

RESOURCE ARTICLE

Leveraging an existing whole-genome resequencing population data set to characterize toll-like receptor gene diversity in a threatened bird

Molly Magid¹  | Jana R. Wold¹  | Roger Moraga²  | Ilina Cubrinovska¹ |
Dave M. Houston³ | Brett D. Gartrell⁴  | Tammy E. Steeves¹ 

¹School of Biological Sciences, University of Canterbury, Christchurch, New Zealand

²Tea Break Bioinformatics, Ltd, Palmerston North, New Zealand

³Department of Conservation, Biodiversity Group, Auckland, New Zealand

⁴Institute of Veterinary, Animal, and Biomedical Sciences, Wildbase, Massey University, Palmerston North, New Zealand

Correspondence

Molly Magid, School of Biological Sciences, University of Canterbury, 20 Kirkwood Avenue, Upper Riccarton, Christchurch NZ 8140, New Zealand.
Email: mollycmagid@gmail.com

Funding information

New Zealand Department of Conservation; New Zealand Ministry of Business, Innovation and Employment Endeavour Fund, Grant/Award Number: UOCX1602; Fulbright New Zealand; Ministry of Business, Innovation and Employment; University of Canterbury

Handling Editor: Paul A. Hohenlohe

Abstract

Species recovery programs are increasingly using genomic data to measure neutral genetic diversity and calculate metrics like relatedness. While these measures can inform conservation management, determining the mechanisms underlying inbreeding depression requires information about functional genes associated with adaptive or maladaptive traits. Toll-like receptors (TLRs) are one family of functional genes, which play a crucial role in recognition of pathogens and activation of the immune system. Previously, these genes have been analysed using species-specific primers and PCR. Here, we leverage an existing short-read reference genome, whole-genome resequencing population data set, and bioinformatic tools to characterize TLR gene diversity in captive and wild tchūriwat'/tūturuatu/shore plover (*Thinornis novaeseelandiae*), a threatened bird endemic to Aotearoa New Zealand. Our results show that TLR gene diversity in tchūriwat'/tūturuatu is low, and forms two distinct captive and wild genetic clusters. The bioinformatic approach presented here has broad applicability to other threatened species with existing genomic resources in Aotearoa New Zealand and beyond.

KEYWORDS

conservation genomics, immune genes, shore plover, toll-like receptors, whole-genome sequences

1 | INTRODUCTION

Advances in high-throughput sequencing technologies and bioinformatic tools are enabling scientists to generate and analyse whole-genome resequencing data (Auwera et al., 2013; Ekblom & Wolf, 2014). This progress has largely increased the measurement of neutral genetic diversity in nonmodel organisms often to inform the conservation management of threatened species (Allendorf et al., 2010; Angeloni et al., 2012; Primmer, 2009). However,

studies suggest that neutral genetic diversity may be a poor proxy for functional diversity (Grueber et al., 2015; Marsden et al., 2013; Sommer, 2005). Further, whereas neutral genetic diversity can be used to measure inbreeding and assess inbreeding depression, determining the genetic mechanisms that underlie inbreeding depression requires information about specific functional loci associated with adaptive or maladaptive traits (Kohn et al., 2006; Mable, 2019; Ouborg et al., 2010; Won et al., 2021). When whole-genome resequencing data and a reference genome for a population of interest

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial](https://creativecommons.org/licenses/by-nc/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

© 2022 The Authors. *Molecular Ecology Resources* published by John Wiley & Sons Ltd.

already exists, bioinformatic tools may be used to analyse these data to characterize functional genetic diversity (Hoelzel et al., 2019; Kohn et al., 2006).

There is a growing interest in characterizing the functional diversity of immune genes within intensively managed species to improve outcomes and reduce the burden of wildlife disease (Morris et al., 2015; Zhu et al., 2020). Low immune gene diversity increases susceptibility to a variety of pathogens (Spielman et al., 2004). Inbreeding leads to an overall decrease in genetic diversity at immune genes, loss of rare and potentially advantageous alleles, and a decreased ability to adapt to novel or rapidly evolving pathogens (Altizer et al., 2003; Spielman et al., 2004). Research from both wild and laboratory populations show that inbreeding contributes to an increased parasite load, a greater susceptibility to pathogens, a higher likelihood that individuals will act as disease reservoirs, and ultimately, higher rates of mortality due to disease (Acevedo-Whitehouse et al., 2003; Ross-Gillespie et al., 2007; Whitehorn et al., 2011; Whiteman et al., 2005). Further, increased immune gene diversity may help populations respond more rapidly to pathogens (Bonneaud et al., 2012; Davies et al., 2021).

Toll-like receptor (TLR) genes are an innate immune gene family involved in pathogen recognition and immune response (Acevedo-Whitehouse & Cunningham, 2006; Grueber et al., 2012; Vinkler & Albrecht, 2009). TLRs are an ancient part of the immune system, present in nearly all multicellular organisms (Singh et al., 2003). They are located either on the cell membrane or intracellularly on the membrane of lysosomes or endosomes within innate immune or somatic cells (Behzadi et al., 2021; Takeda & Akira, 2005). TLRs recognize conserved patterns of pathogens by binding residues, known as antigens, from pathogens that enter the body (Singh et al., 2003). Pathogen recognition by TLRs is necessary for the proper activation and direction of the adaptive immune response (Chen et al., 2021; Clark & Kupper, 2005; Kawasaki & Kawai, 2014; Pasare & Medzhitov, 2005; Xia et al., 2021). Studies have found that TLR gene diversity tends to be low in small, highly threatened populations compared to large populations of species of least concern (Alcaide & Edwards, 2011; Dalton et al., 2016; Knafler et al., 2017; Morrison et al., 2020). In previous studies, having higher TLR gene diversity has been associated with greater disease resistance, survival, and reproduction (Antonides et al., 2019; Bateson et al., 2016; Davies et al., 2021; Heng et al., 2011; Podlaszczuk et al., 2021; Quéméré et al., 2021). Having a robust TLR gene diversity within a population allows for selection of the most beneficial variant(s) to effectively deal with pathogens. Also, immune gene diversity may not correlate with neutral measures of genome-wide diversity, so directly characterizing TLR genes will explicitly reveal how diversity at innate immune genes are affected in highly threatened avian populations (Grueber et al., 2015; Hartmann et al., 2014; Marsden et al., 2013).

To date, TLR gene diversity has been mostly characterized using species-specific TLR primers and PCR protocols to perform SNP genotyping of individuals in a population (Allendorf et al., 2010; Dalton et al., 2016; Grueber et al., 2015; Lara et al., 2020; Morrison

et al., 2020). However, these approaches can be expensive and time-consuming. When whole-genome resequencing data and a reference genome for the species of interest already exists, researchers can use bioinformatic tools to analyse these existing resources to identify TLR genes and characterize TLR gene diversity (Hoelzel et al., 2019; Wang et al., 2005).

TLR genes may be more suited for identification using bioinformatic tools, in comparison to other immune genes like MHC, because they are relatively conserved and gene duplications are rare and well-defined (Grueber et al., 2015). The TLR protein is composed of three protein domains: the extracellular binding domain, transmembrane protein, and intracellular toll/interleukin 1 (TIR) signalling domain (Yilmaz et al., 2005). There is mostly conserved evolution and synonymous substitutions within TLR sequences, and the majority of variation is within the region coding for the extracellular binding protein, which is what comes into contact with antigens (Werling et al., 2009). The binding domain has a pattern of leucine-rich repeats, and so is also known as the leucine-rich repeat (LRR) domain (Alcaide & Edwards, 2011). Inserts of leucines within the LRR domain may impact pathogen recognition (Offord et al., 2010), and single nucleotide polymorphisms (SNPs) within this region can also affect the binding affinity of TLRs (Kestra et al., 2008; Matsushima et al., 2007). In contrast, the TIR signalling domain is largely conserved across the TLR family, and phylogenetically among related species (Beutler & Rehli, 2002; Narayanan & Park, 2015; Yilmaz et al., 2005). The implications of these protein structures are twofold: (1) the pattern of repeats within the LRR domain and conserved TIR domain are ideal for bioinformatic identification due to the similarity of sequences within taxonomically related species, and (2) the sequences within the LRR region are variable and may contain SNPs that are adaptive for the recognition of particular pathogens.

Advances in bioinformatic tools utilizing comparative genomics and sequence similarity comparison and the growth of online repositories and databases like NCBI with extensive genomic data that includes functional genes from a diversity of species, together facilitate the identification of functional genes within the genomes of nonmodel organisms (Feng et al., 2020; Grueber, 2015; Ren et al., 2021). However only a few studies to date have used bioinformatic tools to identify functional genes and characterize functional genetic diversity within a whole-genome resequencing population data set (Brandies et al., 2020; Zhang et al., 2014). Previous research on avian TLRs have used tools such as NCBI Basic Local Alignment Search Tool (BLAST) to design primer sequences to amplify these genes (Chávez-Treviño, 2017; Grueber et al., 2015; Morrison et al., 2020) and to compare sequences among closely-related species (Raven et al., 2017), so using established methods for primer design may also help in the identification of TLR genes (Yilmaz et al., 2005). A study on immune gene diversity in Tasmanian devils used bioinformatics to characterize diversity of several immune genes previously identified through a combination of transcriptome analysis and comparative genomes (Morris et al., 2015). Research using genomes from the B10K consortium highlights the efficacy of comparative genomics to identify orthologues and conserved regions using

bird species in the same taxonomic class (Feng et al., 2020; Zhang et al., 2014). Increasingly, immune gene databases are available to identify and annotate immune genes within specific taxonomic groups (Grueber, 2015; Mueller et al., 2020; Wong et al., 2011). These resources provide an avenue for characterizing immune gene diversity in nonmodel organisms.

Tchūriwat'/tūturuatu/shore plover (*Thinornis novaeseelandiae*) is a threatened shorebird endemic to Aotearoa New Zealand (Robertson et al., 2021). Tchūriwat'/tūturuatu is thought to have been widespread on the coastal areas of mainland Aotearoa, but populations decreased dramatically during the 19th century because of introduced mammalian predators (Davis, 1994). Tchūriwat'/tūturuatu was eventually extirpated from the mainland and the remaining wild population is now restricted to the Chatham Islands archipelago (Davis, 1994). The current wild population totals 234 birds and is mostly confined to a single predator-free island, Hokoreoro/Rangatira (Department of Conservation, Shore Plover Recovery Group herein DOC SPRG). A conservation breeding program was established between 1991–1996 with eggs brought from Hokoreoro/Rangatira and was augmented in 2003 by a single adult male from Western Reef (Dowding et al., 2005). At the time of this study, the captive population had not been augmented since 2003 and the number of individuals in the captive population was 41. Based on the pedigree of the captive population, individuals were highly related to one another (average mean kinship = 0.119) and mean inbreeding was relatively high (0.048) (SPRG, unpublished data).

Captive birds are also highly vulnerable to contracting avian pox (DOC SPRG). Avian pox is caused by avipoxvirus (APV), a large dsDNA virus with hundreds of strains that can infect a wide range of bird species (Bolte et al., 1999; Boyle, 2007). Infection with APV causes lesions on the body at the site where it enters the skin and on feather-free areas of the body. APV is generally transmitted through insect vectors, but if there is a break in epithelial integrity, it can also spread through contact with an infected bird or via contact with shared objects like feeders used by infected birds (Hansen, 1999). In rare instances, APV can also be transmitted through the inhalation of viral particles (Hansen, 1999). Aspects of captive breeding programmes like shared aviaries and resources make them particularly vulnerable for the spread of infection (van Riper & Forrester 2007).

In addition to being highly transmissible within a captive setting, avian pox infections in captive birds are often severe, can last for several months, and may never resolve (DOC SPRG). In contrast, wild tchūriwat'/tūturuatu contract mild avian pox infections which are most often cleared in 1–2 weeks (DOC SPRG). Longer infection periods for captive birds allow pox lesions more time to develop, making it more likely that birds will contract secondary bacterial infections which increase the severity of and rate of mortality from APV (Alley & Gartrell, 2019; Gartrell et al., 2002; Hansen, 1999; Welj & Tryland, 2011). While death from the virus remains low, the severity of infection is high. Also, chicks and juveniles are more likely to contract APV (DOC SPRG), and without a fully developed adaptive

immune system, the innate immune system may be especially important for young birds in fighting off poxvirus infections (Fellah et al., 2008; Palacios et al., 2009). Protecting juveniles from infection is important for the timely release of young birds into the wild (Dowding, 2013, DOC SPRG).

Ongoing research in an aligned project based on approximately 50K single nucleotide polymorphisms (SNPs) generated using a reduced-representation sequencing approach shows that wild and captive tchūriwat'/tūturuatu populations are genetically distinct (I. Cubrinovska, unpublished data). These SNPs reflect genome-wide diversity, but may also be indicative of differences at the gene level, including functional genes like TLRs that contribute to the immune response to poxvirus. If the wild population—which appears to be less susceptible to avian pox (DOC SPRG)—has higher immune gene diversity, then genetic rescue of the captive population through augmentation with wild individuals may improve future disease outcomes.

Sourcing individuals from genetically diverse or different populations is recognized as a way to improve both genome-wide diversity and immune gene diversity (Glasscock et al., 2021; Grueber et al., 2017; Heber et al., 2013; McLennan et al., 2020). Leveraging existing genomic resources and bioinformatic tools, we identify TLR genes and characterize TLR gene diversity in captive and wild tchūriwat'/tūturuatu. In addition to informing future research to determine whether a recent translocation of tchūriwat'/tūturuatu from the wild to captivity has increased TLR gene diversity and an associated immune response to the avipox infection, the bioinformatic approach presented here is broadly applicable to threatened birds with existing genomic resources.

2 | MATERIALS AND METHODS

2.1 | Sample extraction and sequencing

Originally generated for an aligned project (see Benefit-sharing statement), we used short-read sequence data for 66 tchūriwat'/tūturuatu (39 captive, 27 wild), including a wild male named Maui that was chosen for the tchūriwat'/tūturuatu reference genome.

Blood samples for captive birds were collected during routine health checks by Brett Gartrell and Isaac Conservation and Wildlife Trust staff. Blood samples for wild birds were collected as part of the 2018/2019 breeding season juvenile banding trip to Maung' Rē/Mangere and Hokoreoro/Rangatira in the Chathams Island archipelago. Blood was taken from juveniles and adults and stored in lysis buffer for shipment. All samples were subsequently stored at -0°C until extraction. High quantity and quality DNA was extracted using a tailored lithium chloride extraction method (Galla, 2019). Extractions were assessed for quality by running 2 μl of DNA on a 2% agarose gel. A Qubit 2.0 fluorometer (Fisher Scientific) was used for DNA quantification. Blood samples and DNA extractions are stored at the University of Canterbury on behalf of the Shore Plover Recovery Programme.

Libraries were prepared with the Nextera DNA Flex Library Prep Kit according to the manufacturer's specifications and sequenced across one lane of an Illumina Novaseq 6000 to achieve an average coverage depth of approximately 10× (Galla et al., 2020), excluding the individual used for the reference genome which was sequenced to an average depth of approximately 100×. Libraries for 21 individuals that had low depth of coverage were subsequently resequenced as above to achieve an average coverage depth of at least 10×.

2.2 | Reference genome assembly

FASTQC version 0.11.8 (FastQC, 2015) was used to evaluate the quality of the raw Illumina data and assess potential sample contamination. Initial read trimming was performed using TRIMGALORE version 0.6.6 (Kreuger, 2021), using pair-end mode, a minimum length of 54bp, and with the --nextseq two-colour chemistry option. A kmer abundance plot was created using JELLYFISH 2.3.0 (Marçais & Kingsford, 2011) with a kmer length of 31 prior to assembly to assess heterozygosity and contamination. The genome was assembled in two stages using two assemblers, MASURCA version 3.3.4 (Zimin et al., 2013), and MERACULOUS version 2.2.6 (Goltsman et al., 2017). While the assemblies with MaSuRCA give the best contiguity, the assembler can have problems with two-colour chemistry Illumina data when finding runs of low-quality bases or reads with a high proportion of "G" bases (De-Kayne et al., 2020). In those cases, the assembler tends to create regions where the consensus is simply a string of "G" bases and/or unknown "N" bases, as a result of both the error-correction module and the superread creation during the assembly process. Meraculous is much less prone to these artefacts due to their agnostic approach to error correction, which relies on coverage and bubble resolution in de Bruijn graphs, at the expense of contiguity when compared to MaSuRCA. Thus, both assemblers were combined.

In the first stage, MaSuRCA was used to create the main assembly, using the trimmed reads padded with low quality "N" bases to a uniform length of 150bp. Parameter adjustments include a grid batch size of 500,000,000, a Jellyfish hash size of 8,000,000,000, use of the Celera assembler for the final step, and the inclusion of scaffold gap closing. All other parameters were set to default for nonbacterial Illumina assemblies. For Meraculous, a kmer size of 61 with a minimum depth cutoff of five was selected, and the assembly was run on Diploid mode 1 (to create a single haploid reference). Finally the Meraculous assembly was used to correct the ambiguities and gaps in the MaSuRCA assembly by first aligning both genomes using Last version 980 (Kielbasa et al., 2011), filtering for contigs with low quality regions (due to the presence of poly-C/poly-G regions right next to a poly-N region) using awk, then using the merge_and_replace.pl Perl script available at <https://github.com/Lanilen/SemHelpers>. The hardware used for the assembly was a 12-core, 24-thread workstation with 128 Gb of RAM. Assembly took 5 days for MaSuRCA and 10 h for Meraculous.

2.3 | TLR identification and characterization overview

Figure 1 provides an overview of the methods used for TLR identification and characterization. We used a virtual machine (VM; 16 vCPUs, 9 TB of memory, 128GB of RAM) in the host Research Compute Cluster (RCC) at the University of Canterbury.

2.4 | TLR identification

2.4.1 | BLAST alignment

To identify TLR genes, we used NCBI Basic Local Alignment Search Tool (BLAST), given its prior use to design primer sequences that amplify TLR genes (Chávez-Treviño et al., 2017; Yilmaz et al., 2005) and previous comparisons of TLR sequences among closely-related avian species (Mueller et al., 2020; Raven et al., 2017). We used bird TLR nucleotide (BLASTn) and protein (tBLASTn) sequences from close relatives and/or species with high quality genome assemblies to search for similar nucleotide sequences within the reference genome for *tchūriwat'*/*tūturuatu* (Madden, 2013). Unlike other analyses that focused solely on the most variable LRR binding region, we chose to identify the entire sequence for each TLR, because it provided additional confirmation that the whole TLR was captured. For inclusion in our search through the *tchūriwat'*/*tūturuatu* reference genome, TLR sequences were either characterized for other species in the laboratory through targeted amplification and sequencing, or identified using the NCBI genome annotation pipeline (Table 1).

BLAST uses an algorithm to map input query sequences to the most similar region within the *tchūriwat'*/*tūturuatu* genome. If BLAST is able to map the input query sequence to a region or regions within the database, it outputs the region(s) as a list. The list is ordered by the expect value (*e*-value) of each alignment, which is the number of alignments with a similar score that are expected by chance. The lower the *e*-value of the alignment, the less likely that the alignment is due to chance and the more likely it is that the alignment reflects a biological similarity between the sequences (NCBI, 2020). In addition to the *e*-value, the percent query cover and percent identity of the alignment help the user to judge the quality of each alignment. The percent query cover is percent of the query sequence that is aligned to a database sequence, out of the whole length of the query sequence, regardless of the identity of the bases within a sequence. The percent identity is how many of the bases within the query sequence alignment match the bases within the database sequence, out of the whole length of the query sequence. The closer both of these percentages are to 100%, the greater the likelihood that the alignment is biologically relevant.

To further ensure that agreement between the query sequence and individual target sequences was not due to chance, two strategies were implemented: (1) a comparative approach

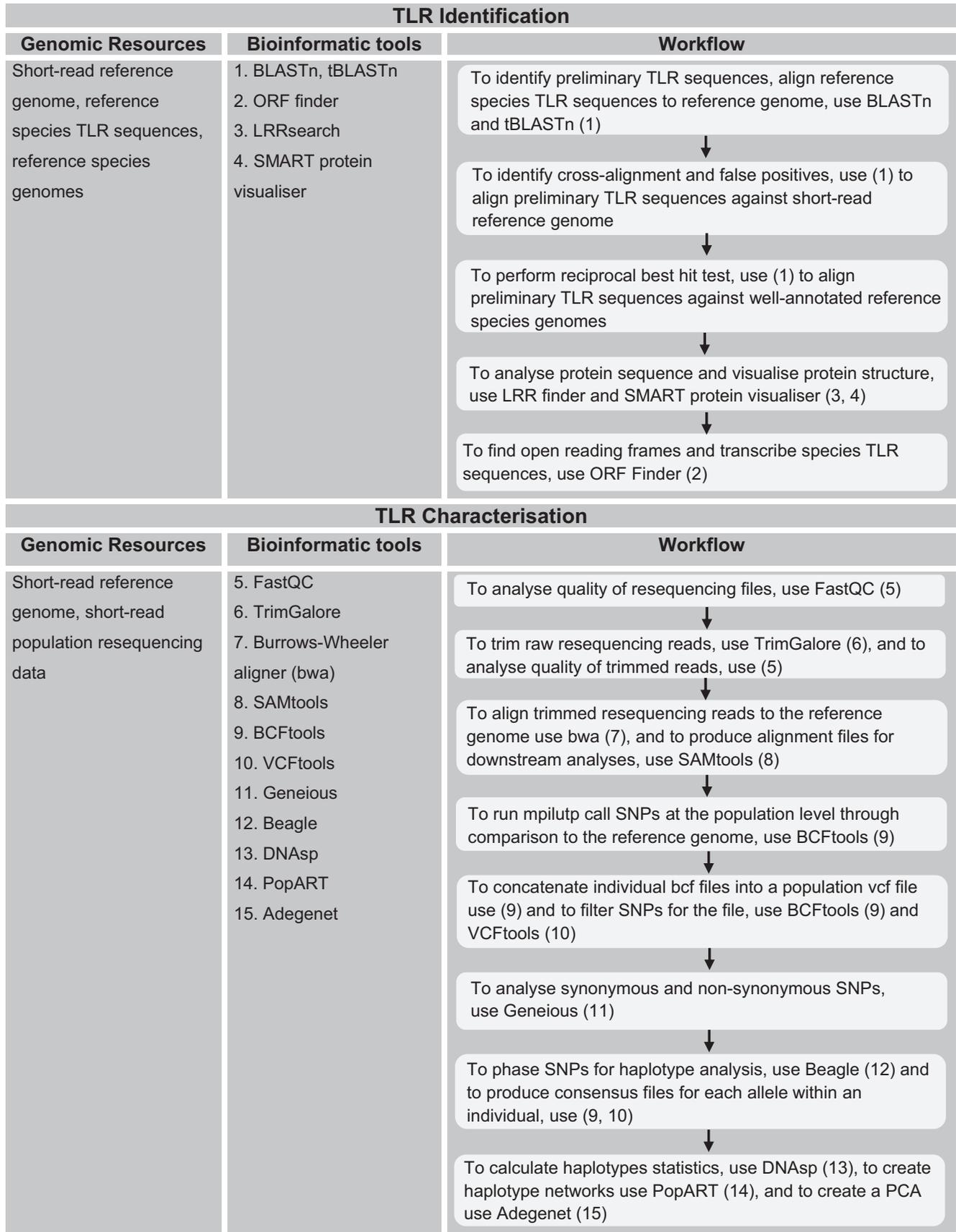


FIGURE 1 Workflow schematic to show how bioinformatic tools and genomic resources are used to identify and characterize TLR genes in threatened birds. See text for details. bcf, BIM collaboration format; LRR, leucine rich repeat; ORF, open reading frame; SNP, single nucleotide polymorphism; TLR, toll-like receptor. All remaining abbreviations refer to names of bioinformatic tools utilized

TABLE 1 Reference species for toll-like receptor (TLR) identification in tchūriwat'/tūturuatu, listed alongside the taxonomic group shared with tchūriwat'/tūturuatu. For each reference species, TLR sequences used for comparison were either identified through the use of the NCBI annotation pipeline^a or through amplification and sequencing of targeted TLRs in the laboratory

Reference TLR species	Shared taxonomic group	Method of TLR identification
Killdeer (<i>Charadrius vociferus</i>)	Family Charadriidae	NCBI genome annotation ^a
Black-headed Gull (<i>Chroicocephalus ridibundus</i>)	Order Charadriiformes	Targeted amplification and sequencing ^b
Chicken (<i>Gallus gallus</i>)	Class Aves	Targeted amplification and sequencing ^c
Zebra Finch (<i>Taeniopygia guttata</i>)	Class Aves	NCBI genome annotation excluding targeted amplification and sequencing of TLR4 ^d

^aThe NCBI Eukaryotic Genome Annotation Pipeline, https://www.ncbi.nlm.nih.gov/genome/annotation_euk/process/.

^bPodlaszczuk et al. (2021).

^cYilmaz et al. (2005).

^dVinkler et al. (2009).

seeking agreement among multiple species' TLR sequences aligning to a specific region in the tchūriwat'/tūturuatu genome; and (2) the reciprocal best hit (RBH) test. In the first instance, if the same region in the tchūriwat'/tūturuatu genome was being mapped to with the same type of TLR gene from different bird species, there was greater confidence in that region. In the second, the RBH test is an approach that has been used within comparative genomics to confirm the identity of orthologs in nonmodel species using well-annotated reference genomes (Kristensen et al., 2011). This limited the test to just chicken and zebra finch genomes, since both are high-quality genome assemblies and are well-annotated. This test involves taking the "best hit" region identified within tchūriwat'/tūturuatu using the TLR sequence from a reference species and BLASTing it against the genome of that reference species (Irizarry et al., 2016). If the "best hit" of this BLAST search matches the original reference TLR, it provides greater support that the genes are orthologs of one another (Kristensen et al., 2011). Each preliminary TLR sequence identified in tchūriwat'/tūturuatu passed the RBH test, mapping to the original chicken and zebra finch reference TLR sequences within each genome. This helped to confirm not only that a TLR gene was likely identified, but also that the specific TLR gene of interest had been located (i.e., a BLAST search with preliminary TLR3 sequence brought up only the TLR3 sequence in reference species).

After conducting these tests of gene alignment, each preliminary TLR sequence was BLASTed against the remainder of the tchūriwat'/tūturuatu genome to find whether there was alignment of this sequence to other regions within the genome. This test was conducted to determine whether there were nontarget sequences within the tchūriwat'/tūturuatu genome that may align to these TLR sequences during the process of whole-genome alignment. There was little cross-alignment of one type of TLR mapping to other TLR genes (e.g., TLR1A reference aligning to the TLR3 region). When this did happen, it was often within pairs of duplicated TLRs (TLR1A & TLR1B, TLR2A & TLR2B), and it was only a partial alignment (at most 50%), so these cross alignments were easy to distinguish from the

true alignment. False positives were also investigated within the genome, to find out whether there were TLR sequences aligning to other, non-TLR regions. There were no false positive sequences within the tchūriwat'/tūturuatu genome. Based on these tests, the Burrows-Wheeler Aligner version 0.7.17 (bwa) (Li & Durbin, 2009) with a default maximum mismatch value of 4% for read alignments (allowing only 4% of bases to differ in identity within sequence alignments) would be sufficient to prevent the cross-alignment of duplicated TLRs.

2.4.2 | Protein analysis

Once the preliminary TLR regions passed these quality control measures, each sequence was entered into the NCBI Open Reading Frame (ORF) finder (<https://www.ncbi.nlm.nih.gov/orffinder>). This tool searches for reading frames within the query DNA sequence and transcribes it into a protein sequence with each possible reading frame. The result is a graphic showing the protein sequence resulting from each reading frame that is used for transcription (Wheeler et al., 2003). Then we used BLASTp searches with the resultant protein as the input sequence and the BLAST protein database as comparison sequences. When the correct reading frame for the protein was identified, the search would bring up the reference TLR protein sequences that we had originally aligned with.

Two additional web applications were used to examine TLR protein sequences and investigate the protein products of TLR genes. The first was LRRfinder (Offord et al., 2010), which uses a database of toll-like receptor TLR sequences acquired from NCBI to identify the LRR regions within a TLR (LRRs, LRRNT, LRRCT, transmembrane protein, TIR signalling domain). The search tool focuses on predicting potential LRR regions, because this region is the most variable and may affect the binding specificity and affinity of TLRs (Keestra et al., 2008; Matsushima et al., 2007). All searches revealed multiple LRR regions within the sequences, a transmembrane protein region in almost all sequences, and a TIR domain in all sequences. Also the LRR identifications matched the

type of TLR gene they were from (i.e., all LRRs for TLR3 matched TLR3 LRRs in reference species).

To further visualize these protein products, we used the simple modular architecture research tool (SMART version 9.0) (Letunic et al., 2021; Schultz et al., 1998). SMART contains a protein database that uses markov modes to identify protein domains within an input protein sequence by calculating the expected value (e-value) SWise score, which is the output of the established SWise protein search algorithm (Birney et al., 1996), for each alignment between the query sequence and sequences in the SMART database (Letunic et al., 2021; Schultz et al., 1998). The protein domains assigned with low e-values are less likely to be assigned by chance, and at a predetermined e-value threshold, the protein domain will appear as a visual block on a 2D schematic of the protein. This visualization revealed whether we had captured all expected protein domains within the TLR protein sequences and had not missed parts of the sequence. This also facilitated a visual comparison between the tchūriwat/tūturuatu TLR protein schematics to the TLR SMART protein schematics in chicken (Temperley et al., 2008).

2.5 | TLR characterization

2.5.1 | Population resequencing alignment and analysis

Once TLR regions were identified within the genome, whole-genome alignment of population resequencing data to the tchūriwat/tūturuatu reference genome was performed to characterize TLR SNP diversity. To assess read quality, FASTQC version 0.11.9 (FastQC, 2015) was run to identify low quality regions and baseline quality scores, and then TRIM GALORE version 0.6.5 (Kreuger, 2021) was used to trim Illumina paired end 150bp reads and remove low quality reads. The two-colour chemistry option was chosen, which supports trimming and removal of low-quality sequences of non-G bases and improves the removal of low quality G's by removing them regardless of their quality (Kreuger, 2021).

Trimmed reads were aligned to the reference genome with BWA version 0.7.17 (Li & Durbin, 2009). SAMTOOLS version 1.10. (Li et al., 2009) was used to sort BAM files prior to SNP discovery. A custom perl script ("split_bamfiles_tasks.pl") (Moraga, 2018) was used to split the bam files into chunks that could be processed more quickly with BCFTOOLS version 1.11 (Li et al., 2009). BCFTOOLS mpileup was run with annotations GT, PL, DP, SP, ADF, ADR, AD to allow for downstream filtering. VCFTOOLS version 0.1.16 (Danecek et al., 2011) was used to filter the data set as follows: minor allele frequency (maf) ≥ 0.05 , Phred-score (quality) >20 , max-missingness = 0.90, minimum depth >5 , minGQ >10 , and maximum depth <200 . BCFTOOLS was also used to filter the vcf for strand bias using the parameter of strand-bias adjusted Phred-score <60 . Hardy-Weinberg equilibrium filtering was not applied because two assumptions are likely to be violated in this study (no selection and random mating). In addition, VCFTOOLS was used

to examine the site and individual depth and missingness and to remove individuals with low depth and high missingness from further analysis ($n = 1$ captive individual). Final SNPs were analysed in Geneious Prime 2020 (<https://www.geneious.com/>) to determine whether they made synonymous or nonsynonymous changes to TLR protein products. Nonsynonymous SNPs were further investigated in Geneious to analyse the amino acid change that resulted from each SNP. Then, a preliminary analysis of nonsynonymous SNPs was conducted to determine which physiochemical changes that might occur due to the change in amino acid identity, using the well-defined physicochemical attributes of charge and polarity.

BEAGLE version 5.2 (Browning et al., 2021) was used with default settings to phase haplotypes for both populations. Each TLR gene was phased separately by phasing along the whole contig that contained that TLR gene. For each individual, genotype sequences were generated using BCFTOOLS. To increase quality and decrease errors within haplotypes, sites with low depth and high strand bias were set to missing using BCFTOOLS +SETGT with options -i 'FORMAT/DP<5' and 'FORMAT/SP>60'. Then, the filtered VCF was used to produce TLR consensus sequences for each individual in the population. SAMTOOLS FAIDX was used to target each TLR region in the genome, and BCFTOOLS CONSENSUS was used with parameters -M to output any missing genotypes as "N", and -H 1plu and 2plu to produce both phased haplotypes for an individual. If individuals had any sites within the TLR sequence where they were missing a genotype, they were not included in the haplotype analysis for that TLR gene. While the accepted standard for SNP analysis is that nonsynonymous SNPs have significant influence on the functionality of a protein, synonymous SNPs may also change protein function (Sauna & Kimchi-Sarfaty, 2011), and within TLR genes, synonymous SNPs have been associated with disease resistance and immune response (Cho et al., 2013; Junjie et al., 2012). Given this, both synonymous and nonsynonymous SNPs were used in the construction of haplotypes.

The resulting TLR consensus sequence for each individual was output into a population mega file for either the captive or wild population, and the process was repeated for every polymorphic TLR gene. The resulting mega files were imported into DNASP v. 6.12 (Rozas, 2017) and were analysed to calculate haplotype diversity, nucleotide diversity (π), and Tajima's D. Haplotype nexus files for each TLR were also created in DNAsp and imported into POPART (Leigh & Bryant, 2015) to construct minimum-joining haplotype networks. We also visualized differences in TLR haplotypes between populations with a PCA. Adegnet (Jombart, 2008) was used to construct the PCA with scaled mean haplotype frequencies for all TLR genes, using code adapted from the introductory tutorial to adegenet (Jombart, 2015).

3 | RESULTS

Nine TLRs were identified within the tchūriwat/tūturuatu genome: TLR1A, TLR1B, TLR2A, TLR2B, TLR3, TLR4, TLR5, TLR7, and TLR21. A partial sequence of the TLR15 gene was identified within the

genome, but was not included in the analysis. The protein schematic structure for each TLR identified is shown in Figure 2.

There were 28 SNPs in the TLR genes of the captive tchūriwat/tūturuatu population ($n = 38$) (Table 2). The SNPs were unevenly distributed among TLR genes, with two that were monomorphic (TLR2B, TLR 21), three with one SNP (TLR1A, TLR1B, TLR2A), one with two SNPs (TLR4), one with three SNPs (TLR3), one with 7 SNPs (TLR5), and one with 13 SNPs (TLR7). Out of the total SNPs, half (14) were synonymous, and the remaining half (14) were nonsynonymous. The majority of SNPs (22) were within the LRR binding domain. The majority of nonsynonymous SNPs (12) were also located in the LRR region and half of them (7) caused a physicochemical change due to the change in identity of the amino acid (Table 3). A total of 24 haplotypes were observed across all TLR genes in the captive population, with the TLR genes that have higher SNP diversity also having a higher number of inferred haplotypes (Table 2). Tajima's D was nonsignificant for most loci, except for TLR5 and TLR7, which were both significantly positive ($p < .01$).

A total of 29 SNPs were observed in the TLR genes of the wild tchūriwat/tūturuatu population ($n = 26$) (Table 2). There were two monomorphic genes (TLR2B, TLR 21), three with one SNP each (TLR1B, TLR2A, TLR4), one with two SNPs (TLR3), one with three

SNPs (TLR1A), one with eight SNPs (TLR5), and one with 13 SNPs (TLR7). Out of the total SNPs, slightly more than half (15) were synonymous, and the remaining 14 were nonsynonymous. The majority of SNPs (22) were within the LRR binding domain. The majority of nonsynonymous SNPs (12) were also located in the LRR region and half of them (7) caused a physicochemical change due to the change in identity of the amino acid (Table 3). A total of 24 haplotypes were observed across all TLR genes in the wild population (Table 2). Tajima's D was nonsignificant for most loci, except for TLR5 and TLR7, which were both significantly positive ($p < .01$).

The wild and captive populations of tchūriwat/tūturuatu both had low overall TLR SNP and haplotype diversity. Each had a similar ratio of nonsynonymous to synonymous SNPs. Across all loci, the populations shared 25 SNPs and 19 haplotypes (Table 2). The captive population had three private SNPs and five private haplotypes in four genes (TLR1A, TLR2A, TLR3, TLR4). The wild population had four private SNPs and five private haplotypes in four genes (TLR1A, TLR2B, TLR4, TLR5). Haplotype networks reveal there are closely related haplotypes for each TLR gene, and that many of the TLR haplotypes shared between captive and wild tchūriwat/tūturuatu occur at different frequencies with each population (Figure S1). The PCA reveals two relatively distinct genetic clusters corresponding to wild and captive populations (Figure 3).

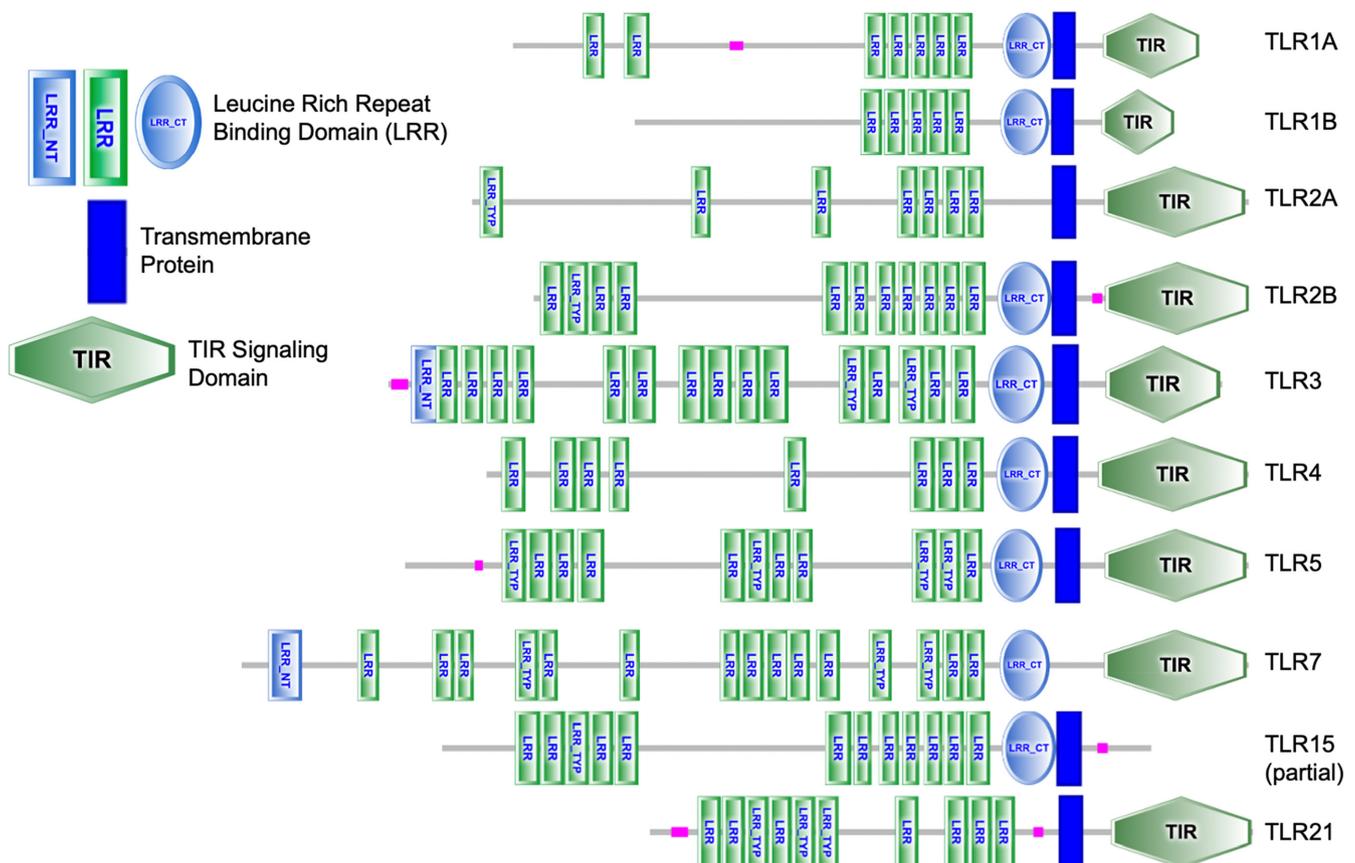


FIGURE 2 SMART protein schematics for each tchūriwat/tūturuatu toll-like receptor identified (available via licence: CC BY 2.0). Key shows the visual representation for each protein domain. The pink boxes are areas of low compositional complexity

TABLE 2 Tchūriwat'/tūturuatu toll-like receptor gene diversity statistics and comparisons for captive ($n = 38$) and wild ($n = 26$) tchūriwat'/tūturuatu populations

Locus	Length bp (aa)	SNPS (syn:Nsyn)			LRR SNPS (syn:Nsyn)		
		Captive	Wild	Shared	Captive	Wild	Shared
TLR1A	2289 (761)	1 (0:1)	3 (2:1)	1 (0:1)	1 (0:1)	3 (2:1)	1 (0:1)
TLR1B	1857 (592)	1 (1:0)	1 (1:0)	1 (1:0)	1 (1:0)	1 (1:0)	1 (1:0)
TLR2A	2423 (806)	1 (1:0)	0	0	1 (1:0)	0	0
TLR2B	2156 (717)	0	1 (1:0)	0	0	0	0
TLR3	2367 (783)	3 (3:0)	2 (2:0)	2 (2:0)	1 (1:0)	1 (1:0)	1 (1:0)
TLR4	2259 (752)	2 (1:1)	1 (1:0)	1 (0:1)	1 (0:1)	0	0
TLR5	2585 (860)	7 (4:3)	8 (4:4)	7 (4:3)	4 (2:2)	4 (2:2)	4 (2:2)
TLR7	3141 (1045)	13 (4:9)	13 (4:9)	13 (4:9)	13 (4:9)	13 (4:9)	13 (4:9)
TLR21	1649 (548)	0	0	0	0	0	0
Total Private		3 (2:1)	4 (3:1)	-	2 (1:1)	2 (2:0)	-
Total		28 (14:14)	29 (15:14)	25 (11:14)	22 (9:13)	22 (10:12)	20 (8:12)

Note: Asterisks denote significance at $p < .01$.

Abbreviations: bp(aa), base pairs (amino acids); syn:nsyn, ratio of synonymous to nonsynonymous SNPs.

TABLE 3 Amino acid analysis of nonsynonymous SNPs within TLR genes of captive and wild tchūriwat'/tūturuatu populations

TLR	Binding specificity ^a	AA site (codon)	TLR region	AA change	Physicochemical change	Population
TLR 1A	Bacteria, lipoproteins	503	LRR	Asp/His	Charge	Captive, Wild
TLR 4	LPS, gram negative bacteria	372	LRR	Asn/Asp	Charge	Captive
TLR 5	Flagellin	9	LRR	Phe/Leu	None	Captive, Wild
		326	LRR	Val/Ile	Charge	Wild
		664	TIR	His/Arg	None	Captive, Wild
		843	TIR	Lys/Glu	Charge	Captive, Wild
TLR7	ssRNA, virus	4	LRR	Ala/Pro	None	Captive, Wild
		79	LRR	Thr/Ile	Polarity	Captive, Wild
		100	LRR	Met/Leu	None	Captive, Wild
		411	LRR	Leu/Phe	None	Captive, Wild
		459	LRR	Gln/His	Charge	Captive, Wild
		465	LRR	Ala/Val	None	Captive, Wild
		467	LRR	Glu/Gly	Charge, Polarity	Captive, Wild
		469	LRR	Asn/Ser	None	Captive, Wild
755	LRR	Gln/Lys	Charge	Captive, Wild		

Abbreviations: AA, amino acid, standard abbreviations for amino acids used; LRR, leucine rich repeat binding domain; SNPs, single nucleotide polymorphisms; TIR, TIR signalling domain; TLR- toll-like receptor.

^aAlcaide and Edwards (2011).

4 | DISCUSSION

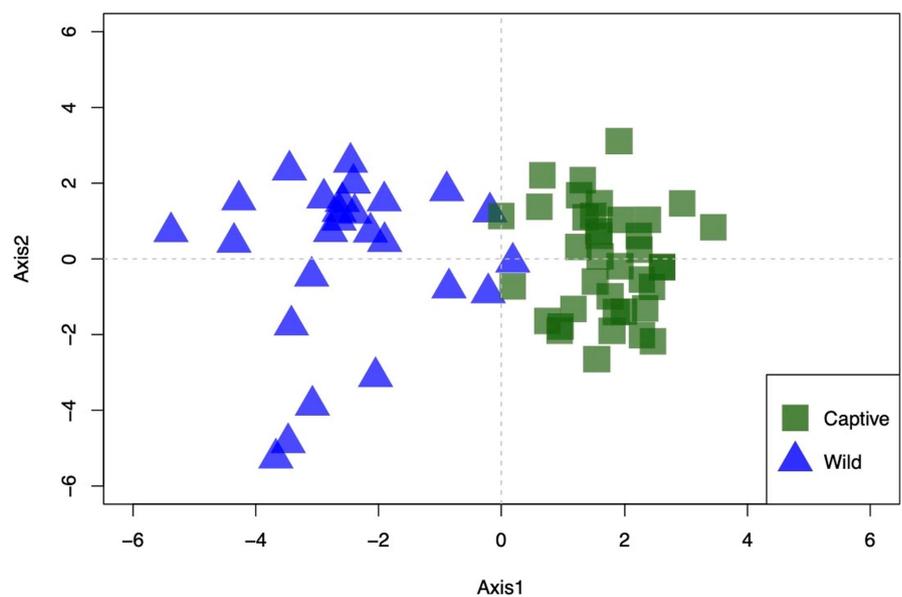
The bioinformatic approach developed here was successful for identifying TLR genes and characterizing TLR gene diversity in tchūriwat'/tūturuatu. This research provides a critical first step towards using TLR gene diversity to inform conservation action for a threatened Aotearoa New Zealand endemic bird. We were able to identify and characterize the same number of or more TLR genes as compared to similar studies done using amplicon sequencing

of TLRs within threatened species (Dalton et al., 2016; Grueber et al., 2015; Morrison et al., 2020). Further, as the generation of whole-genome resequencing data for threatened species becomes routine (Allendorf et al., 2010; Forcina & Leonard, 2020), identifying and characterizing TLRs using existing genomic resources will be more cost effective and more efficient than traditional amplicon sequencing.

All complete TLRs identified display the structure of TLR proteins in other species: a region of leucine rich repeats (the binding domain),

# inferred haplotypes (sample size)			Nucleotide diversity (π)		Haplotype diversity		Tajima's D	
Captive	Wild	Shared	Captive	Wild	Captive	Wild	Captive	Wild
2 (35)	3 (25)	1	0.00011	0.00042	0.248	0.483	0.24403	0.91554
2 (35)	2 (25)	2	0.00024	0.00023	0.437	0.429	1.37017	1.23557
2 (37)	1 (26)	1	0.00018	0	0.425	0	1.4606	0
1 (38)	2 (26)	1	0	0.0001	0	0.216	0	-0.0466
4 (38)	3 (26)	2	0.00045	0.00029	0.478	0.521	0.79912	0.46712
3 (38)	2 (24)	2	0.00027	0.00013	0.537	0.284	0.77496	0.35171
3 (35)	4 (23)	3	0.00126	0.00146	0.551	0.673	3.08884*	2.89296*
6 (38)	6 (26)	6	0.000206	0.00206	0.665	0.681	4.05164*	3.69096*
1 (38)	1 (26)	1	0	0	0	0	0	0
5	5	-						
24	24	19						

FIGURE 3 Principal component analysis genetic clustering of captive ($n = 38$) and wild ($n = 26$) tchūriwat/tūturuatu. Produced using scaled mean allele frequencies for each individual (see text for details)



then a transmembrane protein (carboxyl-terminal tail) where the TLR sits within the cell or lysosome membrane, and then the TIR (toll-like/interleukin receptor) signalling domain (Kannaki et al., 2010; Yilmaz et al., 2005). While LRRs vary in number between TLR sequences, the overall structure and approximate amount of LRRs within each type of TLR is similar to what is seen in chicken TLRs (Temperley et al., 2008). A partial sequence of TLR 15 was identified, but this sequence did not include the TIR signalling domain. The sequence ended prematurely at the end of a contig, so the scaffolding of the short-read genome may have prevented identification of the full gene. Morris et al. (2015) encountered a similar problem identifying immune gene sequences when they were fragmented or split between contigs in the genome assembly.

In nonpasserine avian species, there are most often 10 avian TLRs, eight of which are orthologous to other vertebrate TLRs (TLR1A/B, TLR2A/B, TLR3, TLR4, TLR5, TLR7), one that is orthologous to bony fish and *Xenopus* (TLR21), and one that is unique to reptiles and birds (TLR15) (Alcaide & Edwards, 2011; Grueber et al., 2014). Studies of avian TLR evolution show a pattern of both gene loss and duplication in these regions (Kannaki et al., 2010; Temperley et al., 2008; Velová et al., 2018). Recent research suggests there is also a duplication of TLR7 in some avian taxa. The duplicated TLR7 is thought to have a similar function, though with slight difference, and this is an area of ongoing research (Raven et al., 2017). To date, it has been found in Charadriiformes, Cuculiformes, Mesitornithiformes, and some Passeriiformes (Velová et al., 2018). Tūturuatu are within the order

Charadriiformes, and this species is likely to have a duplication of TLR7 and thus have 11 TLRs total (Raven et al., 2017; Velová et al., 2018).

However, we did not find evidence of the duplication of TLR7 within tchūriwat'/tūturuatu. After identifying the first gene coding for TLR7, this region was removed from the genome and the BLAST search with reference TLR7 sequences was repeated with the remainder of the genome. This second search was done to ensure that the reference TLR sequence would not align to the previously identified TLR7 region. There was no additional alignment to other sequences in the genome. It is possible that the duplication of TLR7 within the reference genome is not fully resolved, meaning the duplicated genes are collapsed into one region. Alternatively, the duplicated region may be incomplete with gaps in the sequence or it may be of low quality, so that it is either partially or fully cut out of the assembly. In the case of a low quality or partial assembly, the duplicated region would not be easily identified through the use of BLAST. In either case, only one TLR7 gene was identified. If the duplication does exist and is of low quality within the tchūriwat'/tūturuatu genome, this means the population resequencing data may not properly align to the low-quality TLR7 gene and may instead align to the other TLR7 gene. In this case, SNPs found within TLR7 may not be true SNPs, but instead may be artefacts of the duplicated TLR7 reads misaligned to this region.

Although it was not possible to make direct comparisons to all birds with TLR data, our findings are consistent with the lower TLR gene diversity found within small, isolated populations of other threatened species (African penguins (*Spheniscus demersus*); Dalton et al., 2016; Mohua (*Mohoua ochrocephala*) Grueber et al., 2015; Orange-bellied parrot (*Neophema chrysogaster*); Morrison et al., 2020).

Comparing captive and wild tchūriwat'/tūturuatu reveals that both populations have low TLR gene diversity. These findings are consistent with Dalton et al. (2016) which show that both captive and wild populations of African penguins (*Spheniscus demersus*) have low levels of TLR SNP diversity. Unlike tchūriwat'/tūturuatu, captive African penguin populations have a lower number of nonsynonymous SNPs compared to wild populations (Dalton et al., 2016). However, it is also important to note that there is a difference in sample size between the captive and wild tchūriwat'/tūturuatu populations, because the captive population ($n = 38$) has more samples than the wild population ($n = 26$). The inclusion of more wild individuals may reveal additional TLR diversity, especially if there are rare TLR SNPs or haplotypes in the wild population. Both captive and wild populations have nonsynonymous SNPs that mostly fall within the LRR binding region, which may have an effect on binding affinity and TLR response (Keestra et al., 2008; Matsushima et al., 2007). Additionally, some of the nonsynonymous SNPs do cause physiochemical changes in the amino acid sequence, but further research is needed to determine how these SNPs may influence functionality of TLR proteins.

In captive and wild populations, genetic diversity is unevenly distributed across TLR genes. Most TLR genes have a low number of SNPs and haplotypes. SNP diversity within TLR7 was the highest of all TLR genes. However, this result may not reflect the true diversity

of TLR7 within each population, if there is a TLR7 duplication that is not well resolved within the reference genome. Both captive and wild tchūriwat'/tūturuatu populations also had two TLRs (TLR5 and TLR7) with a value of Tajima's D that was significant and positive, indicating that rare alleles are scarce. The positive, significant values may reflect a rapid decrease in population size and/or balancing selection (Tajima, 1989). Since this result was the same for both wild and captive populations, it may reflect the overall history of species decline, given that the size and distribution of the species decreased dramatically in a short amount of time after mammalian predators were introduced (Davis, 1994) and has remained small and restricted for more than 100 years (DOC SPRG). Nevertheless, it is also possible that this result reflects balancing selection on one or both TLRs (but see below).

The finding that captive and wild tchūriwat'/tūturuatu fall into two relatively distinct genetic clusters based on TLR haplotype diversity is consistent with a PCA based on 50K SNPs (I. Cubrinovska, unpublished data). While most alleles are shared between captive and wild populations, the frequencies of these alleles differ between each population, and there are a small number of private alleles in both populations. Given that there has been little to no genetic exchange between the populations for decades, these differences can most likely be attributed to genetic drift. Previous research in other bird species has also found genetic differentiation in both neutral (microsatellites) and TLR loci, which generally indicates that genetic drift may be stronger than selection (Gonzalez-Quevedo et al., 2015; Grueber et al., 2013; Knafler et al., 2016). However, it is possible that selection at TLR loci may contribute to the genetic differentiation between populations (Grueber et al., 2013; Knafler et al., 2016, 2017). Further analysis is necessary to determine the impacts of genetic drift and selection on TLR gene diversity in tchūriwat'/tūturuatu. In the meantime, low overall TLR gene diversity combined with the genetic differentiation of captive and wild tchūriwat'/tūturuatu lends support to the recent decision to augment the captive population with birds from the wild. Future research will investigate how this augmentation affects TLR gene diversity in the captive population and whether it leads to increased immune response in captive individuals.

5 | CONCLUDING REMARKS

Below we highlight a few technical considerations that will influence the broad applicability of our bioinformatic approach to other bird species. If the species reference genome is relatively incomplete, it may not be sufficient for identifying all TLR genes. Undertaking strategies to improve completeness of the genome may help, including alignment of multiple assemblies to fill in sequencing gaps. It may also be necessary to undertake deliberate sequencing strategies to maximize depth and minimize missingness in resequencing data. For example, if preliminary screening of resequencing data shows low depth and high missingness, it may be necessary to invest in resequencing of relevant libraries. Also, if there are individuals who are particularly important, it may be beneficial to (re)sequence them

to greater depth. Further, we acknowledge that our approach was used to identify TLR genes and characterize TLR gene diversity in a threatened bird species, so it is unknown whether it may be useful in other species with more complex genomes or heterozygous populations. Nevertheless, advancing technologies such as long-read sequencing may provide opportunities to apply our approach to species with different levels of heterozygosity and genome complexity (Bayer et al., 2020; Pollard et al., 2018; Tettelin et al., 2005). Despite these caveats, we reiterate that if a whole-genome resequencing population data set is readily available for a species of interest, these genomic resources can be used to provide a first look at TLR genes and TLR gene diversity. Further, even if TLR gene identification and characterization is incomplete or wholly unsuccessful, a bioinformatic approach is likely to produce useful information about TLR regions that can inform targeted sequencing of TLR genes (e.g., even partial information about TLR sequences can inform primer design).

ACKNOWLEDGEMENTS

We acknowledge Hokotehi Moriori Trust and Ngāti Mutunga o Wharekauri, especially Susan Thorpe and Gail Amaru, for their guidance and support. We also thank the students of Kaingaroa and Te One for selecting Maui as the individual used for the tchūriwat'/tūturuatu reference genome. We are grateful to all members and associates of the the New Zealand Department of Conservation Shore Plover Recovery Group (formerly Specialist Group), including Troy Makan (DOC), Kate McInnes (DOC), Rose Collen (DOC), John Dowding (DM Consultants), Anne Richardson, and Leigh Percasky (Isaacs Conservation and Wildlife Trust), and Todd Jenkinson and Mireille Hicks (Pūkaha National Wildlife Centre). Many thanks also to Francois Bissey for access to the University of Canterbury's Research Compute Cluster, Natalie Forsdick and Levi Collier-Robinson for assistance with bioinformatic scripts, and to Stephanie Galla for outreach support on Rēkohu/Wharekauri. This research was funded by the New Zealand Ministry of Business, Innovation and Employment Endeavour Fund (UOCX1602 awarded to TES), the New Zealand Department of Conservation (contract awarded to TES) and Fulbright New Zealand (Fulbright US Graduate Award awarded to MM). Open access publishing facilitated by University of Canterbury, as part of the Wiley - University of Canterbury agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST

Author declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Scripts for TLR identification and resequencing pipeline have been made available on GitHub (https://github.com/UC-ConSERT/Magid_et_al.git). Tchūriwat'/tūturuatu are culturally significant to Moriori and Māori (the Indigenous Peoples of the Chatham Islands archipelago and Aotearoa New Zealand, respectively), all genomic data obtained from species like tchūriwat'/tūturuatu is culturally significant in its own right (Collier-Robinson et al., 2019). To ensure that the accessibility of these data are consistent with the CARE data principles

(Carroll et al., 2020, 2021), the short-read reference genome, unfiltered population vcf file, and associated metadata used in this study will be made available from a local genome browser (<http://www.ucconsert.org/data/>) on the recommendation of Hokotehi Moriori Trust and Ngāti Mutunga o Wharekauri Iwi Trust.

BENEFIT-SHARING STATEMENT

Tchūriwat'/tūturuatu/shore plover is an endemic shorebird classified as Nationally Critical according to the New Zealand Threat Classification System criteria (Robertson et al., 2021). This cocreated research is part of a long-standing research partnership between the Conservation Systematics and Evolution Research Team (ConSERT, including TES, MM, IC) and Aotearoa New Zealand's Department of Conservation Shore Plover Recovery Group, which is a team of conservation scientists and practitioners (including DH, BG, TES) that provide guidance regarding the development and execution of the Shore Plover Recovery Plan to ensure the recovery of tchūriwat'/tūturuatu in the wild. This research was also conducted in collaboration with Hokotehi Moriori Trust, Ngāti Mūtunga o Wharekauri Iwi Trust, and the local community on Rēkohu/Wharekauri in the Chatham Island archipelago. For example, following engagement with both trusts regarding the selection of an individual bird for the tchūriwat'/tūturuatu reference genome, IC and Stephanie Galla visited the two schools on Rēkohu/Wharekauri: They first visited Kaingaroa School (one class, Years 1–8) and asked students to name both candidate birds for the reference genome. Then they visited Te One School (three classes, Years 1–2, Years 3–5, Years 6–8), and asked students to vote on whether “Kina” or “Maui” should be used for the reference genome, and Maui won. MM also developed an immune system outreach module that highlights our ongoing research with tchūriwat'/tūturuatu that TES subsequently shared with imi and iwi, and the principals of both schools during a recent visit to Rēkohu/Wharekauri.

ORCID

Molly Magid  <https://orcid.org/0000-0002-6598-2530>
 Jana R. Wold  <https://orcid.org/0000-0001-8095-4282>
 Roger Moraga  <https://orcid.org/0000-0001-7806-1135>
 Brett D. Gartrell  <https://orcid.org/0000-0002-8062-9313>
 Tammy E. Steeves  <https://orcid.org/0000-0003-2112-5761>

REFERENCES

- Acevedo-Whitehouse, K., & Cunningham, A. (2006). Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution*, 21(8), 433–438. <https://doi.org/10.1016/j.tree.2006.05.010>.
- Acevedo-Whitehouse, K., Gulland, F., Greig, D., & Amos, W. (2003). Disease susceptibility in California sea lions. *Nature*, 422(6927), 35–35. <https://doi.org/10.1038/422035a>.
- Alcaide, M., & Edwards, S. V. (2011). Molecular evolution of the toll-like receptor multigene family in birds. *Molecular Biology and Evolution*, 28(5), 1703–1715. <https://doi.org/10.1093/molbev/msq351>
- Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, 11(10), 697–709. <https://doi.org/10.1038/nrg2844>

- Alley, M., & Gartrell, B. (2019). Wildlife diseases in New Zealand: Recent findings and future challenges. *New Zealand Veterinary Journal*, 67(1), 1–11. <https://doi.org/10.1080/00480169.2018.1520656>
- Altizer, S., Harvell, D., & Friedle, E. (2003). Rapid evolutionary dynamics and disease threats to biodiversity. *Trends in Ecology & Evolution*, 18(11), 589–596. <https://doi.org/10.1016/j.tree.2003.08.013>
- Angeloni, F., Wagemaker, N., Vergeer, P., & Ouborg, J. (2012). Genomic toolboxes for conservation biologists. *Evolutionary Applications*, 5(2), 130–143. <https://doi.org/10.1111/j.1752-4571.2011.00217.x>
- Antonides, J., Mathur, S., Sundaram, M., Ricklefs, R., & DeWoody, J. A. (2019). Immunogenetic response of the bananaquit in the face of malarial parasites. *BMC Evolutionary Biology*, 19(1), 107. <https://doi.org/10.1186/s12862-019-1435-y>
- Auweru, G. A., Carneiro, M. O., Hartl, C., Poplin, R., delAngel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ data to high-confidence variant calls: The genome analysis toolkit best practices pipeline. *Current Protocols in Bioinformatics*, 43(1), 11. <https://doi.org/10.1002/0471250953.bi1110s43>
- Bateson, Z. W., Hammerly, S. C., Johnson, J. A., Morrow, M. E., Whittingham, L. A., & Dunn, P. O. (2016). Specific alleles at immune genes, rather than genome-wide heterozygosity, are related to immunity and survival in the critically endangered Attwater's prairie-chicken. *Molecular Ecology*, 25(19), 4730–4744. <https://doi.org/10.1111/mec.13793>
- Bayer, P. E., Golicz, A. A., Scheben, A., Batley, J., & Edwards, D. (2020). Plant pan-genomes are the new reference. *Nature Plants*, 6(8), 914–920. <https://doi.org/10.1038/s41477-020-0733-0>
- Behzadi, P., García-Perdomo, H. A., & Karpiński, T. M. (2021). Toll-like receptors: General molecular and structural biology. *Journal of Immunology Research*, 2021, e9914854. <https://doi.org/10.1155/2021/9914854>
- Beutler, B., & Rehli, M. (2002). Evolution of the TIR, tolls and TLRs: Functional inferences from computational biology. In B. Beutler & H. Wagner (Eds.), *Toll-like receptor family members and their ligands* (pp. 1–21). Springer.
- Birney, E., Thompson, J. D., & Gibson, T. J. (1996). PairWise and SearchWise: Finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Research*, 24(14), 2730–2739. <https://doi.org/10.1093/nar/24.14.2730>
- Bolte, A. L., Meurer, J., & Kaleta, E. F. (1999). Avian host spectrum of avipoxviruses. *Avian Pathology*, 28(5), 415–432. <https://doi.org/10.1080/03079459994434>
- Bonneaud, C., Balenger, S. L., Zhang, J., Edwards, S. V., & Hill, G. E. (2012). Innate immunity and the evolution of resistance to an emerging infectious disease in a wild bird. *Molecular Ecology*, 21(11), 2628–2639. <https://doi.org/10.1111/j.1365-294X.2012.05551.x>
- Boyle, D. B. (2007). Genus *Avipoxvirus*. In A. A. Mercer, A. Schmidt, & O. Weber (Eds.), *Poxviruses. Birkhäuser Advances in Infectious Diseases*. Birkhäuser.
- Brandies, P. A., Wright, B. R., Hogg, C. J., Grueber, C. E., & Belov, K. (2020). Characterisation of reproductive gene diversity in the endangered Tasmanian devil. *Molecular Ecology Resources*, 21, 721–732. <https://doi.org/10.1111/1755-0998.13295>
- Browning, B. L., Tian, X., Zhou, Y., & Browning, S. R. (2021). Fast two-stage phasing of large-scale sequence data. *The American Journal of Human Genetics*, 108(10), 1880–1890. <https://doi.org/10.1016/j.ajhg.2021.08.005>
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materchera, S., Parsons, M., Raseroka, K., Rodríguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for indigenous data governance. *Data Science Journal*, 19, 43. <https://doi.org/10.5334/dsj-2020-043>
- Carroll, S. R., Herczog, E., Hudson, M., Russell, K., & Stall, S. (2021). Operationalizing the CARE and FAIR principles for indigenous data futures. *Scientific Data*, 8(1), 108. <https://doi.org/10.1038/s41597-021-00892-0>
- Chávez-Treviño, A., Canales-del-Castillo, R., Ruvalcaba-Ortega, I., Reséndez-Pérez, D., González-Rojas, J. I., Guzmán-Velasco, A., & Panjabi, A. O. (2017). Primer development for amplification of toll-like genes for the assessment of adaptive genetic diversity in vulnerable grassland bird species. *Conservation Genetics Resources*, 9(3), 385–387. <https://doi.org/10.1007/s12686-017-0690-8>
- Chen, A. Y.-A., Huang, C.-W., Liu, S.-H., Liu, A.-C., & Chung, H.-C. (2021). Single nucleotide polymorphisms of immunity-related genes and their effects on immunophenotypes in different pig breeds. *Genes*, 12(9), 1377. <https://doi.org/10.3390/genes12091377>
- Cho, P., Gelinis, L., Corbett, N. P., Tebbutt, S. J., Turvey, S. E., Fortuno, E. S., & Kollmann, T. R. (2013). Association of common single-nucleotide polymorphisms in innate immune genes with differences in TLR-induced cytokine production in neonates. *Genes & Immunity*, 14(4), 199–211. <https://doi.org/10.1038/gene.2013.5>
- Clark, R., & Kupper, T. (2005). Old meets new: The interaction between innate and adaptive immunity. *Journal of Investigative Dermatology*, 125(4), 629–637. <https://doi.org/10.1111/j.0022-202X.2005.23856.x>
- Collier-Robinson, L., Rayne, A., Rupene, M., Thoms, C., & Steeves, T. (2019). Embedding indigenous principles in genomic research of culturally significant species: A conservation genomics case study. *New Zealand Journal of Ecology*, 43(3), 1–9. <https://doi.org/10.20417/nzjecol.43.36>
- Dalton, D. L., Vermaak, E., Smit-Robinson, H. A., & Kotze, A. (2016). Lack of diversity at innate immunity toll-like receptor genes in the critically endangered white-winged flufftail (*Sarothrura ayresii*). *Scientific Reports*, 6(1), 1–8. <https://doi.org/10.1038/srep36757>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Davies, C., Taylor, M., Hammers, M., Burke, T., Komdeur, J., Dugdale, H., & Richardson, D. (2021). Contemporary evolution of the viral-sensing TLR3 gene in an isolated vertebrate population [Preprint]. Preprints. <https://doi.org/10.22541/au.160992924.47833640/v1>
- Davis, A. (1994). Status, distribution, and population trends of the New Zealand shore plover. *Notornis*, 41, 179–194.
- De-Kayne, R., Frei, D., Greenway, R., Mendes, S. L., Retel, C., & Feulner, P. G. D. (2020). Sequencing platform shifts provide opportunities but pose challenges for combining genomic data sets. *Molecular Ecology Resources*, 21(3), 653–660. <https://doi.org/10.1111/1755-0998.13309>
- Dowding, J. E., Collen, R., Davis, A. M., O'Connor, S. M., & Smith, M. H. (2005). Gains and losses in the New Zealand Shore Plover (*Thinornis novaeseelandiae*). *Recovery Programme 1993–2003, Volume: International Wader Studies 17: 36–42 Conference: Status and Conservation of Shorebirds in the East Asian-Australasian Flyway*.
- Dowding, J. E. (2013). Reducing the risk of extinction of a globally threatened shorebird: Translocations of the shore plover (*Thinornis novaeseelandiae*), 1990–2012. 15.
- Eklom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, 7(9), 1026–1042. <https://doi.org/10.1111/eva.12178>
- FastQC (2010). FastQC: A Quality Control Tool for High Throughput Sequence Data. In S. Andrews (Ed.), *Brabraham Bioinformatics*. (Version 0.11.8) [Online]. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>

- Fellah, J. S., Jaffredo, T., & Dunon, D. (2008). Development of the avian immune system. In K. A. Schat, B. Kaspers, & P. Kaiser (Eds.), *Avian immunology* (pp. 51–66). Elsevier.
- Feng, S., Stiller, J., & Deng, Y. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature*, *587*, 252–257. <https://doi.org/10.1038/s41586-020-2873-9>
- Forcina, G., & Leonard, J. A. (2020). Tools for monitoring genetic diversity in mammals: Past, present, and future. In J. Ortega & J. E. Maldonado (Eds.), *Conservation genetics in mammals: Integrative research using novel approaches* (pp. 13–27). Springer International Publishing.
- Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoepfner, M. P., Maloney, R. F., Richardson, A., Slater, L., Santure, A. W., & Steeves, T. E. (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: Implications for conservation breeding programmes worldwide. *Evolutionary Applications*, *13*(5), 991–1008. <https://doi.org/10.1111/eva.12916>
- Galla, S. J. (2019). *Conservation Genomic Management of Two Critically Endangered New Zealand Birds*. PhD Thesis. University of Canterbury School of Biological Sciences.
- Gartrell, B. D., Alley, M. R., Pauli, J., Collen, R., & Barlow, K. (2002). An outbreak of poxvirus infection in captive shore plovers. *Kokako*, *9*(2), 9–11.
- Glasscock, G. L., Grueber, C. E., Belov, K., & Hogg, C. J. (2021). Reducing the extinction risk of populations threatened by infectious diseases. *Diversity*, *13*(2), 63. <https://doi.org/10.3390/d13020063>
- Goltsman, E., Ho, I., & Rokhsar, D. (2017). *Meraculous-2D: Haplotype-sensitive assembly of highly heterozygous genomes*. arXiv. <https://doi.org/10.48550/ARXIV.1703.09852>
- Gonzalez-Quevedo, C., Spurgin, L. G., Illera, J. C., & Richardson, D. S. (2015). Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Molecular Ecology*, *24*(23), 5852–5863. <https://doi.org/10.1111/mec.13437>
- Grueber, C. E. (2015). Comparative genomics for biodiversity conservation. *Computational and Structural Biotechnology Journal*, *13*, 370–375. <https://doi.org/10.1016/j.csbj.2015.05.003>
- Grueber, C. E., Knafler, G. J., King, T. M., Senior, A. M., Grosser, S., Robertson, B., Weston, K. A., Brekke, P., Harris, C. L. W., & Jamieson, I. G. (2015). Toll-like receptor diversity in 10 threatened bird species: Relationship with microsatellite heterozygosity. *Conservation Genetics*, *16*(3), 595–611. <https://doi.org/10.1007/s10592-014-0685-x>
- Grueber, C. E., Sutton, J. T., Heber, S., Briskie, J. V., Jamieson, I. G., & Robertson, B. C. (2017). Reciprocal translocation of small numbers of inbred individuals rescues immunogenetic diversity. *Molecular Ecology*, *26*(10), 2660–2673. <https://doi.org/10.1111/mec.14063>
- Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2013). Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. *Molecular Ecology*, *22*(17), 4470–4482. <https://doi.org/10.1111/mec.12404>
- Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2014). Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PLoS One*, *9*(3), e89632. <https://doi.org/10.1371/journal.pone.0089632>
- Grueber, C. E., Wallis, G. P., King, T. M., & Jamieson, I. G. (2012). Variation at innate immunity toll-like receptor genes in a bottlenecked population of a New Zealand Robin. *PLoS One*, *7*(9), e45011. <https://doi.org/10.1371/journal.pone.0045011>
- Hansen, W. (1999). *Avian pox*. Report no. 1999–0001; information and technology report (pp. 163–169). USGS Publications Warehouse. <http://pubs.er.usgs.gov/publication/2001161>
- Hartmann, S. A., Schaefer, H. M., & Segelbacher, G. (2014). Genetic depletion at adaptive but not neutral loci in an endangered bird species. *Molecular Ecology*, *23*(23), 5712–5725. <https://doi.org/10.1111/mec.12975>
- Heber, S., Varsani, A., Kuhn, S., Girg, A., Kempnaers, B., & Briskie, J. (2013). The genetic rescue of two bottlenecked south Island robin populations using translocations of inbred donors. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1752), 20122228. <https://doi.org/10.1098/rspb.2012.2228>
- Heng, J., Su, J., Huang, T., Dong, J., & Chen, L. (2011). The polymorphism and haplotype of TLR3 gene in grass carp (*Ctenopharyngodon idella*) and their associations with susceptibility/resistance to grass carp reovirus. *Fish & Shellfish Immunology*, *30*(1), 45–50. <https://doi.org/10.1016/j.fsi.2010.09.004>
- Hoelzel, A. R., Bruford, M. W., & Fleischer, R. C. (2019). Conservation of adaptive potential and functional diversity. *Conservation Genetics*, *20*(1), 1–5. <https://doi.org/10.1007/s10592-019-01151-x>
- Irizarry, K. J. L., Bryant, D., Kalish, J., Eng, C., Schmidt, P. L., Barrett, G., & Barr, M. C. (2016). Integrating genomic data sets for knowledge discovery: An informed approach to Management of Captive Endangered Species. *International Journal of Genomics*, 2016, e2374610. <https://doi.org/10.1155/2016/2374610>
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T. (2015). *An introduction to adegenet 2.0.0* (p. 79). Imperial College London. MRC Centre for Outbreak Analysis and Modelling.
- Junjie, X., Songyao, J., Minmin, S., Yanyan, S., Baiyong, S., Xiaying, D., Jiabin, J., Xi, Z., & Hao, C. (2012). The association between toll-like receptor 2 single-nucleotide polymorphisms and hepatocellular carcinoma susceptibility. *BMC Cancer*, *12*, 57. <https://doi.org/10.1186/1471-2407-12-57>
- Kannaki, T. R., Reddy, M. R., Shanmugam, M., Verma, P. C., & Sharma, R. P. (2010). Chicken toll-like receptors and their role in immunity. *World's Poultry Science Journal*, *66*(4), 727–738. <https://doi.org/10.1017/S0043933910000693>
- Kawasaki, T., & Kawai, T. (2014). Toll-like receptor signaling pathways. *Frontiers in Immunology*, *5*, 461. <https://doi.org/10.3389/fimmu.2014.00461>
- Keestra, A. M., de Zoete, M. R., van Aubel, R. A. M. H., & van Putten, J. P. M. (2008). Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. *Molecular Immunology*, *45*(5), 1298–1307. <https://doi.org/10.1016/j.molimm.2007.09.013>
- Kielbasa, S. M., Wan, R., Sato, K., Horton, P., & Frith, M. C. (2011). Adaptive seeds tame genomic sequence comparison. *Genome Research*, *21*(3), 487–493. <https://doi.org/10.1101/gr.113985.110>
- Knafler, G., Grueber, C., Sutton, J., & Jamieson, I. (2017). Differential patterns of diversity at microsatellite, MHC, and TLR loci in bottlenecked south Island saddleback populations. *New Zealand Journal of Ecology*, *41*(1), 98–106. <https://doi.org/10.20417/nzjecol.41.8>
- Knafler, G. J., Ortiz-Catedral, L., Jackson, B., Varsani, A., Grueber, C. E., Robertson, B. C., & Jamieson, I. G. (2016). Comparison of beak and feather disease virus prevalence and immunity-associated genetic diversity over time in an island population of red-crowned parakeets. *Archives of Virology*, *161*(4), 811–820. <https://doi.org/10.1007/s00705-015-2717-3>
- Kohn, M. H., Murphy, W. J., Ostrander, E. A., & Wayne, R. K. (2006). Genomics and conservation genetics. *Trends in Ecology & Evolution*, *21*(11), 629–637. <https://doi.org/10.1016/j.tree.2006.08.001>
- Kreuger, F. (2021). Trim Galore (0.6.4) [Computer software]. GitHub repository, <https://github.com/FelixKreuger/TrimGalore>
- Kristensen, D. M., Wolf, Y. I., Mushegian, A. R., & Koonin, E. V. (2011). Computational methods for gene orthology inference. *Briefings in Bioinformatics*, *12*(5), 379–391. <https://doi.org/10.1093/bib/bbr030>
- Lara, C. E., Grueber, C. E., Holtmann, B., Santos, E. S. A., Johnson, S. L., Robertson, B. C., Castaño-Villa, G. J., Lagisz, M., & Nakagawa, S. (2020). Assessment of the dunnocks' introduction to New Zealand

- using innate immune-gene diversity. *Evolutionary Ecology*, 34, 803–820. <https://doi.org/10.1007/s10682-020-10070-0>
- Leigh, J. W., & Bryant, D. (2015). Popart: Full-feature software for haplotype network construction. *Methods in Ecology and Evolution*, 6(9), 1110–1116. <https://doi.org/10.1111/2041-210X.12410>
- Letunic, I., Khedkar, S., & Bork, P. (2021). SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Research*, 49(D1), D458–D460. <https://doi.org/10.1093/nar/gkaa937>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, 25(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Mable, B. K. (2019). Conservation of adaptive potential and functional diversity: Integrating old and new approaches. *Conservation Genetics*, 20(1), 89–100. <https://doi.org/10.1007/s10592-018-1129-9>
- Madden, T. (2013). The BLAST sequence analysis tool. In *The NCBI handbook [internet]* (2nd ed.). National Center for Biotechnology Information (US). <https://www.ncbi.nlm.nih.gov/books/NBK153387/>
- Marçais, G., & Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 27(6), 764–770. <https://doi.org/10.1093/bioinformatics/btr011>
- Marsden, C. D., Verberkmoes, H., Thomas, R., Wayne, R. K., & Mable, B. K. (2013). Pedigrees, MHC and microsatellites: An integrated approach for genetic management of captive African wild dogs (*Lycaon pictus*). *Conservation Genetics*, 14(1), 171–183. <https://doi.org/10.1007/s10592-012-0440-0>
- Matsushima, N., Tanaka, T., Enkhbayar, P., Mikami, T., Taga, M., Yamada, K., & Kuroki, Y. (2007). Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics*, 8(1), 124. <https://doi.org/10.1186/1471-2164-8-124>
- McLennan, E. A., Grueber, C. E., Wise, P., Belov, K., & Hogg, C. J. (2020). Mixing genetically differentiated populations successfully boosts diversity of an endangered carnivore. *Animal Conservation*, 23(6), 700–712. <https://doi.org/10.1111/acv.12589>
- Moraga, R. (2018). *SubSampler_SNPcaller*. Tea Break Bioinformatics. [Computer software] Available at: https://github.com/Lanilen/SubSampler_SNPcaller
- Morris, K. M., Wright, B., Grueber, C. E., Hogg, C., & Belov, K. (2015). Lack of genetic diversity across diverse immune genes in an endangered mammal, the Tasmanian devil (*Sarcophilus harrisii*). *Molecular Ecology*, 24(15), 3860–3872. <https://doi.org/10.1111/mec.13291>
- Morrison, C. E., Hogg, C. J., Gales, R., Johnson, R. N., & Grueber, C. E. (2020). Low innate immune-gene diversity in the critically endangered orange-bellied parrot (*Neophema chrysogaster*). *Emu - Austral Ornithology*, 120(1), 56–64. <https://doi.org/10.1080/01584197.2019.1686994>
- Mueller, R. C., Mallig, N., Smith, J., Eöery, L., Kuo, R. I., & Kraus, R. H. S. (2020). Avian immunome DB: An example of a user-friendly interface for extracting genetic information. *BMC Bioinformatics*, 21(1), 502. <https://doi.org/10.1186/s12859-020-03764-3>
- Narayanan, K. B., & Park, H. H. (2015). Toll/interleukin-1 receptor (TIR) domain-mediated cellular signaling pathways. *Apoptosis*, 20(2), 196–209. <https://doi.org/10.1007/s10495-014-1073-1>
- NCBI (2020). *BLAST results: Expect values, part 1*. The National Library of Medicine. <https://www.youtube.com/watch?v=ZN3RrXAe0uM>
- O'Connor, E. A., Westerdahl, H., Burri, R., & Edwards, S. V. (2019). Avian MHC evolution in the era of genomics: Phase 1.0. *Cells*, 8(10), 1152. <https://doi.org/10.3390/cells8101152>
- Offord, V., Coffey, T. J., & Werling, D. (2010). LRRfinder: A web application for the identification of leucine-rich repeats and an integrative toll-like receptor database. *Developmental & Comparative Immunology*, 34(10), 1035–1041. <https://doi.org/10.1016/j.dci.2010.05.004>
- Ouborg, N. J., Pertoldi, C., Loeschcke, V., Bijlsma, R. K., & Hedrick, P. W. (2010). Conservation genetics in transition to conservation genomics. *Trends in Genetics*, 26(4), 177–187. <https://doi.org/10.1016/j.tig.2010.01.001>
- Palacios, M. G., Cunnick, J. E., Vleck, D., & Vleck, C. M. (2009). Ontogeny of innate and adaptive immune defense components in free-living tree swallows, *Tachycineta bicolor*. *Developmental & Comparative Immunology*, 33(4), 456–463. <https://doi.org/10.1016/j.dci.2008.09.006>
- Pasare, C., & Medzhitov, R. (2005). Toll-like receptors: Linking innate and adaptive immunity. In S. Gupta, W. E. Paul, & R. Steinman (Eds.), *Mechanisms of lymphocyte activation and immune regulation X* (pp. 11–18). Springer US.
- Podlasczyc, P., Indykiewicz, P., Kamiński, M., & Minias, P. (2021). Physiological condition reflects polymorphism at the toll-like receptors in a colonial waterbird. *Ornithology*, 138, ukab052. <https://doi.org/10.1093/ornithology/ukab052>
- Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: Their purpose and place. *Human Molecular Genetics*, 27(R2), R234–R241. <https://doi.org/10.1093/hmg/ddy177>
- Primmer, C. R. (2009). From conservation genetics to conservation genomics. *Annals of the New York Academy of Sciences*, 1162(1), 357–368. <https://doi.org/10.1111/j.1749-6632.2009.04444.x>
- Quéméré, E., Hesseuauer, P., Galan, M., Fernandez, M., Merlet, J., Chaval, Y., Morellet, N., Verheyden, H., Gilot-Fromont, E., & Charbonnel, N. (2021). Pathogen-mediated selection favours the maintenance of innate immunity gene polymorphism in a widespread wild ungulate. *Journal of Evolutionary Biology*, 34(7), 1156–1166. <https://doi.org/10.1111/jeb.13876>
- Raven, N., Lisovski, S., Klaassen, M., Lo, N., Madsen, T., Ho, S. Y. W., & Ujvari, B. (2017). Purifying selection and concerted evolution of RNA-sensing toll-like receptors in migratory waders. *Infection, Genetics and Evolution*, 53, 135–145. <https://doi.org/10.1016/j.meegid.2017.05.012>
- Ren, Y., MacPhillamy, C., To, T.-H., Smith, T. P. L., Williams, J. L., & Low, W. Y. (2021). Adaptive selection signatures in river buffalo with emphasis on immune and major histocompatibility complex genes. *Genomics*, 113, 3599–3609. <https://doi.org/10.1016/j.ygeno.2021.08.021>
- Robertson, H. A., Baird, K. A., Elliott, G. P., Hitchmough, R. A., McArthur, N. J., Makan, T. D., O'Donnell, M. C., Sagar, P. M., Scofield, R. P., Taylor, G. A., & Michel, P. (2021). *Conservation status of birds in Aotearoa New Zealand, 2021. Report number: New Zealand classification series 36*. Affiliation: New Zealand Department of Conservation.
- Ross-Gillespie, A., O'Riain, M. J., & Keller, L. F. (2007). Viral epizootic reveals inbreeding depression in a habitually inbreeding mammal. *Evolution*, 61(9), 2268–2273. <https://doi.org/10.1111/j.1558-5646.2007.00177.x>
- Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299–3302. <https://doi.org/10.1093/molbev/msx248>
- Sauna, Z. E., & Kimchi-Sarfaty, C. (2011). Understanding the contribution of synonymous mutations to human disease. *Nature Reviews Genetics*, 12(10), 683–691. <https://doi.org/10.1038/nrg3051>
- Schultz, J., Milpetz, F., Bork, P., & Ponting, C. P. (1998). SMART, a simple modular architecture research tool: Identification of signaling domains. *Proceedings of the National Academy of Sciences of the United States of America*, 95(11), 5857–5864. <https://doi.org/10.1073/pnas.95.11.5857>
- Singh, B. P., Chauhan, R. S., & Singhal, L. K. (2003). Toll-like receptors and their role in innate immunity. *Current Science*, 85(8), 1156–1164.
- Sommer, S. (2005). The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Frontiers in Zoology*, 2(1), 1–18. <https://doi.org/10.1186/1742-9994-2-16>

- Spielman, D., Brook, B. W., Briscoe, D. A., & Frankham, R. (2004). Does inbreeding and loss of genetic diversity decrease disease resistance? *Conservation Genetics*, 5(4), 439–448. <https://doi.org/10.1023/b:coge.0000041030.76598.cd>.
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585–595.
- Takeda, K., & Akira, S. (2005). Toll-like receptors in innate immunity. *International Immunology*, 17(1), 1–14. <https://doi.org/10.1093/intimm/dxh186>
- Temperley, N. D., Berlin, S., Paton, I. R., Griffin, D. K., & Burt, D. W. (2008). Evolution of the chicken toll-like receptor gene family: A story of gene gain and gene loss. *BMC Genomics*, 9, 62. <https://doi.org/10.1186/1471-2164-9-62>
- Tettelin, H., Masignani, V., Cieslewicz, M. J., Donati, C., Medini, D., Ward, N. L., Angiuoli, S. V., Crabtree, J., Jones, A. L., Durkin, A. S., DeBoy, R. T., Davidsen, T. M., Mora, M., Scarselli, M., Ros, I. M. y, Peterson, J. D., Hauser, C. R., Sundaram, J. P., Nelson, W. C., ... Fraser, C. M. (2005). Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: Implications for the microbial "pan-genome". *Proceedings of the National Academy of Sciences of the United States of America*, 102(39), 13950. <https://doi.org/10.1073/pnas.0506758102>, 13955
- van Riper, C., & Forrester, D. J. (2007). Avian pox. In N. J. Thomas, D. B. Hunter, & C. T. Atkinson (Eds.), *Infectious diseases of wild birds* (pp. 131–176). John Wiley & Sons, Ltd.
- Velová, H., Gutowska-Ding, M. W., Burt, D. W., & Vinkler, M. (2018). Toll-like receptor evolution in birds: Gene duplication, pseudogenization, and diversifying selection. *Molecular Biology and Evolution*, 35(9), 2170–2184. <https://doi.org/10.1093/molbev/msy119>
- Vinkler, M., & Albrecht, T. (2009). The question waiting to be asked: Innate immunity receptors in the perspective of zoological research. *Folia Zoologica*, 58, 15–28.
- Vinkler, M., Bryjová, A., Albrecht, T., & Bryja, J. (2009). Identification of the first Toll-like receptor gene in passerine birds: TLR4 orthologue in zebra finch (*Taeniopygia guttata*). *Tissue Antigens*, 74(1), 32–41. <https://doi.org/10.1111/j.1399-0039.2009.01273.x>.
- Wang, X., Ishimori, N., Korstanje, R., Rollins, J., & Paigen, B. (2005). Identifying novel genes for atherosclerosis through mouse-human comparative genetics. *The American Journal of Human Genetics*, 77(1), 1–15. <https://doi.org/10.1086/431656>
- Weli, S. C., & Tryland, M. (2011). Avipoxviruses: Infection biology and their use as vaccine vectors. *Virology Journal*, 8(1), 49. <https://doi.org/10.1186/1743-422X-8-49>
- Werling, D., Jann, O. C., Offord, V., Glass, E. J., & Coffey, T. J. (2009). Variation matters: TLR structure and species-specific pathogen recognition. *Trends in Immunology*, 30(3), 124–130. <https://doi.org/10.1016/j.it.2008.12.001>
- Wheeler, D. L., Church, D. M., Federhen, S., Lash, A. E., Madden, T. L., Pontius, J. U., Schuler, G. D., Schriml, L. M., Sequeira, E., Tatusova, T. A., & Wagner, L. (2003). Database resources of the National Center for Biotechnology. *Nucleic Acids Research*, 31(1), 28–33.
- Whitehorn, P. R., Tinsley, M. C., Brown, M. J. F., Darvill, B., & Goulson, D. (2011). Genetic diversity, parasite prevalence and immunity in wild bumblebees. *Proceedings of the Royal Society B: Biological Sciences*, 278(1709), 1195–1202. <https://doi.org/10.1098/rspb.2010.1550>
- Whiteman, N. K., Matson, K. D., Bollmer, J. L., & Parker, P. G. (2005). Disease ecology in the Galápagos Hawk (*Buteo galapagoensis*): host genetic diversity, parasite load and natural antibodies. *Proceedings of the Royal Society B: Biological Sciences*, 273(1588), 797–804. <https://doi.org/10.1098/rspb.2005.3396>.
- Won, H., Jeon, H.-B., Kim, D.-Y., & Suk, H. Y. (2021). Differential patterns of diversity at neutral and adaptive loci in endangered *Rhodeus pseudosericeus* populations. *Scientific Reports*, 11(1), 15953. <https://doi.org/10.1038/s41598-021-95385-w>
- Wong, E. S., Papenfuss, A. T., & Belov, K. (2011). Immunome database for marsupials and monotremes. *BMC Immunology*, 12(1), 48. <https://doi.org/10.1186/1471-2172-12-48>
- Xia, P., Wu, Y., Lian, S., Yan, L., Meng, X., Duan, Q., & Zhu, G. (2021). Research progress on toll-like receptor signal transduction and its roles in antimicrobial immune responses. *Applied Microbiology and Biotechnology*, 105, 5341–5355. <https://doi.org/10.1007/s00253-021-11406-8>
- Yilmaz, A., Shen, S., Adelson, D. L., Xavier, S., & Zhu, J. J. (2005). Identification and sequence analysis of chicken toll-like receptors. *Immunogenetics*, 56(10), 743–753. <https://doi.org/10.1007/s0025-1-004-0740-8>
- Zhang, G., Li, C., Li, Q., Li, B., Larkin, D. M., Lee, C., Storz, J. F., Antunes, A., Greenwald, M. J., Meredith, R. W., Odeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., ... Froman, D. P. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, 346(6215), 1311–1320. <https://doi.org/10.1126/science.1251385>
- Zhu, Y., Grueber, C., Li, Y., He, M., Hu, L., He, K., Liu, H., Zhang, H., & Wu, H. (2020). MHC-associated *Baylisascaris schroederi* load informs the giant panda reintroduction program. *International Journal for Parasitology: Parasites and Wildlife*, 12, 113–120. <https://doi.org/10.1016/j.ijppaw.2020.05.010>
- Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, 29(21), 2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Magid, M., Wold, J. R., Moraga, R., Cubrinovska, I., Houston, D. M., Gartrell, B. D., & Steeves, T. E. (2022). Leveraging an existing whole-genome resequencing population data set to characterize toll-like receptor gene diversity in a threatened bird. *Molecular Ecology Resources*, 22, 2810–2825. <https://doi.org/10.1111/1755-0998.13656>