

A high-quality, long-read *de novo* genome assembly to aid conservation of Hawaii's last remaining crow species

Jolene T. Sutton^{1,*}, Martin Helmkamp¹, Cynthia C. Steiner², M. Renee Bellinger¹, Jonas Koriach³, Richard Hall³, Primo Baybayan³, Jill Muehling³, Jenny Gu³, Sarah Kingan³, Bryce M. Masuda⁴, and Oliver Ryder².

¹ University of Hawaii at Hilo, Hilo, Hawaii 96720, USA

² Institute for Conservation Research, San Diego Zoo, Escondido, California 92027, USA

³ Pacific Biosciences, Menlo Park, California 94025, USA

⁴ Institute for Conservation Research, San Diego Zoo Global, Volcano, Hawaii 96785, USA

* Correspondence: jtsutton@hawaii.edu; Tel.: +1-808-932-7183

Abstract: Genome-level data can provide researchers with unprecedented precision to examine the causes and genetic consequences of population declines, and to apply these results to 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 Alala. As the only remaining native crow species in Hawaii, the Alala 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 Alala genome is substantially more homozygous. We also provide annotations for a subset of immunity genes that are likely to be important for conservation applications, and we discuss how this genome is currently being used as a roadmap for downstream conservation applications.

Key words: Hawaiian crow; 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 enables 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]. 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 3,7-10].

Here we describe long-read, whole-genome sequencing and *de novo* assembly for the critically endangered Alala (Hawaiian crow; *Corvus hawaiiensis*). 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 can help to improve estimates of relatedness and individual inbreeding coefficients [e.g. 11-14]. 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 [15]. The genome will also be used for comparative studies aimed at understanding the evolution of tool-use and other behaviors [e.g. 16]. Such comparative genome analyses could be especially important for conservation purposes, as they offer the potential to identify the genomic basis of traits associated with inbreeding depression both within and across species.

1.1. Study species and aims

Historically widespread within mesic and dry forest habitats on the Island of Hawaii, the Alala population declined rapidly during the twentieth century [17]. Fewer than 100 individuals were estimated to be alive in the 1970s, and the population continued to decline to fewer than 20 individuals during the 1990s. By 2002 the species was described as “extinct in the wild” by the International Union for Conservation of Nature and Natural Resources (IUCN) Red List of Threatened Species [reviewed in 18]. Today, the Alala is one of the most endangered endemic bird species in Hawaii, having existed entirely in captivity from 2002-2016 [15]. All extant individuals are descended from nine genetic founders that established the conservation breeding program [described in 18,19]. In 2016, a long-term reintroduction program was initiated in 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 [15]. For example, the species suffers from low hatching success, and the majority of inbreeding in the population appears to be attributed to a single founding pair [18]. 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 [20,21]). These studies identified extremely low genetic diversity, which suggested that conservation efforts would benefit from a whole genome approach that could generate resources for targeting additional polymorphic regions (*e.g.* SNPs, and structural variation).

In this study we highlight the quality of the Alala 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 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 (*f*ROH; [22]; *i.e.* identical by descent). Finally, we briefly discuss the goals and perceived challenges for the next stages of data generation and applications to Alala 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 Alala, named Hōike 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 [18]). Library construction protocol followed the workflow for ultra large insert libraries [23]. The DNA was sheared to target 50 kb fragments (resulting distribution 30 – 80 kb) by using a Megaruptor® (Diagenode), and assessed for quality by pulsed-field gel electrophoresis (PFGE) on the CHEF Mapper® system (Bio-Rad). A total of 86 µg of DNA were then recovered from the 50 kb shearing condition. Sheared DNA was constructed into SMRTbell template by following the > 30 kb library construction protocol [23] with minor modifications (*e.g.* 1X AMPure PB purification; 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) 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 and C4 chemistry kits (P6-C4) and 6-hour run time movies, which yielded 128 Gb whole-genome sequence data. 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 [24] to generate long-range phased haplotypes. During the assembly process, sequence reads were overlapped to form long consensus sequences [6,24]. 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 [25]. Primary contigs were formed by using the sequences of non-branching paths, while associated contigs (“haplotigs”) represent the sequences of branching paths. The resulting assembly thus represents a phased diploid genome [24,26].

To assess the quality of the final assembly, we compared the number and length of Alala contigs to those of other avian assemblies. In addition, we used BUSCO v2.0.1 [27] to assess the completeness of the gene space in the Alala assembly based on the detection of conserved single-copy orthologs. For comparison, we included genome assemblies of domestic chicken (*Gallus gallus*, GenBank accession GCF_000002315.4 [28]), Anna’s hummingbird (*Calypte anna*, GCA_002021895.1 [26]), zebra finch (*Taenopygia guttata*, GCF_000151805.1 [29]), hooded crow (*Corvus cornix cornix*, GCF_000738735.1 [30]), and American crow (*C. brachyrhynchos*, GCF_000691975.1 [31]) in the analysis. As lineage datasets, we chose eukaryota_odb9 (303 genes) and a 250-gene eukaryotic subset (pers. comm. Felipe Simão), which is highly congruent with the CEGMA dataset [32]. 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 Alala assembly, we generated a *de novo* repeat library using RepeatModeler v1.0.11 (repeatmasker.org). This software package primarily integrates RECON v1.08 [33] and RepeatScout v1.0.5 [34] 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 (girinst.org). Additional, more detailed repeat classification was performed with CENSOR [35]. The Alala assembly was then screened for repetitive DNA using RepeatMasker v4.0.7 (repeatmasker.org) based on RMBlast and two repeat libraries: 1) the Alala 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 (accessed May 2, 2018). In addition to the version implemented in RepeatMasker, simple repeats were also assessed using the stand alone version of Tandem Repeats Finder v4.0.9 [36] with the following settings: Match = 2, Mismatching penalty = 7, Delta = 7, PM = 80, PI = 10, Minscore = 50, and MaxPeriod = 2,000. Along with the Alala 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 genes annotations

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 Alala assembly, we first

performed Blast (tblastn) searches using homologous protein sequences of other bird species as queries with an e-value cut-off of 1×10^{-5} . For MHC, queries were obtained from zebra finch, the currently best annotated passerine genome [37] (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 Alala 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 gene structure and coding sequence predicted by the AUGUSTUS web server v3.3 [38]. 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 [39]. 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 (http://www.ebi.ac.uk/Tools/st/emboss_transeq/). 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 Alala 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 Alala and other corvids. We included all functional and ambiguous Alala genes, as well as complete MHC class II B gene sets of single individuals each of American crow (13 genes), the jungle crow (*C. macrorhynchos japonensis*; 14 genes), the Asian rook (*C. frugilegus*; 11 genes) and the azure-winged magpie (*Cyanopica cyanus*, 7 genes) to allow for within-genome diversity comparisons. These data were obtained by Eimes *et al.* [40 and pers. comm.] using a targeted 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 [41]. 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 Alala 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. 42,43]. 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 [44-46]. As the Alala 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 [47]. First, we used MUSCLE [48,49] implemented in Geneious vR10 [50] to align our putatively functional Alala exon 2 sequences to the 237 sequences from Eimes *et al.* [47], 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 [44-46]: 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 [51], to reveal clusters of sequences likely to have similar functional properties. We then used discriminant analysis of principle components (DAPC) to describe the clusters [52].

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 [53]. As mutations accumulate over time, ROH segments break down. We assessed ROH in the Alala genome for three purposes: 1) to estimate the autozygous fraction of the genome f_{ROH} [22], *i.e.* the total fraction of the genome that is perfectly autozygous (zero heterozygosity); 2) to estimate f_{ROH} on a per-contig basis; and 3) to evaluate the effect of allowance for low levels of heterozygosity on estimates of whole-genome f_{ROH} and ROH segment lengths. These analyses were conducted using 100 kb sliding windows, with the number of SNPs per sliding window calculated with *vcftools* [54], using the *p.vcf* file (*i.e.* from the primary contigs) generated during the assembly. A total of four categories of allowable heterozygosity thresholds were tested: ≤ 1 , 2, 4, or 10 SNPs per sliding window, which corresponds to 1 SNP per 100kb, 50kb, 25kb or 10kb (*i.e.* ≤ 0.01 , ≤ 0.02 , ≤ 0.04 , and ≤ 0.1 SNPs/kb). Only contigs ≥ 500 kb were included in this analysis, which permits a minimum of five consecutive sliding windows to be included in the ROH segment length estimates. The autozygous fraction of the genome (f_{ROH}) was calculated by taking the sum of the number of ROH segments multiplied by the corresponding ROH length and dividing these by the total number of sliding windows considered in the analysis. The ROH lengths were calculated by summing consecutive sliding windows that met criteria of perfect autozygosity or fell within the heterozygosity threshold.

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 suggest that 75% and 100% of the genomes of two more outbred species, zebra finch and Anna's hummingbird (*Calypte anna*), contained sufficient heterozygosity to be phased into haplotypes [26] (Table 1). Compared to other published short-read based avian genomes of similar size, the Alala 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 [26]. The BUSCO analysis indicated that gene completeness was among the highest of any avian genome to date (Figure 1). Collectively, these results suggest that this Alala long-read genome assembly is one of the highest quality avian genomes currently available.

3.2. Mobile and repetitive elements

De novo repeat modeling resulted in an Alala-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 Alala repeat families were partially similar to the large tandem repeat "crowSat1", a 14 kb satellite that is suspected to be a major heterochromatin component in the hooded crow [55]. In contrast, extended matches to Swiss-Prot entries, which might indicate co-opted transposable elements [56], were not discovered. RepeatMasker screening identified 10.1% of the Alala 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 by 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 max 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%), or 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 Alala). Combined, the RepeatMasker and Tandem Repeats Finder estimates suggest that the Alala assembly contains approximately 15% repetitive DNA. Since large parts of heterochromatic regions routinely elude current-generation sequencing efforts, and repetitive DNA that is sequenced cannot often not be assembled reliably, this is likely an underestimation of the true genome repeat content.

3.3. Candidate gene annotations and analysis

We identified the full complement of avian TLR genes in the Alala genome (Table S2), and found them to be highly conserved in number and sequence with respect to the chicken reference [57]. Gene structure also proved identical to other birds [58], since only a single untranslated 5' exon seemed to be missing from one prediction. All genes could be classified as functional with full-length reading frames by comparison to other birds. Compared to chicken, notable features included the loss of 140 aa at the 5' end in *TLR1A*, a ~50 aa indel in *TLR2B*, and a tandem arrangement of two *TLR7* duplicates differing by 16 aa, which we here annotate as *TLR7a* and *TLR7b*. Although a single *TLR7* copy exists in most other birds with annotated TLR repertoires, including zebra and house finch, duplication does occur in several passerine species [58-62], indicating that duplicates likely predate the split of the corvid family from other passerines. We did not find evidence that *TLR5* was pseudogenized in Alala, as it is in some passerine species [63]; however, based on a single individual this conclusion should be taken with caution.

The MHC class II beta repertoire of the Alala 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 (Table S3). This places the Alala within the lower end of the range of known MHC class II B genes relative to other corvids (e.g. 7–20 alleles per individual in American crow, jungle crow, and carrion crow [47], but note those numbers may include non-functional variants). The Alala genes appear largely conserved compared to zebra finch at the amino acid level, with sequence identity values slightly above 80%. However, it should be noted that this value is derived across entire gene regions, rather than at putative peptide binding regions (in exon 2 for class II B genes), where diversity may be expected to be large. We evaluated exon 2 diversity on a preliminary basis during this supertype analysis, with future work planned to better assess population-level diversity of this region.

The Alala genome also appears to contain a large number of MHC class II B pseudogenes, consistent with expectations for passerines [e.g. 64]. In addition to three presumably non-functional genes comprised of exons 2 and 3 (or identifiable parts thereof), we discovered more than 130 sequences homologous to exon 2 on primary contigs, which encodes the highly variable class II histocompatibility antigen beta domain. An additional ~ 30 homologs were found of the more conserved exon 3, which contains the immunoglobulin C1-set domain that mediates T lymphocyte binding (Table S4). Many of these pseudogenes appeared to be highly fragmented, i.e. homology could only be established over a short length of 150 bp or less. Sequence identity of these fragments fluctuated widely, ranging from near perfect matches with functional genes to less than 40% at the amino acid level, suggesting a broad age distribution with regard to the time of pseudogenization, and making assembly artefacts inflating the number of pseudogenes unlikely. This might be a reflection of the dynamic evolution of this hyper-variable gene family including repeated expansions and contractions over evolutionary timescales [65]. Sequence homology was also visible in adjacent genomic sequence to a lesser degree, suggesting that retrotransposition was not a major mechanism of gene duplication (alignment of 20 randomly chosen pseudogenes ± 250 bp up- and downstream).

Phylogenetically, six of the functional and ambiguous Alala genes comprised a strongly supported clade (Figure S2). These copies only differed by 0–3 amino acids, which were found at or

very close to the positively selected sites (PSS) described for other corvids [47]. In contrast, only three genes – highly similar copies “a” and “d”, as well as copy “g” – separated to other locations of the phylogenetic tree, albeit with lower support (Figure S2).

3.4. MHC functional supertypes

Similarity of the Alala MHC class II B repertoire in terms of functional antigen-binding properties to three other corvids was assessed on the basis of nine PSS [47]. 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 Eimes *et al.* [47]. Congruent with the phylogenetic analysis, only three of the eight supertypes corresponded to Alala, and were also shared by the three other corvid species that were compared (Figure 3). However, it must be noted that sequences from only one Alala were used (compared to 4-6 individuals for each of the other species included in the supertype analysis). No Alala supertypes were discovered to be separate from other corvids.

3.5. Runs of homozygosity

A total of 413,114 SNPS were detected across the 209 Alala genome contigs that met our minimum ROH analysis requirement of being at least 500 kb in length (1.02 Gb of sequence data, 96% of the genome). Based on the resulting 10,292 sliding windows, the fraction of the genome that was perfectly autozygous (*i.e.*, fROH) was approximately 5.1% (Table S5). The fROH estimates were highly sensitive to allowable levels of heterozygosity. For example, allowing for 1 SNP per 50kb or 25kb in a ROH increased the proportion of whole-genome autozygosity to 28% or 46%, respectively (Table S5). A general trend for low genetic diversity was present across most of the genome, with SNP/kb median and average values of 0.05 and 0.40, respectively. Only a small proportion of sliding windows, 15.6% (1604), were estimated to contain 1 or more SNPs/kb. In terms of individual contigs, the proportion of perfectly autozygous ROHs relative to all sliding windows was highly variable, ranging from a minimum of 0 % to a maximum of 70 % (median 5.3 %; average 1.9 %). The ROH segment lengths were short when restricting ROH criteria to perfectly autozygous, but became increasingly long when allowing or low- to moderate levels of heterozygosity across contigs (Table S5).

4. Discussion

Using the 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 [26]. By definition, regions of the primary assembly that have corresponding associated contigs identify areas in the genome with high degrees of heterozygosity. The genome assembly of a single Alala (studbook #32, Figure S1) highlights the genomic signatures of small population size and inbreeding, because the Alala associated contigs corresponded to a substantially smaller proportion of the genome compared to more outbred species.

4.1. Candidate genes

By comparison to many other avian genomes, the Alala 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. The phylogenetically close affiliation and high similarity of six putatively functional Alala MHC class II B genes suggests they may have formed recently, *i.e.* after the split from the other corvids, by multiple rounds of gene duplication. Alternatively, convergent evolution due to strong selection acting on exon 2 (*e.g.* by gene conversion) may also have made these genes more similar to each other and thus mask the gene family’s true evolutionary history [66,67]. Genome-scale data of exon 3, or intronic sequence from other corvids and additional Alala individuals may be required to resolve this question. Three Alala MHC class II B sequences did not cluster together. While a lack of support along the backbone of the phylogenetic tree prevented the identification of clear orthologs with other species, these three sequences may represent remnants

of evolutionarily distinct gene lineages. Multiple MHC class II B lineages have been identified in other corvids, suggesting that the gene family expanded prior to the radiation of the corvid family [40,47]. It should be noted that no Alala genes were found to fall into any of several MHC class II B groups consisting of genes from all or almost all other corvids investigated here (Figure S2). It is thus possible that such Alala genes were lost in the evolutionary lineage leading to the Alala, 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 more recent pseudogenization event. However, these observations are based on only a single individual and should be interpreted with caution. Genome data from additional specimens would be required to gauge within-species diversity and distribution of different MHC class II B lineages, and put these data into the context of corvid and Alala evolution. Additionally, our functional supertype analysis suggests that while nucleotide sequences may differ, between Alala and other corvids, much similarity exists when it comes to pathogen-binding properties.

4.2. Runs of homozygosity

The SNP encounter rate estimate in the Alala is about 1 SNP per 2,477 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 [68], 1 SNP per 256 bp in Hawaii Amakihi (*Hemignathus virens* [69]) and 1 SNP per 935 bp in turkey (*Meleagris* spp. [70]). The Anna's hummingbird genome [26] revealed 1,841,030 variants (*i.e.* 1 SNP per 501 bp) across $n = 283$ contigs of minimum length of 500 kb (923.1 Mb). 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 Alala genome is not surprising because of the overall low population size of Alala and this particular bird's high pedigree inbreeding coefficient (0.25). In the examples noted here, the Alala 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 chromosomes showing very low heterozygosity in the Alala is consistent with empirical observations made of ROHs in turkey [70], and large stretches of very low heterozygosity in Hawaii Amakihi [69]. 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 cumulative lengths of ROH are an indication of shared ancestry and can be used to gauge whether inbreeding events occurred within recent or distant generations [53]. Recombination events break long tracts into smaller pieces, thus numerous short tracts are consistent with distant shared ancestry. In this Alala genome the majority of tracts of ROHs were short, even when allowing for 1 SNP per 100 kb (Table S5). Yet, if allowable heterozygosity is increased to 1 SNP per 25,000 bp then 9.8% (104 of 1054) of all tracts would exceed 1 Mb. Given this measure's high sensitivity to filtering criteria interpretation of results would be aided considerably by whole genome sequencing of additional progeny or parents. The rate of sequencing error will also affect the estimations of homozygosity and, correspondingly, the lengths of runs of homozygosity.

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; the variant-calling pipeline; and computational settings [*e.g.* 71-73]. 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 [71,72]. The density of SNPs across the genome also impacts the minimum length of detection for a ROH, for example, reliable detected of ROHs as short as 100 kb versus 510 Mb using a less dense panel of markers [53]. Comparability of ROH results between avian and mammalian study species are 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 Alala conservation

The Alala genome assembly resulting from long-read sequencing data provides a high-quality reference genome that is enabling downstream comparative, population, and conservation applications. Prior to this study, molecular work using limited genetic markers identified low diversity in the species [20,21], and pedigree analysis identified inbreeding effects on hatching success, as well as skewed founder representation [15,18]. Based on these studies, we identified a need to generate genomic resources to aid conservation management. These included a more detailed picture of population-level genomic diversity and genetic load, as well as more accurate estimates of molecular relatedness to assist with choosing breeding pairs. Genomic resources were also desired to begin investigations of the basis of traits such as poor hatching success, as ~ 61% of fertile Alala eggs fail [15], compared with ~ 10% in most birds [74]. Since the generation of this Alala assembly, several projects have been initiated that rely heavily on use of this new resource. For example, we are currently using targeted SNP-capture to compare genomic diversity in museum and modern Alala, to better understand the impact of population bottlenecks over the past 100 years, and to provide a clearer picture of how much diversity can likely be maintained into the future (manuscript in preparation). So far, these analyses have benefitted from our high-quality genome assembly by allowing identification and removal of sequence reads that map to multiple genomic locations, and through testing and control for linkage disequilibrium (*i.e.* better quality filters). More recently, we have begun reduced representation library sequencing using restriction site-associated DNA sequencing (RAD-seq) strategies to genotype every individual Alala, so that we can inform the choice of breeding pairs in captivity by using realized relatedness estimates that likely correlate better with phenotypic trait information compared to pedigree-based relatedness estimates [*e.g.* 11]. Analytic methods for RADseq data show less bias when a suitable reference genome is included [75]. We have also initiated population-level assessments of candidate genes, which were aided by our annotations. Long-term, we plan to establish whole-genome data for multiple individuals, and to incorporate all of these datasets into models that test for genomic basis of particular phenotypic traits (*e.g.* poor hatching success), as well as models of mate choice and other behaviors.

Genomic data derived from our analyses are an essential component of the current and future recovery of the Alala. Although pair selection and managed breeding using the pedigree has kept the inbreeding level of the Alala population at a relatively low level over the past 20 years [18], the intensive and ongoing conservation management of the species requires a more detailed approach [*e.g.* 11]. In conjunction with ongoing conservation breeding efforts, a comprehensive reintroduction program is underway 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 founder population surviving in the wild at the time of writing. Ongoing management decisions for the breeding and release of particular individuals will have implications for the recovery of the species. As the size of both the captive and wild Alala 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 www.mdpi.com/xxx/s1, Tables S1-S5, Figures S1- S2. Primary and secondary genome assemblies will be available from Genbank, Accessions: (to be provided during review).

Ethics Approval: 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).

Author Contributions: Conceptualization, Jolene T. Sutton, Cynthia C. Steiner, Jenny Gu, Bryce M. Masuda and Oliver Ryder; Data curation, Richard Hall, Primo Baybayan and Jill Muehling; Formal analysis, Jolene T. Sutton, Martin Helmkampf, Cynthia C. Steiner, M. Renee Bellinger, Jonas Korchach and Sarah Kingan; Methodology, Jonas Korchach, Richard Hall, Primo Baybayan and Jill Muehling; Writing – original draft, Jolene T. Sutton, Martin Helmkampf, Cynthia C. Steiner, M. Renee Bellinger, Jonas Korchach, Bryce M. Masuda and Oliver Ryder.

Funding: Funding for the conservation breeding and reintroduction efforts has been provided by the San Diego Zoo Global, U.S. Fish and Wildlife Service, Hawaii 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 Alala at the facilities on Maui (Maui Bird Conservation Center) and Hawaii (Keauhou Bird Conservation Center).

Conflicts of Interest: JK, RH, PB, JM, JG, and SK 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

Tables

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
Alala		
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 ¹		
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 ¹		
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

¹ 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.

Table S1. Accessions of references used to annotate candidate immunity genes.

Gene group	Reference species	Reference accession	Source
MHC	zebra finch	UniProtKB: H1A2M7_TAEGU	29
		UniProtKB: H0ZVH6_TAEGU	29
		UniProtKB: H0ZU54_TAEGU	29
		UniProtKB: H1A3G3_TAEGU	29
		UniProtKB: H0ZXA5_TAEGU	29
		UniProtKB: H0ZZB1_TAEGU	29
		UniProtKB: H1A0J1_TAEGU	29
		UniProtKB: H0ZWA2_TAEGU	29
		UniProtKB: H0ZWI5_TAEGU	29
		UniProtKB: H0ZWL2_TAEGU	29
		UniProtKB: H0ZXL7_TAEGU	29
TLR	chicken	GenBank: NM_001007488.4	76
		GenBank: NM_001081709.3	77
		GenBank: NM_204278.1	78
		GenBank: NM_001161650.1	79
		GenBank: NM_001011691.3	80
		GenBank: NM_001030693.1	81
		GenBank: NM_001024586.1	82
		GenBank: NM_001011688.2	83
		GenBank: NM_001037835.1	84
GenBank: NM_001030558.1	81		

Table S2. Alala toll-like receptor (TLR) genes, following the nomenclature suggested by Temperley et al. (2008). 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.

Gene	Contig	Start	Stop	Orient.	No. aa	Exons	Status	Notes
Coha_TLR1A	000007F	17637771	17640173	+	661	1 (2)	F	5' end 140 aa shorter
Coha_TLR1B	000007F	17626125	17628053	+	643	1 (1)	F	
Coha_TLR2A	000039F	7428319	7430634	+	772	1 (2)	F	
Coha_TLR2B	000039F	7435849	7438197	+	771	1 (2)	F	~50 aa indel
Coha_TLR3	000007F	8442693	8448953	+	896	4 (5)	F	
Coha_TLR4	000034F	6443202	6446898	-	843	3 (3)	F	
Coha_TLR5	000035F	2853399	2855912	-	838	1 (1)	F	
Coha_TLR7a	000000F	13894219	13897341	-	1027	1 (2)	F	tandem duplicate (diff. 16 aa)
Coha_TLR7b	000000F	13880285	13883407	-	1027	1 (2)	F	tandem duplicate (diff. 16 aa)
Coha_TLR15	000045F	6705567	6708188	-	874	1 (1)	F	
Coha_TLR21	000253F	183362	186423	+	968	2 (2)	F	

Table S3. Alala MHC class II beta genes. Beginning and end coordinates refer to initial, uncurated gene models, and may or may not coincide with start and stop codons. The Exons column provides the number of exons identified in the absence of transcriptional evidence. The Status column indicates whether the predicted gene is presumably functional (F), ambiguous (A) or a pseudogene (P) according to the definition given in the main text. Only pseudogenes containing both exon 2 and 3 are listed. “Frameshift fixed” denotes gene models that were fully recovered by the insertion of a single nucleotide at a homopolymer (suggesting a sequencing error).

Gene	Contig	Begin	End	Orient.	No. aa	Exons	Status	Notes
Coha_MHCIIB_a	000257F	17839	19521	–	250	5	F	
Coha_MHCIIB_b	000357F	54501	55649	+	250	5	F	
Coha_MHCIIB_c	000357F	69097	70245	+	250	5	F	
Coha_MHCIIB_d	000357F	80536	81696	+	250	5	F	
Coha_MHCIIB_e	000485F_001	9214	10361	+	250	5	A	Frameshift fixed
Coha_MHCIIB_f	000485F_001	29737	30884	+	250	5	F	
Coha_MHCIIB_g	000485F_001	46119	47225	+	250	5	F	
Coha_MHCIIB_h	000868F	15490	16637	+	250	5	F	
Coha_MHCIIB_i	000868F	26421	27564	+	249	5	A	Frameshift fixed
Coha_MHCIIB_p1	000868F	4152	5188	+	251	5	P	Divergent, multiple frameshifts, non-canonical splice site
Coha_MHCIIB_p2	000251F	76819	77558	+	212	3	P	Fragmentary (exons 1–3), frameshift, mutated splice site
Coha_MHCIIB_p3	000219F	136202	137011	+	99	3	P	Fragmentary (exons 1–3), multiple frameshifts

Table S4. Homologs of MHC class II B exons 2 and 3 in the Alala 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 S2). The remaining hits are fragments that may represent ancient, pseudogenized copies.

Exon	Contig	%id	Length	Mism.	Gaps	Q-start	Q-end	H-start	H-end	E-value	Score
2	000114F	48	44	23	0	41	84	3271249	3271118	3.00E-07	53
2	000114F	48	44	23	0	41	84	3293850	3293981	4.00E-07	53
3	000135F	31	74	50	1	21	94	994087	993869	5.00E-06	49
3	000152F	46	68	35	2	21	86	2087390	2087187	8.00E-09	58
3	000152F	47	68	34	2	21	86	2094071	2093868	6.00E-09	58
2	000171F	48	44	23	0	41	84	107048	106917	6.00E-08	55
3	000219F	67	24	8	0	71	94	100963	100892	2.00E-07	37
3	000219F	65	26	9	0	49	74	101028	100951	2.00E-07	39
3	000219F	89	35	4	0	21	55	101411	101307	4.00E-13	71
2	000219F	41	37	22	0	4	40	136105	136215	3.00E-12	39
2	000219F	53	47	20	1	46	90	136202	136342	3.00E-12	53
3	000219F	85	33	5	0	19	51	136913	137011	5.00E-09	59
3	000219F	74	27	7	0	48	74	144963	145043	8.00E-09	43
3	000219F	67	24	8	0	71	94	145031	145102	8.00E-09	37
2	000219F	54	50	23	0	38	87	147239	147090	2.00E-06	50
3	000219F	32	74	50	0	21	94	221397	221618	7.00E-06	49
2	000219F	54	59	27	0	2	60	246230	246054	3.00E-11	65
3	000219F	63	24	9	0	71	94	267676	267605	2.00E-25	34
3	000219F	85	55	8	0	20	74	267828	267664	2.00E-25	102
3	000219F	34	74	49	0	21	94	331051	330830	3.00E-06	50
2	000219F	52	33	16	0	2	34	484799	484897	1.00E-10	40
2	000219F	46	48	26	0	38	85	484906	485049	1.00E-10	47
2	000219F	46	48	26	0	38	85	507922	507779	1.00E-10	47
2	000219F	52	33	16	0	2	34	508029	507931	1.00E-10	40
2	000220F	55	58	26	0	2	59	427986	428159	3.00E-12	68
2	000220F	53	58	27	0	2	59	466889	466716	9.00E-12	66
2	000220F	53	58	27	0	2	59	508632	508805	2.00E-11	66
2	000220F	54	54	25	0	2	55	511512	511351	3.00E-09	59
2	000227F	52	33	16	0	2	34	237850	237948	6.00E-11	41
2	000227F	46	48	26	0	38	85	237957	238100	6.00E-11	47
2	000227F	46	48	26	0	38	85	292983	292840	1.00E-10	47
2	000227F	52	33	16	0	2	34	293090	292992	1.00E-10	40
2	000227F	46	48	26	0	38	85	362829	362686	8.00E-11	47
2	000227F	52	33	16	0	2	34	362936	362838	8.00E-11	40
2	000227F	52	33	16	0	2	34	409156	409254	2.00E-10	40

2	000227F	46	48	26	0	38	85	409263	409406	2.00E-10	47
2	000227F	47	53	28	0	38	90	497804	497962	5.00E-07	52
3	000232F	75	24	6	0	71	94	48497	48426	5.00E-24	41
3	000232F	77	56	10	1	19	74	48643	48485	5.00E-24	91
2	000232F	48	48	25	0	38	85	111418	111275	3.00E-11	47
2	000232F	55	33	15	0	2	34	111525	111427	3.00E-11	42
2	000232F	48	44	23	0	41	84	166515	166646	7.00E-08	55
2	000232F	48	44	23	0	41	84	183460	183329	7.00E-08	55
2	000239F	53	40	19	0	41	80	116209	116328	3.00E-07	53
2	000239F	51	41	20	0	41	81	135721	135843	1.00E-07	54
2	000239F	36	70	45	0	15	84	179257	179048	5.00E-07	52
2	000239F	48	44	23	0	41	84	201332	201463	8.00E-07	52
2	000239F	39	70	43	0	15	84	214416	214625	6.00E-11	64
2	000239F	52	27	13	0	8	34	330793	330873	6.00E-10	34
2	000239F	45	49	27	0	37	85	330879	331025	6.00E-10	50
2	000239F	48	44	23	0	41	84	384018	384149	2.00E-07	54
2	000245F	43	44	25	0	41	84	38823	38954	3.00E-06	50
2	000250F	52	33	16	0	2	34	21954	22052	1.00E-10	40
2	000250F	46	48	26	0	38	85	22061	22204	1.00E-10	47
2	000250F	52	33	16	0	2	34	80889	80987	1.00E-10	40
2	000250F	46	48	26	0	38	85	80996	81139	1.00E-10	47
2	000250F	46	48	26	0	38	85	167760	167617	1.00E-10	47
2	000250F	52	33	16	0	2	34	167867	167769	1.00E-10	40
2	000250F	46	48	26	0	38	85	188616	188473	1.00E-10	47
2	000250F	52	33	16	0	2	34	188723	188625	1.00E-10	40
2	000250F	52	33	16	0	2	34	217588	217686	1.00E-10	40
2	000250F	46	48	26	0	38	85	217695	217838	1.00E-10	47
2	000251F	52	33	16	0	2	34	28304	28402	9.00E-11	40
2	000251F	46	48	26	0	38	85	28411	28554	9.00E-11	47
2	000251F	66	89	30	0	2	90	76819	77085	1.00E-30	120
3	000251F	83	54	9	0	21	74	77397	77558	7.00E-27	98
3	000251F	75	24	6	0	71	94	77545	77616	7.00E-27	44
3	000257F	100	76	0	0	19	94	18066	17839	8.00E-44	159
2	000257F	83	90	15	0	1	90	19521	19252	2.00E-42	155
2	000258F	46	48	26	0	38	85	71149	71006	1.00E-10	47
2	000258F	52	33	16	0	2	34	71256	71158	1.00E-10	40
2	000258F	52	33	16	0	2	34	143799	143897	1.00E-10	40
2	000258F	46	48	26	0	38	85	143906	144049	1.00E-10	47
2	000258F	46	48	26	0	38	85	168170	168027	9.00E-11	47
2	000258F	52	33	16	0	2	34	168277	168179	9.00E-11	40
2	000258F	46	52	28	0	38	89	222063	221908	1.00E-11	50
2	000258F	52	33	16	0	2	34	222170	222072	1.00E-11	40

2	000264F_001	54	57	26	0	3	59	28478	28308	2.00E-11	66
2	000264F_005	52	58	28	0	2	59	5985	6158	8.00E-11	64
2	000264F_005	52	58	28	0	2	59	8863	8690	7.00E-11	64
2	000264F_005	55	58	26	0	2	59	25285	25112	8.00E-13	69
2	000264F_005	53	58	27	0	2	59	41788	41615	1.00E-11	66
2	000264F_005	53	58	27	0	2	59	70310	70483	7.00E-12	67
2	000264F_005	54	54	25	0	2	55	73186	73025	9.00E-10	61
2	000264F_005	53	58	27	0	2	59	97560	97733	1.00E-11	66
2	000264F_005	53	58	27	0	2	59	100451	100278	1.00E-11	66
2	000264F_005	53	58	27	0	2	59	136272	136099	3.00E-12	68
2	000264F_005	53	58	27	0	2	59	169011	169184	1.00E-11	66
2	000264F_005	52	58	28	0	2	59	171893	171720	8.00E-11	64
2	000264F_005	50	44	22	0	2	45	189479	189610	1.00E-06	51
2	000264F_007	52	58	28	0	2	59	4288	4461	1.00E-10	63
2	000264F_007	52	58	28	0	2	59	20033	20206	1.00E-10	63
2	000264F_007	53	58	27	0	2	59	22924	22751	1.00E-11	66
2	000276F	36	70	45	0	15	84	87607	87816	3.00E-08	56
2	000276F	39	69	42	0	16	84	151735	151529	1.00E-07	54
2	000295F	48	44	23	0	41	84	19846	19977	1.00E-07	54
2	000295F	48	44	23	0	41	84	70504	70635	2.00E-07	53
2	000295F	45	44	24	0	41	84	91040	90909	8.00E-07	52
2	000295F	48	44	23	0	41	84	116480	116611	2.00E-07	53
2	000295F	48	44	23	0	41	84	153048	152917	1.00E-07	54
2	000309F	45	44	24	0	41	84	92564	92695	5.00E-07	52
2	000315F	48	44	23	0	41	84	36308	36177	4.00E-07	52
2	000315F	48	44	23	0	41	84	56092	56223	2.00E-06	50
2	000332F	50	44	22	0	41	84	60499	60630	2.00E-08	56
3	000344F	31	74	51	0	21	94	34098	33877	3.00E-06	50
2	000348F	52	33	16	0	2	34	19721	19819	7.00E-11	41
2	000348F	46	48	26	0	38	85	19828	19971	7.00E-11	47
2	000348F	46	48	26	0	38	85	40978	40835	1.00E-10	47
2	000348F	52	33	16	0	2	34	41085	40987	1.00E-10	40
2	000353F	53	58	27	0	2	59	15730	15903	5.00E-12	67
3	000357F	34	74	49	0	21	94	47879	47658	6.00E-06	49
2	000357F	99	90	1	0	1	90	54501	54770	7.00E-56	194
3	000357F	100	76	0	0	19	94	55422	55649	2.00E-44	160
2	000357F	98	90	2	0	1	90	69097	69366	5.00E-57	197
3	000357F	100	76	0	0	19	94	70018	70245	2.00E-44	160
2	000357F	74	90	23	0	1	90	80536	80805	3.00E-35	134
3	000357F	100	76	0	0	19	94	81469	81696	9.00E-44	159
2	000358F	38	84	52	0	2	85	32443	32694	9.00E-11	63
2	000366F	52	58	28	0	2	59	9616	9789	8.00E-11	64

2	000366F	53	58	27	0	2	59	12500	12327	3.00E-12	68
2	000366F	52	58	28	0	2	59	28579	28406	1.00E-10	63
2	000366F	52	56	27	0	4	59	44455	44288	1.00E-10	63
2	000366F	53	58	27	0	2	59	77094	77267	9.00E-12	66
2	000369F	53	58	27	0	2	59	2455	2282	2.00E-11	66
2	000369F	53	58	27	0	2	59	41486	41313	1.00E-11	66
2	000369F	53	58	27	0	2	59	58158	58331	1.00E-11	66
2	000369F	53	58	27	0	2	59	61040	60867	1.00E-11	66
2	000369F	53	58	27	0	2	59	77661	77834	1.00E-11	66
2	000369F	53	58	27	0	2	59	80544	80371	2.00E-11	66
2	000371F	45	47	26	0	41	87	31915	31775	1.00E-06	51
2	000375F	48	44	23	0	41	84	73626	73495	3.00E-07	53
3	000381F	75	24	6	0	71	94	64203	64132	7.00E-24	41
3	000381F	77	56	10	1	19	74	64349	64191	7.00E-24	90
2	000410F	48	44	23	0	41	84	67033	66902	2.00E-06	50
2	000413F	46	48	26	0	38	85	19766	19623	1.00E-10	47
2	000413F	52	33	16	0	2	34	19873	19775	1.00E-10	40
2	000413F	52	33	16	0	2	34	70704	70802	9.00E-11	40
2	000413F	46	48	26	0	38	85	70811	70954	9.00E-11	47
2	000421F	52	27	13	0	8	34	40362	40442	8.00E-12	34
2	000421F	47	53	28	0	38	90	40451	40609	8.00E-12	56
2	000421F	48	44	23	0	41	84	79625	79756	2.00E-07	54
2	000447F	46	48	26	0	38	85	32873	32730	1.00E-10	46
2	000447F	52	33	16	0	2	34	32980	32882	1.00E-10	40
2	000454F_001	48	44	23	0	41	84	7404	7273	2.00E-07	54
2	000454F_001	45	44	24	0	41	84	30019	29888	3.00E-06	50
2	000485F_001	94	90	5	0	1	90	9214	9483	5.00E-52	182
3	000485F_001	100	76	0	0	19	94	10134	10361	3.00E-44	160
2	000485F_001	98	90	2	0	1	90	29737	30006	9.00E-57	196
3	000485F_001	100	76	0	0	19	94	30657	30884	3.00E-44	160
2	000485F_001	78	90	20	0	1	90	46119	46388	1.00E-39	147
3	000485F_001	100	76	0	0	19	94	46998	47225	8.00E-44	159
2	000501F	53	58	27	0	2	59	301	474	3.00E-12	67
2	000501F	53	58	27	0	2	59	3183	3010	1.00E-11	66
2	000501F	53	58	27	0	2	59	30031	30204	9.00E-12	66
2	000542F	52	56	27	0	4	59	7230	7063	1.00E-10	63
2	000577F	48	44	23	0	41	84	35282	35413	3.00E-08	56
2	000826F	48	44	23	0	41	84	11352	11221	2.00E-08	56
2	000868F	66	87	29	1	1	87	4152	4409	7.00E-29	115
3	000868F	96	56	2	0	19	74	5021	5188	9.00E-36	115
3	000868F	100	24	0	0	71	94	5176	5247	9.00E-36	56
2	000868F	98	90	2	0	1	90	15490	15759	2.00E-55	192

3	000868F	100	76	0	0	19	94	16410	16637	3.00E-44	160
2	000868F	98	90	1	1	1	90	26421	26687	2.00E-54	189
3	000868F	100	76	0	0	19	94	27337	27564	2.00E-44	160
2	000952F_002	52	33	16	0	2	34	13410	13508	1.00E-10	40
2	000952F_002	46	48	26	0	38	85	13517	13660	1.00E-10	47

Table S5. Analysis of runs of homozygosity across 209 Alala genome contigs with lengths \geq 500kb. Each row indicates the number ROH segments observed per ROH length category. Data are provided under conditions of complete autozygosity and four heterozygosity thresholds. The autozygous fraction of the genome (fROH) was calculated by taking the sum of the number of ROH segments multiplied by the corresponding ROH length. The ROH segments are based on 100kb sliding window intervals.

ROH Length (kb)	Number of Sliding Windows per Length Category				
	Perfectly autozygous (0 het per 100 kb)	≤ 1 het per 100 kb	≤ 1 het per 50 kb	≤ 1 het per 25 kb	≤ 1 het per 10 kb
100	407	812	755	309	178
200	45	221	328	179	68
300	9	70	178	104	30
400	1	23	88	98	19
500	0	7	57	74	8
600	0	2	18	57	8
700	0	1	15	43	16
800	0	0	4	34	11
900	0	0	4	30	11
1,000	0	0	3	22	7
1,100	0	0	0	24	5
1,200	0	0	1	15	8
1,300	0	0	0	9	3
1,400	0	0	0	10	5
1,500	0	0	0	10	6
1,600	0	0	0	6	4
1,700	0	0	0	4	3
1,800	0	0	0	5	4
1,900	0	0	0	4	5
2,000	0	0	0	3	2
2,100	0	0	0	1	4
2,200	0	0	0	3	3
2,300	0	0	0	2	3
2,400	0	0	0	1	0
2,500	0	0	0	1	4
2,600	0	0	0	0	0
2,700	0	0	0	1	0
2,800	0	0	0	2	1
2,900	0	0	0	0	4
3,000	0	0	0	1	2
3,100	0	0	0	0	1
3,200	0	0	0	0	1

3,300	0	0	0	1	3
3,500	0	0	0	0	2
3,600	0	0	0	1	1
3,800	0	0	0	0	1
3,900	0	0	0	0	1
4,000	0	0	0	0	2
4,100	0	0	0	0	2
4,200	0	0	0	0	2
4,300	0	0	0	0	1
4,400	0	0	0	0	2
4,600	0	0	0	0	1
4,700	0	0	0	0	1
4,800	0	0	0	0	4
5,100	0	0	0	0	2
5,300	0	0	0	0	2
5,400	0	0	0	0	4
5,500	0	0	0	0	1
6,100	0	0	0	0	1
6,300	0	0	0	0	2
6,400	0	0	0	0	3
6,500	0	0	0	0	1
7,100	0	0	0	0	1
7,400	0	0	0	0	1
7,600	0	0	0	0	2
8,000	0	0	0	0	1
8,400	0	0	0	0	2
8,900	0	0	0	0	1
9,000	0	0	0	0	1
9,300	0	0	0	0	1
10,200	0	0	0	0	1
11,300	0	0	0	0	1
14,300	0	0	0	0	1
14,700	0	0	0	0	1
18,400	0	0	0	0	1
19,100	0	0	0	0	1
24,900	0	0	0	0	1
Numbers of ROHs	462	1136	1451	1054	480
fROH	5.1%	15.6%	28.2%	46.2%	58.4%

Figures

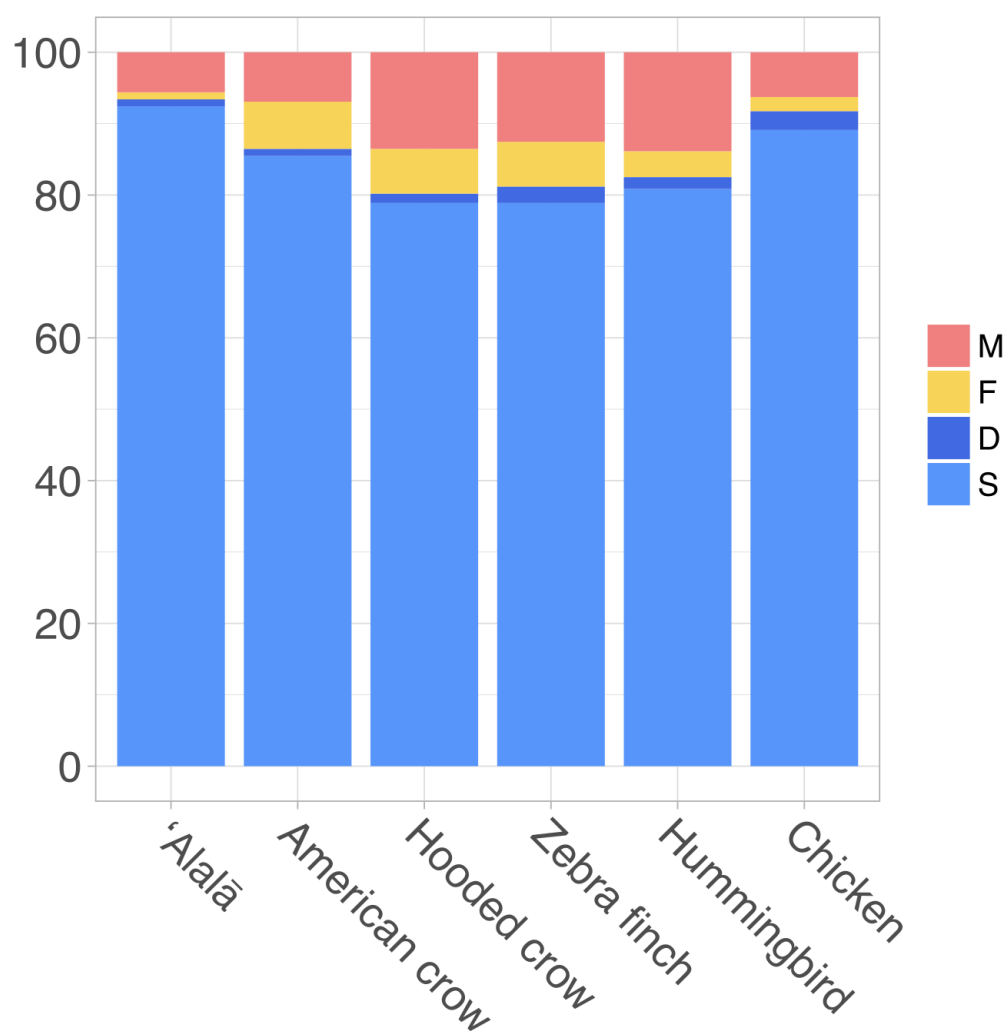


Figure 1. Genome assembly completeness assessed by the recovery of universal single-copy genes (BUSCOs). Percentages refer to complete genes that were found as single (S) or multiple copies (D), as well as fragmented (F) and missing (M) genes. Analyses were based on the BUSCO eukaryote dataset (n = 303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

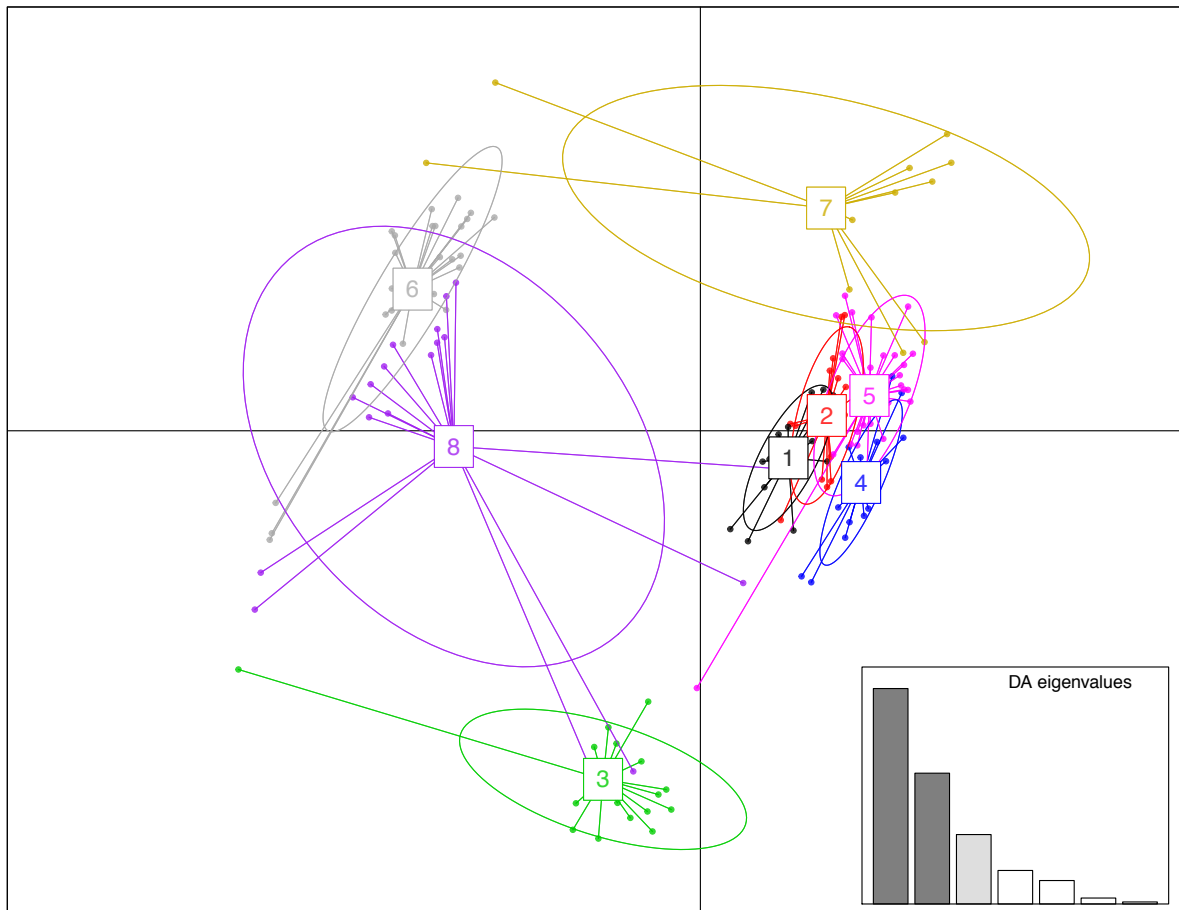


Figure 2. Discriminant Analysis of Principle Components (DAPC) scatterplot of the 8 MHC supertypes. 10 principle components (PC) and three discriminant functions (dimensions) were used to describe the relationship between the clusters. The scatterplot show only the first two discriminant functions ($d = 2$). The bottom graph displays the barplot of eigenvalues for the discriminant analysis. 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.

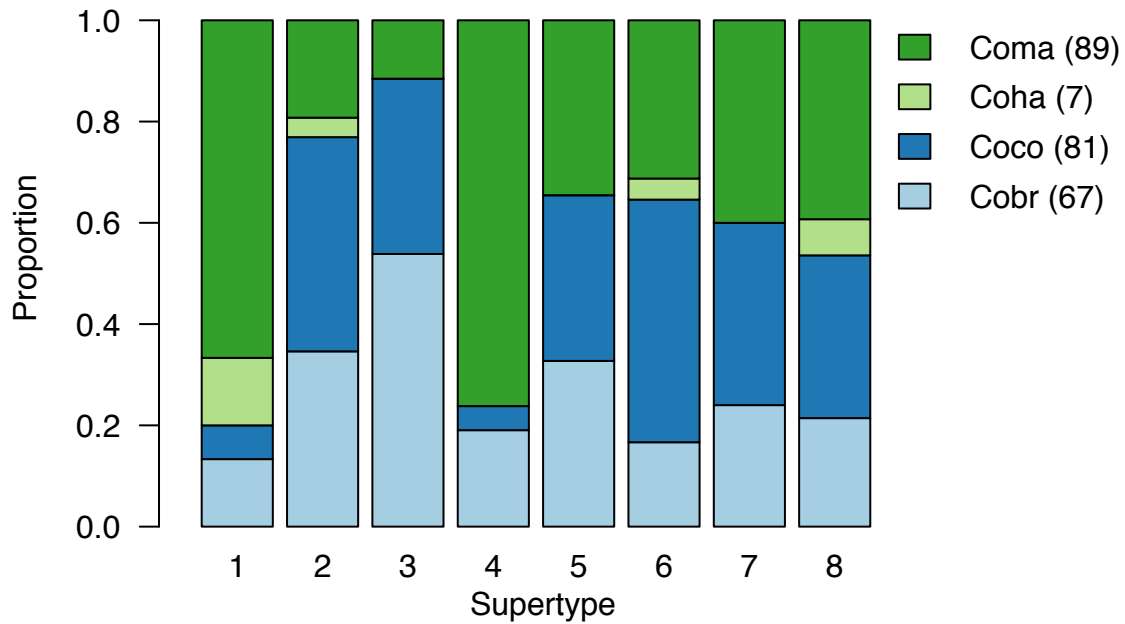


Figure 3. Stacked barplot indicating the representation of each corvid species within each MHC supertype. The three other corvid species were represented across all eight supertypes, while the Alala was represented by three supertypes. Note however that the Alala data were established from a single individual, while the other species' data represent 4-6 individuals per species [47]. In the legend, Coma = jungle crow; Coha = Alala; Coco = carrion crow; Cobr = American crow.

Tree scale: 0.1

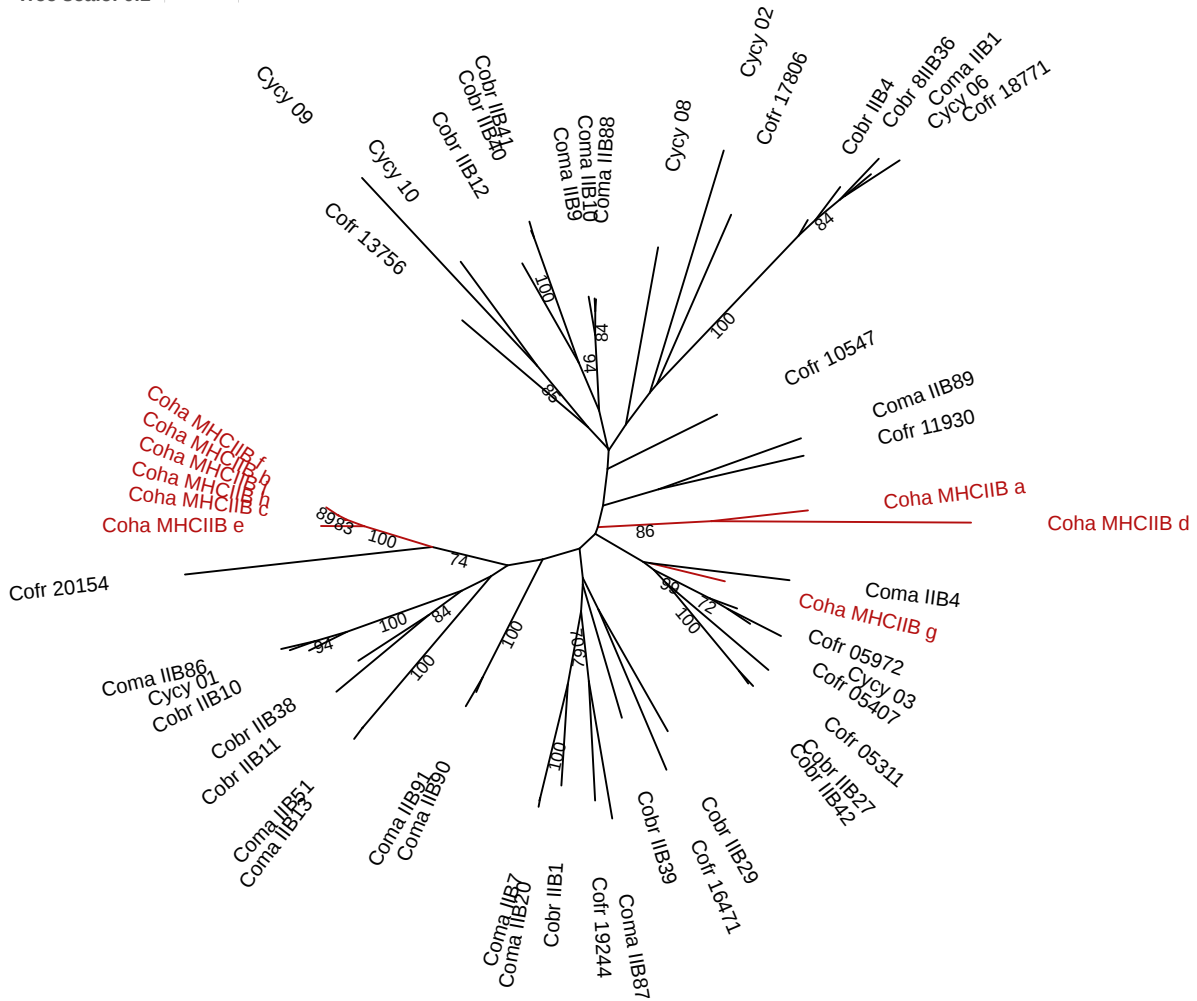


Figure S2. Unrooted maximum likelihood tree of corvid MHC class II beta genes, exon 2. Functional genes predicted from the Alala assembly are highlighted in red (prefix Coha). Other species represented include *C. brachyrhynchus* (American crow, Cobr), *C. macrorhynchus* (jungle crow, Coma), *C. frugilegus* (Asian rook, Cofr) and *Cyanopica cyanus* (azure-winged magpie, Cycy). All genes were obtained from one individual per species (Eimes et al. 2016 and pers. comm.). Confidence values are given for nodes with bootstrap support > 70%.

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.
2. Allendorf, F. W.; Hohenlohe, P. A. & Luikart, G. Genomics and the future of conservation genetics. *Nat. Rev. Genet.* **2010** 11, 697-709.
3. Grueber, C. E. Comparative genomics for biodiversity conservation. *Comput. Struct. Biotechnol. J.* **2015** 13, 370-375.
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.
5. Hayden, E. C. Technology: the \$1,000 genome. *Nature* **2014** 507, 294-295.
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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.* **2016** 113, 3585-3590.
12. 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.
13. 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.
14. 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.
15. 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.
16. Rutz, C.; Klump, B. C.; Komarczyk, L.; Leighton, R.; Kramer, J.; Wischnewski, S.; Sugawara, 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.
17. Culliney, S.; Pejchar, L.; Switzer, R. & Ruiz-Gutierrez, V. Seed dispersal by a captive corvid: the role of the 'Alala (*Corvus hawaiiensis*) in shaping Hawaii's plant communities. *Ecol. Appl.* **2012** 22, 1718-1732.
18. 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 'Alalā, the Hawaiian crow. *Conserv. Genet.* **2015**, 1-10.

19. U.S. Fish and Wildlife Service. Revised Recovery Plan for the `Alalā (*Corvus hawaiiensis*). Available online: URL https://www.fws.gov/pacific/ecoservices/documents/Alala_Revised_Recovery_Plan.pdf. **2009**, 1-104.
20. Fleischer, R. Genetic analysis of captive `Alala (*Corvus hawaiiensis*). Report to the U.S. Fish and Wildlife Service, Pacific Islands Fish and Wildlife Office, Honolulu, Hawaii. **2003**, 1-21.
21. Jarvi, S. I. & Bianchi, K. R. Genetic analyses of captive `Alalā (*Corvus hawaiiensis*) using AFLP analyses. US Geological Survey Open-File Report 2006-1349. **2006**, 1-40.
22. 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.
23. PacBio. Preparing > 30 kb SMRTbell™ Libraries Using the Megaruptor® Shearing and BluePippin™ Size-Selection System. Pacific Biosciences Unsupported Protocol. Available online: URL <http://www.pacb.com/wp-content/uploads/2015/09/Unsupported-Preparing-Greater-than-30kb-SMRTbell-Libraries-Megaruptor-Shearing.pdf>. **2015**.
24. 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.
25. Myers, E. W. The fragment assembly string graph. *Bioinformatics* **2005** 21, ii79-ii85.
26. Korfach, 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.
27. 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**, 3210-3212.
28. Consortium, I. C. G. S. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **2004** 432, 695.
29. 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.
30. 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.
31. 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.
32. Parra, G.; Bradnam, K. & Korf, I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics* **2007** 23, 1061-1067.
33. Bao, Z. & Eddy, S. R. Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res.* **2002** 12, 1269-1276.
34. Price, A. L.; Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **2005** 21, i351-i358.
35. Kohany, O.; Gentles, A. J.; Hankus, L. & Jurka, J. Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **2006** 7, 474.
36. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **1999** 27, 573.

37. 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.
38. Stanke, M.; Steinkamp, R.; Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **2004** 32, W309-W312.
39. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **2013** 30, 772-780.
40. 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.
41. Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **2014** 30, 1312-1313.
42. Trachtenberg, E.; Korber, B.; Sollars, C.; Kepler, T. B.; Hraber, P. T.; Hayes, E.; Funkhouser, R.; Fugate, M.; Theiler, J.; Hsu, Y. S.; Kunstman, K.; Wu, S.; Phair, J.; Erlich, H. & Wolinsky, S. Advantage of rare HLA supertype in HIV disease progression. *Nature Medicine* **2003** 9, 928-935.
43. 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 Evolutionary Biology* **2012** 12, 68, doi:10.1186/1471-2148-12-68.
44. Schwensow, N.; Fietz, J.; Dausmann, K. H. & Sommer, S. Neutral versus adaptive genetic variation in parasite resistance: importance of major histocompatibility complex superotypes in a free-ranging primate. *Heredity* **2007** 99, 265-277.
45. Doytchinova, I. A. & Flower, D. R. In silico identification of superotypes for class II MHCs. *The Journal of Immunology* **2005** 174, 7085-7095, doi:10.4049/jimmunol.174.11.7085.
46. 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. *Journal of Medicinal Chemistry* **1998** 41, 2481-2491, doi:10.1021/jm9700575.
47. 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.
48. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* **2004** 5, 113.
49. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004** 32, 1792-1797.
50. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Mentjies, P. & Drummond, A. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012** 28, 1647-1649.
51. Jombart, T. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* **2008** 24, 1403-1405, doi:10.1093/bioinformatics/btn129.
52. 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.
53. 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.
54. 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.

55. 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.
56. Toll-Riera, M.; Castelo, R.; Bellora, N. & Alba, M. M. Evolution of primate orphan proteins. *Biochem. Soc. Trans.* **2009**, 778-782.
57. 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 Genomics* **2008** 9, 62.
58. Alcaide, M. & Edwards, S. V. Molecular evolution of the Toll-like receptor multigene family in birds. *Mol. Biol. Evol.* **2011** 28, 1703-1715.
59. 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.
60. 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.
61. 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.
62. 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.
63. 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.
64. 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.
65. Klein, J.; Ono, H.; Klein, D. & O'hUigin, C. in *Progress in Immunology Vol. VIII* 137-143 (Springer, 1993).
66. 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.
67. 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.
68. 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.
69. Callicrate, T. E. *Population declines and genetic variation: Effects of serial bottlenecks*, University of Maryland, College Park, (2015).
70. 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 Genomics* **2012** 13, 391.
71. Howrigan, D. P.; Simonson, M. A. & Keller, M. C. Detecting autozygosity through runs of homozygosity: a comparison of three autozygosity detection algorithms. *BMC Genomics* **2011** 12, 460.
72. 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. *Journal of Animal Breeding and Genetics* **2013** 130, 286-293.

73. 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.
74. Koenig, W. D. Ecological and social factors affecting hatchability of eggs. *The Auk* **1982**, 526-536.
75. Shafer, A.; Peart, C. R.; Tusso, S.; Maayan, I.; Brelsford, A.; Wheat, C. W. & Wolf, J. B. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods Ecol. Evol.* **2017** 8, 907-917.
76. 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.
77. 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. *Bing du xue bao= Chinese journal of virology* **2011** 27, 18-25.
78. 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.
79. 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.
80. 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.
81. Karaffová, V.; Marcinková, E.; Bobíková, K.; Herich, R.; Revajová, V.; Stašová, D.; Kavulová, 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. (Praha)* **2017** 62, 89-97.
82. 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.
83. 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.
84. 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.