polymorphic regions (e.g., SNPs, and structural variation).

In this study, we highlight the quality of the ʻAlalā genome assembly and compare it to other avian assemblies that were also generated from whole-genome shotgun-sequencing approaches. We provide details for a subset of candidate immunity genes that we hypothesize will have important conservation implications, and we examine the repeat composition of the genome. We also describe analyses of runs of homozygosity (ROH) and the fraction of the genome estimated to be completely autozygous (fROH [25]; i.e., identical by descent). Finally, we briefly discuss the goals and perceived challenges for the next stages of data generation and applications to ʻAlalā conservation and recovery.

2. Materials and Methods

2.1. Library Construction and Sequencing

Phenol-chloroform was used to extract high molecular weight genomic DNA from a blood sample taken from a single male ʻAlalā, named Hōʻike i ka pono (Figure S1, studbook #32). This individual was chosen for genome sequencing because (1) His high inbreeding coefficient (0.25) would allow for simplified genome assembly; and (2) He is a great-grandson of the two genetic founders that constitute approximately 45% of the ancestry in the captive population (i.e., his genome would be a good representation for most birds in the population [22]). Note that all procedures on live animals were approved by the Institutional Animal Care and Use Committee (IACUC) of San Diego Zoo Global (15-012 and 16-009). Library construction protocol followed the workflow for ultra-large insert libraries [26]. The DNA was sheared to target 50 kb fragments (resulting distribution 30–80 kb) by using a Megaruptor (Diagenode, Denville, NJ, USA), and assessed for quality by pulsed-field gel electrophoresis (PFGE) on the CHEF Mapper system (Bio-Rad, Hercules, CA, USA). A total of 86 µg of DNA were then recovered from the 50 kb shearing condition. Sheared DNA was constructed into SMRTbell templates (PacBio, Menlo Park, CA, USA) by following the >30 kb library construction protocol [26] with minor modifications (e.g., 1× AMPure PB purification (Beckman Coulter, Brea, CA, USA); room temperature rotation instead of vortexing; two-step elution process during AMPure PB elution to maximize recovery). Final SMRTbell library qualities were assessed by PFGE and Pippin Pulse (Sage Science, Beverly, MA, USA) to determine the optimal size-selection cut-off of 20 kb. Size selection was done using the BluePippin system (Sage Science), with targeted exclusion of small fragments (<20 kb) that would otherwise preferentially load during sequencing. Following size selection, the library fragments had a mode size of approximately 30 kb and comprised approximately 8.6 µg of DNA; enough to sequence 133 single-molecule, real-time (SMRT) cells at Pacific Biosciences (PacBio). Sequence data were generated using the PacBio RSII instrument with P6v2 polymerase binding, C4 chemistry kits (P6-C4) and 6 h run time movies, which yielded 9,859,413 reads, totaling 128,622,819,749 bp whole-genome sequence data. The average read length was 13,045 bp (max = 78,477 bp; standard deviation = 8972 bp). Reads less than 500 bp (1.7% of reads; <1% of total bases) were removed. Post-filtering, the N50 subread length was 18,661 bp.

2.2. Genome Assembly and Quality

De novo assembly followed the PacBio string graph assembler process, using FALCON and FALCON-Unzip [27] to generate long-range phased haplotypes. During the assembly process, sequence reads were overlapped to form long consensus sequences [6,27]. These longer reads were used to generate a string graph, and the graph was reduced so that multiple edges formed by heterozygous structural variation were replaced to represent a single haplotype [28]. Primary contigs were formed by using the sequences of non-branching paths, while associated contigs (i.e., haplotigs) represent the sequences of branching paths. The resulting assembly thus represents a phased diploid genome [27,29]. Primary and secondary genome assemblies are available on GenBank, Accession: QORP00000000. Raw reads are available at https://www.ncbi.nlm.nih.gov/sra/SRP151284.

To assess the quality of the final assembly, we compared the number and length of ʻAlalā contigs to those of other avian assemblies. In addition, we used BUSCO v2.0.1 [30] to assess the completeness of the gene space in the ʻAlalā assembly based on the detection of conserved single-copy orthologs. For comparison, we included genome assemblies of the domestic chicken (*Gallus gallus*, GenBank accession GCF_000002315.4 [31]), Anna's hummingbird (*Calypte anna*, GCA_002021895.1 [29]), zebra finch (*Taenopygia guttata*, GCF_000151805.1 [32]), hooded crow (*C. cornix cornix*, GCF_000738735.1 [33]), and American crow (*C. brachyrhynchos*, GCF_000691975.1 [34]). As lineage datasets, we chose eukaryota_odb9 (303 genes) and a 250-gene eukaryotic subset [35], which is highly congruent with the core eukaryotic genes mapping approach (CEGMA) dataset [36]. Gene finding parameters in the AUGUSTUS analysis step were based on the chicken genome.

2.3. Repeat Composition Analysis

To identify mobile and repetitive DNA in the ʻAlalā assembly, we generated a *de novo* repeat library using RepeatModeler v1.0.11 [37]. This software package primarily integrates RECON v1.08 [38] and RepeatScout v1.0.5 [39] to find interspersed repeats. Repeat models with 50% sequence identity over at least half their length to Swiss-Prot entries with known function were removed from the library, and remaining models were assigned to repeat classes by reference to Repbase [40]. Additional, more detailed repeat classification was performed with CENSOR [41]. The ʻAlalā assembly was then screened for repetitive DNA using RepeatMasker v4.0.7 [37] based on RMBlast and two repeat libraries: (1) The ʻAlalā repeat library described above; and (2) An expanded library also containing all chicken and ancestral eukaryotic repeats, as well as all zebra finch repeats provided by Repbase. In addition to the version implemented in RepeatMasker, simple repeats were assessed using the stand alone version of Tandem Repeats Finder v4.0.9 [42] with the following settings: Match = 2, Mismatching penalty = 7, Delta = 7, PM = 80, PI = 10, Minscore = 50, and MaxPeriod = 2000. Along with the ʻAlalā assembly, we also analyzed the assemblies of the domestic chicken, Anna's hummingbird zebra finch, hooded crow, and American crow listed above.

2.4. Candidate Gene Annotation and Analysis

We focused on annotating particular genes associated with adaptive and innate immunity, as diversity at such genes is predicted to be especially relevant to fitness. Specifically, we were interested in genes of the major histocompatibility complex class II beta (MHC class II B) and toll-like receptor (TLR) genes. To identify candidate immunity genes in the primary ʻAlalā assembly, we first performed Blast (tblastn) searches using homologous protein sequences of other bird species as queries, with an e-value cut-off of $1 \times 10e^{-5}$. For MHC, queries were obtained from the zebra finch, the currently best annotated passerine genome [43] (Table S1). For the more conserved TLRs, the full gene repertoire of the domestic chicken was used (Table S1). In the absence of transcriptional evidence, individual ʻAlalā genes were located by comparing genomic coordinates of high-scoring segment pairs on each contig, which often corresponded to exons. Genomic sequence including 1500 bp up- and downstream of each putative gene was extracted, and the gene structure and coding sequence were predicted by the AUGUSTUS web server v3.3 [44]. In the case of MHC class II B, only putative genes including exons 2 and 3 were considered for this step, due to a large number of single-exon or fragmentary hits. Portions of nucleotides that appeared to be missing from the predicted coding sequence were identified by aligning the predicted sequence to the reference using MAFFT v7 [45]. Using short sequence motifs taken from the reference, we then attempted to find missing homologous parts by translating the genomic sequence into all three reading frames in the coding direction by using EMBOSS Transeq [46]. Finally, manually completed gene predictions were tentatively classified as functional, ambiguous or pseudogenized, depending on the integrity and length of the reading frame. Genes for which a complete reading frame, including start and stop codons, could be identified were considered functional, while genes that required the insertion or deletion of a single nucleotide to recover the complete reading frame (suggestive of a sequencing error) were categorized as ambiguous. Fragmentary reading frames or multiple frameshift mutations were regarded as indicative of pseudogenes. Untranslated 5′ and 3′ regions could not be annotated due to the lack of transcriptional evidence. MHC class II B Blast searches were later repeated to assess the number and divergence of gene fragments with increased sensitivity. Exons 2 and 3 of the ʻAlalā gene, Coha_MHCIIB_b (see Results), were used as query sequences.

To shed light on the evolutionary history of the gene family, we performed a phylogenetic analysis based on exon 2 of MHC class II B genes in ʻAlalā and other corvids. We included all functional and ambiguous ʻAlalā genes, as well as complete MHC class II B gene sets of single individuals, each of American crow (13 sequences), the jungle crow (*C. macrorhynchos japonensis*; 14 sequences), the Asian rook (*C. frugilegus*; 11 sequences) and the azure-winged magpie (*Cyanopica cyanus*, 7 sequences), to allow for within-genome diversity comparisons. These data were obtained by [47] using a targeted polymerase chain reaction (PCR) approach. Exon 2 nucleotide sequences were aligned with MAFFT

v7, and 10 maximum likelihood trees computed under the GTRCAT model in RAxML v8.1.20 [48]. Confidence values were estimated from 500 rapid bootstrap replicates and drawn onto the best maximum likelihood tree (-f a algorithm).

2.5. MHC Functional Supertypes

To assess how similar the ʻAlalā MHC class II B repertoire might be to other corvids in terms of properties of the antigen-binding regions, we relied on comparisons of functional supertypes (e.g., [49,50]). Briefly, functional supertype analysis involves identifying codons under selection (positively-selected sites; PSS), and then grouping sequences according to descriptor variables that reflect the physical and chemical properties of the amino acids at these selected positions [51–53]. As the ʻAlalā MHC class II B sequences from this study were generated from a single individual, we based our analysis on the locations of nine PSS shared among three other crow species (jungle crow, carrion crow (*C. corone orientalis*), and American crow), which were previously identified through HYPHY analysis [54]. First, we used MUSCLE [55,56] implemented in Geneious vR10 [57] to align our putatively functional ʻAlalā exon 2 sequences to the 237 sequences from Eimes et al. [54], for a total of 244 nucleotide sequences. We then trimmed all sequences to exclude non-PSS codons, translated them, removed duplicate sequences (sequences remaining: 153), and converted the information into a matrix of five physiochemical descriptor variables that reflect the physical and chemical properties of each amino acid [51–53]: z1 (hydrophobicity), z2 (steric bulk), z3 (polarity), z4 and z5 (electronic effects). Using the matrix of z-descriptors, we performed *k*-means clustering with the adegenet package in R [58] to reveal clusters of sequences likely to have similar functional properties. We then used discriminant analysis of principle components (DAPC) to describe the clusters [59].

2.6. Runs of Homozygosity

Runs of homozygosity (ROH) are stretches of identical haplotypes that occur across homologous chromosomes within the same individual. The length of ROH segments within an individual's genome depends on whether shared ancestry is recent or ancient; recent inbreeding results in relatively long ROH segments, because recombination has not yet broken up the segments that are identical by descent [60]. As mutations accumulate over time, the ROH segments further break down. We assessed ROH in the ʻAlalā genome for three purposes: (1) To estimate the autozygous fraction of the genome fROH [25], i.e., the total fraction of the genome that is perfectly autozygous (zero heterozygosity); (2) To evaluate the effect of allowance for low levels of heterozygosity on estimates of whole-genome fROH and ROH segment lengths; and (3) To estimate fROH on a per-contig basis. For comparison to an outbred species, the fROH and ROH segment length analyses were also conducted using an Anna's hummingbird genome that was sequenced and assembled similarly to the ʻAlalā genome assembly [29]. The ROH segment lengths were calculated by summing consecutive sliding windows that met criteria of perfect autozygosity or fell within the one of four heterozygosity thresholds: 1 SNP per 100 kb, 50 kb, 25 kb or 10 kb (i.e., $\leq$0.01, $\leq$0.02, $\leq$0.04, and $\leq$0.1 SNPs/kb). The fROH was calculated by taking the sum of ROH segment lengths and dividing this number by the cumulative length of all sliding windows. The sliding windows, set to 100 kb, were calculated with vcftools [61] using SNPs called from the primary contigs. Only contigs $\geq$500 kb were included in this analysis, permitting a minimum of five consecutive sliding windows to assess ROH segment lengths. As sliding window analysis incorporates rounding, the cumulative length of sliding windows exceeds the cumulative length of contigs.

3. Results

3.1. Genome Assembly and Quality

The FALCON assembler generated a 1.06 Gbp primary assembly with a contig N50 of 7,737,654 bp across 671 total contigs (Table 1). The diploid assembly process produced 2082 associated haplotype

contigs (haplotigs) with an estimated length of 0.43 Gb and contig N50 of 455,082 bp (Table 1), implying that about 40% of the genome contained sufficient heterozygosity to be phased into haplotypes by FALCON-Unzip. For comparison, the same assembly process suggested that 75% and 100% of the genomes of two more outbred species, zebra finch and Anna's hummingbird, contained sufficient heterozygosity to be phased into haplotypes [29] (Table 1). Compared to other published short-read based avian genomes of similar size, the ʻAlalā assembly represents a dramatic decrease in assembly fragmentation, with substantially fewer and longer contigs, and is similar in quality to other long-read *de novo* assemblies (Figure 1; Table S2). The BUSCO analysis indicated that gene completeness was among the highest of any avian genome to date (Figure 1 and Table S2). Collectively, these results suggest that this ʻAlalā long-read genome assembly is one of the highest quality avian genomes currently available.

Table 1. *De novo* long-read genome assembly statistics comparing PacBio-based primary and secondary haplotypes in three avian species.

Species	PacBio-Based Primary Haplotype	PacBio-Based Secondary Haplotype
ʻAlalā (this study)		
Number of contigs	671	2082
Contig N50	7,737,654 bp	455,082 bp
Total size	1,064,991,496 bp	432,637,353 bp
Zebra finch [29]		
Number of contigs	1159	2188
Contig N50	5,807,022 bp	2,740,176 bp
Total size	1,138,770,338 bp	843,915,757 bp
Anna's hummingbird [29]		
Number of contigs	1076	4895
Contig N50	5,366,327 bp	1,073,631 bp
Total size	1,007,374,986 bp	1,013,746,550 bp

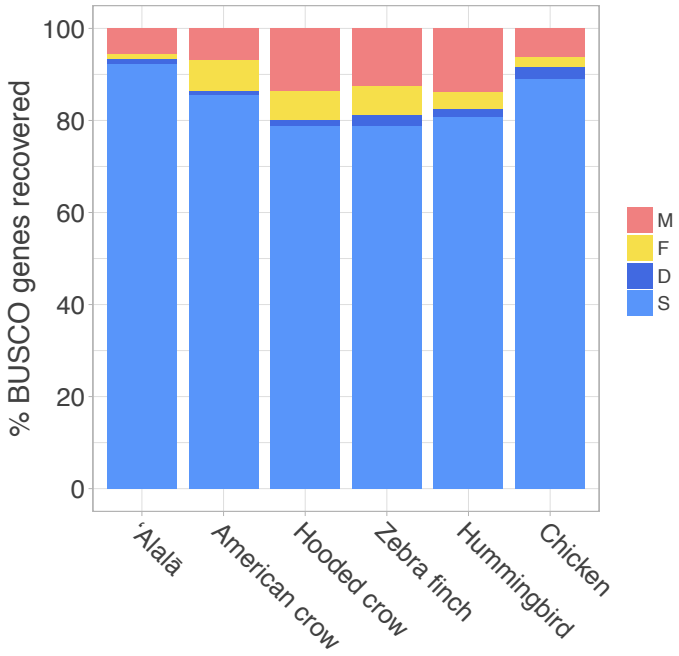

Figure 1. Genome assembly completeness assessed by the recovery of universal single-copy genes (BUSCOs). Percentages refer to complete genes that were found as single (S) or multiple copies (D), as well as fragmented (F) and missing (M) genes. Analyses were based on the BUSCO eukaryote dataset (n = 303 genes).

3.2. Mobile and Repetitive Elements

De novo repeat modeling resulted in an ʻAlalā-specific repeat library containing 260 families, including 50 LINE (long interspersed element) and 23 LTR (long terminal repeat) families. Only 12% of these had matching entries in Repbase, mostly to endogenous retroviruses (ERVs) and CR1 retrotransposons previously identified in other passerine birds. In addition, several ʻAlalā repeat families were partially similar to the large tandem repeat ʻcrowSat1', a 14 kb satellite that is suspected to be a major heterochromatin component in the hooded crow [62]. In contrast, extended matches to Swiss-Prot entries [63], which might indicate co-opted transposable elements [64], were not discovered. RepeatMasker screening identified 10.1% of the ʻAlalā assembly as mobile or repetitive sequence, including 3.3% LINEs (exclusively of the CR1 class), 1.1% LTRs (various endogenous retroviruses), and 4.5% unclassified interspersed repeats. The remaining 1.2% was made up of simple repeats and low complexity sequence, including satellites homologous to crowSat1. This estimate did not change noticeably when using a repeat library expanded with avian and ancestral eukaryotic repeats provided by Repbase. The stand-alone analysis of Tandem Repeats Finder revealed 303,030 tandem repeats with a maximum unit length of 2000 bp, making up 6.9% of the assembly (max. repeat size was ~100 kb). This fraction is substantially lower than in the domestic chicken (16.1%), but higher than in the zebra finch (3.7%), Anna's hummingbird (3.1%), the American crow (2.8%), and the hooded crow (1.8%). However, these results might partially reflect differences in assembly completeness and contiguity, which affect repeat identification (e.g., the American crow and hooded crow genome assemblies are short-read based; the Anna's hummingbird genome was generated and assembled using a similar process as for the ʻAlalā). When combined, the RepeatMasker and Tandem Repeats Finder estimates suggest that the ʻAlalā assembly contains approximately 15% repetitive DNA. Since heterochromatic regions often cannot be assembled reliably, this is likely an underestimate of the true genome repeat content.

3.3. Candidate Gene Annotation and Analysis

We identified eleven (including duplicates) TLR genes in the ʻAlalā genome (Table S3). Sequence and gene structure, which could be reliably assessed with only a single untranslated 5′ exon missing from one prediction, were highly conserved with regard to the chicken reference [65] and other birds [66]. All genes were classified as functional based on possessing complete or nearly complete open reading frames. Notable features relative to the reference include the loss of 140 amino acids at the 5′ end in *TLR1A*, a ~50 amino acid indel in *TLR2B*, and a tandem arrangement of two *TLR7* copies differing by 16 amino acids. This duplication has previously been observed in other passerine birds [66–70] and was annotated here as *TLR7a* and *TLR7b*.

The MHC class II B repertoire of the ʻAlalā proved to be more complex. We identified seven presumably functional and two ambiguous, but in all likelihood equally functional, genes with open reading frames across all five expected exons in the primary assembly. Additionally, we discovered three genes with incomplete or disrupted reading frames comprising exons 2 and 3 (Table S4). Most of these genes were located in tandem arrays of 20–40 kb (b-c-d, e-f-g, and h-i-p1). Uniform read coverage, small but detectable differences in flanking regions, and reads spanning multiple genes suggest that these genes represent individual loci, rather than alleles or assembly artifacts. Overall, the ʻAlalā MHC class II B genes proved largely conserved compared to the zebra finch reference (sequence identity > 80% at the amino acid level), with the highest variability found in exon 2 as expected. In addition to the three more complete putative pseudogenes above, the assembly also contains a large number of MHC class II B fragments. More than 130 sequences homologous to exon 2, and about 30 homologous of exon 3 (containing the immunoglobulin C1-set domain) were found scattered throughout the primary assembly, usually in the form of tandem arrays of 2–5 copies (Table S5). According to phylogenetic analysis, six of the functional and ambiguous ʻAlalā MHC class II B genes comprise a strongly supported clade (Figure S2). These copies only differ by 0–3 amino acids, which were found at or very close to the positively selected sites (PSS) described for other corvids [54].

In contrast, only three genes—highly similar copies 'a' and 'd', as well as copy 'g'—were placed in other locations of the phylogenetic tree, albeit with lower support.

3.4. MHC Functional Supertypes

Similarity of the 'Alalā MHC class II B repertoire in terms of functional antigen-binding properties to three other corvids was assessed on the basis of nine PSS [54]. Focusing on these PSS in 244 nucleotide sequences, we identified 153 unique amino acid variants. From these, we identified eight functional supertypes (Figure 2), consistent with [54]. Three of the eight supertypes corresponded to 'Alalā and were also shared by the three other corvid species that were compared (Figure 3). No 'Alalā supertypes were discovered to be separate from other corvids. It must be noted, however, that sequences from only one 'Alalā were used (compared to 4–6 individuals for each of the other species included in the supertype analysis). Our purpose in performing this supertype analysis was not to fully characterize the repertoire of MHC class II B diversity or functionality in 'Alalā, but rather to establish a sense of how similar (or different) these *might* be compared to related species. The information gained has been used to design primers for targeted-amplicon approaches that are now being used to better assess 'Alalā MHC class II B diversity at the population level.

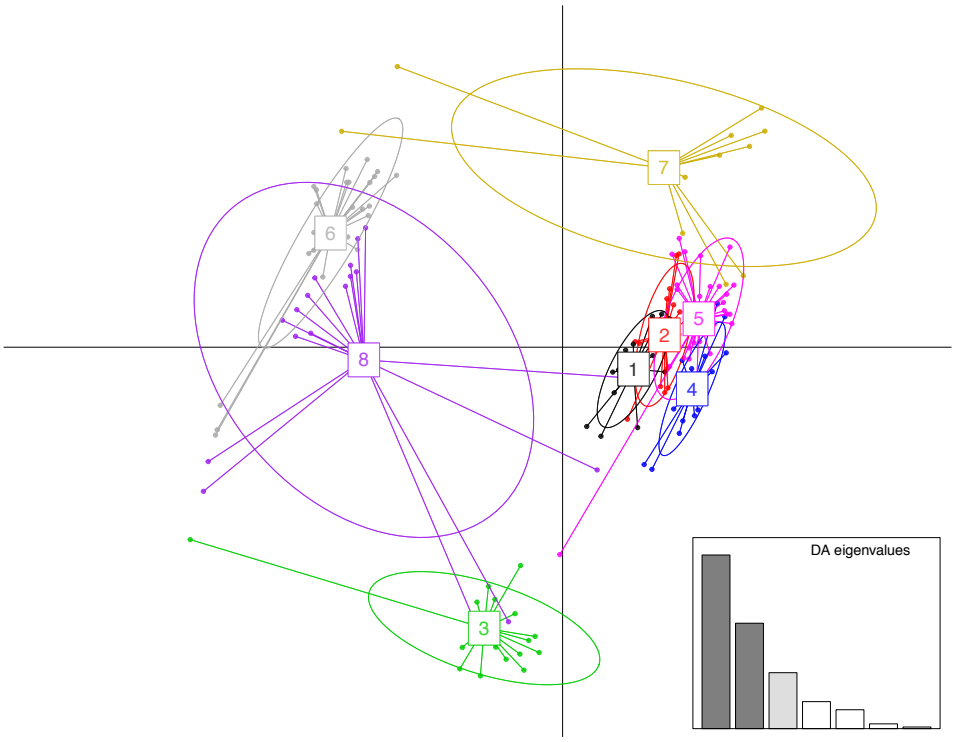

Figure 2. Discriminant Analysis of Principle Components (DAPC) scatterplot of the 8 major histocompatibility complex (MHC) supertypes identified here. 10 principle components (PC) and three discriminant functions (dimensions) were used to describe the relationship between the clusters. The scatterplot show only the first two discriminant functions (d = 2). The bottom graph displays the barplot of eigenvalues for the discriminant analysis (DA). Dark grey, light grey and white bars indicate eigenvalues that were used in the scatterplot, not used in the scatterplot but retained for the analysis, and not retained for the analysis, respectively. Each allele is represented as a dot, and the supertypes as ellipses.

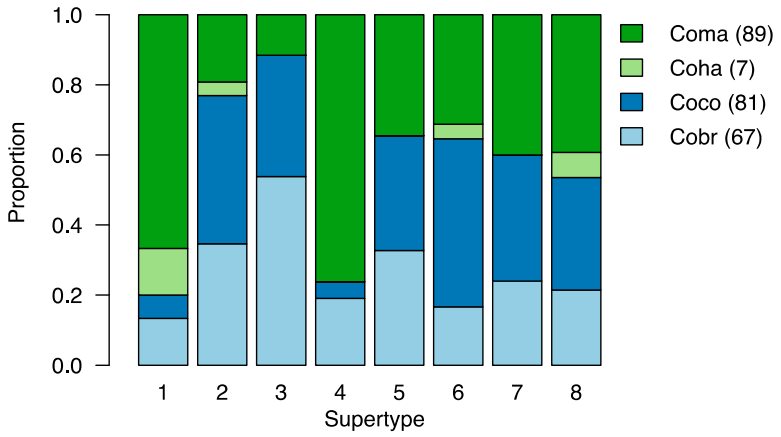

Figure 3. Stacked barplot indicating the representation of each corvid species within each MHC supertype identified here. The three other corvid species were represented across all eight supertypes, while the ʻAlalā was represented by three supertypes. Note, however, that the ʻAlalā data were established from a single individual, while the other species' data represent 4–6 individuals per species [54]. In the legend: Coma = jungle crow; Coha = ʻAlalā; Coco = carrion crow; Cobr = American crow; Numbers in brackets indicate the number of nucleotide sequences included in the analysis.

3.5. Runs of Homozygosity

A total of 413,114 SNPs were detected across the 209 ʻAlalā genome contigs >500 kb in length (1.02 Gb of sequence data, 96% of the genome). Based on sliding windows intervals of 100 kb, the fraction of the genome that was perfectly autozygous (i.e., fROH) was approximately 5.6% (574 of 10,338 sliding windows were completely homozygous, Table S6). Allowing for ROH to include low level heterozygosity resulted in a substantial increase in the number of ROH segments and fROH. For example, with a heterozygosity threshold of $\leq$1 SNP per 50 kb, the number of sliding windows counted as ROH increased to 2496 and fROH increased to 46.1% (Table S6). With regard to ROH segment lengths, the ROH tended to be short under the criterion of perfect autozygosity, but increased in length as the allowable heterozygosity threshold was raised. In terms of individual contigs, the proportion of perfectly autozygous ROH relative to all sliding windows was highly variable and ranged from a minimum of 0% to a maximum of 63% (median 2.7%; average 5.7%). For the hummingbird, 1,841,031 SNPs were detected across 283 genome contigs >500 kb in length (0.92 Gb of sequence data, 92% of the genome). The fraction of the genome that was perfectly autozygous was 4.4% (416 of 9373 sliding windows, Table S6), fairly similar to that of the ʻAlalā. However, in contrast to the ʻAlalā, allowing for low levels of heterozygosity had little impact on the number of ROH segments and fROH. For example, raising the allowable heterozygosity threshold to $\leq$1 SNP per 50 kb only increased the number of sliding windows counted as ROH to 557 and fROH to 5.9%. The ʻAlalā's sensitivity to allowing for low-levels of heterozygosity in ROH calculations is explained by a general trend for low genetic diversity across most of its genome, with median and mean values of 0.05 and 0.40 SNP/kb across the 100 kb sliding windows. For comparison, hummingbird median and mean values were 1.93 and 1.96 SNP/kb, which explains why the heterozygosity threshold had little effect on ROH and fROH calculations.

4. Discussion

Using PacBio SMRT sequencing technology and FALCON assembly, we generated a high-quality, long-read *de novo* genome assembly for one of the world's most endangered birds. During the assembly process, FALCON stipulates that if overlapping regions differ by $\geq$5% over extensive distances then the assembler separates the regions into primary and associated (secondary) contigs [29]. By definition, regions of the primary assembly that have corresponding associated contigs identify areas in the

genome with relatively high heterozygosity. The genome assembly of a single ʻAlalā (studbook #32, Figure S1) highlights the genomic signatures of small population size and inbreeding, because the ʻAlalā associated contigs corresponded to a substantially smaller proportion of the genome compared to more outbred species.

4.1. Candidate Immunity Genes

Toll-like receptors are an integral component of the innate immune system in animals. Recognizing pathogen-associated molecular patterns (PAMPs) at the cell surface, their role consists of activating the organism's inflammatory response through a cascade of intracellular signals. Because variation in TLRs is associated with resistance or susceptibility to infectious diseases, the gene family is especially relevant regarding fitness and inbreeding-related conservation matters. In the ʻAlalā, we discovered a full complement of functional TLR genes (Table S3), with high conservation of gene number, structure, and sequence in comparison to other birds [62]. Although only a single *TLR7* copy exists in most other birds with annotated TLR repertoires, including the zebra and house finch, duplicates have been reported in several passerine species [66–69,71], indicating that duplication likely predates the split of the corvid family from other passerines. We did not find evidence that *TLR5* was pseudogenized in ʻAlalā, as it is in some passerine species [72]. However, as this result is based on a single individual it should be taken with caution.

In contrast to TLRs, the MHC, which activates the adaptive branch of the vertebrate immune system, includes some of the most variable genes found in vertebrates. MHC class II B genes, on which we focused in this study, encode proteins that can bind and present a large range of pathogen-derived peptides. Studying MHC diversity, particularly of the peptide-binding domain located on exon 2 (class II histocompatibility antigen B domain), can therefore be highly relevant for the conservation of endangered species. The number of functional MHC class II B genes we identified in the ʻAlalā (~7–9; Table S4) places it within the range known from other corvids (7–20 alleles per individual were described in the American crow, jungle crow, and carrion crow [54]). The MHC class II B diversity in the assembly appears to be low, with six out of nine putatively functional genes being almost identical at the amino acid level. Although we cannot completely rule out that some of these genes represent allelic variants of each other, evidence obtained from read coverage and flanking regions suggest that all genes are derived from genuine and separate loci. Several recent episodes of gene duplication may account for the pattern of high similarity, possibly including segmental duplications that gave rise to the observed tandem arrays with relatively high sequence homology up- and downstream of the MHC copies. According to our phylogenetic analysis, these events must have occurred after the divergence of the ʻAlalā from the other represented corvids (Figure S2). Alternatively, the true evolutionary history of the MHC class II B gene family may have become obscured by gene conversion, which can homogenize gene copies within a species, and has been implicated in other birds [73,74]. Intronic and exon 3 sequences from additional corvids and ʻAlalā individuals could shed light on this issue in the future (e.g., [73]). The three remaining ʻAlalā MHC class II B genes that did not fall within the main cluster, on the other hand, may represent remnants of evolutionarily distinct gene family lineages. However, a lack of support along the backbone of the phylogenetic tree prevented the identification of clear orthologs within other species. The phylogenetic analysis and previous studies [47,54] also revealed multiple MHC class II B lineages shared by other corvids, suggesting that the gene family expanded prior to the radiation of the corvid family. Notably, no ʻAlalā genes were found within several of these lineages consisting of genes from all or almost all other corvids investigated here (Figure S2). In all likelihood, these genes were lost in the evolutionary lineage leading to the ʻAlalā, or even very recently in the species' population history. This hypothesis is supported by the high number of gene fragments and high sequence similarity between copy 'g' and pseudogene 'p1', suggesting at least one evolutionarily recent pseudogenization event. More generally, the large number of putative MHC class II B pseudogenes (Table S5) is consistent with expectations for passerines (e.g., [75]), and the evolution of a gene family characterized by repeated gene gain and loss. Most of these pseudogenes

appeared to be highly fragmented, i.e., homology could only be established over a short length of exon 2 or 3 (150 bp or less). Sequence identity to the functional Coha_MHCIIB_b copy fluctuated widely, ranging from near perfect matches to less than 40% at the amino acid level, suggesting a broad age distribution with regard to the time of pseudogene origin. This might be a reflection of the dynamic evolution of this hyper-variable gene family, which included repeated expansions and contractions over evolutionary timescales [76]. The localization of most fragments on tandem arrays with high sequence homology in adjacent regions (alignment of 20 randomly chosen pseudogenes ±250 bp up- and downstream) also suggests that frequent segmental duplication events contributed to the abundance of MHC class II B pseudogenes. What remains unclear is why most pseudogenes seem limited to fragments of exons 2 or 3, rather than full-length genes including more conserved exons. A few fragments, especially those on shorter contigs, may represent assembly artifacts. Another possibility is that some are functional parts of other genes that originated by exon-shuffling. Further improvements to the assembly and a comprehensive annotation of the entire ʻAlalā gene content may bring more clarity on this matter in the future.

In summary, the present long-read ʻAlalā genome assembly includes more complete gene sequences than are available for many avian genomes, a crucial factor for annotating complex genomic regions, such as the MHC. While the observations described here are based on only a single individual, and should thus be interpreted with caution, our results imply that MHC class II B diversity in the ʻAlalā is likely to be somewhat similar by comparison to other corvids. Additionally, our functional supertype analysis suggests that while nucleotide sequences may differ between ʻAlalā and other corvids, similarity exists among species when it comes to pathogen-binding properties. However, genome data from additional specimens are required to gauge within-species diversity and distribution of different MHC class II B lineages, which would further place these data into the context of corvid and ʻAlalā evolution.

4.2. Runs of Homozygosity

If SNPs were evenly distributed across the genome, the SNP encounter rate in ʻAlalā would be 1 SNP per 2477 bp (413,114 SNPs identified from 1.02 Gb of sequence data). This value is considerably lower than empirical estimates obtained from genomes of other avian species, for example, 1 SNP per 330 bp in *Ficedula* flycatchers [77], 1 SNP per 256 bp in Hawai ʻi Amakihi (*Hemignathus virens* [78]), 1 SNP per 935 bp in turkey (*Meleagris* spp. [79]), and 1 SNP per 501 bp in Anna's hummingbird [29] (based on contigs ≥500 kb). While caution should be used when making comparisons between genomes that differ by sequencing technologies, genome assembly pipelines, and other computational settings (addressed in more detail below), the paucity of SNPs in the ʻAlalā genome is not surprising because of the overall low population size of ʻAlalā and this particular bird's high pedigree inbreeding coefficient (0.25). By comparison to the examples noted here, the ʻAlalā genome was generated and assembled in a similar fashion to that of the Anna's hummingbird, the latter of which had a SNP encounter rate almost five times as frequent. Certainly, the presence of contigs showing very low heterozygosity in the ʻAlalā is consistent with empirical observations made of ROHs in turkey [79] and large stretches of very low heterozygosity in Hawai ʻi Amakihi [78]. The contrast between highly variable sliding windows and regions with modest variability suggests that the PacBio assembly pipeline used here is sensitive to calling SNPs across a range of heterozygosities, and that low diversity observed for this genome is not solely an artifact of the assembly pipeline.

The lengths of ROH are an indication of shared ancestry and can be used to gauge whether inbreeding events occurred within recent or distant generations [60]. Recombination events break long autozygous segments into smaller pieces, thus numerous short ROH are consistent with distant shared ancestry. In this ʻAlalā genome, the numerous short ROH collapse into much longer segments when low levels of heterozygosity are allowed. This indicates both mutations and recombination have a strong impact on ROH measures in the ʻAlalā. In contrast, the hummingbird had far fewer ROH, and relaxing the heterozygosity threshold had relatively less impact on ROH numbers and lengths. Thus,

recombination, not mutations, was the dominant force in ROH segment lengths in the hummingbird. This comparison between two species that differ by demographic histories highlights the sensitivity of ROH to patterns of mutations in the genome, along with SNP filtering criteria. Moreover, the rate of sequencing error (related to depth of sequencing and sequencing platform) will also affect the estimations of homozygosity and, correspondingly, the length of ROH segments.

Several factors confound comparability of ROH across studies and taxa. These include: Lack of consensus definition for ROH; differences in sequencing platforms and associated sequencing errors; variant-calling pipeline; and computational settings (e.g., [80–82]). The ROH estimates in this study are drawn from a single genome with high depth of sequencing. In contrast, measures of ROH can be obtained from high-density SNP arrays by quantifying the length of rows of homozygous SNPs relative to a reference, the results of which are sensitive to SNP chip density, and may miss unmeasured variants between the markers [80,81]. The density of SNP markers across the genome also impacts what can be reliably detected as a minimum length ROH. For example, in humans, a panel of 3 million SNP markers identified ROH as short as 100 kb, and by applying this denser SNP panel to the Han Chinese population, their estimated total ROH increased from 130 Mb, measured with a 0.4 million SNP panel, to 510 Mb [60]. Comparability of ROH results between species can be further diminished by biological variability in chromosome lengths. Birds, with numerous microchromosomes, are expected to have shorter ROH than mammals, simply because of differences in chromosome lengths.

4.3. Applications to ʻAlalā Conservation

The ʻAlalā genome assembly resulting from long-read sequencing data provides a high-quality reference genome that will enable downstream comparative, population, and conservation applications. Prior to this study, molecular work using limited genetic markers identified low diversity in the species [23,24], and pedigree analysis identified inbreeding effects on hatching success, as well as skewed founder representation [18,22]. Thus, we identified a need to generate genomic resources, particularly to inform strategic pairings in an effort to slow the rate of inbreeding (and increase the population growth rate) and to preserve remaining genetic diversity in order to maximize the likelihood that the species will be able to adapt to environmental changes. Work is also ongoing to better understand the fitness associations of particular candidate genes; for example whether these, in addition to genome-wide diversity, are linked to hatching success. Similarly, identifying the genetic basis of other fitness-related phenotypes is a critical research goal that could be incorporated with strategic pairings. For example, the California condor (*Gymnogyps californianus*) has a relatively high frequency of heritable, embryonic lethal dwarfism, and genomics are currently being applied to identify carriers of the disease in an effort to eliminate it from managed populations [83–87]. Using genomic information, a similar strategy could be applied to managing maladaptive phenotypes in ʻAlalā.

Genomic data derived from our analyses are an essential component of the current and future recovery of the ʻAlalā. The use of genomic information to assist strategic pairing and minimize inbreeding is the most practical and immediate application of genomics at this time, and is predicted to have a positive impact on population growth (e.g., [22], but also see [88]). In conjunction with ongoing conservation-breeding activities, a reintroduction program was recently initiated in an effort to re-establish this formerly extinct-in-the-wild species into its native forest habitat. Early indications of the reintroduction effort are promising, with a small population of recently released individuals surviving in the wild at the time of writing. Because the entire extant ʻAlalā population, both in captivity and the wild, is still relatively small, ongoing management decisions for the breeding and release of particular individuals will have implications for the long-term recovery of the species. As the size of both the captive and wild ʻAlalā populations continue to increase, the integration of genomic data as part of the conservation management effort will help to maximize the genetic health of the species well into the future.

Supplementary Materials: The following are available online at http://www.mdpi.com/2073-4425/9/8/393/s1. References [89–97] are cited in the Supplementary Materials. Table S1. Accessions of references used to annotate candidate immunity genes. Table S2. ʻAlalā assembly results compared to other avian genome assemblies. Statistics are given for contigs, and for scaffolds in parentheses. Table S3. ʻAlalā toll-like receptor (TLR) genes, following the nomenclature suggested by [65]. Start and stop refer to the position of the start and stop codons on the contig. Exons indicates the number of exons identified in the absence of transcriptional evidence, with the number of exons in the *G. gallus* reference given in parentheses. All predicted genes appear to be complete, suggesting they are functional (F). Notes are provided with respect to the *G. gallus* reference, where applicable. aa: amino acids. Table S5. Homologs of MHC class II B exons 2 and 3 in the ʻAlalā genome. Identity (% id), mismatches (Mism.), gaps, query start and end positions (Q-start, Q-end), hit start and end positions (H-start, H-end), e-value and score refer to Blast results using Coha_MHCIIB_b exons 2 and 3 as query. Bit scores > 100 are highlighted in yellow. Consecutive hits on the same contig within 1500 bp (in green blocks) were annotated further as MHC class II B candidate genes (see Table S3). The remaining hits are fragments that may represent, pseudogenized copies. Table S6. Analysis of runs of homozygosity (ROH) for Alala (*Corvus hawaiiensis*, Coha) and Anna's hummingbird (*Calypte anna*, Caan) [29] genome contigs of minimum length 500 kb (n = 209 and n = 283, respectively). The ROH segment length is the cumulative length of uninterrupted 100 kb sliding windows (SW) that were completely autozygous or fell within one of four heterozygosity thresholds. The autozygous fraction of the genome (fROH) is calculated as the total number of sliding windows in ROHs (# SW-ROH) divided by the total number of sliding windows (n = 10,338 Coha; n = 9373 Caan) across all genome contigs. Figure S1. ʻAlalā pedigree. The sequenced individual, studbook 32 (named Hōʻike i ka pono) is shaded. Dashed lines indicate individuals that are represented in multiple positions in the pedigree (e.g., overlapping generations). Figure S2. Unrooted maximum likelihood tree of corvid MHC class II B genes, exon 2. Functional genes predicted from the ʻAlalā assembly are highlighted in red (prefix Coha). Other species represented include *C. brachyrhynchos* (American crow, Cobr), *C. macrorhynchos* (jungle crow, Coma), *C. frugilegus* (Asian rook, Cofr) and *Cyanopica cyanus* (azure-winged magpie, Cycy). All genes were obtained from one individual per species [47]. Confidence values are given for nodes with bootstrap support >70%.

Author Contributions: Conceptualization, J.T.S., C.C.S., J.G., B.M.M. and O.R.; Data curation, R.H., P.B. and J.M.; Formal analysis, J.T.S., M.H., C.C.S., M.R.B., J.K. and S.K.; Methodology, J.K., R.H., P.B. and J.M.; Writing–original draft, J.T.S., M.H., C.C.S., M.R.B., J.K., B.M.M. and O.A.R.

Funding: Funding for the conservation-breeding and reintroduction efforts has been provided by the San Diego Zoo Global, U.S. Fish and Wildlife Service, Hawaiʻi Division of Forestry and Wildlife, National Fish and Wildlife Foundation, the Max and Yetta Karasik Foundation, the Moore Family Foundation, American Forests, and numerous anonymous donors. The material presented here is partially based upon work supported by the National Science Foundation under Grant No. 1345247. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Acknowledgments: Many thanks to the hard-working staff, interns, and volunteers who care for and propagate ʻAlalā at the facilities on Maui (Maui Bird Conservation Center) and Hawaiʻi Island (Keauhou Bird Conservation Center). We thank three anonymous reviewers for their constructive and insightful feedback.

Conflicts of Interest: J.K., R.H., P.B., J.M., J.G. and S.K. are full-time employees at Pacific Biosciences, a company developing single-molecule sequencing technologies. All other authors declare that they have no competing financial interests.

References

1. Ouborg, N.J.; Pertoldi, C.; Loeschcke, V.; Bijlsma, R.K.; Hedrick, P.W. Conservation genetics in transition to conservation genomics. *Trends Genet.* **2010**, *26*, 177–187. [CrossRef] [PubMed]
2. Allendorf, F.W.; Hohenlohe, P.A.; Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **2010**, *11*, 697–709. [CrossRef] [PubMed]
3. Grueber, C.E. Comparative genomics for biodiversity conservation. *Comput. Struct. Biotechnol. J.* **2015**, *13*, 370–375. [CrossRef] [PubMed]
4. Steiner, C.C.; Putnam, A.S.; Hoeck, P.E.; Ryder, O.A. Conservation genomics of threatened animal species. *Annu. Rev. Anim. Biosci.* **2013**, *1*, 261–281. [CrossRef] [PubMed]
5. Hayden, E.C. Technology: The $1,000 genome. *Nature* **2014**, *507*, 294–295. [CrossRef] [PubMed]
6. Gordon, D.; Huddleston, J.; Chaisson, M.J.; Hill, C.M.; Kronenberg, Z.N.; Munson, K.M.; Malig, M.; Raja, A.; Fiddes, I.; Hillier, L.W. Long-read sequence assembly of the gorilla genome. *Science* **2016**, *352*, 52–59. [CrossRef] [PubMed]
7. Glenn, T.C. Field guide to next-generation DNA sequencers. *Mol. Ecol. Resour.* **2011**, *11*, 759–769. [CrossRef] [PubMed]

8. Quail, M.A.; Smith, M.; Coupland, P.; Otto, T.D.; Harris, S.R.; Connor, T.R.; Bertoni, A.; Swerdlow, H.P.; Gu, Y. A tale of three next generation sequencing platforms: Comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genom.* **2012**, *13*, 341. [CrossRef] [PubMed]
9. Goodwin, S.; McPherson, J.D.; McCombie, W.R. Coming of age: Ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* **2016**, *17*, 333–351. [CrossRef] [PubMed]
10. 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. Genomics and the challenging translation into conservation practice. *Trends Ecol. Evol.* **2015**, *30*, 78–87. [CrossRef] [PubMed]
11. Garner, B.A.; Hand, B.K.; Amish, S.J.; Bernatchez, L.; Foster, J.T.; Miller, K.M.; Morin, P.A.; Narum, S.R.; O'Brien, S.J.; Roffler, G. Genomics in conservation: Case studies and bridging the gap between data and application. *Trends Ecol. Evol.* **2016**, *31*, 81–83. [CrossRef] [PubMed]
12. Taylor, H.R.; Dussex, N.; van Heezik, Y. Bridging the conservation genetics gap by identifying barriers to implementation for conservation practitioners. *Glob. Ecol. Conserv.* **2017**, *10*, 231–242. [CrossRef]
13. Britt, M.; Haworth, S.E.; Johnson, J.B.; Martchenko, D.; Shafer, A.B. The importance of non-academic coauthors in bridging the conservation genetics gap. *Biol. Conserv.* **2018**, *218*, 118–123. [CrossRef]
14. Huisman, J.; Kruuk, L.E.; Ellis, P.A.; Clutton-Brock, T.; Pemberton, J.M. Inbreeding depression across the lifespan in a wild mammal population. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 3585–3590. [CrossRef] [PubMed]
15. Kardos, M.; Taylor, H.R.; Ellegren, H.; Luikart, G.; Allendorf, F.W. Genomics advances the study of inbreeding depression in the wild. *Evol. Appl.* **2016**, *9*, 1205–1218. [CrossRef] [PubMed]
16. Hammerly, S.C.; Morrow, M.E.; Johnson, J.A. A comparison of pedigree-and DNA-based measures for identifying inbreeding depression in the critically endangered Attwater's Prairie-chicken. *Mol. Ecol.* **2013**, *22*, 5313–5328. [CrossRef] [PubMed]
17. Ivy, J.A.; Putnam, A.S.; Navarro, A.Y.; Gurr, J.; Ryder, O.A. Applying SNP-derived molecular coancestry estimates to captive breeding programs. *J. Hered.* **2016**, *5*, 403–412. [CrossRef] [PubMed]
18. Hoeck, P.E.; Wolak, M.E.; Switzer, R.A.; Kuehler, C.M.; Lieberman, A.A. Effects of inbreeding and parental incubation on captive breeding success in Hawaiian crows. *Biol. Conserv.* **2015**, *184*, 357–364. [CrossRef]
19. Rutz, C.; Klump, B.C.; Komarczyk, L.; Leighton, R.; Kramer, J.; Wischnewski, S.; Sugasawa, S.; Morrissey, M.B.; James, R.; St Clair, J.J.H.; Switzer, R.A.; Masuda, B.M. Discovery of species-wide tool use in the Hawaiian crow. *Nature* **2016**, *537*, 403–407. [CrossRef] [PubMed]
20. Culliney, S.; Pejchar, L.; Switzer, R.; Ruiz-Gutierrez, V. Seed dispersal by a captive corvid: The role of the ʻAlala (*Corvus hawaiiensis*) in shaping Hawai'i's plant communities. *Ecol. Appl.* **2012**, *22*, 1718–1732. [CrossRef] [PubMed]
21. U.S. Fish and Wildlife Service. Revised Recovery Plan for the ʻAlalā (*Corvus hawaiiensis*). 2009; pp. 1–104. Available online: https://www.fws.gov/pacific/ecoservices/documents/Alala_Revised_Recovery_Plan.pdf (accessed on 30 July 2018).
22. Hedrick, P.W.; Hoeck, P.E.; Fleischer, R.C.; Farabaugh, S.; Masuda, B.M. The influence of captive breeding management on founder representation and inbreeding in the ʻAlalā, the Hawaiian crow. *Conserv. Genet.* **2016**, *17*, 369–378. [CrossRef]
23. Fleischer, R. Genetic analysis of captive ʻAlalā (*Corvus hawaiiensis*). In *Report to the U.S. Fish and Wildlife Service*; Pacific Islands Fish and Wildlife Office: Honolulu, HI, USA, 2003; pp. 1–21.
24. Jarvi, S.I.; Bianchi, K.R. *Genetic Analyses of Captive ʻAlalā (Corvus hawaiiensis) Using AFLP Analyses*; Open-File Report 2006-1349; US Geological Survey: Reston, VA, USA, 2006; pp. 1–40.
25. McQuillan, R.; Leutenegger, A.-L.; Abdel-Rahman, R.; Franklin, C.S.; Pericic, M.; Barac-Lauc, L.; Smolej-Narancic, N.; Janicijevic, B.; Polasek, O.; Tenesa, A. Runs of homozygosity in European populations. *Am. J. Hum. Genet.* **2008**, *83*, 359–372. [CrossRef] [PubMed]
26. PacBio. Preparing > 30 kb SMRTbellTM Libraries Using the Megaruptor®Shearing and BluePippinTM Size-Selection System. Pacific Biosciences Unsupported Protocol. 2015. Available online: https://www.pacb.com/wp-content/uploads/Procedure-Checklist-Preparing-Greater-Than-30-kb-SMRTbell™-Libraries-Using-Megaruptor®-Shearing-and-BluePippin™-Size-Selection-on-Sequel®-and-RSII-Systems.pdf (accessed on 30 July 2018).
27. Chin, C.-S.; Peluso, P.; Sedlazeck, F.J.; Nattestad, M.; Concepcion, G.T.; Clum, A.; Dunn, C.; O'Malley, R.; Figueroa-Balderas, R.; Morales-Cruz, A. Phased diploid genome assembly with single-molecule real-time sequencing. *Nat. Methods* **2016**, *13*, 1050–1054. [CrossRef] [PubMed]

28. Myers, E.W. The fragment assembly string graph. *Bioinformatics* **2005**, *21*, ii79–ii85. [CrossRef] [PubMed]
29. Korlach, J.; Gedman, G.; Kingan, S.B.; Chin, C.-S.; Howard, J.T.; Audet, J.-N.; Cantin, L.; Jarvis, E.D. *De novo* PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience* **2017**, *6*, 1–16. [CrossRef] [PubMed]
30. 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] [PubMed]
31. International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **2004**, *432*, 695–716. [CrossRef] [PubMed]
32. Warren, W.C.; Clayton, D.F.; Ellegren, H.; Arnold, A.P.; Hillier, L.W.; Künstner, A.; Searle, S.; White, S.; Vilella, A.J.; Fairley, S. The genome of a songbird. *Nature* **2010**, *464*, 757–762. [CrossRef] [PubMed]
33. Poelstra, J.W.; Vijay, N.; Bossu, C.M.; Lantz, H.; Ryll, B.; Müller, I.; Baglione, V.; Unneberg, P.; Wikelski, M.; Grabherr, M.G. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* **2014**, *344*, 1410–1414. [CrossRef] [PubMed]
34. Zhang, G.; Li, C.; Li, Q.; Li, B.; Larkin, D.M.; Lee, C.; Storz, J.F.; Antunes, A.; Greenwold, M.J.; Meredith, R.W. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **2014**, *346*, 1311–1320. [CrossRef] [PubMed]
35. Simão, F. University of Geneva. Personal communication, 2016.
36. Parra, G.; Bradnam, K.; Korf, I. CEGMA: A pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **2007**, *23*, 1061–1067. [CrossRef] [PubMed]
37. Smit, A.F.A.; Hubley, R.; Green, P. RepeatMasker Open-4.0. 2013-2015. Available online: www.repeatmasker.org (accessed on 2 May 2018).
38. Bao, Z.; Eddy, S.R. Automated *de novo* identification of repeat sequence families in sequenced genomes. *Genome Res.* **2002**, *12*, 1269–1276. [CrossRef] [PubMed]
39. Price, A.L.; Jones, N.C.; Pevzner, P.A. *De novo* identification of repeat families in large genomes. *Bioinformatics* **2005**, *21*, i351–i358. [CrossRef] [PubMed]
40. Genetic Information Research Institute (GIRI). Giri REPBASE. Available online: girinst.org (accessed on 2 May 2018).
41. Kohany, O.; Gentles, A.J.; Hankus, L.; Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinform.* **2006**, *7*, 474. [CrossRef] [PubMed]
42. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **1999**, *27*, 573–580. [CrossRef] [PubMed]
43. Balakrishnan, C.N.; Ekblom, R.; Völker, M.; Westerdahl, H.; Godinez, R.; Kotkiewicz, H.; Burt, D.W.; Graves, T.; Griffin, D.K.; Warren, W.C. Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biol.* **2010**, *8*, 29. [CrossRef] [PubMed]
44. Stanke, M.; Steinkamp, R.; Waack, S.; Morgenstern, B. AUGUSTUS: A web server for gene finding in eukaryotes. *Nucleic Acids Res.* **2004**, *32*, W309–W312. [CrossRef] [PubMed]
45. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [CrossRef] [PubMed]
46. Rice, P.; Longden, I.; Bleasby, A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet.* **2000**, *16*, 276–277. [CrossRef]
47. Eimes, J.A.; Townsend, A.K.; Jablonski, P.; Nishiumi, I.; Satta, Y. Early duplication of a single MHC IIB locus prior to the passerine radiations. *PLoS ONE* **2016**, *11*, e0163456. [CrossRef] [PubMed]
48. Stamatakis, A. RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014**, *30*, 1312–1313. [CrossRef] [PubMed]
49. Trachtenberg, E.; Korber, B.; Sollars, C.; Kepler, T.B.; Hraber, P.T.; Hayes, E.; Funkhouser, R.; Fugate, M.; Theiler, J.; Hsu, Y.S.; et al. Advantage of rare HLA supertype in HIV disease progression. *Nat. Med.* **2003**, *9*, 928–935. [CrossRef] [PubMed]
50. Sepil, I.; Moghadam, H.K.; Huchard, E.; Sheldon, B.C. Characterization and 454 pyrosequencing of Major Histocompatibility Complex class I genes in the great tit reveal complexity in a passerine system. *BMC Evol. Biol.* **2012**, *12*, 68. [CrossRef] [PubMed]

51. Schwensow, N.; Fietz, J.; Dausmann, K.H.; Sommer, S. Neutral versus adaptive genetic variation in parasite resistance: Importance of major histocompatibility complex supertypes in a free-ranging primate. *Heredity* **2007**, *99*, 265–277. [CrossRef] [PubMed]
52. Doytchinova, I.A.; Flower, D.R. In silico identification of supertypes for class II MHCs. *J. Immunol.* **2005**, *174*, 7085–7095. [CrossRef] [PubMed]
53. Sandberg, M.; Eriksson, L.; Jonsson, J.; Sjöström, M.; Wold, S. New chemical descriptors relevant for the design of biologically active peptides. A multivariate characterization of 87 amino acids. *J. Med. Chem.* **1998**, *41*, 2481–2491. [CrossRef] [PubMed]
54. Eimes, J.A.; Townsend, A.K.; Sepil, I.; Nishiumi, I.; Satta, Y. Patterns of evolution of MHC class II genes of crows (*Corvus*) suggest trans-species polymorphism. *PeerJ* **2015**, *3*, e853. [CrossRef] [PubMed]
55. Edgar, R.C. MUSCLE: A multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.* **2004**, *5*, 113. [CrossRef] [PubMed]
56. Edgar, R.C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004**, *32*, 1792–1797. [CrossRef] [PubMed]
57. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012**, *28*, 1647–1649. [CrossRef] [PubMed]
58. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008**, *24*, 1403–1405. [CrossRef] [PubMed]
59. Jombart, T.; Devillard, S.; Balloux, F. Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genet.* **2010**, *11*, 94. [CrossRef] [PubMed]
60. Kirin, M.; McQuillan, R.; Franklin, C.S.; Campbell, H.; McKeigue, P.M.; Wilson, J.F. Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE* **2010**, *5*, e13996. [CrossRef] [PubMed]
61. 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. The variant call format and VCFtools. *Bioinformatics* **2011**, *27*, 2156–2158. [CrossRef] [PubMed]
62. Weissensteiner, M.H.; Pang, A.W.; Bunikis, I.; Höijer, I.; Vinnere-Petterson, O.; Suh, A.; Wolf, J.B. Combination of short-read, long-read, and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. *Genome Res.* **2017**, *27*, 697–708. [CrossRef] [PubMed]
63. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **2017**, *45*, D158–D169. [CrossRef] [PubMed]
64. Toll-Riera, M.; Castelo, R.; Bellora, N.; Alba, M.M. Evolution of primate orphan proteins. *Biochem. Soc. Trans.* **2009**, *37*, 778–782. [CrossRef] [PubMed]
65. Temperley, N.D.; Berlin, S.; Paton, I.R.; Griffin, D.K.; Burt, D.W. Evolution of the chicken Toll-like receptor gene family: A story of gene gain and gene loss. *BMC Genom.* **2008**, *9*, 62. [CrossRef] [PubMed]
66. Alcaide, M.; Edwards, S.V. Molecular evolution of the Toll-like receptor multigene family in birds. *Mol. Biol. Evol.* **2011**, *28*, 1703–1715. [CrossRef] [PubMed]
67. Grueber, C.E.; Knafler, G.J.; King, T.M.; Senior, A.M.; Grosser, S.; Robertson, B.; Weston, K.A.; Brekke, P.; Harris, C.L.; Jamieson, I.G. Toll-like receptor diversity in 10 threatened bird species: Relationship with microsatellite heterozygosity. *Conserv. Genet.* **2015**, *16*, 595–611. [CrossRef]
68. Cormican, P.; Lloyd, A.T.; Downing, T.; Connell, S.J.; Bradley, D.; O'Farrelly, C. The avian Toll-Like receptor pathway—Subtle differences amidst general conformity. *Dev. Comp. Immunol.* **2009**, *33*, 967–973. [CrossRef] [PubMed]
69. Grueber, C.E.; Wallis, G.P.; King, T.M.; Jamieson, I.G. Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS ONE* **2012**, *7*, e45011. [CrossRef] [PubMed]
70. Knafler, G.; Grueber, C.E.; Sutton, J.T.; Jamieson, I.G. Differential patterns of diversity at microsatellite, MHC, and TLR loci in South Island saddleback populations impacted by translocation and disease. *N. Z. J. Ecol.* **2017**, *41*, 98–106.
71. Hartmann, S.A.; Schaefer, H.M.; Segelbacher, G. Genetic depletion at adaptive but not neutral loci in an endangered bird species. *Mol. Ecol.* **2014**, *23*, 5712–5725. [CrossRef] [PubMed]
72. Bainová, H.; Králová, T.; Bryjová, A.; Albrecht, T.; Bryja, J.; Vinkler, M. First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds. *Dev. Comp. Immunol.* **2014**, *45*, 151–155. [CrossRef] [PubMed]

73. Burri, R.; Hirzel, H.N.; Salamin, N.; Roulin, A.; Fumagalli, L. Evolutionary patterns of MHC class II B in owls and their implications for the understanding of avian MHC evolution. *Mol. Biol. Evol.* **2008**, *25*, 1180–1191. [CrossRef] [PubMed]
74. Miller, H.C.; Lambert, D.M. Gene duplication and gene conversion in class II MHC genes of New Zealand robins (Petroicidae). *Immunogenetics* **2004**, *56*, 178–191. [CrossRef] [PubMed]
75. Zagalska-Neubauer, M.; Babik, W.; Stuglik, M.; Gustafsson, L.; Cichoń, M.; Radwan, J. 454 sequencing reveals extreme complexity of the class II Major Histocompatibility Complex in the collared flycatcher. *BMC Evol. Biol.* **2010**, *10*, 395. [CrossRef] [PubMed]
76. Klein, J.; Ono, H.; Klein, D.; O'hUigin, C. The Accordion Model of *Mhc* Evolution. In *Progress in Immunology Vol. VIII, Proceedings of the 8th International Congress of Immunology, Budapest, Hungary, 1992*; Gergely, J., Benczúr, M., Erdei, A., Falus, A., Füst, G., Medgyesi, G., Petrányi, G., Rajnavölgyi, E., Eds.; Springer: Berlin\Heidelberg, Germany, 1993; pp. 137–143.
77. Ellegren, H.; Smeds, L.; Burri, R.; Olason, P.I.; Backström, N.; Kawakami, T.; Künstner, A.; Mäkinen, H.; Nadachowska-Brzyska, K.; Qvarnström, A. The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **2012**, *491*, 756–760. [CrossRef] [PubMed]
78. Callicrate, T.E. *Population Declines and Genetic Variation: Effects of Serial Bottlenecks*; University of Maryland: College Park, MD, USA, 2015.
79. Aslam, M.L.; Bastiaansen, J.W.; Elferink, M.G.; Megens, H.-J.; Crooijmans, R.P.; Blomberg, L.A.; Fleischer, R.C.; Van Tassell, C.P.; Sonstegard, T.S.; Schroeder, S.G. Whole genome SNP discovery and analysis of genetic diversity in Turkey (*Meleagris gallopavo*). *BMC Genom.* **2012**, *13*, 391. [CrossRef] [PubMed]
80. Howrigan, D.P.; Simonson, M.A.; Keller, M.C. Detecting autozygosity through runs of homozygosity: A comparison of three autozygosity detection algorithms. *BMC Genom.* **2011**, *12*, 460. [CrossRef] [PubMed]
81. Ferenčaković, M.; Hamzić, E.; Gredler, B.; Solberg, T.; Klemetsdal, G.; Curik, I.; Sölkner, J. Estimates of autozygosity derived from runs of homozygosity: Empirical evidence from selected cattle populations. *J. Anim. Breed. Genet.* **2013**, *130*, 286–293. [CrossRef] [PubMed]
82. Hwang, S.; Kim, E.; Lee, I.; Marcotte, E.M. Systematic comparison of variant calling pipelines using gold standard personal exome variants. *Sci. Rep.* **2015**, *5*, 17875. [CrossRef] [PubMed]
83. Ralls, K.; Ballou, J.D.; Rideout, B.A.; Frankham, R. Genetic management of chondrodystrophy in California condors. *Anim. Conserv. Forum* **2000**, *3*, 145–153. [CrossRef]
84. Frankham, R. Challenges and opportunities of genetic approaches to biological conservation. *Biol. Conserv.* **2010**, *143*, 1919–1927. [CrossRef]
85. Ralls, K.; Ballou, J.D. Genetic status and management of California condors. *Condor* **2004**, *106*, 215–228. [CrossRef]
86. Romanov, M.N.; Koriabine, M.; Nefedov, M.; de Jong, P.J.; Ryder, O.A. Construction of a California condor BAC library and first-generation chicken–condor comparative physical map as an endangered species conservation genomics resource. *Genomics* **2006**, *88*, 711–718. [CrossRef] [PubMed]
87. Romanov, M.N.; Tuttle, E.M.; Houck, M.L.; Modi, W.S.; Chemnick, L.G.; Korody, M.L.; Mork, E.M.; Otten, C.A.; Renner, T.; Jones, K.C. The value of avian genomics to the conservation of wildlife. *BMC Genom.* **2009**, *10*, S10. [CrossRef] [PubMed]
88. Hedrick, P.W.; Garcia-Dorado, A. Understanding inbreeding depression, purging, and genetic rescue. *Trends Ecol. Evol.* **2016**, *31*, 940–952. [CrossRef] [PubMed]
89. Oven, I.; Rus, K.R.; Dušanić, D.; Benčina, D.; Keeler, C.L.; Narat, M. Diacylated lipopeptide from *Mycoplasma synoviae* mediates TLR15 induced innate immune responses. *Vet. Res.* **2013**, *44*, 99. [CrossRef] [PubMed]
90. Yang, Q.; Chen, H.; Wei, P. Marek's disease virus can infect chicken brain microglia and promote the transcription of toll-like receptor 15 and 1LB genes. *Chin. J. Virol.* **2011**, *27*, 18–25.
91. Ruan, W.; Wu, Y.; An, J.; Cui, D.; Li, H.; Zheng, S. Toll-like receptor 2 type 1 and type 2 polymorphisms in different chicken breeds. *Poult. Sci.* **2012**, *91*, 101–106. [CrossRef] [PubMed]
92. Tian, W.; Zhao, C.; Hu, Q.; Sun, J.; Peng, X. Roles of Toll-like receptors 2 and 6 in the inflammatory response to *Mycoplasma gallisepticum* infection in DF-1 cells and in chicken embryos. *Dev. Comp. Immunol.* **2016**, *59*, 39–47. [CrossRef] [PubMed]
93. Hu, X.; Zou, H.; Qin, A.; Qian, K.; Shao, H.; Ye, J. Activation of Toll-like receptor 3 inhibits Marek's disease virus infection in chicken embryo fibroblast cells. *Arch. Virol.* **2016**, *161*, 521–528.

94. Karaffová, V.; Marcinková, E.; Bobíková, K.; Herich, R.; Revajová, V.; Stašová, D.; Kavuľová, A.; Levkutová, M.; Levkut, M.; Lauková, A. TLR4 and TLR21 expression, MIF, IFN-β, MD-2, CD14 activation, and sIgA production in chickens administered with EFAL41 strain challenged with *Campylobacter jejuni*. *Folia Microbiol.* **2017**, *62*, 89–97.
95. St. Paul, M.; Paolucci, S.; Sharif, S. Treatment with ligands for toll-like receptors 2 and 5 induces a mixed T-helper 1-and 2-like response in chicken splenocytes. *J. Interferon Cytokine Res.* **2012**, *32*, 592–598. [CrossRef] [PubMed]
96. Wu, G.; Liu, L.; Qi, Y.; Sun, Y.; Yang, N.; Xu, G.; Zhou, H.; Li, X. Splenic gene expression profiling in White Leghorn layer inoculated with the *Salmonella enterica* serovar Enteritidis. *Anim. Genet.* **2015**, *46*, 617–626. [CrossRef] [PubMed]
97. Zhou, Z.; Wang, Z.; Cao, L.; Hu, S.; Zhang, Z.; Qin, B.; Guo, Z.; Nie, K. Upregulation of chicken TLR4, TLR15 and MyD88 in heterophils and monocyte-derived macrophages stimulated with *Eimeria tenella* in vitro. *Exp. Parasitol.* **2013**, *133*, 427–433. [CrossRef] [PubMed]