



Episodic positive diversifying selection on key immune system genes in major avian lineages

Jennifer Antonides¹ · Samarth Mathur² · J. Andrew DeWoody^{1,2}

Received: 18 September 2019 / Accepted: 19 November 2019 / Published online: 28 November 2019
© Springer Nature Switzerland AG 2019

Abstract

The major histocompatibility complex (MHC) of the adaptive immune system and the toll-like receptor (TLR) family of the innate immune system are involved in the detection of foreign invaders, and thus are subject to parasite-driven molecular evolution. Herein, we tested for macroevolutionary signatures of selection in these gene families within and among all three major clades of birds (Paleognathae, Galloanserae, and Neoaves). We characterized evolutionary relationships of representative immune genes (*Mhc1* and *Tlr2b*) and a control gene (ubiquitin, *Ubb*), using a relatively large and phylogenetically diverse set of species with complete coding sequences (34 orthologous loci for *Mhc1*, 29 for *Tlr2b*, and 37 for *Ubb*). Episodic positive diversifying selection was found in the gene-wide phylogenies of the two immune genes, as well as at specific sites within each gene (8.5% of codon sites in *Mhc1* and 2.7% in *Tlr2b*), but not in the control gene (*Ubb*). We found 20% of lineages under episodic diversifying selection in *Mhc1* versus 9.1% in *Tlr2b*. For *Mhc1*, selection was relaxed in the Galloanserae and intensified in the Neoaves relative to the other clades, but no differences were detected among clades in the *Tlr2b* gene. In summary, we provide evidence of episodic positive diversifying selection in key immune genes and demonstrate differential strengths of selection within Class Aves, with the adaptive gene showing an increased divergence and evolutionary rate over the innate gene, contributing to the growing understanding of vertebrate immune gene evolution.

Keywords Major histocompatibility complex · Toll-like receptor · Paleognathae · Galloanserae · Neoaves

Introduction

Genes involved in the immune response directly affect survival and are targets for natural selection as a consequence of the evolutionary arms race between host and pathogen (Apanius et al. 1997; Best et al. 2009; Hamilton 1980; Shultz and Sackton 2019). Many bird species migrate and thus must constantly cope with a plethora of pathogens and parasites endemic to multiple environments (Gill 2007; O'Connor et al. 2018). Thus, signatures of positive selection for amino

acid diversification in avian immune genes should be detectable on an evolutionary time scale.

In this study, the immune loci of interest represent both major vertebrate immune gene families: one of the adaptive immune system (the Major Histocompatibility Complex, MHC) and one of the more ancient innate immune system (the Toll-like Receptor family, TLRs). Both encode receptors that bind microbial ligands and initiate attacks on foreign invaders (Acevedo-Whitehouse and Cunningham 2006; Vinkler and Albrecht 2009). Evidence in birds, primarily at the population level but also across species, suggests the highly polymorphic MHC and to some extent, the TLRs, are subject to positive selection for amino acid diversification particularly in the gene regions responsible for binding parasite-derived ligands (Alcaide and Edwards 2011; Borghans et al. 2004; Downing et al. 2010; Grueber et al. 2014; Khan et al. 2019; Minias et al. 2019; O'Connor et al. 2018; Sommer 2005). In sharp contrast, most functionally constrained genes (e.g., histones, ubiquitins, and heatshock proteins) are highly conserved within and among species, as they evolve under purifying selection

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s10709-019-00081-3>) contains supplementary material, which is available to authorized users.

✉ Jennifer Antonides
jantonid@purdue.edu

¹ Department of Forestry and Natural Resources, Purdue University, 715 W. State St, West Lafayette, IN, USA

² Department of Biological Sciences, Purdue University, 915 W. State St, West Lafayette, IN, USA

that purges mutations that cause amino acid substitutions (Nei et al. 1997, 2000; Piontkivska et al. 2002). Thus, they can effectively serve as control genes for studies of diversifying selection.

The MHC multigene family of the adaptive immune system plays a key role in the immunological and evolutionary response to novel pathogens, with selection often favoring particular alleles or numbers of alleles that confer resistance while maintaining high levels of genetic variation in a population in the form of allelic diversity and heterozygosity (Hedrick 1994; Piertney and Oliver 2006). In this study we focused on the primary classical MHC Class I locus, referred to herein as *Mhc1* (sometimes named *Uaa*), which recognizes intracellular parasites such as viruses and certain protozoa such as haemosporidian parasites, thus affecting a host's resistance to specific diseases such as avian malaria (Antonides et al. 2019; Bonneaud et al. 2006; Hess and Edwards 2002; Sepil et al. 2013). *Mhc1* encodes a polypeptide (the alpha chain) which forms a receptor which recognizes and binds particular parasite-derived ligands; the antigens are then displayed on the cell's surface for recognition by T-cells and subsequent attack on the foreign invader (Janeway 2005; Klein 1986). While most studies of MHC Class I diversity and evolution focus on the peptide-binding region of the alpha chain, we considered the entire coding sequence as different regions of the gene are likely under different selection pressures and evolve via different mechanisms (Yang et al. 2000).

The TLR multigene family recognizes conserved molecular signatures of microbial classes, and upon binding a foreign ligand, induce a signal cascade for the inflammatory response (Kobe and Kajava 2001). Mechanisms governing intraspecific TLR diversity have been under exploration, including in avian species of conservation concern (Dalton et al. 2016; Gilroy et al. 2017; Grueber et al. 2013). TLR evolution has been thought to be governed primarily by purifying selection (Mukherjee et al. 2009; Roach et al. 2005), but recent research has suggested that TLR evolution also involves positive selection particularly in ligand-binding regions (Alcaide and Edwards 2011; Grueber et al. 2014; Khan et al. 2019). Herein we utilize the *Tlr2b* locus (present only in Class Aves, a duplication of eukaryotic *Tlr2*), which evidence suggests forms heterodimers with *Tlr1a* and *Tlr1b* to recognize peptidoglycans of Gram-positive bacteria and glycolipids on the surface of protozoan parasites such as *Plasmodium* spp. (Alcaide and Edwards 2011; Campos 2001; Eriksson et al. 2014; Grueber 2015; Higuchi 2008; Krishnegowda 2005). In a study of a population of Bananaquits (*Coereba flaveola*) subject to avian malarial parasites, *Tlr2b* was one of three TLR genes found to have significant associations between allelic composition and infection status (Antonides et al. 2019). Thus, like *Mhc1*, *Tlr2b* may play an important role in the evolutionary arms race.

In addition to the immune genes, we utilized the housekeeping gene Polyubiquitin B (*Ubb*) to serve as a control (i.e., a gene which is not expected to evolve under any positive diversifying selection). *Ubb* is a conserved eukaryotic precursor to Ubiquitin, an abundant protein involved in regulation of the concentration of cell-cycle signaling proteins and the degradation of damaged proteins. We chose *Ubb* because the cellular function of the protein is well known, the gene is generally well-characterized, the mode of selection (purifying) is known, and orthologs can easily be identified among related species (Kimura and Tanaka 2010; Nei et al. 2000).

The avian lineages in this study encompass all extant birds within Class Aves, which are divided into the infra-classes Paleognathae (the tinamous and flightless ratites) and Neognathae (all other modern birds), which diverged around 100 MYA in the late Cretaceous. The Neognathae subsequently diverged about 88 MYA into the superorders Galloanserae (ducks, geese, chickens, and kin) and the prolific Neoaves, which underwent a rapid radiation around the Cretaceous–Paleogene (K–Pg) boundary (Jarvis et al. 2014). If the evolutionary rates of immune genes within and among these clades (Paleognathae, Galloanserae, and Neoaves) differ, it would be indicative of differential selection pressures that may be manifested by different mechanisms (e.g., purifying selection versus diversifying selection).

Evolutionary processes, both neutral and non-neutral, are reflected in genome-wide nucleotide substitution rates, which have been estimated as an average background rate of 0.0019 substitutions per site per million years in birds (as compared to the average evolutionary background rate of 0.0027 in mammals), and is not significantly different among clades (Jarvis et al. 2014). The evolutionary rates of genes or portions of genes subject to natural selection differ from that expected under neutrality, and are often quantified by nonsynonymous-to-synonymous substitution rate ratios ($\omega = dN/dS$). Neutral evolution is reflected by $\omega = 1$, an excess of synonymous mutations signals negative (purifying) selection ($\omega < 1$), and an excess of nonsynonymous mutations signals positive (diversifying) selection ($\omega > 1$). To explore mechanisms of avian immune gene evolution, we utilized full coding sequences primarily derived from whole genome sequences and bioinformatics methods that detect transient excesses of nonsynonymous substitutions to synonymous substitutions.

We expected to find evidence of positive diversifying selection in the ligand-binding regions of the adaptive MHC gene (*Mhc1*) and, to a lesser extent (due to functional constraints), the innate TLR gene (*Tlr2b*), on evolutionary time scales within and among all major clades of birds, but not in the housekeeping gene *Ubb*. We additionally expected to identify a differential intensity of immune gene evolution among the avian clades. For example, the highly variable

Neoaves clade occupies a large variety of ecosystems, and therefore likely under more diverse selection pressures, which may lead to more rapid evolution of their immune genes relative to the other two clades [e.g. see O'Connor et al. (2018)]. To these ends, we quantified avian divergence at the representative genes, and tested for selection within the genes as a whole. For those that demonstrated episodic positive diversifying selection, we examined selection at codon sites within the genes and among avian lineages (branches), and finally compared rates of gene evolution among the major avian clades. This study comprises a diverse and relatively large subset of avian species (34 for *Mhc1*, 29 for *Tlr2b*, and 37 for *Ubb*, representing all major clades) studied at orthologous loci representing two major immune gene families. Our approach provides enhanced detection of signatures of selection of at the macroevolutionary scale, providing new insights into the evolutionary processes which shape the avian immunogenetic repertoire.

Methods

Sequence selection

We obtained the full coding sequence (CDS) of orthologous sequences for each of the three genes of interest (*Mhc1*, *Tlr2b*, and *Ubb*) for as many avian species as possible. Due to the fragmented nature of draft genome assemblies, the complete sequences are not present in all species with sequenced genomes, and many sequences available from targeted sequencing strategies are not intended to span the full CDS. For *Mhc1*, we used the CDS for the entire alpha chain. The alpha chain encodes both the variable regions $\alpha 1$ and $\alpha 2$, which forms the peptide-binding groove, and the conserved region $\alpha 3$ which encodes the alpha chain immunoglobulin domain (Janeway 2005; Klein 1986). For the *Tlr2b* locus, we obtained the full CDS, including the variable region, the extracellular N-terminal LRR (leucine-rich repeat) region which is involved in pathogen recognition, the conserved TIR (Toll/interleukin-1 receptor) region, and the intracellular domain that initiates a signal cascade for downstream immune response (Alcaide and Edwards 2011; Kobe and Kajava 2001). We obtained complete CDS for *Ubb*, which codes for a polymer of repeat conserved ubiquitin domains (Nei et al. 2000).

First we identified avian orthologs of *Mhc1*, *Tlr2b*, and *Ubb* using protein similarity from the reference chicken genome at www.orthoDB.org (Zdobnov et al. 2017). This corresponded to orthoDB group IDs of EOG090F0AQT for *Mhc1*, EOG090F02AU for *Tlr2b*, and EOG090F07RP for *Ubb*. We then used the chicken protein for each locus as the query on the NCBI nucleotide database (www.blast.ncbi.nlm.nih.gov) using the *tblastn* algorithm to obtain nucleotide

sequences of avian orthologs. The results were hand-curated to choose only those avian orthologs which contain Ref-seq CDS sequences with all functional domains present as described above, the majority of which were derived from whole genome sequences. When alternative transcripts were present, the one with the longest sequence was chosen. Species in which the *Tlr2* sequence was not distinguished between *Tlr2a* and *Tlr2b* were not included for that locus.

Multiple sequence alignments

An in-frame codon-based multiple sequence alignment (MSA) was constructed for each gene (i.e., one for *Mhc1*, one for *Tlr2b*, and for *Ubb*) in the TranslatorX server (Abascal et al. 2010). First the nucleotide CDS from each species was translated according to the standard genetic code, and an amino acid MSA numbered according to the chicken sequence was generated with T-Coffee v. 12.00 (Di Tommaso 2011). After protein alignment, poorly aligned or ambiguous regions (e.g. containing many large gaps) were removed with Gblocks (Castresana 2000) using the “less stringent” criteria of a minimum block length of 5 amino acids, a minimum number of sequences for a conserved or flanking position of 55%, and a maximum of 50% for gap positions in the sequences in the final alignments. The curated peptide alignment was then back-translated to the corresponding codon-based multiple nucleotide sequence alignment numbered according to the chicken sequence. This approach to DNA alignment ensures that gaps cannot be inserted between the first or second nucleotide positions of a codon, providing a biologically relevant alignment consisting of in-frame codons.

Construction of individual gene trees

For each gene, codon-based nucleotide gene trees for the selected avian species were reconstructed by providing the multiple sequence alignments to IQ-TREE v.1.6.11 (Trifunopoulos et al. 2016). First we used ModelFinder within IQ-TREE to choose the appropriate model of sequence evolution (Kalyaanamoorthy et al. 2017), as assessed by Bayesian Information Criterion (BIC) support measures among 185 codon models. The best substitution model for each locus was then used to infer the best gene tree using a maximum likelihood approach. For branch support analysis, we performed 1000 replicates for a non-parametric Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) (Guindon et al. 2010; Shimodaira and Hasegawa 1999) as well as 1000 alignments of ultrafast bootstrapping (Hoang et al. 2018). A 50% majority-rule consensus tree was constructed for each gene based on 1000 bootstrap trees. Graphical representation of gene trees was performed with FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>).

To allow comparison of the gene tree and species tree topologies, we created species cladograms corresponding to the species represented in each gene using the NCBI taxonomy database (www.ncbi.nlm.nih.gov/taxonomy). The Robinson–Foulds symmetric distance (RF) was calculated as the number of edges in the species tree but not in the gene tree plus the number of splits in the gene tree but not in the species tree (Robinson and Foulds 1981). Comparison and distance calculations were performed and visualized in Phylo.io (Robinson et al. 2016).

Detecting selection

Selection tests were performed using the HyPhy package via the DataMonkey Adaptive Evolution server (Pond et al. 2004; Weaver et al. 2018), based on the produced codon-based nucleotide MSAs (which allow silent substitutions to be taken into consideration) and the majority-rule consensus ML trees. The tests for selection are each based on calculations of the nonsynonymous-to-synonymous substitution rate ratios ($\omega = dN/dS$) using codon models and Likelihood Ratio Tests (LRTs), with null hypotheses based on neutral evolution ($\omega = 1$); negative (purifying) selection indicated by $\omega < 1$ and positive (diversifying) selection indicated by $\omega > 1$. The HyPhy methods first perform an initial global MG94xREV fit to optimize branch length and nucleotide substitution, which are used as initial parameter values during model fitting for hypothesis testing. An advantage of the inclusion of synonymous rate variation whereby dS is allowed to vary across sites and branches in the phylogeny, permitting more powerful detection of positive selection and reducing false discovery (Weaver et al. 2018). We sought to detect hierarchical signals of selection across entire genes, at particular codon sites within a gene, and at particular branches of the gene trees. Finally, we tested for differential selection among subsets of species to determine relative selection intensity by clade for each locus.

Gene-wide

To identify positive selection anywhere on the three gene trees, we used the algorithm BUSTED (Branch-site Unrestricted Statistical Test for Episodic Diversification) (Murrell et al. 2015). BUSTED fits a codon model with three rate classes, constrained as $\omega_1 \leq \omega_2 \leq 1 \leq \omega_3$ and estimates the proportion of sites belonging to each ω class. Positive selection is then detected by comparing this model fit to a null model where positive selection is not allowed ($\omega_3 = 1$). If the null hypothesis is rejected, then there is evidence that at least one site has experienced positive selection, at least in one branch.

At sites

To determine at which codon sites, if any, positive selection was detected, the algorithm MEME (Mixed Effects Model of Evolution) was used (Murrell et al. 2012). MEME uses a mixed-effects maximum likelihood approach to test the hypothesis that individual sites have been subject to episodic positive or diversifying selection in a proportion of branches. Two ω rate classes were inferred per site, and corresponding weights (the proportion of branches evolving under that rate class) were calculated. The two rate classes were inferred by a single dS value (α) and two separate dN values (β^- and β^+) per site. In the null model, β^- and β^+ were constrained to be less than or equal to α , but in the alternative model β^+ was not constrained. If $\beta^+ > \alpha$ at a site, and was significant using the likelihood ratio test, positive selection was inferred for the site.

At branches

To test which individual branches (lineages) are subject to selection at one or more sites, we used the algorithm aBSREL (adaptive Branch-Site Random Effects Likelihood) was used (Smith et al. 2015). aBSREL models both site-level and branch-level ω heterogeneity and infers the optimal number of ω classes for each branch using AIC_c (small sample AIC). The alternative model was compared to a null model in which positive selection was not allowed in the rate classes, and a Likelihood Ratio Test was performed at each branch. p-values at each branch were corrected for multiple testing using the Holm-Bonferroni correction before determining significance (Holm 1979).

To detected differential selection among each of the three clades, we used the algorithm RELAX to determine if selection pressure was significantly relaxed or intensified on those branches relative to the rest of the tree (Wertheim et al. 2015). RELAX uses a random effects branch-site model to test whether a set of test branches evolves under a different stringency of selection than a set of reference branches. For the null model, a codon model with three ω classes was fitted to the entire phylogeny. The test for changes in selection stringency involved the selection intensity parameter k (≥ 0), which served as an exponent to the ω classes in the alternative model. Upon likelihood ratio testing between the null and alternative models, a significant result of $k > 1$ indicated that selection strength was intensified along the test branches relative to the reference branches (i.e. the clade in question relative to the rest of the gene tree), and a significant result of $k < 1$ indicated that selection strength was relaxed along the test branches. The RELAX algorithm shows a Type I error rate of 0.052 (Wertheim et al. 2015).

Results

Sequence selection

Our search of available avian sequences produced suitable full CDS sequences for 34 avian species for *Mhc1*, 29 species for *Tlr2b*, and 37 species of *Ubb*. See Supplementary Tables S1–S3 for the summaries of the species and their sequences, and Supplementary Figs. S1–S3 for the fasta CDS sequences. Overall, 18 species common between both immune loci and 13 overlapped at all three loci.

For *Mhc1* these included four species of Paleognathae (in three orders), six species of Galloanserae (in two orders), and 24 species of Neoaves (in 10 orders). The lengths of the *Mhc1* sequences ranged from 759 to 1422 bp. In *Tlr2b* the sequences included four species of Paleognathae (in three orders), five species of Galloanserae (in two orders), and 20 species of Neoaves (in 12 orders). The length of the *Tlr2b* locus ranged from 1464 to 2440 bp. In *Ubb*, sequences

ranged from 462 to 1378 bp, and included three species of Paleognathae (in three orders), five species of Galloanserae (in two orders), and 29 species of Neoaves (in 22 orders). The differences in sample sizes between the clades are reflective of the differences in the real number of species in those clades: of the 10,000 extant birds, ~0.5% are in the Paleognathae clade, 4.5% are in Galloanserae, and 95% are in the Neoaves clade (Eo et al. 2009; Sibley and Monroe 1990).

Multiple sequence alignments

The CDS sequences for each gene were translated to protein sequences, which after protein alignment and curation produced amino acid MSAs of 295aa for *Mhc1*, 783aa for *Tlr2b*, and 302aa for *Ubb* (Supplementary Figs. S4–S6). Back translation to DNA sequences produced in-frame codon-based nucleotide MSAs of 885 bp for *Mhc1* (Fig.

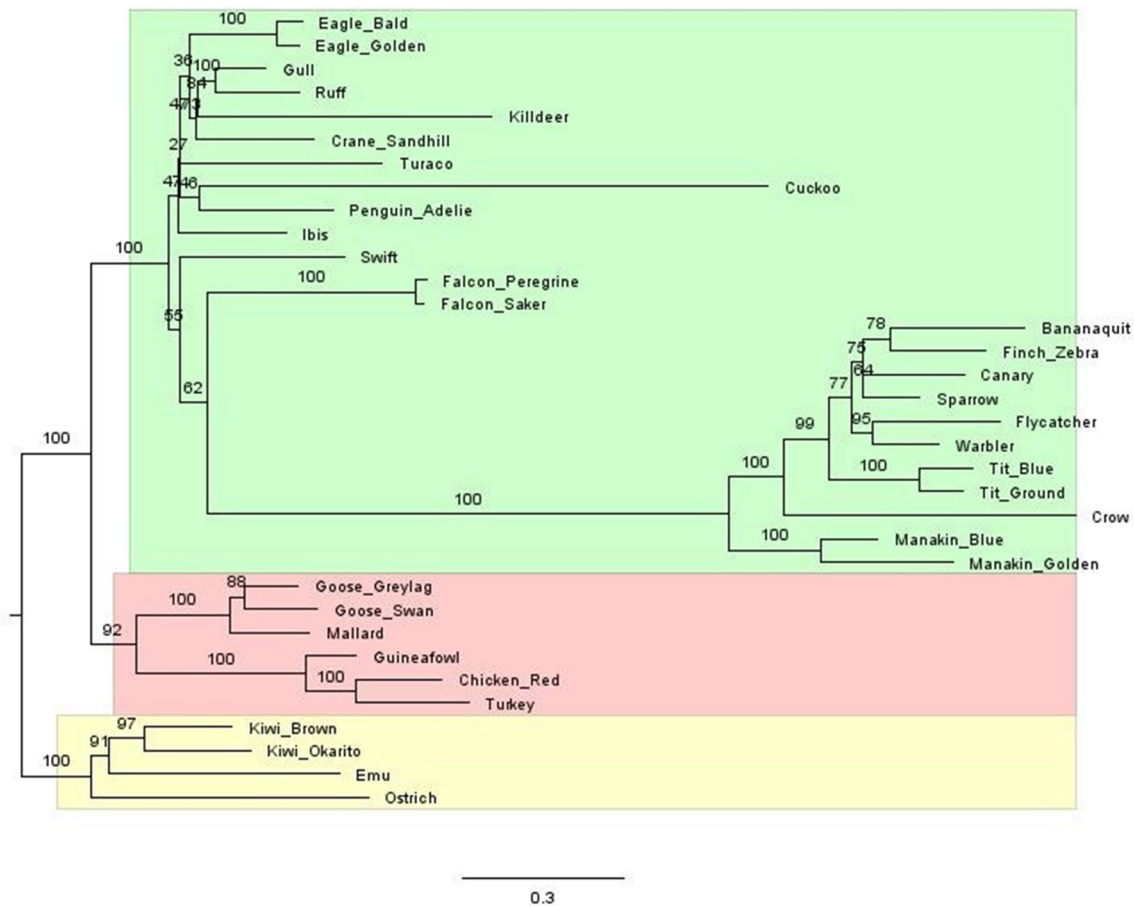


Fig. 1 The gene tree of avian *Mhc1* based on the majority-rule consensus maximum-likelihood tree of 34 species, produced by IQ-Tree (Trifinopoulos et al. 2016) and visualized with FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). Yellow highlighting indicates species in the Paleognathae clade, pink indicates the Gal-

loanserae clade, and green indicates the Neoaves clade. Branches are labeled with Ultrafast bootstrap supports (Hoang et al. 2018). Scale bar shows molecular distance (number of nucleotide substitutions per codon site)

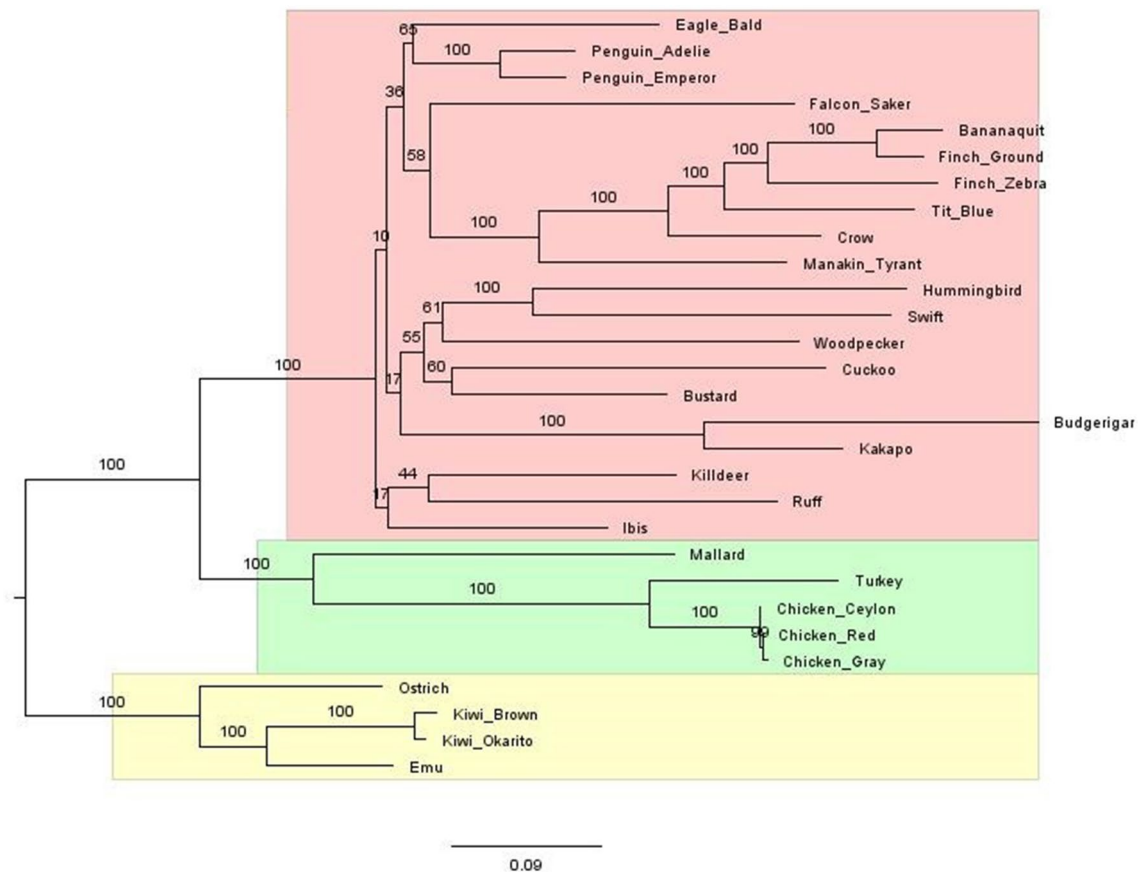


Fig. 2 The gene tree of avian *Tlr2b* based on the majority-rule consensus maximum-likelihood tree of 29 species, produced by IQ-Tree (Trifinopoulos et al. 2016) and visualized with FigTree v. 1.4.3 (<http://tree.bio.ed.ac.uk/software/figtree/>). Yellow highlighting indicates species in the Paleognathae clade, pink indicates the Gal-

loanserae clade, and green indicates the Neoaves clade. Branches are labeled with Ultrafast bootstrap supports (Hoang et al. 2018). Scale bar shows molecular distance (number of nucleotide substitutions per codon site)

S4), 2325 bp for *Tlr2b* (Fig. S5), and 906 bp for *Ubb* (Supplementary Figs. 7–9). Each are numbered according to the chicken sequence. For *Mhc1*, 1.7% of the 295 codon sites were invariant (constant or ambiguous constant) and 267 sites were parsimony informative. For *Tlr2b*, 13.3% of the 783 codon sites were invariant and 558 sites were parsimony informative. For *Ubb*, 11.9% of the 302 codon sites were invariant and 219 sites were parsimony informative.

Construction of individual gene trees

For *Mhc1*, the best-fit codon substitution model determined by BIC within 95% confidence was MGK + F3X4 + R4, and for *Tlr2b* it was MGK + F3X4 + R3. In this nomenclature, MG represents the codon substitution model of Muse and Gaut (1994), with K representing the addition of a dN/dS rate ratio as well as a transition/transversion (ts/tv) rate ratio. The frequency type F3X4 denotes unequal nucleotide frequencies overall as well as unequal nucleotide frequencies

over the three codon positions. The rate type R represents “free rate” heterogeneity, which relaxes the assumption of Gamma distributed rates (Soubrier et al. 2012; Yang 1995). Thus, the differences in the MHC and TLR models reflect different number of categories of rates. In *Mhc1*, these four categories in the form of (proportion of sites, relative rate) were (0.0999, 0.1638), (0.3670, 0.5176), (0.4459, 1.2027), and (0.0872, 2.9511). The three categories among *Tlr2b* were (0.3651, 0.2744), (0.5185, 1.0806), and (0.1164, 2.9175).

The best-fit codon substitution model for *Ubb* was KOSI07 + F + G4. Here KOSI07 represents the empirical codon model of Kosiol et al. (2007), with F representing empirical codon frequencies counted from the data. The rate type G4 indicates the discrete Gamma model of Yang (1994) with four rate categories. These were modeled as: (0.25, 0.2406), (0.25, 0.6047), (0.25, 1.0559), and (0.25, 2.0988).

The best Maximum Likelihood trees were inferred from the codon-based nucleotide alignment and substitution model. For *Mhc1* the best ML tree had a total tree length

(the sum of branch lengths, each representing number of nucleotide substitutions per codon site) of 13.8, with the sum of the internal branch lengths representing 31.3% of the tree length. For *Tlr2b*, the total gene tree length was 4.8, with the internal branch lengths constituting 30.4% of the total length. For *Ubb*, the total gene tree length was 5.3 and 20.2% of the total length was represented by the internal branches. The majority-rule consensus gene trees produced by subsequent bootstrapping are shown in Fig. 1 for *Mhc1* and Fig. 2 for *Tlr2b*.

Comparison of the gene trees with corresponding species trees (cladograms) showed differences at each locus, as visualized in Fig. S10 for *Mhc1*, Fig. S11 for *Tlr2b*, and Fig. S12 for *Ubb*. The RF distance and Euclidean distance between the *Mhc1* gene tree and its species tree were 18 and 14.3, respectively, while for *Tlr2b* the distances were 15 and 13.5. For *Ubb* the RF distance was 6 and the Euclidean distance was 4.0.

Detecting selection

Gene-wide

For both *Mhc1* and *Tlr2b*, the BUSTED algorithm found evidence (LRT, p value ≤ 0.05) of episodic diversifying selection. Evidence shows that for each gene tree, at least one site on one branch has experienced diversifying selection. For *Ubb*, no evidence for diversifying selection was found anywhere in the gene (LRT, $p = 0.40$) (Table 1). Therefore, no expanded testing for diversifying selection was performed with the *Ubb* sequences.

At sites

The MEME algorithm found evidence of episodic positive/diversifying selection (LRT, p -value ≤ 0.05) at particular codon sites in both immune loci. For the *Mhc1* phylogeny,

25 out of 295 codon sites (8.5%) showed evidence for selection (Table 2), and 21 of 783 codon sites (2.7%) showed evidence for selection in *Tlr2b* (Table 3). These codon sites correspond to the numbering of the curated protein alignments in Figs. S3 and S4.

At branches

The algorithm aBSREL tested each locus for evidence of branch-specific selection, finding multiple lineages under selection in each locus (Table 4). In *Mhc1*, selection was detected at 13 (of the 65 = 20%) branches/nodes in the phylogeny (LRT, p -value ≤ 0.05). These included the American crow (*Corvus brachyrhynchos*), Golden-collared manakin (*Manacus vitellinus*), Red-billed gull (*Chroicocephalus scopulinus*), Sandhill crane (*Antigone Canadensis*), North Island brown kiwi (*Apteryx australis*), Common ostrich (*Struthio camelus*), Killdeer (*Charadrius vociferous*), Bananaquit (*Coereba flaveola*), Blue-crowned manakin (*Lepidothrix coronata*), Ruff (*Calidris pugnax*), and Chimney swift (*Chaetura pelagica*). Additionally selection was found at the node representing the divergence of the Peregrine falcon and the Saker falcon (*Falco peregrinus* and *Falco cherrug*), and the node representing the divergence of the Common ostrich from the node that represents the divergence of the Emu (*Dromaius novaehollandiae*) from the North Island brown kiwi and the Okarito kiwi (*Apteryx rowi*).

In *Tlr2b*, selection was detected at 5 (of the 55 = 9.1%) branches/nodes in the gene tree (LRT, p -value ≤ 0.05). In *Tlr2b*, these include the Common starling (*Sturnus vulgaris*), Bald eagle (*Haliaeetus leucocephalus*), Emperor penguin (*Aptenodytes forsteri*), and Budgerigar (*Melopsittacus undulatus*), and the node representing the divergence of the Chimney swift and Anna's hummingbird (*Calypte anna*). Selection was also detected at the node representing the divergence of the clade consisting of the Bald eagle/Emperor penguin/

Table 1 Statistical results of model fits for the BUSTED (Branch-site Unrestricted Statistical Test for Episodic Diversification) algorithm (Murrell et al. 2015) performed with the Datamonkey server (Weaver et al. 2018) showing that for both *Mhc1* and *Tlr2b*, there is evidence of episodic diversifying selection in the gene (the null model of no positive selection, $\omega_3 = 1$, is rejected; LRT: $p < 0.05$)

Model	log L	#. params	AIC _c	ω_1 (Negative)	ω_2 (Neutral)	ω_3 (Positive)
<i>Mhc1</i> (34 sequences, 295 codon sites)						
Alternative Model	-14078.6	84	28,326.6	0.02 (63.46%)	1.00 (32.83%)	16.79 (3.71%)
Null Model	-14197.1	83	28,561.5	0.00 (59.31%)	1.00 (4.03%)	1.00 (36.66%)
<i>Tlr2b</i> (29 sequences, 783 codon sites)						
Alternative Model	-18860.2	74	37,868.9	0.00 (62.78%)	1.00 (37.09%)	656.19 (0.13%)
Null Model	-18920.0	73	37,986.4	0.00 (63.32%)	1.00 (2.09%)	1.00 (34.59%)
<i>Ubb</i> (37 sequences, 302 codon sites)						
Alternative Model	-5969.4	90	12,120.3	0.00 (0.99%)	0.00 (98.29%)	1.70 (0.72%)
Null Model	-5970.3	89	12,120.1	0.00 (1.37%)	0.00 (97.58%)	1.00 (1.04%)

For the control gene *Ubb*, there is no evidence of diversifying positive selection in the gene (the null model of $\omega_3 = 1$ is not rejected; LRT: $p = 0.40$)

Table 2 The results of the MEME (Mixed Effects Model of Evolution) algorithm (Murrell et al. 2012) performed with the Datamonkey server (Weaver et al. 2018) for the *Mhc1*, for the 25 out of 295 (= 8.5%) codon sites under positive/diversifying selection ($p \leq 0.050$)

Site	α	β^-	p^-	β^+	p^+	LRT	p-value
2	0.09	0.09	0.96	52.02	0.04	16.1	0.00
10	0.28	0	0.87	5.16	0.13	6.05	0.02
22	2.02	0	0.71	19.1	0.29	8.98	0.00
30	0	0	0.61	2.91	0.39	7.75	0.01
32	1.15	0	0.75	10.57	0.25	4.91	0.04
37	0	0	0.9	5.32	0.1	4.74	0.04
46	0.52	0.49	0.93	21.36	0.07	5.03	0.04
60	0.5	0	0.67	7.83	0.33	6.36	0.02
64	1.02	0	0.61	18.5	0.39	10.2	0.00
71	0.78	0.42	0.64	12.55	0.36	7.38	0.01
78	0.56	0	0.85	11.27	0.15	10.33	0.00
88	0.25	0.25	0.9	18.59	0.1	7.98	0.01
91	0	0	0.65	17.72	0.35	12.43	0.00
109	0.25	0.25	0.57	53.1	0.43	12.55	0.00
111	1.46	1.46	0.27	9.05	0.73	7.31	0.01
126	0	0	0.66	3.02	0.34	7.97	0.01
146	0.66	0.26	0.82	11.88	0.18	8.76	0.01
149	0.46	0.37	0.97	88.71	0.03	7.07	0.01
159	0.48	0.48	0.96	558.24	0.04	12.76	0.00
169	0.56	0	0.71	3.9	0.29	4.47	0.05
180	0.17	0	0.78	5.26	0.22	7.56	0.01
190	0.2	0	0.8	2.68	0.2	4.8	0.04
217	0.53	0	0.8	5.96	0.2	4.98	0.04
241	0.07	0.07	0.91	6.59	0.09	6.98	0.01
272	0.14	0.14	0.93	39.36	0.07	5.85	0.02

Site numbers correspond to codon sites as numbered in the curated peptide alignment (Fig. S4). α = synonymous substitution rate; β^- = non-synonymous substitution rate for the negative/neutral evolution component; p^- = proportion of tree evolving neutrally or under negative selection; β^+ = non-synonymous substitution rate for the positive/neutral evolution component; p^+ = proportion of tree evolving neutrally or under positive selection; LRT = likelihood ratio test statistic

Adelie penguin (*Pygoscelis adeliae*) with the clade consisting of the Saker falcon/American crow/Bananaquit/Eurasian blue tit (*Cyanistes caeruleus*)/Zebra finch (*Taeniopygia guttata*)/Ground finch (*Geospiza fortis*)/Saffron-crested tyrant-manakin (*Neopelma chrysocephalum*).

The results of the RELAX algorithm (Fig. 3 and Table S4) indicate that for *Mhc1* the Galloanserae clade evolved at a relaxed rate ($K=0.80$) relative to the rest of the tree, while the Neoaves clade evolved at an intensified rate ($K=1.28$) relative to the rest of the tree. The Paleognathae did not show significant relaxation or intensification of selection. For *Tlr2b*, none of the three clades showed significant differences in evolutionary rates among them.

Discussion

Our study is unique in comparing molecular evolution among representative of both the MHC and TLR gene families, and our results indicate that the MHC gene is under stronger diversifying selection than the TLR gene. In summary, in both immune genes, we detected episodic diversifying positive selection at the overall gene level, at the lineage level, and at the individual codon site level, with a stronger effect on *Mhc1* than *Tlr2b*. We additionally found evidence of differential selection pressures among the clades for *Mhc1*, with evolution intensified in the Neoaves but relaxed in the Galloanserae. Diversifying

Table 3 The results of the MEME (Mixed Effects Model of Evolution) algorithm (Murrell et al. 2012) performed with the Datamonkey server (Weaver et al. 2018) for the *Tlr2b* locus, for the 21 out of 783 (=2.7%) of codon sites under positive/diversifying selection ($p \leq 0.050$)

Site	α	β^-	p^-	β^+	p^+	LRT	p-value
7	0	0	0	2.4	1	8.66	0.01
99	0.97	0.52	0.74	15.26	0.26	6.76	0.02
124	0.36	0.36	0.95	12.28	0.05	4.7	0.04
162	0	0	0	2.81	1	8.53	0.01
176	0	0	0	1.64	1	6.21	0.02
208	0.33	0.33	0	2.71	1	4.93	0.04
274	0	0	0	1	1	4.63	0.05
297	0	0	0	3.24	1	10.43	0.00
298	0	0	0	2.79	1	5.59	0.03
302	0	0	0	1	1	5.21	0.03
317	0	0	0.97	34.45	0.03	10.09	0.00
331	0	0	0.14	2.98	0.86	10.54	0.00
343	0.36	0	0.51	6.44	0.49	6.15	0.02
374	0.48	0.45	0.96	150.3	0.04	14.58	0.00
430	0.36	0.24	0.92	315.91	0.08	11.72	0.00
452	1.21	0.31	0.88	21.75	0.12	4.82	0.04
679	0	0	0.98	155.24	0.02	17.34	0.00
746	1.27	0.16	0.98	310.19	0.02	5.98	0.02
748	0	0	0.98	550.2	0.02	22.59	0.00
766	0.71	0.09	0.87	93.77	0.13	5.01	0.04
775	0.51	0	0.88	7102	0.12	23.99	0.00

Site numbers correspond to codon sites as numbered in the curated peptide alignment (Fig. S5). α =synonymous substitution rate; β^- =non-synonymous substitution rate for the negative/neutral evolution component; p^- =proportion of tree evolving neutrally or under negative selection; β^+ =non-synonymous substitution rate for the positive/neutral evolution component; p^+ =proportion of tree evolving neutrally or under positive selection; LRT=likelihood ratio test statistic

selection at the *Mhc1* and *Tlr2b* loci is apparent when their patterns of molecular evolution are contrasted with *Ubb*, which did not show evidence for selection anywhere in the gene.

Mhc1 and *Tlr2b* have diverged at different rates, with *Mhc1* having almost three times the gene tree length (the sum of the branches which represent substitutions per codon site) of *Tlr2b* and thus evolving almost thrice as fast (Figs. 1 and 2). The topologies between gene trees and species trees differ more in *Mhc1* than *Tlr2b* based on distance metrics, and both differ more than *Ubb* (Figs. S10–S12). These differences occur primarily in the Neoaves clade, which rapidly radiated at the Cretaceous–Paleogene (K–Pg) boundary about 66MYA, and incomplete lineage sorting is indicated at these loci (Suh et al. 2015).

We found episodic positive diversifying selection operating on particular codon sites in both loci, with a higher proportion at the adaptive immune system gene *Mhc1* (8.5%) than the innate immune system gene *Tlr2b* (2.7%). Another recent report in birds found that in exon 3 of *Mhc1*, which encodes a portion of the variable peptide-binding region, 29% of codon sites were under positive selection in a group of 23 passerine families and 14% of sites were under positive selection in a group of 10 non-passerine Neognathae

families (Minias et al. 2018). The increased percentage of sites under selection compared to our study is attributable at least in part to the sequences representing only exon 3, and not the full alpha chain including conserved domains, as well as their choice of species, which is consistent with our findings of increased selection intensity in *Mhc1* in Neoaves. Studies of interspecific detection of site-level selection in *Mhc1* genes in other taxa are scarce. In one, among six species of frogs from three families, approximately 8.0% of sites showed evidence of positive selection in the full alpha chain, based on average results from four different methods (5.0–9.7%) (Kiemnec-Tyburczy et al. 2012), on par with the 8.5% we found in birds.

While we found a larger portion of the *Mhc1* locus under diversifying selection than the *Tlr2b* in Class Aves, the percentage of the *Tlr2b* gene showing significant positive selection is higher in this study (2.7%) than in most previously reported, such as 1.6% in a comparison of seven species representing (Alcaide and Edwards 2011) and 0.3–3.0% in a comparison of 14 avian species (Grueber et al. 2014), with sequences representing only the variable LRR region and with no species included from the Galloanserae clade. The higher percentage of the *Tlr2b* gene under positive selection is in accordance with a recent study showing 4.5% of

Table 4 Results for the aBSREL (adaptive Branch-Site Random Effects Likelihood) algorithm (Smith et al. 2015) as determined in the Datamonkey server (Weaver et al. 2018)

Branch/Node	LRT test statistic	p-value (corrected for multiple testing)	ω distribution over sites	
<i>Mhc1</i>				
CROW	27.27	<0.00	$\omega_1 = 0.520$ (88%)	$\omega_2 = 5000$ (12%)
MANAKIN_GOLDEN	30.18	<0.00	$\omega_1 = 0.762$ (93%)	$\omega_2 = \infty$ (7.1%)
GULL	24.96	<0.00	$\omega_1 = 0.220$ (94%)	$\omega_2 = 33.2$ (5.7%)
CRANE_SANDHILL	23.71	<0.00	$\omega_1 = 0.334$ (94%)	$\omega_2 = 5000$ (5.6%)
KIWI_BROWN	22.65	<0.00	$\omega_1 = 0.346$ (96%)	$\omega_2 = 168$ (4.4%)
OSTRICH	23.41	<0.00	$\omega_1 = 0.159$ (87%)	$\omega_2 = 37.2$ (13%)
KILLDEER	19.45	0.00	$\omega_1 = 0.206$ (85%)	$\omega_2 = 5000$ (15%)
Node42	18.90	0.00	$\omega_1 = 0.370$ (88%)	$\omega_2 = 16.0$ (12%)
BANANAQUIT	18.25	0.00	$\omega_1 = 0.437$ (95%)	$\omega_2 = 62.9$ (4.8%)
MANAKIN_BLUE	17.40	0.00	$\omega_1 = 0.231$ (94%)	$\omega_2 = 28.8$ (5.9%)
Node12	17.25	0.00	$\omega_1 = 0.0596$ (94%)	$\omega_2 = 41.6$ (6.4%)
RUFF	17.09	0.00	$\omega_1 = 0.0586$ (85%)	$\omega_2 = 9.54$ (15%)
SWIFT	13.67	0.02	$\omega_1 = 0.359$ (95%)	$\omega_2 = 141$ (5.4%)
<i>Tlr2b</i>				
EAGLE_BALD	57.47	<0.00	$\omega_1 = 0.850$ (97%)	$\omega_2 = \infty$ (2.5%)
PENGUIN_EMPEROR	29.93	<0.00	$\omega_1 = 0.303$ (100%)	$\omega_2 = 660$ (0.50%)
Node18	24.25	<0.00	$\omega_1 = 0.216$ (100%)	$\omega_2 = \infty$ (0.14%)
Node40	20.14	<0.00	$\omega_1 = 0.394$ (99%)	$\omega_2 = \infty$ (0.71%)
BUDGERIGAR	19.04	0.00	$\omega_1 = 0.242$ (97%)	$\omega_2 = 21.9$ (3.0%)

For *Mhc1* evidence of episodic positive selection occurred in 13 branches/nodes (20%) i(LRT, p-value ≤ 0.050). Node42 is the node representing the divergence of the Peregrine falcon and the Saker falcon, and Node12 is the node representing the divergence of the ostrich from the node that represents the divergence of the emu from the kiwis. For *Tlr2b* evidence of episodic positive selection occurred at 5 branches/nodes (9.1%) (LRT, p-value ≤ 0.050). Node40 is the node representing the divergence of the Swift and Hummingbird, and Node18 is the node representing the divergence of the Bald Eagle/Adelie Penguin/Emperor penguin with the Saker Falcon/Zebra finch/Ground finch/Bananaquit/Blue tit/Crow/Tyrant manakin

sites under selection in *Tlr2b* in 42 Neoaves species (Velova et al. 2018).

In mammals, 13 TLRs exist compared with 10 in birds; however, birds recognize a similarly broad range of microbial ligands as mammals. Dimerization between either of the two avian TLR2s (TLR2a and TLR2b) with the two avian TLR1s (TLR1La and TLR1b) mimics the pattern recognition of mammalian TLR2 dimerized with mammalian TLR1, TLR6, or TLR10 (Brownlie and Allan 2011). Site-level interspecific positive selection in TLR2 is reportedly lower in mammals than in our findings, with the percentage of the gene under selection ranging from 0.8% with 23 mammalian species to 2% with 27 mammalian species (Areal et al. 2011; Huang et al. 2011). A study on 11 primate species found no significant positive selection in TLR2 (Wlasiuk and Nachman 2010). This indicates that the avian *Tlr2b* could be under stronger evolutionary pressure than mammalian TLR2, despite the higher overall background substitution rate in mammals. However, it has been suggested the functional ortholog of mammalian TLR2 in birds is TLR2A as the ortholog of *Tlr2b* has been lost in mammals (Huang et al. 2011).

Our branch site tests showed 20% of branches/nodes in MHC and 9.1% in TLR under positive diversifying selection. These data agree with the site-level data in indicating that MHC evolves more rapidly than TLRs, presumably in response to both the intensity of selection and the underlying evolutionary mechanisms (positive vs. negative selection), consistent with adaptive immunity being less constrained than innate immunity. To our knowledge this study is the first to report tests for positive selection at the branch (lineage) level across taxa at either *Mhc1* or *Tlr2b* genes, indicating particular species in which positive selection is detectable (Table 4).

One of the lineages showing episodic diversifying selection at both immune loci was the bananaquit, a Caribbean passerine bird which is a common host of avian malaria. Its demography, evolutionary history, and parasite prevalence have been the subject of previous study (e.g. Bellemain et al. (2008), Ricklefs et al. (2011)) and its genome and immune genes have been characterized and studied at the population level (Antonides et al. 2017, 2019). The *Mhc1* and *Tlr2b* loci are key players in recognizing the ligands of avian malarial parasites, and the positive diversifying selection

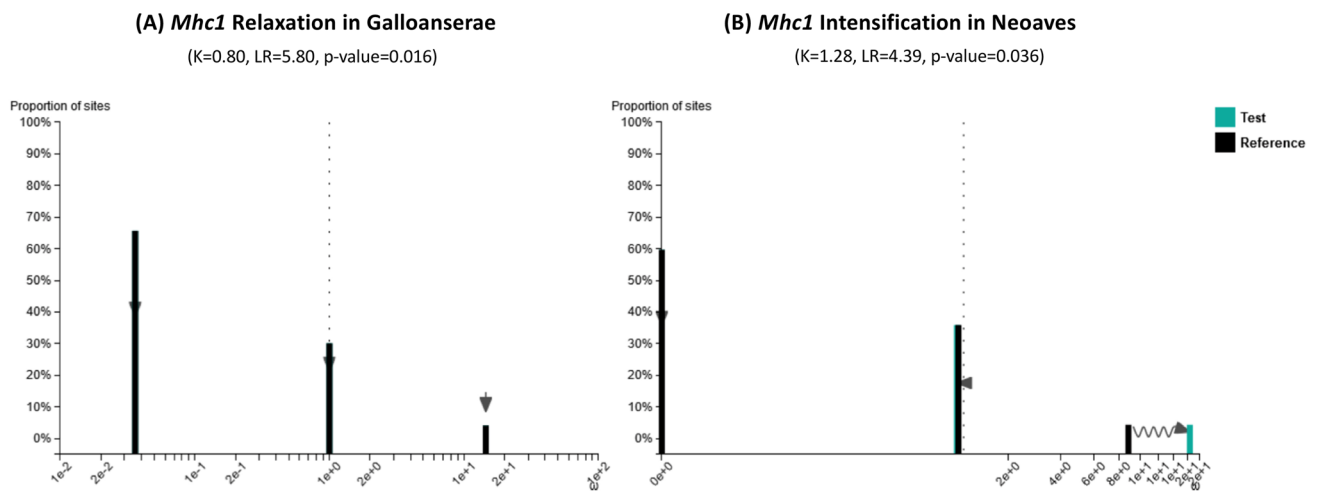


Fig. 3 Omega ($\omega=dn/ds$) distributions under the RELAX alternative model for the significant clades within *Mhc1* (LRT, $p \leq 0.05$) as determined by the RELAX algorithm (Wertheim et al. 2015) in the Datamonkey server (Weaver et al. 2018). In *Mhc1*, selection inten-

sity is significantly relaxed in Galloanserae and increased in Neoaves relative to the rest of the tree. Table S4 shows the full RELAX model results. *K* relaxation/intensification parameter; *LT* likelihood ratio

detected at these loci in the bananaquit lineage (without a priori specification) likely reflects the adaptability of a species whose population continues to expand in the face of haemosporidian parasites (Fahey et al. 2012). The knowledge of the evolutionary rates of immune genes in a species such as the bananaquit could potentially inform long-term viability estimates of species of conservation concern such as Hawaiian honeycreepers (Fringillidae), who, due to the range expansion of dipteran vectors, have been exposed to malarial parasites only recently in their evolutionary history (Foster et al. 2007; Liao et al. 2017).

Avian lineages occupy a wide range of habitat preferences and ecological niches, and their genotypic and phenotypic diversity is a reflection of that (e.g. Jetz et al. (2012); O'Connor et al. (2018)). Variation in evolutionary rates between lineages or clades would suggest evolution under varying selection pressures. Among the three major clades, the average genome-wide background substitution rates vary but are not significantly different: the mean rates and their ranges (in substitutions per site per million years) are 0.0015 (0.0010–0.0020) for Paleognathae, 0.0019 (0.0017–0.0020) for Galloanserae, and 0.0019 (0.0013–0.0036) for Neoaves (Jarvis et al. 2014). We found that *Mhc1* evolved at a relaxed rate in Galloanserae, a rate 80% of the rest of the species as a whole. We found the Neoaves *Mhc1* evolved at a more rapid rate, about 1.7 times that of the rest of the tree as a whole. This is consistent with a study on 16 Neoaves species and 3 Galloanserae on exon 3 of *Mhc1*, which found evidence of stronger adaptive evolution in passerines than in non-passerines (Alcaide et al. 2013). We did not find differential selection pressure among clades in the *Tlr2b* gene. Taken together, these data indicate the adaptive immune gene *Mhc1*

is more responsive to selection pressures among clades and has more functional significance over evolutionary timeframes than the innate immune gene *Tlr2b*.

Behavior, life history traits, and demography are all among factors that influence the selection pressures a particular lineage will encounter, and each has the potential to affect the evolution of their immune genes (e.g. (Lee et al. 2008; O'Connor et al. 2018)). The Neoaves underwent the highest known diversification rate among vertebrate radiations upon the K–T mass extinction that opened up an unprecedented variety of ecological niches (Alfaro et al. 2009; Feduccia 2003). Therefore, an increased likelihood of encountering parasites and a wider diversity of parasites could drive more rapid evolution of the *Mhc1* gene in Neoaves. Similarly, the Galloanserae and Paleognathae may have a decreased risk of encountering novel parasites (e.g. more limited dispersal in the flightless lineages of those clades) which may relax selection pressure in Galloanserae (*Mhc1*). However, as with the interpretation of the rest of our results, the necessarily small sample sizes (i.e., number of species) representing the Galloanserae and Paleognathae may impede detection of differences within and among them.

The framework utilized herein could be applied to detect varying modes and intensity of selection based on phylogenetic hypotheses. For example, migratory species are likely exposed to a wider variety of pathogens on average than non-migratory species, but migratory behavior has been shown to reduce disease risk (Hall et al. 2014), so one might hypothesize intensification of selection in immune genes in the migratory group. Conservation status is another example: vulnerable and endangered species often exhibit reduced genetic diversity in general (Willoughby et al. 2015), and

one might expect relaxation of selection on immune genes in those species relative to species of least concern because of the inefficiency of natural selection in small populations.

Conclusion

This study amassed a relatively large and phylogenetically diverse set of species for inference of avian divergence and signatures of selection at three loci, including representatives from both the adaptive and innate immune gene families. We found strong evidence of episodic diversifying selection in *Mhc1* and *Tlr2b* gene-wide, at codon sites, and at branches (lineages). The control *Ubb* gene did not show evidence of selection, which is expected given its conserved nature. This indicates our results reflect biological mechanisms: selection promotes amino acid diversity in immune genes in such a way as to widen the spectrum of pathogen ligands that can be recognized. *Mhc1* and *Tlr2b* have diverged at different rates, with *Mhc1* having almost three times the gene tree distance of *Tlr2b*, and over twice the percentage of sites and lineages under selection than *Tlr2b*, and thus evolving at least twice as fast. Our data provide insights into avian immune gene evolution, consistent with the adaptive immune system playing a more critical role in avian evolution than the innate immune system. Our data are also consistent with the adaptive immune system evolving more rapidly in the Neoaves than in other avian clades.

Acknowledgements We thank M. Christie, J. Dunning, R. Ricklefs, C. Searle and members of the DeWoody Lab for critical review of a previous version of this manuscript.

Funding This work was funded by the U.S. National Institute of Food and Agriculture, Purdue's Department of Forestry & Natural Resources, and the University Faculty Scholar program.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abascal F, Zardoya R, Telford MJ (2010) TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Res* 38:W7–13. <https://doi.org/10.1093/nar/gkq291>

Acevedo-Whitehouse K, Cunningham AA (2006) Is MHC enough for understanding wildlife immunogenetics? *Trends Ecol Evol* 21:433–438. <https://doi.org/10.1016/j.tree.2006.05.010>

Alcaide M, Edwards SV (2011) Molecular evolution of the Toll-like receptor multigene family in birds. *Mol Biol Evol* 28:1703–1715

Alcaide M, Liu M, Edwards SV (2013) Major histocompatibility complex class I evolution in songbirds: universal primers, rapid evolution and base compositional shifts in exon 3. *PeerJ* 1:e86. <https://doi.org/10.7717/peerj.86>

Alfaro ME et al (2009) Nine exceptional radiations plus high turnover explain species diversity in jawed vertebrates. *Proc Natl Acad Sci USA* 106:13410–13414. <https://doi.org/10.1073/pnas.0811087106>

Antonides J, Ricklefs R, DeWoody JA (2017) The genome sequence and insights into the immunogenetics of the bananaquit (Passeriformes: *Coereba flaveola*). *Immunogenetics* 69:175–186. <https://doi.org/10.1007/s00251-016-0960-8>

Antonides J, Mathur S, Sundaram M, Ricklefs R, DeWoody JA (2019) Immunogenetic response of the bananaquit in the face of malarial parasites. *BMC Evol Biol* 19:107. <https://doi.org/10.1186/s12862-019-1435-y>

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK (1997) The nature of selection on the major histocompatibility complex. *Crit Rev Immunol* 17:179–224. <https://doi.org/10.1615/CritRevImmunol.v17.i2.40>

Areal H, Abrantes J, Esteves PJ (2011) Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol Biol*. <https://doi.org/10.1186/1471-2148-11-368>

Bellemain E, Bermingham E, Ricklefs RE (2008) The dynamic evolutionary history of the bananaquit (*Coereba flaveola*) in the Caribbean revealed by a multigene analysis. *BMC Evol Biol* 8:240. <https://doi.org/10.1186/1471-2148-8-240>

Best A, White A, Boots M (2009) The implications of coevolutionary dynamics to host–parasite interactions. *Am Nat* 173:779–791. <https://doi.org/10.1086/598494>

Bonneaud C, Perez-Tris J, Federici P, Chastel O, Sorci G (2006) Major histocompatibility alleles associated with local resistance to malaria in a passerine. *Evolution* 60:383–389. <https://doi.org/10.1554/05-409.1>

Borghans JA, Beltman JB, De Boer RJ (2004) MHC polymorphism under host-pathogen coevolution. *Immunogenetics* 55:732–739

Brownlie R, Allan B (2011) Avian toll-like receptors. *Cell Tissue Res* 343:121–130. <https://doi.org/10.1007/s00441-010-1026-0>

Campos MA et al (2001) Activation of Toll-like receptor-2 by glycosylphosphatidylinositol anchors from a protozoan parasite. *J Immunol* 167:416–423

Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol Biol Evol* 17:540–552. <https://doi.org/10.1093/oxfordjournals.molbev.a026334>

Dalton DL, Vermaak E, Smit-Robinson HA, Kotze A (2016) Lack of diversity at innate immunity Toll-like receptor genes in the critically endangered white-winged flufftail (*Sarothrura ayresii*). *Sci Rep*. <https://doi.org/10.1038/srep36757>

Di Tommaso P et al (2011) T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Res* 39:W13–17. <https://doi.org/10.1093/nar/gkr245>

Downing T, Lloyd AT, O'Farrelly C, Bradley DG (2010) The differential evolutionary dynamics of avian cytokine and TLR gene classes. *J Immunol* 184:6993–7000. <https://doi.org/10.4049/jimmunol.0903092>

Eo SH, Bininda-Emonds ORP, Carroll JP (2009) A phylogenetic super-tree of the fowls (Galloanserae, Aves). *Zool Scr* 38:465–481. <https://doi.org/10.1111/j.1463-6409.2008.00382.x>

Eriksson EM, Sampaio NG, Schofield L (2014) Toll-like receptors and malaria—sensing and susceptibility. *J Trop Dis* 2:2

Fahey Ricklefs RE, Latta SC, DeWoody JA (2012) Comparative historical demography of migratory and nonmigratory birds from the Caribbean island of Hispaniola. *Evol Biol* 39:400–414

Feduccia A (2003) 'Big bang' for tertiary birds? *Trends Ecol Evol* 18:172–176. [https://doi.org/10.1016/S0169-5347\(03\)00017-X](https://doi.org/10.1016/S0169-5347(03)00017-X)

Foster JT, Woodworth BL, Eggert LE, Hart PJ, Palmer D, Duffy DC, Fleischer RC (2007) Genetic structure and evolved malaria resistance in Hawaiian honeycreepers. *Mol Ecol* 16:4738–4746. <https://doi.org/10.1111/j.1365-294X.2007.03550.x>

- Gill FB (2007) Ornithology. Macmillan, New York
- Gilroy DL, van Oosterhout C, Komdeur J, Richardson DS (2017) Toll-like receptor variation in the bottlenecked population of the endangered Seychelles warbler. *Anim Conserv* 20:235–250. <https://doi.org/10.1111/acv.12307>
- Grueber CE, Wallis GP, Jamieson IG (2013) Genetic drift outweighs natural selection at toll-like receptor (TLR) immunity loci in a reintroduced population of a threatened species. *Mol Ecol* 22:4470–4482. <https://doi.org/10.1111/mec.12404>
- Grueber CE, Wallis GP, Jamieson IG (2014) Episodic positive selection in the evolution of avian Toll-like receptor innate immunity genes. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0089632>
- Grueber CE et al (2015) Toll-like receptor diversity in 10 threatened bird species: relationship with microsatellite heterozygosity. *Conserv Genet* 16:595–611. <https://doi.org/10.1007/s10592-014-0685-x>
- Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 59:307–321. <https://doi.org/10.1093/sysbio/syq010>
- Hall RJ, Altizer S, Bartel RA (2014) Greater migratory propensity in hosts lowers pathogen transmission and impacts. *J Anim Ecol* 83:1068–1077. <https://doi.org/10.1111/1365-2656.12204>
- Hamilton WD (1980) Sex versus non-sex versus parasite. *Oikos* 35:282–290. <https://doi.org/10.2307/3544435>
- Hedrick PW (1994) Evolutionary genetics of the major histocompatibility complex. *Am Nat* 143:945–964. <https://doi.org/10.1086/285643>
- Hess CM, Edwards SV (2002) The evolution of the major histocompatibility complex in birds. *Bioscience* 52:423–431
- Higuchi M et al (2008) Combinational recognition of bacterial lipoproteins and peptidoglycan by chicken Toll-like receptor 2 subfamily. *Dev Comp Immunol* 32:147–155. <https://doi.org/10.1016/j.dci.2007.05.003>
- Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS (2018) UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 35:518–522. <https://doi.org/10.1093/molbev/msx281>
- Holm S (1979) A simple sequentially rejective multiple test procedure. *Scand J Stat* 6:65–70
- Huang YH, Temperley ND, Ren LM, Smith J, Li N, Burt DW (2011) Molecular evolution of the vertebrate TLR1 gene family—a complex history of gene duplication, gene conversion, positive selection and co-evolution. *BMC Evol Biol*. <https://doi.org/10.1186/1471-2148-11-149>
- Janeway C (2005) Immunobiology: the immune system in health and disease, 6th edn. Garland Science, New York
- Jarvis ED et al (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331. <https://doi.org/10.1126/science.1253451>
- Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. *Nature* 491:444–448. <https://doi.org/10.1038/nature11631>
- Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 14:587–589. <https://doi.org/10.1038/nmeth.4285>
- Khan I et al (2019) The vertebrate TLR supergene family evolved dynamically by gene gain/loss and positive selection revealing a host-pathogen arms race in birds. *Diversity* 11:131
- Kiemiec-Tyburczy KM, Richmond JQ, Savage AE, Lips KR, Zamudio KR (2012) Genetic diversity of MHC class I loci in six non-model frogs is shaped by positive selection and gene duplication. *Heredity* 109:146–155. <https://doi.org/10.1038/hdy.2012.22>
- Kimura Y, Tanaka K (2010) Regulatory mechanisms involved in the control of ubiquitin homeostasis. *J Biochem* 147:793–798. <https://doi.org/10.1093/jb/mvq044>
- Klein J (1986) Natural history of the major histocompatibility complex. Wiley, New York
- Kobe B, Kajava AV (2001) The leucine-rich repeat as a protein recognition motif. *Curr Opin Struct Biol* 11:725–732. [https://doi.org/10.1016/S0959-440x\(01\)00266-4](https://doi.org/10.1016/S0959-440x(01)00266-4)
- Kosiol C, Holmes I, Goldman N (2007) An empirical codon model for protein sequence evolution. *Mol Biol Evol* 24:1464–1479
- Krishnegowda G et al (2005) Induction of proinflammatory responses in macrophages by the glycosylphosphatidylinositols of *Plasmodium falciparum*—cell signaling receptors, glycosylphosphatidylinositol (GPI) structural requirement, and regulation of GPI activity. *J Biol Chem* 280:8606–8616. <https://doi.org/10.1074/jbc.m413541200>
- Lee KA, Wikelski M, Robinson WD, Robinson TR, Klasing KC (2008) Constitutive immune defences correlate with life-history variables in tropical birds. *J Anim Ecol* 77:356–363. <https://doi.org/10.1111/j.1365-2656.2007.01347.x>
- Liao W, Atkinson CT, LaPointe DA, Samuel MD (2017) Mitigating future avian malaria threats to hawaiian forest birds from climate change. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0168880>
- Minias P, Pikus E, Whittingham LA, Dunn PO (2018) A global analysis of selection at the avian MHC. *Evolution* 72:1278–1293. <https://doi.org/10.1111/evo.13490>
- Minias P, Pikus E, Anderwald D (2019) Allelic diversity and selection at the MHC class I and class II in a bottlenecked bird of prey, the White-tailed Eagle. *BMC Evol Biol* 19:2
- Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP (2009) Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc Natl Acad Sci USA* 106:7073–7078. <https://doi.org/10.1073/pnas.0811357106>
- Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL (2012) Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* 8:e1002764. <https://doi.org/10.1371/journal.pgen.1002764>
- Murrell B et al (2015) Gene-wide identification of episodic selection. *Mol Biol Evol* 32:1365–1371. <https://doi.org/10.1093/molbev/msv035>
- Muse SV, Gaut BS (1994) A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol* 11:715–724
- Nei M, Gu X, Sitnikova T (1997) Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proc Natl Acad Sci USA* 94:7799–7806
- Nei M, Rogozin IB, Piontkivska H (2000) Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proc Natl Acad Sci USA* 97:10866–10871
- O'Connor EA, Cornwallis CK, Hasselquist D, Nilsson JA, Westerdahl H (2018) The evolution of immunity in relation to colonization and migration. *Nat Ecol Evol* 2:841–849. <https://doi.org/10.1038/s41559-018-0509-3>
- Piertney SB, Oliver MK (2006) The evolutionary ecology of the major histocompatibility complex. *Heredity* 96:7–21. <https://doi.org/10.1038/sj.hdy.6800724>
- Piontkivska H, Rooney AP, Nei M (2002) Purifying selection and birth-and-death evolution in the histone H4 gene family. *Mol Biol Evol* 19:689–697. <https://doi.org/10.1093/oxfordjournals.molbev.a004127>
- Pond SLK, Frost SDW, Muse SV (2004) HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 21:676–679. <https://doi.org/10.1093/bioinformatics/bti079>
- Ricklefs RE, Dodge Gray J, Latta SC, Svensson-Coelho M (2011) Distribution anomalies in avian haemosporidian parasites in the southern Lesser Antilles. *J Avian Biol* 42:570–584

- Roach JC et al (2005) The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci USA* 102:9577–9582. <https://doi.org/10.1073/pnas.0502272102>
- Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. *Math Biosci* 53:131–147
- Robinson O, Dylus D, Dessimoz C (2016) Phylo.io: interactive viewing and comparison of large phylogenetic trees on the web. *Mol Biol Evol* 33:2163–2166. <https://doi.org/10.1093/molbev/msw080>
- Sepil I, Lachish S, Hinks AE, Sheldon BC (2013) Mhc supertypes confer both qualitative and quantitative resistance to avian malaria infections in a wild bird population. *Proc R Soc B*. <https://doi.org/10.1098/rspb.2013.0134>
- Shimodaira H, Hasegawa M (1999) Multiple comparisons of log-likelihoods with applications to phylogenetic inference. *Mol Biol Evol* 16:1114
- Shultz AJ, Sackton TB (2019) Immune genes are hotspots of shared positive selection across birds and mammals. *eLife* 8:e41815
- Sibley CG, Monroe BL (1990) *Distribution and taxonomy of birds of the world*. Yale University Press, New Haven
- Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL (2015) Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* 32:1342–1353. <https://doi.org/10.1093/molbev/msv022>
- Sommer S (2005) The importance of immune gene variability (MHC) in evolutionary ecology and conservation. *Front Zool* 2:16. <https://doi.org/10.1186/1742-9994-2-16>
- Soubrier J, Steel M, Lee MSY, Der Sarkissian C, Guindon S, Ho SYW, Cooper A (2012) The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol Biol Evol* 29:3345–3358. <https://doi.org/10.1093/molbev/mss140>
- Suh A, Smeds L, Ellegren H (2015) The dynamics of incomplete lineage sorting across the ancient adaptive radiation of neoavian birds. *PLoS Biol*. <https://doi.org/10.1371/journal.pbio.1002224>
- Trifinopoulos J, Nguyen LT, von Haeseler A, Minh BQ (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res* 44:W232–235. <https://doi.org/10.1093/nar/gkw256>
- Velova H, Gutowska-Ding MW, Burt DW, Vinkler M (2018) Toll-like receptor evolution in birds: gene duplication, pseudogenisation and diversifying selection. *Mol Biol Evol*. <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 Zool* 58:15–28
- Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Pond SLK (2018) Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol* 35:773–777. <https://doi.org/10.1093/molbev/msx335>
- Wertheim JO, Murrell B, Smith MD, Kosakovsky Pond SL, Scheffler K (2015) RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol* 32:820–832. <https://doi.org/10.1093/molbev/msu400>
- Willoughby JR et al (2015) The reduction of genetic diversity in threatened vertebrates and new recommendations regarding IUCN conservation rankings. *Biol Conserv* 191:495–503. <https://doi.org/10.1016/j.biocon.2015.07.025>
- Wlasiuk G, Nachman MW (2010) Adaptation and constraint at Toll-like receptors in primates. *Mol Biol Evol* 27:2172–2186. <https://doi.org/10.1093/molbev/msq104>
- Yang Z (1994) Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 39:306–314
- Yang Z (1995) A space-time process model for the evolution of DNA sequences. *Genetics* 139:993–1005
- Yang ZH, Nielsen R, Goldman N, Pedersen AMK (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449
- Zdobnov EM et al (2017) OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Res* 45:D744–D749. <https://doi.org/10.1093/nar/gkw1119>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.