Infection, Genetics and Evolution

Research paper

Purifying selection and concerted evolution of RNA-sensing toll-like receptors in migratory waders

Nynke Raven a, Simeon Lisovski a, Marcel Klaassen a, Nathan Lo b, Thomas Madsen a, Simon Y.W. Ho b, Beata Ujvari a,⁎

a Centre for Integrative Ecology, School of Life and Environmental Sciences, Deakin University, Waurn Ponds, VIC 3216, Australia
b School of Life and Environmental Sciences, The University of Sydney, Sydney, NSW 2006, Australia

A B S T R A C T

Migratory birds encounter a broad range of pathogens during their journeys, making them ideal models for studying immune gene evolution. Despite the potential value of these species to immunecology and disease epidemiology, previous studies have typically focused on their adaptive immune gene repertoires. In this study, we examined the evolution of innate immune genes in three long-distance migratory waders (order Charadriiformes).

We analysed two parts of the extracellular domains of two Toll-like receptors (TLR3 and TLR7) involved in virus recognition in the Sanderling (Calidris alba), Red-necked Stint (Calidris ruficollis), and Ruddy Turnstone (Arenaria interpres). Our analysis was extended to 50 avian species for which whole-genome sequences were available, including two additional waders. We found that the inferred relationships among avian TLR3 and TLR7 do not match the whole-genome phylogeny of birds. Further analyses showed that although both loci are predominantly under purifying selection, the evolution of the extracellular domain of avian TLR3 has also been driven by episodic diversifying selection. TLR7 was found to be duplicated in all five wader species and in two other orders of birds, Cuculiformes and Passeriformes. The duplication is likely to have occurred in the ancestor of each order, and the duplicated copies appear to be undergoing concerted evolution. The phylogenetic relationships of wader TLR7 matched those of the five wader species, but that of TLR3 did not. Instead, the tree inferred from TLR3 showed potential associations with the species’ ecology, including migratory behaviour and exposure to pathogens. Our study demonstrates the importance of combining immunological and ecological knowledge to understand the impact of immune gene polymorphism on the evolutionary ecology of infectious diseases.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

The regular, directed movements of large numbers of migratory animals have been shown to facilitate the long-distance dispersal of zoonotic pathogens (Altizer et al., 2011; Fritzschke McKay and Hoyle, 2016). Recent examples include the spread of Ebola virus by migratory fruit bats (Leroy et al., 2009), West Nile virus by migratory songbirds (Owen et al., 2006), and avian influenza viruses (AIV) by migratory waterbirds (Krauss et al., 2010; Marius et al., 2006; Verhagen et al., 2015). Long-distance migration also leads to increased metabolic activity (Altizer et al., 2011; Klaassen et al., 2012), and a concomitant reduction in immune function (Møller et al., 2004; Møller and Erritzøe, 1998; Røberg et al., 1998).

Each breeding area, stopover site, and wintering ground has its own unique pathogen community (Altizer et al., 2011), so that migratory birds encounter a greater diversity of pathogens than do non-migratory birds (Figuerola and Green, 2000). In combination with high bird density at stopover sites (Krauss et al., 2010), these factors increase the risk of pathogen exposure, and consequently the overall infection risk (Korpivnirak and Leung, 2015). Pathogens have been shown to impair migratory ability (Bradley and Altizer, 2005; van Gils et al., 2007), which ultimately results in reduced reproductive success (Asghar et al., 2011). The physiological and pathological impacts associated with the migratory lifestyle act as strong evolutionary drivers of avian immune systems (Minias et al., 2016; Møller and Erritzøe, 1998). Therefore, immune gene variation is likely to have important implications for understanding the disease ecology of migratory birds.

The innate immune system initiates host immune and inflammatory responses designed to eliminate a diverse community of pathogens (Ademre and Ulevitch, 2000; Hellgren, 2015; Takeda et al., 2003; Tschirren et al., 2013), and hence provides the first line of immunological defence in vertebrates (Medzhitov and Janeway, 1997). Innate immune genes such as the Toll-like receptors are highly conserved across species (Beutler and Rehli, 2002), but still exhibit high species-specific genetic variation (Cormican et al., 2009). Therefore, these genes are ideal...
During the last decade, environmental and anthropogenic factors have led to dramatic population declines (primarily due to habitat destruction) in migrants, with waders being a particularly noticeable example (Altizer et al., 2011; Wilcoze and Wikelski, 2008). These declines have also been linked to increased exposure and susceptibility to disease (Klaassen et al., 2012), underscoring an urgent need for studies of relevant factors such as innate gene diversity and function.

The Ruddy Turnstone (Arenaria interpres), Red-necked Stint (Calidris ruficollis), and Sanderling (Calidris alba) are long-distance migratory waders in the sandpiper family (Scolopacidae). These three species winter on different coastline areas in Australia, but follow similar migratory paths through eastern Asia to breeding grounds in northern Russia and the Arctic Circle, making stopovers at the same or similar sites (Lisovski et al., 2015; Minton et al., 2010; Minton et al., 2011). Recent studies in south-east Australia found a high prevalence (25–30%) of antibodies against AIV in the Ruddy Turnstone and Red-necked Stint, suggesting that these species frequently encounter the virus (Curran et al., 2014; Ferenczi, 2016). However, the Sanderling had no or a low level of AIV antibodies detected (Curran et al., 2014; Ferenczi, 2016). Recent research on AIV infections in migratory birds, notably regarding interspecific variation in susceptibility and dispersal potential, has focused on variation in ecological parameters or acquired immunity (Maxted et al., 2016). In contrast, the potential role of innate immune gene variation has largely been overlooked.

In this study we aimed to compare the levels of polymorphism in TLR3 and TLR7 across bird species. As TLRs have a diverse evolutionary history (Alcaide and Edwards, 2011), we commenced our study by cloning and sequencing loci encoding two distinct parts of the extracellular domain of TLR3 and TLR7 in A. interpres, C. ruficollis, and C. alba. We then investigated the diversity of these genes in the three waders. By including sequences from an additional 50 published avian genomes (Jarvis et al., 2015; Romanov et al., 2011; Zhang et al., 2014) we provide a comprehensive analysis of the selective forces involved in the evolution of avian TLR3 and TLR7.

### 2. Materials and methods

#### 2.1. Sampling and sequencing

We studied three migratory wader species from the sandpiper family (Scolopacidae): Ruddy Turnstone (Arenaria interpres), Sanderling (Calidris alba), and Red-necked Stint (Calidris ruficollis). All live birds were captured on beaches located within a 30 km radius, during a 3-day period in April 2013, from their wintering grounds on the coast of south-east Australia (breeding origin not known). Birds were caught using cannon nets by the Victorian Waders Study Group. Using a

### Table 1

| Comparison of the two RNA-recognizing Toll-like receptors (TLRs) in avian taxa. |
|---------------------------------|---------------------------------|
| **Location of TLR**             | **Across avian taxa**           |
| Length of gene                  | Endosomal membrane*             |
| Structure of the gene           | 2600 bp*                        |
| Duplicates discovered           | Conserved across species*       |
| PAMPs recognized                | No                              |
| Diseases in birds associated with TLRs | diRNA, approximately 40–50 bp long* |
|                                 | Duck reovirus, avian influenza virus*, pigeon paramyxovirus type 1*, West Nile virus, Marek's disease virus* |
|                                 | Endosomal membrane:*            |
|                                 | 3200 bp*                        |
|                                 | Conserved across species*       |
|                                 | Passerine birds*                |
|                                 | ssRNA, small interfering RNAs, self-RNA* |
|                                 | Pigeon paramyxovirus type 1*, avian influenza virus* |

* Bell et al. (2005)
  b Tanji et al. (2013).
  d Cormican et al. (2009).
  e Liu et al. (2008).
  f Zhang et al. (2015).
  g Downing et al. (2010); Koprivnikar and Leung (2015).
  h Li et al. (2015).
  i Wang et al. (2004).
  j Hu et al. (2015).
  k MacDonald et al. (2007).
capillary tube, a blood sample was taken (approx. 50 μl) from the brachi- 
al vein from each bird. All bird catching and sampling was in accordance 
with animal ethics and state research permits. Blood samples were 
stored at 4 °C for 24 h before being centrifuged to separate plasma and 
red blood cells. All samples were stored at − 80 °C. We obtained a total 
of 18 samples of A. interpres, 22 samples of C. alba, and 17 samples of C. 
ruficollis. Two specimens per species were used in the detailed analyses 
of TLR gene evolution, while the remaining specimens were used in the 
population-level comparisons.

Sequenced bird genomes were retrieved from the study by Zhang et al. 
(2014), along with those sequenced subsequently from the White-
throated Sparrow (Zonotrichia albicollis; Romanov et al., 2011) and Ruff 
(Calidris pugnax; genome sequencing and assembly project number 
PRJNA281024, released November 2015). A total of 50 bird genomes 
were analysed in the current study. Sequences of TLR3 and TLR7 from 
A. interpres that were generated during the course of this study were 
BLASTed against these 50 bird genomes. Scaffolds containing matches 
to the TLR3/TLR7 sequences from A. interpres were retrieved from 
GenBank and checked for completeness (i.e., containing complete open 
reading frames) and for the potential presence of multiple copies of the 
same gene. Only complete sequences were used in the study (Supple-
mentary Material S1, Tables S1 and S2).

DNA was extracted from 20 μl of red blood cells using the phenol/ 
chloroform technique (Maniatis et al., 1982). The quantity and quality 
of the extracted DNA was established using a NanoDrop 2000 
(ThermoFisher Scientific, Waltham, MA, USA), and the samples were di-
luted to 50 ng/μl for further analyses. The primers were designed to am-
ify regions of the ECD, TLR3 primers amplified 800 bp in the central 
region of the ECD, the primers for TLR7 amplified a 1000 bp region at 
the 3′ end of the ECD and a fraction of 5′ TIR. A detailed description of 
primer design, polymerase chain reactions, cloning, and sequencing 
can be found in the Supplementary material S1 (Table S3). All sequences 
generated in this study have been deposited in NCBI GenBank, with ac-
ccess numbers KX823453–KX823575.

2.2 Characterization of wader TLR3 and TLR7 genes

Clones generated from the two specimens per wader species were included 
in the analyses described below. Sequences were aligned and 
checked in BioEdit 7.0 (Hall, 1999). Clone sequences sharing 100% iden-
tity were discarded. The molecular structures of wader TLR3 and TLR7 
were determined by aligning the clone sequences generated in this 
study with sequences from modern human (GenBank NP_003256, 
TLR3; NP_057646, TLR7), mouse (GenBank NP_569054, TLR3; 
NP_573474 XP_918297, TLR7), and Killdeer (GenBank XP_009882203, 
TLR3; XP_009893509, TLR7). LRR regions were determined using a com-
bination of Simple Modular Architecture Research Tool (SMART; Schultz 
et al., 1998) and previous LRR structural descriptions (Liu et al., 2008; 
Philbin et al., 2005; Tanji et al., 2013).

Three-dimensional protein structures were constructed using SWISS-
MODEL (Basィini et al., 2014). Templates for the ECD crystal structures of 
TLR3 (3CIG; Liu et al., 2008) and TLR8 (3W3K; Tanji et al., 2013) 
were retrieved from the Protein Data Bank (Berman et al., 2000) and were used 
with the cloned sequences. The entire TLR7 sequences generated by us 
were included in all the analyses except the 3D modelling. Since the tem-
plate available in the Protein Databank (PDB) only contained a truncated 
TLR molecule, only regions of our TLR7 sequences that corresponded to 
the PDB template were visualized in 3D models. LRR regions were iden-
tified following the descriptions by Liu et al. (2008) and Tanji et al. 
(2013) and by using SMART. Swiss Pdb-Viewer 4.1.0 (Guex et al., 
2009) was used to view the TLR molecular structures in 3D.

Genetic diversity indices of TLR3 and TLR7 within and between 
wader species were calculated using MEGA 6.0 (Tamura et al., 2013). 
Within- and between-group genetic distances were calculated using the 
Tajima-Nei substitution model (Tajima and Nei, 1984), with 500 
bootstraps and with a value of 1 for the alpha parameter of the gamma 
distribution. Amino acid (AA) changes were identified between the 
cloned sequences. AA side-chain charges, polarity, and hydropathy indi-
ces were retrieved from the study by Weast and Lide (1991). Physico-
chemical distances among AA states at variable sites were quantified 
using Grantham’s distance matrix (Grantham, 1974). Protein electrostatic 
and hydrophilic potentials were calculated using cloned sequences in 
PDB2PQR (Dolinsky et al., 2007; Dolinsky et al., 2004) using the PARSE 
force-field and electrostatic calculation performed in the APBS web solv-
er (Baker et al., 2001). Surface charge distribution was visualized using 
PyMOL educational v 1.7.4 (DeLano, 2002).

2.3. Phylogenetic analysis of TLR3 and TLR7 genes across birds

In our phylogenetic analyses, we included TLR3 and TLR7 sequences 
identified from the 50 published bird genomes as well as representative 
sequences from the three wader species generated in this study. Since 
TLR3 has not been found to be duplicated in any of the 50 bird genomes, 
only one TLR3 sequence per wader specimen was included in the anal-
yses. Owing to two copies of TLR7 being identified in some of the 50 bird 
genomes and in the waders studied here, the two TLR7 clone sequences 
with the largest number of identical clones per wader specimen were used 
in the TLR7 analyses.

Phylogenetic analyses of sequences from TLR3 and TLR7 were per-
formed using maximum likelihood in RAxML B (Stamatakis, 2014). The 
GTR + G substitution model was identified as the best-fitting model of 
nucleotide substitution using the Bayesian information criterion. For 
each data set, the phylogeny was inferred using 10 random starts, and 
node support was evaluated using 1000 bootstrap replicates.

The phylogeny and node times were jointly estimated using Bayesian 
analysis in BEAST 1.8 (Drummond et al., 2012). For each data set, the anal-
ysis was performed using the GTR + G substitution model and with a 
Yule prior for the tree. To account for rate variation across branches, we 
used an uncorrelated lognormal relaxed clock (Drummond et al., 2006).
To calibrate the estimate of the evolutionary timescale, we used the fos-
sil-based age constraints described by Jarvis et al. (2014). For the TLR3 
and TLR7 trees, we constrained the ages of eight and six nodes, respecti-
vely (Supplementary material S1, Table S4). Calibrations were only applied 
to nodes with posterior probabilities >0.90, and were implemented as 
uniform priors on node times.

The posterior distributions of parameters, including the tree and 
node times, were estimated by MCMC sampling. Samples were drawn 
every 2000 steps from a total of 20,000,000 steps, with the first 10% of 
samples discarded as burn-in. From the posterior sample of trees, we 
identified the maximum-clade-credibility tree and rescaled the node 
heights to match the median posterior estimates.

2.4. Signatures of selection in TLR3 and TLR7 across birds

As with the phylogenetic reconstruction, evolutionary analyses were 
conducted on TLR3 and TLR7 sequences identified from the 50 bird ge-
nomes and the three additional wader species. These included two spec-
imens per species, one sequence of the TLR3 gene per wader specimen, 
and two sequences of the TLR7 gene per wader specimen.

Phylogenetic relationships among the sequences were estimated using neighbour-joining in the HyPhy package (Pond and Muse, 2005). 
These were followed by tests to detect different selection scenarios. For 
each TLR gene, we identified the best-fitting substitution model using 
the model-selection tool (Delport et al., 2010a) on the Datamonkey 
webserver (http://www.datamonkey.org, accessed March 2016; Delport 
et al., 2010a).

The Datamonkey webserver was used to test for evidence of positive, 
negative, and episodic selection using the single-likelihood ancestor 
counting (SLAC), the random–effects likelihood (REL; Pond and Frost, 
2005), fast unconstrained Bayesian approximation (FUBAR; Murrell 
et al., 2013), and the mixed-effects model of evolution (MEME; Murrell 
et al., 2012) methods. Additionally, the branch-site REL (BSR) method
was used to identify individual lineages under diversifying selection (Pond et al., 2011). The three different codon-based maximum-likelihood methods, SLAC, FUBAR, and REL, estimate \( dN/dS \) at every codon in the sequence alignment.

2.5. Analyses of selective forces on TLR3 and TLR7 in waders

Data generated by direct sequencing of 18 specimens of *A. interpres*, 22 of *C. alba* and 17 of *C. ruficollis* were used to elucidate the type and level of selection affecting the three species. Sequences were processed, assembled, and aligned using BioEdit, and all polymorphisms were examined by eye. TLR3 and TLR7 haplotypes were reconstructed with PHASE 2.1 (Stephens and Scheet, 2005; Stephens et al., 2001), implemented in DnaSP 5.1 (Librado Sanz et al., 2009), using the default settings: a thinning interval of 1100 burn-in iterations, and 100 main iterations. Genetic diversity indices of TLR3 and TLR7 within and between wader species were calculated using MEGA, as described above.

The haplotypes of the three species were tested for effects of positive selection, using different models implemented in CODEML in the package PAML 4.7 (Yang, 2007). The M1/M2 model tests for positive selection through \( dN/dS \) (Wong et al., 2004; Yang et al., 2005). For a conservative detection of sites under ongoing positive selection, the M8a/M8 model was used (Swanson et al., 2003; Wong et al., 2004). Bayes empirical Bayes (BEB; Yang et al., 2005) was used to calculate posterior probabilities of positively selected sites in cases where the likelihood-ratio test was significant. BEB uses maximum-likelihood estimates of parameters and accounts for sampling errors by applying a Bayesian prior. Pairs of models were compared using a likelihood-ratio test.

As a further test of selection at the population level, we used the internal fixed-effects likelihood (IFEL) method on the Datamonkey webserver. IFEL is a codon-based maximum-likelihood method that tests whether sequences sampled from a population have been subjected to selective pressure at the population level (i.e., along the branches internal to each species) (Pond et al., 2006). Sites that were detected to be under positive or negative selection were mapped onto the 3D protein structures that were inferred for TLR3 and TLR7.

3. Results and discussion

3.1. Characterization of TLR3 and TLR7 genes in waders

Comparison of the TLR3 and TLR7 sequences from the three waders with those from the closely related Killdeer (*Charadrius vociferus*), human (*Homo sapiens*), and house mouse (*Mus musculus*) revealed the amplification of slightly different regions of the two genes in the waders. The TLR3 primers amplified the central area of the ECD, including several leucine-rich repeat regions (LRR7–LRR18). In contrast, the TLR7 primers amplified the second part of the ECD, covering the last 8 LRRs in the ECD, as well as the LRR C-terminal cap (LRR—CT) transmembrane and Toll Interleukin-1 signalling (TIR) regions (Supplementary material S2, Fig. S1–S2).

TLR3 and TLR7 sequences from the three waders had regions with high similarities to those of orthologous mammalian sequences, indicating the presence of structural constraints. We also found polymorphic regions that potentially represent ligand-recognition sites across the two taxonomic groups (Echave et al., 2016) (Supplementary material S2, Fig. S1–S2). Based on TLR3 from mouse and TLR7 from human as templates, the inferred 3D structures of TLR3 and TLR7 in the three waders revealed typical TLR structures: alternating concave and convex descending and ascending lateral surfaces (3CG, Fig. 1a; 3W3K, Fig. 1b).

3.2. Diversity in TLR3 and TLR7 genes in waders

Previous studies have shown that TLR7 might be duplicated in Passeriformes (Cormican et al., 2009), leading to the presence of two or more TLR7 loci and more than two TLR7 AA haplotypes. The results from the present study showed that there were more than two AA haplotypes among the TLR7 clones, confirming that TLR7 is duplicated in *C. alba* and *C. ruficollis* and appears to be triplicated in *A. interpres* (although analysing a larger number of clones would be necessary to confirm the exact number of TLR7 copies in the three species) (Supplementary material S3, Table S5). Although we analysed nearly three times as many TLR3 clones (21–28 clones/specimen) as TLR7 clones (8 clones/specimen), the sequence alignments consistently showed the presence of two unique AA TLR3 haplotypes in each bird. This suggests that each of the three waders has only one copy of the TLR3 gene.

Consistent with the duplication of TLR7, the sequences of this receptor had greater numbers of segregating sites and higher nucleotide diversity than the TLR3 sequences in all of the three waders. The same results were obtained when analysing sequences obtained by directly sequencing multiple specimens from the three species (Supplementary material S3, Tables S6–S9).

To determine whether the observed AA changes within and across the three wader species have led to alterations in the TLR3 and TLR7 protein structures, we investigated the physicochemical properties and mean chemical distances of AA substitutions in the cloned sequences (Supplementary material S4, Tables S10 and S11, Figs. S3–S6). These analyses revealed that several of the observed AA substitutions indeed caused changes in the side chains’ charge, polarity, and the surface electrostatic potential, as well as in the hydrophathy indices. The across-

![Fig. 1. Comparison of 3D structures of the extracellular domains of the (a) TLR3 gene and the (b) TLR7 gene in three waders studied here. The models from mouse (*Mus musculus*) and human (*Homo sapiens*) were used as templates for TLR3 and TLR7, respectively. Grey lines show the template, while colours indicate the locations of the identified wader sequences within the structures. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)](image)
species comparison showed that TLR7 sequences had slightly fewer AA polymorphisms than the TLR3 sequences (21 vs 26) (Fig. 2 illustrates one angle of the physiochemical properties of TLR3 and TLR7 AA substitutions, further information can be found in Figs. S3–S6). These analyses also show high conservancy of TLRs (potentially maintained by negative selection) within the same family. Vinkler et al. (2014) have found higher variation in physiochemical differences, a result that could have originated from comparing TLR structures within class or between taxonomic groups (i.e. birds vs mammals), rather than within a single family.

3.3. Phylogenetic analysis of TLR3 and TLR7 genes across 50 avian taxa

The branching pattern of the TLR3 phylogenetic tree was inconsistent with the evolutionary history of the 50 birds inferred from their whole genomes (Jarvis et al., 2014) (Fig. 3). Furthermore, our molecular-clock analysis indicates that the TLR3 sequences of most waders diverged from those of other birds >60 million years ago (MYA). The age estimates for major nodes of the TLR3 phylogenetic tree were similar to those reported by Jarvis et al. (2014) and Cracraft et al. (2015), probably because

Fig. 2. Physiochemical properties of TLR AA substitutions across species presented by 3D models, (a) TLR3 and (b) TLR7 (for further details, see Supplementary 4, Figs. S3–S6). Blue and red models indicate variation of surface electrostatic potentials, red shows negative surface charge, blue shows positive surface charge and white shows neutral. Red and white models indicate hydrophobic surface variation, red shows hydrophobic, white shows hydrophilic. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
a very similar set of calibrations was used for both dating analyses. A notable exception among the waders is that the four migratory waders (C. alba, C. ruficollis, A. interpres, and Calidris pugnax) formed a monophyletic group, whereas the only non-migratory wader, Charadrius vociferus, grouped with three very distantly related species: Downy Woodpecker (Picoides pubescens), Rock Dove (Columbia livia), and Barn Owl (Tyto alba). This branching pattern suggests that TLR3 is subject to different selection pressures in different wader species.

Charadrius vociferus is a philopatric, occasionally short-distance intra-continental migratory species that only moves within the North American continent under certain conditions (del Hoyo et al., 1992–2013; Sanzenbacher and Haig, 2001). The other four wader species, including Calidris pugnax, are long-distance migrants that use similar inland and coastal stopover sites when flying from their high-latitude breeding areas in the north to their tropical or Southern Hemisphere wintering destinations (del Hoyo et al., 1992–2013; Lisovski et al., 2015; Minton et al., 2011). Thus, they should be exposed to a similar range of pathogens. This difference in migration pattern and habitat use between Charadrius vociferus and the other waders is not reflected in the TLR7 phylogenetic analysis, further indicating different pathogen selection pressures between the TLR3 and TLR7 genes (Barton, 2007), as discussed below.

As in the tree inferred from the TLR3 sequences, we found that the relationships among the TLR7 sequences did not match the phylogenomic relationships among the 50 birds (Fig. 4). In contrast with the TLR3 tree, however, all TLR7 sequences from waders grouped with those from the closely related Charadrius vociferus. Our Bayesian phylogenetic analyses clearly show the duplications of TLR7 in the three waders, but also in Charadrius vociferus and Calidris pugnax, and in two other bird orders, Passeriformes (Taeiopygia guttata, Zebra Finch; Corvus brachyrhynchos, American Crow; and Manucus vitellinus, Golden-collared Manakin) and Cuculiformes (Cuculus canorus, Common Cuckoo). Although TLR7 duplications have previously been described in several passerines (Cormican et al., 2009; Grueber et al., 2012), our study is the first to describe a third parologue of TLR7 in the Medium Ground-finch (Geospiza fortis), as well as TLR7 duplications in additional bird species in Charadriiformes and Cuculiformes. Although a genome sequence is available for an additional passerine, the White-crowned Sparrow (Zonotrichia leucophrys), the poor assembly and annotation of the genome meant that it was not possible to confirm whether TLR7 is duplicated in this species. Improved annotations across all bird genomes could lead to the discovery of further TLR7 duplications in other species (Salzberg and Yorke, 2005).

3.4. Evolution of avian TLR7 and TLR3

The phylogenetic tree of avian TLR7 revealed that duplicate copies of TLR7 clustered according to the phylogenetic relationships of the species. The TLR7 sequences from both Passeriformes and Charadriiformes formed individual clades. Importantly, the TLR7 paralogues were more similar within than between species, suggesting recent duplication events.

Multiple duplications across species are known to occur (Pohl et al., 2009), but independent duplications usually display a random pattern across the phylogenetic tree (Bainova et al., 2014). The pattern of duplications across the three avian orders appears to be random, but the duplication pattern within the orders is not. Although this does not make recent, independent duplications impossible, it does make them extremely unlikely. It is possible that Charadriiformes experienced similar selection pressures over the last 10 million years or so, causing this grouping of independent duplications. This does not, however, explain why both the Passeriformes and Cuculiformes show a similar pattern of duplications that occurred around the same time. It is unlikely that
all three orders experienced the same selection pressures, given the diversity of habitats used and the different lifestyles displayed by these birds. Thus, the most likely explanation is that these duplications occurred ancestrally within each order.

An ancestral duplication event, even within an order, should still show a divergent pattern between the copies of the gene. This would produce a phylogenetic pattern in which the paralogues cluster by gene copies rather than by species (Schirrmeister et al., 2012). The observed clustering by species indicates homogenization of the genes, which is a signature of concerted evolution. Concerted evolution maintains the similarity between the two genes, using a combination of crossing-over events and homogenization, with each gene being used as a template (Hurles, 2004; Nei and Hughes, 1992). This process is not unique to avian TLR7, but has also been found to drive the genetic variation of the duplicated TLR1 and TLR2 genes across avian taxa (Alcaide and Edwards, 2011).

Gene duplication has been suggested to lead to an increased amount of gene product, although it does not necessarily double the dosage (Hurles, 2004; Stark and Wahl, 1984). Therefore, extra copies of TLR7 might allow a faster and a more specific response to viral infections in migratory waders. This might provide a selective advantage in the pathogen-host arms race, not only for long-distance migrants but for all birds.

3.5. Signatures of selection in TLR3 and TLR7 across birds

Comparison of rates of synonymous substitution ($dS$) and non-synonymous substitution ($dN$) were similar to those previously calculated in avian TLR3 and TLR7 (Alcaide and Edwards, 2011; Grueber et al., 2014; Mikami et al., 2012), and confirmed the predominance of purifying selection in both genes in all 50 birds ($dN/dS < 1$ for all loci; Table 2). Despite this evidence of purifying selection at the whole-gene level, we detected signals of positive selection at a small number of sites (TLR3 3.2%; TLR7 2.5%; Table 2). Nearly twice as many sites were found to be under episodic diversifying selection in TLR3 (9.2%) compared with TLR7 (5.4%). In contrast, close to 50% more sites were under negative selection.
in TLR7 (43.0%) than in TLR3 (29.6%). Many of the sites under selection (54% of the TLR3 and 66% of TLR7) detected by different models were identified by two or more of the methods used (Supplementary material S5, Tables S12 and S13).

The branch-site model (BSR) found evidence of diversifying selection along a single branch of the TLR3 tree. These results further support that the evolution of the two genes is most likely driven by responses to different types and levels of pathogen exposure. A potential additional explanation for the difference in selection between the two genes, and specifically the larger number of codons under negative selection in the TLR7 ECD, lies in the involvement of this gene in self-recognition. TLR7 recognizes the smallest molecules of any of the TLRs (Tanji et al., 2013) and can recognize self RNA from sick and dying cells signalling apoptosis (Krieg, 2007). Therefore, excessive variation in the TLR7 ECD could potentially cause autoimmune diseases, and would thus be selected against (Richez et al., 2011).

Although the ECD regions are the primary sites of adaptive responses to pathogens, and hence are expected to show high variation, PAMP binding actually requires a rigid structural framework (Bryant et al., 2015). Therefore, the pervasive force of purifying selection on TLR ECDs maintains the conserved structure of these domains to facilitate efficient PAMP recognition and binding (Echave et al., 2016). Despite the presence of strong purifying selection, the mean ratio of nonsynonymous to synonymous substitution rates \( (dN/dS = 0.36) \) in these two genes is still higher than that found in most other genes (Ellegren, 2008; Zhang and Li, 2004).

We found that 9 of 15 sites (60%) under selection in TLR3 were identical in the bird genomes to those found in other animals, including other birds and mammals (Alcaide and Edwards, 2011; Areal et al., 2011; Escalera-Zamudio et al., 2015; Grueber et al., 2014; Wlasiuk and Nachman, 2010). Therefore, the pervasive force of purifying selection on TLR ECDs maintains the conserved structure of these domains to facilitate efficient PAMP recognition and binding (Echave et al., 2016). Despite the presence of strong purifying selection, the mean ratio of nonsynonymous to synonymous substitution rates \( (dN/dS = 0.36) \) in these two genes is still higher than that found in most other genes (Ellegren, 2008; Zhang and Li, 2004).

Across the three wader species, we found evidence of higher selection pressure in TLR7 than in TLR3 when analysing the population-level data (data originating from direct sequencing of several specimens/species; Supplementary material S6, Table S16). Moreover, across the three species, six sites were found to be under positive selection and three sites were found to be affected by negative selection in TLR3. Sequences from C. alba had the largest number of positively and negatively selected sites (0.88% of nucleotides analysed). There was no evidence of selection in C. ruficollis (likelihood-ratio test, \( p = 0.13 \)), so we could not conduct a Bayes empirical Bayes test. Only 0.25% of sites were under selection in the sequences from A. interpres. Within TLR3, half of the positively selected sites were in LRR14, whereas the negatively selected sites were located in LRR9 and LRR10. TLR7 showed 14 and 22 sites under positive and negative selection, respectively, across the three species. Similar to the results from our analysis of TLR3, the largest number of sites under selection was detected in C. alba (1.8%), followed by A. interpres (1%) and C. ruficollis (0.8%). It is important to note that the observed differences between TLR3 and TLR7 may have originated from different regions being analysed in the two TLRs.

The 3D structural modelling of TLR3 revealed a localized patch of purifying selection corresponding to LRR9 and LRR10, indicating that this region is important for maintaining protein structure (Figs. 1 and 5). Positively selected sites are somewhat scattered in LRR9, LRR12, LRR17, with three sites in LRR14. In TLR7, the negatively and positively selected sites were spread throughout the last part of the ECD. Calidris alba had the largest number of sites under selection in both TLR3 and TLR7 across the three species. These results suggest an association between the lack of AIV infection and the observed TLR variations in the species. Further analyses, such as direct correlation of individual TLR3 and TLR7 polymorphism to AIV prevalence, are needed to identify the specific nucleotide variations associated with AIV infection.

4. Conclusions

Studies of pathogen dynamics and the underlying evolutionary mechanisms are urgently needed to predict future disease risks for...
wildlife and humans alike. By describing the major forms of selection shaping the evolution of avian RNA-sensing Toll-like receptors, our study has laid the groundwork for future investigations of the evolutionary mechanisms underlying infection dynamics and epidemiology. Furthermore, our study is the first to examine innate immune gene polymorphism in long-distance migratory waders in the context of viral infections. AVI infections are prevalent in both A. interpres and C. tuficolis but have rarely been recorded in C. alba. Our results point towards a possible association between TLR3 polymorphism in C. alba and the lack of AVI in this species. This finding contributes to the identification of candidate loci for future avian eco-immunogenetics research, opening the way for further insights into the impact of innate gene polymorphism on the evolutionary ecology of infectious diseases.

Acknowledgements

We are grateful to the Victorian Wader Study Group for assistance in collecting the samples. We thank Bethany Hoye, Alice Risely, Mejiluan Zhao, and Adam Miller for valuable discussions throughout the project and two anonymous reviewers for their valuable comments.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2017.05.012.

References


