Parasitology Open

cambridge.org/pao

Molecular evolution in immune genes across the avian tree of life

Diana C. Outlaw[1], V. Woody Walstrom[1], Haley N. Bodden[1], Chuan-yu Hsu[2], Mark Arick II[2] and Daniel G. Peterson[2]

[1]Department of Biological Sciences, Mississippi State University, PO Box GY, Mississippi State, MS 39762, USA and [2]Institute for Genomics, Biocomputing & Biotechnology, Mississippi State University, 2 Research Boulevard, Box 9627, Starkville, MS 39759, USA

Research Article

Cite this article: Outlaw DC, Walstrom VW, Bodden HN, Hsu Chuan-yu, Arick II M, Peterson DG (2019). Molecular evolution in immune genes across the avian tree of life. *Parasitology Open* **5**, e3, 1–9. https://doi.org/10.1017/pao.2019.3

Received: 11 July 2018
Revised: 22 February 2019
Accepted: 22 February 2019

Key words:
Malaria parasites; haemosporidians; birds; molecular evolution; immune genes; innate immunity; bird transcriptome

Author for correspondence:
Diana C. Outlaw, E-mail: doutlaw@biology.msstate.edu

Abstract

All organisms encounter pathogens, and birds are especially susceptible to infection by malaria parasites and other haemosporidians. It is important to understand how immune genes, primarily innate immune genes which are the first line of host defense, have evolved across birds, a highly diverse group of tetrapods. Here, we find that innate immune genes are highly conserved across the avian tree of life and that although most show evidence of positive or diversifying selection within specific lineages or clades, the number of sites is often proportionally low in this broader context of putative constraint. Rather, evidence shows a much higher level of negative or purifying selection in these innate immune genes – rather than adaptive immune genes – which is consistent with birds' long coevolutionary history with pathogens and the need to maintain a rapid response to infection. We further explored avian responses to haemosporidians by comparing differential gene expression in wild birds (1) uninfected with haemosporidians, (2) infected with *Plasmodium* and (3) infected with *Haemoproteus* (*Parahaemoproteus*). We found patterns of significant differential expression with some genes unique to infection with each genus and a few shared between 'treatment' groups, but none that overlapped with the genes included in the phylogenetic study.

Introduction

Our understanding of the evolution of the avian genome has recently been updated by the landmark avian phylogenomic paper by Jarvis *et al.* (2014). This phylogeny includes all extant orders from Neoaves (e.g. songbirds, parrots, pigeons) and species from several clades in the more basal orders including Anseriformes (waterfowl), Galliformes (landfowl), Tinamiformes (tinamous) and Struthioniformes (ostrich). The phylogeny supports the rapid radiation of the most derived order Passeriformes (i.e. perching birds) and their sister taxon, the parrots (Psittaciformes). This diversification occurred around 60–55 million years ago during the Late Paleocene in Gondwana (Gill, 1995; Jarvis *et al.*, 2014). Passeriformes account for more than half of all avian species, are found on every continent except for Antarctica, and occupy a diverse range of habitats (Gill, 1995). Although birds as a whole appear to have evolved at a conservative molecular evolutionary rate when contrasted with other large groups of organisms (notably mammals), passeriform songbirds have an average mutation rate that is close to twice that of other avian species (Jarvis *et al.*, 2014; Zhang *et al.*, 2014).

Immune genes are shaped by selection pressure from an onslaught of coevolving pathogens (Hendrick, 1998; Acevedo-Whitehouse and Cunningham, 2006; Piertney and Oliver, 2006). The genes of the major histocompatibility complex (MHC) have been the most useful for understanding the adaptive immune response in wildlife, particularly in birds (Acevedo-Whitehouse and Cunningham, 2006; Turner *et al.*, 2012), and many other systems (natural and model or experimental) have been used to explore gene-for-gene coevolutionary processes (reviewed in Brockhurst and Koskella, 2013). However, the role of selection in shaping diversity in other aspects of the immune system, such as innate immunity, remains poorly understood (Acevedo-Whitehouse and Cunningham, 2006; Vinkler and Albrecht, 2009; Grueber *et al.*, 2012, 2014).

To date, the only non-MHC immune genes to be extensively studied in birds are Toll-like receptors (TLR). TLRs are membrane-bound sensors of the innate immune system that recognize invariant and distinctive molecular features of invading microbes and are essential for initiating adaptive and innate immune responses in vertebrates. These types of receptors are part of a large family of proteins called pattern-recognition receptors (PRRs). The genetic variation at TLR genes has been directly related to differential pathogen responses in birds, humans and livestock (Alcaide and Edwards, 2011; Grueber *et al.*, 2012; Grueber *et al.*, 2014; Bateson, *et al.*, 2016; Raven *et al.*, 2017; Kannaki *et al.*, 2018; Nelson-Flower *et al.*, 2018). Prior research has shown that vertebrate TLR evolution is mostly characterized by purifying (negative) selection (Barreiro *et al.*, 2009; Alcaide and Edwards, 2011; Nelson-Flower *et al.*, 2018) and, to a lesser extent, balancing selection (Ferrer-Admetlla *et al* 2008). However, Grueber *et al.* (2014) also found evidence of episodic positive selection in avian TLRs. This discovery was aided by the doubling of avian transcriptomic and genomic data made available at the time of their

CAMBRIDGE UNIVERSITY PRESS

research (Grueber *et al*., 2014). These new data allowed Grueber *et al*. (2014) to evaluate up to 23 bird species per gene and gave added resolution in discerning patterns of selection.

More generally, PRRs are highly conserved and found in a diverse range of animal species. They provide the foundation upon which the hosts' immune response to pathogens has evolved (e.g. Bagheri and Zahmatkesh, 2018). PRRs account for many other families of sensors beside TLRs (Acevedo-Whitehouse and Cunningham, 2006). These include soluble components (e.g. lipopolysaccharide-binding protein, collectins and pentraxins) and cell-associated components (e.g. TLRs, C-type lectins, NOD-like receptors, RIG-I like receptors, scavenger receptors, formyl-peptide receptors and various intracellular receptors; Chen *et al*., 2013). Pattern recognition is considered to occur at three different levels. These are interactions involving soluble extracellular PRRs, membrane-bound PRRs and PRRs found in the cytoplasm (see also Zhang *et al*., 2018). Most major pathogen-response groups are included in one or more of these three levels. The redundancy in recognition of particular pathogen-associated molecular patterns and coordination of different PRRs to sense any single invading class of microorganism is an important feature of the PRR network. The outcome of PRR-mediated responses is to initiate downstream immune cascades that represent the first stage in establishing the immune response. These cascades are in turn mediated by a large group of regulatory proteins called cytokines. Cytokines include interleukins, interferons, tumour necrosis factors and chemokines (Chen *et al*., 2013). Variation in cytokine genes has been associated with certain pathogens in humans, but relatively little research is available for birds, and what is available is primarily limited to fowl (e.g. Chhabra *et al*., 2015; Kannaki *et al*., 2018).

PRR immune genes have shown signatures of selection that can be associated with specific bird species or evolutionary families throughout the avian phylogenetic tree (Grueber *et al*., 2014). For example, chickens, unlike ducks, geese and finches, lack RIG-I, the PRR sensor used to detect highly pathogenic avian influenza. Instead, they express a gene called the melanoma differentiation-associated gene 5 (MDA5), which compensates for the missing RIG-I (Chen *et al*., 2013; Magor *et al*., 2013). Some PRR-induced interferon genes are missing from birds completely. These include interferon genes ISG15, ISG54 and ISG56 (see Velova *et al*., 2018). Birds as a whole have a reduced repertoire of immune genes when compared to mammals (Magor *et al*., 2013), and because of this, we could predict that avian PRRs are evolutionarily constrained (Chen *et al*., 2013; Magor *et al*., 2013; Grueber *et al*., 2014).

To explore PRR evolution in birds, we conducted an *in silico* analysis of bird gene sequences in the GenBank. Additionally, we examined gene expression differences in Northern cardinals uninfected by haemosporidians, infected by *Plasmodium* haemosporidians, or infected by *Haemoproteus* haemosporidians. The computational analyses suggest that some avian PRRs are under positive, diversifying selection, while the differential gene expression analysis highlighted the differences in response among naturally-occurring 'treatment' groups.

Methods

Sequence mining and alignment

Using annotations of immune genes from the GenBank, we used BLAST against the known genomes and transcriptomes of birds and aligned 47 individual gene datasets (i.e. fasta files). We included most PRRs that we could align with at least 250 base pairs (bp), some non-PRR genes and some from the adaptive immune system with at least 250 bp; we did not include all the TLRs because they had been analysed in another study (Grueber *et al*., 2014). We then trimmed each down to contiguous alignments for each taxon that were aligned at a first codon position and ended at a third codon position using webPrank (https://www.ebi.ac.uk/goldman-srv/webprank/). There were no ambiguities, stop codons or missing data in any alignment, and all genes were checked for recombination using RDP4 (Martin *et al*., 2015). Numbers of taxa per dataset ranged from 18 to 57, and gene regions ranged in length from 279 to 2346 bp. Only coding regions were used in these analyses. All gene alignments and MEME results are freely available in the Dryad Depository (doi:10.5061/dryad.r5594f0).

Characterization of gene evolution

We analysed each dataset to determine patterns of selection using several methods within the HyPhy platform (via the Adaptive Evolution server at DataMonkey.org; Kosakovsky Pond and Frost, 2005; Delport *et al*., 2010). The single-likelihood ancestor counting (SLAC; Kosakovsky Pond and Frost, 2005) method was used to determine the overall d_N/d_S per gene dataset. The fixed-effects likelihood (FEL; Kosakovsky Pond and Frost, 2005) method was used to determine the number of sites (codons) under negative selection; FEL fits a site-specific likelihood model on a per-codon basis and then determines whether d_N is significantly greater than d_S. FEL reconstructs the phylogeny using maximum likelihood while optimizing parameters. The mixed effects model of evolution (MEME; Murrell *et al*., 2012) was then used to determine the number of positively-selected sites; MEME essentially fits a likelihood model assuming a background d_N/d_S and determines, on a site-by-site basis, whether d_N is greater than the background rate. Like FEL, MEME reconstructs the phylogeny using maximum likelihood while optimizing parameters. Finally, even with the limitations recently pointed out by Venkat *et al*. (2018), we used adaptive branch-site REL (aBSREL; Kosakovsky Pond *et al*., 2011; Smith *et al*., 2011), a branch-based rather than site-based model, to determine if and where episodic diversifying selection has occurred in branches within trees reconstructed with distance methods (per the programme).

Comparative transcriptome analyses

In a previous study (Walstrom and Outlaw, 2016), we collected blood samples from Northern cardinals (*Cardinalis cardinalis*). All samples were adult (after hatch-year) females. We determined whether each animal was (1) uninfected with haemosporidians, (2) infected with *Plasmodium* or (3) infected with *Haemoproteus* (*Parahaemoproteus*) parasites. As only a small amount of blood was required to determine the haemosporidian infection status, excess blood for each cardinal was stored at −80°C. In the current study, we extracted RNA from blood samples of three birds uninfected with haemosporidians, three birds infected with *Plasmodium* (two with OZ03OZ01 and one with SIAMEX01) and three birds infected with *Haemoproteus* (two with OZ45_MEX19 and one with TUTI233) for a total of nine samples. RNA was extracted using a chloroform/ethanol protocol followed by a DNase digestion, and then cleaned using a Qiagen RNeasy Mini kit (Qiagen.com). cDNA libraries were prepared using a NEBNext® Ultra™ II RNA Library Prep Kit for Illumina and run on a HiSeq (Illumina.com). Salmon (v0.9.1, Patro *et al*., 2017) generated transcript expression estimates for each library by mapping the raw Illumina reads to the zebra finch (*Taeniopygia guttata*) transcriptome (v3.2.4, Warren *et al*., 2010). We recovered ~11 200 genes from each 'treatment'. Gene expression estimates were produced by

https://doi.org/10.1017/pao.2019.3 Published online by Cambridge University Press

Table 1. Summary of results for each gene from MEME (# + sites), FEL (# – sites), SLAC (d_N/d_S) and aBSREL (diversifying selection) analyses in HyPhy

Gene name	Number of taxa	Base pairs	Number of positive sites	Proportion of positive sites	Number of negative sites	Proportion of negative sites	d_N/d_S (ω)	diversifying selection
Basigin (Ok blood group)	36	660	16	0.024	81	0.123	0.407	+
Complement component 1, *q* subcomponent, A chain	22	717	9	0.013	87	0.121	0.336	+
CD36 antigen (collagen type-I receptor, thrombospondin receptor)	34	834	10	0.012	121	0.145	0.302	–
CD47 antigen (Rh-related antigen, integrin-associated signal transducer)	27	810	27	0.033	59	0.073	0.781	+
Collectin sub-family member 10 (a C-type lectin pattern recognition receptor)	44	810	2	0.002	127	0.156	0.197	–
DEXH-box helicase 58 (also known as RIG-I-like receptor 3)[a]	**23**	**1116**	**11**	**0.010**	**162**	**0.145**	**0.224**	–
Hexose-6-phosphate dehydrogenase (glucose 1-dehydrogenase)	57	816	3	0.004	151	0.185	0.186	–
Interferon induced with helicase C domain 1 (RIG-I-like receptor 2)	25	1497	16	0.011	159	0.106	0.344	–
Interferon alpha and beta receptor subunit 1	35	561	11	0.020	74	0.012	0.383	–
Interleukin 1 receptor, type I	32	1167	7	0.006	131	0.112	0.368	–
Interleukin 1 receptor, type 2	47	666	11	0.017	81	0.122	0.417	–
Interleukin-2 receptor alpha chain	31	492	24	0.049	37	0.075	0.790	–
Interleukin-5 receptor alpha chain	44	546	10	0.018	82	0.150	0.370	–
Interleukin 6	39	450	11	0.024	56	0.124	0.389	–
Interleukin 7 receptor	45	699	7	0.010	99	0.142	0.327	–
Interleukin 13 receptor subunit alpha 1	47	510	6	0.018	81	0.159	0.329	–
Interleukin 13 receptor subunit alpha 2	48	510	6	0.012	81	0.159	0.317	–
Interleukin 15	38	375	6	0.016	89	0.237	0.457	–
Interleukin 16	47	339	1	0.003	52	0.153	0.296	–
Interleukin 17 receptor A	42	819	12	0.015	114	0.139	0.343	–
Interleukin 18	42	450	6	0.013	79	0.176	0.285	–
Interleukin 18 receptor 1	43	1323	16	0.012	174	0.132	0.358	–
Interleukin 20 receptor subunit alpha	48	636	7	0.011	102	0.160	0.309	–
Interleukin 21	22	279	10	0.036	16	0.057	0.789	–
Interleukin 21 receptor	31	987	14	0.014	102	0.103	0.432	–
Interleukin 31 receptor A	42	1095	20	0.002	132	0.121	0.470	+
Interferon, gamma	40	417	14	0.034	45	0.108	0.602	+
Interleukin 1 receptor associated kinase 2	45	786	9	0.011	134	0.170	0.260	–
Interferon regulatory factor 8	47	615	0	0	127	0.207	0.136	–
Lipopolysaccharide-induced TNF factor (tumour necrosis factor)	41	420	6	0.014	59	0.140	0.289	–
Mannan binding lectin serine peptidase 2 (a C-type lectin pattern recognition receptor)	42	522	6	0.011	87	0.167	0.282	–
Mannose receptor, C type 1 (a C-type lectin pattern recognition receptor)	32	1338	19	0.014	156	0.117	0.368	–
Mannose receptor, C type 2 (a C-type lectin pattern recognition receptor)	31	1560	0	0	322	0.206	0.040	–
Nuclear factor, interleukin 3 regulated[b]	**48**	**1347**	**1**	**0.001**	**243**	**0.180**	**0.100**	–
NOD-like receptor P3[c]	**20**	**1434**	**29**	**0.020**	**120**	**0.084**	**0.441**	**+**
NOD-like receptor X1	48	1245	1	0.001	293	0.235	0.098	–
NOD-like receptor 1	46	1602	11	0.007	229	0.143	0.276	–
Nitric oxide synthase 2	48	2346	13	0.006	398	0.170	0.208	–

(*Continued*)

https://doi.org/10.1017/pao.2019.3 Published online by Cambridge University Press

Table 1. (*Continued.*)

Gene name	Number of taxa	Base pairs	Number of positive sites	Proportion of positive sites	Number of negative sites	Proportion of negative sites	d_N/d_S (ω)	diversifying selection
Phospholipase A2 receptor 1 (a C-type lectin pattern recognition receptor)	41	2085	6	0.003	261	0.125	0.335	–
Scavenger receptor class A, member 5	36	672	1	0.001	128	0.190	0.129	–
Scavenger receptor class B, member 1	46	1242	9	0.007	216	0.174	0.221	–
Scavenger receptor class B, member 2	43	1287	8	0.006	247	0.192	0.128	–
Solute carrier family 4, anion exchanger, member 1 (Diego blood group)	18	906	1	0.001	131	0.145	0.131	+
Stabilin-1	39	1461	11	0.008	246	0.168	0.239	–
Toll-like receptor 3	48	903	14	0.016	128	0.142	0.447	–
Toll-like receptor 7[d]	47	2019	24	0.012	356	0.176	0.363	+
X-linked Kell blood group precursor antigen	42	735	1	0.001	139	0.189	0.091	–
Average: Non-innate	*38*	*968*	*7*	*0.007*	*161*	*0.164*	*0.209*	
Average: Innate	*40*	*926*	*11*	*0.014*	*132*	*0.137*	*0.378*	

[a]DEXH-box helicase 58 is involved in nucleic acid binding and helicase activity (https://www.genecards.org/cgi-bin/carddisp.pl?gene=DHX58&keywords=DEXH-box).
[b]NFIL3 is involved in DNA-binding transcription factor activity and transcription corepressor activity (https://www.genecards.org/cgi-bin/carddisp.pl?gene=NFIL3&keywords=NFIL3).
[c]NOD-like receptor P3 is involved in peptidoglycan binding (https://www.genecards.org/cgi-bin/carddisp.pl?gene=NLRP3&keywords=NOD-like,P3).
[d]TLR7 is involved in transmembrane signalling receptor activity and double-stranded RNA binding (https://www.genecards.org/cgi-bin/carddisp.pl?gene=TLR7&keywords=TLR7).
Trees from genes in bold are shown in Fig. 1.

combining the transcript expression estimates from the same gene for each sample (Sonenson *et al.*, 2015). Genes expressed in fewer than three of the nine samples were removed from further analysis (Bourgon *et al.*, 2010). The gene expression estimates were normalized based on sequencing depth of each library. EdgeR (v3.12.1; Robinson *et al.*, 2009) was used to find differentially expressed genes. Generalized linear models were used to test for differential expression (DE) based on contrasting each treatment against each other, i.e. *Plasmodium*-infected *vs* uninfected, *Haemoproteus*-infected *vs* uninfected and *Plasmodium*-infected *vs Haemoproteus*-infected. Genes with an FDR adjusted *P* value ≤ 0.05 were considered differentially expressed. DE results are freely available in the Dryad Depository (doi:10.5061/dryad.r5594f0).

Results

Gene evolution

The number of taxa in each aligned immune-associated gene dataset ranged from 18 to 57 with an average number of 39 species. The number of bps included in each dataset ranged from 279 to 2346 with an average of 940 bps. The number of positively selected sites ranged from 0 to 29 (detected using MEME) and the number of negatively selected sites (detected using FEL) ranged from 0 to 398. Across all datasets d_N/d_S ranged from 0.040 to 0.789. Eight phylogenetic trees generated for specific gene datasets showed statistically significant evidence of positive, diversifying selection and 39 showed no evidence of positive, diversifying selection (detected using aBSREL). These results are listed in Table 1. Two trees with evidence of positive selection – one with high taxon and high bp values and one with low taxon and low bp values, and two trees with no evidence of positive selection – one with high taxon and high bp values and one with low taxon and low bp values are shown in Fig. 1 (from aBSREL analyses). Note that MEME and aBSREL analyses returned similar topologies.

Differential gene expression

All samples in this study were collected from free-living birds, and although our polymerase chain reaction-based tests were repeated several times and have been shown in our lab and in others to be highly reliable, we did not control or specifically account for infection with other pathogens. Relevant values for significantly differentially expressed genes are listed in Table 2. Six genes were uniquely downregulated and four were uniquely upregulated in samples classified as being infected solely with *Haemoproteus* parasites. Nine genes were uniquely downregulated and seven genes were uniquely upregulated in samples putatively infected solely with *Plasmodium* parasites. Three genes were upregulated in both *Haemoproteus*-infected samples and *Plasmodium*-infected samples. The first of these, SIX Homeobox 4 is found to be involved in eye development in *Drosophila* and is a transcription factor in mouse development (NCBI gene). The second, immunoglobulin superfamily2C member 11, is involved in cell adhesion (NCBI gene), and the third, tetratricopeptide repeat domain 37, is involved in protein–protein interactions (NCBI gene). The small sample sizes almost certainly bias our results and affect the substantiation of our interpretation, but regardless of these limitations, DE is evident between groups (see also Supplementary Table 1). Surprisingly, none of the genes included in the phylogenetic analyses was found in the DE analyses. Neither did our DE analyses include overlap with the results from Videvall *et al.* (2015), in which the authors examined patterns of DE in response to experimental infection of birds with *Plasmodium*.

Discussion

Recently, the genomic sequences of 45 species have been made available from across the avian tree of life (Zhang, 2014). These data should provide a more complete picture of immune gene evolution in birds, including both TLRs and other non-MHC-related immune genes. To date, the only examination of molecular

https://doi.org/10.1017/pao.2019.3 Published online by Cambridge University Press

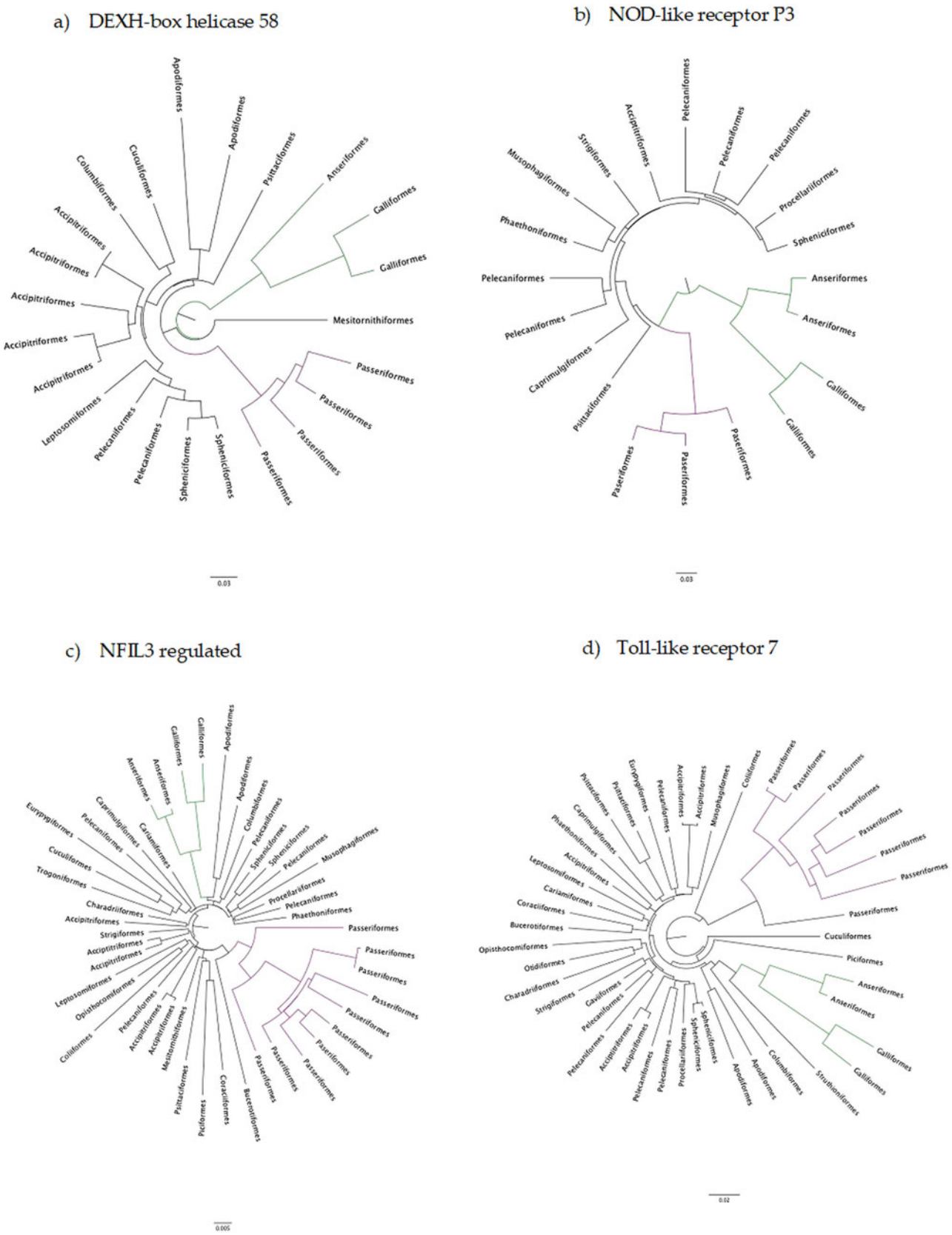

Fig. 1. Gene trees reconstructed with aBSREL. Branch lengths are proportional to the number of codon substitutions. Songbirds (Passeriformes) are the purple clade and game birds (Anseriformes and Galliformes) are the green clade.

evolution in avian immune genes has been researched by making pairwise comparisons of orthologous genes between bird species (see Ekblom and Galindo, 2010; Ekblom *et al.*, 2011, but see Minias *et al.*, 2018). Here we evaluated signatures of natural selection in avian PRR gene families using a phylogenetic framework.

Like humans, birds are highly susceptible to malaria (haemosporidian) parasite infection and are confronted with no fewer than four genera of malaria parasites (Valkiunas, 2004; see also Bichet *et al.*, 2015). Because malaria has had such a profound effect on the human genome (see Kwiatkowski, 2005) it would

https://doi.org/10.1017/pao.2019.3 Published online by Cambridge University Press

Table 2. Comparative patterns of gene expression between uninfected birds and those infected with *Plasmodium* or *Parahaemoproteus* parasites

	log FC	log CPM	LR	*P* Value	FDR	Description
Parahaemoproteus						
Gene ID: 100221957	−9.36275	2.92109	35.29946	0.00000	0.00002	SIX homeobox 4
Gene ID: 105759718	−7.26418	−0.70744	17.23340	0.00003	0.03692	LETM1 and EF-hand domain-containing protein 12C mitochondrial-like
Gene ID: 751976	−7.15826	9.93214	17.04873	0.00004	0.03699	Histidine triad nucleotide binding protein 1
Gene ID: 100232311	−4.67308	9.00162	17.94572	0.00002	0.03104	Immunoglobulin superfamily2C member 11
Gene ID: 100227340	−3.93052	0.12766	17.76342	0.00003	0.03104	Arachidonate 5-lipoxygenase-activating protein
Gene ID: 100229287	−1.96561	4.50633	20.32498	0.00001	0.01824	Enoyl-CoA delta isomerase 1
Gene ID: 100219531	−1.83781	4.48235	19.21280	0.00001	0.02176	Tetratricopeptide repeat domain 37
Gene ID: 105760653	1.63847	4.32007	18.20766	0.00002	0.03104	Prefoldin subunit 6
Gene ID: 100231330	2.64436	4.41541	19.58085	0.00001	0.02154	Integrin2C beta 4
Gene ID: 100225098	3.86951	3.55359	27.93441	0.00000	0.00047	ATPase2C H+ transporting2C lysosomal accessory protein 1
Gene ID: 100230529	4.55704	3.31428	48.87034	0.00000	0.00000	Family with sequence similarity 1712C member A12C transcript variant X2
Plasmodium						
Gene ID: 100221957	−9.12695	2.92109	38.15712	0.00000	0.00001	SIX homeobox 4
Gene ID: 100219559	−7.90354	−0.38295	17.99303	0.00002	0.02063	Histone H2B 1/2/3/4/6
Gene ID: 100224924	−6.35962	−0.88948	19.50257	0.00001	0.01402	1-Phosphatidylinositol 42C5-bisphosphate phosphodiesterase delta-3-like
Gene ID: 100232311	−5.02708	9.00162	20.08035	0.00001	0.01342	Immunoglobulin superfamily2C member 11
Gene ID: 100225454	−4.75489	4.30713	19.11349	0.00001	0.01528	Spermidine synthase
Gene ID: 100224843	−3.72758	−0.45130	17.04934	0.00004	0.02905	Interferon-induced transmembrane protein 5
Gene ID: 100227340	−3.62338	0.12766	17.60857	0.00003	0.02331	Arachidonate 5-lipoxygenase-activating protein
Gene ID: 100232731	−3.22187	3.93025	15.33878	0.00009	0.04944	ADP-ribosylation factor-like protein 8B
Gene ID: 100231715	−2.54681	3.06869	18.20384	0.00002	0.02015	Heat shock 70 kDa protein 4L-like
Gene ID: 100229287	−2.20351	4.50633	25.46898	0.00000	0.00251	Enoyl-CoA delta isomerase 1
Gene ID: 100219531	−1.88625	4.48235	20.47453	0.00001	0.01342	Tetratricopeptide repeat domain 37
Gene ID: 100229704	−1.88178	4.37649	19.84160	0.00001	0.01342	Runt-related transcription factor 13B translocated to2C 1 (cyclin D-related)
Gene ID: 100192321	−1.67426	6.37482	15.97917	0.00006	0.04768	Tropomyosin 42C transcript variant X5
Gene ID: 100226858	−1.51119	7.38191	15.32339	0.00009	0.04944	Nuclear export mediator factor
Gene ID: 100232218	2.05376	2.52090	15.44468	0.00008	0.04944	Zinc finger CCCH-type containing 7A
Gene ID: 100222996	2.17997	3.04243	18.65439	0.00002	0.01750	TSEN15 tRNA splicing endonuclease subunit
Gene ID: 100227101	5.39183	2.09989	22.52959	0.00000	0.00770	Myeloid/lymphoid or mixed-lineage leukaemia (trithorax homolog2C Drosophila)3B translocated to2C 42C transcript variant X8
Gene ID: 100229846	5.89259	1.83186	20.53874	0.00001	0.01342	Crystallin2C mu
Gene ID: 100224354	6.39599	−1.32883	15.66410	0.00008	0.04944	Chemokine-like factor
Gene ID: 101234193	7.05925	−0.61655	15.83667	0.00007	0.04820	TELO2-interacting protein 1 homologue
Gene ID: 100218691	7.56559	−0.01905	15.27458	0.00009	0.04944	Histone H2A-IV

Upregulated genes are in light grey and downregulated genes are in darker grey. See also Supplementary Table 1.

seem necessary that we look for the effects of molecular evolution in avian genomes. However, given that it is relatively unknown which genes are turned on or off in response to infection with malaria parasites, studies – like this one – must first compare birds with different types of parasite infections. Many genetic mechanisms of malaria resistance in humans were shaped by evolution in such a way that they have even become harmful to the constituent populations. The most well-known example is heterozygote advantage in the HBB (haemoglobin beta) gene, which confers such resistance to human *Plasmodium falciparum* malaria that the deleterious allele remains at 10% frequency or higher even though the homozygous recessive is lethal (Kwiatkowski, 2005). Such costly mechanisms have arisen independently many times and in response to different species of human *Plasmodium* (summarized in Kwiatkowski, 2005).

General evolutionary trends

Innate immune genes of birds appear to evolve in a highly conserved manner under strong purifying selection (see Table 1)

https://doi.org/10.1017/pao.2019.3 Published online by Cambridge University Press

which is entirely consistent with their evolutionary role. In the eight phylogenetic trees with at least one branch under positive selection, selection occurred mostly on terminal branches and there was no identifiable pattern with which taxa were under selection.

Haemosporidian malaria parasites elicited different transcriptomic responses from our wild-caught Northern cardinal samples. None of the significantly differentially expressed genes overlap with the immune genes in the first part of the study. Of note, a few of the differentially expressed genes in the birds are known to be involved in immune responses and signal transduction. With further sampling of other bird species, differentially expressed genes may be candidates for further studies of avian–haemosporidian interactions. However, considerable caution is warranted as the birds were caught in the wild and not tested for infection with other parasites, and the sample size is very small. Thus, the observed DE may reflect the effects of many other environmental factors.

Adaptive strategy

The adaptive strategy of avian innate immune genes is their conservation: a d_N/d_S less than 1. Overwhelmingly, genes of the innate immune system have been under intense selection to stay the same, although it is likewise clear than many sites are under positive selection even against the background of constraint. Therefore, why then, have other studies found immune genes to be rapidly evolving? The most comprehensive study was conducted between a galliform (chicken) and a passerine (zebra finch), two branches of the avian tree of life separated by ~40% genomic sequence divergence (Ekblom *et al*., 2011). In a follow-up study, Ekblom *et al*. (2011) summarized across gene ontology groups within the zebra finch genome and found d_N/d_S values about half of what we see here. On the other hand, analyses of selection in the MHC genes show rampant evidence of positive selection, a d_N/d_S greater than 1 within passerines (Minias *et al*., 2018), which is much higher than we found. More recently, big data from across the avian tree of life have shown that songbirds radiated in a big bang – even with low overall values of d_N/d_S (Zhang *et al*., 2014) – as we have seen in our gene trees (Fig. 1a–d). This big bang would at least partially explain the elevated rates that Ekblom *et al*. (2011) found between these very divergent lineages, except that their overall estimate of selection was closer to neutral, with positive selection showing up at specific residues rather than across whole genes.

Our analyses paint a more complex picture given that we could leverage more data than in previous studies. Immune genes are a blend of both positively and negatively selected sites in a constrained background, and their complex patterns of selection reflect their complex roles. The complexity of evolutionary interactions between hosts and pathogens is becoming increasingly apparent, particularly within the plant literature (e.g. Thrall *et al*., 2016 and references therein), and mathematical models to understand these interactions increasingly refer to them as 'tangled'. Detailed analyses of sites under selection and the interpretation thereof are beyond our expertise, but the myriad parasites that differentially affect avian clades suggest that this pursuit would unlock many secrets about the evolution of avian immune genes.

Although studies of parasites in wild birds are extremely patchy in geographic location and host taxonomy, there is ample evidence to suggest that some parasites infect all birds but that different birds are affected by different parasites in many cases. *Plasmodium* (Atkinson, 2008), *Toxoplasma* (Dubey, 2008) and capillarid nematodes (Yabsley, 2009) are cosmopolitan and seemingly ubiquitous across birds (and other vertebrate taxa). On the other hand, *Trichomonas* primarily infects doves, raptors and parrots (Forrester and Foster, 2008), and the *Dispharynx*/*Echinuria*/*Streptocara* complex and tracheal worms (Fernando and Barta, 2008) are most commonly found in songbirds and galliformes. There are many other examples of host-specificity (at the order and family level) in bird parasites (see Wobeser, 2008; see also Kumar *et al*., 2017) and coevolutionary relationships between these hosts and parasites presumably over millions of years have shaped the adaptive responses within and between avian clades. Differences between the innate and adaptive immune systems are profound, and therefore, our expectations of patterns of selection between these groups of genes are likewise very different. Here, we focused on (primarily) genes of the innate immune system, and found an overall signature of purifying selection. However, where there are clear distinctions between groups of birds that are infected by different parasite groups, there is evidence of molecular evolution that may correspond with these differences.

Author ORCIDs. Diana C. Outlaw, 0000-0002-9072-4090

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/pao.2019.3.

Financial support. Funding for the project was provided by the National Institutes of Health R03AI117223-01A1 to DCO and U.S. Department of Agriculture ARS 58-6066-6-046 to DGP.

Conflict of interest. None.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals. Birds were mist-netted and released under both United States Federal (MB6214) and Mississippi State collecting permits to DCO, and under MSU's Institutional Animal Care and Use Committee (IACUC).

References

Acevedo-Whitehouse K and Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution* **2**, 433–438.

Alcaide M and Edwards SV (2011) Molecular evolution of the toll-like receptor multigene family in birds. *Molecular Biology and Evolution* **28**, 1703–1715.

Atkinson CT (2008) Haemoproteus. In Atkinson CT, Thomas NJ and Hunter DB (eds), *Parasitic Diseases of Wild Birds*. Ames, IO: Wiley-Blackwell, pp. 13–34.

Bagheri M and Zahmatkesh A (2018) Evolution and species-specific conservation of toll-like receptors in terrestrial vertebrates. *International Reviews of Immunology* **37**, 217–228. DOI: 10.1080/08830185.2018.1506780.

Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, Kidd JR, Kidd KK, Alcaïs A, Ragimbeau J, Pellegrini S, Abel L, Casanova J-L and Quintana-Murci L (2009) Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genetics* **5**, e1000562.

Bateson ZW, Hammerly SC, Johnson JA, Morrow ME, Whittingham LA and Dunn PO (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**, 4730–4744.

Bichet C, Moodley Y, Penn DJ, Sorci G and Garnier S (2015) Genetic structure in insular and mainland populations of house sparrows (*Passer domesticus*) and their hemosporidian parasites. *Ecology and Evolution* **5**, 1639–1652.

Bourgon R, Gentleman R and Huber W (2010) Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences* **107**, 9546–9551.

Brockhurst MA and Koskella B (2013) Experimental coevolution of species interactions. *Trends in Ecology and Evolution* **28**, 367–375.

Chen S, Cheng A and Wang M (2013) Innate sensing of viruses by pattern recognition receptors in birds. *Veterinary Research* **44**, 82.

Chhabra R, Chantrey J and Ganapathy K (2015) Immune responses to virulent and vaccine strains of infectious bronchitis viruses in chickens. *Viral Immunology* **28**, 478–488.

Delport W, Poon AF, Frost SDW and Kosakovsky Pond SL (2010) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics (Oxford, England)* **26**, 2455–2457.

Dubey JP (2008) Toxoplasma. In Atkinson CT, Thomas NJ and Hunter DB (eds), *Parasitic Diseases of Wild Birds*. Ames, IO: Wiley-Blackwell, pp. 204–224.

Ekblom R and Galindo J (2010) Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity* **107**, 1–15. doi: https://doi.org/10.1038/hdy.2010.152.

Ekblom R, French L, Slate J and Burke T (2011) Evolutionary analysis and expression profiling of zebra finch immune genes. *Genome Biology and Evolution* **2**, 781–790.

Fernando MA and Barta JR (2008) Tracheal worms. In Atkinson CT, Thomas NJ and Hunter DB (eds), *Parasitic Diseases of Wild Birds*. Ames, IO: Wiley-Blackwell, pp. 343–354.

Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J and Casals F (2008) Balancing selection is the main force shaping the evolution of innate immunity genes. *Journal of Immunology* **181**, 1315–1322.

Forrester DJ and Foster GW (2008) Trichomonosis. In Atkinson CT, Thomas NJ and Hunter DB (eds), *Parasitic Diseases of Wild Birds*. Ames, IO: Wiley-Blackwell, pp. 120–153.

Gill FB (1995) *Ornithology*. New York, NY: Macmillan.

Grueber CE, Wallis GP, King T and Jamieson IG (2012) Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS ONE* **7**, e45011.

Grueber CE, Wallis GP and Jamieson IG (2014) Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes. *PLoS ONE* **9**, e89632.

Hedrick PW (1998) Balancing selection and MHC. *Genetica* **104**, 207–214.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SY, Faircloth BC, Nabholz B, Howard JT, Suh A, Weber CC, da Fonseca RR, Li J, Zhang F, Li H, Zhou L, Narula N, Liu L, Ganapathy G, Boussau B, Bayzid MS, Zavidovych V, Subramanian S, Gabaldón T, Capella-Gutiérrez S, Huerta-Cepas J, Rekepalli B, Munch K, Schierup M, Lindow B, Warren WC, Ray D, Green RE, Bruford MW, Zhan X, Dixon A, Li S, Li N, Huang Y, Derryberry EP, Bertelsen MF, Sheldon FH, Brumfield RT, Mello CV, Lovell PV, Wirthlin M, Schneider MP, Prosdocimi F, Samaniego JA, Vargas Velazquez AM, Alfaro-Núñez A, Campos PF, Petersen B, Sicheritz-Ponten T, Pas A, Bailey T, Scofield P, Bunce M, Lambert DM, Zhou Q, Perelman P, Driskell AC, Shapiro B, Xiong Z, Zeng Y, Liu S, Li Z, Liu B, Wu K, Xiao J, Yinqi X, Zheng Q, Zhang Y, Yang H, Wang J, Smeds L, Rheindt FE, Braun M, Fjeldsa J, Orlando L, Barker FK, Jønsson KA, Johnson W, Koepfli KP, O'Brien S, Haussler D, Ryder OA, Rahbek C, Willerslev E, Graves GR, Glenn TC, McCormack J, Burt D, Ellegren H, Alström P, Edwards SV, Stamatakis A, Mindell DP, Cracraft J, Braun EL, Warnow T, Jun W, Gilbert MT and Zhang G (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331.

Kannaki TR, Reddy MR, Raja Ravindra KS and Chatterjee RN (2018) Molecular and functional characterization of Toll-like receptor 5 (TLR5) in Aseel and White Leghorn chicken. *Indian Journal of Animal Research* **52**, 235–241.

Kosakovsky Pond SL and Frost SDW (2005) Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Molecular biology and Evolution* **22**, 1208–1222.

Kosakovsky Pond SL, Murrell B, Fourment M, Frost SD, Delport W and Scheffler K (2011) A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution* **28**, 3033–3043.

Kumar a, Vijayakumar P, Gandhale PN, Ranaware PB, Kumar H, Kulkarni DD, Raut AA and Mishra A (2017) Genome-wide gene expression pattern underlying differential host response to high or low pathogenic H5N1 avian influenza virus in ducks. *Acta Virologica* **61**, 66–76.

Kwiatkowski DP (2005) How malaria has affected the human genome and what human genetics can teach us about malaria. *American Journal of Human Genetics* **77**, 171–192.

Luo W, Friedman M, Shedden K, Hankenson KD and Woolf PJ (2009) GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161.

Magor KE, Navarro DM, Barber MR, Petkau K, Fleming-Canepa X, Blyth GA and Blaine AH (2013) Defense genes missing from the flight division. *Developmental and Comparative Immunology* **41**, 377–388.

Martin DP, Murrell B, Golden M, Khoosal A and Muhire B (2015) RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evolution* **1**, vev003.

Minias P, Pikus E, Whittingham LA and Dunn PO (2018) A global analysis of selection at the avian MHC. *Evolution* **72**, 1278–1293.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K and Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics* **8**, e1002764. doi: https://doi.org/10.1371/journal.pgen.1002764.

Nelson-Flower MJ, Germain RR, MacDougall-Shackleton EA, Taylor SS and Arcese P (2018) Purifying selection in the Toll-like receptors of song sparrows *Melospiza melodia*. *Journal of Heredity* **109**, 501–509.

Patro R, Duggal G, Love MI, Irizarry RA and Kingsford C (2017) Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods* **14**, 417–419.

Piertney SB and Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* **96**, 7–21.

Raven N, Lisovski S, Klaassen, Lo N, Madsen T, Ho SYW and Ujvari B (2017) Purifying selection and concerted evolution of RNA-sensing toll-like receptors in migratory waders. *Infection, Genetics and Evolution* **53**, 135–145.

Robinson MD, McCarthy DJ and Smyth GK (2009) Edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139–140.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K and Kosakovsky Pond SL (2011) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Molecular Biology and Evolution* **32**, 1342–1353.

Soneson C, Love MI and Robinson MD (2015) Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**, 1521.

Thrall PH, Barrett LG, Dodds PN and Burdon JJ (2016) Epidemiological and evolutionary outcomes in gene-for-gene and matching allele models. *Frontiers in Plant Science* **6**, 1084.

Turner AK, Begon M, Jackson JA and Paterson S (2012) Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). *Molecular Ecology* **21**, 1632–1646.

Valkiunas G (2004) *Avian Malaria Parasites and Other Haemosporidia*. Boca Raton, FL: CRC Press.

Velová H, Gutowska-Ding MW, Burt DW and Vinkler M (2018) Toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection. *Molecular Biology and Evolution* **35**, 2170–2184.

Venkat A, Hahn MW and Thornton JW (2018) Multinucleotide mutations cause false inferences of lineage-specific positive selection. *Nature Ecology and Evolution* **2**, 1280–1288.

Videvall E, Cornwallis CK, Palinauskas V, Valkiūnas G and Hellgren O (2015) The avian transcriptome response to malaria infection. *Molecular Biology and Evolution* **32**, 1255–1267.

Vinkler M and Albrecht T (2009) The question waiting to be asked: innate immunity receptors in the perspective of zoological research. *Folia Zoologica* **58**, 15–28.

Walstrom VW and Outlaw DC (2016) Distribution and prevalence of haemosporidian parasites in the Northern Cardinal (*Cardinalis cardinalis*). *Parasitology* **103**, 63–68.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong L, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TA, Ferris M, Balakrishnan CN, Sinha S, Blatti C, London SE, Li Y, Lin YC, George J, Sweedler J, Southey B, Gunaratne P, Watson M, Nam K, Backström N, Smeds L, Nabholz B, Itoh Y, Whitney O, Pfenning AR, Howard J, Völker M, Skinner BM, Griffin DK, Ye L, McLaren WM, Flicek P, Quesada V, Velasco G, Lopez-Otin C, Puente XS, Olender T, Lancet D, Smit AF, Hubley R, Konkel MK, Walker JA, Batzer MA, Gu W, Pollock DD, Chen L, Cheng Z, Eichler EE, Stapley J, Slate J, Ekblom R, Birkhead T, Burke T, Burt D, Scharff C, Adam I, Richard H, Sultan M, Soldatov A, Lehrach H, Edwards SV, Yang SP, Li X, Graves T, Fulton L, Nelson J, Chinwalla A, Hou S, Mardis ER and Wilson RK (2010) The genome of a songbird. *Nature* **464**, 757–762.

Wobeser GA (2008) Parasitism: costs and effects. In Atkinson CT, Thomas NJ and Hunter DB (eds), *Parasitic Diseases of Wild Birds*. Ames, IO: Wiley-Blackwell, pp. 3–12.

https://doi.org/10.1017/pao.2019.3 Published online by Cambridge University Press

Yabsley MJ (2008) Eimeria. In Atkinson CT, Thomas NJ and Hunter DB (eds), *Parasitic Diseases of Wild Birds*.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, Ödeen A, Cui J, Zhou Q, Xu L, Pan H, Wang Z, Jin L, Zhang P, Hu H, Yang W, Hu J, Xiao J, Yang Z, Liu Y, Xie Q, Yu H, Lian J, Wen P, Zhang F, Li H, Zeng Y, Xiong Z, Liu S, Zhou L, Huang Z, An N, Wang J, Zheng Q, Xiong Y, Wang G, Wang B, Wang J, Fan Y, da Fonseca RR, Alfaro-Núñez A, Schubert M, Orlando L, Mourier T, Howard JT, Ganapathy G, Pfenning A, Whitney O, Rivas MV, Hara E, Smith J, Farré M, Narayan J, Slavov G, Romanov MN, Borges R, Machado JP, Khan I, Springer MS, Gatesy J, Hoffmann FG, Opazo JC, Håstad O, Sawyer RH, Kim H, Kim KW, Kim HJ, Cho S, Li N, Huang Y, Bruford MW, Zhan X, Dixon A, Bertelsen MF, Derryberry E, Warren W, Wilson RK, Li S, Ray DA, Green RE, O'Brien SJ, Griffin D, Johnson WE, Haussler D, Ryder OA, Willerslev E, Graves GR, Alström P, Fjeldså J, Mindell DP, Edwards SV, Braun EL, Rahbek C, Burt DW, Houde P, Zhang Y, Yang H, Wang J, Avian Genome Consortium, Jarvis ED, Gilbert MT and Wang J (2014) Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320.

Zhang L, Gao Z, Yu L, Zhang B, Wang J and Zhou J (2018) Nucleotide-binding and oligomerization domain (NOD)-like receptors in teleost fish: current knowledge and future perspectives. *Journal of Fish Diseases* **41**, 1317–1330.

https://doi.org/10.1017/pao.2019.3 Published online by Cambridge University Press