Charles University
Faculty of Science

Study programme: Zoology

Mgr. Hana Velová (Bainová)

Adaptive evolution of Toll-like receptors in birds

Adaptivní evoluce Toll-like receptorů u ptáků

Doctoral thesis

Supervisor: RNDr. Michal Vinkler, Ph.D.

Supervisor-consultant: Prof. doc. Mgr. Tomáš Albrecht, Ph.D.

Prague, 2020

Declaration

Prohlašuji, že jsem nepředložila tuto práci ani její podstatnou část k získání jiného nebo stejného akademického titulu a že jsem práci zpracovala samostatně a uvedla všechny použité informační zdroje a literaturu. Na všech publikacích, které jsou součástí této disertace, se podílelo více spoluautorů. Můj podíl je vždy uveden a vysvětlen přímo v poděkování k příslušným publikovaným článkům zahrnutým do této disertační práce.

Hana Velová, Prague, 30th January 2020

Contents

Acknowledgement

The research included in this thesis would never be possible without contribution of many people.

Firstly, my special thanks belong to Michal Vinkler, not only because he was a very supportive supervisor throughout my entire university study, he provided me with great mentoring and he always found time to give me useful advises, but also because it was he who definitely redirected my future scientific career. From the very beginning of our cooperation, I very appreciate his kind and humanistic approach to all students. Moreover, he always tried to find enough financial support not only for increasing of student scholarships, but also for international meetings, where he was very opened to establishing new international collaborations involving student forein research visits. For all of this he has my sincere thanks.

Furthermore, I would like to thank to all my colleagues and co-authors for their cooperation in all our projects dealing with TLRs. Especially, I am very grateful to Anna Bryjová, Marta Promerová and Dagmar Čížková for introducing me to the rules and secrets of molecular genetics lab work. My great thanks belong also to Tereza Králová, Zuzana Swiderská and Marian Novotný for very inspirative and helpful disscussions about methodology and the meanings of obtained results. I thank to Josef Bryja that I could work, gain knowledge and friends in excellent laboratories in the external research facility of the Institute of Vertebrate Biology ASCR in Studenec. I am grateful to Dave Burt who invited me to come to the Roslin Institute, Division of Genetics and Genomics at the University of Edinburgh to analyse the genomic data in avian genomes project and to Weronika Gutowska-Ding who taught me how to do it. Without them the crucial paper of my thesis could be hardly created. Vladimír Beneš gave me the opportunity to spend short but productive and very inspiring time in the Genomics Core Facility laboratories in EMBL in Heidelberg, for which I am very grateful. My great thanks belong also to Robert Settlage who introduce me to the secrets of bioinformatical analyses (not only) of transcriptomic data during my stay at Advanced Research Computing, Virginia Tech University, USA. I would like to thank to all members of our Evolutionary and Ecological Immunology Group for creating a friendly and inspirative environment. Great thank belongs to Jon Fjeldså who allowed me to (re)use his beautiful painting.

Surely, this thesis and my PhD studies could not be finished without the financial and material support. Here I would like to thank not only to the grant agencies (GA UK and GA CR) for money support of our research, but also to other foundations (namely to Hlávka foundation and Mobility Fund of the Charles University) for awarding me with travel grants and so enabling me to gain international experience. Without these my PhD project would shrink by half.

Special thanks belong to my parents who supported me not only during my studies but also during my entire life so far. And most importantly, they decided to let me find my own way to go and live, which I appreciate very much. Above all, however, I would like to thank to Honza, my life partner and best friend, who support me fully in everything I try to achieve already more than ten years.

Preface

Biology and the the laws of nature fascinated me from the very beginning of my life. Moreover, I was lucky to have a brilliant biology teacher at our grammar school Gymnasium Jírovcova in České Budějovice – Jarmila Ichová. She was very enthusiastic, and this enthusiasm passed to almost all of us in the class. Although I was later quite indecisive which university to choose (during my grammar school studies I was also involved in a small chemical research focused on heavy metal content of fish meat), at the very last moment, I decided for studying rather Biology at Faculty of Science, Charles University. I attended many interesting courses during my bachelor studies there, but I was mainly attracted by various ecological lectures. Therefore, I decided to study Ecology as my master program.

However, at the same time, I met Dr. Michal Vinkler, which meant a turning point in my scientific career. I changed the topic of my bachelor thesis and also the supervisor (to Michal). I was very curious, how the PHA-skin swelling test we wished to use in one field experiment was actually working (I had my doubts about it) and Michal was just designing an experiment to resolve it (and by lucky coincident, he had no students to help him to do it at that time). Which meant I switched from almost a pure field work to more experimental one, followed by moving from ecology to evolution and immunology and thus to the molecular genetics' laboratory for my master thesis. Only the birds as model species remained. Molecular genetics was for me a completely new field, I had to study a lot, but I had great a menthor especially in Anna Bryjova. Moreover, I had the opportunity to start in an excellent laboratory of the Institute of Vertebrate Biology ASCR in Studenec, where there was a very kind, friendly and inspiring environment (and I am sure, still is). To deepen my laboratory skills, I conducted a half-year stay in laboratories of the Natural History Museum at the University of Oslo, Norway under the kind supervision of prof. Arild Johnsen. Later I further improved and applied this knowledge during my doctoral studies at the Department of Zoology, Faculty of Science, Charles University. During this period, I was namely inspired by my visit to the Genomics Core Facility laboratories in EMBL, Heidelberg where I worked under the supervision of Dr. Vladimír Beneš.

With the time of our research, also the methods used were moving on. This opened new challenges for me and as the time passed, I was switching more and more from the wet laboratory to bioinformatics. Most of the analyses we decided to run were totally new not only to me, but also to my supervisor. I accepted this as my new mission in our team. Very valuable advices on how to correctly handle the protein structures provided me Dr. Marian Novotný, Department of Cell Biology, Faculty of Science, Charles University. Besides self-teaching, I attended also several courses, out of which the most useful were the NGS workshop in Hinxton, Cambridge, UK and the NGS workshop lead by Libor Mořkovský (a brilliant mentor) at our department. I applied and improved the gained knowledges on genomic data analysis during my research visits to prof. Dave Burt, the Roslin Institute, University of Edinburgh, UK and to Dr. Robert Settlage, Virginia Tech University, USA. Unfortunately, not all results and all what I learned during my doctoral studies could be involved in my PhD thesis, but I strongly believe that everything will be useful for my future carrer.

The PhD title was (and still is) not the necessary aim for me, but rather a journey. Some kind of a by-product of my work that I am glad to do, and which makes sense to me.

Abstract

Toll-like receptors (TLRs) are one of the key and presumably also evolutionary most original components of animal immune system. As Pattern recognition receptors they form the first line of innate immune defence against various pathogens. The proper receptor binding of pathogenic ligands is crucial for their correct recognition and for subsequent triggering of an appropriate immune response. Because there exists a direct interaction between the receptor surface and the pathogenic ligand, host-pathogen coevolution on molecular level can be predicted. Thus, through variability of their ligands, TLRs are exposed to extensive selective pressures that may be detected on both genetic and protein levels. Surprisingly, the variability we revealed in birds is even higher than previously expected based on the reports from other vertebrates, mainly mammals. In my doctoral thesis I summarise the results of my contribution to the avian TLR research. We were the first who experimentally verify the absence of functional TLR5 in several avian species and duplication of TLR7 in others. We finally resolved the origin of duplication in TLR1 and in TLR2 family. An important part of my research project focused on the prediction of potentially functionally important positions in TLRs. We have outlined an investigation strategy universally applicable to any coding genes. Moreover, we found that some of the positively selected positions importantly affect the surface charge distribution. In passerine birds, we also attempted to find ecological factors determining the adaptive evolution in TLRs. However, this attempt was unsuccessful. Besides that, our methodological research improved the knowledge of the molecular background of the PHA-skin swelling test used for assessing the inflammatory responsiveness related to the TLR function. Since further research is highly needed to test the real functional effect of the TLR genetic variation, at the end of this thesis I outline several possible future directions.

Key words:

adaptive evolution, birds, gene duplication, gene expression, host-pathogen coevolution, inflammation, protein structure, pseudogenisation, selection, surface electrostatic potencial, TLRs

Abstrakt

Toll-like receptory (TLR) patří mezi klíčové a evolučně staré součásti imunitního systému zvířat. Jakožto receptory vrozené imunity tvoří první obrannou linii proti nejrůznějším patogenům. Správné navázání ligandů receptorem je přitom zcela zásadní pro bezchybné rozpoznání patogenů a následné spuštění přiměřené imunitní reakce. Protože zde dochází k přímé interakci mezi povrchem daného receptoru a patogenními strukturami, můžeme předpokládat molekulární koevoluci mezi hostitelem a jeho patogeny. TLR jsou tudíž díky variabilitě svých ligandů vystaveny značnému selekčnímu tlaku, který může být detekován jak na genetické, tak i proteinové úrovni. I tak je ale míra variability, kterou jsme popsali u ptačích TLR, překvapivě vysoká. Dokonce vyšší, než se dalo očekávat z dříve publikovaných prací u ostatních obratlovců, především savců. Ve své doktorské práci shrnuji výsledky, kterými jsem přispěla k výzkumu TLR u ptáků. Jako první jsme například experimentálně ověřili absenci funkčního TLR5 u několika ptačích druhů, či duplikaci genu pro TLR7 u druhů jiných. Také jsme definitivně rozřešili původ genové duplikace v rodině TLR1 a TLR2. Důležitá část mého výzkumného projektu se týkala predikce potenciálně funkčně významných pozic u TLR. Navrhli jsme postup identifikace těchto pozic, který je univerzálně použitelný i pro ostatní kódující geny. Navíc jsme zjistili, že některé z těchto pozic významně ovlivňují distribuci povrchového náboje. U pěvců jsme se pokoušeli najít ekologické faktory určující adaptivní evoluci TLR, nicméně bez úspěchu. Kromě toho jsme prohloubili znalosti o molekulární podstatě kožní zánětlivé odpovědi vyvolané aplikací PHA, která může být použita k ověření funkčních rozdílů mezi různými variantami TLR. Jelikož je však navazující výzkum v tomto směru stále potřeba a to hlavně, abychom otestovali skutečný funkční význam genetické variability popsané u ptačích TLR, na konci této práce navrhuji několik možných směrů, kudy se ubírat dál.

Klíčová slova:

adaptivní evoluce, genová duplikace, genová exprese, koevoluce hostitele a patogena, povrchový elektrostatický potenciál, pseudogenizace, ptáci, selekce, struktura proteinů, TLRs, zánět

List of Papers

This thesis consists of the following papers:

I. **Velová H.**, Gutowska-Ding M. W., Burt D. W. & Vinkler M. (2018): Toll-like receptors in birds: gene duplication, pseudogenization, and diversifying selection. Molecular Biology and Evolution, 35(9): 2170–2184. (IF_{2018}= 14.797)

II. Vinkler M., **Bainová H.** & Bryja J. (2014): Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. Genetics Selection Evolution 46:72. (IF_{2014}= 3.821)

III. Králová T., Albrecht T., Bryja J., Hořák D., Johnsen A., Lifjeld J. T., Novotný M., Sedláček O., **Velová H.** & Vinkler M. (2018): Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an ecological context. Molecular Ecology, 27(13):2871–2883. (IF_{2018}= 5.855)

IV. **Bainová H.**, Králová T., Bryjová A., Albrecht T., Bryja J. & Vinkler M. (2014): First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds. Developmental & Comparative Immunology 45(1):151-155. (IF_{2014}= 2.815)

V. Vinkler M., **Bainová H.**, Bryjová A., Tomášek O., Albrecht T. & Bryja J. (2015): Characterization of Toll-like receptors 4, 5 and 7 and their polymorphism in grey partridge. Genetica 143(1):101-112. (IF_{2015}= 1.343)

VI. Vinkler M., Svobodová J., Gabrielová B., **Bainová H.** & Bryjová A. (2014): Cytokine expression in phytohaemagglutinin-induced skin inflammation in a galliform bird. Journal of Avian Biology 45(1):43–50. (IF_{2014}= 1.971)

Papers not included in the thesis:

VII. Vinkler M., **Bainová H.** & Albrecht T. (2010): Do we know what we measure?: Functional analysis of the skin-hypersensitivity response to phytohemagglutinin in Zebra Finch (*Taeniopygia guttata*). Functional Ecology 24(5):1081-1086. (IF_{2010}= 4.645)

VIII. Kropáčková L., Pechmanová H., Vinkler M., Svobodová J., **Velová H.**, Těšičký M., Martin J.F. & Kreisinger J. (2017): Variation between the oral and faecal microbiota in a free-living passerine bird, the great tit (*Parus major*). Plos One. (IF_{2017}= 2.766)

IX. Svobodová J., Bauerová P., Eliáš J., **Velová H.**, Vinkler M. & Albrecht T. (2018): Sperm variation in great tit males (*Parus major*) is linked to a haematological health related trait, but not ornamentation. Journal of Ornithology 159(3): 815-822. (IF_{2018}= 1.472)

X. Tešický M., **Velová H.**, Novotný M., Kreisinger J., Beneš V. and Vinkler M.: "Positive selection and convergent evolution shape molecular phenotypic traits of innate immunity receptors in tits (Paridae)." Submitted to Evolution. (IF_{2019}= 3.573)

XI. Bauerová P., Krajzingrová T., Těšický M., **Velová H.**, Hraníček J., Musil S., Svobodová J., Albrecht T. and Vinkler M.: "Lifetime bioaccumulation and health effects of heavy metals in longitudinally monitored urban bird blood." Submitted to Science of the Total Environment. (IF_{2019}= 5.589)

Introduction

Host-pathogen coevolution

My PhD thesis focuses on the host-pathogen coevolution as one of the most widespread evolutionary associations in the living world. Dealing with such a broad topic I decided to start with the hosts and their defence against possible threats from various pathogens, where birds may serve as model hosts and Toll-like receptors (TLRs) as model components of the host's immune defence system. By definition, pathogens reduce host fitness through their negative effect on host survival and reproduction. In general, infectious diseases in wild living animals present huge risk for both the natural populations of other species (Robinson et al. 2010; Staley and Bonneaud 2015), as well as for domestic animals and humans (Tsiodras et al. 2008; Reperant et al. 2009; Cunningham et al. 2017). This is true especially for wild living birds, which may be colonised by many diseases' agents (Thomas et al. 2007) and are able to move for long distances to rapidly spread any possible threats to geographically distant regions (Dhondt et al. 2005; Lawson et al. 2011). Very well-known and described is the case of highly pathogenic avian influenza viruses (not only the subtype H5N1) which are able to switch the host species and infect not only birds, but also various mammals including humans (Reperant et al. 2009). Another widespread disease affecting numerous avian species is caused by a protozoan *Trichomonas gallinae* (Amin et al. 2014). Well documented is the recent case of the trichomonosis spread in the Great Britain, where the population of finches decreased in few years by up to 35%, representing mortality of half a million birds (Robinson et al. 2010). Another avian disease example is from USA, where mortality of more than 20% was revealed in passerine birds caused by *Salmonella* infection with an increasing trend in the last decades (Hall and Saito 2008). At the same time, up to 60% decrease in the size of US populations of free-living house finches have been attributed to another bacterial pathogen, *Mycoplasma gallisepticum* (Hochachka and Dhondt 2000; Dhondt et al. 2005). This epidemic spread very rapidly across North America followed by dramatic evolution towards increase in the pathogen's virulence (Hawley et al. 2013).

Infectious diseases activate host immune system and subsequently affect other traits including life histories, condition or reproduction of infected individuals and thus also the population dynamics (Clayton and Moore 1997). The functionality of immune system thus directly influences the total individual fitness and thus there must be strong selective pressure to adaptive changes in immunity leading towards increasing host's resistance against particular pathogens causing the diseases. At the same time, the selection is acting also on pathogens increasing their ability to successfully overcome the host's immune defence. The host-pathogen coevolution is nicely expressed by the Red queen's metaphor (van Valen 1973), which proposes that species must „run" (means constantly evolve) as fast as they can to stay at least at the same place and evade extinction. This famous phenomenon is originally based on Carroll's idea depicted in the fantasy book for children „Through the Looking-Glass and What Alice Found There" (Carroll 1993), where Alice realizes that in nature „it takes all the running you can do, to keep in the same place" (Fig. 1). According to this view, in host-pathogen coevolution a never-ending evolutionary arms race is going on between hosts and pathogens (Woolhouse et al. 2002). This dynamic process is based on the adaptation in one species that is followed by a selection acting in the second species giving rise to the adaptive coevolution. Immune system of a host must constantly evolve the capacity of defence against pathogens, because pathogens are continuously selected on invention of new weapons and strategies to overcome the host's immunity. The visible consequence of this interaction is besides others the establishment and

maintenance of genetic variability in both, the hosts and also the pathogens. At microevolutionary level this means that nucleotide substitutions causing functional changes enabling faster and easier host recovery from the infection become advantageous and undergo rapid changes in allele frequencies within a host's population. Thus, the immune genotype plays a crucial role in anti-pathogen defence.

Fig. 1: Host-pathogen coevolution. Alice is trying to run faster than pathogens attacking her. In scientific words, coevolution is leading to adaptations in all involved species. Illustration by Kristina Hedrick, Lightray Productions, Los Angeles, California reused from Hedrick (2004).

Selection acting on hosts

Diversifying selection

Natural selection is the most important process driving adaptive evolution. Diversifying (disruptive) selection describes a scenario of the positive selection when the selection acts against mainstream allele and towards the extremes. This increases genetic variability in a population, which is necessarily a starting point for making new innovations and thus novel evolutionary adaptations (Yang et al. 2000). In most of the cases, the adaptive evolution acts in a particular codon rather than in the whole gene locus. To measure the selective forces acting on molecular level at a particular site in the genome we can calculate the parameter omega (ω), i. e. the ratio of non-synonymous to synonymous substitution rate (dN/dS) among investigated sequences, where the value of $\omega > 1$ indicates positive diversifying selection, $\omega = 1$ neutral evolution and $\omega < 1$ negative (purifying) selection (Miyata and Yasunaga 1980). A powerful and generally popular approach for detecting signatures of natural selection from molecular data is a comparative sequence analysis using codon-substitution models (Yang et al. 2000; Yang 2007; Murrell et al. 2012; Murrell et al. 2013). Particular examples of commonly used models I have mentioned in the subchapter Selection analysis in Material and Methods section (p. 26) and in the papers included in this thesis.

Balancing selection

The high level of genetic polymorphism in immune genes is maintained mainly by balancing selection (Trowsdale and Parham 2004). Among tree major mechanisms maintaining the long-lasting variability in a population belong i) negative frequency-dependent selection, ii) the advantage of heterozygotes and iii) selection varying in time and space. ***The negative frequency-dependent selection*** occurs when possessing a less frequent genotype becomes advantageous, i.e. in cases where fitness of a particular genotype at a locus declines with its increasing frequency (Dobson and Hudson 1992; Woolhouse et al. 2002; Papkou et al. 2016). So, virtually, hosts with beneficial immune genotype which successfully recognize the current common pathogen are in advantage and thereby they increase their frequency in a population. Certainly, the genotype frequency is growing only until the pathogen finds the way, how to overcome such host immune defence, or until another pathogen appears as the common threat. At that moment such genotype starts to decrease in frequency, being replaced with another (until then less common) genotype that starts to be advantageous. And since the negative frequency-dependent selection is acting, this is running all over again in consequent cycles (Fig. 2). In the case of ***heterozygote advantage***, where selection favours individuals heterozygous at a particular locus to both homozygotes, resulting in a significantly higher relative fitness of heterozygotes in the population (Fisher 1922; Hedrick 2012). This might be for example the case, where the higher variability given by two different alleles of a particular immune gene (receptor) enable improved detection of possible threats (e.g. recognition of greater spectrum of pathogenic variants) and thus more effectively resist infectious diseases, then homozygotes (Penn et al. 2002; Savage and Zamudio 2011). ***Selection varying in space and time*** is other form of balancing selection in which different genotypes have an advantage in different environments. For example, the substitution D299G in human TLR4 decreases mortality due to malaria (Mockenhaupt et al. 2006), but at the same time it is linked with higher risk of septic shock (Lorenz E et al. 2002). Therefore, distinct genotypes are beneficial in human populations living in malarian and non-malarian regions (Ferwerda et al. 2007).

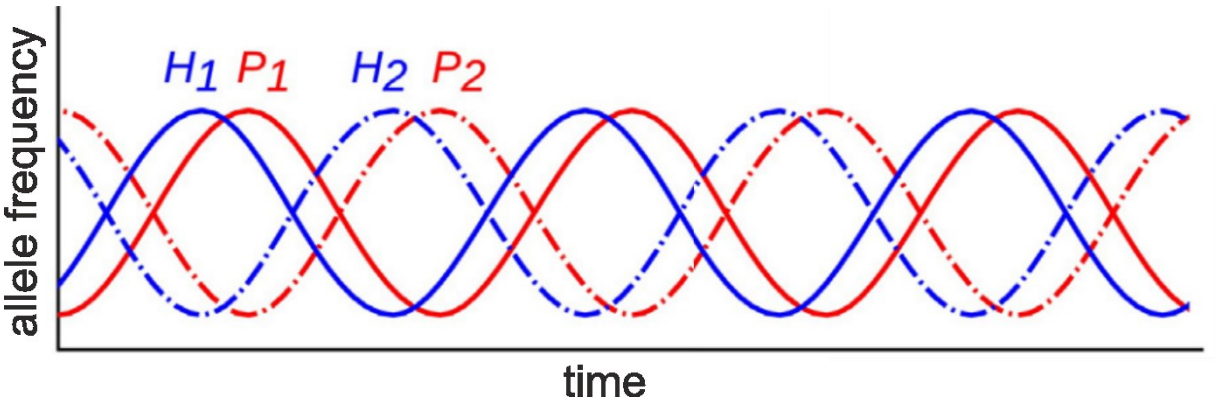


Fig. 2: Negative frequency-dependent selection. Theoretical visualisation of the host-pathogen coevolution is shown. Specifically, the situation, where two different hosts genotypes (alleles: H1 and H2) come into the contact with two different pathogens genotypes (P1 and P2), is visualised. The host genotype H1 successfully recognize pathogen allele P2 but not P1, in contrast to genotype H2 which is able to recognize P1 but not P2. The effective immune response is triggered only after successful pathogen recognition. Therefore, hosts with currently beneficial genotype (allele) become more common in the population. But at the same time pathogens, which are capable to overcome the immune response of the more frequent host genotype, are in selective advantage and thus are these pathogens also rising in numbers, which results in decrease of the frequency in the previously more abundant host genotype in the population. This mechanism drives the selection into permanent cyclic oscillations. The figure is adopted from Papkou et al. (2016).

Birds as model taxon in evolutionary immunology

Birds are very interesting group of vertebrates showing many unique adaptations. Being very similar to mammals in many physiological aspects (including many important traits like homeothermy and endothermy), they also share with mammals many ecological life strategies. Modern birds are very diverse - this vertebrate lineage contains approximately 10,000 extant species (Jetz et al. 2012). Most of the recent avian lineages originated from a rapid radiation which started on the Cretaceous–Tertiary boundary 65 million years ago (Mya) and graduated especially during the last 50 Mya (Jetz et al. 2012). This closely paralleled mammalian explosive phylogenetic evolution in the early Tertiary (Feduccia 1995). Regardless, birds are still surprisingly much less studied than mammals, which is true also in the field of immunogenetics. Broader investigation in birds was recently enabled by the whole avian genome sequencing projects including the Bird 10,000 genomes project (http://b10k.genomics.cn/; Eöry et al. 2015). This public sequencing data allow resolving many questions concerning avian adaptations and their evolution (Zhang, Li, et al. 2014) including investigation of evolution in particular immune genes (PAPER I.).

Given their ecology, abundance and diversity, birds represent excellent taxon for evolutionary ecology studies of host-pathogen interactions (Clayton and Moore 1997). For example, the avian models play a central role in studies of sexual selection, which has been proposed to be one of the basic and chief process involved in evolution of anti-parasite resistance (Hamilton and Zuk 1982). Even though birds are confronted with similar types of pathogens as mammals, they often solve their threat differently and during evolutionary history they evolved diverse adaptations how to defend against pathogenic challenges. Moreover, often moving on long distances, birds are important pathogen vectors and considering TLRs as the key molecules essential for resistance to the most threatening pathogens (including malaria or avian influenza; Ferwerda et al. 2007; Bodewes and Kuiken 2018), it is important to investigate the role of TLRs polymorphism in free-living avian populations. This may be interesting not only from the perspective of evolutionary ecology but in the same time bring highly valuable information to veterinary and hygiene praxis.

Immune system

Immune genes belong among the genes under strongest positive selection acting in animal genomes (Bustamante et al. 2005; Kosiol et al. 2008; Shultz and Sackton 2019). Among them especially genes coding immunity receptors are most often the target of positive selection because they are interacting directly with various pathogen associated molecular patterns (PAMPs; Akira et al. 2006). Even within a single group of pathogens the variability is enormous (Zdorovenko et al. 2007; Dentovskaya et al. 2008; Wang et al. 2015), so how is it possible that hosts are able to successfully protect themselves against most of them? Although pathogens evolve their traits improving host intrusion more rapidly (thanks to their shorter lifetime), hosts have several important advantages including increased variability in immune genes. Resulting high variation within a population ensures that at least some individuals from the population possess the immune genotype suitable for appropriate detection of attacking pathogen which ensures resistance. In other words, rapid adaptive changes in pathogens determine the evolution of variation in host immunity that prevents negative effect of the pathogen eventually eliminated by the function of the immune system. Host immunity comprises the sophisticated network of various defensive mechanisms operating on different levels, in vertebrates represented by innate and adaptive immunity (see Fig. 3 and e.g.

Danilova 2006). The functionally heterogeneous mechanisms of innate immunity are evolutionary original and typically they form the first and fast host defence, often including inflammation, whereas the immunoglobin-based adaptive (acquired) immunity is present since jaw vertebrates, being highly specific and involving immunological memory.

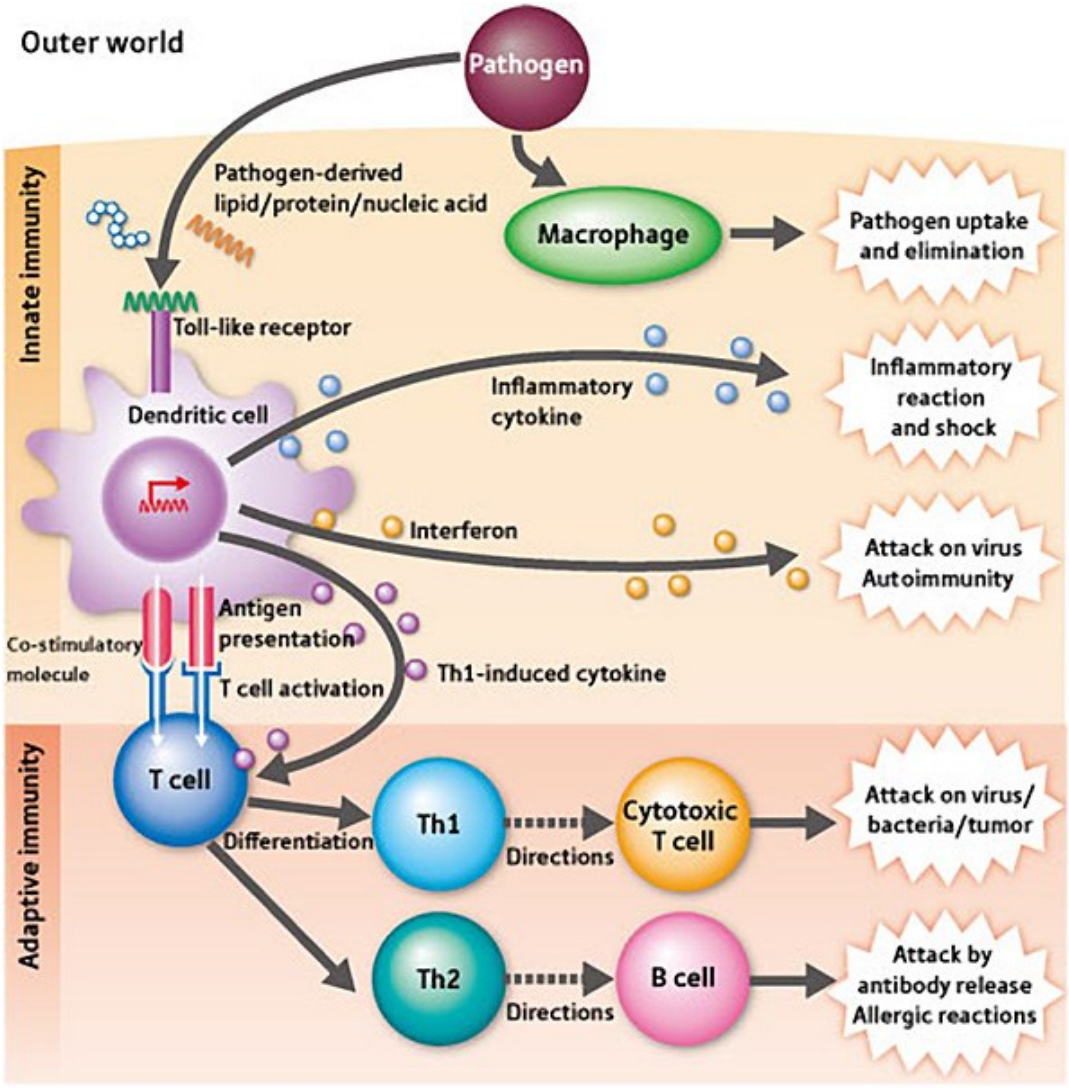


Fig. 3: Innate and adaptive immunity. Schematic figure depicts the detection of a pathogen through PAMPs by innate immune system mediated by TLRs inducing inflammation as well as activating adaptive immunity through improved antigen presentation (via MHC molecule) to T cell. The figure is adopted from Kaisho (2008).

MHC

Many different genes are involved in host defence against pathogens. However, despite the complexity of the immune system, in eco-evolutionary studies the undoubtedly most widely explored immune-related gene cluster is the Major histocompatibility complex (MHC; Kaufman 1999; Milinski 2006), the part of adaptive immune system. The role of MHC is to allow distinguishing self from non-self antigens by presenting short peptides at the cell surface to various T-cell receptors (Neefjes et al. 2011). To be able to bind high diversity of peptides, MHC evolved extreme polymorphism (especially in the peptide binding region), where up to hundreds of alleles are described even in a single population (Klein et al. 2007; Promerová et al. 2009; Sepil et al. 2012). Although highly polymorphic,

MHC is not the only one molecule responsible for potential pathogen threat detection. Furthermore, based on mapping and association studies it is known that at least half of the genetic variability responsible for resistance to infections can be attributed to other immune genes than MHC (Acevedo-Whitehouse and Cunningham 2006), suggesting that MHC-independent immune responses are also important for the host-pathogen coevolution.

PRRs

As good candidate genes for the eco-evolutionary research among non-MHC genes seem to be those coding receptors of innate immunity - Pattern recognition receptors (PRRs; Gordon 2002), because these receptors form the direct molecular interface between the host and the pathogen structures (various PAMPs; Akira et al. 2006; Kang and Lee 2011) and they are thus supposed to be exposed to the highest selective pressures from the constantly evolving pathogens. Among innate immunity receptors, TLRs belong to the crucial and most-well described PRRs (Kawai and Akira 2010). When I started my PhD studies, TLRs were already very well-known molecules with general function understood in humans and several mammal species (Smirnova et al. 2000; Ferwerda et al. 2007; Shen et al. 2010; Areal et al. 2011; Shen et al. 2012; Smith et al. 2012; Quach et al. 2013). The situation in birds was known mainly from the studies in chicken (Leveque et al. 2003; Boyd et al. 2007; Temperley et al. 2008; Kannaki et al. 2010; Brownlie and Allan 2011), later followed by other avian species (de la Lastra and de la Fuente 2007; MacDonald et al. 2008; Cormican et al. 2009; Vinkler et al. 2009; Gopinath et al. 2011). Even though TLRs have been intensively studied from the immunological and biomedicine perspective in last two decades, the evolutionary point of view dealing with the genetic variability within birds retained untouched until recently (Vinkler and Albrecht 2009). This was the point where we decided to uncover this missing piece of the puzzle and explore the evolution in TLRs in wild-living birds, which was since that time largely overlooked.

Toll-like receptors

As a part of innate immune system, TLRs belong among the first elements responsible for the detection of potential danger (Kawai and Akira 2010) and thus their importance for successful infection defeat is crucial. After specific ligand binding and thus recognising the pathogen through PAMPs, TLRs trigger the first (mainly inflammatory) immune response as the reaction to this present threat (Fig. 3; Akira et al. 2006; Takeuchi and Akira 2010). Subsequently this signalling also initiates adaptive immune response (Iwasaki and Medzhitov 2015) and thereby TLRs build the virtual bridge between innate and adaptive immunity. TLRs represent the basic and one of the evolutionary most original components of animal immune system responsible for pathogen-recognition. They are present not only in all vertebrates, but their homologues named Toll receptors or TLR-like molecules are known also in invertebrates (Coscia et al. 2011). These receptors are structurally highly conserved (PAPER II.). All TLRs are transmembrane proteins composed from a horse-shoe-shaped ectodomain (ECD) built-up by multiple leucine-rich repeat (LRR) motifs that form the interface for the direct contact between the receptor and PAMPs, transmembrane domain and intracellular Toll–interleukin 1 receptor (TIR) domain which enables the downstream signal transmission leading towards proinflammatory changes (Fig. 4; Jin and Lee 2008; Kawai and Akira 2010).

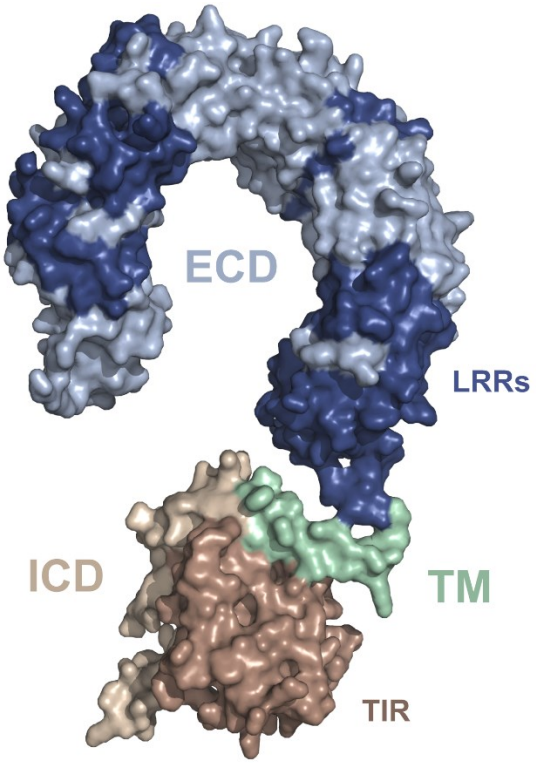


Fig. 4: Toll-like receptor.
Structural visualisation of TLR4 protein, where extracellular domain (ECD) is coloured in blue, transmembrane region (TM) in green, intracellular domain (ICD) in brown. TLR important structure motives are also highlighted - leucine-rich repeats (LRRs) in dark blue as a part of ECD and Toll-interleukin 1 receptor (TIR) domain in darker brown as a part of ICD. TLR structure was predicted using I-TASSER server (Roy et al. 2010), receptor domains by SMART tool (Letunic and Bork 2018) and subsequently graphically modified in PyMOL Molecular Graphics System (Schrödinger, LLC).

Inflammation

Inflammation belongs to the most essential mechanisms of host immunity serving to protection of host health through rapid and intensive pathogen clearance (Medzhitov 2008). Inflammation is therefore a hallmark of various infectious diseases. It is a tissue-destroying process, in which the first step involves recognition of the infection through detection of pathogen- or damage-associated molecular patterns (PAMPs or DAMPs; Ashley et al. 2012). The PAMPs comprise of general motifs of pathogen molecules that are essential for their survival. These are recognized by PRRs of innate immunity, including TLRs (Takeuchi and Akira 2010). After pathogen-derived ligand recognition, TLRs trigger intracellular signalling pathway leading to expression of inflammatory cytokines (e.g. IL-1β; Fig. 5), which results in inflammatory response and subsequent clearance of the infection (Veerdonk et al. 2011).

However, if the inflammatory response is not regulated properly, especially in the case of excessive response, this process can have devastating effects through collateral damage and lead to serious pathology (e.g. rheumatoid arthritis, multiple sclerosis, diabetes; reviewed in Ashley et al. 2012). In some cases, the damage from self-harm is even greater than the damage caused by the pathogen itself, for instance most of the brain neuronal damage from bacterial meningitis or cerebral malaria is probably caused by over-reactive inflammatory responses (Ashley et al. 2012). Thus, immunological regulation optimises (not maximises) the level of inflammatory response. Cytokines belongs among crucial regulatory molecules in this process (Thomson and Lotze 2003). Based on their function, they can be divided into two groups: pro-inflammatory and anti-inflammatory.

The paramount role of TLRs in inflammation is supported by many studies describing the relationship between genetic variability in TLRs and various inflammatory diseases in humans (reviewed e.g. by Misch and Hawn 2008). Contrary to mammals, in birds these mechanisms are far worse understood. So far there is only a single piece of evidence in birds linking disease variation with *TLR* polymorphism (Leveque et al. 2003). More frequently, the infection is experimentally mimicked by sterile

inflammation (as typical for eco-evolutionary studies dealing with individual fitness, capability to survival, etc.). Among one of the most common agents used to trigger tissue inflammation belongs phytohaemagglutinin (PHA; Vinkler et al., 2010). Application of PHA-treatment into subcutaneous tissue leads to activation of Th17-mediated immunological pathway due to proinflammatory cytokines (IL-1β and IL-6; PAPER VI.), and thus enable us to measure the inflammatory response of individual.

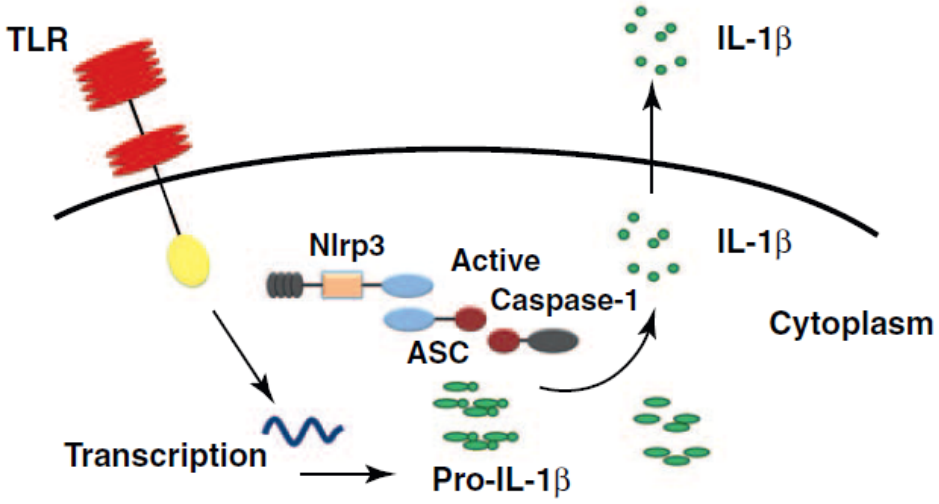


Fig. 5: TLR signalling pathway triggers proinflammatory cytokine (IL-1β) expression which induces inflammatory response involving phagocytosis. The figure is adopted from Veerdonk et al. (2011).

Avian TLRs

Each vertebrate species is equipped with about a dozen of TLRs (Mikami et al. 2012). In its phylogenetically original state the avian TLR family consists of ten receptors which are situated on cell surface (TLR10[1A], TLR1[1B], TLR2A, TLR2B, TLR4, TLR5 and TLR15) or in various cellular compartments as are endosomes, lysosomes and endolysosomes (TLR3, TLR7 and TLR21; Fig. 6; Brownlie and Allan 2011). Avian *TLR* genes are located on both macro and micro chromosomes (PAPER I.). Most of avian TLRs have their homologues with conserved function also in mammals (Roach et al. 2005; Coscia et al. 2011; Wang, Zhang, Liu, et al. 2016), except for TLR21 presented besides birds also in fish (Keestra et al. 2010) and sauropsid's unique TLR15 (Boyd et al. 2012). As in other immune genes, the process of gene duplication plays an important role in diversification of the TLR family. Most obviously, the duplication occurred in vertebrate TLR1 subfamily given rise TLR10[1A], TLR1[1B] and TLR2A, TLR2B in birds (PAPER I.; Coscia et al. 2011; Huang et al. 2011; Wang, Zhang, Liu, et al. 2016). Interesting is the unique and very recent gene duplication of TLR7 documented in several avian clades (PAPER I.; Cormican et al. 2009; Grueber et al. 2012; Raven et al. 2017). Not only gene duplication, but also pseudogenisation happened in avian TLR family during the evolution, since some of the avian species loss the functional of *TLR5* gene and possess the pseudogene instead (PAPER I.; PAPER IV.).

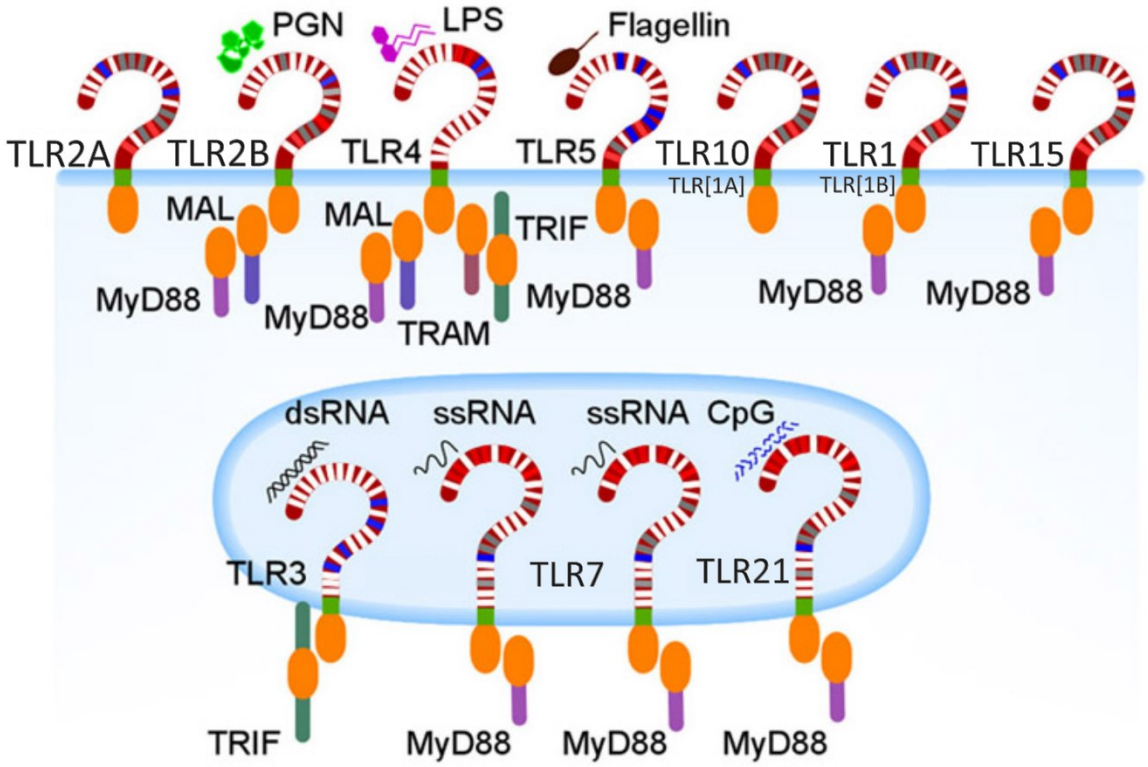


Fig. 6: Avian TLR repertoire. Schematic illustration of all members of avian TLR family with some of their ligands and signalling molecules is given. TLR10[1A], TLR1[1B], TLR2A, TLR2B, TLR4, TLR5 and TLR15 are situated on cell surface, whereas TLR3, TLR7 and TLR21 in various cellular compartments (e.g. endosomes). The possible duplication of TLR7 is also depicted. Original figure from Sarkar Lab (2018) is modified according to avian TLR repertoire (Brownlie and Allan 2011).

TLR ligand recognition

The nature and beauty of TLR recognition is based on the fact, that their ligands play essential role for the survival of pathogens, and, therefore, pathogens cannot avoid expressing these structures. Intriguingly, possibly in response to the selective pressures from the host immunity, it has been many times shown that PAMPs significantly vary in their structures not only across pathogen species, but even among different strains of the same species (see e.g. Zdorovenko et al. 2007; Dentovskaya et al. 2008; Wang et al. 2015). Through balancing selection, the spectrum of various PAMPs may maintain the variability in each host TLR. This system of TLR-ligand recognition enables us to relatively easily observe the ongoing host-pathogen coevolution on molecular level at sites where the receptor surface comes into a direct contact with various ligands (Fig. 7; Manavalan et al. 2011). Thus, TLRs represent an excellent model for studying host-pathogen coevolution.

The signalling through TLRs is always initiated by (homo- or hetero-) dimerization of the receptors that is followed by the binding of a TLR ligand to this receptor complex (for example see Fig. 7). Another, very important mechanism regulating the function of receptors involved in nucleic acid recognition (i.e. TLR3, TLR7 and TLR15 in birds) is the proteolytic cleavage of receptor ECD's part (Ewald et al. 2011; de Zoete et al. 2011; Garcia-Cattaneo et al. 2012; Kanno et al. 2013; Toscano et al. 2013; Murakami et al. 2014). Each avian TLR is responsible for specific ligand recognition: for example di/triacylated lipopeptides are recognized by heterodimers of TLR10[1A], TLR1[1B], TLR2A, TLR2B (Higuchi et al. 2008), lipopolysaccharides (LPS) by TLR4 (Keestra and van Putten 2008), flagellin by TLR5 (Keestra et al. 2008), and various nucleic acids such as dsRNA by TLR3 (Zou et al. 2017),

ssRNA by TLR7 (MacDonald et al. 2008), or CpG DNA by TLR21 (Keestra et al. 2010). TLR15 appears to be unique given its capacity to detect proteases (de Zoete et al. 2011). Thus TLR-based recognition allows detection of ligands derived from a wide range of microbes such as bacteria, viruses, parasites and fungi.

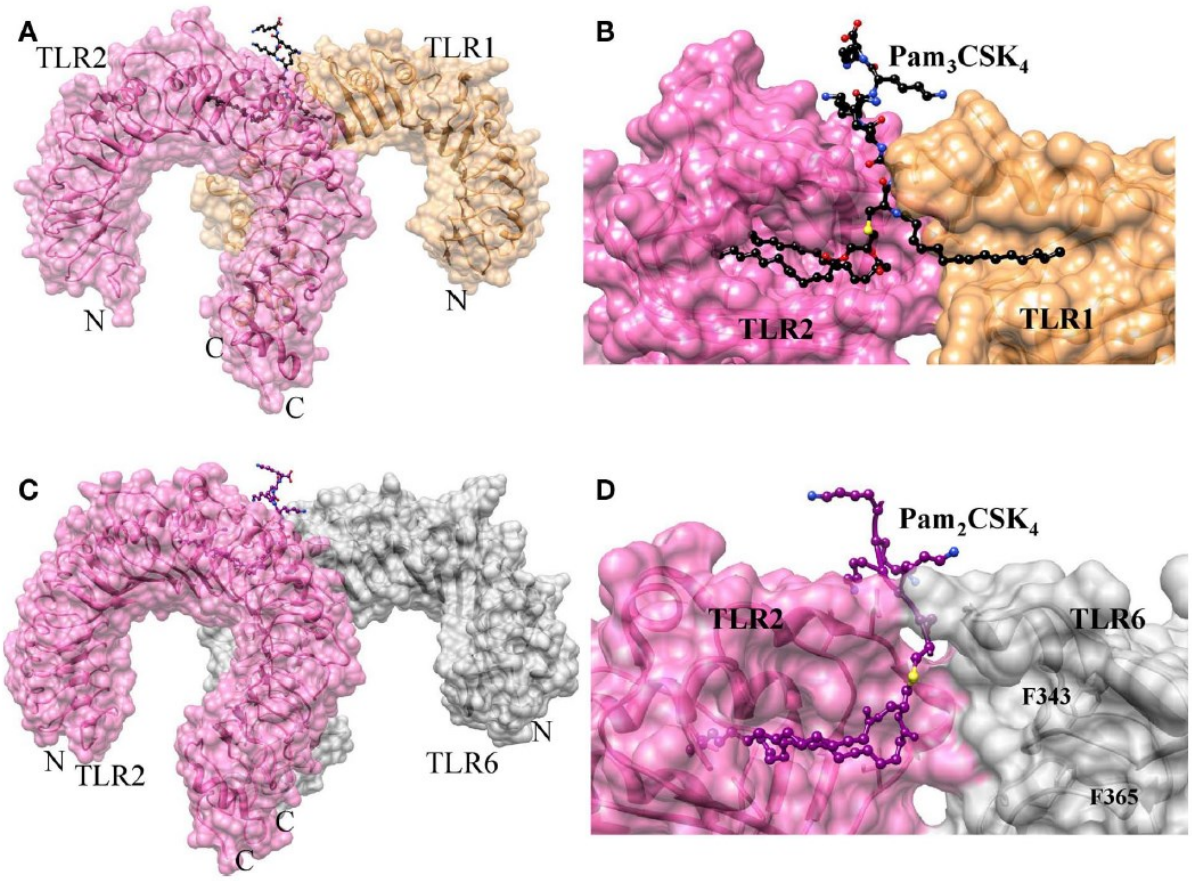


Fig. 7: TLR-ligand complex. An example of direct contact between TLRs and their ligands is given by the case of TLR2 and TLR1 or TLR6 heterodimer (pink, orange and grey) binding various lipopeptides (ball-and-stick models) into their lipid-binding pocket (see B and D for detailed view). The figure is adopted from Manavalan et al. (2011).

Vertebrates TLRs can be divided into three different groups based on the continuous hydrogen-bond network (asparagine ladder) formed among the asparagine residues on the concave surfaces in the neighbourhood of LRRs stabilizing the shape of ECD (Wang, Zhang, Liu, et al. 2016). In birds only two of these groups occur: the single-domain TLRs possessing complete asparagine ladder (avian TLR3, TLR5, TLR7, TLR15 and TLR21) and the three-domain TLRs where the central part of the horseshoe-shaped ECD is lacking the asparagine ladder (avian TLR10[1A], TLR1[1B], TLR2A, TLR2B, TLR4; Wang, Zhang, Liu, et al. 2016). The ECD architecture correspond with TLR ligand binding: the single-domain receptors recognize hydrophilic ligands, such as nucleic acids and proteins and the three-domain ones then hydrophobic ligands, such as lipids, lipoproteins and LPS (Wang, Zhang, Liu, et al. 2016). The single-domain TLRs appear to be under stronger purifying selection than the three-domain ones in mammals, but this has not been shown in birds due to a small TLR dataset available in the time of the analysis (Wang, Zhang, Liu, et al. 2016). Although TLRs are structurally conserved proteins (PAPER II.), they exhibit very high inter- and also intraspecific genetic variation in birds (Alcaide and Edwards 2011; Huang et al. 2011; Grueber et al. 2014). But not all of this genetic variability is evolutionary adaptive enabling specific PAMPs recognition and thus maintained by balancing selection forced by

these pathogens (Ferrer-Admetlla et al. 2008). Selection analysis may help us to distinguish between adaptive (pathogen-driven) variability and non-functional neutral mutations.

TLR genotypes and diseases

The associations between *TLR* genotypes and susceptibility to various infections, allergies, inflammatory and autoimmune diseases or cancers has been described in humans on many examples (summarised in Misch and Hawn 2008; Netea et al. 2012; Medvedev 2013). Contrary to humans (and some other mammals), in birds the evidence currently available is scarce and moreover all our knowledge comes from studies in domestic chickens. Leveque and colleagues (2003) described the association between *TLR4* polymorphism and susceptibility to *Salmonella* infection in Leghorn chicken lines. However, much deeper investigation of this relationship is needed, especially in the light of the fact that salmonellosis is a significant contributor to mortality in many avian species (Hall and Saito 2008). In passerines more than 20% of all mortality events involve salmonellosis and transmission to other species might have potentially negative impact on food production, biosafety and human health (Brankston et al. 2007; Mellata 2013; Kalmar et al. 2014). One of the possibilities, how to contribute to better protection against such infections in birds and also to improve our understanding of avian host-pathogen coevolution, would be to focus on investigation of adaptive variability in immune genes in free-living species, where the diversity is enormous and the natural selection operates unbiased.

TLR evolution

TLRs, as molecules of ancient innate immunity, are relatively structurally conserved across species (PAPER II.), their homologues are presented even in invertebrates (Coscia et al. 2011), but still exhibit sufficient species-specific and intraspecific genetic variation. As mentioned previously, TLRs are useful model molecules for studying host-pathogen coevolution, because they enable the direct contact between host receptor surface and pathogen-derived ligand molecules (Gay and Gangloff 2007). So, it has been repeatedly shown, that even a single amino acid substitution may change the receptor capability to appropriately recognize the ligand (Keestra et al. 2008; Walsh et al. 2008; Resman et al. 2009; Meng et al. 2010; Meng et al. 2011). For example, even a single point mutation in horse *TLR4* resulted in an amino acid substitution that affected the surface shape and electrostatic potential, which dramatically altered the lipid IVa recognition (Fig. 8.; Walsh et al. 2008). A useful way, how to predict which amino acid(s) could play such crucial role in receptor function, may be adopting the selection analysis involving non-model and rarely studied taxa, such as birds. The present evidence to this approach consists of several published papers reporting the detection of positively selected sites on interspecific level in various avian taxa (e.g. PAPER I.; PAPER II.; PAPER III.; Alcaide and Edwards 2011; Huang et al. 2011; Grueber et al. 2014; Wang, Zhang, Liu, et al. 2016). However, only some of these contributions solve the possible functional effects of the detected positively selected positions in the investigated species. Is the amino acid substitution present at the detected positively selected site nonconservative, i.e. is it likely changing the physico-chemical properties of the molecule? Where is the particular amino acid substitution located, exposed on the receptor surface, or buried inside the molecule? How far is the detected positively selected site from the previously described functionally important region of the receptor? After answering such questions, we should be able to predict suitable candidate amino acid positions for further research (for exact methodology see PAPER I. or PAPER II.). For example, such candidate positions can be subjected to targeted mutagenesis and subsequent *in vitro* testing of their functional significance,

which is important to validate their real function in optimising the responsiveness of avian immune system (Voogdt et al. 2018).

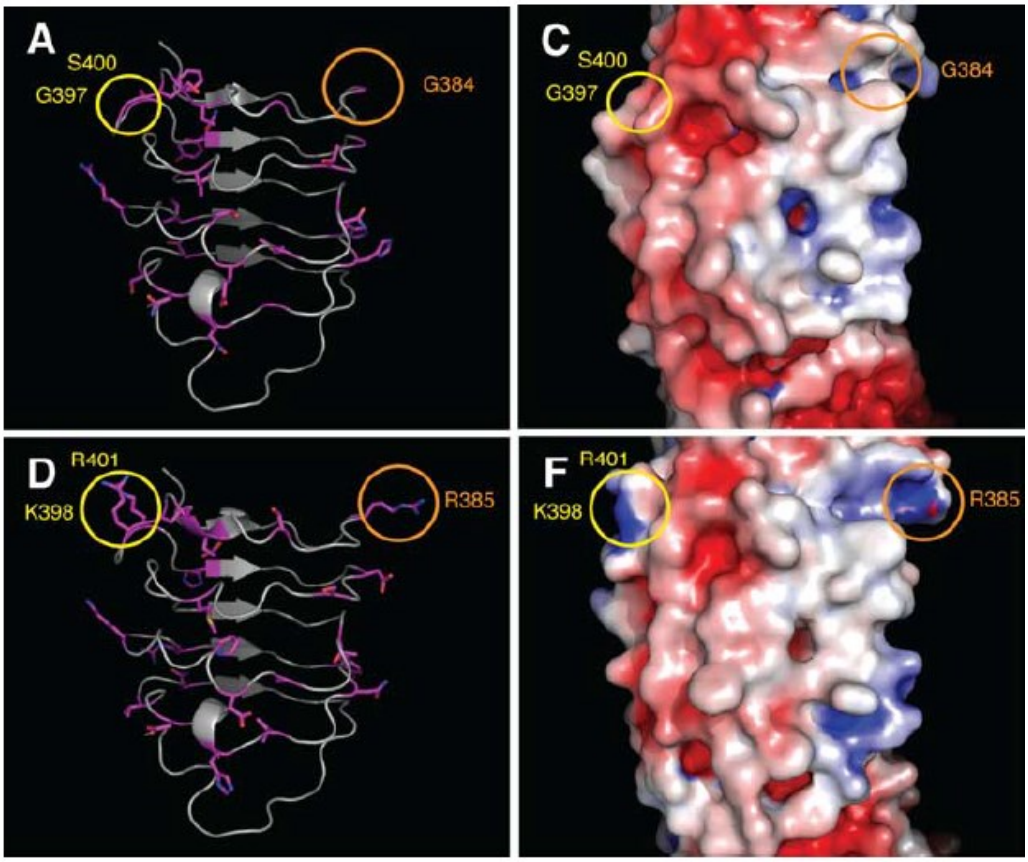


Fig. 8: A single point mutation changing the ligand-binding capacities in TLR4. Species differences in receptor shape and distribution of the electrostatic potential on receptor surface are visualised - the part of human (A and C) and horse (D and F) TLR4 are shown. Encircled residues have been swapped between human and horse proteins. In yellow circles, horse TLR4 K398G and R401S "humanized mutants" on the left flank of the LRR motif did not affect lipid IVa activity. In orange circles, the horse TLR4 R385G "humanized" mutant localized on the right flank lost the ability to signal with lipid IVa. The figure is adopted from Walsh et al. (2008).

Taken altogether, the genetic variability seems to be crucial for function of TLRs, but only little attention is paid to evolution of TLR variability in birds. This was also the main reason, why I have focused on evolution of TLRs in free-living birds during my PhD studies.

Aims of the Thesis

The general aim of my doctoral thesis is to illuminate the adaptive evolution of Toll-like receptors in birds. To achieve this goal, my colleagues and I first decided to focus on various aspects of evolution of these key innate immunity receptors, aiming to provide the description of genetic and also protein variability of TLRs. We predicted variability both on intra and interspecific levels high in non-model free-living birds. Hypothesising the TLRs to be directly influenced by host-pathogen coevolution, we strived to identify sites under positive selection in avian TLRs. Expecting high levels of false-positive results, we proposed that key adaptive variants should differ in their functional effects. Given that TLRs enter into direct molecular contacts with the pathogenic ligand, we predicted especially the receptor surface to play an important role in receptor function. Furthermore, we were interested in answering the question, how various ecological patterns could influence evolution of avian TLRs. As a special part of the project, I also aimed to uncover the molecular regulation of the inflammatory process in skin, because measuring this response could be used to study the TLR functional variation in birds.

Therefore, as partial aims of my doctoral thesis I set:

1) **To describe the history and general patterns of TLR evolution on interspecific level in birds**
2) **To assess the level of avian TLR polymorphism on intraspecific level**
3) **To test for selection acting in avian TLRs**
4) **To predict the functional significance of adaptive changes in TLR structures in birds**
5) **To identify ecological aspects influencing evolution of avian TLRs**
6) **To characterize molecular regulation of inflammation in birds**

Material and Methods

In this chapter, I would like to generally introduce the materials used in my research and all main methods adopted in the papers included in my doctoral thesis. For clear arrangement I divided this section into eight main subchapters: Genetic material, Molecular genetic laboratory procedures, Gene expression analysis, Bioinformatics, Phylogenetic analysis, Selection analysis, Conservativeness of amino acid substitution, and Protein structure analyses. I mostly describe the used methods only briefly, for detailed information see methods in the papers included in this thesis and cited throughout the text.

Genetic material

Consistent with the focus of my work, the genetic samples were used as the initial and crucial material for most aspects of my work in all of the studies. We analysed genetic samples collected during sampling procedures ***in the field*** (PAPER VI.; PAPER V.). Other very useful tissues for further genetic analyses, especially from the species not inhabiting the territory of the Czech Republic, came as scientific loans from various ***genetic banks*** (PAPER I.; PAPER III.; PAPER IV.), namely from The DNA Bank of the Natural History Museum of Oslo, Norway (NHMO), The Museum of Zoology, University of Michigan, USA (UMMZ), The Institute of Vertebrate Biology AS CR (IVB) and to these we often included also samples of the free-living birds from the Czech Republic we had previously collected for our Genetic bank of the Department of Zoology, Charles University (ZCU). Furthermore, whenever possible, we downloaded target ***sequences*** for at least some of the investigated species from publicly available resources, e.g. the National Centre for Biotechnology Information (NCBI) GenBank or The RCSB Protein Data Bank (PDB; PAPER I.; PAPER II.; PAPER III.; PAPER IV.). Especially in the case of the crucial paper of my thesis (PAPER I.) were most of the investigated sequences extracted from the whole-genome data generated by the Avian Phylogenomics Consortium (Zhang, Jarvis, et al. 2014), genomes included also in the Bird 10,000 genomes project (http://b10k.genomics.cn/; Avianbase, Eöry et al. 2015).

Molecular genetic laboratory procedures

Most of the molecular-genetic work was done in laboratories of Institute of Vertebrate Biology, the Czech Academy of Sciences in the external research facility in Studenec. Among the main laboratory procedures belong: DNA/RNA extraction, PCR product amplification, target sequencing, gene expression and copy number analysis.

DNA/RNA extraction

The processing of genetic samples started always with DNA/RNA extraction based on the specimen type. Working with birds, for majority of the genetic analyses we used blood samples readily available for ***DNA extraction***. Compared to mammals, avian erythrocytes possess nucleus (Lucas and Jamroz 1961) and thus serve as a rich source of DNA, with very effective processing and yield. DNA was extracted in all included papers identically using the DNeasy Blood and Tissue Kit (QIAGEN) following the manufacturer's protocol and subsequently stored at -20 °C. In the case of RNA studies included in this thesis (PAPER VI.; PAPER V.), various tissue samples were collected and stored in RNAlater (QIAGEN) at -80°C until the ***RNA extraction***. Total RNA was extracted using High Pure RNA Tissue Kit (Roche) and the RNA was subsequently reverse-transcribed into cDNA using the

Transcriptor First Strand cDNA Synthesis Kit (Roche) and random hexamer primers according to the manufacturer's protocol and stored deeply frozen.

PCR amplification

Before Polymerase Chain Reactions (PCR) product amplification, the specific ***primers*** must be designed. I have typically attempted to locate the primers into conservative regions of the target genes that were identified based on alignment of previously known sequences of species as related as possible. When designing primers (e.g. using OligoAnalyzer 3.1, Integrated DNA Technologies, Inc.) the crucial properties I have considered were: melting temperature during annealing phase of PCR, which is influenced by GC content and primer length; predicted secondary structure and the ability to form hairpins; and the capacity of homo- and hetero-dimerization corresponding to the free energy of the oligo sequence binding to the complement of its own or to the primer-pair sequence. The primers designed for all studies included in this thesis were synthetized commercially by Generi Biotech. The optimal ***PCR conditions*** for all sets of primers were then tested on annealing temperature gradient using thermocycler Mastercycler (Eppendorf) and visually checked on gel electrophoresis. For 454 sequencing (used in PAPER III. and PAPER V.) the target-specific primers were elongated by individual-specific barcodes (Caporaso et al. 2012) as well as by the key sequence (TCAG) and specific 454 adaptor sequence (Roche). PCR was in all cases performed using Multiplex PCR Master Mix (QIAGEN). Before sequencing, each successfully amplified ***PCR product*** was ***purified*** using Exo-CIP PCR clean up protocol, where Exo (Exonuclease I - E. coli, New England Biolabs) is an enzyme that will cut up single-stranded DNA such as primers and unfinished PCR products and CIP (Calf Intestinal Alkaline Phosphatase, New England Biolabs) removes dNTPs that were not incorporated into the PCR DNA; or using magnetic beads (Agencourt AMPure XP, Beckman Coulter Life Sciences) in the case of NGS approach (PAPER III.; PAPER V.).

Sequencing

For sequencing we used two different approaches in papers included in this thesis: Sanger sequencing and target amplicon next-generation sequencing (NGS). ***Sanger sequencing*** was used not only for gene description (PAPER IV.; PAPER V.), but also as the first step before expression analysis (PAPER IV.; PAPER V.; PAPER VI.), copy number variation (CNV) analysis (PAPER I.), or NGS method (PAPER III.; PAPER V.) to sequence the broader region surrounding the region of interest in all genes with the aim to ensure the specificity of any subsequently designed primers. For Sanger sequencing the purified PCR products were prepared using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems, Life Technologies Corporation) following the standard protocol and subsequently processed on an ABI Genetic Analyser machine (Applied Biosystems). For ***target NGS*** we used two 454 platforms (Roche): GS Junior (PAPER V.) and GS FLX+ (PAPER III.). For detailed description of the procedures used see the Section 1.2 in the ESM of PAPER V. or the Subchapter TLR4 sequencing in Material and methods of PAPER III. Application of target NGS approach is very useful, because it allowed us to avoid the costly cloning step in our effort to obtain single-haplotype gene sequences.

Copy number variation analysis

The copy number variation (CNV) analysis was used to resolve the number of copies occurring the particular gene in the genomes of investigated species (PAPER I.). Real-time quantitative PCR (qPCR) is one of the methods usually used to assess the CNV (Weaver et al. 2010) because of its easy use, sensitivity, and scalability. The relative quantification principle is applied, where at least one

reference gene occurring only in a single copy in the whole genome (single-copy gene) must be defined and serve as a reference for the examined gene of interest. For this type of analysis, it is important to design specific primers to the conserved regions to reach the ideal qPCR effectivity. Furthermore, it is crucial to amplify PCR products of similar lengths in all genes included in the CNV analysis to achieve result as precise as possible. In our study (PAPER I.), each sample was run in technical triplicate in LightCycler 480 Instrument II (Roche) using EvaGreen® Dye (Biotum; Mao et al. 2007). The qPCR efficiency was calculated based on a dilution series for each gene and sample in LightCycler® 480 software v1.5.1 using both 2nd Derivate function and automatic Fit Point method. The target gene copy number was then calculated based on a modified version of formula proposed by Pfaffl (2001):

$$R = (2*Eff_{TLR7}\text{^}(-cp_{TLR7}))/(Eff_{TLR3}\text{^}(-cp_{TLR3})+Eff_{TLR4}\text{^}(-cp_{TLR4})),$$

where R means the relative ratio of target potentially duplicated gene (*TLR7*) and both single-copy genes (*TLR3* and *TLR4*) for each investigated species. The modified formula calculates with PCR efficiency (Eff) for each gene raised to the power of the negative number of cycles (Cp) in which the amplified gene reaches the fixed threshold.

Gene expression analysis

We used the gene expression analysis to answer three different questions. First, in the case of *TLR5* pseudogene in birds (PAPER IV.) we wanted to check for presence or absence of a functional copy of this gene in the genome based on ***mRNA expression*** of the sequence. It was important to treat the extracted RNA with RNase-free DNase I (Thermo Fisher Scientific Inc.) to remove any possible DNA contamination. Reverse transcription of total mRNA to complementary DNA (cDNA) was performed immediately afterwards with the Transcriptor First Strand cDNA Synthesis Kit (Roche) using random hexamer primers. The cDNA product was used as a template for PCR amplification of target (pseudo)gene region, where the specific primers were designed based on the sequence of an interspecifically conserved region. Always two independent PCR reactions (duplicates) were tested per individual to support the results. After the PCR amplification, the presence of a PCR product was verified using 1.5 % agarose gel electrophoresis and visualised using GoodView (SBS Genetech) and UV Transilluminator (GenoPlex, VWR International). Second, to measure *TLR* expression in various tissues of the grey partridge (PAPER V.) we performed a ***semi-quantitative PCR*** with the Kapa2G Robust PCR kit (Kapa Biosystems, Inc.). The PCR amplified products (in triplicates) were visualised using gel electrophoresis and subsequently the relative differences in gene expression were measured by relative band intensity (GenoSoft v. 4.00, VWR International) and normalised by b-actin values (b-actin was used as a reference gene treated in the same way as *TLRs*). Third, more recently in the paper describing the differential expression in selected cytokines before and after PHA treatment (PAPER VI.), we quantified the gene expression using classical ***real-time qPCR*** performed with the LightCycler 480 Detection System (Roche) with LightCycler 480 SYBR Green I Master chemistry (Roche) and for each investigated gene and tissue in triplicates. Based on the results of a geNorm analysis from six tested candidate genes, as a housekeeping (reference) gene we selected the TATA-box-binding protein (TBP). The gene expression was calculated based on the relative quantification ΔΔCt method (Bustin 2000; Livak and Schmittgen 2001) in LightCycler 480 SW 1.5 (Roche) using both the comparative Ct method (ΔCt) and Fit Points method, respectively. The PCR efficiency was calculated for all individual genes across all treatment and control samples.

Bioinformatics

During my PhD studies, I have used for bioinformatic analysis of the sequenced data many different pieces of software, tools and approaches. The raw ***Sanger sequences*** I first analysed (since 2015) using Sequencing Analysis 5.2, SeqScape v2.5 (both Applied Biosystems) and then with BioEdit Sequence Alignment Editor (Tom Hall, Ibis Biosciences). Later I switch to a more universal and user-friendly software environment provided by Geneious (http://www.geneious.com, Kearse et al. 2012). The sequence data from ***NGS*** must be firstly filtered based on quality, trimmed and de-multiplexed according to the individual barcode using e.g. R packages seqinr, Biostrings and ShortRead (last two developed in BioConductor scientific community) running in the open statistical software and programming language R (R Core Team). ***BLAST*** (NCBI, (Zhang et al. 2000) was used to find the genes of interest according to a given reference sequence in all accessible species (publicly available on NCBI GenBank server, https://www.ncbi.nlm.nih.gov/; or kindly provided by the Avian Phylogenomics Consortium as a part of the Bird 10,000 genomes project, http://b10k.genomics.cn/, Avianbase, Eöry et al. 2015). CDS and also genomic scaffold databases were used as the source of information. For ***multiple sequence alignments*** I used different approaches in different studies always based on the data type: mostly ClustalW (implemented in BioEdit or Geneious), MAFFT (EMBL-EBI; Katoh and Standley 2013) for highly unrelated sequences, PAL2NAL webtool (Suyama et al. 2006) for the construction of nucleotide codon alignments corresponding to the protein sequence alignments, or the EMBOSS Needle pairwise alignment tool (https://www.ebi.ac.uk/Tools/psa/emboss_needle/) for secondary structure protein alignment. ***Translation*** of nucleotide into the correct protein sequence was performed using the Translate Tool (https://web.expasy.org/translate/, ExPASy, SIB Bioinformatics Resource Portal; Artimo et al. 2012). The particular allele sequences were in heterozygote individuals reconstructed using PHASE algorithm (Stephens et al. 2001) implemented in the DnaSP software (Librado and Rozas 2009). The occurrence of ***gene conversion*** between duplicated genes is possible to statistically test using GENECONV tool (S. A. Sawyer, Washington University in St. Louis, http://www.math.wustl.edu/~sawyer/geneconv/).

Phylogenetic analysis

Phylogenetic approach we use in our studies very often, mainly to reconstruct the evolutionary history of the target gene/protein. Before any phylogeny analysis is necessary to ***determine the evolutionary model*** that suits the particular sequence alignment the best. For this purpose, we use the FindModel tool (https://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html), the web implementation of the Modeltest (Posada and Crandall 1998). The best fitted model for each alignment is afterwards selected based on the given Akaike information criterion (AIC). For the phylogenetic analysis we use two approaches: ***maximum likelihood*** calculated in PhyML software package (Guindon et al. 2010) and ***Bayesian estimation*** of phylogeny calculated using MrBayes (Ronquist et al. 2012). To compare gene tree with the species phylogeny is necessary to reconstruct also the species tree, where if you do not have your own data (our case), two approaches could be used. One option is to reconstruct the schematic consensus species tree based on up to date published avian phylogeny (e.g. Jarvis et al. 2014) but without the information about the length of individual branches. The second approach we use is to generate the phylogeny tree for all investigated species only from the global phylogeny of birds using a web-based tool application available at http://birdtree.org/ (Jetz et al. 2012). We used the Hackett sequenced species backbone

(Hackett et al. 2008) as the source tree with 1,000 randomly generated trees. The maximum clade credibility tree is subsequently produced using the TreeAnnotator tool incorporated in the BEAST software (Drummond et al. 2012). All generated trees are afterwards graphically adjusted in FigTree (A. Rambaut, University of Edinburgh, UK; http://tree.bio.ed.ac.uk/software/figtree/).

Selection analysis

To investigate if adaptive evolution is acting on target genes/gene regions, the diversifying positive selection was tested. For testing ***site-by-site selection*** the codon alignment must be used, where all codons/regions containing gaps in any species in the alignment are removed and are analysed separately. Firstly, we tested the possibility of ***recombination*** influencing evolution of target genes in the studied species using GARD analysis (Genetic Algorithms for Recombination Detection; http://www.datamonkey.org; Kosakovsky Pond et al. 2006). To test positive selection posing on individual residues on the interspecific level we used various ***codon-based maximum likelihood methods***: PAML (Phylogenetic Analysis by Maximum Likelihood; Yang 2007); FUBAR (A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection; Murrell et al. 2013); SLAC (Single Likelihood Ancestor Counting; Kosakovsky Pond and Frost 2005a); and MEME (Mixed Effects Model of Episodic Diversifying Selection; Murrell et al. 2012). The last three mentioned (FUBAR, SLAC and MEME) are implemented into the Datamonkey server (http://www.datamonkey.org/; Kosakovsky Pond and Frost 2005b; Weaver et al. 2018). In **PAML** the codon-based substitution models (codeml) using comparison of neutral M8a (beta&ω=1) with alternative M8 (beta&ω) model. The likelihood ratio test (LRT) for comparison of two nested models was calculated using chi-square approximation:

$$Chi^2 = 2*(lnLM8 - lnLM8a),$$

where LM8 and LM8a are likelihood values. The degrees of freedom (df) were established as the difference in the numbers of parameters in the models used. If the LRT is significant (≤0.05), positive selection is considered as supported. The hierarchical Bayes (Bayes Empirical Bayes, BEB) approach implementing Markov chain Monte Carlo routine (Yang 2007) was then used to determine site-specific posterior probabilities indicating positive selection (on the level ≥0.9). **FUBAR** algorithm was always used with the default significance level of posterior probability established to 0.9. This method is more robust and much faster than other available types of the selection analysis based on random effect likelihood (REL methods; Murrell et al. 2013). **SLAC** is a counting method, which tests whether the number of nonsynonymous changes per nonsynonymous site (dN) is significantly different from the number of synonymous changes per synonymous site (dS). Counting methods are very fast and thus suitable for larger datasets, but may lack power, especially for data sets comprising a small number of sequences or low divergence, as the power of the test is limited by the total number of inferred substitutions at a site (Kosakovsky Pond and Frost 2005a). **MEME** is a branch-site model of selection, which fits a mixture model with two classes of dN/dS ratio to each branch in the phylogeny and it is, therefore, able to identify both the pervasive and also the episodic selection acting only on individual lineages of investigated dataset (Murrell et al. 2012), but the number of detected positively selected sites should be taken with caution, because it can be overestimated (evinces a very high rate of false positives; Lu and Guindon 2014).

Conservativeness of amino acid substitution

Along with positive selection we tested the degree of dissimilarity of amino acid substitutions according to their ***physiochemical properties*** using the tool PRIME (PRoperty Informed Model of Evolution) tool available on the Datamonkey server (http://www.datamonkey.org/; Kosakovsky Pond and Frost 2005b). This tool builds on the same conceptual frameworks as MEME (Murrell et al. 2012) but allows the non-synonymous substitution rate β depend not only on the site in question, but also on which residues are being exchanged. Two predefined sets of 5 amino-acid composite properties can be used for PRIME analysis: Polarity index, Secondary structure factor, Volume, Refractivity/Heat Capacity and Charge/Iso-electric point (Atchley et al. 2005) and Chemical Composition, Polarity, Volume, Iso-electric point and Hydropathy (Conant et al. 2007). The level of significance was always settled as default (posterior probabilities ≥0.9). Amino acid physicochemical properties (chemistry, charge and hydrophobicity) at all positively selected sites were graphically visualised using a web-based application Weblogo 3 (http://weblogo.threeplusone.com; Crooks et al. 2004). To assess the ***possible function*** of detected positively selected sites, the comprehensive review of already published literature describing residues with any function as well as other studies detecting positive selection on interspecific level was done (for the most recent version see Supplementary Material 6: Table S27 in PAPER I.). Afterwards the distances of detected positively selected sites from these already known functionally important positions were measured using 3D structural protein models (for details see the next section Protein structure analyses, p. 27) in the PyMOL Molecular Graphics System (Schrödinger, LLC) using python command iterate and plugin distancetoatom. Positively selected sites lying up to 5Å were supposed as closely connected to functionally important residue with the potential to influence the protein function.

Protein structure analyses

The key stone of papers included in my thesis (PAPER I.; PAPER II.; PAPER III.; PAPER V.) comprises the protein structure description and functional predictions based on protein structure modelling. To predict the general distribution of ***structural domains*** in the proteins we used a combination of the following web-based tools: SMART (Simple Modular Architecture Research Tool; http://smart.embl-heidelberg.de/; Letunic and Bork 2018) for the general identification and annotation of protein domains and the analysis of protein domain architectures; SignalP (http://www.cbs.dtu.dk/services/SignalP/; Petersen et al. 2011) for prediction of the presence and location of signal peptide cleavage sites in amino acid sequences; LRRfinder (http://www.lrrfinder.com/; Offord et al. 2010) for the identification of LRRs in protein sequences; DAS-TMfilter (http://www.enzim.hu/DAS/DAS.html; Cserzo et al. 2004) for detection the presence and location of transmembrane region of protein; PFAM (http://pfam.xfam.org/; Finn et al. 2016) for definition specific domains in our sequence compared to the database including a large collection of defined protein families.

For prediction of ***tertiary structure*** model of target proteins, we use two approaches: I-TASSER (Iterative Threading ASSEmbly Refinement; https://zhanglab.ccmb.med.umich.edu/I-TASSER/; Roy et al. 2010) and Modeller (Webb and Sali 2014). The I-TASSER server uses a hierarchical protein-structure modelling approach based on secondary structure enhanced profile–profile threading alignment and iterative implementation of the threading assembly refinement program (Zhang 2008). Modeller is used for homology or comparative modelling of protein three-dimensional structures including e.g. the possibility of creating the multiple alignment of protein sequences and

structures or clustering and comparison of protein structures (Webb and Sali 2014). As a template for a new structural prediction, a structure of a related protein with known amino acid sequence and empirically verified crystallographic structure is used. For TLRs ECD, TM and ICD were modelled separately. For I-TASSER predictions the model with the highest C-value reflecting the confidence score for estimating the quality of the predicted models was chosen, while for Modeller predictions the model with lowest DOPE (Discrete Optimized Protein Energy; Shen and Sali 2006) score was always selected for further analysis. The quality of each 3D protein structure model was evaluated using the ModFOLD Model Quality Assessment Server (http://www.reading.ac.uk/bioinf/ModFOLD/; Maghrabi and McGuffin 2017). The pairwise structure comparison was done on DALI protein comparison server (http://ekhidna2.biocenter.helsinki.fi/dali/; Holm and Laakso 2016) using the pairwise structure comparison. The graphical visualisation of 3D protein structures was done using the PyMOL Molecular Graphics System (Schrödinger, LLC).

Although we realize that the accuracy of the predicted models is limited, the aim of the analysis was not to describe the proteins' tertiary structures precisely (and substitute the experimental crystallography studies), but to assess the properties characterising and distinguishing the given proteins in the investigated species. Except for the structural protein similarity analysis and the visualisation of the positions of positively selected sites, we focused also on analysis of ***evolutionary non-conservativeness of amino acid positions*** to estimate location the functionally variable regions. This was achieved using an online software ConSurf (http://consurf.tau.ac.il; Glaser et al. 2003; Ashkenazy et al. 2016). Furthermore, we also predicted ***protein surface electrostatic properties*** (PAPER II.; PAPER III.), because the surface charge distribution could play a crucial role in defining the protein ligand-binding functional properties. Protein electrostatic potentials were calculated using PDB2PQR server (http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/; Dolinsky et al. 2004) based on the PARSE force-field and electrostatic calculation on the APBS web (Adaptive Poisson-Boltzmann Solver; http://www.poissonboltzmann.org/; Baker et al. 2001). Surface charge distribution was afterwards visualised using Jmol, an open-source Java viewer for chemical structures in 3D (http://www.jmol.org/) or UCSF Chimera software (developed by UCSF Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco; Pettersen et al. 2004). The UCSF Chimera software was used also for ***aligning the 3D protein structures*** and such aligned dataset was uploaded to the web-based PIPSA tool (Protein Interaction Property Similarity Analysis; http://pipsa.h-its.org/pipsa/; Richter et al. 2008) to acquire a matrix of species with the ***pairwise comparisons of their surface electrostatic potential distances***. The cluster analysis was performed in R (R Core Team, 2017) using RStudio (RStudio Team, 2015) based on the matrix of pairwise electrostatic distances from PIPSA and using the upgma function (package phangorn; Schliep 2011). The phenetic tree (dendrogram) of surface charge distribution was then compared with a neutral phylogenetic species tree.

Results and Discussion

Although Toll receptors were discover and described already two decades ago in insects (*Drosophila*, (Medzhitov et al. 1997) and TLRs have been described in model vertebrate species few years after that (mainly in human, mouse, followed by other mammals; Smirnova et al. 2000; Ferwerda et al. 2007; Shen et al. 2010; Areal et al. 2011; Shen et al. 2012; Smith et al. 2012; Quach et al. 2013), in birds the situation remained for a long time largely overlooked. The avian TLR research oriented mainly for studies on the domestic chicken (Leveque et al. 2003, Boyd et al. 2007, Temperley et al. 2008, Kannaki et al. 2010, Brownlie and Allan 2011), with only few first non-chicken avian TLRs described later (de la Lastra and de la Fuente 2007; MacDonald et al. 2008; Cormican et al. 2009; Vinkler et al. 2009; Gopinath et al. 2011). In fact, from the evolutionary point of view it is preferable to focus mainly on the situation in free-living animals to better understand the coevolution of hosts and their pathogens in the natural, non-constrained system. Although recently the situation is improving and the number of studies dealing with TLRs in free-living birds is slowly increasing (see e.g. Alcaide and Edwards 2011; Grueber et al. 2014; Raven et al. 2017), most authors still focus mainly on the description of the *TLR* genetic variability without exploring further functional effects. Therefore, in my PhD project I tried to go a little bit further and predict the impact of the variation detected on functionally important regions and positions (e.g. in PAPER I.; PAPER II.). Thus, in the articles my co-authors and I created predictions which could be subsequently experimentally tested. In particular, we found interspecifically important physicochemical changes on the receptor surface in the ligand binding regions (PAPER II.; PAPER III.), which functionally differentiate individual group of species and so serves as a base for formulation of wider ecological hypotheses on TLR adaptive evolution. We also tested experimentally the phenomena of copy number variation (gene duplication previously described based on sequence data only; PAPER I.) and pseudogenisation (PAPER IV.) and verified the gene functionality based on gene expression analysis (PAPER V.; PAPER VI.).

Variability in avian Toll-like receptors

Although few case studies dealing with the description of complete TLR repertoire concerning the whole CDS sequence in birds exists (e.g. comparing chicken and zebra finch TLRs, Cormican et al. 2009; Brownlie and Allan 2011), the key general study including more avian species was missing until we have published the comprehensive paper about all TLR members across the bird phylogeny (PAPER I.). The TLR family contains initially ten receptors in most of the avian species. There are, however, exceptions, since some of the birds have only nine TLRs, lacking TLR5 (for examples see Fig. 10 and the subchapter Gene pseudogenisation, p. 33), while others possess eleven TLRs due to the TLR7 duplication (for examples see Fig. 10 and the subchapter TLR7 in Gene duplication, p. 32). Even though all avian TLRs are very conservative in their extracellular “horse shoe”-shaped structure (PAPER II.), individual receptors differ in their gene/protein ***sequence lengths*** (Tab. 1; the length of individual TLRs for each investigated avian species is in Supplementary Material 2: Table S25 in PAPER I.). The longest avian TLR is TLR7 (in average 1050 amino acids long), the shortest then TLR1B (in average 653 amino acids long). The most variable TLRs in sequence length within birds are TLR7, TLR15 and TLR1A, where the interspecific difference between shortest and longest full-length protein sequence within birds comprise 54, 44 and 43 amino acids, respectively. This huge variability in protein length is caused, e. g. in TLR7, the unknown exact protein start codon in the case of more methionines at the beginning of the sequence. To resolve which methionine is the real initiation amino acid a functional study is needed. It is possible, that in this case there could be also more

splicing variants of a single gene, which differ in start codon and thus also in protein length (two different splicing variants of TLR7 are described at least in chicken; Philbin et al. 2005). Contrary to TLR7, in TLR15 the start methionine is evident, but decisive for the protein length is large amount of indels at various places following whole sequence in different avian species (PAPER I.). The avian TLR1B represents the combination of both cases mentioned above. On the other site, the most conservative TLRs, which are almost similar in sequence length within all investigated birds, are TLR5, TLR3 and TLR2A, which differs interspecifically only about 5, 7 and 9 amino acids, respectively (Tab. 1). But this is very taxa dependent, for example TLR5 varied in length between 859 and 862 amino acids even only in Galloanserae clade (PAPER II.) and there is also a variation in length (one codon indel) within the grey partridge population in this receptor (PAPER V.). The phylogeny of the *TLR* genes (the phenetic tree) investigated in our studies mirrors very faithfully the known phylogeny of bird species (PAPER II.; PAPER III.; PAPER V.).

Tab. 1: The sequence length of avian TLRs. CDS lengths of avian TLRs is given. The length characteristics are listed both in nucleotides (nt) and amino acids (aa). The mean and median values are given together with the lengths of the shortest (min) and longest (max) representative of each TLR. The number of investigated species per each TLR is shown in the bottom line. The table is adopted from PAPER I.

length	TLR1A		TLR1B		TLR2A		TLR2B		TLR3		TLR4		TLR5		TLR7		TLR15		TLR21	
	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa	nt	aa
mean	2441	813	1962	653	2380	792	2352	783	2689	895	2527	841	2585	861	3152	1050	2617	871	2802	933
median	2457	818	1959	652	2382	793	2352	783	2688	895	2532	843	2586	861	3150	1049	2625	874	2787	928
min	2394	797	1932	643	2355	784	2340	779	2679	892	2478	825	2577	858	3120	1039	2505	834	2778	925
max	2523	840	2016	671	2382	793	2379	792	2700	899	2550	849	2592	863	3282	1093	2637	878	2838	945
species	51		52		50		50		53		54		47		57		54		14	

Another mechanism which modify protein length is ***proteolytic cleavage***. It is present in receptors involved in nucleic acid recognition (i.e. TLR3, TLR7 and TLR15 in birds) and by the cleavage of receptor ECD's part is regulating their function (Ewald et al. 2011; de Zoete et al. 2011; Garcia-Cattaneo et al. 2012; Kanno et al. 2013; Toscano et al. 2013; Murakami et al. 2014). Surprisingly, in the case of avian TLR3, there is a gap in the position of the cleavage site within the glycosylation-free side of LRR12 (residues 335–342 in human TLR3, KQSISLAS; in chicken TLR3 between aa 334-335; PAPER I.) previously described in humans (Choe et al. 2005; Garcia-Cattaneo et al. 2012; Toscano et al. 2013). However, the question on the influence of this missing residues in avian TLR3 cleavage site on receptor function is still not answered.

Despite avian TLRs are coding by usually highly variable immune genes (PAPER I.; Alcaide and Edwards 2011; Grueber et al. 2014; Świderská et al. 2018), surprisingly we found only a ***limited genetic polymorphism*** in grey partridge population (PAPER V.). Specifically, we revealed only 11 *PePeTLR4*, 8 *PePeTLR5* and 6 *PePeTLR7* nucleotide polymorphic positions. Unfortunately given their location and chemical features, most of the substitutions probably have only a minor impact on receptor function. However, this genetic polymorphism formed 10 (*PePeTLR4*), 6 (*PePeTLR5*) and 3 (*PePeTLR7*) alleles, resulted in 5, 4 and 3 non-synonymous haplotypes, respectively in 10 investigated individuals, unrelated birds from four different wintering flocks in the Czech Republic. As the same major haplotypes were found repeatedly also within this limited dataset, we have almost certainly described the most frequent protein-coding diversity in these genes in our study population. The generally low level of intraspecific genetic variation in partridge TLRs compared for example to related domestic chickens (W. Ruan et al. 2012; W.K. Ruan et al. 2012) and junglefowl (Downing et al.

2010) may be a result of either negative selection or a population bottleneck caused by local inbreeding. We suppose the later scenario is more probable, because the European grey partridge has declined severely in numbers in last few decades (De Leo et al. 2004; Kuijper et al. 2009), including the population inhabiting the Czech Republic, which decreased dramatically between the 1930s and 1990s from ca. 6 million individuals to less than 20 thousand (Šťastný et al. 2010), resulting in a highly fragmented population. This low genetic variability in immune-related genes may ultimately explain the relative susceptibility of the grey partridge population to infectious diseases (Sedlák et al. 2000; Vitula et al. 2011).

Gene duplication

Gene duplication is a very common mechanism acting in fast evolving genes, including immune genes. Therefore, we have focused on this issue also in the avian TLRs. Our main aims were to resolve the duplication event in TLR1 family in vertebrate evolution and also to describe the recent duplication in avian TLR7.

TLR1 family

The duplications in *TLR1* family has been previously well documented in vertebrates, birds including (Temperley et al. 2008; Cormican et al. 2009; Huang et al. 2011; Wang, Zhang, Liu, et al. 2016). But there has been inconsistency in the estimated time of these duplication events in the literature. The main question is, if the ***TLR1 gene duplication*** had occurred before mammals diversified from the sauropsids (Huang et al. 2011), or this duplication followed only after this divergence, independently in avian and mammalian lineage (Temperley et al. 2008; Cormican et al. 2009; Mikami et al. 2012; Wang, Zhang, Liu, et al. 2016). We assume (PAPER I.), that this contradiction is caused methodologically by omitting the phenomenon of gene conversion between the two copies in most of the studies (namely in Temperley et al. 2008; Cormican et al. 2009; Mikami et al. 2012; Wang, Zhang, Liu, et al. 2016), even though the gene conversion occurred in avian *TLR1* and also in avian *TLR2* without doubt (for detailed info see Supplementary Material 3 in PAPER I.). When the phylogenetic analysis is performed using the whole gene sequence, the converted part of the duplicated genes masks the effect of phylogeny and this brings both paralogues incorrectly into one clade. Therefore, we constructed the TLR1 tree based on the non-converted region only (Fig. 9; PAPER I.) and our results supported the previous findings of Huang et al. (2011), that the duplication in TLR1 family arose before the Synapsids diversified from Sauropsids, while the duplication of *TLR1* in amphibians (namely in *Xenopus*) arose apparently independently on the TLR1 duplication in amniotes. On the contrary, it seems that the ***duplication of avian TLR2*** arose in birds independently on mammals, where the second copy of *TLR2* is present only as non-functional pseudogene in several species (for more info see PAPER I.).

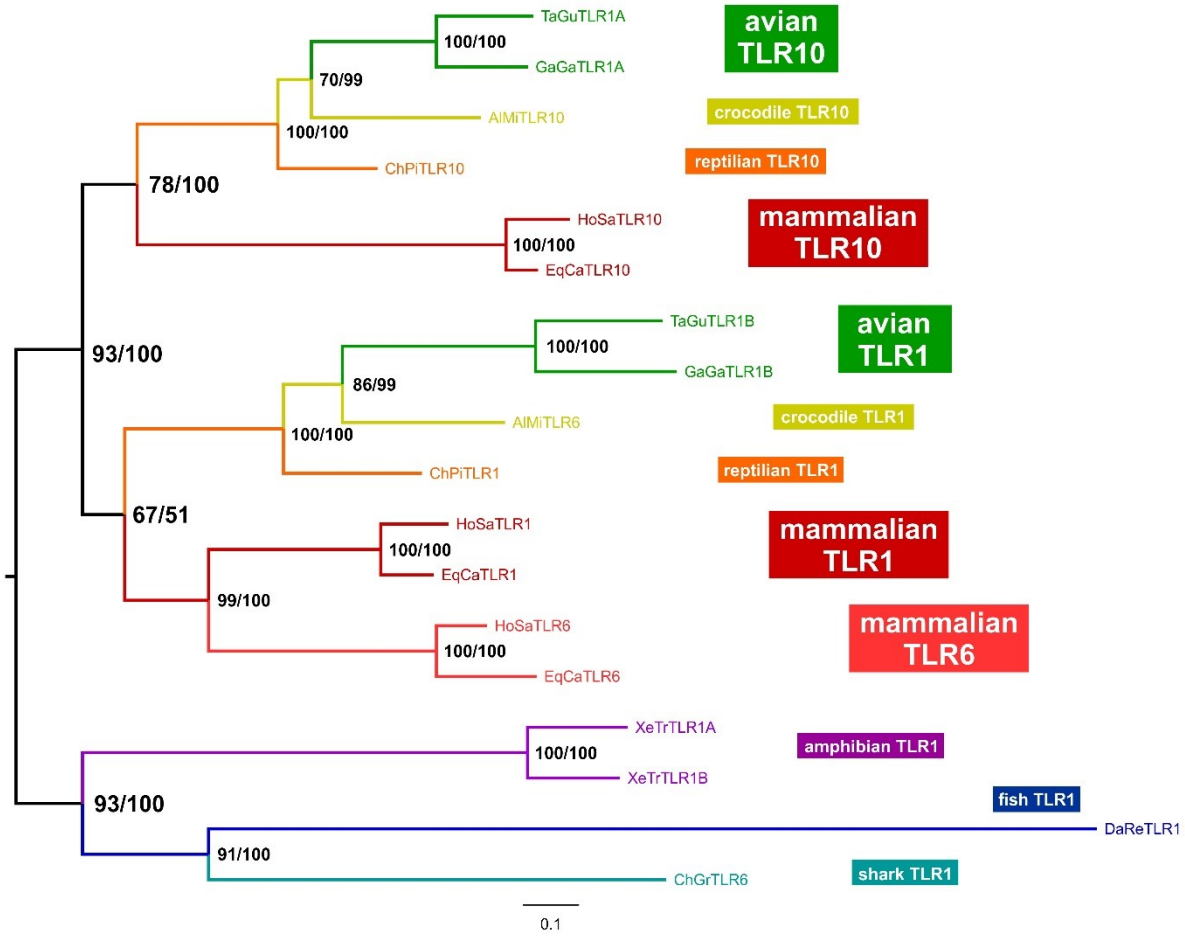


Fig. 9: Duplication in TLR1 family. Tree based on non-converted regions of *TLR1* subfamily members is visualised. PhyML/MrBayes values in percentage per each node are provided. Birds are represented by zebra finch (TaGu, *Taeniopygia guttata*) and chicken (GaGa, *Gallus gallus*), crocodiles by alligator (AlMi, *Alligator mississippiensis*), reptiles by turtle (ChPi, *Chrysemys picta*), mammals by human (HoSa, *Homo sapiens*) and horse (EqCa, *Equus caballus*). Amphibians represented by clawed frog (XeTr, *Xenopus tropicalis*), fish by zebrafish (DaRe, *Danio rerio*) and chondrichthyes by sharks (ChGr, *Chiloscyllium griseum*) were used as outgroups. The analysis was performed using a single amino acid sequence per TLR per species. Based on the results we suggest renaming TLR1A to TLR10 and TLR1B to TLR1 in birds. The figure is adopted from PAPER I.

TLR7

Birds lack two of the three vertebrate *TLR7* family members (TLR8, TLR9; Philbin et al. 2005; Temperley et al. 2008). Instead, avian *TLR7* locus appears to be duplicated in several avian species belonging to various orders: Passeriformes, Charadriiformes, Cuculiformes and Mesiornithiformes (Fig. 10; PAPER I.; Cormican et al. 2009; Grueber et al. 2012; Raven et al. 2017). Moreover, we found out that this duplication is evolutionarily very recent (in comparison to the duplications in TLR1 family) and arose in each clade independently (PAPER I.). Furthermore, we predict that the two avian *TLR7* copies can probably generally slightly differ in their function, because 11 from the sites identified as distinct between the two copies in any avian species match the known ligand binding positions of TLR7 (Wei et al. 2009; Gupta et al. 2016; Zhang et al. 2016), 5 are situated into the z-loop region responsible for TLR7 binding and dimerization (Zhang et al. 2016) and other 10 variable sites are identical with positively selected sites detected in birds (Supplementary Material 1: Table S3 in PAPER I.). We are also first, who supported the existence of the two copies of avian *TLR7*, described previously based on sequence data only, also experimentally using CNV analysis (PAPER I.).

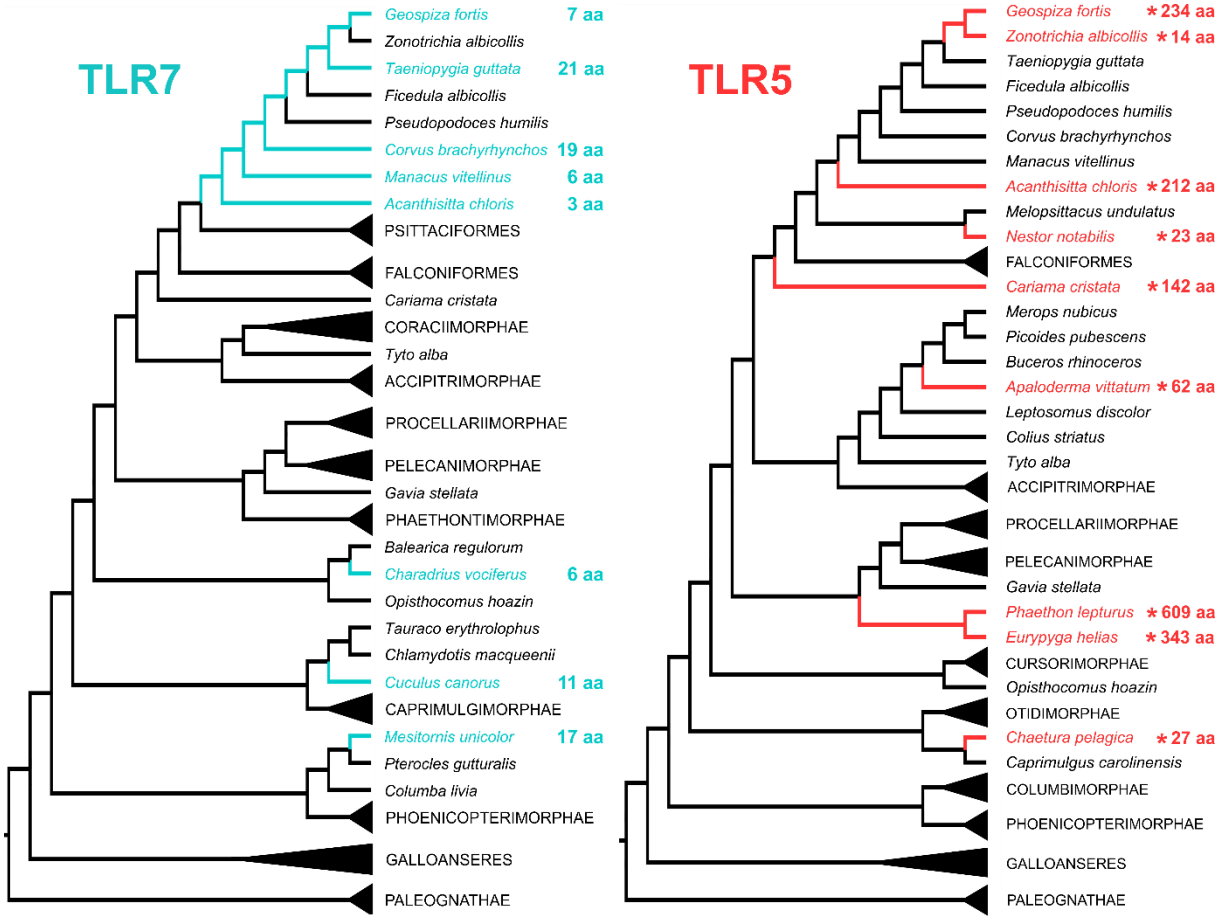


Fig. 10: *TLR7* duplication and *TLR5* pseudogenisation in birds. In the schematic representations of the avian phylogenetic tree the species with duplicated *TLR7* are highlighted in blue and species possessing *TLR5* pseudogene only are highlighted in red. The numbers of amino acid substitutions distinguishing the two copies of the duplicated *TLR7* are shown behind the species name. The position of the first stop codon in *TLR5* is indicated by the number provided behind the species name. Position numbering is according to the chicken reference. Both figures are adopted from PAPER I.

Gene pseudogenisation

Not only gene gain, but also functional gene loss, so called pseudogenisation, is a very common mechanism in immune genes, *TLRs* including. Until present, some examples of this phenomenon have been described in mammals, where e.g. *TLR2* (Roach et al. 2005) or *TLR5* (Hawn et al. 2003) pseudogene is known.

Despite TLR5 belongs to one of the most important PRRs allowing the recognition of invading bacteria based on their flagellin recognition, the occurrence of ***TLR5 pseudogene*** was described first in humans (Hawn et al. 2003) and even recently by our group also in birds (PAPER I.; PAPER IV.). Species possessing *TLR5* pseudogene are present in different avian orders: Passeriformes, Psittaciformes, Cariamiformes, Trogoniformes, Phaethontiformes, Eurypygiformes and Apodiformes (Fig. 10; PAPER I.). Moreover, also within some orders are present species with and without functional *TLR5* (e.g. in passerines; PAPER IV.). The sequence data show that the stop codon positions differs among the investigated species with differences also in the pseudogenisation mechanism (single nucleotide substitution or frame-shifting indels, see Supplementary Material 1: Table S8 in PAPER I.), but always affect mainly the extracellular part of the receptor responsible for bacterial flagellin recognition and binding (Keestra et al. 2008). The stop-codons in *TLR5*, thus, arose independently in the recent

evolutionary history of different avian taxa (PAPER I.; PAPER IV.), e.g. at least during seven independent events of *TLR5* pseudogenisation in passerines (for details see Fig. 1 in PAPER IV.). To test the presence of potential functional copy of *TLR5* in species possessing pseudogene, we performed the gene expression analysis (PAPER IV.), which fully supported our previous results based on gene sequencing. We found no PCR product using cDNA template compared to genomic DNA template in species, where we previously described the *TLR5* pseudogene (Fig. 11).

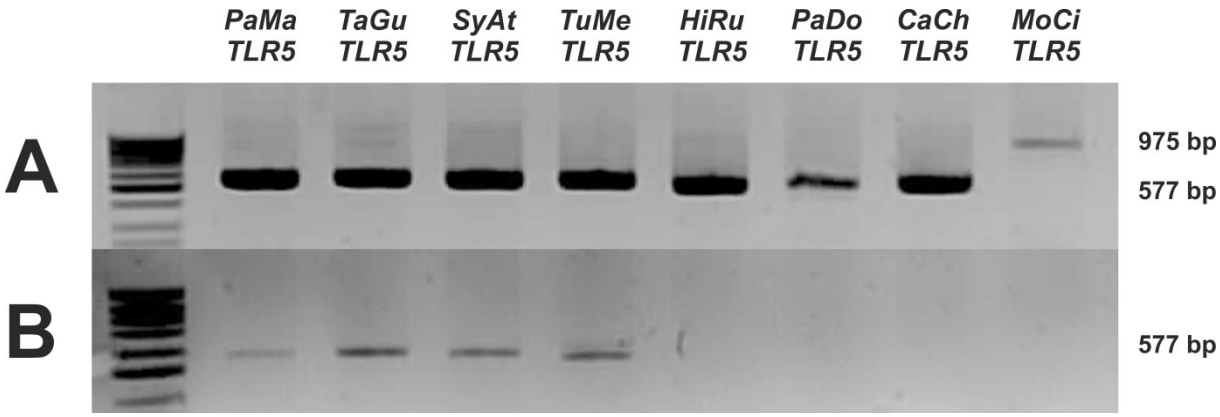


Fig. 11: *TLR5* expression in selected passerine species. Gel electrophoresis image of *TLR5* PCR products amplified with the same primer pair for all investigated species (PaMa – *Parus major*, TaGu – *Taeniopygia guttata*, SyAt – *Sylvia atricapilla*, TuMe – *Turdus merula*, HiRu – *Hirundo rustica*, PaDo – *Passer domesticus*, CaCh – *Carduelis chloris*, MoCi – *Motacilla cinerea*). Genomic (A) and complementary (B) DNA has been used as a template. Longer PCR product in MoCiTLR5 is caused by 398 bp long insertion in the amplified region of *TLR5* in this species. The figure is adopted from PAPER IV.

The loss of *TLR5* functional gene is limited not only to birds, the pseudogene has been described also in one human allele, where its presence is associated with diseases caused by flagellated bacteria (Hawn et al. 2003; Zhang et al. 2013), while at the same time possessing *TLR5* non-functional allele might be advantageous for decreasing the probability of autoimmune disease development (Hawn et al. 2005). The *TLR5* pseudogenisation may be allowed by the high redundancy of pathogen detection, where flagellin is recognized also by other PRRs, e.g. NLRC4 (Zhao et al. 2011; Yang et al. 2014). But unfortunately, in birds other flagellin-recognising PRRs have not yet been sufficiently studied to support this explanation. The same is true also for the investigation of avian symbiotic microbiota. Therefore, despite it may seem that the species possessing *TLR5* pseudogene do not share any common ecological features, there may be certain component selecting for TLR5 loss of function in these avian evolutionary lineages. Pseudogenisation might be associated with impaired flagellin detection or signalling overreaction and may occur only in taxa where either the risk of infection by flagellated bacteria is low or where other PRRs are sufficient to ensure an adequate immune response. Further research is, however, needed to functionally address the impact of TLR5 loss on immune responsiveness to bacterial flagellin in the TLR5-deficient species.

Diversifying selection

In our research we tested positive diversifying selection acting on individual positions in all avian TLRs. The numbers of detected ***positively selected sites*** (PSS) differ between individual TLRs. Consistently with other vertebrates (Smirnova et al. 2000; Nakajima et al. 2008; Vinkler et al. 2009; Wlasiuk et al. 2009; Wlasiuk and Nachman 2010; Areal et al. 2011; Huang et al. 2011; Shen et al. 2012; Smith et al. 2012; Fornuskova et al. 2013; Wang, Zhang, Chang, et al. 2016), our results in birds

confirm that positive selection is acting more strongly on ECD than ICD part of the TLR protein (PAPER I.; PAPER II.). ECD is responsible for direct and specific PAMPs binding and recognition, contrary to ICD which is responsible for signal transmission due to binding of TLR adapters (Jin and Lee 2008; Kawai and Akira 2010) and thus there is a selective pressure for conservative nature of the receptor domain. Our comparative studies in avian TLR family also suggest, that the positive selection is acting more in receptors exposed towards the cell surface (mainly in TLR1[1B], TLR2A, TLR5, TLR4 and TLR2B), than in TLRs situated in endosomes (TLR21, TLR3 and TLR7; precise numbers in Tab. 2) on both intraspecific (PAPER V.) and interspecific levels (PAPER I.; PAPER II.), which is in concordance with the general findings in vertebrate TLRs (Mikami et al. 2012). This could be explained by the fact that the endosomal TLRs are adapted for detection of ligands showing low structural variation (such as ssRNA in the case of TLR7, dsRNA in TLR3 or CpG DNA regions in TLR21; Diebold et al. 2006; Brownlie and Allan 2011), while the transmembrane TLRs recognize structurally more variable complex bacterial ligands (see e.g. Andersen-Nissen et al. 2005; Marr et al. 2010; DeMarco and Woods 2011). The only exception from this apparently universal rule represents TLR15 which is unique to birds and reptiles (Boyd et al. 2012). Although situated on the cell surface, TLR15 is activated by the ECD proteolytic cleavage with pathogen-derived proteases (de Zoete et al. 2011). We hypothesise that this pathogen-recognition mechanism may be linked to reduce positive selection acting in ECD. Nonetheless, the cleavage site possesses genetic variability including PSS in birds (Supplementary Material 1: Tab. S11 in PAPER I.), which might play important role in receptor function.

	species	aa length	PSS	PSS/TLR (%)
TLR1A	45	794	**31**	3.9
TLR1B	44	638	**37**	5.8
TLR2A	40	793	**43**	5.4
TLR2B	42	781	**35**	4.5
TLR3	51	895	**19**	2.1
TLR4	54	843	**38**	4.5
TLR5	46	861	**45**	5.2
TLR7	51	1041	**31**	3.0
TLR15	53	868	**24**	2.8
TLR21	14	907	**3**	0.3

Tab. 2: Positive selection detected in avian *TLRs*. Species - number of species (one sequence per species), aa length - the protein sequence length in chicken TLRs (NCBI accession numbers are listed in Supplementary Material 1: Table S14 in PAPER I.), PSS - number of positively selected sites detected in investigated species per each gene, PSS/TLR (%) – the percentage of PSS per whole receptor. The table is adopted from PAPER I.

Our results in avian TLRs (PAPER I.) are also mostly consistent with the findings of (Wang, Zhang, Liu, et al. 2016), who suggested grouping of vertebrate TLRs based on their ***ECD architecture*** and showed that at least in mammals the single-domain TLRs (i.e. in birds: TLR3, TLR5, TLR7, TLR15 and TLR21) are under stronger purifying selection than the three-domain ones (i.e. in birds: TLR1A, TLR1B, TLR2A, TLR2B and TLR4). Based on our results (PAPER I.), the positive selection is acting more strongly in the three-domain TLRs, with the exception of TLR5, which is under strong positive selection in birds (Tab. 2), supported also by results from other studies in avian TLRs (PAPER II.; Alcaide and Edwards 2011; Grueber et al. 2014). The role of TLR5 seems to be special in avian immunity compared to e.g. in mammals, where several studies proposed relaxed purifying selection in this receptor (Wlasiuk et al. 2009; Areal et al. 2011). Besides, the nonsense stop-codons preventing the TLR5 function has been

described in both birds (PAPER I.; PAPER IV.) and mammals (Wlasiuk et al. 2009). One of the explanations for diverse selective pressures acting on avian TLR5 might be that the variation reflects differences between species in gut immunity overactivation through flagellated symbiotic microbiota (Iqbal et al. 2005). Alternatively, there are present also other activation pathways involving other immune receptors that may be adapted for recognising flagellated bacteria, e.g. NLRC4 mentioned above (in section Gene pseudogenisation, p. 33) may be one of them (Zhao et al. 2011; Yang et al. 2014).

In our comparative avian TLRs paper (PAPER I.) we show, that the PSS frequently evolve towards similar amino acid physicochemical properties being gained in distantly related avian taxa, while distinct properties can be found even in closely related species (see Supplementary Material 8 in PAPER I.). This ***convergent evolution*** may probably result from selective pressures induced by similar pathogen communities in different taxa (Waite and Taylor 2014). We also revealed that most of the identified positively selected amino acid residues tend to be exposed to the receptor surface to allow the direct contact with the ligand (PAPER III.).

Non-conservative PSS

We may predict stronger functional effects of PSS with non-conservative amino acid substitutions that importantly change the ***physicochemical properties*** of the particular residues and thus significantly influence the PAMPs recognition. Based on our results reported in PAPER I., among the TLRs possessing the most non-conservative PSS in birds belong TLR4 and TLR2A (Fig. 12 and Supplementary Material 1: Tab. S9 in PAPER I.). Especially in TLR4, most PSS with dramatic changes in charge and hydrophobicity fall into the known functionally important sites of the receptor, with high probability of changing the TLR4 ligand-binding properties (Fig. 12). On the other side there are TLR15, TLR3, TLR1A, TLR7 and surprisingly TLR5, where majority of the amino acid substitutions at PSS does not change dramatically the amino acid properties (e. g. the charge or hydrophobicity; Fig. 12), furthermore at least in the case of TLR7 PSS are conservative in amino acid properties even between birds and mammals (PAPER II.).

Despite avian TLR1B is significantly shorter than TLR1A (Tab. 1), we found more PSS in TLR1B and these were less conservative in their charge than in TLR1A (Supplementary Material 1: Tab. S9 in PAPER I.), especially in the predicted ligand-binding region that is spanning the region avoiding gene conversion (Fig. 12). This, together with high TLR2A PSS variation, can be possibly explained by distinct binding capacities of the heterodimers formed by the duplicated TLR1/TLR2 subfamily members, where the combination TLR1B/TLR2A (unlike any other combination) is able to recognize peptidoglycans and efficiently recognizes diacylated bacterial lipoproteins (Higuchi et al. 2008).

In avian TLR3 most PSS are very conservative (Fig. 12, for details see Supplementary Material 1: Table S12 and Supplementary Material 7 both in PAPER I.), which is consistent with the results reported by Wang and colleagues (Wang, Zhang, Liu, et al. 2016) who identified the TLR3 family as the most conservative TLR family within vertebrates. Two of these conservative PSS in the TLR3 are situated in transmembrane region (Fig. 12 and Supplementary Material 7 in PAPER I.), with the known receptor di/trimerization function (Mineev et al. 2014). Since form of TLR3 dimerization depends on the dsRNA length (Pirher et al. 2008) and the number of TLR3 molecules involved in the interaction (receptor di- or trimer; Mineev et al., 2014), selection on these particular residues may play crucial role in ligand recognition even they are situated into TM domain.

Surprisingly, regardless the enormous number of PSS, TLR5 is the most conservative in the physicochemical properties of the amino acid substitutions at the PSS across all avian taxa (PAPER I.). Thus, the variation present in this receptor likely does not cause many dramatic changes in the receptor ligand-binding properties in most avian taxa. Given that flagellin is the only known TLR5 ligand (Hayashi et al. 2001), this may indicate functional constrains in those species with preserved functional TLR5, limiting any adaptations to larger numbers of relatively minor changes.

Despite of low number of PSS in the known functional sites of avian TLR7 (Fig. 13, detailed in Supplementary Material 9 in PAPER I.), which may result from the constraints of conservative ligand binding, two non-conservative PSS located on the TLR7 Z-loop, which is oriented into the inner concave space of the receptor's horseshoe and is crucial for receptor dimerization and formation of a second ligand-binding site (Zhang et al. 2016). The Z-loop directly follows the TLR7 cleavage site (Ewald et al. 2008; Kanno et al. 2013) and these two PSS may, therefore, affect the receptor cleavage. Similarly in TLR15, the change of hydrophobicity of one PSS (N353; PAPER I.), which is situated at the top of the proline-rich loop included in the TLR15 cleavage region, might play some role during proteolytic cleavage of the receptor (de Zoete et al. 2011) and thus importantly influence the receptor function.

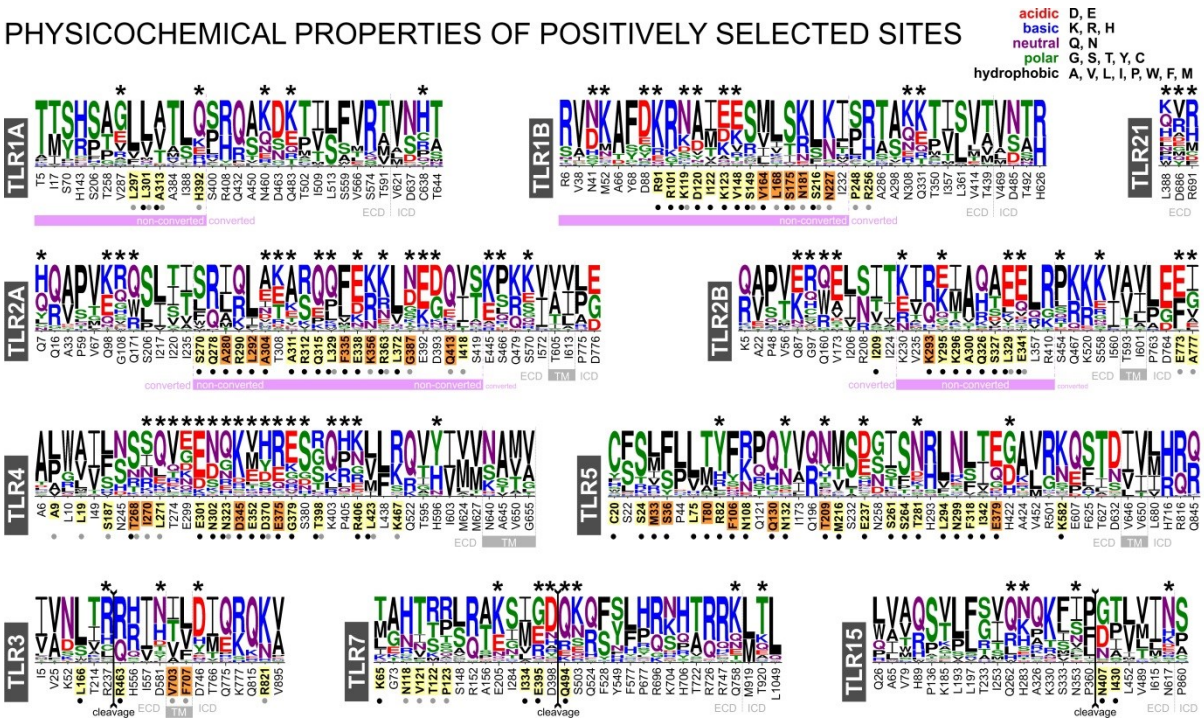


Fig. 12: Physicochemical properties of the positively selected sites (PSS). All PSS are shown in all avian TLRs - amino acid substitutions are coloured according to their physicochemical properties: acidic in red, basic in blue, neutral in purple, polar in green and hydrophobic in black. The size of a letter corresponds to the percentual proportion of that particular amino acid within sequence alignment. The numbering is adopted from chicken TLRs. PSS which correspond to functionally important residues (black dot - ligand binding; grey dot – dimerization) are highlighted in bold and orange (identical site) or yellow (topological proximity closer than 5 Å from a functionally important residue). Non-converted region in TLR1 and TLR2 is highlighted in pink, transmembrane (TM) region in grey, ECD - extracellular domain, ICD - intracellular domain; cleavage site in TLR3, TLR7 and TLR15 is indicated by a black line tipped with arrows (Supplementary Material 1: Table S11 and Text S1 in PAPER I.). Non-conservative PSS are marked by star (Supplementary Material 1: Table S12 in PAPER I.). The figure is adopted from PAPER I.

Amino acid positions in avian TLRs candidate for relevant functional effects

The detected PSS in avian TLRs can be suggested as candidate for relevant functional effect based on the non-conservativeness of the changes in their physicochemical properties (Fig. 12). To even better assess the ***possible functional effects***, we visualised the detected PSS in 3D protein structures with the highlighted previously described ligand-binding and dimerization sites and measured the distances of PSS from these sites (PAPER I.; PAPER II.; PAPER III.; PAPER V.). We found the best agreement between the PSS and functional site distribution in TLR5, TLR4, TLR2A and TLR1B (PAPER I.). The lack of agreement between the identified PSS and previously described TLR functional sites in other TLRs may be caused either by the lack of functional studies known in some TLRs (e.g. TLR15 and TLR21), or it may indicate interspecific variation in TLR-ligand binding (e.g. TLR1, TLR3 and TLR7, where the structural and experimental studies were performed in mammals and not in birds; see Supplementary Material 6: Table S27 in PAPER I.), or even false positive predictions of PSS. We also compared the PSS we identified in avian TLRs with PSS described by other studies in vertebrates, where the best consensus we found especially in TLR15, TLR4 and TLR5 (Fig. 13; PAPER I.; PAPER II.). For both these purposes we reviewed all up to date literature describing TLR functional positions and regions (mainly including the crystallography studies and also site-mutations functionally studies) and also sources investigated positive selection on interspecific level in vertebrates TLRs (for detailed and most comprehensive table see the Supplementary Material 6: Table S27 in PAPER I., but see also PAPER II.; PAPER III.). The main idea is that the consistency between PSS identified in several different studies based on different datasets supports the power of selection acting on that particular position in broader evolutionary context. Thus, this approach should minimize the probability of detection of false positive results. That PSS might play important role in receptor ligand recognition and binding even though not laying in topological proximity to any of the already known functionally important positions. This is supported by the findings in humans, that even amino acid substitutions located far from the functionally important regions (in the case of TLR4, the MD-2-dimerization or LPS-binding sites) can influence the responsiveness to a given pathogenic ligand (here LPS; Arbour et al. 2000). Therefore, one of the aims of our research was to predict new potentially crucial positions not yet described by crystallography or mutation studies. Our predicted PSS might serve well as a starting point for such experimental studies in birds.

We predicted the key amino acid positions based on the criteria mentioned above: non-conservativeness of PSS (especially considering surface charge and hydrophobicity; Fig. 12; Supplementary Material 7 in PAPER I.), the proximity of PSS to already known functionally important positions (Fig. 13; Supplementary Material 9 in PAPER I.) and coincidence of PSS with PSS previously described in other datasets (Fig. 13; Supplementary Material 6: Table S27 in PAPER I.). List and description of such positions is given for each avian TLR in Supplementary Material 1: Text S2 in PAPER I. but see also Additional file 3: Section S4 in PAPER II. and the discussion part in PAPER V. For instance, there were identified 5 positions where are amino acid substitutions related to *Salmonella enterica* resistance variation in chickens (Leveque et al. 2003). One of them (Y383H) was detected as PSS also within Galloanserae clade (PAPER II.) and the other (E301D) even within all birds (PAPER I.), where moreover negatively charged E/D are substituted with positively charged K/R (Fig. 12) supported also by other studies in birds (Grueber et al. 2014) and mammals (Vinkler et al. 2009; Wlasiuk and Nachman 2010) and partially with the results of numerous human studies (Arbour et al. 2000; Zareparsi et al. 2005; Shen et al. 2010; Cario 2013; Rupasree et al. 2015; but see also Ohto, Yamakawa, et al. 2012) showing potential importance of variation at a neighbouring position D299G

(i.e. D303 according to chicken numbering). Based on other our findings, it seems that avian TLR4 is identical to equine, but distinct from human TLR4 in the lipid IVa recognising by TLR4-MD2 complex, because most of the investigated avian species possess arginine at the crucial position 393, similarly to horses (Walsh et al. 2008).

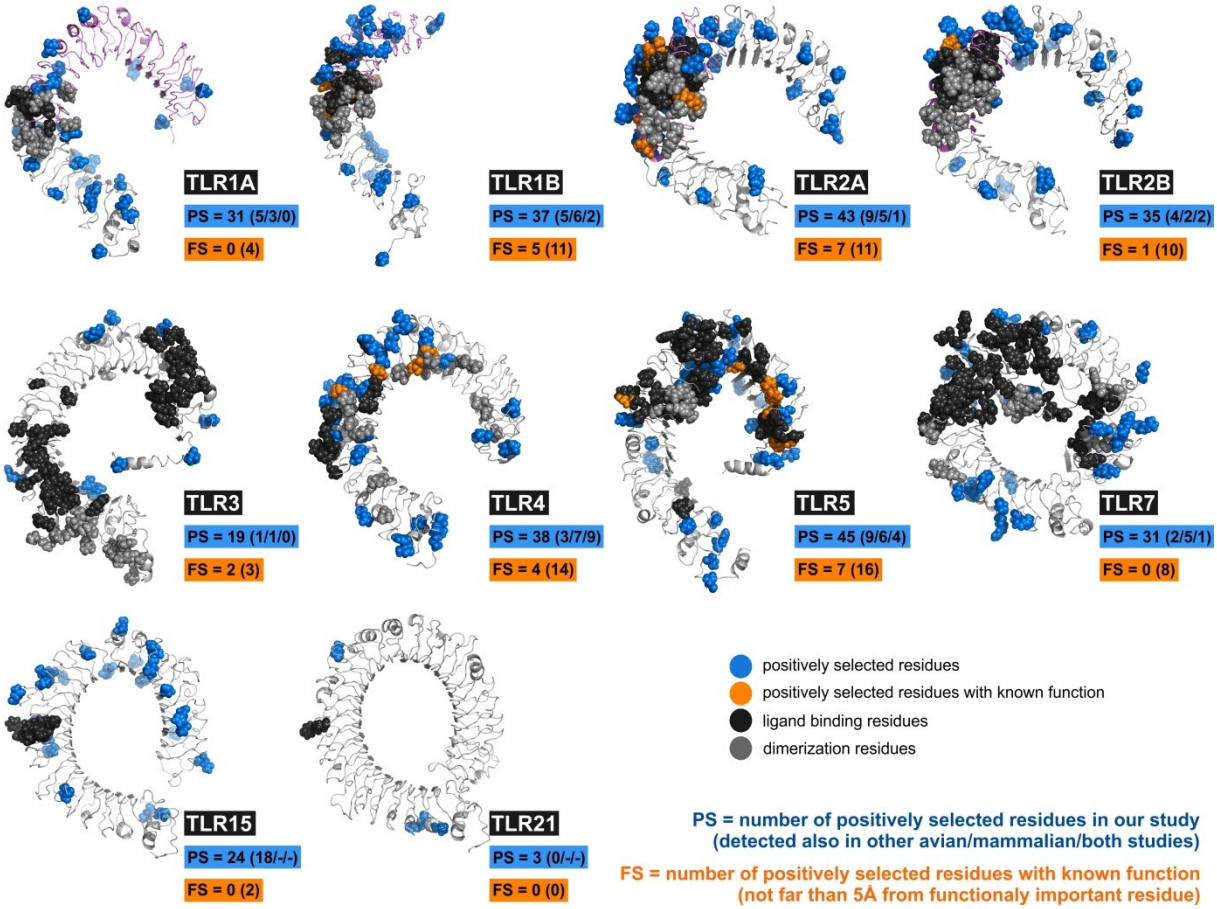


Fig. 13: Positively and functionally important residues visualised on 3D extracellular domain structures of avian TLRs. PSS detected in this study are shown in blue or in orange (sites with known function). Known functionally important residues are highlighted in black (ligand binding residues) or in grey (dimerization residues). The total number of PSS for each TLR is given in blue rectangle, where in parentheses are numbers of PSS detected also in other avian/mammalian studies (for references see Supplementary Material 6: Table S27 in PAPER I.). The numbers of PSS with known function are provided in orange rectangles, where the number of PSS in topological proximity <5 Å is shown in parentheses. The figure is adopted from PAPER I.

Protein evolution

We did not focus only on gene evolution, but also on proteins encoded by the *TLR* genes (PAPER I.; PAPER II.; PAPER III.; PAPER V.). In TLRs, amino acids exhibited on the protein surface play important role in receptor recognition, because there occurs a direct physical contact between selected receptor's residues and pathogenic ligands (see e.g. Manavalan et al. 2011). The host-pathogen coevolution is going on as the micro-evolutionary process acting on molecular level. Especially TLR 3D structure and surface charge might be thus shaped by pathogen-mediated natural selection.

Protein structure evolution

To test the potential selection acting on TLR protein structure, we modelled secondary and tertiary structures of TLR4, TLR5 and TLR7 for all investigated Galloanserae species to find any interspecific structural differences (PAPER II.). However, ***secondary structures*** of selected TLRs revealed only very low levels of variability. Furthermore, even the ECDs did not differ importantly from the intracellular regions in their levels of variability. As expected, all modelled ***3D*** extracellular domains in avian TLRs have the horseshoe-like shape, in which the concave surface comprised β-sheets and the convex surface parallel loops and short helices (PAPER I.; PAPER II.; PAPER III.; PAPER V.), similarly to mammals and other vertebrates. On the contrary, intracellular TLR domain has always a globular shape and transmembrane region is formed by α-helix (both predicted separately). Phenetic analysis of RMSD distances obtained by superposition of the predicted domain structures revealed that not surprisingly there is interspecifically more variability among ECD than among ICD of avian TLRs (PAPER II.). This is probably caused by different selective pressures acting on these two domains, where ICD tends to be more conservative because of their role in cell-signalling contrary to ligand binding ECD (Bell et al. 2003). But more importantly, the RSMD of all our superposed models were lower than the accuracy of individual models (PAPER II.; PAPER III.), so the conclusion is that the structural variability in avian TLRs is generally very low and thus probably unimportant. These findings are supported also by other results in wild rodents (Fornuskova et al. 2013) and seems to be valid for TLRs in general.

Surface electrostatic potential

Since TLRs are structurally almost identical, the question we asked was in which other molecular trait is the variability important for various pathogenic ligands binding and recognition manifested. It is known that even distribution of variability in charge can influence protein conformation and domain composition and thus produce variation in ligand-binding features (Keestra et al. 2008; Walsh et al. 2008; Resman et al. 2009; Meng et al. 2010; Meng et al. 2011). By modelling ***electrostatic potential distribution*** on the TLR protein surfaces we have ascertained that intracellular TIR domain remained interspecifically relatively conserved, contrary to highly variable surface charge distribution in receptor ECD in birds (see Additional file 2: Figure S5 in PAPER II.). The ECD surface charge varied not only among birds, but there is also a great difference between birds and mammals at least in the case of TLR4 (Fig. 14, PAPER II.). Moreover, apparently the highest variability in surface charge is focused into the functionally important regions of receptor ECD (PAPER II.; PAPER III.). From the results is also obvious, that the charge of known ligand binding region in avian TLR4 is more similar to murine TLR4 than to the one of humans (Fig. 14a; PAPER II.), which supports also the previously published results by Keestra and van Putten (2008) on the recombinant TLR4 variant responsiveness pattern to different ligands. In contrast, avian TLR7 electrostatic potential distribution suited more human than mice surface charge. This might indicate the convergent evolution ongoing in given species of birds and mammals resulting in similar ligand binding capabilities of the receptors possessing similar charge on the crucial regions of TLR surface. Contrary to TLR4, interspecific charge variation was much lower in the case of avian TLR5 and TLR7 (Fig. 14; PAPER II.). This is not surprising if we consider that both these receptors show only very limited number of non-conservative positively selected amino acid positions in birds (PAPER I.), which is then mirrored in less variable charge of surface.

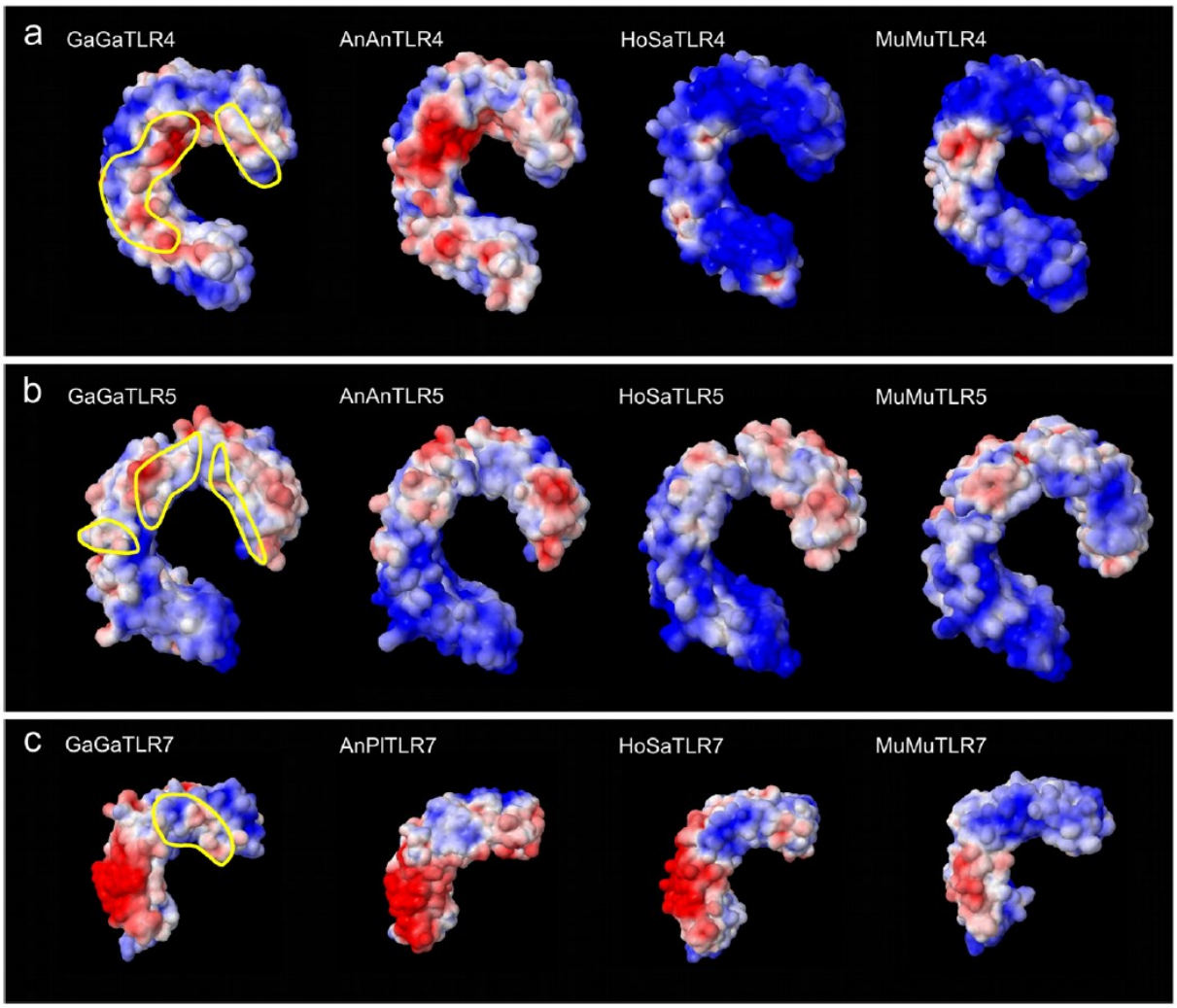


Fig. 14: Differences in ECD surface electrostatic potentials between birds and mammals (for a full comparison see Additional file 2: Figure S5 in PAPER II.). (a) TLR4, (b) TLR5, (c) TLR7; positive surface charge is highlighted in red, negative charge in blue; the predicted functional sites in ECD are outlined in yellow in the GaGaTLR models; GaGa = chicken (*Gallus gallus*), AnAn = goose (*Anser anser*), AnPl = duck (*Anas platyrhynchos*), HoSa = human (*Homo sapiens*), MuMu = mouse (*Mus musculus*). The figure is adopted from PAPER II.

In passerine TLR4 we revealed the pattern of ligand binding surface charge distribution, which enabled us to divide all investigated species into four clearly separated clusters based on PIPSA algorithm (Fig. 15; Supplementary Material: Fig. S4 in PAPER III.). Moreover, there are two key positions (namely 267 and 374) in passerines harbouring non-conservative amino acid substitutions in passerines, based on which we could distinguish four clusters of the TLR4 molecular phenotype (to A+D and B+C, for details see PAPER III.). The functional importance of above-mentioned positions is moreover supported by the fact that both these sites are known to be involved in ligand binding (267; Kim et al. 2007), or lying in the close proximity to previously described ligand binding site (374; Ohto, Fukase, et al. 2012).

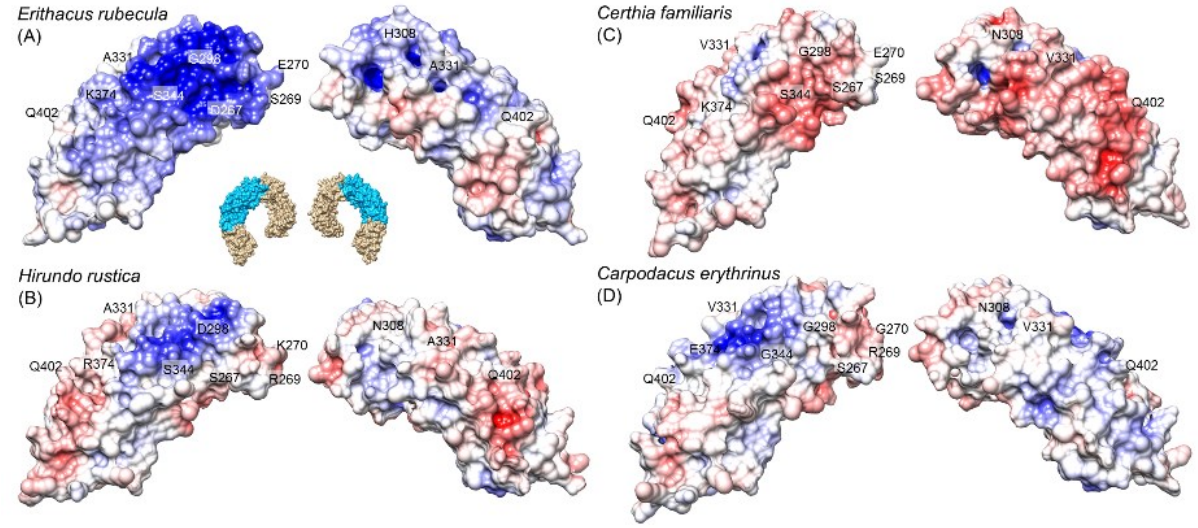


Fig. 15: Variability in the surface charge of the TLR4 LBR in four different passerine species. One species from each charge cluster (Figure S3 in PAPER III.) is displayed – *Erithacus rubecula* (cluster A), *Hirundo rustica* (cluster B), *Certhia familiaris* (cluster C), and *Carpodacus erythrinus* (cluster D). *Certhia familiaris* and *Erithacus rubecula* represent two species with the greatest distance based on the surface electrostatic potential of the whole TLR4 LBR (d = 1.284; Figure S3 in PAPER III.). The inner (left-hand side) and outer (right-hand side) surface of each TLR4 LBR is shown, the ligand-binding surface is represented by the left-hand side view. Positive charge (10) is highlighted in red colour and negative (-10) in blue. Positively selected amino acids confirmed by three independent tests (Table 1 and Table S7 in PAPER III.) are labelled (the site numbering corresponds to *Taeniopygia guttata* protein sequence; positions 272 and 351 are hidden). The position of LBR on the whole ECD is indicated by blue colour in the schematic representation of *Taeniopygia guttata* TLR4. The figure is adopted from PAPER III.

Considering the theory of adaptive evolution, the natural selection mediated by pathogens should play the main role in separation of individual variants of TLR4 that may adapt to optimal recognition of distinct pathogenic challenges. Therefore, we tested (PAPER III.) if there is any relationship between various ***ecological traits*** which characterise the investigated species and the four surface charge clusters of passerine TLR4. As ecological traits we decided to choose commonly used variables such as the migratory strategy, latitudinal dispersion, or diet - all these strategies could be hypothesised to bring different pathogenic threats. In our dataset we have not found any association between TLR4 clusters and latitudinal dispersion, while in the case of the diet such a link cannot be ruled out (Fig. 16). Nevertheless, it seems that phylogeny is the most important factor, which influences the surface charge of TLR4 ligand binding region (Fig. 4 in PAPER III.). Despite this, the limited phylogenetical relatedness of species grouped in cluster D (involving besides passerines also the domestic chicken that was used as an outgroup) suggests convergent adaptive evolution of the TLR4 surface charge.

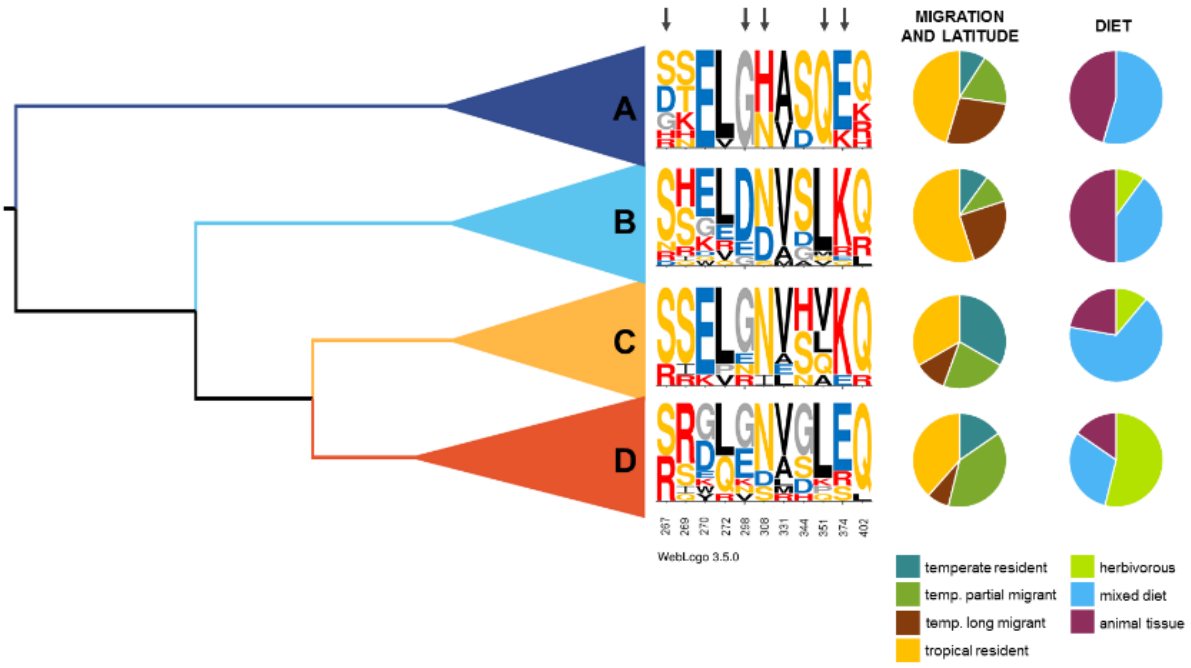


Fig. 16: Projection of ecological characteristics to TLR4 LBR surface charge clustering. The dendrogram showing the surface charge clustering of TLR4 LBR was collapsed into four main clusters (A-D) and for each group we determined the proportion of two basic ecological characteristics: (1) migration and latitude, (2) diet (Table S1 in PAPER III.). The letter diagrams show the proportion of specific amino acids on the 11 positively selected sites with their physicochemical properties indicated by colour (red – positive charge, blue – negative charge, orange – polar uncharged, grey – special cases, black – hydrophobic; Table S8 in PAPER III.). The arrows indicate those amino acid sites that appear to be important for differentiation of individual clusters. The figure is adopted from PAPER III.

Measuring of inflammation

Inflammation belongs to essential mechanisms of innate immunity serving to parasite clearance (Veerdonk et al. 2011). The process is dependent on TLR pathogen detection followed by induction of signalling through proinflammatory cytokines that trigger local leukocyte infiltration linked to phagocytosis and pathogen clearance. Therefore, the ability to appropriately respond to the pathogen recognition is reflected by detectable changes in cytokine expression. Measuring the immune response in ecological-evolutionary studies in free-living non-laboratory birds is very important, but at the same time quite challenging (Millet et al. 2007; Martin et al. 2011). One of the most favourite methods used in birds is the ***phytohaemagglutinin (PHA) skin-swelling test*** (Smits et al. 1999; Kennedy and Nager 2006) because it is fast, efficient and simple, allowing to obtain valuable data in a brief period of time and directly in the field. However, despite a common usage of this skin-swelling test in avian ecological studies, the precise immunological mechanism triggering the swelling response induced by PHA has been previously only a rarely studied (Vinkler et al. 2010).

We decided to test the mechanism regulating induction of the inflammatory response caused by PHA in the grey partridge (*Perdix perdix*) based on measuring the ***expression of nine cytokines*** (interferon: IFN-γ; transforming growth factor: TGF-β2; and interleukins: IL-1β, IL-2, IL-4, IL-6, IL-10, IL-12α, IL-17) and also B2M (β-2-microglobulin subunit of MHC class I) as an internal control with no expected differential expression after the PHA treatment (PAPER VI.). Both the pro-inflammatory and also anti-inflammatory cytokines were differentially expressed in the initial phase of the PHA-induced immune response (6 hours after PHA treatment; shown in Fig. 17). Where three cytokines (mainly IL-6, but

also IL-1β and TGF-β) seems to play a crucial role, because they were substantially upregulated in more of investigated individuals. Moreover, there is a relationship between expressions of these three cytokines. This cytokine composition (generally pro-inflammatory IL-1β and IL-6) indicates the activation of Th17-mediated immunological pathway (Kaiser 2010; Veerdonk et al. 2011). So, the primary PHA-induced swelling is not very probably caused due to the T-cell proliferation and thus adaptive immunity as was previously wrongly thought (Tella et al. 2008), but more likely by innate immune Th17-mediated ***inflammatory response*** as we proved (PAPER VI.). This statement supports also the fact that the cytokine IL-2 which is known to be produced by activated T cells (Lin and Leonard 2003) and to stimulate T-cell proliferation (Lillehoj et al. 2001) was in most of the investigated individual after PHA-treatment down-regulated in its expression (Fig. 17). Contrary to previous two interleukins (IL-1β and IL-6), TGF-β is a commonly known negative regulator of the inflammation (Kaiser 2010). In our PHA experiment we found weak (marginally non-significant, p=0.051) negative association between TGF-β expression and the metrically measured swelling. Based on this we assume that TGF-β negatively regulates also the inflammation underlying the swelling response to PHA, which may result in reduced leukocyte infiltration into the inflamed skin tissue (Vinkler et al. 2012). Although we improved our knowledge about the molecular mechanism behind the PHA skin swelling test, our results show only a limited association between the cytokine signalling and the ***swelling intensity***. The swelling response seems to be probably more dependent on the numbers of cells that infiltrate into the treated tissue (Vinkler et al. 2012), so by measuring the tissue swelling we estimate the responsiveness of such cells to the cytokine signalling and not exactly the differences in the initial cellular signalling triggered in the tissue by the PHA treatment.

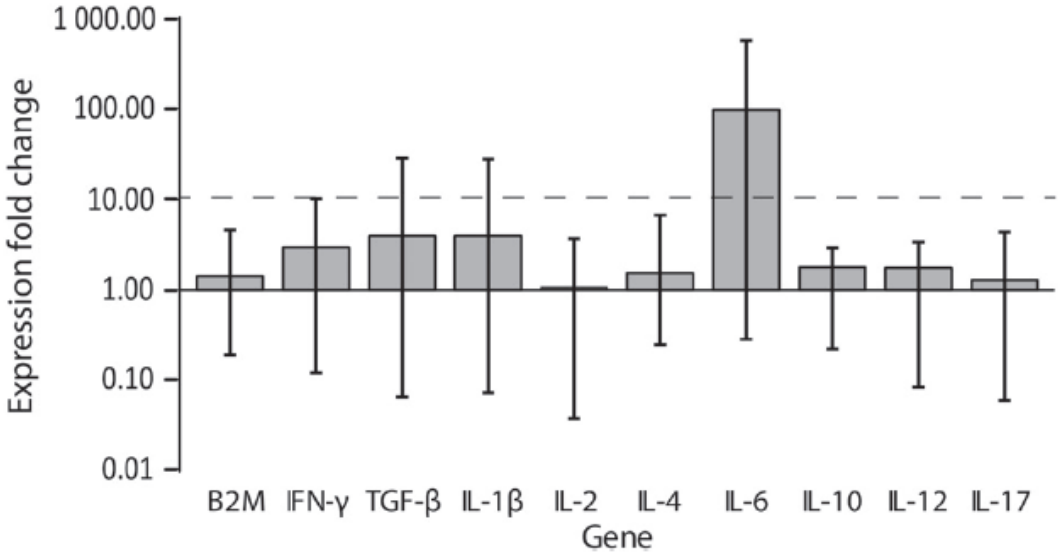


Fig. 17: Cytokine expression after PHA treatment. Fold change of cytokine expression in PHA-treated wings compared to control wings six hours after the PHA treatment (n=10). The fold change was calculated as $2^{-\Delta\Delta Ct}$ (standardised on expression of B2M): up-regulation indicated by values above 1, downregulation indicated by values below 1. We revealed the following fold changes in expression of our cytokine genes (mean fold change, minimum-maximum): IFN-γ 3x (0.4x-10x), TGF-β 4x (0.06x-29x); IL-1β 4x (0.7x-28x), IL-2 1x (0.03x-4x), IL-4 2x (0.2x-7x), IL-6 96x (0.3x-574x), IL-10 2x (0.2x-3x), IL-12α 2x (0.08x-3x), IL-17 1x (0.06x-4x). In the figure, mean is represented by bar, differences between maximum and minimum by whiskers. Dashed line represents methodologically meaningful threshold of expression up-regulation (for rationale see main text in PAPER VI.) and genes with expression above this threshold in at least one individual were included into the statistical analyses. The figure is adopted from PAPER VI.

Conclusions

During my PhD studies I focused primarily on the patterns of evolution in avian Toll-like receptors. I contributed to uncovering the story of gene gain and gene loss in the avian TLR family (Aim 1; PAPER I.; PAPER IV.), to description of the TLR sequence variability (Aim 1, 2; PAPER I.; PAPER II.; PAPER III.; PAPER IV.; PAPER V.) and diversifying selection acting in these genes (Aim 3; PAPER I.; PAPER II.; PAPER III.). Furthermore, I participated in investigation of the receptor phenotypes evolution on protein level (Aim 4; PAPER I.; PAPER II.; PAPER III.; PAPER IV.), in attempt to find out ecological patterns related to TLR variability (Aim 5; PAPER III.) and in improvement of methods to study the immunological processes tightly connected with TLR evolution (Aim 6; PAPER VI.).

As shown by our comprehensive analysis of all members belonging to the TLR family covering all major clades of avian phylogeny, the TLR family is typically composed from ten members in birds. There are, however, exceptions since some of the species possess only nine TLRs, lacking functional TLR5 because of its pseudogenization, while others recently duplicated their TLR7. To both these phenomena we brought gene expression evidence. Interestingly, these ***gene loss and gain events*** occurred in the TLR family several times independently during avian evolution. This suggests that similar selection forces may have acted in different avian clades, leading to adaptive loss or gain of the receptors. The question waiting to be answered is what selective forces could be responsible for these adaptations. Future research should attempt to answer this question by focusing on immune responses in groups of species with and without functional TLR5 and duplicated TLR7. Also searching for shared ecological traits (e.g. presence of similar groups of pathogens, inhabiting similar environments, feeding on the same diet, or exhibiting related life-history traits) may bring light into this investigation. As a first step, however, it is necessary to experimentally test the potential functional effect of TLR5 lacking/TLR7 duplication on the capability of the immune system to correctly recognize their ligands. Further research involving experimental infections will face the problem of the limited knowledge of pathogens present in wild living birds. Nevertheless, before more specific investigation is conducted, the experiments in TLR5 could begin using currently known widespread flagellated bacteria (e.g. *Salmonella typhimurium*, which recombinant purified flagellin is already commercially available). With respect to more ancient duplications that occurred in TLR evolutionary history, we helped to resolve the history of main duplication events in the TLR1 family. Our results show that TLR1 duplication arose even before sauropsids and synapsids split into two separate lineages (giving rise to *TLR1* [avian *TLR1B*] and *TLR10* [avian *TLR1A*]), while in the TLR2 family the duplication probably occurred in parallel to TLR2 duplication in some mammals that resulted in pseudogenisation of the second *TLR2* copy. This finding may help us to better search for functional and evolutionary parallels in vertebrate TLRs.

Generally, in birds, TLRs are ***highly polymorphic molecules***. They exhibit potentially functionally important variability that is shaped by presumably pathogen-mediated selection. On the molecular level the host-pathogen coevolution is acting between the amino acid residues of the host receptors exposed on the surface and the ligand structures. As a result, most of the relevant variability is present at the ligand-binding interface at the extracellular domain. TLRs exposed to the cell surface contain more positively selected sites (mainly in TLR1 [TLR1B], TLR2A, TLR2B, TLR4 and TLR5), than receptors expressed into the endosomes, probably because they bind structurally more diverse ligands. Our results in birds supported the hypothesis proposed previously in mammals that more positively selected sites are present in the three-domain TLRs than in the single-domain ones (except

for avian TLR5 that is under very strong positive selection). We also detected convergent evolution presented in avian TLRs which probably resulted from similar pathogenic selective pressures acting in unrelated host species. Despite the generally high sequence variability detected in avian TLRs on both interspecific and intraspecific levels, we also found exceptions. We revealed only limited *TLR* genetic polymorphism present in a grey partridge population. Apparently, this polymorphism has only a minor functional impact. The low genetic variability of grey partridge *TLRs* is probably caused by a recent population bottleneck and local inbreeding in this species. This may also explain the relatively high susceptibility of the grey partridges to certain infectious diseases, such as putatively e.g. the *Mycoplasma gallisepticum* infection (Vitula et al. 2011).

We may predict that among the ***positively selected sited***, those harbouring the non-conservative amino acid substitutions with potentially dramatic changes in the receptor physicochemical properties, may have the most important functional impact. On interspecific level, we detected most of such non-conservative positively selected sites in two bacteria-sensing avian TLRs (TLR2A and TLR4) with highly variable ligands known in mammals. Interestingly, several potentially functional sites were shared between birds and mammals. However, the real functional effect of each of these positions must be experimentally tested. One way, how to do this, is to measure signalling activation potential of the recombinant receptors (differing in only one target amino acid) expressed in model cells and stimulated with a standard ligand panel under *in vitro* conditions.

Furthermore, we focused on ***TLR protein phenotypic evolution*** that may mirror the imprints of the ongoing or past pathogen-mediated natural selection. Although, we confirmed that avian TLRs are interspecifically very conservative in both their secondary and tertiary structures, we described significant differences in various surface characteristics, mainly in the electrostatic potential. The most important physicochemical changes we found in the ligand binding regions, which functionally differentiates individual groups of avian species. The intracellular part of TLRs responsible for molecular signalling remained interspecifically conserved. Based on our results, it also appears that there might be an ongoing convergent evolution in the TLR surface charges between birds and mammals. In passerine TLR4 we revealed distinctive variation in the patterns of the surface charge distribution, which enabled us to divide all the investigated species into four clearly separated clusters. Unfortunately, we did not succeed in identifying any ecological explanation for these clusters. While partially the clusters reflected the phylogeny, an important part of the variation remained unresolved.

Major disadvantage of studying host-pathogen system in birds is, that we know only a little about wild bird's pathogens. To test the responsiveness of the host's immune system regardless on any specific pathogens, ecological immunology is using various *in vivo* tests. The ***PHA skin-swelling test*** has been frequently adopted by researchers investigating free-living birds. We have designed an improved protocol for this test involving a cytokine mRNA expression profiling in skin. Comparing expression of cytokines involved in distinct regulatory immunological pathways, we were able to show that the test is directed mainly towards Th17 inflammatory immune response, rather than towards adaptive T-cell proliferation, as was previously assumed. This test may be later adopted with specific TLR ligands to describe the relationship between structural variability in TLRs and their capability to trigger an appropriate immune response.

Taken altogether, the results reported in my doctoral thesis indicate strong positive selection driving TLR evolution in birds on both gene and protein levels. However, to better understand the real

significance of this adaptive evolution in avian TLR, further research is highly needed. Above, I have outlined some suggestions on the future way to take. Other possible directions of investigation may involve transcriptome sequencing applicable to revealing multigenic effects of TLR activation. Finally, further research should describe the precise selective pressures shaping the TLR variability and repertoire. To resolve this question the researchers will need to focus more also on pathogens in the wild birds. The current lack of the quality pathogen data may be the reason, why studies investigating the relationship between the TLR variability and susceptibility to various diseases are missing in wild birds. In any case and without doubts, TLRs are crucial molecules of the avian immune system to which we should focus more of our scientific attention.

List of Tables

List of Figures

List of Abbreviations

cDNA - complementary DNA
CDS - coding DNA sequence
CNV - copy number variation
Cp - number of cycles
CpG DNA - cytosine followed by guanine linked by phosphodiester DNA
DAMPs - damage-associated molecular patterns
DOPE - Discrete Optimized Protein Energy
dsRNA - double-strand RNA
ECD - ectodomain / extracellular domain
Eff - efficiency
ESM - electronic supplementary material
FS - forward strain
ICD - intracellular domain
IFN - interferon
IL - interleukin
IVB - The Institute of Vertebrate Biology
LBR - ligand-binding region
LPS - lipopolysaccharide
LRR - leucine-rich repeat
LRT - likelihood ratio test
MD2 - Myeloid differentiation factor-2
MHC - Major histocompatibility complex
mRNA - messenger RNA
Mya - million years ago
NCBI - the National Centre for Biotechnology Information
NGS - next-generation sequencing
NHMO - The DNA Bank of the Natural History Museum of Oslo
NLRC4 - Nucleotide oligomerization domain (NOD)-like receptor C4
PAMPs - pathogen associated molecular patterns
PCR - Polymerase Chain Reactions
PDB - Protein Data Bank
PePe - *Perdix perdix*
PHA - phytohaemagglutinin
PIPSA - Protein Interaction Property Similarity Analysis
PRR - Pattern recognition receptor
PSS - positively selected sites
qPCR - quantitative PCR
RS - reverse strain
TGF - transforming growth factor
TLR - Toll-like receptor
TM - transmembrane
ssRNA - single-stranded RNA
TBP - TATA-box-binding protein
TGF - transforming growth factor
TIR - Toll–interleukin 1 receptor
UMMZ - The Museum of Zoology, University of Michigan
ZCU - Genetic bank of the Department of Zoology, Charles University

References

PAPER I.: Velová H, Gutowska-Ding MW, Burt DW, Vinkler M. 2018. Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection. Mol Biol Evol 35:2170–2184.

PAPER II.: Vinkler M, Bainova H, Bryja J. 2014. Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. Genet Sel Evol 46:72.

PAPER III.: Králová T, Albrecht T, Bryja J, Hořák D, Johnsen A, Lifjeld JT, Novotný M, Sedláček O, Velová H, Vinkler M. 2018. Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an ecological context. Mol Ecol 27:2871–2883.

PAPER IV.: Bainova H, Kralova T, Bryjova A, Albrecht T, Bryja J, Vinkler M. 2014. First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds. Dev Comp Immunol 45:151–155.

PAPER V.: Vinkler M, Bainova H, Bryjova A, Tomasek O, Albrecht T, Bryja J. 2015. Characterisation of Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge. Genetica 143:101–112.

PAPER VI.: Vinkler M, Svobodova J, Gabrielova B, Bainova H, Bryjova A. 2014. Cytokine expression in phytohaemagglutinin-induced skin inflammation in a galliform bird. J Avian Biol 45:43–50.

Acevedo-Whitehouse K, Cunningham AA. 2006. Is MHC enough for understanding wildlife immunogenetics? Trends Ecol Evol 21:433–438.

Akira S, Uematsu S, Takeuchi O. 2006. Pathogen recognition and innate immunity. Cell 124:783–801.

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

Amin A, Bilic I, Liebhart D, Hess M. 2014. Trichomonads in birds – a review. Parasitology 141:733–747.

Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, Aderem A. 2005. Evasion of Toll-like receptor 5 by flagellated bacteria. Proc Natl Acad Sci U S A 102:9247–9252.

Arbour N, Lorenz E, Schutte B, Zabner J, Kline J, Jones M, Frees K, Watt J, Schwartz D. 2000. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat Genet 25:187—191.

Areal H, Abrantes J, Esteves PJ. 2011. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. BMC Evol Biol 11:368.

Artimo P, Jonnalagedda M, Arnold K, Baratin D, Csardi G, de Castro E, Duvaud S, Flegel V, Fortier A, Gasteiger E, et al. 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res 40:W597–W603.

Ashkenazy H, Abadi S, Martz E, Chay O, Mayrose I, Pupko T, Ben-Tal N. 2016. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. Nucleic Acids Res 44:W344–W350.

Ashley NT, Weil ZM, Nelson RJ. 2012. Inflammation: Mechanisms, Costs, and Natural Variation. Annu Rev Ecol Evol Syst 43:385–406.

Atchley WR, Zhao J, Fernandes AD, Drüke T. 2005. Solving the protein sequence metric problem. Proc Natl Acad Sci U S A 102:6395–6400.

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA. 2001. Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci 98:10037–10041.

Bell JK, Mullen GED, Leifer CA, Mazzoni A, Davies DR, Segal DM. 2003. Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends Immunol 24:528–533.

Bodewes R, Kuiken T. 2018. Changing Role of Wild Birds in the Epidemiology of Avian Influenza A Viruses. In: Kielian, M and Mettenleiter, TC and Roossinck, MJ, editor. ADVANCES IN VIRUS RESEARCH, VOL 100. Vol. 100. Advances in Virus Research. p. 279–307.

Boyd A, Philbin VJ, Smith AL. 2007. Conserved and distinct aspects of the avian Toll-like receptor (TLR) system: implications for transmission and control of bird-borne zoonoses. Biochem Soc Trans 35:1504–1507.

Boyd AC, Peroval MY, Hammond JA, Prickett MD, Young JR, Smith AL. 2012. TLR15 is unique to avian and reptilian lineages and recognizes a yeast-derived agonist. J Immunol 189:4930–4938.

Brankston G, Gitterman L, Hirji Z, Lemieux C, Gardam M. 2007. Transmission of influenza A in human beings. Lancet Infect Dis 7:257–265.

Brownlie R, Allan B. 2011. Avian toll-like receptors. Cell Tissue Res 343:121–130.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Todd Hubisz M, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, et al. 2005. Natural selection on protein-coding genes in the human genome. Nature 437:1153–1157.

Bustin S. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25:169–193.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N, Owens SM, Betley J, Fraser L, Bauer M, et al. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6:1621–1624.

Cario E. 2013. The human TLR4 variant D299G mediates inflammation-associated cancer progression in the intestinal epithelium. Oncoimmunology 2:e24890.

Carroll L. 1993. Through the Looking-Glass and What Alice Found There. 2nd ed. HarperCollins

Choe J, Kelker MS, Wilson IA. 2005. Crystal structure of human toll-like receptor 3 (TLR3) ectodomain. Science 309:581–585.

Clayton DH, Moore J. 1997. Host-Parasite Evolution: General Principles and Avian Models. USA: Oxford University Press

Conant GC, Wagner GP, Stadler PF. 2007. Modeling amino acid substitution patterns in orthologous and paralogous genes. Mol Phylogenet Evol 42:298–307.

Cormican P, Lloyd AT, Downing T, Connell SJ, Bradley D, O'Farrelly C. 2009. The avian Toll-Like receptor pathway-Subtle differences amidst general conformity. Dev Comp Immunol 33:967–973.

Coscia MR, Giacomelli S, Oreste U. 2011. Toll-like receptors: an overview from invertebrates to vertebrates. ISJ-Invertebr Surviv J 2011:210–226.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. Genome Res 14:1188–1190.

Cserzo M, Eisenhaber F, Eisenhaber B, Simon I. 2004. TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter. Bioinformatics 20:136–137.

Cunningham AA, Daszak P, Wood JLN. 2017. One Health, emerging infectious diseases and wildlife: two decades of progress? Philos Trans R Soc B Biol Sci 372:20160167.

Danilova N. 2006. The evolution of immune mechanisms. J Exp Zoolog B Mol Dev Evol 306B:496–520.

De Leo GA, Focardi S, Gatto M, Cattadori IM. 2004. The decline of the grey partridge in Europe: comparing demographies in traditional and modern agricultural landscapes. Ecol Model 177:313–335.

DeMarco ML, Woods RJ. 2011. From agonist to antagonist: Structure and dynamics of innate immune glycoprotein MD-2 upon recognition of variably acylated bacterial endotoxins. Mol Immunol 49:124–133.

Dentovskaya SV, Bakhteeva IV, Titareva GM, Shaikhutdinova RZ, Kondakova AN, Bystrova OV, Lindner B, Knirel YA, Anisimov AP. 2008. Structural diversity and endotoxic activity of the lipopolysaccharide of *Yersinia pestis*. Biochem Mosc 73:192–199.

Dhondt AA, Altizer S, Cooch EG, Davis AK, Dobson A, Driscoll MJL, Hartup BK, Hawley DM, Hochachka WM, Hosseini PR, et al. 2005. Dynamics of a novel pathogen in an avian host: Mycoplasmal conjunctivitis in house finches. Acta Trop 94:77–93.

Diebold SS, Massacrier C, Akira S, Paturel C, Morel Y, Reis e Sousa C. 2006. Nucleic acid agonists for Toll-like receptor 7 are defined by the presence of uridine ribonucleotides. Eur J Immunol 36:3256–3267.

Dobson AP, Hudson PJ. 1992. Regulation and stability of a free-living host-parasite system: Trichostrongylus tenuis in red grouse. II. Population models. J Anim Ecol 61:487–498.

Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA. 2004. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. Nucleic Acids Res 32:W665–W667.

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.

Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973.

Eöry L, Gilbert MTP, Li C, Li B, Archibald A, Aken BL, Zhang G, Jarvis E, Flicek P, Burt DW. 2015. Avianbase: a community resource for bird genomics. Genome Biol 16:21.

Ewald SE, Engel A, Lee J, Wang M, Bogyo M, Barton GM. 2011. Nucleic acid recognition by Toll-like receptors is coupled to stepwise processing by cathepsins and asparagine endopeptidase. J Exp Med 208:643–651.

Ewald SE, Lee BL, Lau L, Wickliffe KE, Shi G-P, Chapman HA, Barton GM. 2008. The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. Nature 456:658-U88.

Feduccia A. 1995. Explosive evolution in tertiary birds and mammals. Science 267:637–638.

Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J, et al. 2008. Balancing selection is the main force shaping the evolution of innate immunity genes. J Immunol 181:1315.

Ferwerda B, McCall MBB, Alonso S, Giamarellos-Bourboulis EJ, Mouktaroudi M, Izagirre N, Syafruddin D, Kibiki G, Cristea T, Hijmans A, et al. 2007. TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. Proc Natl Acad Sci 104:16645.

Finn RD, Coggill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A, et al. 2016. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res 44:D279–D285.

Fisher RA. 1922. On the dominance ratio. Proc Roy Soc 56:321–341.

Fornuskova A, Vinkler M, Pages M, Galan M, Jousselin E, Cerqueira F, Morand S, Charbonnel N, Bryja J, Cosson J-F. 2013. Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). BMC Evol Biol 13.

Garcia-Cattaneo A, Gobert F-X, Mueller M, Toscano F, Flores M, Lescure A, Del Nery E, Benaroch P. 2012. Cleavage of Toll-like receptor 3 by cathepsins B and H is essential for signaling. Proc Natl Acad Sci U S A 109:9053–9058.

Gay NJ, Gangloff M. 2007. Structure and function of toll receptors and their Ligands. Annu Rev Biochem 76:141–165.

Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Ben-Tal N. 2003. ConSurf: Identification of Functional Regions in Proteins by Surface-Mapping of Phylogenetic Information. Bioinformatics 19:163–164.

Gopinath VP, Biswas M, Raj GD, Raja A, Kumanan K, Elankumaran S. 2011. Molecular Cloning and Tissue-Specific Expression of Toll-Like Receptor 5 Gene from Turkeys. Avian Dis 55:480–485.

Gordon S. 2002. Pattern Recognition Receptors. Cell 111:927–930.

Grueber CE, Wallis GP, Jamieson IG. 2014. Episodic positive selection in the evolution of avian Toll-like receptor innate immunity genes. Plos One 9:e89632.

Grueber CE, Wallis GP, King TM, Jamieson IG. 2012. Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. Plos One 7:e45011.

Guindon S, Dufayard J-F, 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.

Gupta CL, Akhtar S, Sayyed U, Pathak N, Bajpai P. 2016. In silico analysis of human Toll-like receptor 7 ligand binding domain. Biotechnol Appl Biochem 63:441–450.

Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, et al. 2008. A Phylogenomic Study of Birds Reveals Their Evolutionary History. Science 320:1763–1768.

Hall AJ, Saito EK. 2008. Avian wildlife mortality events due to salmonellosis in the United States, 1985-2004. J Wildl Dis 44:585–593.

Hamilton W, Zuk M. 1982. Heritable true fitness and bright birds: a role for parasites? Science 218:384.

Hawley DM, Osnas EE, Dobson AP, Hochachka WM, Ley DH, Dhondt AA. 2013. Parallel patterns of increased virulence in a recently emerged wildlife pathogen. PLoS Biol 11:e1001570–e1001570.

Hawn TR, Verbon A, Lettinga KD, Zhao LP, Li SS, Laws RJ, Skerrett SJ, Beutler B, Schroeder L, Nachman A, et al. 2003. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to Legionnaires' disease. J Exp Med 198:1563.

Hawn TR, Wu H, Grossman JM, Hahn BH, Tsao BP, Aderem A. 2005. A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. Proc Natl Acad Sci U S A 102:10593–10597.

Hayashi F, Smith K, Ozinsky A, Hawn T, Yi E, Goodlett D, Eng J, Akira S, Underhill D, Aderem A. 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature 410:1099—1103.

Hedrick PW. 2012. What is the evidence for heterozygote advantage selection? Trends Ecol Evol 27:698–704.

Hedrick SM. 2004. The Acquired Immune System. Immunity 21:607–615.

Higuchi M, Matsuo A, Shingai M, Shida K, Ishii A, Funami K, Suzuki Y, Oshiumi H, Matsumoto M, Seya T. 2008. Combinational recognition of bacterial lipoproteins and peptidoglycan by chicken Toll-like receptor 2 subfamily. Dev Comp Immunol 32:147–155.

Hochachka WM, Dhondt AA. 2000. Density-dependent decline of host abundance resulting from a new infectious disease. Proc Natl Acad Sci 97:5303.

Holm L, Laakso LM. 2016. Dali server update. Nucleic Acids Res 44:W351–W355.

Huang Y, Temperley ND, Ren L, 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 11:149.

Iqbal M, Philbin V, Withanage G, Wigley P, Beal R, Goodchild M, Barrow P, McConnell I, Maskell D, Young J, et al. 2005. Identification and functional characterization of chicken Toll-like receptor 5 reveals a fundamental role in the biology of infection with Salmonella enterica serovar typhimurium. Infect Immun 73:2344–2350.

Iwasaki A, Medzhitov R. 2015. Control of adaptive immunity by the innate immune system. Nat Immunol 16:343–353.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346:1320–1331.

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. Nature 491:444–448.

Jin MS, Lee J-O. 2008. Structures of the Toll-like receptor family and its ligand complexes. Immunity 29:182–191.

Kaiser P. 2010. Advances in avian immunology—prospects for disease control: a review. Avian Pathol 39:309–324.

Kaisho T. 2008. Elucidating the mechanism behind immunity using dendritic cells. RIKEN Res 2008:13–16.

Kalmar ID, Dicxk V, Dossche L, Vanrompay D. 2014. Zoonotic infection with *Chlamydia psittaci* at an avian refuge centre. Vet J 199:300–302.

Kang JY, Lee J-O. 2011. Structural Biology of the Toll-Like Receptor Family. In: Kornberg, RD and Raetz, CRH and Rothman, JE and Thorner, JW, editor. ANNUAL REVIEW OF BIOCHEMISTRY, VOL 80. Vol. 80. Annual Review of Biochemistry. p. 917–941.

Kannaki TR, Reddy MR, Shanmugam M, Verma PC, Sharma RP. 2010. Chicken toll-like receptors and their role in immunity. Worlds Poult Sci J 66:727–738.

Kanno A, Yamamoto C, Onji M, Fukui R, Saitoh S, Motoi Y, Shibata T, Matsumoto F, Muta T, Miyake K. 2013. Essential role for Toll-like receptor 7 (TLR7)-unique cysteines in an intramolecular disulfide bond, proteolytic cleavage and RNA sensing. Int Immunol 25:413–422.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 30:772–780.

Kaufman J. 1999. Co-evolving genes in MHC haplotypes: the “rule” for nonmammalian vertebrates? Immunogenetics 50:228–236.

Kawai T, Akira S. 2010. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol 11:373–384.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Keestra AM, van Putten JPM. 2008. Unique properties of the chicken TLR4/MD-2 complex: Selective lipopolysaccharide activation of the MyD88-dependent pathway. J Immunol 181:4354–4362.

Keestra AM, de Zoete MR, van Aubel RAMH, van Putten JPM. 2008. Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. Mol Immunol 45:1298–1307.

Keestra AM, de Zoete MR, Bouwman LI, van Putten JPM. 2010. Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. J Immunol 185:460–467.

Kennedy MW, Nager RG. 2006. The perils and prospects of using phytohaemagglutinin in evolutionary ecology. Trends Ecol Evol 21:653–655.

Kim HM, Park BS, Kim J-I, Kim SE, Lee J, Oh SC, Enkhbayar P, Matsushima N, Lee H, Yoo OJ, et al. 2007. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. Cell 130:906–917.

Klein J, Sato A, Nikolaidis N. 2007. MHC, TSP, and the Origin of Species: From Immunogenetics to Evolutionary Genetics. Annu Rev Genet 41:281–304.

Kosakovsky Pond SL, Frost SDW. 2005a. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol Biol Evol 22:1208–1222.

Kosakovsky Pond SL, Frost SDW. 2005b. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531–2533.

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. Automated Phylogenetic Detection of Recombination Using a Genetic Algorithm. Mol Biol Evol 23:1891–1901.

Kosiol C, Vinař T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A. 2008. Patterns of Positive Selection in Six Mammalian Genomes. PLOS Genet 4:1–17.

Kuijper DPJ, Oosterveld E, Wymenga E. 2009. Decline and potential recovery of the European grey partridge (Perdix perdix) population—a review. Eur J Wildl Res 55:455–463.

de la Lastra JMP, de la Fuente J. 2007. Molecular cloning and characterisation of the griffon vulture (Gyps fulvus) toll-like receptor 1. Dev Comp Immunol 31:511–519.

Lawson B, Robinson RA, Neimanis A, Handeland K, Isomursu M, Agren EO, Hamnes IS, Tyler KM, Chantrey J, Hughes LA, et al. 2011. Evidence of Spread of the Emerging Infectious Disease, Finch Trichomonosis, by Migrating birds. EcoHealth 8:143–153.

Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res 46(D1):D493–D496.

Leveque G, Forgetta V, Morroll S, Smith A, Bumstead N, Barrow P, Loredo-Osti J, Morgan K, Malo D. 2003. Allelic variation in *TLR4* is linked to susceptibility to *Salmonella enterica* serovar *typhimurium* infection in chickens. Infect Immun 71:1116–1124.

Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Lillehoj HS, Min W, Choi KD, Babu US, Burnside J, Miyamoto T, Rosenthal BM, Lillehoj EP. 2001. Molecular, cellular, and functional characterization of chicken cytokines homologous to mammalian IL-15 and IL-2. Vet Immunol Immunopathol 82:229–244.

Lin JX, Leonard WJ. 2003. Interleukin-2. In: Thomson AW, Lotze MT, editors. The cytokine handbook. Vol. 2003. Academic Press. p. 167 – 199.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T)(-Delta Delta C) method. Methods 25:402–408.

Lorenz E, Mira J, Frees KL, Schwartz DA. 2002. Relevance of mutations in the tlr4 receptor in patients with gram-negative septic shock. Arch Intern Med 162:1028–1032.

Lu A, Guindon S. 2014. Performance of Standard and Stochastic Branch-Site Models for Detecting Positive Selection among Coding Sequences. Mol Biol Evol 31:484–495.

Lucas AM, Jamroz C. 1961. Atlas of avian hematology. Washington, D.C.: United States Department of Agriculture

MacDonald MRW, Xia J, Smith AL, Magor KE. 2008. The duck toll like receptor 7: Genomic organization, expression and function. Mol Immunol 45:2055–2061.

Maghrabi AHA, McGuffin LJ. 2017. ModFOLD6: an accurate web server for the global and local quality estimation of 3D protein models. NUCLEIC ACIDS Res 45:W416–W421.

Manavalan B, Basith S, Choi S. 2011. Similar Structures but Different Roles – An Updated Perspective on TLR Structures. Front Physiol 2:41.

Mao F, Leung W-Y, Xin X. 2007. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. BMC Biotechnol 7:76.

Marr N, Novikov A, Hajjar AM, Caroff M, Fernandez RC. 2010. Variability in the Lipooligosaccharide Structure and Endotoxicity among Bordetella pertussis Strains. J Infect Dis 202:1897–1906.

Martin LB, Hawley DM, Ardia DR. 2011. An introduction to ecological immunology. Funct Ecol 25:1–4.

Medvedev AE. 2013. Toll-like receptor polymorphisms, inflammatory and infectious diseases, allergies, and cancer. J Interferon Cytokine Res Off J Int Soc Interferon Cytokine Res 33:467–484.

Medzhitov R. 2008. Origin and physiological roles of inflammation. Nature 454:428.

Medzhitov R, Preston-Hurlburt P, Janeway Jr CA. 1997. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature 388:394–397.

Mellata M. 2013. Human and avian extraintestinal pathogenic *Escherichia coli*: infections, zoonotic risks, and antibiotic resistance trends. Foodborne Pathog Dis 10:916—932.

Meng J, Gong M, Björkbacka H, Golenbock DT. 2011. Genome-wide expression profiling and mutagenesis studies reveal that lipopolysaccharide responsiveness appears to be absolutely dependent on TLR4 and MD-2 expression and is dependent upon intermolecular ionic interactions. J Immunol 187:3683–3693.

Meng J, Lien E, Golenbock DT. 2010. MD-2-mediated ionic interactions between lipid A and TLR4 are essential for receptor activation. J Biol Chem 285:8695–8702.

Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N. 2012. Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains. Gene 503:235–243.

Milinski M. 2006. The Major Histocompatibility Complex, Sexual Selection, and Mate Choice. Annu Rev Ecol Evol Syst 37:159–186.

Millet S, Bennett J, Lee KA, Hau M, Klasing KC. 2007. Quantifying and comparing constitutive immunity across avian species. Dev Comp Immunol 31:188–201.

Mineev KS, Goncharuk SA, Arseniev AS. 2014. Toll-like receptor 3 transmembrane domain is able to perform various homotypic interactions: An NMR structural study. FEBS Lett 588:3802–3807.

Misch EA, Hawn TR. 2008. Toll-like receptor polymorphisms and susceptibility to human disease. Clin Sci 114:347–360.

Miyata T, Yasunaga T. 1980. Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23–36.

Mockenhaupt FP, Cramer JP, Hamann L, Stegemann MS, Eckert J, Oh N-R, Otchwemah RN, Dietz E, Ehrhardt S, Schröder NWJ, et al. 2006. Toll-like receptor (TLR) polymorphisms in African children: Common TLR-4 variants predispose to severe malaria. Proc Natl Acad Sci U S A 103:177.

Murakami Y, Fukui R, Motoi Y, Kanno A, Shibata T, Tanimura N, Saitoh S, Miyake K. 2014. Roles of the cleaved N-terminal TLR3 fragment and cell surface TLR3 in double-stranded RNA sensing. J Immunol 193:5208–5217.

Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Mol Biol Evol 30:1196–1205.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting Individual Sites Subject to Episodic Diversifying Selection. PLOS Genet 8.

Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H, Ishida T, Kimura A. 2008. Natural selection in the TLR-related genes in the course of primate evolution. Immunogenetics 60:727–735.

Neefjes J, Jongsma MLM, Paul P, Bakke O. 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11:823.

Netea MG, Wijmenga C, O'Neill LAJ. 2012. Genetic variation in Toll-like receptors and disease susceptibility. Nat Immunol 13:535–542.

Offord V, Coffey TJ, Werling D. 2010. LRRfinder: A web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database. Dev Comp Immunol 34:1035–1041.

Ohto U, Fukase K, Miyake K, Shimizu T. 2012. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. Proc Natl Acad Sci U S A 109:7421–7426.

Ohto U, Yamakawa N, Akashi-Takamura S, Miyake K, Shimizu T. 2012. Structural analyses of human Toll-like receptor 4 polymorphisms D299G and T399I. J Biol Chem 287:40611–40617.

Papkou A, Gokhale CS, Traulsen A, Schulenburg H. 2016. Host–parasite coevolution: why changing population size matters. SI Host-Parasite Coevol 119:330–338.

Penn DJ, Damjanovich K, Potts WK. 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc Natl Acad Sci 99:11260–11264.

Petersen TN, Brunak S, von Heijne G, Nielsen H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth 8:785–786.

Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera—A visualization system for exploratory research and analysis. J Comput Chem 25:1605–1612.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT–PCR. Nucleic Acids Res 29:e45–e45.

Philbin VJ, Iqbal M, Boyd Y, Goodchild MJ, Beal RK, Bumstead N, Young J, Smith AL. 2005. Identification and characterization of a functional, alternatively spliced Toll-like receptor 7 (TLR7) and genomic disruption of TLR8 in chickens. Immunology 114:507–521.

Pirher N, Ivicak K, Pohar J, Bencina M, Jerala R. 2008. A second binding site for double-stranded RNA in TLR3 and consequences for interferon activation. Nat Struct Mol Biol 15:761–763.

Posada D, Crandall K. 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14:817–818.

Promerová M, Albrecht T, Bryja J. 2009. Extremely high MHC class I variation in a population of a long-distance migrant, the Scarlet Rosefinch (Carpodacus erythrinus). Immunogenetics 61:451–461.

Quach H, Wilson D, Laval G, Patin E, Manry J, Guibert J, Barreiro LB, Nerrienet E, Verschoor E, Gessain A, et al. 2013. Different selective pressures shape the evolution of Toll-like receptors in human and African great ape populations. Hum Mol Genet 22:4829–4840.

Raven N, Lisovski S, Klaassen M, Lo N, Madsen T, Ho SYW, Ujvari B. 2017. Purifying selection and concerted evolution of RNA-sensing toll-like receptors in migratory waders. Infect Genet Evol 53:135–145.

Reperant L, Rimmelzwaan G, Kuiken T. 2009. Avian influenza viruses in mammals. Rev Sci Tech Int Off Epizoot 28:137–159.

Resman N, Vasl J, Oblak A, Pristovsek P, Gioannini TL, Weiss JP, Jerala R. 2009. Essential roles of hydrophobic residues in both MD-2 and toll-like receptor 4 in activation by endotoxin. J Biol Chem 284:15052—15060.

Richter S, Wenzel A, Stein M, Gabdoulline RR, Wade RC. 2008. webPIPSA: a web server for the comparison of protein interaction properties. Nucleic Acids Res 36:W276–W280.

Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The evolution of vertebrate Toll-like receptors. Proc Natl Acad Sci U S A 102:9577–9582.

Robinson RA, Lawson B, Toms MP, Peck KM, Kirkwood JK, Chantrey J, Clatworthy IR, Evans AD, Hughes LA, Hutchinson OC, et al. 2010. Emerging Infectious Disease Leads to Rapid Population Declines of Common British Birds. PLOS ONE 5:e12215.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542.

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat Protoc 5:725–738.

Ruan W, Wu Y, Zheng SJ. 2012. Different genetic patterns in avian Toll-like receptor (TLR)5 genes. Mol Biol Rep 39:3419–3426.

Ruan WK, Wu YH, An J, Zheng SJ. 2012. Polymorphisms of chicken Toll-like receptors 4, 15, and 21 in different breeds. Poult Sci 91:2512–2516.

Rupasree Y, Naushad SM, Rajasekhar L, Uma A, Kutala VK. 2015. Association of TLR4 (D299G, T399I), TLR9-1486T < C, TIRAP S180L and TNF-alpha promoter (-1031,-863,-857) polymorphisms with risk for systemic lupus erythematosus among South Indians. LUPUS 24:50–57.

Sarkar Lab. 2018. http://labs.mmg.pitt.edu/sarkar/Images/TLRs_large.jpg. 2018.

Savage AE, Zamudio KR. 2011. MHC genotypes associate with resistance to a frog-killing fungus. Proc Natl Acad Sci 108:16705.

Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics 27:592–593.

Sedlák K, Literák I, Vitula F, Benák J. 2000. High susceptibility of partridges (*Perdix perdix*) to toxoplasmosis compared with other gallinaceous birds. Avian Pathol 29:563–569.

Sepil I, Lachish S, Sheldon BC. 2012. Mhc-linked survival and lifetime reproductive success in a wild population of great tits. Mol Ecol 22:384–396.

Shen M-Y, Sali A. 2006. Statistical potential for assessment and prediction of protein structures. Protein Sci 15:2507–2524.

Shen T, Xu S, Wang X, Yu W, Zhou K, Yang G. 2012. Adaptive evolution and functional constraint at TLR4 during the secondary aquatic adaptation and diversification of cetaceans. BMC Evol Biol 12:39.

Shen X, Shi R, Zhang H, Li K, Zhao Y, Zhang R. 2010. The Toll-like receptor 4 D299G and T399I polymorphisms are associated with Crohn's disease and ulcerative colitis: a meta-analysis. Digestion 81:69—77.

Shultz AJ, Sackton TB. 2019. Immune genes are hotspots of shared positive selection across birds and mammals.Landry CR, Tautz D, editors. eLife 8:e41815.

Smirnova I, Poltorak A, Chan EKL, McBride C, Beutler B. 2000. Phylogenetic variation and polymorphism at the Toll-like receptor 4 locus (TLR4). Genome Biol 1:research002.1.

Smith SA, Jann OC, Haig D, Russell GC, Werling D, Glass EJ, Emes RD. 2012. Adaptive evolution of Toll-like receptor 5 in domesticated mammals. BMC Evol Biol 12:122.

Smits JE, Bortolotti GR, Tella JL. 1999. Simplifying the phytohaemagglutinin skin-testing technique in studies of avian immunocompetence. Funct Ecol 13:567–572.

Staley M, Bonneaud C. 2015. Immune responses of wild birds to emerging infectious diseases. Parasite Immunol 37:242–254.

Šťastný K, Bejček V, Hudec K. 2010. Atlas hnízdního rozšíření ptáků v České republice. Praha: Aventinum

Stephens M, Smith N, Donnelly P. 2001. A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res 34:W609–W612.

Świderská Z, Šmídová A, Buchtová L, Bryjová A, Fabiánová A, Munclinger P, Vinkler M. 2018. Avian Toll-like receptor allelic diversity far exceeds human polymorphism: an insight from domestic chicken breeds. Sci Rep 8:17878.

Takeuchi O, Akira S. 2010. Pattern Recognition Receptors and Inflammation. Cell 140:805–820.

Tella JL, Lemus JA, Carrete M, Blanco G. 2008. The PHA Test Reflects Acquired T-Cell Mediated Immunocompetence in Birds. PLOS ONE 3:1–9.

Temperley ND, Berlin S, Paton IR, Griffin DK, Burt DW. 2008. Evolution of the chicken Toll-like receptor gene family: A story of gene gain and gene loss. BMC Genomics 9:62.

Thomas NJ, Hunter DB, Atkinson CT eds. 2007. Infectious Diseases of Wild Birds. United States: Blackwell Publishing

Thomson A, Lotze M. 2003. The Cytokine Handbook. 4th ed. Academic Press

Toscano F, Estornes Y, Virard F, Garcia-Cattaneo A, Pierrot A, Vanbervliet B, Bonnin M, Ciancanelli MJ, Zhang S-Y, Funami K, et al. 2013. Cleaved/associated TLR3 represents the primary form of the signaling receptor. J Immunol 190:764–773.

Trowsdale J, Parham P. 2004. Mini-review: Defense strategies and immunity-related genes. Eur J Immunol 34:7–17.

Tsiodras S, Kelesidis T, Kelesidis I, Bauchinger U, Falagas ME. 2008. Human infections associated with wild birds. J Infect 56:83–98.

van Valen L. 1973. A new evolutionary law. Evol Theory 1:1–30.

Veerdonk FL van de, Netea MG, Dinarello CA, Joosten LAB. 2011. Inflammasome activation and IL-1β and IL-18 processing during infection. Trends Immunol 32:110–116.

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.

Vinkler M, Bainová H, Albrecht T. 2010. Functional analysis of the skin-swelling response to phytohaemagglutinin. Funct Ecol 24:1081–1086.

Vinkler M, Bryjova A, Albrecht T, Bryja J. 2009. Identification of the first Toll-like receptor gene in passerine birds: TLR4 orthologue in zebra finch (Taeniopygia guttata). Tissue Antigens 74:32–41.

Vinkler M, Schnitzer J, Munclinger P, Albrecht T. 2012. Phytohaemagglutinin skin-swelling test in scarlet rosefinch males: low-quality birds respond more strongly. Anim Behav 83:17–23.

Vitula F, Peckova L, Bandouchova H, Pohanka M, Novotny L, Jira D, Kral J, Ondracek K, Osickova J, Zendulkova D, et al. 2011. Mycoplasma gallisepticum infection in the grey partridge *Perdix perdix*: outbreak description, histopathology, biochemistry and antioxidant parameters. BMC Vet Res 7:34–34.

Voogdt CGP, Merchant ME, Wagenaar JA and van Putten JPM. 2018. Evolutionary regression and species-specific codon usage of TLR15. Front Immunol 9: 2626.

Waite DW, Taylor MW. 2014. Characterizing the avian gut microbiota: membership, driving influences, and potential function. Front Microbiol 5:223.

Walsh C, Gangloff M, Monie T, Smyth T, Wei B, McKinley TJ, Maskell D, Gay N, Bryant C. 2008. Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa. J Immunol 181:1245–1254.

Wang J, Zhang Z, Chang F, Yin D. 2016. Bioinformatics analysis of the structural and evolutionary characteristics for toll-like receptor 15. PeerJ 4:e2079.

Wang J, Zhang Z, Liu J, Zhao J, Yin D. 2016. Ectodomain architecture affects sequence and functional evolution of vertebrate Toll-like receptors. Sci Rep 6:26705.

Wang X, Quinn PJ, Yan A. 2015. Kdo2-lipid A: structural diversity and impact on immunopharmacology. Biol Rev 90:408–427.

Weaver S, Dube S, Mir A, Qin J, Sun G, Ramakrishnan R, Jones RC, Livak KJ. 2010. Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. Methods 50:271–276.

Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. 2018. Datamonkey 2.0: A Modern Web Application for Characterizing Selective and Other Evolutionary Processes. Mol Biol Evol 35:773–777.

Webb B, Sali A. 2014. Protein Structure Modeling with MODELLER. In: Kihara D, editor. Protein Structure Prediction. New York, NY: Springer New York. p. 1–15.

Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Roessle SC. 2009. Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. Protein Sci 18:1684–1691.

Wlasiuk G, Khan S, Switzer WM, Nachman MW. 2009. A history of recurrent positive selection at the toll-like receptor 5 in primates. Mol Biol Evol 26:937–949.

Wlasiuk G, Nachman MW. 2010. Adaptation and constraint at Toll-like receptors in primates. Mol Biol Evol 27:2172–2186.

Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat Genet 32:569.

Yang J, Zhang E, Liu F, Zhang Y, Zhong M, Li Y, Zhou D, Chen Y, Cao Y, Xiao Y, et al. 2014. Flagellins of Salmonella typhi and nonpathogenic Escherichia coli are differentially recognized through the NLRC4 pathway in macrophages. J Innate Immun 6:47–57.

Yang Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol Biol Evol 24:1586–1591.

Yang Z, Nielsen R, Goldman N, Pedersen A-MK. 2000. Codon-Substitution Models for Heterogeneous Selection Pressure at Amino Acid Sites. Genetics 155:431.

Zareparsi S, Buraczynska M, Branham KEH, Shah S, Eng D, Li M, Pawar H, Yashar BM, Moroi SE, Lichter PR, et al. 2005. Toll-like receptor 4 variant D299G is associated with susceptibility to age-related macular degeneration. Hum Mol Genet 14:1449—1455.

Zdorovenko GM, Zdorovenko EL, Varbanets LD. 2007. Composition, structure, and biological properties of lipopolysaccharides from different strains of *Pseudomonas syringae* pv. *atrofaciens*. Microbiology 76:683–697.

Zhang F, Gao X-D, Wu W-W, Gao Y, Zhang Y-W, Wang S-P. 2013. Polymorphisms in toll-like receptors 2, 4 and 5 are associated with *Legionella pneumophila* infection. Infection 41:941–948.

Zhang G, Jarvis ED, Gilbert MTP. 2014. A flock of genomes. Science 346:1308.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. Science 346:1311.

Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9:40.

Zhang Z, Ohto U, Shibata T, Krayukhina E, Taoka M, Yamauchi Y, Tanji H, Isobe T, Uchiyama S, Miyake K, et al. 2016. Structural analysis reveals that Toll-like receptor 7 is a dual receptor for guanosine and single-stranded RNA. Immunity 45:737–748.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. J Comput Biol 7:203–214.

Zhao Y, Yang J, Shi J, Gong Y-N, Lu Q, Xu H, Liu L, Shao F. 2011. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. Nature 477:596-U257.

de Zoete MR, Bouwman LI, Keestra AM, van Putten JPM. 2011. Cleavage and activation of a Toll-like receptor by microbial proteases. Proc Natl Acad Sci U S A 108:4968–4973.

Zou H, Su R, Ruan J, Shao H, Qian K, Ye J, Yao Y, Nair V, Qin A. 2017. Double-stranded RNA induces chicken T-cell lymphoma apoptosis by TRIF and NF-κB. Sci Rep 7:7547.

Selected publications

PAPER I.

Velová H., Gutowska-Ding M. W., Burt D. W. & Vinkler M. (2018): Toll-like receptors in birds: gene duplication, pseudogenization, and diversifying selection. Molecular Biology and Evolution, 35(9): 2170–2184. (IF_{2018}= 14.797)

This paper provides a comprehensive insight into evolution of avian *TLR* genetic variability. We analysed sequences of all members of the TLR family in 63 bird species covering all major avian clades. In its original state, avian TLRs consists of ten receptors adapted to specifically recognize distinct ligands. We described the story of avian *TLR* gene gain and gene loss. Our results indicate that duplication events in vertebrate TLR1 family have most probably occurred before Synapsids diversified from Sauropsids. Moreover, unlike in mammals, ssRNA-recognising TLR7 has duplicated independently in several avian taxa, while flagellin-sensing TLR5 pseudogenised multiple times in bird phylogeny. Our selection analysis revealed stronger positive diversifying selection acting in TLR5 and three-domain TLRs (TLR1A, TLR1B, TLR2A, TLR2B, TLR4) that face extracellular space and bind more complex ligands than in single-domain TLR15 and others endosomal TLRs (TLR3, TLR7, TLR21). Based on physicochemical properties and their location, we predicted particular amino acid substitutions which might have important effect on receptor function. Moreover, we found the evidence for convergent evolution acting between birds and mammals at many of that positions. They are thus key candidate positions in the receptors that have been very likely shaped by direct molecular host-pathogen coevolutionary interactions and most probably play important functional role in birds.

Toll-Like Receptor Evolution in Birds: Gene Duplication, Pseudogenization, and Diversifying Selection

Hana Velová,*[,1] Maria W. Gutowska-Ding,[2] David W. Burt,[3] and Michal Vinkler[1]

[1]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic

[2]Department of Genomics and Genetics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The Roslin Institute Building, University of Edinburgh, Midlothian, United Kingdom

[3]Office of DVC (Research), University of Queensland, St. Lucia, QLD, Australia

***Corresponding author:** E-mail: bainova@natur.cuni.cz.

Associate editor: Meredith Yeager

Abstract

Toll-like receptors (TLRs) are key sensor molecules in vertebrates triggering initial phases of immune responses to pathogens. The avian TLR family typically consists of ten receptors, each adapted to distinct ligands. To understand the complex evolutionary history of each avian TLR, we analyzed all members of the TLR family in the whole genome assemblies and target sequence data of 63 bird species covering all major avian clades. Our results indicate that gene duplication events most probably occurred in TLR1 before synapsids diversified from sauropsids. Unlike mammals, ssRNA-recognizing TLR7 has duplicated independently in several avian taxa, while flagellin-sensing TLR5 has pseudogenized multiple times in bird phylogeny. Our analysis revealed stronger positive, diversifying selection acting in TLR5 and the three-domain TLRs (TLR10 [TLR1A], TLR1 [TLR1B], TLR2A, TLR2B, TLR4) that face the extracellular space and bind complex ligands than in single-domain TLR15 and endosomal TLRs (TLR3, TLR7, TLR21). In total, 84 out of 306 positively selected sites were predicted to harbor substitutions dramatically changing the amino acid physicochemical properties. Furthermore, 105 positively selected sites were located in the known functionally relevant TLR regions. We found evidence for convergent evolution acting between birds and mammals at 54 of these sites. Our comparative study provides a comprehensive insight into the evolution of avian TLR genetic variability. Besides describing the history of avian TLR gene gain and gene loss, we also identified candidate positions in the receptors that have been likely shaped by direct molecular host–pathogen coevolutionary interactions and most probably play key functional roles in birds.

Key words: **adaptive evolution, amino acid physicochemical properties, convergence, pattern recognition receptors, positive selection, pseudogene.**

Article

Introduction

The species richness and wide range of ecological adaptations are remarkable in birds (Jetz et al. 2012). Despite being similar to mammals in many aspects of their biology, birds evolved many similar traits (including homoiothermy and parental care) convergently through their analogous ecological life strategies (Farmer 2000; Emery and Clayton 2004; Olkowicz et al. 2016; Lovegrove 2017). Recent advance in avian genomic research, which started with the Avian Phylogenomics Project (http://avian.genomics.cn/en/; Zhang, Jarvis, et al. 2014) and continues with the Bird 10,000 Genomes Project (https://b10k.genomics.cn/) has created new possibilities to use publicly available whole genome sequencing data to resolve many questions regarding avian biology, evolution, and adaptations (Zhang, Li, et al. 2014). Among these, the evolution of avian immune function is of particular interest. Being challenged by similar classes of pathogens as found in mammals, highly diversified immune genes in birds are rich in mixed signatures of symplesiomorphy, convergence, and apomorphic adaptations to novel functions (Burri et al. 2010; Cheng et al. 2015).

The coevolution between host and pathogen likely involves mainly molecules that form direct interface between the host and the pathogen structures, for example, the pattern recognition receptors (PRRs) and their ligands, the pathogen-associated molecular patterns (PAMPs; Janeway and Medzhitov 2002). Forming the first line of the host immune defence, PRRs appear to constantly evolve toward specific and appropriate recognition of certain PAMPs (Wang, Zhang, Liu, et al. 2016). The Toll-like receptor (TLR) protein family belongs to one of the most essential and functionally most characterized PRRs (Palm and Medzhitov 2009; Coscia et al. 2011). After specific PAMP binding, TLRs trigger signaling pathways activating transcription factors, such as NF-κB to induce expression of target genes that are key to triggering an inflammatory immune response and subsequent activation of acquired immunity (Iwasaki and Medzhitov 2015).

TLRs are transmembrane proteins each with the characteristic horse-shoe-shaped ectodomain (ECD), where the direct contact between the receptor surface and specific microbe molecules occurs, the transmembrane domain and an intracellular Toll–interleukin 1 receptor (TIR) domain that

enables the downstream signal transmission (Kawai and Akira 2010). The capacity to bind a particular type of ligand in TLRs is dependent on the structure of the ECD. This has been used to divide the vertebrate TLRs into three different groups based on the pattern of hydrogen bonds formed among the asparagine residues in the neighboring leucine-rich repeats, which stabilize the shape of the ECD (Wang, Zhang, Liu, et al. 2016). The ten avian TLR family members belong either to the single-domain TLRs possessing a complete asparagine ladder (TLR3, TLR5, TLR7, TLR15, and TLR21) or to the three-domain TLRs with the ladder interrupted in the central part of the ECD (TLR1A, TLR1B, TLR2A, TLR2B, and TLR4; Wang, Zhang, Liu, et al. 2016). While some avian TLRs conservatively recognize the same ligands as mammalian TLRs (such as TLR4 detecting bacterial lipopolysaccharides, LPS; TLR5 flagellin; TLR3 dsRNA or TLR7 ssRNA; reviewed by Keestra et al. 2013), others were reported to form distinct paralogues with related ligand specificity (e.g., heterodimer-forming TLR1A/TLR1B together with TLR2A/TLR2B, which recognize di/triacylated lipopeptides; Keestra et al. 2007; Higuchi et al. 2008) or achieve recognition of similar ligands as their mammalian analogues through convergence (e.g., avian TLR21 that recognizes CpG DNA similarly as mammalian TLR9; Brownlie et al. 2009; Keestra et al. 2010). Finally, TLR15 that is unique to birds evolved to gain a novel function in recognition of extracellular proteases (de Zoete et al. 2011).

Gene duplication is an important mechanism allowing genes to evolve novel functions (Zhang 2003; Ellegren 2008). Gene duplications are common in TLRs, as with other immune genes. Duplicated TLR1 family members are known in mammals (TLR1, TLR6, TLR10, and TLR2) as well as in birds (TLR1A, TLR1B, TLR2A, and TLR2B; Coscia et al. 2011; Huang et al. 2011; Wang, Zhang, Liu, et al. 2016). Recently, duplication of *TLR7* has been described in several avian taxa (Cormican et al. 2009; Grueber et al. 2012; Raven et al. 2017), while other birds possibly lack the duplicated *TLR7*. This intriguing finding contrasts with that found in mammals, which contain two other TLR7-like family members, *TLR8* and *TLR9*, both missing in birds (Philbin et al. 2005; Temperley et al. 2008), suggesting that gene loss may also play a significant role in avian TLR evolution. *TLR5* pseudogenization occurred independently several times within the passerines (Bainova et al. 2014) and possibly in parrots (Alcaide and Edwards 2011). The general pattern of *TLR* pseudogenization and gene duplication is largely unknown and may be far more common across various *TLR* genes and avian taxa.

Although generally conserved in their structure, TLRs were reported to exhibit very high levels of inter- and intraspecific genetic variation in birds (Alcaide and Edwards 2011; Huang et al. 2011; Grueber et al. 2014). Several authors have documented the functional significance of *TLR* genetic variation (Leveque et al. 2003; Walsh et al. 2008), characterized associations between TLR variation and disease susceptibility (Netea et al. 2012; Medvedev 2013) and suggested the maintenance of intraspecific polymorphism through balancing selection mediated by pathogens (Ferrer-Admetlla et al. 2008). Despite several attempts to understand the evolutionary significance of *TLR* genetic variation in birds (Alcaide and Edwards 2011; Grueber et al. 2014; Vinkler et al. 2014), the question of its adaptive value for avian specific PAMP recognition remains unresolved. Here, we use various approaches to analyze natural selection in the most comprehensive attempt to investigate adaptive pathogen-driven evolution in avian TLRs.

Facilitated by the conservative structures of TLR molecules, they present a suitable model for studying pathogen-driven microevolutionary processes at the DNA level. The direct contact between TLRs and PAMPs is restricted to certain amino acid sites (Gay and Gangloff 2007) where even single amino acid substitutions may have profound effects on receptor binding properties (Keestra et al. 2008; Walsh et al. 2008; Resman et al. 2009; Meng et al. 2010). Thus, positively selected sites can be predicted to emerge from the generally negatively selected background and can be used for prediction of functionally relevant positions in nonmodel animals lacking precise protein crystallographic data. Several studies have investigated the nature of positively selected sites (PSS) at the interspecific level in various TLRs in selected bird taxa (Alcaide and Edwards 2011; Huang et al. 2011; Grueber et al. 2014; Vinkler et al. 2014; Wang, Zhang, Liu, et al. 2016). Yet, colocation study of PSS on the receptor surface with the predicted functional residues already known in the model species (human and mouse), has never been done across the avian phylogeny (but see Vinkler et al. 2014, for a similar approach in Galloanserae birds).

Given the current lack of comprehensive evidence on TLR family molecular evolution across the avian phylogeny, our comparative study reported here performed evolutionary analysis on whole-genome sequences for 48 species representing 34 avian orders (supplementary material S2: table S23, Supplementary Material online) mainly gained through the Avian Phylogenomics Project (Zhang, Jarvis, et al. 2014). The TLR sequences extracted from these genomes together with those added from other public resources (in total 63 species) allowed us to infer the history of avian *TLR* gene gain and gene loss in the context of TLR evolution in other vertebrates. By critical assessment of positive selection acting on all members of the TLR family in birds, we describe the adaptive microevolutionary changes in these immune receptors in their molecular context. Thus, this study is the first to comprehensively predict functionally relevant genetic variation in avian TLRs providing insights into the coadaptation between host and pathogen through evolution of ligand recognition.

Results and Discussion

Evolution of *TLR1* Family Gene Duplications

The *TLR1* gene family is widely duplicated in vertebrates including birds (*TLR1A*, *TLR1B*, *TLR2A*, and *TLR2B*; Temperley et al. 2008; Cormican et al. 2009; Huang et al. 2011; Wang, Zhang, Liu, et al. 2016). Presently, however, the timing of these gene duplication events is unclear: the paralogues within the *TLR1* and *TLR2* subfamily were suggested to have duplicated before the mammalian divergence from sauropsids (Huang et al. 2011) or after this divergence, independently in birds and mammals (Temperley et al. 2008; Cormican et al. 2009;

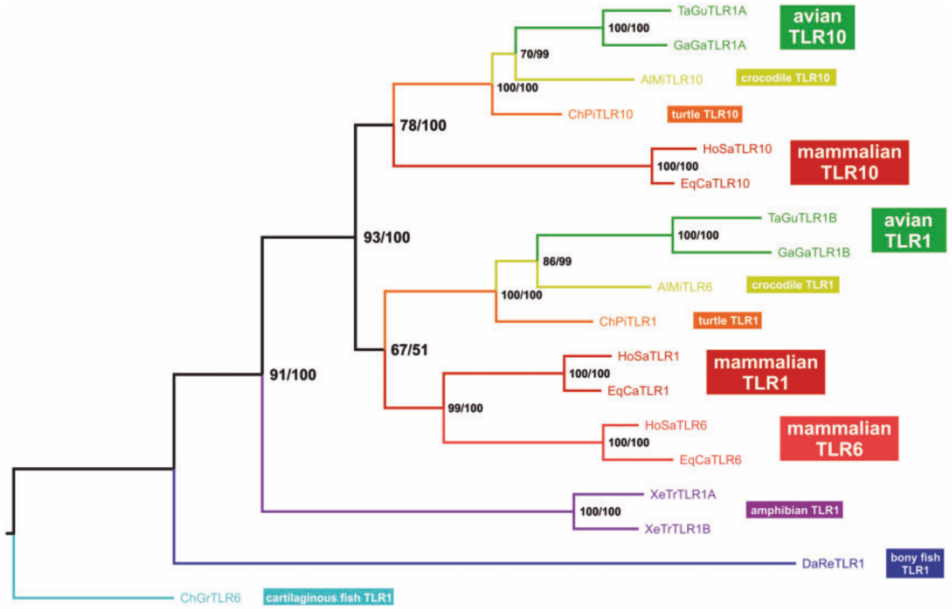


Fig. 1. Phylogenetic tree based on nonconverted regions of *TLR1* subfamily members. The bootstrap values of maximum likelihood analysis obtained using PhyML and the posterior probability of Bayesian analysis obtained using MrBayes (in percentage per each node) are provided. Birds are represented by zebra finch (TaGu, *Taeniopygia guttata*) and chicken (GaGa, *Gallus gallus*), crocodiles by alligator (AlMi, *Alligator mississippiensis*), turtles by painted turtle (ChPi, *Chrysemys picta*), mammals by human (HoSa, *Homo sapiens*) and horse (EqCa, *Equus caballus*), amphibians by clawed frog (XeTr, *Xenopus tropicalis*), bony fish by zebrafish (DaRe, *Danio rerio*) and cartilaginous fish by shark (ChGr, *Chiloscyllium griseum*). The analysis was performed using a single amino acid sequence per TLR and species. Based on the results, we suggest renaming TLR1A to TLR10 and TLR1B to TLR1 in birds.

Mikami et al. 2012; Wang, Zhang, Liu, et al. 2016). This disagreement results from the lack of accounting for gene conversion as a mechanism of paralogue sequence homogenization in some studies (Temperley et al. 2008; Cormican et al. 2009; Wang, Zhang, Liu, et al. 2016). We found significant support for gene conversion between the two paralogues both in the *TLR1* subfamily (*TLR1A* and *TLR1B*) and *TLR2* subfamily (*TLR2A* and *TLR2B*; supplementary material S1: table S1, Supplementary Material online; for detailed info see supplementary material S3, Supplementary Material online) in birds. Therefore, we constructed both *TLR1* and *TLR2* phylogenetic trees based on the protein sequences of the nonconverted regions only (fig. 1; supplementary material S1: figs. S1 and S2, Supplementary Material online).

Our results based on the analysis of 179 sequences of 83 vertebrate species (for the list see supplementary material S2: table S24, Supplementary Material online) show that in the *TLR1* family the sequences of avian species cluster together based on paralogue identity (supplementary material S1: fig. S1, Supplementary Material online). Avian *TLR1A* and reptilian *TLR10* cluster together with mammalian *TLR10*, while avian *TLR1B* and reptilian *TLR1*/*TLR6* (inconsistent nomenclature in reptilian *TLR1*) cluster together with mammalian *TLR1* and *TLR6* paralogues (shown in fig. 1; detailed in supplementary material S1: fig. S1, Supplementary Material online). This is consistent with the results previously published by Huang et al. (2011) confirming that the first duplication event within the *TLR1* subfamily occurred before mammal–reptile divergence. On the basis of these findings, we suggest renaming avian *TLR1A* to *TLR10* (from now on marked as *TLR10* [*TLR1A*] in the text), while modifying the name of the avian *TLR1B* gene to *TLR1* (marked as *TLR1* [*TLR1B*]). Our analysis also indicates that the duplication of *TLR1* in amphibians (namely in *Xenopus*) apparently arose independently of the *TLR1* duplication in amniotes (supplementary material S1: fig. S1, Supplementary Material online).

In addition, in the *TLR2* subfamily there are two copies present in birds (*TLR2A* and *TLR2B*). In mammalian genomes, on the other hand, only one functional *TLR2* is maintained and for some species *TLR2*-like pseudogenes (a second copy) were described (Roach et al. 2005). The duplication of *TLR2* in sauropsids could have arisen independently of mammals, or alternatively, a duplication event predating divergence between birds and mammals could have occurred (similar to the *TLR1* subfamily), followed with pseudogenization leaving only a single *TLR2* copy in mammals (Huang et al. 2011). Our analysis of 137 homologous sequences in 78 vertebrate species (for the list see supplementary material S2: table S24, Supplementary Material online) contradicts the second scenario (supplementary material S1: fig. S2, Supplementary Material online) and supports the independent duplication of *TLR2* in sauropsids. This is also supported by the phylogenetic analysis of a ca. 80 amino-acid-long alignment of the conserved sequence in human *TLR2*/*TLR2*-like pseudogene

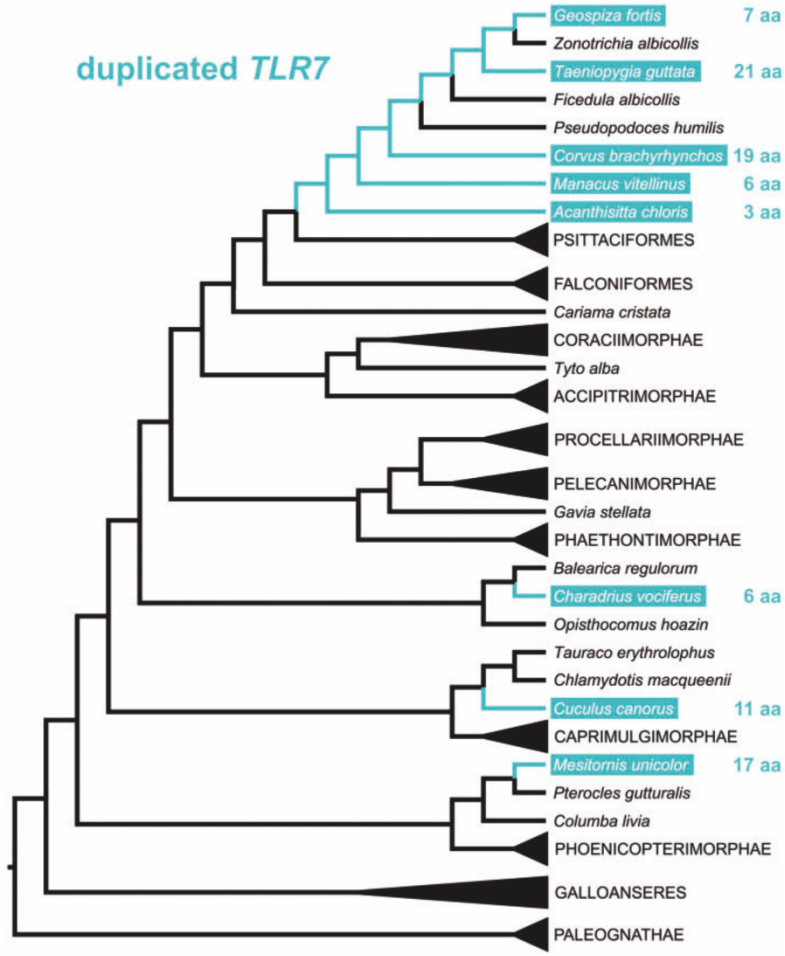


Fig. 2. Avian *TLR7* duplication. In the schematic avian phylogenetic tree the species with duplicated *TLR7* are highlighted in teal rectangles. The numbers of amino acid substitutions distinguishing the two copies of the duplicated *TLR7* are shown behind the species name. The analysis was performed using a single sequence per gene and species.

and chicken TLR2A/TLR2B (supplementary material S1: fig. S3, Supplementary Material online); the resulting phylogenetic tree suggests that *TLR2* duplication in birds was independent of a parallel mammalian *TLR2* duplication that was followed by a pseudogenization of the second *TLR2* copy (supplementary material S1: fig. S4, Supplementary Material online). However, it is difficult to draw any final conclusions on the validity of the two scenarios proposed for *TLR2* evolution in amniotes since the support for the branching of the avian *TLR2* paralogues is limited (bootstrap support for the mammal-sauropsid split = 47, supplementary material S1: fig. S2, Supplementary Material online).

Avian *TLR7* Gene Duplications

Birds lack two of the three vertebrate *TLR7* family members (i.e., *TLR8* and *TLR9*; Philbin et al. 2005; Temperley et al. 2008). Instead, the avian *TLR7* locus appears to have recently duplicated in some passerines (Cormican et al. 2009; Grueber et al. 2012) and waders (Raven et al. 2017). Within the genome sequences investigated in this study, we also found two *TLR7* copies in other avian taxa: Cuculiformes and Mesiornithiformes (fig. 2). The number of amino acid substitutions distinguishing the two *TLR7* loci ranges between species from 21 in zebra finch (*Taeniopygia guttata*) to only 3 in rifleman (*Acanthisitta chloris*; see fig. 2; supplementary material S1: table S2 and supplementary material S4, Supplementary Material online). Furthermore, the two *TLR7* copies likely differ slightly in their functions, because 11 of the sites identified as distinct between the two copies in any avian species match the known ligand-binding positions (Wei et al. 2009; Gupta et al. 2016; Zhang et al. 2016), five are situated in the Z-loop region responsible for ligand binding and dimerization and the other 10 variable sites are identical with PSS detected in birds (supplementary material S1: table S3, Supplementary Material online). To bring independent support to the evidence of *TLR7* gene duplication, we

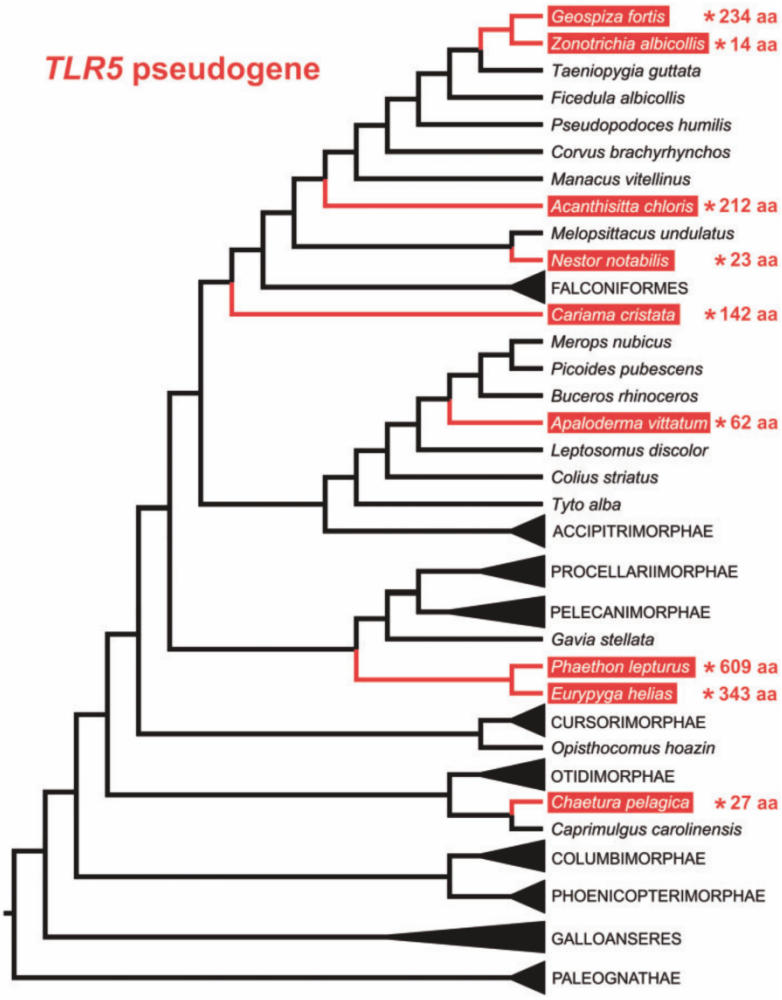


Fig. 3. *TLR5* pseudogenization in birds. Species possessing only *TLR5* pseudogene are highlighted in red rectangles within the schematic representation of avian phylogenetic tree. The position of the first stop codon is indicated by the number provided behind the species name (position numbering according to the chicken reference). The analysis was performed using a single sequence per gene and species.

performed a PCR-based quantitative copy number variation (CNV) analysis comparing *TLR7* with two single-copy *TLRs* (*TLR3* and *TLR4*) in selected species (collared flycatcher, *Ficedula albicollis*; ground tit, *Pseudopodoces humilis*; white-throated sparrow, *Zonotrichia albicollis* and zebra finch). Our results confirm that the zebra finch (a species with two *TLR7* loci based on the genome sequence analysis) possesses two *TLR7* copies (supplementary material S1: tables S4–S6 and fig. S5, Supplementary Material online), while the other investigated species (with a predicted single *TLR7* locus) possess only one *TLR7* gene copy. In contrast to the *TLR1* family, we did not detect any gene conversion between the two copies of the duplicated *TLR7* genes (see supplementary material S3, Supplementary Material online). Despite this, the constructed tree of avian TLR7 does not show any evidence of two separate TLR7 phylogenetic clusters, but surprisingly both copies of *TLR7* always clustered separately for each species (supplementary material S1: fig. S6 and table S7, Supplementary Material online). Therefore, we infer that the duplication in *TLR7* occurred several times independently in recent avian history and possibly even multiple times in some cases (as, e.g., in ruddy turnstone, *Arenaria interpres*, where *TLR7* was reported to be triplicated; Raven et al. 2017).

TLR5 Pseudogenization in Birds

Recently, it has been shown that there is no functional gene for TLR5 in some passerine species (Bainova et al. 2014). In this study, we have also found several avian taxa possessing only nonfunctional *TLR5* pseudogenes with stop-codons in their sequences. Apart from passerines, we revealed *TLR5* pseudogenes in other clades: that is, Psittaciformes, Cariamiformes, Trogoniformes, Phaethontiformes, Eurypygiformes, and Apodiformes (fig. 3). The sequence data show that the stop-codon positions differ among the

investigated species with differences also in the pseudogenization mechanism (single nucleotide substitution or frameshifting indels, see supplementary material S1: table S8, Supplementary Material online). The stop-codons in *TLR5*, thus, arose independently in the evolutionary history of the different avian taxa. This, however, does not appear to be the result of relaxed selection acting in *TLR5*. Firstly, the number of single nucleotide variants (SNVs) in the alignment of the *TLR5* functional sequences is comparable to other *TLRs* (indicating comparable mutation rate; supplementary material S1: table S9, Supplementary Material online). Secondly, the total number of PSS is higher than in other TLRs and ω (dN/dS ratio) is again similar to other TLRs (see below and supplementary material S1: table S10, Supplementary Material online) and higher than the genome average (Zhang, Li, et al. 2014). However, since most of the PSS contain only conservative substitutions without dramatic effects on the amino acid site physicochemical properties, it appears that negative selection may prevent the loss of function in those species where the *TLR5* functional gene is maintained. Altogether this suggests that there may be a selection for TLR5 loss of function in certain avian evolutionary lineages. The loss of a *TLR5* functional gene is not limited only to birds. The *TLR5* pseudogene has also been described in one human allele, where its presence is associated with an increased risk of pneumonia infection caused by flagellated bacteria *Legionella pneumophila* (Hawn et al. 2003; Zhang et al. 2013), while at the same time possessing a *TLR5* nonfunctional allele might be advantageous for decreasing the probability of autoimmune disease development (Hawn et al. 2005). *TLR5* pseudogenization may be allowed by the high redundancy of pathogen detection, where apart from TLR5, flagellin is also recognized by other PRRs, for example, NLRC4 inflammasome (Zhao et al. 2011; Yang et al. 2014). In birds other flagellin-recognizing PRRs have not yet sufficiently been studied to support this hypothesis.

Diversifying Selection in Avian TLRs

We tested pervasive, positive, and diversifying selection acting in all avian TLRs. Numbers of PSS differ between *TLRs* (from 0.3% to 5.8% PSS per TLR; table 1; supplementary material S1: fig. S7 and supplementary material S5: table S26, Supplementary Material online). The results suggest that positive selection is acting more on TLRs exposed toward the cell surface (mainly TLR1 [TLR1B], TLR2A, TLR5, TLR4, and TLR2B), than in the TLRs situated in endosomes (TLR21, TLR3, and TLR7). This may be because the endosomal TLRs are specialized for detection of less complex ligands, which show low structural variation (such as ssRNA in TLR7, dsRNA in TLR3 or CpG DNA regions in TLR21; Brownlie and Allan 2011) interacting with the genetic variability in TLRs. This supports the previous findings by Mikami et al. (2012) in vertebrates. The only exception is TLR15, which is unique to birds and reptiles (Boyd et al. 2012). Although situated on the cell surface, TLR15 is activated (unlike other TLRs) by the ECD proteolytic cleavage with pathogen-derived proteases (de Zoete et al. 2011). Since *TLR15* harbors only low numbers of PSS (table 1), we hypothesize that this pathogen-recognition mechanism may be linked to reduced positive selection acting on the ECD. Nonetheless, the cleavage site is variable across bird taxa and contains two PSS (supplementary material S1: table S11, Supplementary Material online; see also supplementary material S1: text S1 and fig. S8, Supplementary Material online).

Table 1. The Number of Positively Selected Sites in Avian TLRs.

	Species[a]	aa length[b]	PSS[c]	PSS/TLR[d] (%)
TLR10 [TLR1A]	45	794	31	3.9
TLR1 [TLR1B]	44	638	37	5.8
TLR2A	40	793	43	5.4
TLR2B	42	781	35	4.5
TLR3	51	895	19	2.1
TLR4	54	843	38	4.5
TLR5	46	861	45	5.2
TLR7	51	1041	31	3.0
TLR15	53	868	24	2.8
TLR21	14	907	3	0.3

[a]Number of species (one sequence per species).
[b]The protein sequence length in reference chicken TLRs (NCBI accession numbers are listed in supplementary material S1: table S14, Supplementary Material online).
[c]Number of positively selected sites detected in investigated species per each gene.
[d]The percentage of PSS per whole receptor amino acid sequence.

Our results are also mostly consistent with the findings of Wang, Zhang, Liu, et al. (2016), who suggested grouping of TLRs based on their ECD architecture and showed that in mammals the single-domain TLRs (TLR3, TLR5, TLR7, TLR15, and TLR21) are under stronger purifying selection than the three-domain TLRs (TLR10 [TLR1A], TLR1 [TLR1B], TLR2A, TLR2B, and TLR4). Here, we found that also in birds the positive selection is acting more on the three-domain TLRs, with the exception of TLR5, which is under strong positive selection in birds (table 1 and supplementary material S1: fig. S7, Supplementary Material online). Since some avian species have lost a functional TLR5 through pseudogenization (Bainova et al. 2014; and this study, fig. 3), we hypothesize that TLR5 has a specific role in avian immunity when compared, for example, to mammals. The variation in selection acting at TLR5 in different taxa might, for example, reflect the differences in selection against overactivation of gut immunity with flagellated symbiotic microbiota (Iqbal et al. 2005).

Being responsible for direct and specific recognition of structurally heterogeneous PAMPs (Reddick and Alto 2014), the pathogen-mediated selective pressures are particularly diversified in the TLR ECDs. As previously shown in general for vertebrates (Mikami et al. 2012), we also find the majority of TLR PSS in birds to be situated in ligand-binding ECDs (supplementary material S1: fig. S7, Supplementary Material online; the exact positions of PSS are visualized in supplementary material S1: fig. S9, Supplementary Material online and listed in supplementary material S6: table S27, Supplementary Material online). We may predict stronger functional effects of PSS with nonconservative amino acid substitutions that change the physicochemical properties of the particular residues. The highest numbers of these nonconservative PSS are in TLR4 and TLR2A (fig. 4; see also supplementary material S1: table S12 and supplementary materials S7 and S8, Supplementary Material online). Especially in TLR4, most

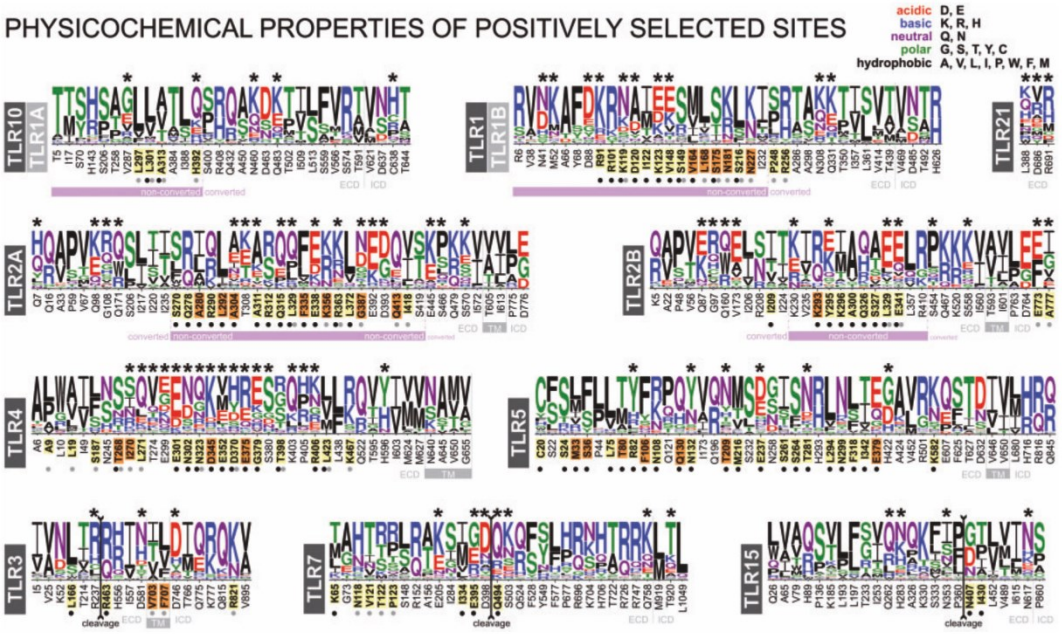


FIG. 4. Physicochemical properties of the positively selected sites (PSS). All PSS are shown in all avian TLRs—amino acid substitutions are colored according to their physicochemical properties: acidic in red, basic in blue, neutral in purple, polar in green, and hydrophobic in black. The size of an letter corresponds to the procentual proportion of that particular amino acid within the sequence alignment. The numbering is adopted from reference chicken TLRs (for NCBI IDs see supplementary material S1: table S14, Supplementary Material online). PSS which correspond to functionaly important residues (black dot—ligand binding; gray dot—dimerization) are highlighted in bold and orange (identical site) or yellow (topological proximity closer than 5 Å from a functionaly important residue). Ectodomain (ECD), intracellular domain (ICD), and transmembrane (TM) region are visualized; nonconverted region in TLR1/2 is highlighted in pink; cleavage site in TLR3, TLR7, and TLR15 is indicated by a black line tipped with arrows (supplementary material S1: table S11 and text S1, Supplementary Material online). Nonconservative PSS are marked by stars (supplementary material S1: table S12, Supplementary Material online).

PSS with dramatic changes in charge and hydrophobicity fall into the known functionally important sites of the receptor, with high probability of changing the TLR4 ligand-binding properties. On the other hand, most substitutions in PSS in TLR15, TLR3, and TLR10 [TLR1A] do not significantly change the amino acid physicochemical properties. Despite avian TLR1 [TLR1B] being significantly shorter than TLR10 [TLR1A], we found more PSS in TLR1 [TLR1B] and these were less conservative in their charge than in TLR10 [TLR1A], especially in the predicted ligand-binding region that is spanning the region avoiding gene conversion (supplementary material S1: text S2 and supplementary material S7, Supplementary Material online). This, together with high TLR2A PSS variation, can be possibly explained by distinct binding capacities of the heterodimers formed by the duplicated TLR1/TLR2 subfamily members, where the combination TLR1 [TLR1B]/TLR2A (unlike any other combination) is able to recognize peptidoglycans and efficiently recognizes diacylated bacterial lipoproteins (BLP; Higuchi et al. 2008). Surprisingly, despite its high number of PSS, TLR5 is the most PSS-conservative TLR across all avian taxa (supplementary material S1: table S12, Supplementary Material online). Given that flagellin is the only known TLR5 ligand (Hayashi et al. 2001), this may indicate functional constraints in those species with preserved functional TLR5, limiting any adaptations to larger numbers of relatively minor changes. The PSS identified in this study are consistent with PSS identified by other studies and other taxa especially in TLR15, TLR4, and TLR5 (PSS listed in supplementary material S6: table S27, Supplementary Material online and those located in ECD visualized in fig. 5; visualization of ECDs along with intracellular domains, ICDs, and transmembrane, TM, regions is given in supplementary material S9, Supplementary Material online). Interestingly, from the five amino acid substitutions related to *Salmonella enterica* resistance variation in chickens (Leveque et al. 2003) only position E301 is also positively selected within birds in general, where negatively charged E/D are substituted with positively charged K/R (fig. 4). The relevance of this position was also supported in other studies of birds (Grueber et al. 2014) and mammals (Vinkler et al. 2009; Wlasiuk and Nachman 2010) and partially with the results of numerous human studies (Arbour et al. 2000; Zareparsi et al. 2005; Shen et al. 2010; Cario 2013; Rupasree et al. 2015; but see also Ohto, Yamakawa, et al. 2012) showing potential importance of variation at a neighboring position D299G (D303 in chicken sequence).

We found the highest agreement between the PSS and known functional site distribution in TLR5, TLR4, TLR2A, and TLR1 [TLR1B] (23, 18, 18, and 16 positions, respectively; for details see supplementary material S1: table S13 and fig. S10,

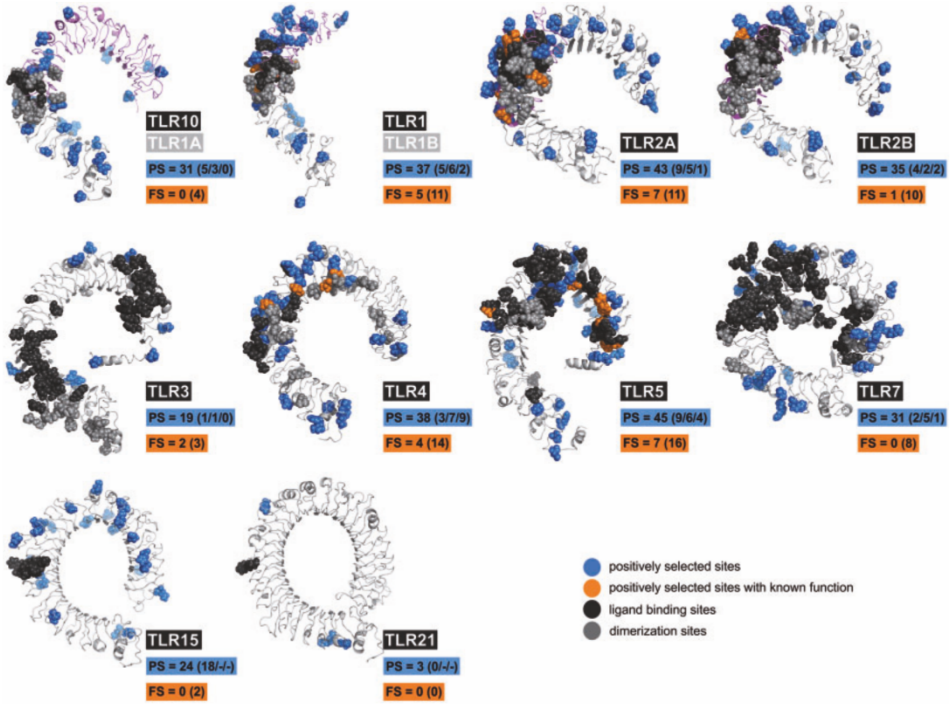


Fig. 5. Positively selected sites and functionaly important sites visualized on 3D extracellular domain structures of avian TLRs. PSS detected in this study are shown in blue. By orange coloration are highlighted the PSS identified at sites with previously described function. Other previously reported functionaly important residues are highlighted in black (ligand binding residues) or in gray (dimerization residues). The total numbers of PSS for each TLR are shown in blue rectangles, where in parentheses are numbers of PSS detected also in other avian/mammalian/both studies (for references see supplementary material S6: table S27, Supplementary Material online). The numbers of PSS with previously described function are provided in orange rectangles, where the number of PSS in topological proximity <5 Å is shown in parentheses. For more detailed information (including ICD and TM), see supplementary material S9, Supplementary Material online.

Supplementary Material online). We show that the PSS frequently evolve toward an amino acid of similar physicochemical properties being gained in distantly related avian taxa, while distinct properties can be found in closely related taxa (supplementary material S1: text S2 and supplementary material S8, Supplementary Material online). This convergent evolution may result from analogous selective pressures induced by partially similar microbial communities in different avian taxa (Waite and Taylor 2014). For example, there are several PSS in avian TLR4 (D345, E375, and G379) that are good candidates for functionally important sites evolving under convergence either toward negative or toward positive charge (supplementary material S8, Supplementary Material online). Being either lipid IVa-binding sites (Meng et al. 2010; Ohto, Fukase, et al. 2012; Scior et al. 2013) or LPS-binding sites (Park et al. 2009; Garate and Oostenbrink 2013; Paramo et al. 2013; Scior et al. 2013), these positions have been identified as positively selected also in several other studies in birds (Grueber et al. 2014) and mammals (Nakajima et al. 2008; Vinkler et al. 2009; Wlasiuk and Nachman 2010; Areal et al. 2011; Shen et al. 2012; Fornůsková et al. 2013).

The lack of agreement between the identified PSS and previously described TLR functional sites in other TLRs may be caused either by the lack of functional studies in some TLRs (e.g., TLR15 and TLR21) or by interspecific variation in TLR-ligand binding (e.g., TLR1, TLR3, and TLR7, where the structural and experimental studies were performed in mammals and not in birds; see supplementary material S1: text S2, Supplementary Material online). In both cases the analysis of PSS may indicate the target sites that are candidates for a functional role and may be the focus for further research. Despite the low number of PSS among the previously reported functional sites of TLR7, which may result from the constraints of conservative ligand binding, two nonconservative PSS located on the TLR7 Z-loop at the proximity of the TLR7 cleavage site (supplementary material S1: text S2, Supplementary Material online; Ewald et al. 2008; Kanno et al. 2013) might affect formation of secondary ligand-binding sites (Zhang et al. 2016).

Although most of the PSS functional sites are located in the ECD and involved in direct ligand binding, in TLR3 we found two PSS that are lying in the transmembrane region and are located in close proximity to the known receptor dimerization residues (fig. 4). Since TLR3 dimerization depends on the dsRNA length (Pirher et al. 2008) and the number of TLR3 molecules involved in the interaction (TLR3 dimer or trimer; Mineev et al. 2014), selection on these residues may play a crucial role in ligand recognition in TLR3. In

general, however, most PSS in avian TLR3 are very conservative (fig. 4; for details see also supplementary material S1: table S12 and supplementary material S7, Supplementary Material online), which is consistent with the results reported by Wang, Zhang, Liu, et al. (2016) who identified the TLR3 family as the most conservative TLR family within vertebrates. The avian PSS which can be suggested as candidates for relevant functional effect based on the nonconservativeness of the changes in their physicochemical properties are described in supplementary material S1: text S2, Supplementary Material online.

Conclusions

To our knowledge, this is the first study investigating adaptive evolution in all members of the avian TLR protein family in high number of species and most avian orders. Although we were unable to extract data to all *TLR* genes from all avian whole-genome assemblies investigated, our study covers 87% of the genes in these species, providing the most comprehensive analysis published so far. Our results provide important novel insights into the history of TLR family evolution including the processes of gene duplication, pseudogenization and diversifying selection. Most importantly, we show that the main duplication event for *TLR1* arose before avian and mammalian lineages split into separate clades (giving rise to *TLR1* [avian *TLR1B*] and *TLR10* [avian *TLR1A*]), while avian gene duplication of *TLR2* to *TLR2A* and *TLR2B* may have occurred in parallel to *TLR2* duplication in some mammals that resulted in pseudogenization of the second copy of *TLR2*. The latter conclusion however remains uncertain due to the power limitation imposed by the short sequences (not affected by gene conversion) aligned in our analysis. Furthermore, we confirmed the unique avian gene duplication in *TLR7* based on qPCR copy-number variation analysis. Surprisingly, this recent *TLR7* gene duplication in birds emerged independently in many species and representing several clades (Passeriformes, Charadriiformes, Cuculiformes, and Mesiornithiformes). Similarly, the *TLR5* pseudogenization, previously reported in several passerines and parrots, is seemingly more widespread in birds. Independent loss of functional TLR5 occurred also in seriemas (Cariamiformes), trogons (Trogoniformes), tropicbirds (Phaethontiformes), sunbitterns (Eurypygiformes), and swifts (Apodiformes).

The results of our analysis of positive selection in avian TLRs allowed us to predict functionally important, interspecifically variable positions. Consistent with some previous findings, these sites were located mainly at the ligand binding extracellular domain, mostly in TLRs where the extracellular domain is exposed to the cell surface and binds structurally diverse ligands (mainly in TLR1 [TLR1B], TLR2A, TLR2B, TLR4, and TLR5). The level of positive selection was low in endosomal TLRs which bind structurally more conservative oligonucleotide ligands (TLR3, TLR7, and TLR21). Positive selection also appears to act more on three-domain than on single-domain TLRs (except for TLR5 that is under very strong positive selection in birds). We also compared the PSS detected in this study with those previously described on an interspecific level in birds and mammals and with all sites previously identified as functionally relevant in TLRs to show strongest evidence for functional effect of the predicted PSS in TLR1 [TLR1B], TLR2A, TLR4, and TLR5. In agreement with the hypothesis that the variation in the TLR ligands is a driving force for selection in TLRs, nonconservative substitutions with the potential to dramatically change surface physicochemical properties were found mainly in two TLRs (TLR2A and TLR4) with highly variable ligands known in mammals. Altogether, our results indicate strong positive selection driving TLR evolution in birds. To better understand the significance of this adaptive evolution in avian *TLR* genes, further research would benefit from both in silico structural modeling allowing closer prediction of the positively selected physicochemical changes and functional testing of the variant effects in model in vitro systems.

Materials and Methods

Data Set

To create our data set, we used sequence data publicly available for up to 63 avian species per gene representing all orders of Neoaves (supplementary material S2: table S23, Supplementary Material online). Most of the *TLR* sequences were extracted from the whole-genome data generated by the Avian Phylogenomics Consortium (http://avian.genomics.cn/en/; Zhang, Jarvis, et al. 2014), genomes included in the B10K Project (Avianbase; http://b10k.genomics.cn/; Eöry et al. 2015). The nucleotide sequences of all avian *TLRs* were obtained by blasting (BLAST v 2.2.25+, NCBI, Zhang et al. 2000; blastn settings: E = 0.1, hits with the highest score and lowest E were accepted) the reference *TLR* sequences of chicken and zebra finch (for sequence ID number see supplementary material S1: table S14, Supplementary Material online) against the CDS database of all avian whole-genome sequences of the species included in the first B10K study (Zhang, Jarvis, et al. 2014). In cases where the blast search in the CDS databases was not successful, we ran blat search against the genomic scaffolds (e.g., the case of *TLR5* pseudogenes; see supplementary material S1: text S3 and table S15, Supplementary Material online; settings: tile size = 11, step size = 5, allowing one mismatch in tile, the number of repetitions of a tile = 1,000,000). In several avian species, where *TLRs* were previously sequenced, the sequences were found using on-line Web BLAST search (blastn searching cds database with default settings; https://blast.ncbi.nlm.nih.gov/Blast.cgi; accessed in March 2013) and downloaded directly from the GenBank nucleotide database (Benson et al. 2012). The list of NCBI accession numbers of the sequences used is given in supplementary material S2: table S23, Supplementary Material online. Basic description of the *TLR* CDS analyzed is provided in supplementary material S1: text S4, tables S16 and S17 and more detailed in supplementary material S2: table S25, Supplementary Material online.

Multiple Sequence Alignment

First, all CDSs were translated to extract the protein sequences of all representatives of each TLR family member. Then,

the multiple amino acid sequence alignment was done for each TLR separately using the ClustalW alignment tool implemented in Geneious v. 9.0.4 (Biomatters Limited; Kearse et al. 2012), or using MAFFT v. 6.850 (Katoh and Standley 2013) in the case of the duplicated genes from the vertebrate *TLR1* family. For the construction of nucleotide codon alignments corresponding to the protein sequence alignments, we used the PAL2NAL webtool v. 14 (http://www.bork.embl.de/pal2nal/; accessed in July 2014; Suyama et al. 2006). All created multiple alignments are shown in supplementary material S10, Supplementary Material online.

Gene Conversion Analysis

The occurrence of gene conversion was statistically tested on nucleotide sequence alignment of all duplicated genes (*TLR1*, *TLR2*, and *TLR7*) using GENECONV v. 1.81 (S. A. Sawyer, Washington University in St. Louis, https://www.math.wustl.edu/∼sawyer/geneconv/). Full program setting is provided for all investigated genes in supplementary material S3, Supplementary Material online.

Phylogenetic Tree Analyses

Phylogenetic analysis was done only in the duplicated avian *TLRs* (i.e., in *TLR1*, *TLR2*, and *TLR7*), also including other vertebrate taxa, that is, mammals, reptiles, amphibians, and fishes (involved species and their sequences IDs are listed in supplementary material S2: table S24, Supplementary Material online). For the *TLR1* family, the phylogenetic analysis was based on the sequence alignment of nonconverted amino acid regions only. Amino acid sequences were used instead of nucleotide sequences to avoid the biasing effect of rapid evolution on the third codon positions on broad evolutionary scales (i.e., higher vertebrate taxa). For the phylogenetic analysis, we used two approaches: maximum likelihood calculated in PhyML v. 3.0 software package (Bootstrap: 1,000; Tree type: SPR&NNI; Substitution model: LG; Guindon et al. 2010) and Bayesian estimation of phylogeny calculated using MrBayes v. 3.2.1 (number of generations: 1,000; burn-in fraction was set to default of 25%; the critical value for topological convergence diagnostic was 0.01; the best suited aa model for TLR1: Fixed(Jones) and for TLR2: Fixed(Wag); Ronquist et al. 2012). Schematic phylogenetic trees for visualization of *TLR7* duplication (fig. 2) and *TLR5* pseudogenization (fig. 3) were constructed using previously published avian phylogeny (Jarvis et al. 2014). All trees were graphically adjusted in FigTree v1.3.1 (A. Rambaut, University of Edinburgh, United Kingdom; http://tree.bio.ed.ac.uk/software/figtree/).

Copy Number Variation Analysis

To verify the increased number of *TLR7* copies in selected passerine species, we performed a qPCR copy number variation (CNV) analysis. Tissue samples of four model species (*Ficedula albicollis*, *Pseudopodoces humilis*, *Taeniopygia guttata*, and *Zonotrichia albicollis*) used for this analysis were obtained from various genetic banks (for details see supplementary material S1: table S18, Supplementary Material online). DNA was extracted using the DNeasy Blood and Tissue Kit (QIAGEN) and stored at −20°C. As reference single copy genes, we used *TLR3* and *TLR4*. Primers were designed to amplify a conserved region of a similar length for all studied genes (for details see supplementary material S1: table S19, Supplementary Material online). The specificity of the primers was previously verified by Sanger sequencing (Applied Biosystems 3130xl Genetic Analyzer) of a broader surrounding region of each gene in all investigated species (for PCR conditions see supplementary material S1: table S20, Supplementary Material online; NCBI accession numbers are listed in supplementary material S1: table S21, Supplementary Material online). Each sample was run in triplicate in LightCycler 480 Instrument II (Roche) using EvaGreen Dye (Biotum; Mao et al. 2007). The qPCR efficiency was calculated based on a dilution series (5-times dilution) for each gene and sample in LightCycler 480 software v1.5.1 using both 2nd Derivate function and automatic Fit Point method (the values measured are shown in supplementary material S1: table S4, Supplementary Material online). The *TLR7* copy numbers were then calculated based on a modified version of the formula proposed by Pfaffl (2001; eq. 1, where Eff stands for the PCR efficiency and Cp stands for the crossing point).

$$\mathbf{R} = 2 * \frac{\mathbf{Eff}_{\mathrm{TLR7}}{}^{-\mathbf{Cp}_{\mathrm{TLR7}}}}{\mathbf{Eff}_{\mathrm{TLR3}}{}^{-\mathbf{Cp}_{\mathrm{TLR3}}} + \mathbf{Eff}_{\mathrm{TLR4}}{}^{-\mathbf{Cp}_{\mathrm{TLR4}}}} \quad (1)$$

Positive Selection Analysis and Estimates of Conservativeness of Amino Acid Substitutions

For the detection of selective pressures acting on each avian *TLR*, the codon alignment generated by PAL2NAL tool (http://www.bork.embl.de/pal2nal/; last accessed July 2014; Suyama et al. 2006) was used to ensure correct alignment of codons. All regions involving gaps were removed before the analysis and tested separately. The problem of missing sequence data did not impact our estimates since the number of sequences sampled was not too small in any of the TLRs and thus the dN/dS ratios could have been estimated. The position numbering of PSS followed the chicken reference sequence (for NCBI IDs see supplementary material S1: table S14, Supplementary Material online). To test for positive selection acting on individual residues at the interspecific level in avian TLRs, we used two methods based on the hierarchical Bayes (Bayes Empirical Bayes, BEB) approach implementing the Markov chain Monte Carlo routine—PAML (Phylogenetic Analysis by Maximum Likelihood; Yang 2007) and FUBAR (A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection; Murrell et al. 2013). Being based on unrelated preconditions, these two tests provide independent estimates of positive selection. In PAML (Version 4.7), the codon-based substitution models (codeml) using comparison of neutral M8a (beta&$\omega = 1$) with alternative M8 (beta&ω) model were adopted. The likelihood ratio test (LRT) for comparison of two nested models was calculated using the chi-square approximation: $\chi^2 = 2\times(\mathrm{lnLM8}-\mathrm{lnLM8a})$, where LM8 and LM8a are the likelihood values. The degrees of freedom (df) were established as the difference in the numbers of parameters in used models (for details see supplementary material S1: table

S10, Supplementary Material online). If the LRT is significant ($\leq$0.05), positive selection is considered to be detected. The BEB approach (Yang 2007) was then used to determine site specific posterior probabilities indicating positive selection ($\geq$0.9) at specific codons. The phylogenetic tree needed for PAML analysis was constructed based on the previously published avian phylogeny (Jarvis et al. 2014). FUBAR (Fast Unconstrained Bayesian AppRoximation) analysis was performed at the Datamonkey server (http://www.datamonkey.org/; accessed in July 2014; Pond and Frost 2005) with the significance level of posterior probability established by default to 0.9. The FUBAR algorithm was used because it is more robust and much faster than other available selection tests, which are based on the random effect likelihood (REL; Murrell et al. 2013) methods. The degree of dissimilarity in biochemical properties of amino acid substitutions was tested using the PRIME tool (PRoperty Informed Model of Evolution; accessed in July 2014) available at the Datamonkey server (Pond and Frost 2005). This tool builds on the same conceptual frameworks as MEME (Murrell et al. 2012), but allows the nonsynonymous substitution rate β to depend not only on the site in question but also on what type of residues are being exchanged. Both predefined sets of five amino-acid composite properties were used for PRIME analysis: 1) Polarity index, Secondary structure factor, Volume, Refractivity/Heat Capacity, and Charge/Iso-electric point (Atchley et al. 2005) and 2) Chemical Composition, Polarity, Volume, Iso-electric point, and Hydropathy (Conant et al. 2007) on the significant level of posterior probabilities $\geq$0.9. Amino acid physicochemical properties (chemistry, charge, and hydrophobicity) at all PSS were graphically visualized using a web-based application Weblogo v. 3.5 (http://weblogo.threeplusone.com/create.cgi; last accessed October 2017; Crooks et al. 2004).

We compared the PSS identified in this study with the results of other studies that focused on avian and mammalian TLR evolution identifying positive selection at the interspecific level (Nakajima et al. 2008; Vinkler et al. 2009; Wlasiuk et al. 2009; Wlasiuk and Nachman 2010; Alcaide and Edwards 2011; Areal et al. 2011; Huang et al. 2011; Shen et al. 2012; Fornůsková et al. 2013; Grueber et al. 2014; Vinkler et al. 2014; Wang, Zhang, Chang, et al. 2016) and studies that described the functionally relevant residues for TLR1 (Jin et al. 2007; Omueti et al. 2007), TLR2 (Underhill et al. 1999; Lorenz et al. 2000; Xu et al. 2000; Kang and Chae 2001; Tao et al. 2002; Gautam et al. 2006; Kang et al. 2009), TLR3 (Sarkar et al. 2003; Choe et al. 2005; de Bouteiller et al. 2005; Bell et al. 2006; Sun et al. 2006; Ranjith-Kumar et al. 2007; Liu et al. 2008; Pirher et al. 2008; Luo et al. 2012; Mineev et al. 2014), TLR4 (Poltorak et al. 1998; Ronni et al. 2003; Nishitani et al. 2006; Kim et al. 2007; Walsh et al. 2008; Park et al. 2009; Resman et al. 2009; Meng et al. 2010; Ohto, Fukase, et al. 2012; Ohto, Yamakawa, et al. 2012; Garate and Oostenbrink 2013; Paramo et al. 2013; Scior et al. 2013; Wang, Su, et al. 2016), TLR5 (Jacchieri et al. 2003; Andersen-Nissen et al. 2007; Keestra et al. 2008; Yoon et al. 2012; Ivicak-Kocjan et al. 2013; Song et al. 2017), TLR7 (Wei et al. 2009; Yu et al. 2013; Tseng et al. 2014; Gentile et al. 2015; Gupta et al. 2016; Zhang et al. 2016), TLR10 (Hasan et al. 2005; Nyman et al. 2008; Guan et al. 2010; Jang and Park 2014), TLR15 (Wang, Zhang, Chang, et al. 2016), and TLR21 (Keestra et al. 2010).

Protein Structure Modeling

To predict 3D structures of all avian TLRs, the I-TASSER v. 5.0 server (https://zhanglab.ccmb.med.umich.edu/I-TASSER/; last accessed November 2016; Roy et al. 2010) was used. Since all TLRs are transmembrane proteins, the ECD and ICD (and in the case of TLR3 also TM region) were modeled separately, always based on the chicken reference sequence (for NCBI IDs see the supplementary material S1: table S14, Supplementary Material online). The domains were identified by SMART v. 7.0 web tool (http://smart.embl-heidelberg.de/; last accessed November 2016; Letunic and Bork 2018), the amino acid ranges of ECD and ICD for each TLR are provided in supplementary material S1: table S22, Supplementary Material online. The I-TASSER model with the highest C value reflecting the confidence score for estimating the quality of predicted models was always downloaded and used for further analysis. The graphical visualization of important residues was then done using the PyMOL Molecular Graphics System (Version 1.7.6, Schrödinger, LLC).

Assessing Function of Positively Selected Sites

The list of the previously reported functionally important positions as well as PSS detected in other interspecific studies for all TLRs were obtained by a detailed review of the published literature (for the complete list of references see supplementary material S6: table S27, Supplementary Material online). The distance of any PSS detected in our study from these previously described functionally important residues was measured on the 3D structural models of TLRs (obtained in previous step) in the PyMOL Molecular Graphics System (Version 1.7.6, Schrödinger, LLC) using python command iterate and plugin distancetoatom, where PSS lying in distance closer to 5 Å were considered as closely connected to the functionally important residues, that is, having potential influence on the receptor function.

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank especially Zuzana Świderská and Anna Bryjová for their helpful advice during laboratory experiments and Tatiana Aghová and Petri Kemmpainen for their advices in phylogenetic analyses. Tissue samples for this analysis were kindly provided from The DNA Bank of the Natural History Museum of Oslo (NHMO), The Museum of Zoology, University of Michigan (UMMZ), and The Genetic bank of the Department of Zoology, Charles University (ZCU). The data sets generated and/or analyzed during the current study are available in the NCBI GenBank repository (see supplementary material S1: tables S14 and S21 and supplementary material S2: tables S23 and S24, Supplementary Material online) or are included in this published article and

its supplementary material files online. This work was supported by the Czech Science Foundation (P505/10/1871 and P506/15-11782S); by institutional research support (SVV-260434/2018); by the Charles University (GAUK 540214, UNCE 204069 and PRIMUS/17/SCI/12); by the travel grant provided by the Charles University Mobility Fund to H.V. and by BBSRC CASE award to M.W.G.D. The Avian Phylogenetic and B10K consortium kindly provided the access to avian genome sequences, with support from Institute Strategic Grant funding from the BBSRC. Computing and data storage facilities were provided by the National Grid Infrastructure MetaCentrum (project CESNET LM2015042). The authors' contribution to this article was as follows: H.V. (70%)—study design, performed all given analyses, prepared all output files, wrote the first draft of the article; M.W.G.D. (5%)—involvement in BLAST of genomic databases and PAML analysis; D.W.B. (10%)—supervision of M.W.G.D. and study design and M.V. (15%)—study design, revised the article. All authors contributed by their comments to the article preparation, read, and approved the final article.

References

Alcaide M, Edwards SV. 2011. Molecular evolution of the Toll-like receptor multigene family in birds. *Mol Biol Evol*. 28(5):1703–1715.

Andersen-Nissen E, Smith KD, Bonneau R, Strong RK, Aderem A. 2007. A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. *J Exp Med*. 204(2):393–403.

Arbour N, Lorenz E, Schutte B, Zabner J, Kline J, Jones M, Frees K, Watt J, Schwartz D. 2000. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. *Nat Genet*. 25(2):187–191.

Areal H, Abrantes J, Esteves PJ. 2011. Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evol Biol*. 11:368.

Atchley WR, Zhao J, Fernandes AD, Drüke T. 2005. Solving the protein sequence metric problem. *Proc Natl Acad Sci U S A*. 102(18):6395–6400.

Bainova H, Kralova T, Bryjova A, Albrecht T, Bryja J, Vinkler M. 2014. First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds. *Dev Comp Immunol*. 45(1):151–155.

Bell JK, Askins J, Hall PR, Davies DR, Segal DM. 2006. The dsRNA binding site of human Toll-like receptor 3. *Proc Natl Acad Sci U S A*. 103(23):8792–8797.

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2012. GenBank. *Nucleic Acids Res*. 41(D1):D36–D42.

Boyd AC, Peroval MY, Hammond JA, Prickett MD, Young JR, Smith AL. 2012. TLR15 is unique to avian and reptilian lineages and recognizes a yeast-derived agonist. *J Immunol*. 189(10):4930–4938.

Brownlie R, Allan B. 2011. Avian Toll-like receptors. *Cell Tissue Res*. 343(1):121–130.

Brownlie R, Zhu J, Allan B, Mutwiri GK, Babiuk LA, Potter A, Griebel P. 2009. Chicken TLR21 acts as a functional homologue to mammalian TLR9 in the recognition of CpG oligodeoxynucleotides. *Mol Immunol*. 46(15):3163–3170.

Burri R, Salamin N, Studer RA, Roulin A, Fumagalli L. 2010. Adaptive divergence of ancient gene duplicates in the avian MHC class II β. *Mol Biol Evol*. 27(10):2360–2374.

Cario E. 2013. The human TLR4 variant D299G mediates inflammation-associated cancer progression in the intestinal epithelium. *Oncoimmunology* 2(7):e24890.

Cheng Y, Prickett MD, Gutowska W, Kuo R, Belov K, Burt DW. 2015. Evolution of the avian β-defensin and cathelicidin genes. *BMC Evol Biol*. 15:188.

Choe J, Kelker MS, Wilson IA. 2005. Crystal structure of human Toll-like receptor 3 (TLR3) ectodomain. *Science* 309(5734):581–585.

Conant GC, Wagner GP, Stadler PF. 2007. Modeling amino acid substitution patterns in orthologous and paralogous genes. *Mol Phylogenet Evol*. 42(2):298–307.

Cormican P, Lloyd AT, Downing T, Connell SJ, Bradley D, O'Farrelly C. 2009. The avian Toll-Like receptor pathway-Subtle differences amidst general conformity. *Dev Comp Immunol*. 33(9):967–973.

Coscia MR, Giacomelli S, Oreste U. 2011. Toll-like receptors: an overview from invertebrates to vertebrates. *Invertebr Surviv J*. 2011:210–226.

Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res*. 14(6):1188–1190.

de Bouteiller O, Merck E, Hasan U, Hubac S, Benguigui B, Trinchieri G, Bates E, Caux C. 2005. Recognition of double-stranded RNA by human Toll-like receptor 3 and downstream receptor signaling requires multimerization and an acidic pH. *J Biol Chem*. 280(46):38133–38145.

de Zoete MR, Bouwman LI, Keestra AM, van Putten JPM. 2011. Cleavage and activation of a Toll-like receptor by microbial proteases. *Proc Natl Acad Sci U S A*. 108(12):4968–4973.

Ellegren H. 2008. Comparative genomics and the study of evolution by natural selection. *Mol Ecol*. 17(21):4586–4596.

Emery NJ, Clayton NS. 2004. The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306(5703):1903–1907.

Eöry L, Gilbert MTP, Li C, Li B, Archibald A, Aken BL, Zhang G, Jarvis E, Flicek P, Burt DW. 2015. Avianbase: a community resource for bird genomics. *Genome Biol*. 16:21.

Ewald SE, Lee BL, Lau L, Wickliffe KE, Shi G-P, Chapman HA, Barton GM. 2008. The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. *Nature* 456(7222):658–662.

Farmer CG. 2000. Parental care: the key to understanding endothermy and other convergent features in birds and mammals. *Am Nat*. 155(3):326–334.

Ferrer-Admetlla A, Bosch E, Sikora M, Marques-Bonet T, Ramirez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J, Casals F. 2008. Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol*. 181(2):1315–1322.

Fornůsková A, Vinkler M, Pagès M, Galan M, Jousselin E, Cerqueira F, Morand S, Charbonnel N, Bryja J, Cosson J-F. 2013. Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). *BMC Evol Biol*. 13:194.

Garate JA, Oostenbrink C. 2013. Lipid a from lipopolysaccharide recognition: structure, dynamics and cooperativity by molecular dynamics simulations. *Proteins* 81(4):658–674.

Gautam JK, Ashish CLD, Krueger JK, Smith MF. Jr. 2006. Structural and functional evidence for the role of the TLR2 DD loop in TLR1/TLR2 heterodimerization and signaling. *J Biol Chem*. 281:30132–30142.

Gay NJ, Gangloff M. 2007. Structure and function of Toll receptors and their Ligands. *Annu Rev Biochem*. 76:141–165.

Gentile F, Deriu MA, Licandro G, Prunotto A, Danani A, Tuszynski JA. 2015. Structure based modeling of small molecules binding to the TLR7 by atomistic level simulations. *Molecules* 20(5):8316–8340.

Grueber CE, Wallis GP, Jamieson IG. 2014. Episodic positive selection in the evolution of avian Toll-like receptor innate immunity genes. *PLoS One* 9(3):e89632.

Grueber CE, Wallis GP, King TM, Jamieson IG. 2012. Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. *PLoS One* 7(9):e45011.

Guan Y, Ranoa DRE, Jiang S, Mutha SK, Li X, Baudry J, Tapping RI. 2010. Human TLRs 10 and 1 share common mechanisms of innate immune sensing but not signaling. *J Immunol*. 184(9):5094–5103.

Guindon S, Dufayard J-F, 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(3):307–321.

Gupta CL, Akhtar S, Sayyed U, Pathak N, Bajpai P. 2016. In silico analysis of human Toll-like receptor 7 ligand binding domain. *Biotechnol Appl Biochem*. 63(3):441–450.

Hasan U, Chaffois C, Gaillard C, Saulnier V, Merck E, Tancredi S, Guiet C, Briere F, Vlach J, Lebecque S, et al. 2005. Human TLR10 is a functional

receptor, expressed by B cells and plasmacytoid dendritic cells, which activates gene transcription through MyD88. *J Immunol*. 174(5):2942–2950.
Hawn TR, Verbon A, Lettinga KD, Zhao LP, Li SS, Laws RJ, Skerrett SJ, Beutler B, Schroeder L, Nachman A, et al. 2003. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to Legionnaires' disease. *J Exp Med*. 198(10):1563–1572.
Hawn TR, Wu H, Grossman JM, Hahn BH, Tsao BP, Aderem A. 2005. A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. *Proc Natl Acad Sci U S A*. 102(30):10593–10597.
Hayashi F, Smith K, Ozinsky A, Hawn T, Yi E, Goodlett D, Eng J, Akira S, Underhill D, Aderem A. 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature* 410(6832):1099–1103.
Higuchi M, Matsuo A, Shingai M, Shida K, Ishii A, Funami K, Suzuki Y, Oshiumi H, Matsumoto M, Seya T. 2008. Combinational recognition of bacterial lipoproteins and peptidoglycan by chicken Toll-like receptor 2 subfamily. *Dev Comp Immunol*. 32(2):147–155.
Huang Y, Temperley ND, Ren L, 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*. 11(149):1–17.
Iqbal M, Philbin VJ, Withanage GSK, Wigley P, Beal RK, Goodchild MJ, Barrow P, McConnell I, Maskell DJ, Young J, et al. 2005. Identification and functional characterization of chicken Toll-like receptor 5 reveals a fundamental role in the biology of infection with *Salmonella enterica* serovar typhimurium. *Infect Immun*. 73(4):2344–2350.
Ivicak-Kocjan K, Panter G, Bencina M, Jerala R. 2013. Determination of the physiological 2: 2 TLR5: flagellin activation stoichiometry revealed by the activity of a fusion receptor. *Biochem Biophys Res Commun*. 435(1):40–45.
Iwasaki A, Medzhitov R. 2015. Control of adaptive immunity by the innate immune system. *Nat Immunol*. 16(4):343–353.
Jacchieri S, Torquato R, Brentani R. 2003. Structural study of binding of flagellin by Toll-like receptor 5. *J Bacteriol*. 185(14):4243–4247.
Janeway CA, Medzhitov R. 2002. Innate immune recognition. *Annu Rev Immunol*. 20:197–216.
Jang T, Park HH. 2014. Crystal structure of TIR domain of TLR6 reveals novel dimeric interface of TIR-TIR interaction for Toll-like receptor signaling pathway. *J Mol Biol*. 426(19):3305–3313.
Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346(6215):1320–1331.
Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO. 2012. The global diversity of birds in space and time. *Nature* 491(7424):444–448.
Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, Lee H, Lee J-O. 2007. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* 130(6):1071–1082.
Kang JY, Nan X, Jin MS, Youn S-J, Ryu YH, Mah S, Han SH, Lee H, Paik S-G, Lee J-O. 2009. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* 31(6):873–884.
Kang T, Chae G. 2001. Detection of Toll-like receptor 2 (TLR2) mutation in the lepromatous leprosy patients. *FEMS Immunol Med Microbiol*. 31(1):53–58.
Kanno A, Yamamoto C, Onji M, Fukui R, Saitoh S, Motoi Y, Shibata T, Matsumoto F, Muta T, Miyake K. 2013. Essential role for Toll-like receptor 7 (TLR7)-unique cysteines in an intramolecular disulfide bond, proteolytic cleavage and RNA sensing. *Int Immunol*. 25(7):413–422.
Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 30(4):772–780.
Kawai T, Akira S. 2010. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol*. 11(5):373–384.
Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28(12):1647–1649.
Keestra AM, de Zoete MR, Bouwman LI, Vaezirad MM, van Putten JPM. 2013. Unique features of chicken Toll-like receptors. *Dev Comp Immunol*. 41(3):316–323.
Keestra AM, de Zoete MR, Bouwman LI, van Putten JPM. 2010. Chicken TLR21 is an innate CpG DNA receptor distinct from mammalian TLR9. *J Immunol*. 185(1):460–467.
Keestra AM, de Zoete MR, van Aubel RAMH, van Putten JPM. 2007. The central Leucine-rich repeat region of chicken TLR16 dictates unique ligand specificity and species-specific interaction with TLR2. *J Immunol*. 178(11):7110.
Keestra AM, de Zoete MR, van Aubel RAMH, van Putten JPM. 2008. Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. *Mol Immunol*. 45(5):1298–1307.
Kim HM, Park BS, Kim J-I, Kim SE, Lee J, Oh SC, Enkhbayar P, Matsushima N, Lee H, Yoo OJ, Lee J-O. 2007. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. *Cell* 130(5):906–917.
Letunic I, Bork P. 2018. 20 years of the SMART protein domain annotation resource. *Nucleic Acids Res*. 46(D1):D493–D496.
Leveque G, Forgetta V, Morroll S, Smith A, Bumstead N, Barrow P, Loredo-Osti J, Morgan K, Malo D. 2003. Allelic variation in TLR4 is linked to susceptibility to Salmonella enterica serovar typhimurium infection in chickens. *Infect Immun*. 71(3):1116–1124.
Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, Segal DM, Davies DR. 2008. Structural basis of Toll-like receptor 3 signaling with double-stranded RNA. *Science* 320(5874):379–381.
Lorenz E, Mira J, Cornish K, Arbour N, Schwartz D. 2000. A novel polymorphism in the Toll-like receptor 2 gene and its potential association with staphylococcal infection. *Infect Immun*. 68(11):6398–6401.
Lovegrove BG. 2017. A phenology of the evolution of endothermy in birds and mammals. *Biol Rev*. 92(2):1213–1240.
Luo J, Obmolova G, Malia TJ, Wu S-J, Duffy KE, Marion JD, Bell JK, Ge P, Zhou ZH, Teplyakov A, et al. 2012. Lateral clustering of TLR3: dsRNA signaling units revealed by TLR3ecd: 3Fabs quaternary structure. *J Mol Biol*. 421(1):112–124.
Mao F, Leung W-Y, Xin X. 2007. Characterization of EvaGreen and the implication of its physicochemical properties for qPCR applications. *BMC Biotechnol*. 7:76.
Medvedev AE. 2013. Toll-like receptor polymorphisms, inflammatory and infectious diseases, allergies, and cancer. *J Interferon Cytokine Res off J Int Soc Interferon Cytokine Res*. 33(9):467–484.
Meng J, Lien E, Golenbock DT. 2010. MD-2-mediated ionic interactions between lipid A and TLR4 are essential for receptor activation. *J Biol Chem*. 285(12):8695–8702.
Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N. 2012. Molecular evolution of vertebrate Toll-like receptors: evolutionary rate difference between their leucine-rich repeats and their TIR domains. *Gene* 503(2):235–243.
Mineev KS, Goncharuk SA, Arseniev AS. 2014. Toll-like receptor 3 transmembrane domain is able to perform various homotypic interactions: an NMR structural study. *FEBS Lett*. 588(21):3802–3807.
Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: a fast, unconstrained Bayesian AppRoximation for inferring selection. *Mol Biol Evol*. 30(5):1196–1205.
Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 8(7):e1002764.
Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H, Ishida T, Kimura A. 2008. Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* 60(12):727–735.
Netea MG, Wijmenga C, O'Neill LAJ. 2012. Genetic variation in Toll-like receptors and disease susceptibility. *Nat Immunol*. 13(6):535–542.

Nishitani C, Mitsuzawa H, Sano H, Shimizu T, Matsushima N, Kuroki Y. 2006. Toll-like receptor 4 region Glu24-Lys47 is a site for MD-2 binding: importance of CYS29 and CYS40. *J Biol Chem*. 281(50):38322–38329.

Nyman T, Stenmark P, Flodin S, Johansson I, Hammarström M, Nordlund P. 2008. The crystal structure of the human Toll-like receptor 10 cytoplasmic domain reveals a putative signaling dimer. *J Biol Chem*. 283(18):11861–11865.

Ohto U, Fukase K, Miyake K, Shimizu T. 2012. Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proc Natl Acad Sci U S A*. 109:7421–7426.

Ohto U, Yamakawa N, Akashi-Takamura S, Miyake K, Shimizu T. 2012. Structural analyses of human Toll-like receptor 4 polymorphisms D299G and T399I. *J Biol Chem*. 287(48):40611–40617.

Olkowicz S, Kocourek M, Lučan RK, Porteš M, Fitch WT, Herculano-Houzel S, Němec P. 2016. Birds have primate-like numbers of neurons in the forebrain. *Proc Natl Acad Sci U S A*. 113(26):7255–7260.

Omueti KO, Mazur DJ, Thompson KS, Lyle EA, Tapping RI. 2007. The polymorphism P315L of human Toll-like receptor 1 impairs innate immune sensing of microbial cell wall components. *J Immunol*. 178(10):6387–6394.

Palm NW, Medzhitov R. 2009. Pattern recognition receptors and control of adaptive immunity. *Immunol Rev*. 227(1):221–233.

Paramo T, Piggot TJ, Bryant CE, Bond PJ. 2013. The structural basis for endotoxin-induced allosteric regulation of the Toll-like receptor 4 (TLR4) innate immune receptor. *J Biol Chem*. 288(51):36215–36225.

Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 458(7242):1191–1195.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res*. 29(9):e45.

Philbin VJ, Iqbal M, Boyd Y, Goodchild MJ, Beal RK, Bumstead N, Young J, Smith AL. 2005. Identification and characterization of a functional, alternatively spliced Toll-like receptor 7 (TLR7) and genomic disruption of TLR8 in chickens. *Immunology* 114(4):507–521.

Pirher N, Ivicak K, Pohar J, Bencina M, Jerala R. 2008. A second binding site for double-stranded RNA in TLR3 and consequences for interferon activation. *Nat Struct Mol Biol*. 15(7):761–763.

Poltorak A, He X, Smirnova I, Liu M-Y, Huffel CV, Du X, Birdwell D, Alejos E, Silva M, Galanos C. 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. *Science* 282(5396):2085–2088.

Pond SLK, Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21(10):2531–2533.

Ranjith-Kumar CT, Miller W, Xiong J, Russell WK, Lamb R, Santos J, Duffy KE, Cleveland L, Park M, Bhardwaj K, et al. 2007. Biochemical and functional analyses of the human Toll-like receptor 3 ectodomain. *J Biol Chem*. 282(10):7668–7678.

Raven N, Lisovski S, Klaassen M, Lo N, Madsen T, Ho SYW, Ujvari B. 2017. Purifying selection and concerted evolution of RNA-sensing Toll-like receptors in migratory waders. *Infect Genet Evol*. 53:135–145.

Reddick LE, Alto NM. 2014. Bacteria fighting back: how pathogens target and subvert the host innate immune system. *Mol Cell* 54(2):321–328.

Resman N, Vasl J, Oblak A, Pristovsek P, Gioannini TL, Weiss JP, Jerala R. 2009. Essential roles of hydrophobic residues in both MD-2 and Toll-like receptor 4 in activation by endotoxin. *J Biol Chem*. 284(22):15052–15060.

Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci U S A*. 102(27):9577–9582.

Ronni T, Agarwal V, Haykinson M, Haberland M, Cheng G, Smale S. 2003. Common interaction surfaces of the Toll-like receptor 4 cytoplasmic domain stimulate multiple nuclear targets. *Mol Cell Biol*. 23(7):2543–2555.

Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol*. 61(3):539–542.

Roy A, Kucukural A, Zhang Y. 2010. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc*. 5(4):725–738.

Rupasree Y, Naushad SM, Rajasekhar L, Uma A, Kutala VK. 2015. Association of TLR4 (D299G, T399I), TLR9-1486T < C, TIRAP S180L and TNF-alpha promoter (-1031,-863,-857) polymorphisms with risk for systemic lupus erythematosus among South Indians. *Lupus* 24:50–57.

Sarkar S, Smith H, Rowe T, Sen G. 2003. Double-stranded RNA signaling by Toll-like receptor 3 requires specific tyrosine residues in its cytoplasmic domain. *J Biol Chem*. 278(7):4393–4396.

Scior T, Lozano-Aponte J, Figueroa-Vazquez V, Yunes-Rojas JA, Zähringer U, Alexander C. 2013. Three-dimensional mapping of differential amino acids of human, murine, canine and equine TLR4/MD-2 receptor complexes conferring endotoxic activation by lipid A, antagonism by Eritoran and species-dependent activities of Lipid IVA in the mammalian LPS sensor system. *Comput Struct Biotechnol J*. 7:e201305003.

Shen T, Xu S, Wang X, Yu W, Zhou K, Yang G. 2012. Adaptive evolution and functional constraint at TLR4 during the secondary aquatic adaptation and diversification of cetaceans. *BMC Evol Biol*. 12:39.

Shen X, Shi R, Zhang H, Li K, Zhao Y, Zhang R. 2010. The Toll-like receptor 4 D299G and T399I polymorphisms are associated with Crohn's disease and ulcerative colitis: a meta-analysis. *Digestion* 81(2):69–77.

Song WS, Jeon YJ, Namgung B, Hong M, Yoon S. 2017. A conserved TLR5 binding and activation hot spot on flagellin. *Sci Rep*. 7:40878.

Sun J, Duffy KE, Ranjith-Kumar CT, Xiong J, Lamb RJ, Santos J, Masarapu H, Cunningham M, Holzenburg A, Sarisky RT, et al. 2006. Structural and functional analyses of the human Toll-like receptor 3 – role of glycosylation. *J Biol Chem*. 281(16):11144–11151.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res*. 34(Web Server):W609–W612.

Tao X, Xu Y, Zheng Y, Beg AA, Tong L. 2002. An extensively associated dimer in the structure of the C713S mutant of the TIR domain of human TLR2. *Biochem Biophys Res Commun*. 299(2):216–221.

Temperley ND, Berlin S, Paton IR, Griffin DK, Burt DW. 2008. Evolution of the chicken Toll-like receptor gene family: a story of gene gain and gene loss. *BMC Genomics* 9:62.

Tseng C-Y, Gajewski M, Danani A, Tuszynski JA. 2014. Homology and molecular dynamics models of Toll-like receptor 7 protein and its dimerization. *Chem Biol Drug Des*. 83(6):656–665.

Underhill D, Ozinsky A, Hajjar A, Stevens A, Wilson C, Bassetti M, Aderem A. 1999. The Toll-like receptor 2 is recruited to macrophage phagosomes and discriminates between pathogens. *Nature* 401(6755):811–815.

Vinkler M, Bainova H, Bryja J. 2014. Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. *Genet Sel Evol*. 46:72.

Vinkler M, Bryjova A, Albrecht T, Bryja J. 2009. Identification of the first Toll-like receptor gene in passerine birds: tLR4 orthologue in zebra finch (Taeniopygia guttata). *Tissue Antigens* 74(1):32–41.

Waite DW, Taylor MW. 2014. Characterizing the avian gut microbiota: membership, driving influences, and potential function. *Front Microbiol*. 5:223.

Walsh C, Gangloff M, Monie T, Smyth T, Wei B, McKinley TJ, Maskell D, Gay N, Bryant C. 2008. Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa. *J Immunol*. 181(2):1245–1254.

Wang J, Zhang Z, Chang F, Yin D. 2016. Bioinformatics analysis of the structural and evolutionary characteristics for Toll-like receptor 15. *PeerJ* 4:e2079.

Wang J, Zhang Z, Liu J, Zhao J, Yin D. 2016. Ectodomain architecture affects sequence and functional evolution of vertebrate Toll-like receptors. *Sci Rep*. 6:26705.

Wang Y, Su L, Morin MD, Jones BT, Whitby LR, Surakattula MMRP, Huang H, Shi H, Choi JH, Wang K-w. 2016. TLR4/MD-2 activation by a synthetic agonist with no similarity to LPS. *Proc Natl Acad Sci U S A*. 113(7):E884–E893.

Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rössle SC. 2009. Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. *Protein Sci.* 18(8):1684–1691.

Wlasiuk G, Khan S, Switzer WM, Nachman MW. 2009. A history of recurrent positive selection at the Toll-like receptor 5 in primates. *Mol Biol Evol.* 26(4):937–949.

Wlasiuk G, Nachman MW. 2010. Adaptation and constraint at Toll-like receptors in primates. *Mol Biol Evol.* 27(9):2172–2186.

Xu Y, Tao X, Shen B, Horng T, Medzhitov R, Manley JL, Tong L. 2000. Structural basis for signal transduction by the Toll/interleukin-1 receptor domains. *Nature* 408(6808):111–115.

Yang J, Zhang E, Liu F, Zhang Y, Zhong M, Li Y, Zhou D, Chen Y, Cao Y, Xiao Y, et al. 2014. Flagellins of Salmonella typhi and nonpathogenic Escherichia coli are differentially recognized through the NLRC4 pathway in macrophages. *J Innate Immun.* 6(1):47–57.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24(8):1586–1591.

Yoon S, Kurnasov O, Natarajan V, Hong M, Gudkov AV, Osterman AL, Wilson IA. 2012. Structural basis of TLR5-flagellin recognition and signaling. *Science* 335(6070):859–864.

Yu H, Jin H, Sun L, Zhang L, Sun G, Wang Z, Yu Y. 2013. Toll-like receptor 7 agonists: chemical feature based pharmacophore identification and molecular docking studies. *PLoS One* 8(3):e56514.

Zareparsi S, Buraczynska M, Branham KEH, Shah S, Eng D, Li M, Pawar H, Yashar BM, Moroi SE, Lichter PR, et al. 2005. Toll-like receptor 4 variant D299G is associated with susceptibility to age-related macular degeneration. *Hum Mol Genet.* 14(11):1449–1455.

Zhang F, Gao X-D, Wu W-W, Gao Y, Zhang Y-W, Wang S-P. 2013. Polymorphisms in Toll-like receptors 2, 4 and 5 are associated with Legionella pneumophila infection. *Infection* 41(5):941–948.

Zhang G, Jarvis ED, Gilbert MTP. 2014. A flock of genomes. *Science* 346(6215):1308–1309.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, et al. 2014. Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* 346:1311–1320.

Zhang J. 2003. Evolution by gene duplication: an update. *Trends Ecol Evol.* 18(6):292–298.

Zhang Z, Ohto U, Shibata T, Krayukhina E, Taoka M, Yamauchi Y, Tanji H, Isobe T, Uchiyama S, Miyake K, et al. 2016. Structural analysis reveals that Toll-like receptor 7 is a dual receptor for guanosine and single-stranded RNA. *Immunity* 45(4):737–748.

Zhang Z, Schwartz S, Wagner L, Miller W. 2000. A greedy algorithm for aligning DNA sequences. *J Comput Biol.* 7(1–2):203–214.

Zhao Y, Yang J, Shi J, Gong Y-N, Lu Q, Xu H, Liu L, Shao F. 2011. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. *Nature* 477(7366):596–600.

PAPER II.

Vinkler M., **Bainová H.** & Bryja J. (2014): Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. Genetics Selection Evolution 46:72. (IF_{2014}= 3.821)

In this paper, we focused on analysis of interspecific variability of three TLR (bacterial-sensing TLR4 and TLR5 and viral-sensing TLR7) within the Galloanserae bird clade, including the investigation of their phylogeny, assessing structural conservation and estimating site-specific selection pressures. Based on the results of these analyses we revealed that the physiochemical properties of protein surface varied between the specific TLR, mainly regarding the surface electrostatic potential distribution. Moreover, the predicted ligand-binding features (mainly in two investigated bacteria-sensing TLRs) differed between the avian proteins and their fish and mammalian counterparts, but also varied within the Galloanserae birds. This might be the evidence for different selective pressures caused by various pathogens. We also tested positive selection, described evolutionarily non-conservative sites and predicted functionally important positions which might play crucial role in pathogen recognition of all these receptors in Galloanserae lineage. Based on our findings, the receptor variability appears to be functionally more conserved for viral-sensing TLR7 than for the bacterial-sensing TLRs. In general context the host-pathogen coevolution has a major effect on the features of host immune receptors. Our results suggest that avian and mammalian TLRs may be differentially adapted to pathogen-derived ligand recognition. We have detected signatures of positive selection even within the Galloanserae lineage. This study is the first one to depict evolutionary pressures on Galloanserae TLR, but also to estimate the validity of current knowledge on TLRs function (based on mammalian and chicken models) for non-model species of this clade.

Vinkler *et al. Genetics Selection Evolution* 2014, **46**:72
http://www.gsejournal.org/content/46/1/72

GSE Genetics Selection Evolution

RESEARCH Open Access

Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds

Michal Vinkler[1*], Hana Bainová[1] and Josef Bryja[2]

Abstract

Background: Toll-like receptors (TLR) are essential activators of the innate part of the vertebrate immune system. In this study, we analysed the interspecific variability of three TLR (bacterial-sensing TLR4 and TLR5 and viral-sensing TLR7) within the Galloanserae bird clade, investigated their phylogeny, assessed their structural conservation and estimated site-specific selection pressures.

Results: Physiochemical properties varied according to the TLR analysed, mainly with regards to the surface electrostatic potential distribution. The predicted ligand-binding features (mainly in TLR4 and TLR5) differed between the avian proteins and their fish and mammalian counterparts, but also varied within the Galloanserae birds. We identified 20 positively selected sites in the three TLR, among which several are topologically close to ligand-binding sites reported for mammalian and fish TLR. We described 26, 28 and 25 evolutionarily non-conservative sites in TLR4, TLR5 and TLR7, respectively. Thirteen of these sites in TLR4, and ten in TLR5 were located in functionally relevant regions. The variability appears to be functionally more conserved for viral-sensing TLR7 than for the bacterial-sensing TLR. Amino-acid positions 268, 270, 343, 383, 444 and 471 in TLR4 and 180, 183, 209, 216, 264, 342 and 379 in TLR5 are key candidates for further functional research.

Conclusions: Host-pathogen co-evolution has a major effect on the features of host immune receptors. Our results suggest that avian and mammalian TLR may be differentially adapted to pathogen-derived ligand recognition. We have detected signatures of positive selection even within the Galloanserae lineage. To our knowledge, this is the first study to depict evolutionary pressures on Galloanserae TLR and to estimate the validity of current knowledge on TLR function (based on mammalian and chicken models) for non-model species of this clade.

Background

Toll-like receptors (TLR) are part of the pattern recognition receptor machinery and play a key role in initial pathogen recognition in vertebrates [1]. Since TLR are responsible for the recognition of microbe-associated molecular patterns that are present on pathogens [2], the efficiency of innate immunity in vertebrates is conditioned by their optimal functioning. Substantial variability has been described for TLR, both at the interspecific and intraspecific level [3]. Variability in the structure and binding features of TLR could significantly influence host resistance to diseases and vulnerability to autoimmune damage. TLR evolution has been intensively studied in vertebrates in general [4-6] and within the mammalian clade in particular [7-10]. In birds, although the number of studies on the evolution of TLR has steadily increased [11-13], there is still a very limited understanding of the functional significance of the putatively adaptive variability observed.

There are ten *TLR* genes in birds [5,11,14], including those encoding predominantly bacterial-sensing TLR, such as TLR4 and TLR5, and viral-sensing TLR, such as TLR7. Although the TLR and their corresponding genes have been well characterised in the domestic chicken (*Gallus gallus domesticus* [15-21]), our knowledge on TLR of other avian species remains rather fragmentary [8,22-31]. Based mainly on data for the domestic chicken, we assume that the ligands recognized by avian TLR are similar to their mammalian counterparts [14], i.e. TLR4 binds lipopolysaccharide (LPS) [32], TLR5 binds flagellin [33] and TLR7 binds viral single-stranded RNA and synthetic antivirals [19,34]. The structure of bacterial ligands, such as LPS or flagellin, varies between species [35-37], which may exert selective

* Correspondence: michal.vinkler@natur.cuni.cz
[1]Department of Zoology, Faculty of Science, Charles University in Prague, Praha, Czech Republic
Full list of author information is available at the end of the article

pressure on TLR and lead to host-pathogen co-evolution of these molecules [3]. Previously, it has been shown that TLR-ligand binding is, in many aspects, species-specific and that TLR protein variation may have a functional significance [32,33,38-41] that could affect resistance to pathogen infections [16,42].

The function of the receptor is determined by its tertiary structure and surface features that confer specificity for ligands. TLR are type I integral membrane glycoproteins that are characterized by an extracellular ligand-binding domain (ECD) and a cytoplasmic signalling Toll/IL-1R homology (TIR) domain. Three-dimensional (3D) molecular structures of the ECD [43-50] and TIR domains [51,52] have been described for several human and mouse TLR. In several other cases, the ECD structures are predicted [53-58]. Most recently, the TLR5 ECD structure has been resolved in zebrafish (*Danio rerio* [59]). These studies have identified TLR ligand-binding sites. All these protein structures, however, represent mammalian or fish TLR only, no avian TLR have been characterised or even predicted in 3D as yet.

In this study, we provide a comprehensive view on the evolution of TLR in Galloanserae birds. Galloanserae is an ancient clade of avian species that includes the orders Galliformes and Anseriformes which are largely separated from all other modern birds of the Neoaves lineage [60]. Currently, this is the most extensively studied avian taxon regarding TLR immunogenetics. Based on published data, we investigated the sequence and structural variability of TLR4, TLR5 and TLR7 within this bird clade. These three TLR were chosen as representatives of the bacterial-sensing and viral-sensing TLR based on published sequence data. Since comparison of 3D structures and protein surface features may reveal biologically interesting similarities not detectable by sequence analysis [61], the 3D tertiary structure of these three proteins was modelled. We then carried out a structural comparison of functionally important regions, a comparison of surface electrostatic potentials and four independent analyses of positive selection. Superposition of the 3D structures allowed us, for the first time, to conduct a phenetic (not only phylogenetic) analysis of avian TLR evolution.

Methods

Input sequences

Coding DNA sequences (CDS) for the selected TLR were downloaded from the National Centre for Biotechnology Information (NCBI) GenBank for all Galloanserae species currently available in full length (n = 13) and for humans and mouse. GenBank ID are given in Additional file 1: Table S1. CDS translation was performed using BioEdit Sequence Alignment Editor (Tom Hall, Ibis Biosciences, Carlsbad, California, USA) and protein sequences for each gene were aligned using ClustalW multiple-sequence alignment software. Alignment of nucleotide codons and their corresponding protein sequences was undertaken using the PAL2NAL webtool [62].

Structural analysis

To predict the distribution of structural domains in the proteins, we applied SMART [63]. Since the analysis revealed differences between species in the number and position of leucine-rich repeats (LRR) we tested LRR distribution by an independent approach using the LRRfinder, with upper and lower boundaries fixed at 95% and 80%, respectively [64]. Because the results of these two approaches differed slightly, SMART predictions were only used to identify N-terminal LRR (LRRNT) and C-terminal LRR (LRRCT) motifs which in most cases were not identified by LRRfinder, while other LRR were identified based on LRRfinder predictions. Molecular weight and charge at pH = 7 were also calculated for each predicted protein. Presence of a transmembrane (TM) region in each protein was checked for on the DAS-TMfilter server [65]. When the TIR domain was not detected by SMART, we used PFAM comparison [63] to check for its presence. Signal peptides were identified using SignalP 4.0 [66]. Finally, secondary and 3D tertiary structures of the three TLR were predicted by applying a comparative modelling approach (see e.g. [56,57] or [55]) using I-TASSER [67], which is currently the leading protein structure prediction server (see [68]) and http://predictioncenter.org/casp10/)). The I-TASSER server uses a hierarchical protein-structure modelling approach based on secondary-structure enhanced profile-profile threading alignment and iterative implementation of the threading assembly refinement program [68]. For the prediction of secondary structures, we used the whole CDS, whereas ECD and TIR domains were modelled separately in order to compare tertiary structures. Since it has been shown that the sequence for TLR7 up to LRR14 is cleaved in the endoplasmic reticulum [69], we modelled the sequence starting at amino-acid 417 only. Models with the highest C-scores and conformation similarities to other modelled structures were used for further analysis. In these models, we excluded regions with limited structural stability (i.e. regions with high modelling errors: signal peptides and regions following LRRCT). Modelling errors in the regions of interest were estimated using ModFOLD [70]; all models had high levels of confidence with P-values less than 0.002 and Global model quality scores greater than 0.37. Although the accuracy of our models may still have been limited, the aim of this study was not to describe the proteins' tertiary structures precisely but to assess average structural similarity between the receptors. The error estimates obtained indicate that the models constructed represent reasonably reliable inputs for further phenetic analysis. Images of the predicted protein 3D structures were visualised using PyMOL software v. 1.5 (http://pymol.org/). Protein electrostatic potentials were calculated using

PDB2PQR v. 1.9.0 [71] (PDB2PQR Server, http://nbcr-222.ucsd.edu/pdb2pqr_1.9.0/) based on the PARSE force-field and electrostatic calculation in the APBS web solver [72] (http://www.poissonboltzmann.org/). Surface charge distribution was visualised using Jmol v. 12.2 (http://jmol.sourceforge.net/).

Phylogenetic and phenetic analysis

Alignments of *TLR4*, *TLR5* and *TLR7* CDS were used for phylogenetic analysis using a maximum likelihood (ML) method. As outgroups, we used orthologous human and mouse sequences. FindModel (http://www.hiv.lanl.gov/content/sequence/findmodel/findmodel.html) was used to evaluate the fit of 28 nested nucleotide substitution models to the data, the best model for each alignment being selected on the basis of the Akaike information criterion. ML analyses were performed using PHYML [73], with the NNI algorithm and BIONJ distance-based tree as the starting tree. Bootstrap analysis (with 1000 replicates) was performed to estimate the robustness of internal nodes. The results were visualised in FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/figtree/). A consensus phylogenetic tree including all investigated species (see Additional file 2: Figure S1) was constructed using the avian phylogenetic tool available at http://birdtree.org/ [74]. We used the Hackett backbone [60] as the source tree with 1000 randomly generated trees. The maximum clade credibility tree was produced using the TreeAnnotator v. 1.8.0 tool in BEAST v. 1.8.0 software [75]. Phenetic similarity analysis of the predicted protein secondary and tertiary structures was performed to detect conserved structures in avian TLR. Secondary structures were compared based on alignments obtained using the EMBOSS Needle pairwise alignment tool (http://www.ebi.ac.uk/Tools/psa/emboss_needle/), using chicken (*Gallus gallus*) GaGaTLR sequences as references. To predict 3D structures, we used the adjusted I-TASSER pdb models for structural superposition in the DALI pairwise comparison tool [61]. The pair-wise root mean square deviations (RMSD) metric was used to compare protein structures [76,77] and to construct distance matrices that subsequently served as matrices of Euclidean distances in cluster analysis using an unweighted pair group method with arithmetic mean (UPGMA) method in STATISTICA v. 6.0 (StatSoft, Inc., Tulsa, OK, USA; [78]; for a similar approach see [10]).

Selection analysis

Before testing for selection, all codons containing gaps in any species in the alignment were removed (this applied to only six codons in TLR5; throughout the text, codon positions are numbered according to the chicken GaGaTLR5 sequence (see Additional file 1: Table S1). We used two methods to test for positive selection on individual residues at the interspecific level within the Galloanserae clade, i.e. (1) the hierarchical Bayes (Bayes Empirical Bayes, BEB) approach with implemented Markov chain Monte Carlo routine - PAML (Phylogenetic Analysis by Maximum Likelihood [79]); and (2) FUBAR (A Fast, Unconstrained Bayesian AppRoximation for inferring selection [80]). For PAML (v. 4.6), we used codon-based substitution models (codeml) to identify amino acid sites under positive selection in the CDS comparing the neutral M8a (beta&ω = 1) model with the alternative M8 (beta&ω) model. The likelihood ratio test (LRT) for the comparison of two nested models was calculated using chi-square approximation: $Chi^2 = 2 \times (lnLM8 - lnLM8a)$, where LM8 and LM8a are likelihood values. The degrees of freedom (df) were defined as the difference in the number of parameters in the models used (see Additional file 1: Table S6). If the LRT was less than 0.05, positive selection was considered significant. The BEB approach [81] was used to determine site-specific posterior probabilities of positive selection (≥0.9). FUBAR analysis was performed on the Datamonkey server (http://www.datamonkey.org/, [82]) using a default significance level of posterior probability set at 0.9. In this study, we applied the FUBAR algorithm because it is more robust and much faster than selection analysis based on random effect likelihood (REL methods [80]).

We tested the degree of dissimilarity of amino acid substitutions according to their physiochemical properties using the new PRIME (PRoperty Informed Model of Evolution) tool available on the Datamonkey server [82]. We used the set of five composite physiochemical properties proposed by Atchley et al. [83], i.e. polarity index, secondary structure factor, volume, refractivity/heat capacity and charge/iso-electric point. A change in these properties was considered significant if the posterior probability was greater than 0.9. The evolutionary conservation of amino acid positions was predicted using the ConSurf tool [84], with the assumption that positively selected residues (functionally important for pathogen binding) were the least conserved. For all ConSurf analyses, we used GaGaTLR protein 3D models (obtained as described above) and the LG substitution matrix [85]. A phylogenetic tree of the three *TLR* genes studied including all investigated species in the PAML and ConSurf analyses was constructed as described above (see Additional file 2: Figure S1).

Results and discussion

We were able to verify the homology of all sequences examined (see Additional file 2: Figure S2) and subsequently to assess amino acid identity and similarity (see Additional file 1: Table S2). The phylogeny of the *TLR* genes was consistent with known phylogeny for the Galloanserae clade [86] (see Additional file 2: Figures S1 and

S2). While the size of both TLR4 and TLR7 was uniform (843 aa and 1047 aa, respectively), TLR5 varied in length between 859 aa and 862 aa (see Additional file 1: Table S3 and Additional file 3: Section S1). The orthologues showed little variation in molecular weights, although some differed markedly in their charge at pH = 7 (see Additional file 1: Table S3). Since localised charge variability can influence protein conformation and domain composition and produce variation in ligand-binding features, we examined charge and structural variation in more detail.

Protein structure evolution

TLR proteins are likely to be involved in host-parasite coevolution, and thus shaped by parasite-mediated natural selection [3]. Interspecific differences in TLR protein structure, therefore, may exhibit imprints of structural evolutionary convergence due to selection. In this study, for the first time in birds, we were able to model secondary and tertiary structures of TLR4, TLR5 and TLR7 for all Galloanserae species with currently known CDS. Secondary structures for the proteins predicted by I-TASSER revealed low levels of interspecific structural variability within individual TLR (see Additional file 2: Figure S3). Although two regions of potential functional interest in TLR4 were polymorphic (see Additional file 3: Section S2), the most important structural motifs in all three receptors were conserved in all species analysed, resulting in more than 90% interspecific identity in secondary structure distribution (see Additional file 1: Table S4). As a result, 3D extracellular domain models invariably had a horseshoe-like shape, in which the concave surface comprised β-sheets and the convex surface parallel loops and short helices. The TIR domain (which was modelled separately) had a globular shape. A phenetic analysis of RMSD distances obtained by superposition of the modelled structures revealed that structural variability in the TIR domain of all three receptors was lower than ECD variability (see Additional file 2: Figures S4d, S4e and S4f). This may be a result of higher conservation in the TIR domain when compared to ECD [5], although domain size may also have played a role as RMSD tends to increase with protein size. Avian TLR4 ECD showed a stronger structural resemblance to human TLR4 than mouse TLR4 (see Additional file 2: Figure S4a), which suggests that avian TLR4 may exhibit similar binding features to human TLR4. At present, however, this must remain a hypothesis since current experimental data provide no support [32]. As the RMSD of our superposed models were below 2.5 Å (i.e. deviation between models was lower than the accuracy of individual models), we conclude that TLR4 and TLR5 ECD structural variability within the Galloanserae is generally low and probably unimportant (see Additional file 2: Figures S4a and S4b). Of more interest is the ECD of TLR7, with the phenogram indicating that Anseriform TLR7 exhibits a close structural relationship to human TLR7, while Galliform TLR7 clusters with murine TLR7 (see Additional file 2: Figure S4c). Despite this, the RMSD were too low to indicate any meaningful structural variation. The low TLR structural variability observed in this study is consistent with recent findings for rodents [10].

Surface electrostatic potential

After showing that the protein tertiary structures were highly conserved, we ascertained interspecific differences in surface features by modelling electrostatic potential distribution on the TLR protein surfaces. While TIR domain surface charge distribution remained relatively conserved, we observed high variability in TLR ECD (see Additional file 2: Figure S5). Anseriform TLR ECD differed from ECD of their Galliform counterparts. Species-specific differences were observed even between individual species within the Galliformes. Avian TLR4 ECD surface-charge distribution was clearly distinguishable from that of mammalian (murine and human) TLR4, although surface charge for the predicted avian ligand-binding region [48,49], in particular, was clearly closer to that of murine TLR4 than human TLR4 (Figure 1a). This is consistent with a previous observation indicating that GaGaTLR4 LPS-binding specificity shows greater similarity to that of murine TLR4 than human TLR4 [32]. Variability was much lower in TLR5 ECD; although, once again, avian TLR5 charge distribution at the flagellin-binding site predicted for mammals [54] was closer to that for murine TLR5 than human TLR5 (Figure 1b). As for TLR4, these results are in concordance with the results of previous functional assays [33]. In contrast, for both TLR7 (Figure 1c) and the flagellin-binding interface-A region identified in zebrafish [59], electrostatic potential distribution at the predicted ligand-binding interface [56] resembled that of human TLR more than murine TLR.

Changes in physiochemical features of predicted ligand-binding residues

To further explore potential interspecific variability in the ligand-binding properties of avian TLR proteins, we compared basic chemical features of residues known or predicted to be functionally important in other vertebrate TLR [38,48,49,54,56,59]. These included residues directly involved in ligand binding, TLR homodimerization and MD-2 heterodimerization. At most of the predicted functional sites in avian TLR, similar chemical properties were preserved as in mammalian TLR (see Additional file 1: Table S5). Such conservation was particularly visible in TLR7, which suggests that most receptor-ligand interactions in TLR7 are highly conserved (see Additional file 3: Section S3). In TLR5, we identified substantial residue differences at the binding region previously described in zebrafish [59] as separating fish from amniotes (see Additional

Vinkler *et al. Genetics Selection Evolution* 2014, **46**:72
http://www.gsejournal.org/content/46/1/72

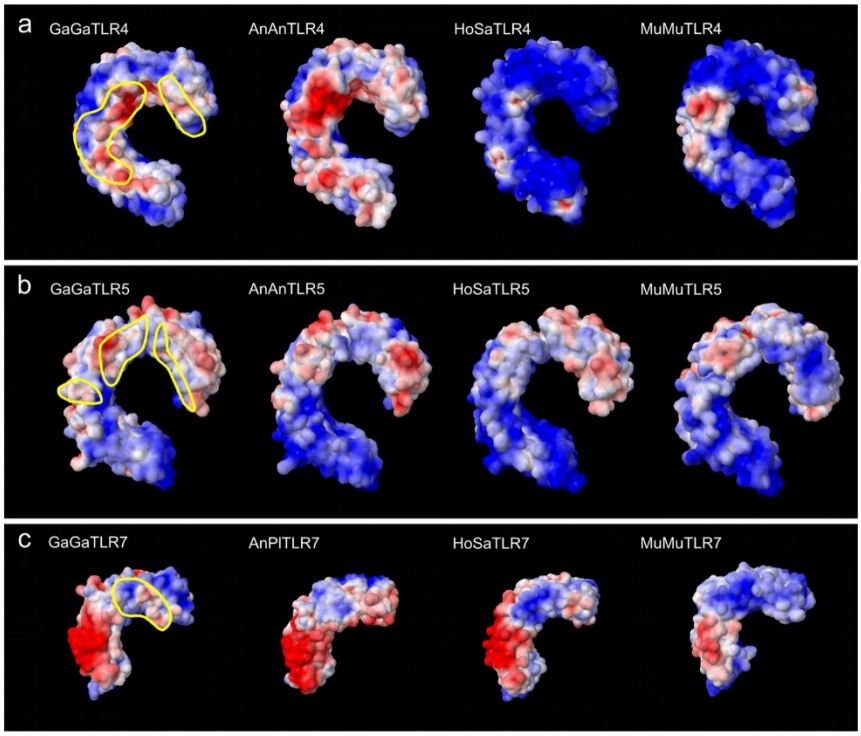


Figure 1 Differences in ECD surface electrostatic potentials between birds and mammals (for a full comparison see Additional file 2: Figure S5). (a) TLR4, **(b)** TLR5, **(c)** TLR7; positive surface charge is highlighted in red, negative charge in blue; the predicted functional sites in ECD are outlined in yellow in the GaGaTLR models; GaGa = chicken *Gallus gallus*, AnAn = goose *Anser anser*, AnPl = duck *Anas platyrhynchos*, HoSa = human *Homo sapiens*, MuMu = mouse *Mus musculus*.

file 1: Table S5). Primary binding interfaces A and B, as well as both dimerization interfaces, appear to be only modestly conserved, indicating that flagellin binding probably differs in fish and amniotes. This is further supported by the fact that avian TLR5 sites generally retain their physiochemical properties at those binding residues predicted for mammals. Intriguingly, we identified potentially important changes in amino acid properties at several TLR4 functional sites, which suggests variability in LPS binding and MD-2 dimerization (see Additional file 1: Table S5). In a number of cases, residue changes precluded the existence of charge (positions 268 and 397) or hydrophobic interactions (positions 449 and 472) known for human TLR4 [49,87]. Interestingly, all avian species possess arginine at position 393; in this respect, avian TLR4 is identical to equine TLR4 [38] but distinct from human TLR4. This suggests that lipid IVa in birds serves as an agonist triggering a TLR4-MD-2-mediated immune response, just as it does in horses [38]. Galliformes also display systematic differences from Anseriformes at several positions (see Additional file 3: Section S3). Thus, our results suggest that the ligand-binding features of avian TLR differ not only from mammalian TLR but also between the various avian taxa. This is in concordance with previously reported experimental results [32,33].

Analysis of positive selection

Chicken *TLR* genes show remarkable differences in the level of sequence polymorphism, most likely as a result of distinct positive and negative selection [6,20,31]. This is further supported by analysis of partial *TLR* CDS in previous studies on avian species [12,13]. In this study, we investigated signatures for positive, diversifying selection acting on individual positions in TLR4, TLR5 and TLR7 within the Galloanserae lineage. We identified one positively selected site in TLR4, 11 sites in TLR5 and eight sites in TLR7, using three different methods (Figures 2, 3 and 4). Two of the approaches used were based on site-by-site synonymous and non-synonymous rate ratios (PAML and FUBAR), which detected seven positively selected sites in both TLR5 and TLR7 but no positively selected sites in TLR4 (see Additional file 1: Table S7). The lack of evidence for positive selection in TLR4 is probably due to the low power of analysis because of a limited number of distantly-related species (all species of the genus *Gallus* are genetically closely-related, and thus display similar genetic

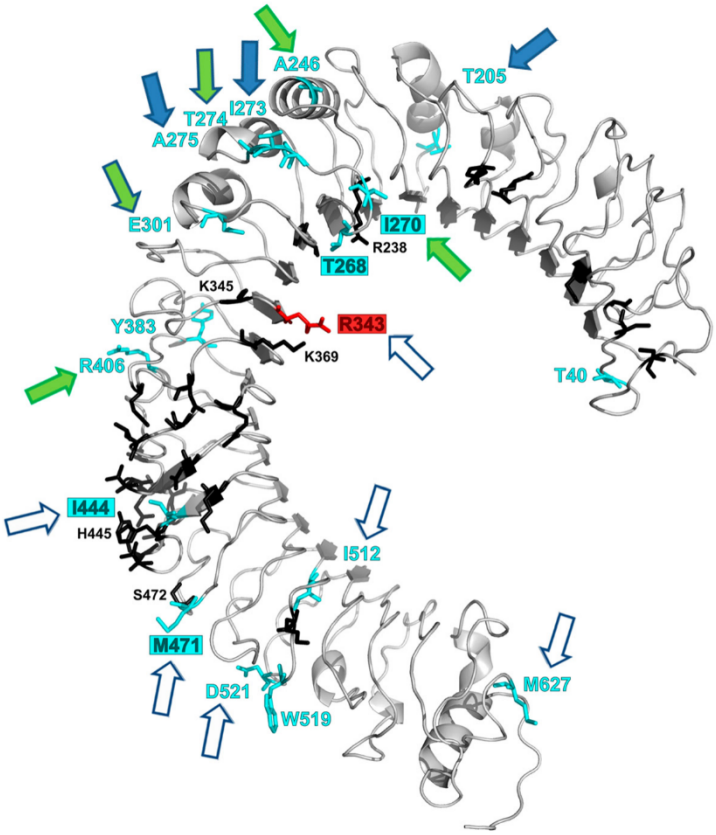


Figure 2 Projection of detected positively selected sites (PAML/FUBAR/PRIME) and evolutionarily non-conservative sites (ConSurf) onto a 3D protein model of chicken GaGaTLR4 ECD. Positively selected sites are highlighted in red, evolutionarily non-conservative sites in blue and functionally important sites known from mammalian studies [38,48-50,87] in black (only positions lying in close proximity to the selected or non-conservative sites are numbered); GaGaTLR residue numbering has been adopted; selected sites or non-conservative residues located in close proximity to any of the predicted functional sites are indicated by a filled red or blue rectangle, respectively; arrows indicate positively selected sites concurrent with those identified in previous studies (filled green = avian, filled blue = mammalian, filled green with a blue border = avian and mammalian, open blue = site in neighbourhood); position 268 has been identified in mice and humans as a site directly responsible for MD-2 dimerization and LPS binding.

sequences); hence, in this case, the lack of evidence should be treated as a probable methodological artefact. The third method which used a physiochemical comparative approach (PRIME), detected positive selection in all three proteins, with one positively selected site in TLR4, five sites in TLR5 and two sites in TLR7 (see Additional file 1: Table S8). Of the 20 positively selected sites detected in total, only one was located in the transmembrane domain, the others being located in ECD. No positive selection signatures were detected in the intracellular TIR domain, which suggests that positive selection acts mainly on ligand-binding regions. This confirms the results of similar analyses undertaken on mammals [7,9,10,88,89] and amniotes [22]. Projection of positively selected and functionally important sites onto 3D protein models of GaGaTLR revealed changes at six positions (TLR4: 343; TLR5: 180, 209, 342, 379; TLR7: 26) that could directly influence receptor expression or function (Figures 2, 3 and 4, see Additional file 1: Tables S7 and S8 and Additional file 3: Section S4). As in other birds [12,13], Galloanserae TLR5 displayed a relatively high accumulation of codons that exhibit positive selection. This is in concordance with the results of Wlasiuk et al. [7] reported for primates. On the other hand, TLR7 has previously been shown to evolve mainly under purifying selection in birds; with no or only limited positively selected sites in the predicted ligand-binding region [9,10,12,13]. In this study, we were able to detect some positively selected sites (Figure 4; see Additional file 1: Tables S7 and S8); however, given their locations (except for position 26), these sites appear to have only limited functional importance and their impact remains unclear.

While the vast majority of residues in TLR were evolutionarily highly conservative, consistent with purifying selection on TLR [6,10,13], ConSurf analysis revealed 26

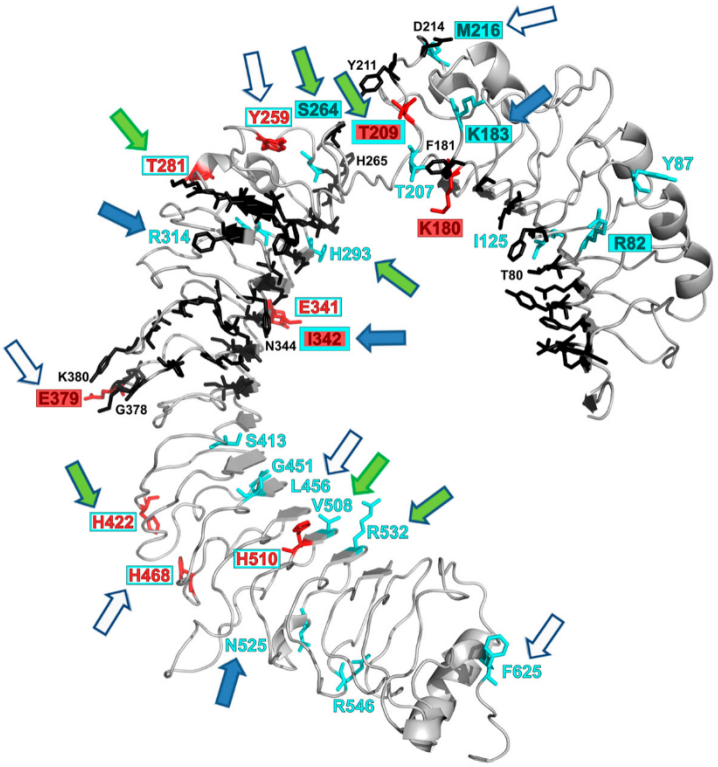


Figure 3 Projection of detected positively selected sites (PAML/FUBAR/PRIME) and evolutionarily non-conservative sites (ConSurf) onto a 3D protein model of chicken GaGaTLR5 ECD. Positively selected sites are highlighted in red, evolutionarily non-conservative sites in blue and functionally important sites known from fish [59], and predicted for mammals [54], are in black (only positions lying in close proximity to the selected or non-conservative sites are numbered); GaGaTLR residue numbering has been adopted; selected sites or non-conservative residues located in close proximity to any of the predicted functional sites are indicated by a filled red or blue rectangle, respectively; overlap of the identified selected positions with non-conservative residues is indicated by a red number in an open blue rectangle; arrows indicate positively selected sites concurrent with those identified in previous studies (filled green = avian, filled blue = mammalian, filled green with blue border = avian and mammalian, open blue = site in neighbourhood); position 209 has been identified as both positively selected and non-conservative; in addition, this residuum (209) is also known as a flagellin-binding site in fish, as are residues 183 and 379; position 342 is both positively-selected and non-conservative and lies in close proximity to residuum 344 which has been recognised as a possible flagellin-binding site in mammals.

reliable evolutionarily non-conservative sites in TLR4, 28 in TLR5 and 25 in TLR7 (Figures 2, 3 and 4 and (see Additional file 1: Table S9 for detailed information). These may also represent positively selected sites since functionally important positively-selected residues are also predicted to be the least conserved. This is supported by the location of six of the non-conservative sites in the functionally important TLR4 signal peptide, four in the predicted TLR4-MD-2-binding regions [48,49], three in the LPS-binding regions [49,50], two in the TLR4-TLR4-dimerization regions [49] and one in the transmembrane region (see Additional file 3: Section S4). One of the non-conservative positions identified in our Galloanserae dataset (383) has previously been identified as a possible single nucleotide polymorphism (SNP) responsible for *Salmonella* resistance in the domestic chicken [16]. In TLR5, one non-conservative site was situated in the signal peptide, five in the flagellin-binding residues identified in fish [59], two in flagellin-binding residues predicted in mammals [54] and two in the trans-membrane region. Most of the non-conservative sites in TLR7 were situated in the excised region [69] (and hence likely to be functionally unimportant), with just one in the potentially relevant signal peptide and one in the TIR domain.

Intriguingly, both the positively-selected and non-conservative sites identified in this study partly match the positively selected sites detected in previous studies in other vertebrate taxa [7,9,10,12,13,22,90,91] (Figures 2, 3 and 4, Additional file 1: Table S10 and Additional file 3: Section S4). In TLR4 and TLR5, 15 and 20 candidate sites, respectively, were either precisely the same or lay in close proximity (up to 2 aa) to a site identified in previous studies. In contrast, only seven sites were close to, or the same as, previously detected sites in TLR7. Most importantly, there was consensus for positive selection on site 270 in the predicted MD-2-binding region in TLR4 in birds, and

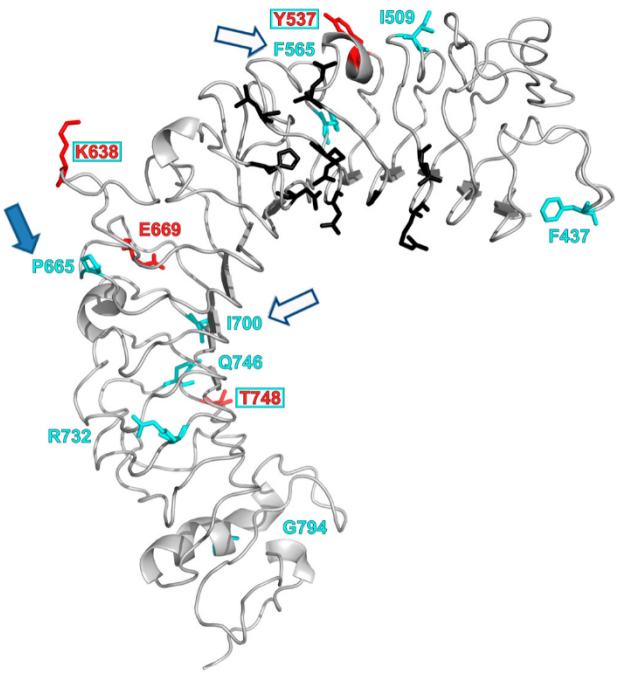


Figure 4 Projection of detected positively selected sites (PAML/FUBAR/PRIME) and evolutionarily non-conservative sites (ConSurf) onto a 3D protein model of chicken GaGaTLR7 ECD. Positively selected sites are highlighted in red, evolutionarily non-conservative sites in blue and functionally important sites predicted for mammals [56] are in black; GaGaTLR residue numbering has been adopted; overlap of the identified selected positions with non-conservative residues is indicated by a red number in an open blue rectangle; arrows indicate positively selected sites concurrent with those identified in previous studies (filled blue = mammalian, open blue = site in neighbourhood).

on sites 183, 209, 264 and 342 in TLR5 both in birds and mammals, all of which are predicted flagellin binding sites or sites in close proximity to these binding sites (see Additional file 3: Section S4). No consensus site was located in the predicted ligand-binding region in TLR7. Independent concordance between the results of selection analysis in several studies that cover different taxa, and topological agreement between the identified sites and predicted functionally important regions, strongly support the proposed importance of diversifying selection or positive selection in the evolution of these residues. In human TLR4, for example, it has been demonstrated that even SNP located far from MD-2-dimerization or LPS-binding sites can modify responsiveness to LPS [92] or influence the binding of other ligands. Thus, other consensus selected sites identified in TLR4 (205, 246, 273–275, 301 and 406, Figure 2) and TLR5 (281, 293, 314, 422, 508, 525 and 532, Figure 3) may play an important role in the evolution of TLR-ligand binding in birds (see Additional file 1: Table S10). However, the potentially important consensus sites identified in TLR7, are located mainly in the excised region [69] (residues at positions 39, 99 and 383) and do not appear to influence receptor function. The only exception to this is site 665 which lies, however, outside the predicted ligand-binding region. Given the generally low concordance in predicted positively-selected sites between studies, we suggest that many of the sites reported for TLR7 represent false-positive predictions resulting from the method applied (most sites were identified by REL or FEL; see [80] for discussion).

In concordance with previous studies, we found stronger signatures of positive selection acting on the ligand-binding regions in TLR4 [9] and TLR5 [7] than in TLR7 [12]. There are several possible explanations for this. First, bacterial-sensing TLR such as TLR4 and TLR5 recognise structurally variable ligands [35-37], while viral-sensing TLR7 detects structurally invariant RNA molecules regardless of their precise sequence [93]. Hence, TLR7 is likely to have evolved mainly under purifying selection, while TLR4 and TLR5 evolved mainly under diversifying selection. Furthermore, since TLR4 is capable of recognising several unrelated ligands [3], it is also possible that this receptor evolved more rapidly. Second, there appears to be greater redundancy in bacterial recognition than viral recognition. In TLR5, for example, several studies have proposed relaxed purifying selection [7,9], while nonsense stop-codons have been described in both birds [94] and mammals [7]. This may be a result of the presence of other flagellin receptors which may compensate for malfunction of TLR5 (discussed in [94]). Finally, the limited number of positively

selected sites detected in the TLR7 ligand-binding region may result from the limited knowledge about its precise location. While the precise mechanism of ligand-binding has been described for both TLR4 [48-50] and TLR5 [59], only crude predictions are available for TLR7 [56]. Although all these explanations may be relevant to some extent, the relatively low concordance between results of multiple studies involving TLR7 (see Additional file 1: Table S10) tends to support the view that there is, indeed, only weak positive selection acting on TLR7.

Conclusion

Interspecific comparisons within the Galloanserae clade revealed relatively high sequence variability in all three TLR investigated. Such variation has been shown to influence the physiochemical properties of proteins. Despite high tertiary-structure conservation, evolutionary changes were manifested by alterations to protein surface characteristics, such as changes in electrostatic potential distribution. Importantly, not only does surface charge in Galloanserae birds differ distinctly from that of mammals, to some extent there are also distinct differences observed within the clade. These variations most likely affect receptor binding features, a theory that is consistent with the idea of a host-pathogen evolutionary arms race [95] in which any adaptation enabling a pathogen to escape host immunity leads inevitably to a counter-adaptation in host receptors that again enables pathogen detection. Co-evolution of this kind has previously been described in both human TLR4 and TLR4 of the bacterium *Pseudomonas aeruginosa* [39]. In fact, there are now numerous examples of individual host TLR adaptations to pathogens known (summarised, for example, in [3]). What we presently lack, however, is a functional understanding of evolutionarily-tested beneficial innovations present at the interspecific level in vertebrates. In this study, selection analysis identified a number of positively selected sites, mostly in the ligand-binding ECD. As in previous studies, we observed stronger positive selection acting on the ligand-binding regions of TLR4 [9] and TLR5 [7] compared to TLR7 [12]. Sites subjected to selection in TLR4 and TLR5 were frequently located either precisely in, or in close topological proximity to, ligand-binding sites known or predicted in mammalian or fish TLR. This suggests that, although avian TLR may be differentially adapted to pathogen-derived ligand recognition compared to that of other vertebrate species, identical regions are responsible for ligand binding. We suggest that future investigations in this field should focus on functional testing of evolutionarily relevant substitutions detected by selection analyses. Based on the evidence summarised above, we propose several sites for more detailed investigation. In particular, sites 270, 343, 444 and 471 in TLR4 and sites 183, 209, 216, 264, 342 and 379 in TLR5 represent key candidates for further research on the functional significance of selection acting on TLR in birds. We suggest that site 268 (and possibly also site 383) in TLR4 and site 180 in TLR5 may be of particular evolutionary importance in Galloanserae birds since no selection on these sites has previously been observed in either mammals or other birds. Furthermore, while we do not know the functional importance of the sites, concordance between selection analyses in both birds and mammals suggests that special attention should be paid to positions 244–246, 273–275, 301 in TLR4 and 293, 294, 314, 342, 422, 525, 532, 533 in TLR5.

Taken together, our results have identified variability in Galloanserae birds that very likely results from pathogen-mediated evolution of species-specific TLR binding features. To our knowledge, this is the first study to depict evolutionary pressures on Galloanserae TLR and to estimate the validity of current knowledge on TLR function (based on mammalian and chicken models) for non-model species of this clade. Functional testing of the importance of individual sites (such as that performed by Walsh et al. [38] in mammals) should provide novel understanding of evolutionary mechanisms increasing resistance to pathogens in avian species. Any knowledge gained would be of great practical relevance, with applications in animal breeding for increased resistance to diseases.

Additional files

Additional file 1: Table S1. Specification of sequences obtained from NCBI GenBank. Detailed information on NCBI GenBank sequences analysed in this study. **Table S2.** Homology of GaGaTLRs with other Galloanserae TLR molecules. Nucleotide and amino acid identity and similarity calculated in NCBI BLAST. **Table S3.** Variation in structure and physical features of Galloanserae TLR. Information of amino protein length, molecular weight, charge, number of LRR and signal peptide length in Galloanserae TLR. **Table S4.** Identity in TLR secondary structures within Galloanserae. Whole protein, extracellular domain and intracellular domain identity of secondary structure motives within alignment of Galloanserae TLR. **Table S5.** List of binding residues identified in other vertebrates (fish and mammals) and their conservation within the Galloanserae lineage. Variability of residues at the predicted functional sites in Galloaserae birds with the prediction of changes in amino acid binding features compared to humans, mice and fish. **Table S6.** PAML codeml site model test for positive selection within Galloanserae TLR. Detailed information on parameters of the PAML model tests for positive selection within Galloanserae TLR. **Table S7.** Positively selected sites identified by PAML and FUBAR. Summary of positively selected sites identified by PAML and FUBAR with details on their location and p-values. **Table S8.** Positively selected sites identified by PRIME. Summary of positively selected sites identified by PRIME analysis with details on their location, p-values and changing properties. **Table S9.** Evolutionarily non-conservative sites identified by ConSurf. Summary of evolutionarily non-conservative sites identified by the ConSurf analysis with details on their location and conservation scores. **Table S10.** Co-location of sites under positive selection in TLR4, TLR5 and TLR7. Agreement of sites under positive selection in TLR4, TLR5 and TLR7 identified in this study with the results obtained by other evolutionarily studies aimed at detecting selection in TLR in vertebrates.

Additional file 2: Figure S1. Phylogeny of *TLR* genes within Galloanserae. Maximum likelihood phylogeny of nucleotide sequences of

TLR4 (a), *TLR5* (b) and *TLR7* (c) in Galloanserae species calculated in PHYML. **Figure S2.** Phylogenetic relationships between the investigated species. Phylogenetic tree of the investigated species constructed based on consensus avian phylogenetic tool. **Figure S3.** Secondary structure alignment of Galloanserae TLR. Alignment of secondary structure motifs (α-helices, β-sheets and connecting sequences) in Galloanserae TLR. **Figure S4.** 3D structural similarity of Galloanserae TLR tertiary structures. Phenograms representing the 3D structural similarity of models of Galloanserae TLR tertiary structures. **Figure S5.** Species-specific differences in surface electrostatic potential. Visualisation of species-specific differences in surface electrostatic potential on 3D tertiary structure models.

Additional file 3: Section S1. Variability in Galloanserae TLR basic protein features. Description of basic protein features in Galloanserae TLR. **Section S2.** Evolution of TLR secondary structures within the Galloanserae lineage. Description of secondary structure similarities in Galloanserae TLR. **Section S3.** Variability in features of predicted ligand-binding residues. Detailed information on amino acid binding features at the functional sites known in other vertebrates. **Section S4.** Analysis of positive selection. Detailed information and discussion on the results of the positive selection analysis.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
MV conceived and designed the study, participated in the protein structure analysis and drafted the manuscript. HB carried out the selection analysis and participated in the protein structure analysis. JB performed the phylogenetic and phenetic analyses. All authors read and approved the final manuscript.

Acknowledgements
We are grateful to Marian Novotný, Zuzana Bainová and Anežka Pojezdná for their help and advice, and Dagmar Vinklerová and Kevin Roche for language correction. This study was supported through Czech Science Foundation Projects P505/10/1871 and P502/12/P179 and through Institutional Research Support (Grant SVV 260 087/2014). The work was also partially supported by the Ministry of Education, Youth and Sport of the Czech Republic within the NextGenProject (CZ.1.07/2.3.00/20.0303).

Author details
[1]Department of Zoology, Faculty of Science, Charles University in Prague, Praha, Czech Republic. [2]Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, v.v.i., Brno, Czech Republic.

Received: 11 June 2014 Accepted: 10 October 2014
Published online: 12 November 2014

References
1. Palsson-McDermott EM, O'Neill LAJ: **Building an immune system from nine domains.** *Biochem Soc Trans* 2007, **35:**1437–1444.
2. Akira S, Uematsu S, Takeuchi O: **Pathogen recognition and innate immunity.** *Cell* 2006, **124:**783–801.
3. Vinkler M, Albrecht T: **The question waiting to be asked: Innate immunity receptors in the perspective of zoological research.** *Folia Zool* 2009, **58:**15–28.
4. Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A: **The evolution of vertebrate Toll-like receptors.** *Proc Natl Acad Sci U S A* 2005, **102:**9577–9582.
5. Temperley ND, Berlin S, Paton IR, Griffin DK, Burt DW: **Evolution of the chicken Toll-like receptor gene family: A story of gene gain and gene loss.** *BMC Genomics* 2008, **9:**62.
6. Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N: **Molecular evolution of vertebrate Toll-like receptors: Evolutionary rate difference between their leucine-rich repeats and their TIR domains.** *Gene* 2012, **503:**235–243.
7. Wlasiuk G, Khan S, Switzer WM, Nachman MW: **A history of recurrent positive selection at the Toll-like receptor 5 in primates.** *Mol Biol Evol* 2009, **26:**937–949.
8. Huang YH, Temperley ND, Ren LM, Smith J, Li N, Burt DW: **Molecular evolution of the vertebrate TLR1 gene family - a complex history of gene duplication, gene conversion, positive selection and co-evolution.** *BMC Evol Biol* 2011, **11:**149.
9. Areal H, Abrantes J, Esteves PJ: **Signatures of positive selection in Toll-like receptor (TLR) genes in mammals.** *BMC Evol Biol* 2011, **11:**368.
10. Fornuskova A, Vinkler M, Pages M, Galan M, Jousselin E, Cerqueira F, Morand S, Charbonnel N, Bryja J, Cosson J-F: **Contrasted evolutionary histories of two Toll-like receptors (*Tlr4* and *Tlr7*) in wild rodents (MURINAE).** *BMC Evol Biol* 2013, **13:**194.
11. Cormican P, Lloyd AT, Downing T, Connell SJ, Bradley D, O'Farrelly C: **The avian Toll-Like receptor pathway-Subtle differences amidst general conformity.** *Dev Comp Immunol* 2009, **33:**967–973.
12. Alcaide M, Edwards SV: **Molecular evolution of the Toll-like receptor multigene family in birds.** *Mol Biol Evol* 2011, **28:**1703–1715.
13. Grueber CE, Wallis GP, Jamieson IG: **Episodic positive selection in the evolution of avian toll-like receptor innate immunity genes.** *PLoS One* 2014, **9:**e89632.
14. Brownlie R, Allan B: **Avian toll-like receptors.** *Cell Tissue Res* 2011, **343:**121–130.
15. Iqbal M, Philbin VJ, Smith AL: **Expression patterns of chicken Toll-like receptor mRNA in tissues, immune cell subsets and cell lines.** *Vet Immunol Immunopathol* 2005, **104:**117–127.
16. Leveque G, Forgetta V, Morroll S, Smith AL, Bumstead N, Barrow P, Loredo-Osti JC, Morgan K, Malo D: **Allelic variation in TLR4 is linked to susceptibility to *Salmonella enterica* serovar *typhimurium* infection in chickens.** *Infect Immun* 2003, **71:**1116–1124.
17. Fukui A, Inoue N, Matsumoto M, Nomura M, Yamada K, Matsuda Y, Toyoshima K, Seya T: **Molecular cloning and functional characterization of chicken Toll-like receptors - A single chicken toll covers multiple molecular patterns.** *J Biol Chem* 2001, **276:**47143–47149.
18. Lynn DJ, Lloyd AT, O'Farrelly C: **In silico identification of components of the Toll-like receptor (TLR) signaling pathway in clustered chicken expressed sequence tags (ESTs).** *Vet Immunol Immunopathol* 2003, **93:**177–184.
19. Philbin VJ, Iqbal M, Boyd Y, Goodchild MJ, Beal RK, Bumstead N, Young J, Smith AL: **Identification and characterization of a functional, alternatively spliced Toll-like receptor 7 (TLR7) and genomic disruption of TLR8 in chickens.** *Immunology* 2005, **114:**507–521.
20. Yilmaz A, Shen SX, Adelson DL, Xavier S, Zhu JJ: **Identification and sequence analysis of chicken Toll-like receptors.** *Immunogenetics* 2005, **56:**743–753.
21. Smith J, Speed D, Law AS, Glass EJ, Burt DW: **In-silico identification of chicken immune-related genes.** *Immunogenetics* 2004, **56:**122–133.
22. Vinkler M, Bryjova A, Albrecht T, Bryja J: **Identification of the first Toll-like receptor gene in passerine birds: *TLR4* orthologue in zebra finch (*Taeniopygia guttata*).** *Tissue Antigens* 2009, **74:**32–41.
23. Gopinath VP, Biswas M, Raj GD, Raja A, Kumanan K, Elankumaran S: **Molecular cloning and tissue-specific expression of Toll-like receptor 5 gene from turkeys.** *Avian Dis* 2011, **55:**480–485.
24. Elfeil WK, Abouelmaatti RR, Sun CJ, Han WY, Li XK, Ma JS, Lei LC, Liu SS, Yang YJ, Wang Y, Mandour M, Fawzy M, Shalaby MN: **Identification, cloning, expression of a novel functional *Anas platyrhynchos* mRNA TLR4.** *J Anim Vet Adv* 2012, **11:**1727–1733.
25. MacDonald MRW, Xia JG, Smith AL, Magor KE: **The duck Toll like receptor 7: Genomic organization, expression and function.** *Mol Immunol* 2008, **45:**2055–2061.
26. Jia H, Li G, Li J, Tian Y, Wang D, Shen J, Tao Z, Xu J, Lu L: **Cloning, expression and bioinformatics analysis of the duck TLR 4 gene.** *Br Poult Sci* 2012, **53:**190–197.
27. Wang F, Lu LZ, Yuan H, Tian Y, Li JJ, Shen JD, Tao ZR, Fu Y: **cDNA cloning, characterization and expression analysis of Toll-like receptor 4 gene in goose.** *Can J Anim Sci* 2011, **91:**371–377.
28. Ruan WK, Wu YH, Zheng SJJ: **Different genetic patterns in avian Toll-like receptor (TLR)5 genes.** *Mol Biol Rep* 2012, **39:**3419–3426.
29. Ramasamy KT, Verma P, Reddy MR, Murugesan S: **Molecular characterization of coding sequence and mRNA expression pattern of Toll-like receptor 15 in Japanese quail (*Coturnix japonica*) and indigenous chicken breeds (Aseel and Kadaknath).** *J Poult Sci* 2011, **48:**168–175.
30. Uno Y, Usui T, Fujimoto Y, Ito T, Yamaguchi T: **Quantification of interferon, interleukin, and Toll-like receptor 7 mRNA in quail splenocytes using real-time PCR.** *Poult Sci* 2012, **91:**2496–2501.

31. Downing T, Lloyd AT, O'Farrelly C, Bradley DG: **The differential evolutionary dynamics of avian cytokine and TLR gene classes.** *J Immunol* 2010, **184:**6993–7000.
32. Keestra AM, van Putten JPM: **Unique properties of the chicken TLR4/MD-2 complex: Selective lipopolysaccharide activation of the MyD88-dependent pathway.** *J Immunol* 2008, **181:**4354–4362.
33. Keestra AM, de Zoete MR, van Aubel RAMH, van Putten JPM: **Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin.** *Mol Immunol* 2008, **45:**1298–1307.
34. Brownlie R, Zhu JZ, Allan B, Mutwiri GK, Babiuk LA, Potter A, Griebel P: **Chicken TLR21 acts as a functional homologue to mammalian TLR9 in the recognition of CpG oligodeoxynucleotides.** *Mol Immunol* 2009, **46:**3163–3170.
35. DeMarco ML, Woods RJ: **From agonist to antagonist: Structure and dynamics of innate immune glycoprotein MD-2 upon recognition of variably acylated bacterial endotoxins.** *Mol Immunol* 2011, **49:**124–133.
36. Marr N, Novikov A, Hajjar AM, Caroff M, Fernandez RC: **Variability in the lipooligosaccharide structure and endotoxicity among *Bordetella pertussis* strains.** *J Infect Dis* 2010, **202:**1897–1906.
37. Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, Aderem A: **Evasion of Toll-like receptor 5 by flagellated bacteria.** *Proc Natl Acad Sci U S A* 2005, **102:**9247–9252.
38. Walsh C, Gangloff M, Monie T, Smyth T, Wei B, McKinley TJ, Maskell D, Gay N, Bryant C: **Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa.** *J Immunol* 2008, **181:**1245–1254.
39. Hajjar AM, Ernst RK, Tsai JH, Wilson CB, Miller SI: **Human Toll-like receptor 4 recognizes host-specific LPS modifications.** *Nat Immunol* 2002, **3:**354–359.
40. Lien E, Means TK, Heine H, Yoshimura A, Kusumoto S, Fukase K, Fenton MJ, Oikawa M, Qureshi N, Monks B, Finberg RW, Ingalls RR, Golenbock DT: **Toll-like receptor 4 imparts ligand-specific recognition of bacterial lipopolysaccharide.** *J Clin Invest* 2000, **105:**497–504.
41. Poltorak A, Ricciardi-Castagnoli P, Citterio S, Beutler B: **Physical contact between lipopolysaccharide and Toll-like receptor 4 revealed by genetic complementation.** *Proc Natl Acad Sci U S A* 2000, **97:**2163–2167.
42. Calenge F, Kaiser P, Vignal A, Beaumont C: **Genetic control of resistance to salmonellosis and to *Salmonella* carrier-state in fowl: a review.** *Genet Sel Evol* 2010, **42:**11.
43. Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, Lee HY, Lee JO: **Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide.** *Cell* 2007, **130:**1071–1082.
44. Kang JY, Nan X, Jin MS, Youn SJ, Ryu YH, Mah S, Han SH, Lee H, Paik SG, Lee JO: **Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer.** *Immunity* 2009, **31:**873–884.
45. Bell JK, Botos I, Hall PR, Askins J, Shiloach J, Segal DM, Davies DR: **The molecular structure of the Toll-like receptor 3 ligand-binding domain.** *Proc Natl Acad Sci U S A* 2005, **102:**10976–10980.
46. Choe J, Kelker MS, Wilson IA: **Crystal structure of human Toll-like receptor 3 (TLR3) ectodomain.** *Science* 2005, **309:**581–585.
47. Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, Segal DM, Davies DR: **Structural basis of Toll-like receptor 3 signaling with double-stranded RNA.** *Science* 2008, **320:**379–381.
48. Kim HM, Park BS, Kim JI, Kim SE, Lee J, Oh SC, Enkhbayar P, Matsushima N, Lee H, Yoo OJ, Lee JO: **Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist eritoran.** *Cell* 2007, **130:**906–917.
49. Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO: **The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex.** *Nature* 2009, **458:**1191–1195.
50. Ohto U, Fukase K, Miyake K, Shimizu T: **Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2.** *Proc Natl Acad Sci U S A* 2012, **109:**7421–7426.
51. Xu YW, Tao X, Shen BH, Horng T, Medzhitov R, Manley JL, Tong L: **Structural basis for signal transduction by the Toll/interleukin-1 receptor domains.** *Nature* 2000, **408:**111–115.
52. Nyman T, Stenmark P, Flodin S, Johansson I, Hammarstrom M, Nordlund P: **The crystal structure of the human Toll-like receptor 10 cytoplasmic domain reveals a putative signaling dimer.** *J Biol Chem* 2008, **283:**11861–11865.
53. Mizel SB, West AP, Hantgan RR: **Identification of a sequence in human Toll-like receptor 5 required for the binding of Gram-negative flagellin.** *J Biol Chem* 2003, **278:**23624–23629.
54. Andersen-Nissen E, Smith KD, Bonneau R, Strong RK, Aderem A: **A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin.** *J Exp Med* 2007, **204:**393–403.
55. Wei TD, Gong J, Rossle SC, Jamitzky F, Heckl WM, Stark RW: **A leucine-rich repeat assembly approach for homology modeling of the human TLR5-10 and mouse TLR11-13 ectodomains.** *J Mol Model* 2011, **17:**27–36.
56. Wei TD, Gong J, Jamitzky F, Heckl WM, Stark RW, Rossle SC: **Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains.** *Protein Sci* 2009, **18:**1684–1691.
57. Kubarenko AV, Ranjan S, Colak E, George J, Frank M, Weber ANR: **Comprehensive modeling and functional analysis of Toll-like receptor ligand-recognition domains.** *Protein Sci* 2010, **19:**558–569.
58. Zhou KF, Kanai R, Lee P, Wang HW, Modis Y: **Toll-like receptor 5 forms asymmetric dimers in the absence of flagellin.** *J Struct Biol* 2012, **177:**402–409.
59. Yoon SI, Kurnasov O, Natarajan V, Hong MS, Gudkov AV, Osterman AL, Wilson IA: **Structural basis of TLR5-flagellin recognition and signaling.** *Science* 2012, **335:**859–864.
60. Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han KL, Harshman J, Huddleston CJ, Marks BD, Miglia KJ, Moore WS, Sheldon FH, Steadman DW, Witt CC, Yuri T: **A phylogenomic study of birds reveals their evolutionary history.** *Science* 2008, **320:**1763–1768.
61. Holm L, Kaariainen S, Rosenstrom P, Schenkel A: **Searching protein structure databases with DaliLite v. 3.** *Bioinformatics* 2008, **24:**2780–2781.
62. Suyama M, Torrents D, Bork P: **PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments.** *Nucleic Acids Res* 2006, **34:**W609–W612.
63. Letunic I, Doerks T, Bork P: **SMART 7: recent updates to the protein domain annotation resource.** *Nucleic Acids Res* 2012, **40:**D302–D305.
64. Offord V, Coffey TJ, Werling D: **LRRfinder: A web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database.** *Dev Comp Immunol* 2010, **34:**1035–1041.
65. Cserzo M, Eisenhaber F, Eisenhaber B, Simon I: **TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMfilter.** *Bioinformatics* 2004, **20:**136–137.
66. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nat Methods* 2011, **8:**785–786.
67. Roy A, Kucukural A, Zhang Y: **I-TASSER: a unified platform for automated protein structure and function prediction.** *Nat Protoc* 2010, **5:**725–738.
68. Zhang Y: **I-TASSER server for protein 3D structure prediction.** *BMC Bioinformatics* 2008, **9:**40.
69. Ewald SE, Lee BL, Lau L, Wickliffe KE, Shi G-P, Chapman HA, Barton GM: **The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor.** *Nature* 2008, **456:**658–662.
70. McGuffin LJ: **The ModFOLD server for the quality assessment of protein structural models.** *Bioinformatics* 2008, **24:**586–587.
71. Dolinsky TJ, Nielsen JE, McCammon JA, Baker NA: **PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations.** *Nucleic Acids Res* 2004, **32:**W665–W667.
72. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA: **Electrostatics of nanosystems: Application to microtubules and the ribosome.** *Proc Natl Acad Sci U S A* 2001, **98:**10037–10041.
73. Guindon S, Lethiec F, Duroux P, Gascuel O: **PHYML Online - a web server for fast maximum likelihood-based phylogenetic inference.** *Nucleic Acids Res* 2005, **33:**W557–W559.
74. Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO: **The global diversity of birds in space and time.** *Nature* 2012, **491:**444–448.
75. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007, **7:**214.
76. Cozzetto D, Kryshtafovych A, Fidelis K, Moult J, Rost B, Tramontano A: **Evaluation of template-based models in CASP8 with standard measures.** *Proteins* 2009, **77:**18–28.
77. Hasegawa H, Holm L: **Advances and pitfalls of protein structural alignment.** *Curr Opin Struct Biol* 2009, **19:**341–348.
78. Kalinowski ST: **How well do evolutionary trees describe genetic relationships among populations?** *Heredity* 2009, **102:**506–513.
79. Yang Z: **PAML 4: Phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24:**1586–1591.
80. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K: **FUBAR: a fast, unconstrained bayesian approximation for inferring selection.** *Mol Biol Evol* 2013, **30:**1196–1205.

81. Yang ZH, Wong WSW, Nielsen R: **Bayes empirical Bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22:**1107–1118.
82. Kosakovsky Pond SL, Frost SDW: **Datamonkey: rapid detection of selective pressure on individual sites of codon alignments.** *Bioinformatics* 2005, **21:**2531–2533.
83. Atchley WR, Zhao JP, Fernandes AD, Druke T: **Solving the protein sequence metric problem.** *Proc Natl Acad Sci U S A* 2005, **102:**6395–6400.
84. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N: **ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.** *Nucleic Acids Res* 2010, **38:**W529–W533.
85. Le SQ, Gascuel O: **An improved general amino acid replacement matrix.** *Mol Biol Evol* 2008, **25:**1307–1320.
86. Eo SH, Bininda-Emonds ORP, Carroll JP: **A phylogenetic supertree of the fowls (Galloanserae, Aves).** *Zoologica Scripta* 2009, **38:**465–481.
87. Resman N, Vasl J, Oblak A, Pristovsek P, Gioannini TL, Weiss JP, Jerala R: **Essential roles of hydrophobic residues in both MD-2 and Toll-like receptor 4 in activation by endotoxin.** *J Biol Chem* 2009, **284:**15052–15060.
88. Smirnova I, Poltorak A, Chan EKL, McBride C, Beutler B: **Phylogenetic variation and polymorphism at the Toll-like receptor 4 locus (TLR4).** *Genome Biol* 2000, **1:**research002.1–research002.10.
89. White SN, Taylor KH, Abbey CA, Gill CA, Womack JE: **Haplotype variation in bovine Toll-like receptor 4 and computational prediction of a positively selected ligand-binding domain.** *Proc Natl Acad Sci U S A* 2003, **100:**10364–10369.
90. Wlasiuk G, Nachman MW: **Adaptation and constraint at Toll-like receptors in primates.** *Mol Biol Evol* 2010, **27:**2172–2186.
91. Smith SA, Jann OC, Haig D, Russell GC, Werling D, Glass EJ, Emes RD: **Adaptive evolution of Toll-like receptor 5 in domesticated mammals.** *BMC Evol Biol* 2012, **12:**122.
92. Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA: **TLR4 mutations are associated with endotoxin hyporesponsiveness in humans.** *Nat Genet* 2000, **25:**187–191.
93. Diebold SS, Massacrier C, Akira S, Paturel C, Morel Y, Reis e Sousa C: **Nucleic acid agonists for Toll-like receptor 7 are defined by the presence of uridine ribonucleotides.** *Eur J Immunol* 2006, **36:**3256–3267.
94. Bainova H, Kralova T, Bryjova A, Albrecht T, Bryja J, Vinkler M: **First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds.** *Dev Comp Immunol* 2014, **45:**151–155.
95. Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR: **Biological and biomedical implications of the co-evolution of pathogens and their hosts.** *Nat Genet* 2002, **32:**569–577.

doi:10.1186/s12711-014-0072-6
Cite this article as: Vinkler *et al.*: Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. *Genetics Selection Evolution* 2014 **46**:72.

PAPER III.

Králová T., Albrecht T., Bryja J., Hořák D., Johnsen A., Lifjeld J. T., Novotný M., Sedláček O., **Velová H.** & Vinkler M. (2018): Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an ecological context. Molecular Ecology, 27(13):2871–2883. (IF_{2018}= 5.855)

In this paper, we provide evidence of phenotypic variability in avian TLR4 ligand-binding region (LBR), which forms the direct interface between host and pathogen molecular structures. We sequenced this region in 55 passerine species using NGS approach and we subsequently measured the distribution of electrostatic potential on the surface of the receptor. This distribution varies substantially among species and moreover it clusters based on distinct patterns of the TLR4 LBR surface charge distribution. We also tested selection acting on this LBR and predicted potentially functionally important residues. Moreover, five of these positions co-determine identity of the charge clusters. Groups of species that host related communities of pathogens were predicted to cluster based on TLR4 LBR charge. Despite some evidence for convergence among taxa, there were no clear associations between the TLR4 LBR charge distribution and any of the general ecological characteristics compared (migration, latitudinal distribution and diet). Closely related species, however, mostly belong to the same surface charge cluster indicating that phylogenetic constraints are key determinants shaping TLR4 adaptive evolution. Our results suggest that host innate immune evolution is consistent with the Fahrenholz's rule on co-speciation of hosts and their parasites.

Received: 20 November 2017 | Revised: 23 April 2018 | Accepted: 24 April 2018

DOI: 10.1111/mec.14724

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an evolutionary context

Tereza Králová[1,2] | Tomáš Albrecht[1,3] | Josef Bryja[1,2] | David Hořák[4] | Arild Johnsen[5] | Jan T. Lifjeld[5] | Marian Novotný[6] | Ondřej Sedláček[4] | Hana Velová[3] | Michal Vinkler[3]

[1]Institute of Vertebrate Biology, The Czech Academy of Sciences, Brno, Czech Republic

[2]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic

[3]Department of Zoology, Faculty of Science, Charles University, Prague, Czech Republic

[4]Department of Ecology, Faculty of Science, Charles University, Prague, Czech Republic

[5]Natural History Museum, University of Oslo, Oslo, Norway

[6]Department of Cell Biology, Faculty of Science, Charles University, Prague, Czech Republic

Correspondence
Tereza Králová, The Czech Academy of Sciences, Institute of Vertebrate Biology, Studenec 122, 675 02 Koněšín, Czech Republic.
Email: kralova.tereza@mail.muni.cz

Funding information
Ministerstvo Školství, Mládeže a Tělovýchovy (ELIXIR CZ), Grant/Award Number: LM2015047; Grantová Agentura, Masarykova univerzita, Grant/Award Number: MUNI/A/0888/2013; Grantová Agentura, Univerzita Karlova, Grant/Award Number: 540214; Grantová Agentura České Republiky, Grant/Award Number: GA17-24782S, P506/15-11782S; Univerzita Karlova v Praze, Grant/Award Number: PRIMUS/17/SCI/12, UNCE 204069; Institutional Research Support, Grant/Award Number: SVV 260 434/2018

Abstract

Positive selection acting on Toll-like receptors (TLRs) has been recently investigated to reveal evolutionary mechanisms of host–pathogen molecular co-adaptation. Much of this research, however, has focused mainly on the identification of sites predicted to be under positive selection, bringing little insight into the functional differences and similarities among species and a limited understanding of convergent evolution in the innate immune molecules. In this study, we provide evidence of phenotypic variability in the avian TLR4 ligand-binding region (LBR), the direct interface between host and pathogen molecular structures. We show that 55 passerine species vary substantially in the distribution of electrostatic potential on the surface of the receptor, and based on these distinct patterns, we identified four species clusters. Seven of the 34 evolutionarily nonconservative and positively selected residues correspond topologically to sites previously identified as being important for lipopolysaccharide, lipid IVa or MD-2 binding. Five of these positions codetermine the identity of the charge clusters. Groups of species that host-related communities of pathogens were predicted to cluster based on their TLR4 LBR charge. Despite some evidence for convergence among taxa, there were no clear associations between the TLR4 LBR charge distribution and any of the general ecological characteristics compared (migration, latitudinal distribution and diet). Closely related species, however, mostly belonged to the same surface charge cluster indicating that phylogenetic constraints are key determinants shaping TLR4 adaptive evolution. Our results suggest that host innate immune evolution is consistent with Fahrenholz's rule on the cospeciation of hosts and their parasites.

KEYWORDS
functional evolution, host–pathogen interaction, surface charge distribution, TLR4 molecular phenotype, Toll-like receptors

1 | INTRODUCTION

Immune genes frequently exhibit remarkable levels of adaptive variation as a result of pathogen-mediated diversifying selection (Schmid-Hempel, 2011). In animal genomes, the most conspicuous and well-studied adaptive immune genetic variation is undoubtedly found in the major histocompatibility complex (*MHC*; Hughes & Yeager, 1998). Research on the *MHC*, however, has not yet provided a full understanding of the basic principles of host–pathogen co-adaptation, for several possible reasons. First, *MHC* variation is typically

highly complex, comprising in most species tens to hundreds of alleles encoded at a variable number of loci (Piertney & Oliver, 2006), making locus-specific research difficult in most taxa. Second, conformational plasticity responsible for structurally flexible ligand binding has been documented for some (though not all) MHC variants, further increasing the complexity of the pathogen detection system (Koch et al., 2007). At last, MHC is not a pathogen-recognizing receptor itself, but rather the detection system is dependent on T-cell receptor clonal somatic variability, which is virtually unlimited and uncharacterizable based on the sequencing methods currently available. Taken together, this model system may be simply too complex to allow the resolving of some particular basic questions of molecular co-adaptations between hosts and their pathogens. It has also been suggested that the *MHC* is not sufficient for an understanding wildlife immunogenetics as approximately half of the genetic variability for resistance to infection is attributable to non-*MHC* genes (Acevedo-Whitehouse & Cunningham, 2006).

Considering their level of interspecific and intraspecific variability with potential functional significance, innate immune receptors, such as Toll-like receptors (TLRs), may be convenient models for describing the basic features of host–pathogen molecular co-evolution (Vinkler & Albrecht, 2009). TLRs are transmembrane proteins that directly bind various host-derived damage-associated molecular patterns and pathogen-derived microbe-associated molecular patterns (MAMPs), both representing important danger signals to immunity (Akira, Uematsu, & Takeuchi, 2006). From an evolutionary perspective, MAMPs are particularly interesting. In a typical manner, MAMPs are structures essential for the survival of broad ranges of potential pathogens (bacteria, viruses, etc.). Although generally conservative, important variability has been described in these structures across pathogen taxa (Maeshima & Fernandez, 2013; Needham & Trent, 2013; Nijland, Hofland, & van Strijp, 2014; Wang, Quinn, & Yan, 2015), including adaptations allowing the avoidance of immune detection. In response to changeable pathogen selective pressure, hosts may be predicted to adaptively modify their TLR structures. Several independent lines of evidence using several tests of positive selection support this assumption in phylogenetically diverse animals (Alcaide & Edwards, 2011; Areal, Abrantes, & Esteves, 2011; Grueber, Wallis, & Jamieson, 2014; Wlasiuk & Nachman, 2010; Yilmaz, Shen, Adelson, Xavier, & Zhu, 2004).

The TLR protein family (typically comprising about a dozen members per vertebrate species) is strictly conservative in its general structure. The TLR ligand-binding N-terminal ectodomain (ECD) consists of several leucine-rich repeat motifs, resulting in its typical horseshoe shape. The ECD is followed by a short transmembrane domain and a C-terminal Toll/interleukin-1 receptor (TIR) cytosolic domain responsible for signal transduction (Takeda, Kaisho, & Akira, 2003). Despite negative selection acting on most parts of the molecule, evolutionarily variable regions interspace the conservative leucine-rich repeat motifs in the ECD to provide variability to the ligand-binding surfaces. It has been shown that the predominantly bacterial-sensing TLRs are more genetically polymorphic than the viral-sensing TLRs, due to their ability to recognize several structurally different ligands and their greater redundancy in bacterial recognition (Vinkler, Bainová, & Bryja, 2014; Wlasiuk & Nachman, 2010; Yilmaz et al., 2004). TLR4 in particular binds a broad spectrum of structurally diverse ligands including lipopolysaccharides (LPS). It exhibits high levels of putatively adaptive variation, mainly in the ligand-binding region (LBR) that is located in the central part of the ECD (Alcaide & Edwards, 2011; Areal et al., 2011; Fornůsková, Bryja, Vinkler, Macholán, & Piálek, 2014; Grueber et al., 2014; Tong et al., 2015; Vinkler, Bryjová, Albrecht, & Bryja, 2009; Vinkler et al., 2014; Wlasiuk & Nachman, 2010). TLR4 ligand variation is increased by strain- or species-specific modifications such as acylation/de-acylation, phosphorylation or hydroxylation of the LPS in Gram-negative bacteria (Needham & Trent, 2013). As a consequence, various forms of LPS can bind with different affinity to different TLR4 variants, and it has been shown that even slight differences in TLR4 structure can have dramatic effects on its signalling (Walsh et al., 2008). For example, it has been suggested that at least two single point substitutions in human *TLR4* evolved in response to distinct pathogen pressures in Africa and Eurasia (Ferwerda et al., 2007). Substitutions in *TLR4* can have either structural effects, changing the shape of the LBR or functional effects, for example altering surface molecular features such as its charge (Walsh et al., 2008). For example, in Galloanserae birds, the electrostatic potential distribution on the surface of the TLR4 molecule was found to be more significant than changes in tertiary structure (Vinkler et al., 2014).

Perching birds (order Passeriformes), with their high species richness and large ecological and phylogenetic diversification (Jetz, Thomas, Joy, Hartmann, & Mooers, 2012), are an ideal model taxon for research on host–pathogen molecular co-evolutionary dynamics. Some of the most important ecological characteristics associated with parasite diversity in birds are latitudinal distribution, seasonal movements and diet (e.g., Hannon, Kinsella, Calhoun, Joseph, & Johnson, 2015; Schemske, Mittelbach, Cornell, Sobel, & Roy, 2009; Waldenström, Bensch, Kiboi, Hasselquist, & Ottosson, 2002). It has been shown that in several vertebrate species, the parasite load and disease incidence increase towards the equator, following differences in the extent of host–pathogen interactions with changing latitude (reviewed in Schemske et al., 2009). With respect to this fact, the migratory species inhabiting diverse habitats during their annual cycle expose themselves to increased parasite diversity compared to both temperate and tropical residents (Gutiérrez, Rakhimberdiev, Piersma, & Thieltges, 2017; Hannon et al., 2015; Leung & Koprivnikar, 2016; Waldenström et al., 2002). Dietary preferences for more diverse food may also increase the parasite richness with trophically transmitted parasite species, that is carnivorous and omnivorous birds are expected to host a wider range of parasites than herbivores. However, evidence supporting this hypothesis has only been obtained through studies focusing on macroparasites such as helminths (Hannon et al., 2015; Santoro, Kinsella, Galiero, degli Uberti, & Aznar, 2012).

The theory of adaptive evolution driven by pathogen selective pressure predicts that species exposed to similar pathogens should

show similarities in the binding properties of their immune receptors. As microbial-richness species-specific differences are scarcely known in birds (but see Kropáčková et al., 2017), analyses of host-associated ecological variables together with their neutral phylogeny may reveal factors driving molecular convergence in immune receptors. Furthermore, the consistency of the immune receptor phenotype clustering with the neutral phylogeny of the host species may provide support for the codiversification of parasites with their hosts, as hypothesized by Fahrenholz's rule (Eichler, 1948). Closely related species of hosts are frequently colonized by closely related communities of microbiota (Kropáčková et al., 2017). As a result of their intimate associations, many pathogens and their hosts continually co-adapt in phylogenetically corresponding patterns.

Combining sequence analyses with protein structural modelling, we applied a novel methodological approach to explore the molecular evolution of the most functionally relevant part of the TLR4 LBR. Patterns of TLR4 LBR adaptive evolution were investigated in a carefully selected data set of 55 species representing both tropical and temperate passerine birds with diverse ecological characteristics. The main aim of this study was to test the adaptive value of sequence variation encoding molecular phenotypic traits attributable to microbiota-determining traits. More specific, we compared the effects of ecological traits, such as migration, latitudinal distribution and diet, on TLR4 LBR variation, while controlling for phylogenetic relatedness. Our approach should be generally applicable to comparative analyses of functional consequences at the protein level arising from interspecific nucleotide sequence variation shaped by natural selection.

2 | MATERIAL AND METHODS

2.1 | Target region of the TLR4 and studied taxa

In this study, we performed targeted amplification of the whole LBR-coding region (681 bp) in the *TLR4* gene, which encodes for all putative LPS-binding sites. For polymerase chain reaction (PCR), a new pair of *TLR4* LBR-specific primers was designed based on an alignment of previously known sequences of passerine *TLR4*s. These primers, passerTLR4fw-764 (5′-GTTTGCAGGT CAGCAGACTC-3′) and passerTLR4rev-1444 (5′-GCAAAGAGTT CAAGCCACAAAA-3′), successfully amplified the required gene region in all 46 de novo sequenced passerine species from the Czech Republic, Norway, Cameroon and Nigeria selected for this study (samples obtained from European biobanks as scientific loans, see Supporting Information Table S1 in the Supplement; for PCR conditions see below). Several of the analysed species form phylogenetically related pairs inhabiting distinct geographical zones. The newly produced sequence data were added to nine currently available passerine *TLR4* sequences freely available in genome data deposited at NCBI GenBank. The complete list of species including the origin of the representative samples and sequences is provided in Supporting Information Table S1. Furthermore, the sequence of the chicken *TLR4* (*Gallus gallus*, *Galgal*; GenBank acc. no. AY064697.1) was used as an outgroup in both phylogenetic and structural analyses, and the crystal structure of the mouse *TLR4* (*Mus musculus*; GenBank acc. no. NM_021297.2; PDB ID 2Z64) was used in the structural analysis.

2.2 | TLR4 sequencing

For the sequencing step, special fusion primers were synthesized that consisted of the *TLR4* LBR-specific primer (see above), a unique 12-bp-long barcode (Caporaso et al., 2012), a key sequence (TCAG) and a specific 21-bp-long next-generation sequencing adaptor (Roche, Basel, Switzerland). Blood samples were stored frozen (20°C) in 96% ethanol or refrigerated (4°C) in Queens Lysis Buffer (QLB) until the DNA extraction using DNeasy Blood & Tissue Kit (QIAGEN, Hilden, Germany) according to the manufacturer's protocol. PCR was first performed with only the *TLR4* LBR-specific primers to confirm successful amplification, and then with fusion primers required for the sequencing step on the 454-pyrosequencing platform, with a final PCR product length of 755 bp. The PCR was mixed using a Multiplex PCR Kit (QIAGEN) under the following conditions: 0.3 μM of each primer, 1× Multiplex PCR Master Mix (including HotStarTaq Polymerase and 3 mM Mg^{2+}) and 1 μl of purified DNA with the addition of deionized water up to 10 μl. The reaction was performed on a Mastercycler ep (Eppendorf, Hamburg, Germany) with initial heat activation at 94°C (15 min) followed by 30 cycles of denaturation at 94°C (30 s), annealing at 56°C (40 s) and extension at 72°C (1 min), ending with a final extension step at 72°C (10 min). All amplified samples were then purified using magnetic beads (Agencourt AMPure XP, Beckman Coulter Life Sciences, Brea, CA, USA), pooled in equimolar amounts and sequenced with an FLX+ instrument (Roche). Application of the next-generation sequencing (NGS) approach allowed us to avoid the costly cloning step in our effort to obtain single-haplotype sequences. Each individual was identified according to its unique combination of the forward and reverse barcode, and the data were demultiplexed using the R packages SEQINR, BIOSTRINGS and SHORTREAD in the open statistical software and programming language R v. 3.1.2 (R Core Team, 2015). All nucleic acid sequences were aligned, edited and translated into protein sequences using GENEIOUS v. 9.1.5. (http://www.geneious.com; Kearse et al., 2012). No interspecifically shared variation was observed among the species in our data set. For methodological reasons, the final data set consisted of one randomly selected haplotype sequence per species (GenBank acc. nos. MG209180-MG209225). Sequences of the Hooded crow (*Corvus cornix cornix*) and Atlantic canary (*Serinus canaria*) had to be excluded from further analyses because of their identity with sequences of the American crow (*Corvus brachyrhynchos*) and European serin (*Serinus serinus*), respectively. Thus, the final data set consisted of haplotype sequences from 53 passerine species covering 28 families across the passerine phylogeny.

2.3 | Analysis of functional variation

A basic description of the structural and physical features of the sequences was performed in Geneious. The molecular weight and charge of the linear sequences at pH = 7 were computed using the web-based tool LRRfinder (www.lrrfinder.com), which compares input protein sequence to those already present in a Protein Data Bank (PDB) database. The protein structure of the whole extracellular domain of TLR4 in the Zebra finch (*Taeniopygia guttata*) was modelled using the I-TASSER server (Yang et al., 2015). The LBR of this protein structure served as a PDB model to identify regions on the surface of the molecule that may be potentially involved in host–pathogen co-evolution based on their level of evolutionary nonconservatism using the online tool CONSURF (http://consurf.tau.ac.il; Glaser et al., 2003). The advantage of CONSURF compared to other similar tools is that it takes into account the phylogenetic relationships among the sequences, which is especially important for the interspecific comparison of species differing in their levels of relatedness. For this analysis, the alignment of the 53 unique amino acid sequences (Supporting Information Figure S1) was uploaded to CONSURF along with the phylogenetic species tree extracted from http://birdtree.org/ (Jetz et al., 2012; see below for more details). We applied a Bayesian calculation method with an LG evolutionary substitution model for the analysis (Le & Gascuel, 2008). As in immune receptors any host adaptations to diversified pathogen structures can occur only at the phylogenetically nonconservative sites, we consider evolutionary nonconservatism as evidence strengthening indication for positive selection.

According to recent research in Galloanserae birds (Vinkler et al., 2014), one of the most important variable features of TLRs is the electrostatic potential distribution on the molecular surface, so we performed a closer examination of this property. To evaluate the anticipated functional differences in TLR4 LBRs between selected passerine species, we predicted the tertiary structure of each species-specific protein using homology modelling tool MODELLER v. 9.14 (Eswar et al., 2006; pairwise sequence alignment performed in Geneious), with mouse TLR4 as a template with known crystal structure (Kim et al., 2007). We generated two models for each sequence, and the one with the lower DOPEscore was selected for further analysis. The quality of the selected model for each species was evaluated using the MODFOLD Model Quality Assessment Server v. 4.0 and 6.0 (McGuffin, Buenavista, & Roche, 2013) and DALI pairwise comparison tool computing similarity of two proteins based on their structural alignment (DALILITE v. 3, http://ekhidna.biocenter.helsinki.fi/dali_lite/start; Hasegawa & Holm, 2009).

All structures (including the chicken as an outgroup) were superimposed in UCSF CHIMERA v. 1.10.1 using the "Match->Align" function (http://www.cgl.ucsf.edu/chimera/; Pettersen et al., 2004), and the whole set was uploaded to PIPSA (Protein Interaction Property Similarity Analysis; Richter, Wenzel, Stein, Gabdoulline, & Wade, 2008) to acquire a matrix of species with pairwise comparisons of their surface electrostatic potential distances. The visualization of the protein surface charge variation in 3D space was displayed in UCSF CHIMERA using input files generated in APBS v. 1.4.1. (Adaptive Poisson-Boltzmann Solver; Baker, Sept, Joseph, Holst, & McCammon, 2001), which solves the equations of continuum electrostatics using structures adapted for it in PDB2PQR v. 2.0.0. (Dolinsky, Nielsen, McCammon, & Baker, 2004; Dolinsky et al., 2007) based on the PARSE force-field (PDB2PQR Server, http://nbcr-222.ucsd.edu/pdb2pqr_2.0.0/). To estimate the surface receptiveness of individual residues in the TLR4 LBR, we calculated the relative solvent accessibility (RSA) from the residue's solvent accessibility surface area (ASA) computed from the PDB models of four selected species using the web server xssp (http://www.cmbi.ru.nl/xssp/; based on Kabsch & Sander, 1983) divided by a corresponding maximum possible ASA for the given amino acid (Tien, Meyer, Sydykova, Spielman, & Wilke, 2013).

At last, a cluster analysis was performed in R v. 3.4.1. (R Core Team, 2017) using RSTUDIO v. 1.0.153 (RStudio Team, 2015) based on the matrix of pairwise electrostatic distances from PIPSA and using the upgma function (package PHANGORN v. 2.2.0, method = ward; Schliep, 2011). This phenetic tree (dendrogram) of surface charge distribution was then visually compared with a neutral phylogenetic species tree using R packages APE v. 4.1 (Paradis, Claude, & Strimmer, 2004), PHANGORN v. 2.2.0 (Schliep, 2011) and DENDEXTEND v. 1.5.2 (Galili, 2015). Both dendrograms were also compared statistically by the "entanglement" function (E) in the package DENDEXTEND, resulting in a value between 1 (different dendrograms) and 0 (identical dendrograms). The phylogenetic tree was extracted from the global phylogeny of birds using a web-based tool at http://birdtree.org/ (Jetz et al., 2012; Supporting Information Figure S2). We generated 1000 random trees (Stage2 Mayr All Hackett backbone), which were then summarized into a single consensus tree with TREEANNOTATOR v. 1.8.2 implemented in BEAST software package v. 1.8.2 (Drummond, Suchard, Xie, & Rambaut, 2012). The TLR4 gene tree was constructed using the maximum-likelihood principle in a web-based phylogeny software PHYML 3.0 (Guindon et al., 2010), and it mostly resembled the phylogenetic species tree (data not shown). All trees were visualized using FIGTREE v. 1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/).

2.4 | Selection analysis

First, the possibility of recombination influencing evolution of the *TLR4* LBR in the studied species was excluded based on the results of GARD analysis (Genetic Algorithms for Recombination Detection, http://www.datamonkey.org; Kosakovsky Pond, Posada, Gravenor, Woelk, & Frost, 2006). Then, we applied four standard codon-based methods to find signatures of diversifying selection acting on the *TLR4* LBR in passerines: (i) FUBAR—A Fast, Unconstrained Bayesian AppRoximation for inferring selection (Murrell et al., 2013); (ii) MEME—Mixed Effects Model of Episodic Diversifying Selection (Murrell et al., 2012); (iii) SLAC—Single Likelihood Ancestor Counting (Pond & Frost, 2005b), all three available on the Datamonkey webserver (http://www.datamonkey.org; Delport, Poon, Frost, & Kosakovsky Pond, 2010; Pond & Frost, 2005a); and (iv) the codon-based

substitution model (codeml) M8a (null model, $\omega = 1$) in comparison with M8 (an alternative model, beta&ω) implemented in PAML v. 4.8 (Phylogenetic Analysis by Maximum Likelihood; Yang, 2007). Using several alternative selection tests allowed us to evaluate the selective pressure occurring in our diverse data set with slightly different assumptions and constraints. The SLAC model combines maximum-likelihood (ML) and counting approaches to find nonsynonymous (dN) and synonymous (dS) substitution rates based on the reconstructed ancestral sequences. This is different from the FUBAR method that infers dN and dS substitution rates using Bayesian approach with the assumption of constant selective pressure along the entire phylogeny, while MEME allows for selection under the proportion of branches to detect sites evolving under episodic selection. Evolutionary signatures in the input sequences were also searched using the identification of conserved/variable physicochemical properties at individual amino acid sites in PRIME (PRoperty Informed Models of Evolution) running on the Datamonkey server (Delport et al., 2010; Pond & Frost, 2005a). We applied a set of five physicochemical properties reflecting the polarity, secondary structure, molecular volume, refractivity/heat capacity and electrostatic charge of the amino acids (Atchley, Zhao, Fernandes, & Drüke, 2005). The location and variability in positively selected sites were then compared with the functional sites of the TLR4 LBR found in previous vertebrate studies using crystallographic methods (Supporting Information Table S2; Kim et al., 2007; Ohto, Fukase, Miyake, & Shimizu, 2012; Park et al., 2009). The variability in the positively selected amino acids with indications of diversity in their physicochemical properties was visualized using WEBLOGO v. 3 (Crooks, Hon, Chandonia, & Brenner, 2004; http://weblogo.threeplusone.com/). Throughout the study, sites are numbered according to the Zebra finch *TLR4* coding sequence (GenBank acc. no. FJ695612).

2.5 | Identifying convergence in structural adaptive evolution

The bioinformatic approach described above allows searching for evidence on evolutionary convergence even in phylogenetically distantly related taxa. In an attempt to interpret the convergence in structural adaptive evolution of the TLR4 LBR, we compared our surface charge clustering with several broad ecological characteristics of the species. We focused on basic ecological factors and characterized the studied species according to: (i) migration and latitudinal distribution (hereafter called latitudinal dispersion: temperate resident, temperate partial or short-distance migrant, temperate long-distance migrant and tropical resident species) and (ii) diet (predominantly plant, animal or mixed diet). These traits represent intersection between those characteristics to which data are presently available for all the analysed species (including the tropical species) and the traits that were shown in earlier studies to exhibit associations with parasite abundance or prevalence (Gutiérrez et al., 2017; Hannon et al., 2015; Leung & Koprivnikar, 2016; Santoro et al., 2012; Waldenström et al., 2002). The categories of each characteristic were mapped onto the dendrogram of electrostatic potentials using a colour code to compare TLR4 phenotypic clustering with the ecological variable distribution. To examine the significance of associations between surface charge clusters and selected ecological categories, we performed the Fisher's exact test for count data in R v. 3.4.1. (R Core Team, 2017). In addition, we applied phylogenetic generalized least squares (PGLS) analysis which corrects for the relatedness of taxa using the electrostatic distances of individual species from the chicken outgroup (caper package v. 0.5.2 in R; Orme et al., 2013). Using the maximum-likelihood method, we also estimated the phylogenetic scaling parameter lambda (λ) and tested its value against $\lambda = 0$ (indicating phylogenetic independence) and $\lambda = 1$ (phylogenetic dependence). The complex PGLS model (with both predictors included), models containing only one predictor and the null model were then compared using the second-order Akaike information criterion (AICc) to identify the best-fit model (i.e., the model with the lowest AICc value; MUMIN package v. 1.15.6; Barton, 2016).

3 | RESULTS

All 53 unique nucleotide sequences of the *TLR4* LBR encode polypeptides without any stop codons and with only one single codon deletion in the Medium ground finch (*Geospiza fortis*; Supporting Information Table S3, Supporting Information Figure S1). The mean interspecific identity to the Zebra finch (*Taeniopygia guttata*) amino acid sequence was 84.3% (range 73.0%–94.2%). In an interesting manner, this sequence variation affected the charge variation in the linearized polypeptide molecules (values from 9.1 to 8.0; Supporting Information Table S3). We also found nonsynonymous substitutions at several LPS-binding positions (267, 344, 368, 392, 396; Supporting Information Table S2) identified in previous studies (Kim et al., 2007; Ohto et al., 2012; Park et al., 2009).

3.1 | No interspecific variation in TLR4 LBR tertiary structures

To assess the effect of sequence variation on the protein structural properties, we modelled the tertiary structures of the TLR4 LBRs. The accuracy of our models was supported with MODFOLD (quality scores of all models with p 0.001). The interspecific pairwise comparison performed in DALI showed only low variability in TLR4 LBR structures, with RMSD values below 1.0 Å.

3.2 | Distinct patterns of TLR4 LBR surface charge distribution in passerines

Contrary to the conservative tertiary protein structure, the distribution of electrostatic potential on the polypeptide surfaces varied substantially among the species (Supporting Information Figure S3). The electrostatic distance (calculated as the absolute difference between the surface electrostatic potential values of two different molecules) ranged from 0.307 (*Phoenicurus phoenicurus* vs. *Ficedula*

albicollis) to 1.284 (*Erithacus rubecula* vs. *Certhia familiaris*) along the whole TLR4 LBR (Figure 1). Based on PIPSA clustering (Supporting Information Figure S4), we divided the whole data set into four distinct groups differing in their patterns of surface charge distribution (Figure 1, Supporting Information Table S3). The PGLS analysis with implemented phylogeny correction showed that PIPSA clustering corresponded to the charges of the linearized polypeptides calculated in LRRfinder (pgls, $F = 9.26$, p 0.001; this pattern was less obvious without phylogeny correction, see Supporting Information Table S3, Supporting Information Figure S5). We found no clear distinctions between surface charge clusters with respect to the physicochemical features of the predicted LPS-binding sites.

3.3 | Positive selection shaped TLR4 LBR surface charge clustering

To show the evolutionary significance of the variation in TLR4 LBR physicochemical properties, we predicted the key amino acid positions under positive selection in our data set. We assumed that sites under positive (diversifying) selection should be also the least evolutionarily conservative sites. Using CONSURF, 18 sites were identified as evolutionarily nonconservative in our data set, four of which lie in close topological proximity to the potentially important LPS-binding sites (Supporting Information Figure S6, Supporting Information Table S4). Furthermore, based on five independent approaches (FUBAR, MEME, SLAC, PAML and PRIME), we identified a total of 32 amino acid sites predicted to be under positive/diversifying selection (Figure 2, Supporting Information Figure S7, for details, see Supporting Information Tables S5 and S6). Sixteen sites of these 32 were also detected in CONSURF as evolutionarily nonconservative and six sites corresponded to the amino acid positions defined as potentially important for LPS or lipid IVa binding. Altogether, we found 34 evolutionarily nonconservative and positively selected residues in the studied region (Supporting Information Table S7). To avoid false-positive results, before proceeding with further analysis, we applied a criterion that only those positions identified by at least three independent approaches (including the CONSURF analysis) would be considered as reliable positively selected sites. Only 11 sites fulfilled this criterion: positions 267, 269, 270, 272, 298, 308, 331, 344, 351, 374 and 402 (for details and consistency with other vertebrate studies, see Supporting Information Table S7). On these 11 positively selected sites, we found distinct patterns in amino acid characteristics differentiating surface charge cluster A from clusters B, C and D (Table 1). Most of the evaluated positions, however, were variable within individual clusters, suggesting functional diversification in various taxa (Supporting Information Table S8, Figure 3). Of 20 positively selected sites identified by MEME, which tests for the clade-specific selection, three sites clearly differed between charge clusters (351, 366 and 374). While positions 351 and 374 were found by multiple selection tests, position 366 was recognized only by MEME

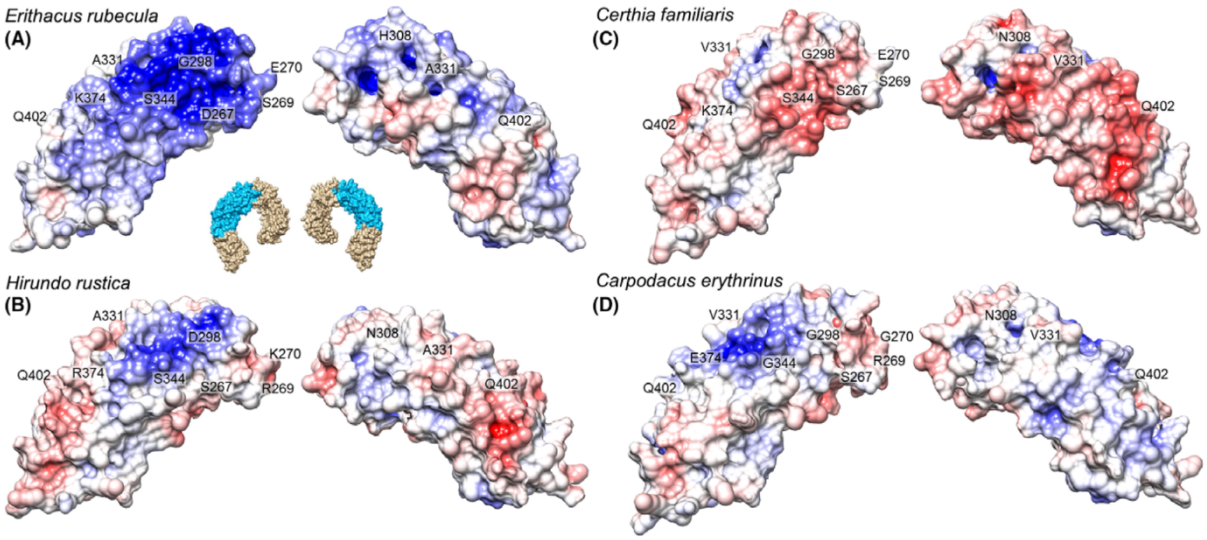


FIGURE 1 Variability in the surface charge of the TLR4 LBR in four different passerine species. One species from each charge cluster (Supporting Information Figure S3) is displayed—*Erithacus rubecula* (cluster A), *Hirundo rustica* (cl. B), *Certhia familiaris* (cl. C) and *Carpodacus erythrinus* (cl. D). *Certhia familiaris* and *Erithacus rubecula* are the two species with the greatest distance based on the surface electrostatic potential of the whole TLR4 LBR ($d = 1.284$, Supporting Information Figure S3). The inner (left-hand side) and outer (right-hand side) surface of each TLR4 LBR is displayed, with the ligand-binding surface shown in the left-hand side view. Positive charge (10) is highlighted in red and negative (10) in blue (colours in electronic version only). Positively selected amino acids confirmed by three independent tests (Table 1, Supporting Information Table S7) are labelled (the site numbering corresponds to the Zebra finch protein sequence; positions 272 and 351 are hidden). The position of the LBR within the ECD is indicated by the blue colour in the schematic representation of Zebra finch TLR4. For the variability in the surface electrostatic potential of all studied species, see Supporting Information Figure S2 [Colour figure can be viewed at wileyonlinelibrary.com]

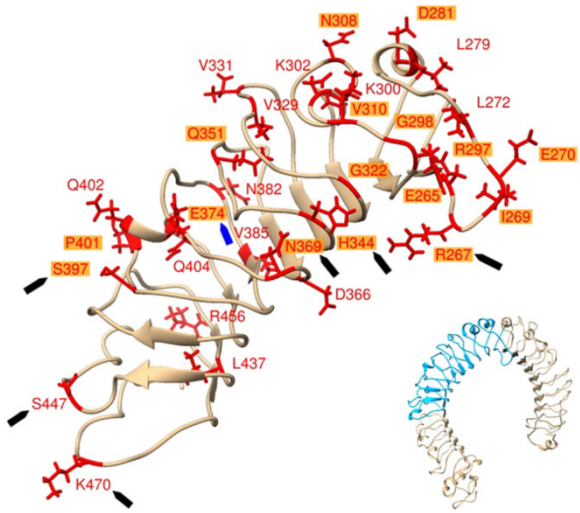


FIGURE 2 Positions of positively selected sites visualized on the 3D protein model of the Zebra finch (*Taeniopygia guttata; Taegut*) TLR4 LBR. Positively selected sites (PAML/FUBAR/MEME/SLAC/PRIME) are highlighted in red (colours in electronic version only); those sites that are also nonconservative (ConSurf) are labelled with an orange background; the sites that are in close proximity of potentially important LPS-binding sites defined in previous studies (Kim et al., 2007; Ohto et al., 2012; Park et al., 2009) are indicated with black arrows; the blue arrow shows the position of the positively selected site potentially important for lipid IVa recognition (Supporting Information Table S7). Site numbering corresponds to the *Taegut* protein sequence. The position of the ligand-binding region (in blue) on the receptor is shown in the small model of the *Taegut* TLR4 extracellular domain [Colour figure can be viewed at wileyonlinelibrary.com]

suggesting its association with branch-specific positive selection acting on charge cluster A. PRIME results showed that the physicochemical properties that are variable among passerine birds were connected with the secondary structure factor (265, 298), refractivity/heat capacity (270, 371) and charge/iso-electric point (267, 374; Supporting Information Table S6). All these sites except two were also identified by at least two other methods that either identify putatively positively selected sites based on Ka/Ks approaches or estimate evolutionary nonconservatism (for the sites 265 and 371 that do not match this criterion, further evidence supporting their functional relevance is shown in Supporting Information Table S7). In an interesting manner, positions 267 and 374 seem to divide the clusters into two distinct groups: A+D and B+C. While clusters B+C are polar uncharged at position 267 and positive at position 374, clusters A+D are variable at 267 and possess mainly negative residues at 374 (Figure 3, Supporting Information Table S8). Hence, despite the indicated relatedness of the deep branches of the surface charge dendrogram (Supporting Information Figure S5), judging by the physicochemical features at the key sites, cluster A is closer to cluster D, and cluster B to C. The importance of these positions is also supported by the fact that site 267 is directly involved in LPS and MD-2 binding and site 374 lies in close proximity to the lipid IVa recognition site (Supporting Information Table S2). Cluster A is then differentiated from cluster D by the properties of position 351 (polar vs. hydrophobic) and cluster B from cluster C mainly by the properties of position 298 (where B is negatively charged, while C is mostly not; Figure 3). Moreover, according to the relative solvent accessibility analysis, the amino acids on most of the identified positively selected sites tend to be exposed on the surface of the molecule, allowing for direct interaction with the ligand (Supporting Information Table S9).

TABLE 1 Predominant characteristics of amino acid variants represented at the well-supported positively selected sites in the four distinct TLR4 LBR surface charge clusters[Colour figure can be viewed at wileyonlinelibrary.com]

Cluster	267	269	270	272	298	308	331	344	351	374	402
A	-	PLR	NEG	HYD	SPE	-	HYD	PLR	PLR	NEG	-
B	PLR	-	-	-	NEG	-	HYD	-	HYD	POS	-
C	PLR	PLR	NEG	HYD	-	PLR	HYD	-	HYD	POS	PLR
D	-	-	-	-	-	PLR	HYD	-	HYD	-	PLR

Notes. Only those sites that were identified as positively selected (or nonconservative) by at least three independent tests (PAML, FUBAR, MEME, SLAC, PRIME, ConSurf) are displayed ($n = 11$; for details see Supporting Information Table S7). A characteristic was considered predominant for a specific cluster only when present in more than 75% of the species belonging to the particular cluster (Supporting Information Table S8). Insufficient data are indicated by a dash (–).
HYD: hydrophobic; NEG: negative charge; PLR: polar uncharged; POS: positive charge; SPE: special cases.

3.4 | Surface physicochemical features reflect phylogeny more than ecology of the species

Based on the theory of adaptive evolution driven by the pathogen-mediated natural selection, we predicted that the TLR4 LBR clusters A-D would group passerine species having distinct pathogen spectra. Therefore, we hypothesized that the clusters would be related to traits such as latitudinal dispersion or diet. In our data set, there was no association of surface charge clusters with latitudinal dispersion (Fisher's exact test, $p = 0.527$), but diet might have had an effect on the physicochemical features of TLR4 LBR molecules (Fisher's exact test, $p = 0.022$), which can also be seen on the projection of these characteristics onto a dendrogram of the TLR4 LBR surface charge distribution (Supporting Information Figure S8). A visual comparison of neutral phylogeny with the surface charge clustering indicates, however, that closely related species mostly belong to the same surface charge group (Figure 4). In accordance with this, the PGLS analysis showed no influence of either latitudinal dispersion or diet on the TLR4 LBR electrostatic distance of individual species from the chicken outgroup (the complex model with both predictors involved; $F_{5,46} = 0.833$, $p = 0.533$, $\lambda = 0.41^{0.61;\ 0.01}$, AICc = 86.85). In concordance with this finding, the null model had a lower AICc than the full model (AICc = 94.93). Models containing only diet or only latitudinal dispersion were also nonsignificantly different from the null model ($F_{2,49} = 1.895$, $p = 0.161$, $\lambda = 0.42^{0.52;\ 0.01}$, AICc = 93.64 and $F_{3,48} = 0.164$, $p = 0.92$, $\lambda = 0.66^{0.04;\ 0.03}$, AICc = 88.69, respectively; also see Supporting Information Table S10). These

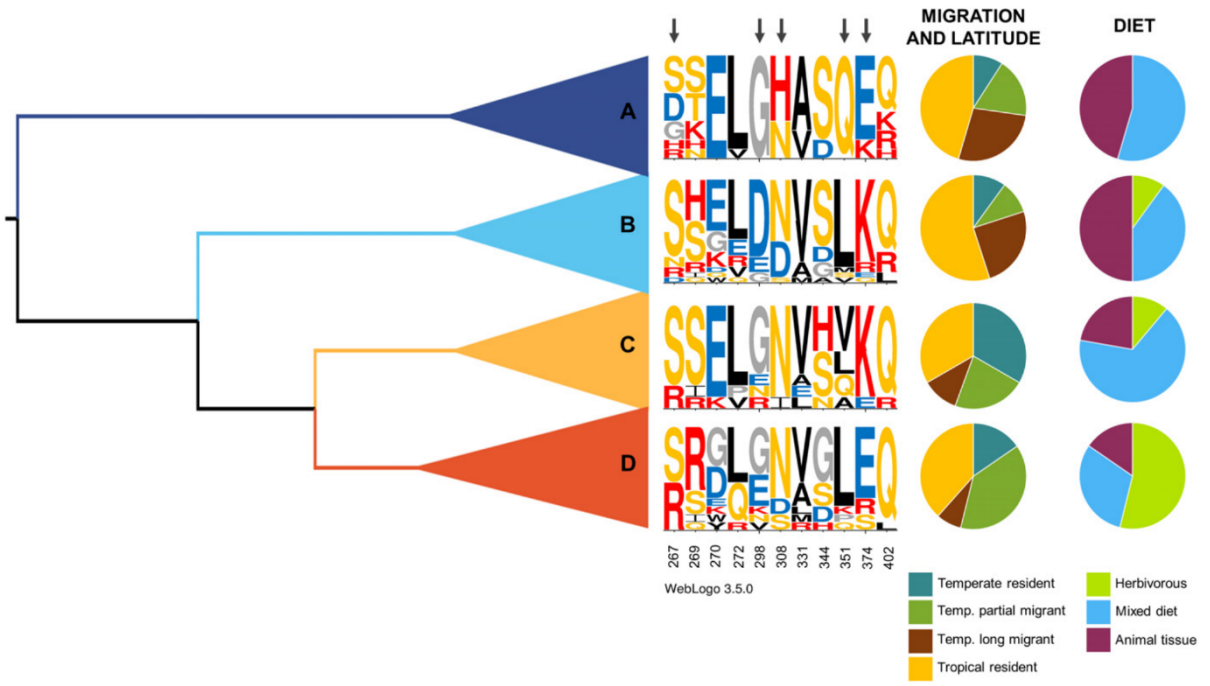


FIGURE 3 Projection of ecological characteristics to TLR4 LBR surface charge clustering. The dendrogram showing the surface charge clustering of the TLR4 LBR (Supporting Information Figures S3 and S4) was collapsed into four main clusters (A–D), and for each group, we determined the proportion of two basic ecological characteristics: (i) migration and latitude, and (ii) diet (Supporting Information Table S1). The letter diagrams show the proportion of specific amino acids on the 11 positively selected sites with their physicochemical properties indicated by colour (in electronic version only; red—positive charge, blue—negative charge, orange—polar uncharged, grey—special cases, black—hydrophobic; Supporting Information Table S8). The arrows specify those amino acid sites that appear to be important for the differentiation of individual clusters [Colour figure can be viewed at wileyonlinelibrary.com]

results imply that phylogeny is most likely the key factor influencing the charge of the TLR4 LBR. On the one hand, the well-distinguished functional cluster A is also the only cluster that is composed of species belonging to the Muscicapoidea monophyletic clade (Supporting Information Figure S9). On the other hand, the least well-defined and relatively heterogeneous cluster D consists of phylogenetically unrelated species, including domestic chicken that was used as an outgroup. In addition, the statistical comparison of both dendrograms showed an incomplete effect of phylogeny on the surface charge clustering (E = 0.578).

4 | DISCUSSION

Despite the earlier view of TLRs as highly conservative proteins (Medzhitov & Janeway, 1998), an increasing body of evidence has supported the idea that certain parts of these innate immune receptors are evolutionarily plastic and presumably shaped by pathogen-mediated adaptive evolution (e.g., Smirnova, Poltorak, Chan, McBride, & Beutler, 2000; Wang, Zhang, Liu, Zhao, & Yin, 2016). Our analysis of 53 unique passerine TLR4 LBRs, covering species inhabiting both the temperate and tropical climatic zones, revealed relatively high levels of sequence variability in *TLR4*, including substitutions at the eleven positions previously shown to bind LPS. Given that six of these sites were at least in close topological proximity to sites also identified as positively selected, our data support the notion of this genetic variation having functional significance. Among the 34 sites detected in the passerine TLR4 LBR as positively selected based on the multiple criteria adopted in this study, ten have also been identified as positively selected in other avian studies (Alcaide & Edwards, 2011; Grueber et al., 2014; Vinkler et al., 2009, 2014), and another eleven were consistent with sites previously detected as being under positive selection in both birds and mammals (Areal et al., 2011; Fornůsková et al., 2013; Grueber et al., 2014; Vinkler et al., 2009, 2014; Wlasiuk & Nachman, 2010; for the consensual sites see Supporting Information Table S7). This may suggest convergence in TLR4 molecular evolution among various vertebrate taxa.

In an interesting manner, the presumably functional sequence variability in *TLR4* is not reflected in the shapes of the predicted tertiary structures of the TLR4 LBR polypeptides, which were found to be invariant. Although we are aware that our models are only predictions, given the structural conservatism in vertebrate TLRs and the high accuracy of our models, we consider them to be trustworthy. Together with similar results previously reported for nonpasserine birds and mammals (Fornůsková et al., 2013; Vinkler et al., 2014),

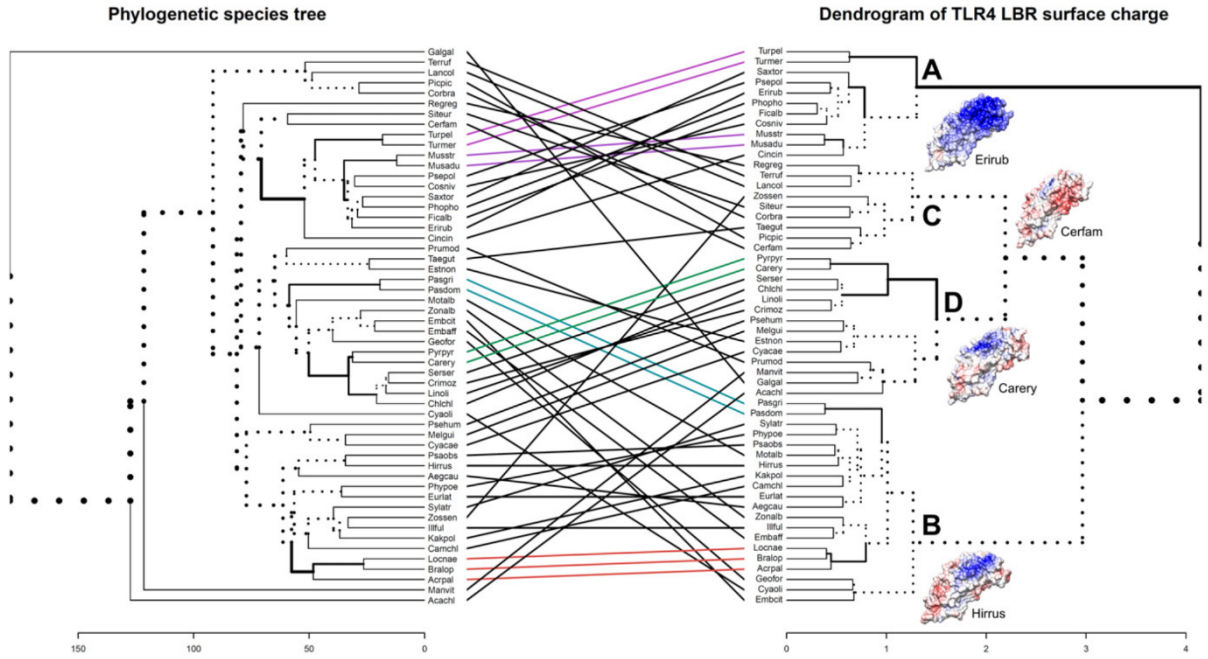


FIGURE 4 Consistency of neutral phylogeny with surface charge clustering. A comparison of the neutral phylogeny (left-hand side) and the surface charge dendrogram (right-hand side) shows that closely related species mostly cluster together in both trees, but the deeper branches do not match. The end branches are labelled with the species name abbreviations (Supporting Information Table S1). The chicken sequence (*Galgal*, GenBank Acc. No. AY064697.1) was used as an outgroup. The end branches of the same species are connected with lines. Coloured lines (electronic version only) indicate common subtrees of both dendrograms. Branches leading to distinct subtrees are highlighted with a dashed line. The surface charge clusters are indicated in the dendrogram by capital letters and miniature of the representative structural models [Colour figure can be viewed at wileyonlinelibrary.com]

our data thus indicate a high structural uniformity in TLRs, which appears to be an evolutionarily stable feature. In contrast (but again similar to nonpasserine birds and mammals; Fornůsková et al., 2013; Vinkler et al., 2014), TLR4 LBR surface charge distributions showed very distinct patterns across passerines, suggesting the importance of this physicochemical feature for ligand binding. We were able to identify four basic clusters of surface charge distribution patterns in passerines. These surface charge clusters were not identified by amino acid substitutions at previously defined mammalian LPS-binding sites (Kim et al., 2007; Ohto et al., 2012; Park et al., 2009), but were clearly recognizable by variation at five of the 11 well-supported positively selected positions (267, 298, 308, 351 and 374) that are candidate for functional effects.

Among these candidate sites, position 374 in particular distinguishes the surface charge clusters A and D from the clusters B and C. It is positioned in close topological proximity to residue 375, which has already been shown to be crucial for lipid IVa recognition in mammals (Ohto et al., 2012). Therefore, substitutions at this position altering the molecular charge from positive to negative may profoundly affect the recognition of diverse pathogens. Given that analogous positive selection has been shown in Galloanserae birds (Vinkler et al., 2014) and cyprinid fish (Tong et al., 2015), it appears likely that charge-changing substitutions at position 267, which is directly involved in LPS and MD-2 binding, may also be of particular evolutionary significance. Furthermore, the B-cluster-determining position 298, which was predicted to affect the receptor's secondary structure, has been identified as a positively selected site in mammals and birds (Areal et al., 2011; Vinkler et al., 2009; Wlasiuk & Nachman, 2010). In the same way, cluster A is differentiated from other charge clusters based on the properties of position 351, a hydrophobic-altering site that has previously been identified as positively selected in other avian studies (Alcaide & Edwards, 2011; Grueber et al., 2014), along with an adjacent site in various mammals (Areal et al., 2011; Fornůsková et al., 2013; Wlasiuk & Nachman, 2010). The interspecific differences in charge on the receptor surface may reflect the result of long-lasting co-evolutionary processes between hosts and their pathogens (Woolhouse, Webster, Domingo, Charlesworth, & Levin, 2002). It appears that also residues other than those previously identified in mammals may play an important role in shaping TLR4 LBR-binding specificity in birds, which is consistent with empirical results published earlier (Leveque et al., 2003). We recommend that the sites 267, 298, 308, 351, 374 be subject to closer investigation in further and especially functional studies.

Given the lack of published data on avian pathogens present in the environments inhabited by the wild passerine host species investigated here (Poulin & Morand, 2000), our effort to explain the observed patterns of variation in *TLR4* focused on possible connections between the TLR4 LBR surface charge clusters and basic

ecological characteristics that should be related to the pathogen pressure. However, we were unable to find any associations between TLR4 surface charge clustering and the ecology of these species. Although both studied parameters have been reported to be associated with parasite richness in previous studies (e.g., Hannon et al., 2015; Schemske et al., 2009; Waldenström et al., 2002) and TLR4 is also capable of detecting parasite ligands from tropical malaria (Eriksson, Sampaio, & Schofield, 2014), the possible effects of latitude and migration are not mirrored in the interspecific TLR4 surface charge distribution. The apparent link of surface charge with the type of diet reflects only the misbalance in numbers of species in each group and ecological similarities between phylogenetically closely related species. Therefore, our analysis suggests that TLR4 surface physicochemical features reflect the phylogeny of species more than their ecology. This is mostly consistent with the results of a recent study in passerines reporting a high association of species phylogeny with their gut microbiota (Kropáčková et al., 2017). According to Fahrenholz's rule (Eichler, 1948), phylogenetically related host species may share related species of pathogens. Thus, a similar selection mode can be predicted in individual clades of related species, allowing for clade-specific molecular evolutionary interactions between the hosts and their parasites. As a result, host compatibility rather than ecological factors may determine the avian host susceptibility to pathogens (Medeiros, Hamer, & Ricklefs, 2013). Nevertheless, Kropáčková et al. (2017) also found significant (though weaker) effects of host ecology and geography on gut microbiota content. In addition, in TLR4, the consistency between the surface charge dendrogram and passerine neutral phylogeny is incomplete, indicating a contribution of some nonphylogenetic factors to the formation of TLR4 molecular properties. The TLR4 LBR charge cluster A was the only well-distinguished cluster formed by species belonging to the monophyletic clade of Muscicapoidea. In contrast, cluster D consists of a heterogeneous spectrum of unrelated species including chicken, which suggests convergence resulting from sharing pathogens with common immunogenic profiles. A very similar result was obtained in MHC class I investigated across distantly related passerine species (Follin et al., 2013). At a larger evolutionary scale, a similar pattern was documented by our previous findings in TLR4 showing a surface charge similarity of chicken LBR to mouse LBR that is distinct from human LBR charge (Vinkler et al., 2014). The adaptive significance of this finding is highlighted by functional test results documenting similar ligand specificity of the chicken and mouse recombinant proteins when experimentally expressed in vitro in model cells (Keestra & van Putten, 2008). Therefore, in our opinion, it may be difficult to distinguish ecological signals from phylogenetic signals in traits that may be involved in host–pathogen cospeciation according to the Fahrenholz's rule, as these traits adaptively evolve in response to parasite ecology, yet in the same time, they generally follow the host phylogeny.

To conclude, our study provides an insight into the potential functionality of evolutionary adaptations in TLR4 that may contribute to understanding of the interaction patterns in the presumed host–parasite molecular co-evolution. We found no associations between TLR4 functional variation and ecological characteristics. Further research on selective pressures acting on the innate immune receptors would benefit from linking immunogenetic population data with a wider set of ecological parameters including data on microbial communities. As the interspecific differences in charge on the receptor surface were found mainly around positively selected sites, indicating functional significance, our results point to the value of in silico structural analyses. We suggest that the methodological approach adopted in this study can have general applicability for studies of the forces driving host–pathogen co-evolution based on sequence data.

ACKNOWLEDGEMENTS

We are grateful to David Hardekopf and anonymous reviewers for helpful comments on an earlier version of the manuscript. We thank Jiří Reif for providing the samples for analysis, as well as Anna Bryjová, Dagmar Čížková and Jana Nekvindová for their appreciated help in the laboratory. We also thank Jakub Kreisinger for his help with bioinformatics. This study was supported through Czech Science Foundation Projects P506/15-11782S and GA17-24782S, the Grant Agency of Masaryk University (Grant no. MUNI/A/0888/2013), the Grant Agency of Charles University (Grant no. 540214), through Institutional Research Support (Grant SVV 260 434/2018); M.V. was further supported by Charles University Research Centre programme No. 204069 and project PRIMUS/17/SCI/12, and M.N. by ELIXIR CZ research infrastructure project (MEYS Grant No: LM2015047).

DATA ACCESSIBILITY

DNA sequences: Genbank accessions MG209180-MG209225.

All data are presented within this manuscript or in online Supporting Information.

AUTHOR CONTRIBUTIONS

T.A., J.B. and M.V. performed the study design and provided funding. T.A., D.H., A.J., J.T.L., O.S. and M.V. provided the samples for analysis. T.K. and H.V. performed laboratory analysis. T.K., H.V., M.N. and M.V. analysed the data. T.K. and M.V. drafted the manuscript. All authors provided helpful comments and recommendations and approved the final version of the manuscript.

ORCID

Tereza Kralova http://orcid.org/0000-0003-1247-2497
Tomas Albrecht http://orcid.org/0000-0002-9213-0034
David Horak https://orcid.org/0000-0002-8073-1617
Jan T. Lifjeld http://orcid.org/0000-0002-9172-9985
Ondrej Sedlacek https://orcid.org/0000-0001-7216-7369
Michal Vinkler http://orcid.org/0000-0003-3572-9494

REFERENCES

Acevedo-Whitehouse, K., & Cunningham, A. (2006). Is MHC enough for understanding wildlife immunogenetics? *Trends in Ecology & Evolution*, *21*(8), 433–438. https://doi.org/10.1016/j.tree.2006.05.010

Akira, S., Uematsu, S., & Takeuchi, O. (2006). Pathogen recognition and innate immunity. *Cell*, *124*(4), 783–801. https://doi.org/10.1016/j.cell.2006.02.015

Alcaide, M., & Edwards, S. V. (2011). Molecular evolution of the Toll-Like receptor multigene family in birds. *Molecular Biology and Evolution*, *28*, 1703–1715. https://doi.org/10.1093/molbev/msq351

Areal, H., Abrantes, J., & Esteves, P. J. (2011). Signatures of positive selection in Toll-like receptor (TLR) genes in mammals. *BMC Evolutionary Biology*, *11*(1), 368. https://doi.org/10.1186/1471-2148-11-368

Atchley, W. R., Zhao, J., Fernandes, A. D., & Drüke, T. (2005). Solving the protein sequence metric problem. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(18), 6395–6400. https://doi.org/10.1073/pnas.0408677102

Baker, N. A., Sept, D., Joseph, S., Holst, M. J., & McCammon, J. A. (2001). Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(18), 10037–10041. https://doi.org/10.1073/pnas.181342398

Barton, K. (2016). *MuMIn: Multi-model inference*. Retrieved from https://CRAN.R-project.org/package=MuMIn

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Huntley, J., Fierer, N., … Knight, R. (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME Journal*, *6*(8), 1621–1624. https://doi.org/10.1038/ismej.2012.8

Crooks, G. E., Hon, G., Chandonia, J.-M., & Brenner, S. E. (2004). WebLogo: A sequence logo generator. *Genome Research*, *14*(6), 1188–1190. https://doi.org/10.1101/gr.849004

Delport, W., Poon, A. F. Y., Frost, S. D. W., & Kosakovsky Pond, S. L. (2010). Datamonkey 2010: A suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics*, *26*(19), 2455–2457. https://doi.org/10.1093/bioinformatics/btq429

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G., & Baker, N. A. (2007). PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research*, *35*(Suppl 2), W522–W525. https://doi.org/10.1093/nar/gkm276

Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., & Baker, N. A. (2004). PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research*, *32*(Suppl 2), W665–W667. https://doi.org/10.1093/nar/gkh381

Drummond, A. J., Suchard, M. A., Xie, D., & Rambaut, A. (2012). Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular Biology and Evolution*, *29*(8), 1969–1973. https://doi.org/10.1093/molbev/mss075

Eichler, W. (1948). XLI.—Some rules in ectoparasitism. *Journal of Natural History*, *1*(8), 588–598. https://doi.org/10.1080/00222934808653932

Eriksson, E. M., Sampaio, N. G., & Schofield, L. (2014). Toll-like receptors and malaria – Sensing and susceptibility. *Journal of Tropical Diseases*, *2*(1), 126. https://doi.org/10.4172/2329-891X.1000126

Eswar, N., Webb, B., Marti-Renom, M. A., Madhusudhan, M. S., Eramian, D., Shen, M., … Sali, A. (2006). Comparative protein structure modeling using modeller. *Current Protocols in Bioinformatics*, *15*, 5.6.1–5.6.30. Unit-5.6. https://doi.org/10.1002/0471250953.bi0506s15

Ferwerda, B., McCall, M. B. B., Alonso, S., Giamarellos-Bourboulis, E. J., Mouktaroudi, M., Izagirre, N., … Netea, M. G. (2007). TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(42), 16645–16650. https://doi.org/10.1073/pnas.0704828104

Follin, E., Karlsson, M., Lundegaard, C., Nielsen, M., Wallin, S., Paulsson, K., & Westerdahl, H. (2013). In silico peptide-binding predictions of passerine MHC class I reveal similarities across distantly related species, suggesting convergence on the level of protein function. *Immunogenetics*, *65*(4), 299–311. https://doi.org/10.1007/s00251-012-0676-3

Fornůsková, A., Bryja, J., Vinkler, M., Macholán, M., & Piálek, J. (2014). Contrasting patterns of polymorphism and selection in bacterial-sensing toll-like receptor 4 in two house mouse subspecies. *Ecology and Evolution*, *4*(14), 2931–2944. https://doi.org/10.1002/ece3.1137

Fornůsková, A., Vinkler, M., Pagès, M., Galan, M., Jousselin, E., Cerqueira, F., … Cosson, J.-F. (2013). Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (MURINAE). *BMC Evolutionary Biology*, *13*(1), 194. https://doi.org/10.1186/1471-2148-13-194

Galili, T. (2015). dendextend: An R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, *31*(22), 3718–3720. https://doi.org/10.1093/bioinformatics/btv428

Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E., & Ben-Tal, N. (2003). ConSurf: Identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, *19*(1), 163–164. https://doi.org/10.1093/bioinformatics/19.1.163

Grueber, C. E., Wallis, G. P., & Jamieson, I. G. (2014). Episodic positive selection in the evolution of avian Toll-like receptor innate immunity genes. *PLoS One*, *9*(3), e89632. https://doi.org/10.1371/journal.pone.0089632

Guindon, S., Dufayard, J.-F., 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. *Systematic Biology*, *59*(3), 307–321. https://doi.org/10.1093/sysbio/syq010

Gutiérrez, J. S., Rakhimberdiev, E., Piersma, T., & Thieltges, D. W. (2017). Migration and parasitism: Habitat use, not migration distance, influences helminth species richness in Charadriiform birds. *Journal of Biogeography*, *44*(5), 1137–1147. https://doi.org/10.1111/jbi.12956

Hannon, E. R., Kinsella, J. M., Calhoun, D. M., Joseph, M. B., & Johnson, P. T. J. (2015). Endohelminths in bird hosts from Northern California and an analysis of the role of life history traits on parasite richness. *Journal of Parasitology*, *102*(2), 199–207. https://doi.org/10.1645/15-867

Hasegawa, H., & Holm, L. (2009). Advances and pitfalls of protein structural alignment. *Current Opinion in Structural Biology*, *19*(3), 341–348. https://doi.org/10.1016/j.sbi.2009.04.003

Hughes, A. L., & Yeager, M. (1998). Natural selection at major histocompatibility complex loci of vertebrates. *Annual Review of Genetics*, *32*, 415–435. https://doi.org/10.1146/annurev.genet.32.1.415

Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K., & Mooers, A. O. (2012). The global diversity of birds in space and time. *Nature*, *491*(7424), 444–448. https://doi.org/10.1038/nature11631

Kabsch, W., & Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, *22*(12), 2577–2637. https://doi.org/10.1002/bip.360221211

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., … Drummond, A. (2012). Geneious basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics (Oxford, England)*, *28*(12), 1647–1649. https://doi.org/10.1093/bioinformatics/bts199

Keestra, A. M., & van Putten, J. P. M. (2008). Unique properties of the chicken TLR4/MD-2 complex: selective lipopolysaccharide activation of the MyD88-dependent pathway. *The Journal of Immunology*, *181*(6), 4354–4362. https://doi.org/10.4049/jimmunol.181.6.4354

Kim, H. M., Park, B. S., Kim, J.-I., Kim, S. E., Lee, J., Oh, S. C., … Lee, J.-O. (2007). Crystal structure of the TLR4-MD-2 complex with bound

endotoxin antagonist eritoran. *Cell*, *130*(5), 906–917. https://doi.org/10.1016/j.cell.2007.08.002

Koch, M., Camp, S., Collen, T., Avila, D., Salomonsen, J., Wallny, H.-J., … Kaufman, J. (2007). Structures of an MHC Class I molecule from B21 chickens illustrate promiscuous peptide binding. *Immunity*, *27*(6), 885–899. https://doi.org/10.1016/j.immuni.2007.11.007

Kosakovsky Pond, S. L., Posada, D., Gravenor, M. B., Woelk, C. H., & Frost, S. D. W. (2006). Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution*, *23*(10), 1891–1901. https://doi.org/10.1093/molbev/msl051

Kropáčková, L., Těšický, M., Albrecht, T., Kubovčiak, J., Čížková, D., Tomášek, O., … Kreisinger, J. (2017). Codiversification of gastrointestinal microbiota and phylogeny in passerines is not explained by ecological divergence. *Molecular Ecology*, *26*(19), 5292–5304. https://doi.org/10.1111/mec.14144

Le, S. Q., & Gascuel, O. (2008). An improved general amino acid replacement matrix. *Molecular Biology and Evolution*, *25*(7), 1307–1320. https://doi.org/10.1093/molbev/msn067

Leung, T. L. F., & Koprivnikar, J. (2016). Nematode parasite diversity in birds: The role of host ecology, life history and migration. *Journal of Animal Ecology*, *85*(6), 1471–1480. https://doi.org/10.1111/1365-2656.12581

Leveque, G., Forgetta, V., Morroll, S., Smith, A. L., Bumstead, N., Barrow, P., … Malo, D. (2003). Allelic variation in TLR4 is linked to susceptibility to Salmonella enterica serovar Typhimurium infection in chickens. *Infection and Immunity*, *71*(3), 1116–1124. https://doi.org/10.1128/IAI.71.3.1116-1124.2003

Maeshima, N., & Fernandez, R. (2013). Recognition of lipid A variants by the TLR4-MD-2 receptor complex. *Frontiers in Cellular and Infection Microbiology*, *3*, 3. https://doi.org/10.3389/fcimb.2013.00003

McGuffin, L. J., Buenavista, M. T., & Roche, D. B. (2013). The ModFOLD4 server for the quality assessment of 3D protein models. *Nucleic Acids Research*, *41*(Web Server issue), W368–W372. https://doi.org/10.1093/nar/gkt294

Medeiros, M. C. I., Hamer, G. L., & Ricklefs, R. E. (2013). Host compatibility rather than vector-host-encounter rate determines the host range of avian Plasmodium parasites. *Proceedings of the Royal Society B: Biological Sciences*, *280*(1760), 20122947–20122947. https://doi.org/10.1098/rspb.2012.2947

Medzhitov, R., & Janeway, C. A. Jr (1998). An ancient system of host defense. *Current Opinion in Immunology*, *10*(1), 12–15. https://doi.org/10.1016/S0952-7915(98)80024-1

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S. L., & Scheffler, K. (2013). FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Molecular Biology and Evolution*, *30*(5), 1196–1205. https://doi.org/10.1093/molbev/mst030

Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., & Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genetics*, *8*(7), e1002764. https://doi.org/10.1371/journal.pgen.1002764

Needham, B. D., & Trent, M. S. (2013). Fortifying the barrier: The impact of lipid A remodelling on bacterial pathogenesis. *Nature Reviews Microbiology*, *11*(7), 467–481. https://doi.org/10.1038/nrmicro3047

Nijland, R., Hofland, T., & van Strijp, J. A. G. (2014). Recognition of LPS by TLR4: Potential for anti-inflammatory therapies. *Marine Drugs*, *12*(7), 4260–4273. https://doi.org/10.3390/md12074260

Ohto, U., Fukase, K., Miyake, K., & Shimizu, T. (2012). Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(19), 7421–7426. https://doi.org/10.1073/pnas.1201193109

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2013). *caper: Comparative analyses of phylogenetics and evolution in R*. Retrieved from http://CRAN.R-project.org/package=caper

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290. https://doi.org/10.1093/bioinformatics/btg412

Park, B. S., Song, D. H., Kim, H. M., Choi, B.-S., Lee, H., & Lee, J.-O. (2009). The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature*, *458*(7242), 1191–1195. https://doi.org/10.1038/nature07830

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13), 1605–1612. https://doi.org/10.1002/jcc.20084

Piertney, S. B., & Oliver, M. K. (2006). The evolutionary ecology of the major histocompatibility complex. *Heredity*, *96*(1), 7–21. https://doi.org/10.1038/sj.hdy.6800724

Pond, S. L. K., & Frost, S. D. W. (2005a). Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics (Oxford, England)*, *21*(10), 2531–2533. https://doi.org/10.1093/bioinformatics/bti320

Pond, S. L. K., & Frost, S. D. W. (2005b). Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution*, *22*(5), 1208–1222. https://doi.org/10.1093/molbev/msi105

Poulin, R., & Morand, S. (2000). The diversity of parasites. *The Quarterly Review of Biology*, *75*(3), 277–293. https://doi.org/10.1086/393500

R Core Team (2015). *R: A language and environment for statistical computing*. Vienna, Austria. Retrieved from https://www.R-project.org/

R Core Team (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Richter, S., Wenzel, A., Stein, M., Gabdoulline, R. R., & Wade, R. C. (2008). webPIPSA: A web server for the comparison of protein interaction properties. *Nucleic Acids Research*, *36*(Suppl 2), W276–W280. https://doi.org/10.1093/nar/gkn181

RStudio Team (2015). *RStudio: Integrated development environment for R*. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Santoro, M., Kinsella, J. M., Galiero, G., degli Uberti, B., & Aznar, F. J. (2012). Helminth community structure in birds of prey (accipitriformes and falconiformes) in Southern Italy. *Journal of Parasitology*, *98*(1), 22–29. https://doi.org/10.1645/ge-2924.1

Schemske, D. W., Mittelbach, G. G., Cornell, H. V., Sobel, J. M., & Roy, K. (2009). Is there a latitudinal gradient in the importance of biotic interactions? *Annual Review of Ecology, Evolution, and Systematics*, *40*(1), 245–269. https://doi.org/10.1146/annurev.ecolsys.39.110707.173430

Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, *27*(4), 592–593. https://doi.org/10.1093/bioinformatics/btq706

Schmid-Hempel, P. (2011). *Evolutionary parasitology: The integrated study of infections, immunology, ecology, and genetics*. Oxford, UK: OUP Oxford.

Smirnova, I., Poltorak, A., Chan, E. K., McBride, C., & Beutler, B. (2000). Phylogenetic variation and polymorphism at the Toll-like receptor 4 locus (TLR4). *Genome Biology*, *1*(1), research002.1–research002.10.

Takeda, K., Kaisho, T., & Akira, S. (2003). Toll-like receptors. *Annual Review of Immunology*, *21*, 335–376. https://doi.org/10.1146/annurev.immunol.21.120601.141126

Tien, M. Z., Meyer, A. G., Sydykova, D. K., Spielman, S. J., & Wilke, C. O. (2013). Maximum allowed solvent accessibilites of residues in proteins. *PLoS One*, *8*(11), e80635. https://doi.org/10.1371/journal.pone.0080635

Tong, C., Lin, Y., Zhang, C., Shi, J., Qi, H., & Zhao, K. (2015). Transcriptome-wide identification, molecular evolution and expression analysis of Toll-like receptor family in a Tibet fish, *Gymnocypris przewalskii*. *Fish & Shellfish Immunology*, *46*(2), 334–345. https://doi.org/10.1016/j.fsi.2015.06.023

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

Vinkler, M., Bainová, H., & Bryja, J. (2014). Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. *Genetics Selection Evolution*, *46*(1), 72. https://doi.org/10.1186/s12711-014-0072-6

Vinkler, M., Bryjová, A., Albrecht, T., & Bryja, J. (2009). Identification of the first Toll-like receptor gene in passerine birds: TLR4 orthologue in zebra finch (*Taeniopygia guttata*). *Tissue Antigens*, *74*(1), 32–41. https://doi.org/10.1111/j.1399-0039.2009.01273.x

Waldenström, J., Bensch, S., Kiboi, S., Hasselquist, D., & Ottosson, U. (2002). Cross-species infection of blood parasites between resident and migratory songbirds in Africa. *Molecular Ecology*, *11*(8), 1545–1554. https://doi.org/10.1046/j.1365-294X.2002.01523.x

Walsh, C., Gangloff, M., Monie, T., Smyth, T., Wei, B., McKinley, T. J., … Bryant, C. (2008). Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa. *The Journal of Immunology*, *181*(2), 1245–1254. https://doi.org/10.4049/jimmunol.181.2.1245

Wang, X., Quinn, P. J., & Yan, A. (2015). Kdo_2-lipid A: Structural diversity and impact on immunopharmacology: Kdo_2-lipid A. *Biological Reviews*, *90*(2), 408–427. https://doi.org/10.1111/brv.12114

Wang, J., Zhang, Z., Liu, J., Zhao, J., & Yin, D. (2016). Ectodomain architecture affects sequence and functional evolution of vertebrate Toll-like receptors. *Scientific Reports*, *6*, 26705. https://doi.org/10.1038/srep26705

Wlasiuk, G., & Nachman, M. W. (2010). Adaptation and constraint at Toll-like receptors in primates. *Molecular Biology and Evolution*, *27*(9), 2172–2186. https://doi.org/10.1093/molbev/msq104

Woolhouse, M. E., Webster, J. P., Domingo, E., Charlesworth, B., & Levin, B. R. (2002). Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics*, *32*(4), 569–577. https://doi.org/10.1038/ng1202-569

Yang, Z. (2007). PAML 4: Phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591. https://doi.org/10.1093/molbev/msm088

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., & Zhang, Y. (2015). The I-TASSER suite: Protein structure and function prediction. *Nature Methods*, *12*(1), 7–8. https://doi.org/10.1038/nmeth.3213

Yilmaz, A., Shen, S., Adelson, D. L., Xavier, S., & Zhu, J. J. (2004). Identification and sequence analysis of chicken Toll-like receptors. *Immunogenetics*, *56*(10), 743–753. https://doi.org/10.1007/s00251-004-0740-8

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Králová T, Albrecht T, Bryja J, et al. Signatures of diversifying selection and convergence acting on passerine Toll-like receptor 4 in an evolutionary context. *Mol Ecol*. 2018;27:2871–2883. https://doi.org/10.1111/mec.14724

PAPER IV.

Bainová H., Králová T., Bryjová A., Albrecht T., Bryja J. & Vinkler M. (2014): First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds. Developmental & Comparative Immunology 45(1):151-155. (IF_{2014}= 2.815)

Toll-like receptor 5 (TLR5) is a Pattern-recognition receptor responsible for microbial flagellin detection in vertebrates and, hence, recognition of potentially pathogenic bacteria. In this paper, we report emergence of *TLR5* pseudogene in several phylogenetic lineages of passerine birds. Out of 47 species examined in this study 18 possessed a *TLR5* pseudogene. Phylogenetic analysis together with the type of mutation responsible for pseudogenization indicate that *TLR5* pseudogene emerged at least seven times independently in passerines. Lack of any functional copy of the gene has been verified based on *TLR5* mRNA blood expression in four species representing the four main passerine lineages possessing the *TLR5* pseudogene. Our results suggest that the non-functional TLR5 variant is fixed in those lineages or, at least, that individuals which are homozygote in the *TLR5* pseudogene are frequent in the investigated species. This is the first evidence of *TLR5* pseudogenization in passerine birds.

Developmental and Comparative Immunology 45 (2014) 151–155

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Short communication

First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds

Hana Bainová [a,*], Tereza Králová [b,c], Anna Bryjová [b], Tomáš Albrecht [a,b], Josef Bryja [b,c], Michal Vinkler [a]

[a] Department of Zoology, Faculty of Science, Charles University in Prague, Viničná 7, 128 43 Prague, Czech Republic
[b] Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, v.v.i., Květná 8, 603 65 Brno, Czech Republic
[c] Department of Botany and Zoology, Faculty of Science, Masaryk University, Kotlářská 267/2, 611 37 Brno, Czech Republic

ARTICLE INFO

Article history:
Received 17 January 2014
Revised 19 February 2014
Accepted 19 February 2014
Available online 6 March 2014

Keywords:
Birds
Expression
Innate immunity
Toll-like receptor 5
Pseudogene
Flagellin

ABSTRACT

Toll-like receptor 5 (TLR5) is a Pattern-recognition receptor responsible for microbial flagellin detection in vertebrates and, hence, recognition of potentially pathogenic bacteria. Herein, we report emergence of *TLR5* pseudogene in several phylogenetic lineages of passerine birds (Aves: Passeriformes). Out of 47 species examined in this study 18 possessed a *TLR5* pseudogene. Phylogenetic analysis together with the type of mutation responsible for pseudogenization indicate that *TLR5* pseudogene emerged at least seven times independently in passerines. Lack of any functional copy of the gene has been verified based on *TLR5* mRNA blood expression in four species representing the four main passerine lineages possessing the *TLR5* pseudogene. Our results suggest that the non-functional TLR5 variant is fixed in those lineages or, at least, that individuals homozygote in the *TLR5* pseudogene are frequent in the investigated species. Further research is needed to assess the impact of the TLR5 loss on immunological performance in birds.

1. Introduction

Toll-like receptor 5 (TLR5) is a Pattern recognition receptor (PRR) responsible for bacterial flagellin detection in birds and other vertebrates (Keestra et al., 2008; Temperley et al., 2008). Appropriate TLR5 function is, therefore, important for timely activation of immunity during various bacterial pathogen infections (Iqbal et al., 2005). Although generally the TLR5 function is much conserved (Keestra et al., 2013), it has been shown that both avian and mammalian TLR5 are interspecifically highly variable and that the *TLR5* sequence is shaped by natural selection leading to conservatism at some positions and diversification at others (Alcaide and Edwards, 2011; Ruan et al., 2012; Smith et al., 2012; Wlasiuk et al., 2009). Interestingly, polymorphism in TLR5 can be found even within individual species investigated. This holds true for humans (Hawn et al., 2003), domestic animals (Kathrani et al., 2011; Seabury et al., 2010; Shinkai et al., 2006; Yang et al., 2013) and free-living species (Alcaide and Edwards, 2011). Interspecific variation as well as intraspecific polymorphism in TLR5 has been shown to influence immune function (Kathrani et al., 2012; Keestra et al., 2008; Merx et al., 2006; Sheridan et al., 2013) and resistance to infectious and inflammatory diseases (Kathrani et al., 2011; Shinkai et al., 2011). Intriguingly, in humans where population variability of *TLR5* gene is well documented a non-sense substitution (R392X) has been reported to be responsible for *TLR5* pseudogenization in one particular allele (Hawn et al., 2003). The frequency of the *TLR5*-pseudogene allele is between 0% and 23% in different human populations (Wlasiuk et al., 2009). Individuals bearing the *TLR5* pseudogene have been shown to be more susceptible to the Legionnaires' disease caused by flagellated bacteria *Legionella pneumophila* (Hawn et al., 2003). Although the non-sense substitution impairs immune response to flagellin it does not seem that this reduced immunological capacity has any substantial effect on the lifetime health of the individuals (Merx et al., 2006). It has been even proposed that the individuals bearing the non-functional allele of *TLR5* may suffer less frequently from some inflammatory disorders such as Systemic lupus erythematosus or Crohn's disease (Gewirtz et al., 2006; Hawn et al., 2005). Nonetheless, there is presently no clear evidence that the non-functional

Abbreviations: CaCh, *Carduelis chloris*; cDNA, complementary DNA; CDS, coding DNA sequence; ESM, Electronic Supplementary Material; gDNA, genomic DNA; HiRu, *Hirundo rustica*; MoAl, *Motacilla alba*; MoCi, *Motacilla cinerea*; mRNA, Messenger RNA; NCBI, National Center for Biotechnology Information; PaDo, *Passer domesticus*; PaMa, *Parus major*; PCR, Polymerase Chain Reaction; PRR, Pattern recognition receptor; SyAt, *Sylvia atricapilla*; TaGu, *Taeniopygia guttata*; TLR, Toll-like receptor; TuMe, *Turdus merula*.

* Corresponding author. Tel.: +420 221951845; fax: +420 221951841.
E-mail addresses: bainova@natur.cuni.cz (H. Bainová), kralova.tereza@mail.muni.cz (T. Králová), bryjova@ivb.cz (A. Bryjová), albrecht@ivb.cz (T. Albrecht), bryja@brno.cas.cz (J. Bryja), michal.vinkler@natur.cuni.cz (M. Vinkler).

http://dx.doi.org/10.1016/j.dci.2014.02.010
0145-305X/

allele of *TLR5* in human has been maintained through selection (Wlasiuk et al., 2009). On the contrary, genetic drift is believed to be responsible for increased frequencies of the *TLR5* pseudogene in some human populations (Wlasiuk et al., 2009). Birds, except for the domestic fowl, are far less studied than mammals. In birds the only report of *TLR5* pseudogenes concerns two alleles described in house finch (*Haemorhous mexicanus*) and in white-fronted amazon (*Amazona albifrons*; Alcaide and Edwards, 2011) and one partial predicted sequence of *TLR5* in medium ground finch (*Geospiza fortis*) available in GenBank encodes clearly a pseudogene (NCBI: XM_005431638).

Encouraged by these accidental reports of *TLR5* pseudogenization in vertebrates, in the present study we searched systematically for the presence of non-sense substitutions in *TLR5* coding DNA sequence (CDS) in passerine birds. Our aim was to identify passerine lineages lacking the functional *TLR5* and assess whether the current state represents a result of a single random pseudogenization event or if there might have been a selection for recurrent loss of the gene. This knowledge is potentially important with respect to investigation of avian innate immunity evolution as it may bring us closer to understanding the past and ongoing positive selection for gene loss. We sequenced selected regions of the *TLR5* CDS in 43 passerine species representing 22 families and examined the pattern of all events of *TLR5* pseudogenization detected. To confirm the presence or absence of a functional copy of *TLR5*, we assessed *TLR5* expression in eight passerine species representing different passerine phylogenetic lineages exhibiting either a putatively functional *TLR5* gene or the *TLR5* pseudogene.

2. Materials and methods

2.1. Genetic samples

For the *TLR5* pseudogene screening we used genetic samples from 47 species distributed throughout passerine phylogeny (43 newly sequenced and 4 downloaded from public databases; one individual per species, for the list see Table S1 in Electronic Supplementary Material, ESM). Based on the sequence data we subsequently selected 4 species possessing a functional copy of the gene (*Parus major*, *Taeniopygia guttata*, *Sylvia atricapilla*, *Turdus merula*) and 4 species with a putative pseudogene (*Hirundo rustica*, *Passer domesticus*, *Carduelis chloris*, *Motacilla cinerea*) for the *TLR5* mRNA expression analysis. Both the blood and embryonic samples used for DNA and RNA extraction in the present study originated from the established collections of genetic samples of the Avian Evolutionary Ecology Group, Charles University in Prague, Czech Republic, EU and the Natural History Museum, University of Oslo, Norway, where samples for DNA extraction are stored in 96% ethanol at −20 °C and samples for RNA extraction are stored in RNA later (Qiagen, Hilden, Germany) at −80 °C.

2.2. Gene identification

Genomic DNA (gDNA) was extracted using the DNeasy Blood & Tissue Kit (Qiagen) from blood samples of 43 passerine species. Primers were designed to amplify part of the single *TLR5* exon and are based on alignment of known avian *TLR5* sequences (especially *Gallus gallus*, NCBI: EU052290 and *Taeniopygia guttata*, NCBI: XM_002188726) with respect to interspecifically conservative regions encoding parts of the extracellular part of *TLR5* (using web tool OligoAnalyzer 3.1, Integrated DNA Technologies, Inc., Coralville, Iowa, USA). The extracellular-domain-encoding sequence was chosen based on our preliminary research showing that the pseudogene stop codons and indels are present in this region (for details on positions and sequences of the primers see Table S2A in ESM). For gDNA amplification the Multiplex PCR Kit was used (Qiagen), the PCR conditions are given in Table S2A in ESM. All PCR products were Sanger-sequenced bidirectionally using the BigDye Terminator chemistry (Applied Biosystems, Life Technologies Corporation, Foster City, California, USA). Raw sequences were aligned and manually checked in SeqScape® Software v2.5 (Applied Biosystems). Translation into the correct protein sequence was performed using the Translate Tool (ExPASy, SIB Bioinformatics Resource Portal; Artimo et al., 2012) and the presence of indels and stop codons was thereafter analysed in BioEdit Sequence Alignment Editor (Tom Hall, Ibis Biosciences, Carlsbad, California, USA). All sequences have been uploaded to the National Center for Biotechnology Information (NCBI) GenBank database (for accession numbers see Table S1 in ESM). In the article, each species-specific sequence is entitled according to the first two letters from the genus and species name, e.g. *TaGuTLR5* for *TLR5* of *Taeniopygia guttata* (for the complete list see Table S1 in ESM).

2.3. Expression analysis

It has been recently reported that avian erythrocytes constitutively express mRNA transcripts for TLR5 (St Paul et al., 2013). In the present study we, therefore, utilised blood samples to verify the TLR5 expression. The blood samples of 8 passerine species (one individual per species, see Section 2.1 above for the list) were stored in RNAlater solution (Qiagen) at −80 °C. Total RNA was extracted from the blood samples using the High Pure RNA Tissue Kit (Roche, Basel, Switzerland), RNA was subsequently treated with RNase-free DNase I (Thermo Fisher Scientific Inc., Waltham, Massachusetts, USA) to remove DNA contamination. Reverse transcription of total mRNA to complementary DNA (cDNA) was performed immediately with the Transcriptor First Strand cDNA Synthesis Kit (Roche) using random hexamer primers. A 577 bp long fragment of cDNA was amplified by PCR using the Multiplex PCR Kit (Qiagen) and the same pair of primers (avianTLR5-F and avianTLR5-R; see Table S2B in ESM) and the same PCR conditions (initial denaturation at 95 °C for 15 min, 40 cycles of template denaturation at 94 °C for 30 s, primer annealing at 58 °C for 60 s, DNA elongation at 72 °C for 40 s and after last cycle the final elongation at 72 °C for 10 min) for all investigated individuals. The presence of PCR product was then verified by gel electrophoresis. We have done at least two independent PCR repetitions per individual to confirm our results. To further verify that our results do not represent merely artefacts resulting from the variability in *TLR5* expression in blood-borne cells we assessed the expression of the gene also in mixed embryonic samples originated from four species (*Hirundo rustica*, *Passer domesticus*, *Parus major* and *Taeniopygia guttata*). The homogenised samples were treated in the same way as the blood samples mentioned above in this paragraph.

2.4. Visualisation of TLR5 pseudogenization in passerine phylogenetic tree

A phylogenetic tree including all investigated species was constructed based on consensus avian phylogenetic tool available at http://birdtree.org/ (Jetz et al., 2012). As outgroups we used two parrot species with publicly available *TLR5* orthologue sequences (*Amazona albifrons*, NCBI: GU904708 and *Melopsittacus undulatus*, NCBI: XM_005151875). As the source of our tree the Hackett All Species tree with 1000 of randomly generated trees was used. The best credible tree was then determined using the tool TreeAnnotator v1.8.0 in the software BEAST v1.8.0 (Drummond and Rambaut, 2007). This tree was then graphically adjusted in FigTree v1.3.1 (Andrew Rambaut, University of Edinburgh, UK; http://tree.bio.ed.ac.uk/software/figtree/).

3. Results and discussion

In the present study we describe novel partial nucleotide sequences of the *TLR5* gene in 43 species representing 22 distinct passerine families (in all cases single individual per species was sequenced). In total, together with sequences currently available in public databases we analysed 47 passerine species representing 23 families. The investigated region spanning between bp 150 and bp 700 of the *TLR5* nucleotide sequence (numbered according to the zebra finch sequence, *TaGuTLR5*, NCBI: XM_002188726) comprises partial CDS of the TLR5 extracellular domain that has a crucial role in appropriate flagellin recognition (Keestra et al., 2008). In this region we detected multiple stop codons in 18 species belonging to 7 families (i.e. Hirundinidae, Fringillidae, Thraupidae, Emberizidae, Motacillidae, Passeridae and Muscicapidae; positions of the detected stop codons are visualised in Fig. 1 and in Fig. S1 in ESM), which allowed opening of frequent frame shift indels. In most cases the frame shift indels follow the initial termination codons, except for the first 13-bp insertion in wagtails (*MoAlTLR5*, NCBI: KF995065; *MoCiTLR5*, NCBI: KF995066) which directly contains the first stop codon. Given the phylogenetic relationships among these species and sequence variability in the position of the termination codons we have clearly detected at least seven independent events of *TLR5* pseudogenization in passerines (Fig. 1). Interestingly, the only other avian *TLR5* pseudogene described in the literature has been reported in the parrot white-fronted amazon (Alcaide and Edwards, 2011), which suggests that pseudogenes may be found also in avian orders closely related to passerines.

To test the existence or non-existence of a functional copy of the *TLR5* gene in 8 selected species (4 species with the functional gene and 4 species possessing the pseudogene) the expression of *TLR5* in blood has been investigated. Blood is a tissue where *TLR5* expression should be detected in birds (St Paul et al., 2013). The expression analysis fully supported our previous sequencing results. Although amplification of the target *TLR5* region was possible in all investigated species when using the gDNA template, this was not the case when the blood-derived cDNA was utilised. The target *TLR5* region could not be amplified from cDNA specifically in the species possessing the putative *TLR5* pseudogene (Fig. 2). In grey wagtail (*MoCiTLR5*) a 398 bp long insertion between bp 2178 and bp 2179 (numbered according to zebra finch, *TaGuTLR5*, NCBI: XM_002188726) prolonged the PCR product from gDNA visualised on the gel (Fig. 2; the insertion was verified by sequencing). The

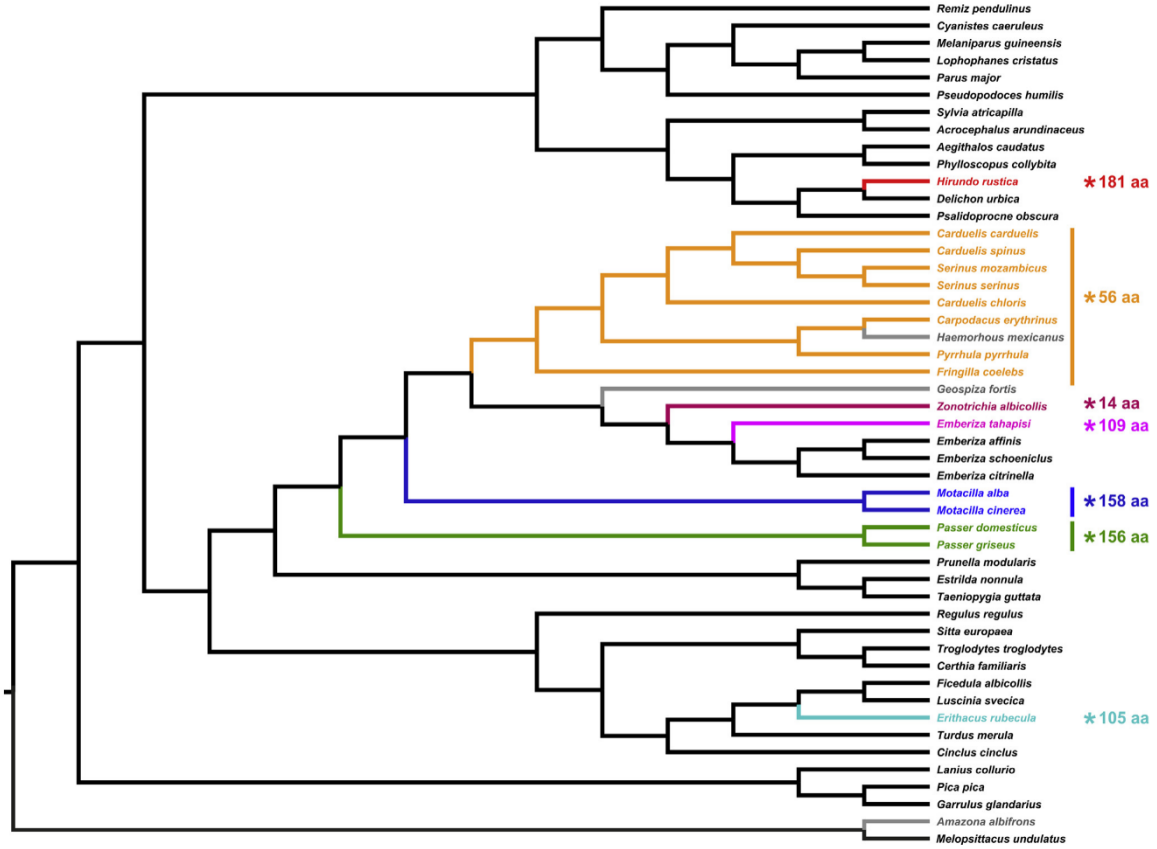


Fig. 1. Independent *TLR5* pseudogenization in passerine phylogeny. The phylogenetic tree of the investigated passerine species has been constructed based on consensus avian phylogenetic tool available at http://birdtree.org/ (Jetz et al., 2012). Species with common origin of the *TLR5* pseudogene are highlighted in identical colours (online version only; the same colours are used also in Fig. S1 in ESM). Species with previously described *TLR5* pseudogene but with another non-overlapping *TLR5* sequenced region available are highlighted in grey (online version only; *Haemorhous mexicanus*, NCBI: GU904827; *Geospiza fortis*, NCBI: XM_005431638; *Amazona albifrons*, NCBI: GU904708) because it was not possible to precisely identify the pseudogene homology in these species. The amino acid positions of the first *TLR5* stop codons are given behind the asterisks (the stop codons are numbered according to the translated zebra finch *TLR5* sequence, NCBI: XM_002188726). Parrots (XM_005151875 and GU904708) have been added to the phylogenetic tree as an out-group. The provided tree does not represent the phylogenetic gene tree of the *TLR5* orthologues.

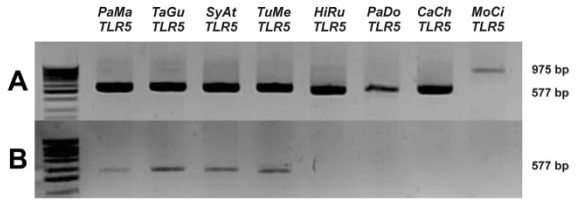


Fig. 2. *TLR5* expression in selected passerine species. Gel electrophoresis image of PCR products amplified with the same primer pair (see Section 2) for all investigated species (*PaMaTLR5* – *Parus major*, *TaGuTLR5* – *Taeniopygia guttata*, *SyAtTLR5* – *Sylvia atricapilla*, *TuMeTLR5* – *Turdus merula*, *HiRuTLR5* – *Hirundo rustica*, *PaDoTLR5* – *Passer domesticus*, *CaChTLR5* – *Carduelis chloris*, *MoCiTLR5* – *Motacilla cinerea*). Genomic (A) and complementary (B) DNA has been used as a template. Longer PCR product in MoCiTLR5 is caused by 398 bp long insertion in the amplified region of *TLR5* in this species.

later analysis in the mixed embryonic samples corroborated that the expression pattern is not an artefact resulting from differential expression of *TLR5* in the blood-borne cells. The results of the expression analysis thus confirm that the identified passerine species possessing the *TLR5* pseudogene are unlikely to possess another functional copy of the *TLR5* gene.

Toll-like receptors represent evolutionarily interesting protein family involved in animal immune defence (Keestra et al., 2013; Vinkler and Albrecht, 2009). Taken altogether, our results show that several passerine lineages independently evolved a non-functional TLR5 receptor which might be associated with impaired flagellin recognition. Whether the *TLR5* pseudogenization is adaptive and whether, indeed, the immune responsiveness to flagellin and flagellated bacteria in the passerine species possessing the *TLR5* pseudogene is weakened, however, remains unresolved. In our opinion adaptive pseudogenization of *TLR5* in passerines is less likely. Firstly, the species possessing the *TLR5* pseudogene are highly diverse and do not seem to share virtually any common ecological features. Secondly, some other receptors may detect the same spectrum of pathogens as TLR5 by redundancy of innate immunity. Some bacterial pathogens are known to avoid TLR5 detection even in species with fully functional TLR5 (Keestra et al., 2013). It has been shown in mammals that bacterial flagellin may be sensed by leukocytes using NLRC4 inflammasome receptors that are functionally independent on TLR5 (Zhao et al., 2011). It is, therefore, possible that some passerine species might have evolved alternative means of detection of infections caused by flagellated bacteria.

4. Conclusions

TLR5 is one of the important PRRs that allow recognition of invading bacteria and initiation of a subsequent innate and adaptive immune response. Although *TLR5* pseudogene has been known in humans, there is currently no example of the fixation of the non-functional *TLR5* pseudogene in mammals or any other vertebrate lineage. Our results suggest that in passerine birds the *TLR5* gene underwent several independent pseudogenization events associated with fixation of the non-functional allele. We hypothesise that the *TLR5* pseudogenization might be associated with impaired flagellin detection and occurs only in taxa where either the risk of infection by flagellated bacteria is low or where other PRRs are sufficient to ensure an adequate immune response. Further research is, however, needed to functionally address the impact of TLR5 loss on immune responsiveness to flagellin in the TLR5-deficient species.

Acknowledgements

We would like to thank especially to Arild Johnsen, Jan T. Lifjeld and Oldřich Tomášek for providing us with several DNA samples, Zdena Csibreiová for her help in laboratory and to Dagmar Vinklerová for language correction. This study was supported by the Czech Science Foundation (project no. P505/10/1871) and SVV-2013-267 201. The authors' contribution to this paper was as follows: HB (35%) – study design, molecular analyses, main part in manuscript preparation, TK (20%) – molecular analyses, AB (15%) – molecular analyses, TA (5%) – samples collection, JB (5%) – study design and MV (20%) – study design, samples collection, manuscript preparation. All authors contributed by their comments to the manuscript preparation and there was no conflict of interests in this research.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.dci.2014.02.010.

References

Alcaide, M., Edwards, S.V., 2011. Molecular evolution of the Toll-like receptor multigene family in birds. Mol. Biol. Evol. 28, 1703–1715.
Artimo, P., Jonnalagedda, M., Arnold, K., Baratin, D., Csardi, G., de Castro, E., Duvaud, S., Flegel, V., Fortier, A., Gasteiger, E., Grosdidier, A., Hernandez, C., Ioannidis, V., Kuznetsov, D., Liechti, R., Moretti, S., Mostaguir, K., Redaschi, N., Rossier, G., Xenarios, I., Stockinger, H., 2012. ExPASy: SIB bioinformatics resource portal. Nucleic Acids Res. 40, W597–W603.
Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7.
Gewirtz, A.T., Vijay-Kumar, M., Brant, S.R., Duerr, R.H., Nicolae, D.L., Cho, J.H., 2006. Dominant-negative TLR5 polymorphism reduces adaptive immune response to flagellin and negatively associates with Crohn's disease. Am. J. Physiol.-Gastroint. Liver Physiol. 290, G1157–G1163.
Hawn, T.R., Verbon, A., Lettinga, K.D., Zhao, L.P., Li, S.S., Laws, R.J., Skerrett, S.J., Beutler, B., Schroeder, L., Nachman, A., Ozinsky, A., Smith, K.D., Aderem, A., 2003. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. J. Exp. Med. 198, 1563–1572.
Hawn, T.R., Wu, H., Grossman, J.M., Hahn, B.H., Tsao, B.P., Aderem, A., 2005. A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. Proc. Natl. Acad. Sci. USA 102, 10593–10597.
Iqbal, M., Philbin, V.J., Withanage, G.S.K., Wigley, P., Beal, R.K., Goodchild, M.J., Barrow, P., McConnell, I., Maskell, D.J., Young, J., Bumstead, N., Boyd, Y., Smith, A.L., 2005. Identification and functional characterization of chicken Toll-like receptor 5 reveals a fundamental role in the biology of infection with *Salmonella enterica serovar typhimurium*. Infect. Immun. 73, 2344–2350.
Jetz, W., Thomas, G.H., Joy, J.B., Hartmann, K., Mooers, A.O., 2012. The global diversity of birds in space and time. Nature 491, 444–448.
Kathrani, A., Holder, A., Catchpole, B., Alvarez, L., Simpson, K., Werling, D., Allenspach, K., 2012. TLR5 risk-associated haplotype for canine inflammatory bowel disease confers hyper-responsiveness to Flagellin. PLoS ONE 7.
Kathrani, A., House, A., Catchpole, B., Murphy, A., Werling, D., Allenspach, K., 2011. Breed-independent Toll-like receptor 5 polymorphisms show association with canine inflammatory bowel disease. Tissue Antigens 78, 94–101.
Keestra, A.M., de Zoete, M.R., Bouwman, L.I., Vaezirad, M.M., van Putten, J.P.M., 2013. Unique features of chicken Toll-like receptors. Dev. Comp. Immunol. 41, 316–323.
Keestra, A.M., de Zoete, M.R., van Aubel, R.A.M.H., van Putten, J.P.M., 2008. Functional characterization of chicken TLR5 reveals species-specific recognition of flagellin. Mol. Immunol. 45, 1298–1307.
Merx, S., Zimmer, W., Neumaier, M., Ahmad-Nejad, P., 2006. Characterization and functional investigation of single nucleotide polymorphisms (SNPs) in the human TLR5 gene. Hum. Mutat. 27, 293.
Ruan, W.K., Wu, Y.H., Zheng, S.J.J., 2012. Different genetic patterns in avian Toll-like receptor (TLR)5 genes. Mol. Biol. Rep. 39, 3419–3426.
Seabury, C.M., Seabury, P.M., Decker, J.E., Schnabel, R.D., Taylor, J.F., Womack, J.E., 2010. Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. Proc. Natl. Acad. Sci. USA 107, 151–156.
Sheridan, J., Mack, D.R., Amre, D.K., Israel, D.M., Cherkasov, A., Li, H.F., Grimard, G., Steiner, T.S., 2013. A non-synonymous coding variant (L616F) in the TLR5 gene is potentially associated with Crohn's disease and influences responses to bacterial flagellin. PLoS ONE 8.
Shinkai, H., Suzuki, R., Akiba, M., Okumura, N., Uenishi, H., 2011. Porcine Toll-like receptors: recognition of *Salmonella enterica serovar Choleraesuis* and influence of polymorphisms. Mol. Immunol. 48, 1114–1120.

Shinkai, H., Tanaka, M., Morozumi, T., Eguchi-Ogawa, T., Okumura, N., Muneta, Y., Awata, T., Uenishi, H., 2006. Biased distribution of single nucleotide polymorphisms (SNPs) in porcine Toll-like receptor 1 (TLR1), TLR2, TLR4, TLR5, and TLR6 genes. Immunogenetics 58, 324–330.

Smith, S.A., Jann, O.C., Haig, D., Russell, G.C., Werling, D., Glass, E.J., Emes, R.D., 2012. Adaptive evolution of Toll-like receptor 5 in domesticated mammals. BMC Evol. Biol. 12.

St Paul, M., Paolucci, S., Barjesteh, N., Wood, R.D., Sharif, S., 2013. Chicken erythrocytes respond to Toll-like receptor ligands by up-regulating cytokine transcripts. Res. Vet. Sci. 95, 87–91.

Temperley, N.D., Berlin, S., Paton, I.R., Griffin, D.K., Burt, D.W., 2008. Evolution of the chicken Toll-like receptor gene family: a story of gene gain and gene loss. BMC Genomics 9, 62.

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.

Wlasiuk, G., Khan, S., Switzer, W.M., Nachman, M.W., 2009. A history of recurrent positive selection at the Toll-like receptor 5 in primates. Mol. Biol. Evol. 26, 937–949.

Yang, X.Q., Li, H.T., Guan, Q.Z., Liu, D., 2013. Genetic diversity of Toll-like receptor 5 among pig populations. Genet. Mol. Biol. 36, 37–42.

Zhao, Y., Yang, J., Shi, J., Gong, Y.-N., Lu, Q., Xu, H., Liu, L., Shao, F., 2011. The NLRC4 inflammasome receptors for bacterial flagellin and type III secretion apparatus. Nature 477, 596-U257.

PAPER V.

Vinkler M., **Bainová H.**, Bryjová A., Tomášek O., Albrecht T. & Bryja J. (2015): Characterization of Toll-like receptors 4, 5 and 7 and their polymorphism in grey partridge. Genetica 143(1):101-112. (IF_{2015}= 1.343)

In this paper, we identified orthologues of three innate immune receptors: TLR4, TLR5 and TLR7 (representing both bacterial- and viral-sensing TLRs) in the grey partridge (*Perdix perdix*). To verify functionality in these genes we mapped their tissue-expression profiles, revealing generally high *PePeTLR4* and *PePeTLR5* expression in the thymus and absence of *PePeTLR4* and *PePeTLR7* expression in the brain. Using target NGS, we then assessed genetic variation within these genes for a wild grey partridge population in the Czech Republic. We identified surprisingly only very limited genetic variation in all investigated receptors. Moreover, given their locations and chemical features, most of the non-synonymous substitutions probably have only minor functional impact. As the intraspecific genetic variation of the three *TLR* genes was low, we assume that either negative selection or a bottleneck may have reduced *TLR* population variability in this species.

Genetica (2015) 143:101–112
DOI 10.1007/s10709-015-9819-4

Characterisation of Toll-like receptors 4, 5 and 7 and their genetic variation in the grey partridge

Michal Vinkler · Hana Bainová · Anna Bryjová · Oldřich Tomášek · Tomáš Albrecht · Josef Bryja

Received: 20 June 2013 / Accepted: 18 January 2015 / Published online: 28 January 2015

Abstract Toll-like receptors (TLRs) are a cornerstone of vertebrate innate immunity. In this study, we identified orthologues of *TLR4*, *TLR5* and *TLR7* (representing both bacterial- and viral-sensing TLRs) in the grey partridge (*Perdix perdix*), a European Galliform game bird species. The phylogeny of all three *TLR* genes follows the known phylogeny of Galloanserae birds, placing grey partridge *TLRs* (*PePeTLRs*) in close proximity to their turkey and pheasant orthologues. The predicted proteins encoded by the *PePeTLR* genes were 843, 862–863 and 1,047 amino acids long, respectively, and clearly showed all TLR structural features. To verify functionality in these genes we mapped their tissue-expression profiles, revealing generally high *PePeTLR4* and *PePeTLR5* expression in the thymus and absence of *PePeTLR4* and *PePeTLR7* expression in the brain. Using 454 next-generation sequencing, we then assessed genetic variation within these genes for a wild grey partridge population in the Czech Republic, EU. We identified 11 nucleotide substitutions in *PePeTLR4*, eight in *PePeTLR5* and six in *PePeTLR7*, resulting in four, four and three amino acid replacements, respectively. Given their locations and chemical features, most of these non-synonymous substitutions probably have a minor functional impact. As the intraspecific genetic variation of the three *TLR* genes was low, we assume that either negative selection or a bottleneck may have reduced TLR population variability in this species.

Keywords Gene transcription · Pattern recognition receptors · Polymorphism · Population decrease · Protein structure · Variability reduction

Electronic supplementary material The online version of this article (doi:10.1007/s10709-015-9819-4) contains supplementary material, which is available to authorized users.

M. Vinkler (✉) · H. Bainová · A. Bryjová · O. Tomášek · T. Albrecht
Department of Zoology, Faculty of Science, Charles University in Prague, Viničná 7, 128 44 Prague, Czech Republic, EU
e-mail: michal.vinkler@natur.cuni.cz

A. Bryjová · O. Tomášek · T. Albrecht · J. Bryja
Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, v.v.i., Květná 8, 603 65 Brno, Czech Republic, EU

Introduction

The efficiency of innate immunity depends a great deal on appropriate initial pathogen recognition and the key role of Pattern Recognition Receptors (PRRs) in this process is now well known (Palsson-McDermott and O'Neill 2007). Of these PRRs, Toll-like receptors (TLRs) represent the most extensively studied group of molecules responsible for detection of microbe-associated molecular patterns (Akira et al. 2006). These receptors are involved both in triggering innate immunity and in regulation of adaptive immunity (Iwasaki and Medzhitov 2010). Although intensively studied at large evolutionary scales (Roach et al. 2005; Temperley et al. 2008), and within the mammalian clade (Wlasiuk et al. 2009; Huang et al. 2011), there is only limited information on interspecific and intraspecific variability in these receptors in non-mammalian vertebrate taxa, including birds (Vinkler et al. 2009; Downing et al. 2010; Alcaide and Edwards 2011; Grueber et al. 2014).

Vertebrate taxa differ in the number and types of TLRs they possess (Roach et al. 2005). In birds, the *TLR* repertoire comprises ten genes (Temperley et al. 2008; Cormican et al. 2009; Brownlie and Allan 2011), with the possible exception of some passerine birds where *TLR5*

pseudogenised (Bainova et al. 2014) and *TLR7* duplication may have occurred (Grueber et al. 2012). All ten *TLR* genes have been identiÞed in the domestic chicken (*Gallus gallus domesticus*; Fukui et al. 2001; Leveque et al. 2003; Lynn et al. 2003; Smith et al. 2004; Iqbal et al. 2005; Philbin et al. 2005; Yilmaz et al. 2005), and most TLRs have also been identiÞed in the red junglefowl (*Gallus gallus gallus*), green junglefowl (*Gallus varius*), Ceylon junglefowl (*Gallus lafayetti*) and grey junglefowl (*Gallus sonneratii*) (Downing et al. 2010). Thanks to recent descriptions of their whole genome sequences, most of the *TLRs* were have also been predicted for zebra Þnch (*Taeniopygia guttata*; Warren et al. 2010), domestic turkey (*Meleagris gallopavo*; Dalloul et al. 2010) and several other avian species (although precise descriptions are presently only available for zebra Þnch *TLR4*, Vinkler et al. 2009, and turkey *TLR1LA*, *TLR1LB*, *TLR2A*, *TLR2B* and *TLR5*, Gopinath et al. 2011; Huang et al. 2011). Aside from these, there have been limited studies describing some *TLRs* in other avian species (e.g. MacDonald et al. 2008; Huang et al. 2011; Ramasamy et al. 2011; Wang et al. 2011; Elfeil et al. 2012; Jia et al. 2012; Ruan et al. 2012b; Uno et al. 2012; Zhao et al. 2013; Xiong et al. 2014). Despite many unique features (Keestra et al. 2013), most avian TLR proteins generally recognise similar ligands as their mammalian counterparts (Brownlie and Allan 2011), e.g. TLR4 binds lipopolysaccharide (Keestra and van Putten 2008), TLR5 ßagellin (Keestra et al. 2008) and TLR7 viral single-stranded RNA and synthetic antivirals (Philbin et al. 2005; Brownlie et al. 2009). It has been previously recognised that there may be variability in TLR-ligand binding and that TLR protein polymorphism may be functional (Vinkler and Albrecht 2009). IntraspeciÞc genetic variation in avian TLRs has only rarely been investigated (e.g. Leveque et al. 2003; Downing et al. 2010; Alcaide and Edwards 2011; Liu et al. 2011; Ruan et al. 2012b), however, leaving many questions on *TLR* variability in free-living bird populations unanswered.

Despite recent advances in the mapping of interspeciÞc and intraspeciÞc variability in avian *TLRs*, the number of species that can be included into comparative studies is still extremely low. This is especially true as regards information on full-length protein-coding DNA sequences (CDS) of *TLR* genes. In this study, we identiÞed the orthologues of *TLR4*, *TLR5* and *TLR7* (three model genes representing predominantly bacterial- and viral-sensing TLRs; Akira et al. 2006) in the grey partridge (*Perdix perdix*). This is the Þrst study to characterise any PRR in grey partridge, a non-model species representing an important clade within the extensively investigated Galliform order. We assessed the expression pattern and intraspeciÞc level of genetic variation in these genes within a natural wild population and evaluated the predicted functional impact of coding variability by estimating the chemical features of the residua and mapping the variable sites on modelled 3D tertiary structures of the PePeTLR proteins.

Materials and methods

Genetic samples

In order to describe the CDSs of *TLR4*, *TLR5* and *TLR7* in grey partridge (*PePeTLR4*, *PePeTLR5* and *PePeTLR7*) we used genetic samples from 10 individuals (unrelated birds from four different wintering ßocks) collected in Milešín, Czech Republic (N 49°22′06″, E 16°12′05″). Genomic DNA was isolated from blood using the DNeasy Blood and Tissue Kit (QIAGEN, Hilden, Germany) and stored in 96 % ethanol at −20 °C. To evaluate *TLR* expression, nine different tissue samples (thymus, bursa of Fabricius, bone marrow, spleen, caecal tonsils, lung, brain, liver and whole blood) were collected from two euthanised males and stored in RNAlater (QIAGEN) at −80 °C until total RNA extraction.

Gene sequencing

Polymerase Chain Reactions (PCR) were performed using Multiplex PCR Master Mix (QIAGEN). A detailed PCR protocol is given in the Electronic supplementary material (ESM; Section 1.2.1). For sequencing, we designed three amplicons for *TLR4* and *TLR7* and two amplicons for *TLR5* (see Table S1 in the ESM). The primers were initially designed using the OligoAnalyzer web tool (Integrated DNA Technologies, Inc., Coralville, IA, USA) based on gene sequence homology between domestic chicken and zebra Þnch and later optimised based on preliminary partial sequences obtained through Sanger sequencing of grey partridge DNA. We sequenced the genes and mapped the intraspeciÞc genetic variation of Þve grey partridges using next-generation pyrosequencing on an in-house bench-top 454 GS Junior platform (Roche, Basel, Switzerland; see Jiang et al. (2012); for a detailed description of the procedure see Section 1.2 in the ESM). We treated all sites with a 40–60 % frequency for the two nucleotide variants within all reads as possible single nucleotide polymorphic (SNP) sites (also termed “genetic variants” throughout the manuscript). As a control, all sequences were also checked manually in GS Reference Mapper via the graphical user interface. Genetic variation revealed by 454 sequencing was veriÞed through Sanger sequencing of those regions containing the putative SNPs using the BigDye Terminator Cycle Sequencing Kit v. 3.1 (Applied Biosystems, Foster City, CA, USA) on an ABI 3130 Genetic Analyser (Applied Biosystems). The resultant sequences were analysed using

SeqScape version 2.5 (Applied Biosystems). Variable positions were assigned to putative alleles based on the succession of variants in individual reads minimising the number of variants when combining regions with low levels of sequence variability (minimal plausible allele approach). Where possible, the original state was identiÞed based on the results of phylogenetic analysis. To further conÞrm population genetic variation in the TLRs (avoiding the possible effects of 454 reading errors) we then took samples from another Þve individuals and performed CDS PCR and Sanger sequencing, as described above (sequencing primers are provided in Table S2 in the ESM). Each putative SNP was independently sequenced at least twice to exclude false positives. The resultant sequences were assigned to alleles using the PHASE method implemented in the DnaSP v.5.10.01 software (Librado and Rozas 2009). We performed GARD analysis via the Datamonkey server (Kosakovsky Pond et al. 2006) in order to reveal recombination between alleles of individual genes. All *PePeTLRs* sequences are available in GenBank (www.ncbi.nlm.nih.gov/nuccore/) under accession numbers: JQ713172–77 and KC808003–06 for *PePeTLR4*, JQ713180–83 and KC808007–08 for *PePeTLR5* and JQ713178–79 and KC808009 for *PePeTLR7*.

Phylogenetic analysis

Orthology in the newly sequenced genes was identiÞed through a BLAST search (http://blast.ncbi.nlm.nih.gov/). Coding regions for the selected *TLRs* in ten other Galloanserae species (red junglefowl; Ceylon junglefowl; grey junglefowl; green junglefowl; domestic turkey; common pheasant *Phasianus colchicus*; guineafowl *Numida meleagris*; Japanese quail *Coturnix japonica*; mallard *Anas platyrhynchos*; and goose *Anser anser*), humans and mice were downloaded from the NCBI GenBank (GenBank IDs are given in Section 1.1 of the ESM). Nucleotide sequence alignments of the CDS regions for *TLR4*, *TLR5* and *TLR7* genes were used for phylogenetic analysis using the maximum likelihood method. Orthologous human and mouse sequences were used as outgroups. FindModel (http://www.hiv.lanl.gov/content/sequence/ÞndmodelÞndmodel.html) was used to evaluate the Þt of 28 nested models of nucleotide substitution to the data and the best model for each alignment was selected on the basis of the Akaike information criterion. Phylogenetic analysis was performed using PHYML (Guindon et al. 2005), using the NNI algorithm and BIONJ distance-based tree as the starting tree. Bootstrap analysis (1,000 replicates) was used to estimate the robustness of internal nodes. TLR phylogenetic trees were then compared with a consensus phylogenetic tree including all investigated species. This tree was constructed using the avian phylogenetic tool available at http://bird-tree.org/ (Jetz et al. 2012). We used the Hackett sequenced species backbone (Hackett et al. 2008) as the source tree with 1,000 randomly generated trees. The maximum clade credibility tree was produced using the TreeAnnotator v. 1.8.0 tool in BEAST v. 1.8.0 software (Drummond and Rambaut 2007). The results were visualised in FigTree v. 1.3.1 (http://tree.bio.ed.ac.uk/software/Þgtree/).

Protein structure prediction

For all structural analyses we used the *PePeTLR* sequences JQ713172, JQ713180 and JQ713178 as grey partridge references. In order to predict the general distribution of structural domains in the proteins we adopted a combination of the SMART (Letunic et al. 2012) and LRRÞnder (Offord et al. 2010) tools. With LRRÞnder, molecular weight was also calculated for each predicted protein. Presence of the transmembrane (TM) region was veriÞed using the DAS-TMÞlter server (Cserzo et al. 2004). Signal peptides were identiÞed using SignalP 4.0 server (Petersen et al. 2011). Finally, we used a composite modelling approach (comparative modelling and threading) using I-TASSER (Roy et al. 2010), a leading protein structure prediction server (see Zhang 2008 and http://predictioncenter.org/casp9/), to predict the 3D tertiary structure of receptor extracellular domains (ECDs). The I-TASSER server uses a hierarchical protein-structure modelling approach based on secondary-structure enhanced proÞle–proÞle threading alignment and iterative implementation of the threading assembly reÞnement program (Zhang 2008). Two models were created for TLR7: (1) a full-length protein, and (2) a sequence beginning with amino acid 417. We modelled the truncated form as it has previously been shown that, up to Leucine rich repeat 14 (LRR14), the sequence is cleaved in the endoplasmic reticulum (Ewald et al. 2008). In each case, we used models with the C-scores. Modelling error at the regions of interest was estimated in ModFOLD (McGufÞn 2008; all models had high conÞdence levels with P values <0.002 and Global model quality scores >0.39). Images of the predicted protein 3D structures were visualised using PyMOL software v. 1.5 (Schrödinger, München, Germany; http://pymol.org/).

Gene expression assay

Total RNA for *TLR* semi-quantitative expression analysis was isolated from the nine selected tissues (thymus, bursa of Fabricius, bone marrow, spleen, caecal tonsils, lung, brain, liver and blood) using the High Pure RNA Tissue Kit (Roche). For each sample, 0.1 µg of the total RNA was reverse-transcribed into cDNA using the Transcriptor First Strand cDNA Synthesis Kit (Roche) and random hexamer primers. Semi-quantitative PCR was performed with the Kapa2G Robust PCR kit (Kapa Biosystems, Inc., Woburn,

MA, USA), with primers designed based on the previously described *PePeTLR* sequences and alignment with Galliform birds (see Table S3 in the ESM). AmpliÞed products were identiÞed using 2 % agarose gel electrophoresis and visualised using GoodView (SBS Genetech, Beijing, China) and a UV Transilluminator (GenoPlex, VWR International, Radnor, PA, USA). PCR and subsequent gel electrophoresis were repeated three times for each gene and each tissue (triplicate design). To quantify relative differences in mRNA expression, we measured relative band intensity (GenoSoft v. 4.00, VWR International), normalised by *β-actin* values (*β-actin* was treated in the same way as *TLRs*; primers and PCR conditions provided in Table S3 in the ESM). For a similar approach see Gopinath et al. (2011).

Results

IdentiÞcation of *TLR4*, *TLR5* and *TLR7* in grey partridge

We sequenced complete CDS regions of *TLR4*, *TLR5* and *TLR7* for 10 grey partridges (non-synonymous allele GenBank IDs: JQ713172–83 and KC808003–KC808009). The CDS regions of *PePeTLR4* was 2,529 bp long, that for *PePeTLR5* 2,589–2,592 bp, and that for *PePeTLR7* 3,141 bp, encoding predicted amino acid (aa) sequences of 843 aa, 862–863 aa and 1,047 aa, respectively. Orthology, identiÞed through a protein BLAST search (Table 1), indicated that grey partridge genes were most closely clustered with those of turkey (*TLR4* and *TLR7*) or pheasant (*TLR5*; see Fig. 1). Comparison with other Galloanserae species revealed that the initial ATG indel in some *PePeTLR5* sequences resulted from the deletion of one of the ATG codons at the beginning of the CDS. Molecular weights of the predicted proteins were 110,887 Da for PePeTLR4, 114,241 Da for PePeTLR5 and 138,974 Da for PePeTLR7. Each protein sequence began with a signal peptide of 30 aa, 20 aa and 25 aa long, respectively. The PePeTLR4 ECD consisted of 18 LRRs, the PePeTLR5 ECD had 19 and the PePeTLR7 ECD had 27 (Table S4 in the ESM). Transmembrane domains were detected in all three proteins and the 3D tertiary structure ECD models displayed a typical horseshoe-shaped LRR distribution (Fig. 2). At the C-terminal end, each protein contained a Toll/interleukin-1 receptor (TIR) domain of

Table 1 *PePeTLR* homology with other Galloanserae TLR molecules

TLR	Species	GenBank ID	nt identity (%)	aa identity (%)	aa positives (%)
PePeTLR4	*Perdix perdix*	JQ713172			
AnAnTLR4	*Anser anser*	HQ436371.1	87	82	89
AnPlTLR4	*Anas platyrhynchos*	JN048668.1	86	82	89
GaGaTLR4	*Gallus gallus*	AY064697.1	96	95	97
GaLaTLR4	*Gallus lafayetii*	FJ915482.1	96	95	97
GaSoTLR4	*Gallus sonneratii*	FJ915508.1	96	95	97
GaVaTLR4	*Gallus varius*	FJ915504.1	96	95	97
MeGaTLR4	*Meleagris gallopavo*	XM_003211211.1	96	95	98
PePeTLR5	*Perdix perdix*	JQ713180			
AnAnTLR5	*Anser anser*	JN641303.1	88	82	90
GaGaTLR5	*Gallus gallus*	FJ915551.1	95	93	97
GaLaTLR5	*Gallus lafayetii*	FJ915530.1	95	93	97
MeGaTLR5	*Meleagris gallopavo*	HQ436463.1	95	93	96
NuMeTLR5	*Numida meleagris*	JF767221.1	93	91	95
PhCoTLR5	*Phasianus colchicus*	JF767220.1	97	96	98
PePeTLR7	*Perdix perdix*	JQ713178			
AnPlTLR7	*Anas platyrhynchos*	DQ888644.1	89	87	91
CoJaTLR7	*Coturnix japonica*	AB553582.1	94	94	97
GaGaTLR7	*Gallus gallus*	AJ632302.1	95	93	96
GaLaTLR7	*Gallus lafayetii*	FJ915556.1	95	93	96
GaSoTLR7	*Gallus sonneratii*	FJ915580.1	95	93	96
GaVaTLR7	*Gallus varius*	FJ915576.1	95	93	96
MeGaTLR7	*Meleagris gallopavo*	XM_003203086.1	97	97	98

Nucleotide (nt) and amino acid (aa) identity and interchangeability (aa positives; i.e. reßection of substitution conservatism) were calculated in BLAST

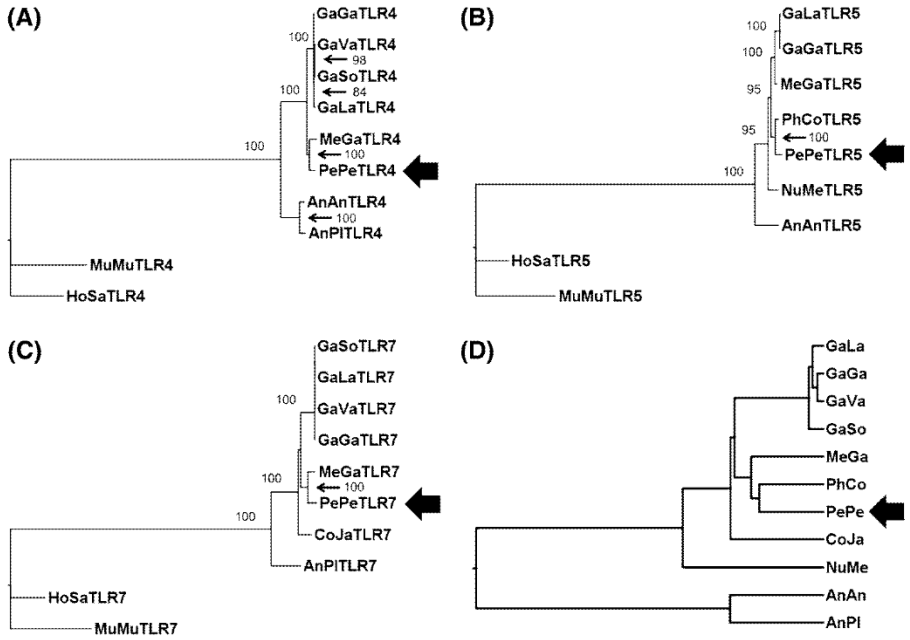


Fig. 1 Phylogeny of *TLR* genes within the Galloanserae. Maximum likelihood phylogeny of nucleotide sequences for *TLR4* (**a**), *TLR5* (**b**) and *TLR7* (**c**) in Galloanserae species calculated using PHYML online software. Orthologous human (HoSa) and mouse (MuMu) sequences were used as outgroups. Small numbers indicate the bootstrap values (only those higher than 75 are shown). **d** A consensus phylogenetic tree for the Galloanserae clade (only species involved in the TLR analysis are included) constructed using the avian phylogenetic tool (Jetz et al. 2012). The position of grey partridge is indicated by the bold arrows. AnAn = *Anser anser*, AnPl = *Anas platyrhynchos*, CoJa = *Coturnix japonica*, GaGa = *Gallus gallus*, GaLa = *Gallus lafayetii*, GaSo = *Gallus sonneratii*, GaVa = *Gallus varius*, MeGa = *Meleagris gallopavo*, NuMe = *Numida meleagris*, PhCo = *Phasianus colchicus*, PePe = *Perdix perdix*

145 aa (*PePeTLR4*), 147 aa (*PePeTLR5*) and 146 aa (*PePeTLR7*) long.

Expression of *PePeTLR4*, *PePeTLR5* and *PePeTLR7*

All three *PePeTLR* genes were expressed in most of the grey partridge tissues investigated, with the exception of *PePeTLR4* and *PePeTLR7*, which were undetected in brain tissue (Fig. 3 and Fig. S1 in the ESM). Expression of *PePeTLR4* was low in both lung and liver tissue samples and in the caecal tonsils of one individual. Highest *PePeTLR4* expression was detected in the thymus. *PePeTLR5* showed a relatively high expression in all samples aside from spleen. *PePeTLR7* expression was high in bone marrow and liver but low in lung. There was, however, signiÞcant variability between the two individuals analysed, with one individual having consistently higher levels of *PePeTLR7* expression and lower *PePeTLR4* expression in most immune organs (for details see Fig. 3).

Genetic variation in *TLR4*, *TLR5* and *TLR7* in grey partridge

In our grey partridge population, we revealed 11 *PePeTLR4*, eight *PePeTLR5* and six *PePeTLR7* nucleotide polymorphic positions (for the description of SNPs and their frequency, see Table 2; ECD amino acid substitutions are visualised on *PePeTLR* protein 3D models in Fig. 2). These genetic variants were clustered in 10 (*PePeTLR4*), 6 (*PePeTLR5*) and 3 (*PePeTLR7*) alleles, forming 5, 4 and 3 non-synonymous haplotypes, respectively (Table S5 in the ESM). As the same major haplotypes were found repeatedly, we have almost certainly described the most frequent protein-coding diversity in these genes in our study population (allele and haplotype frequencies are provided in Table S5 in the ESM). GARD analysis revealed one recombination breakpoint in *PePeTLR4* (following nucleotide position 675) and two possible breakpoints in *PePeTLR7* (following nucleotide positions 279 and 482; a signiÞcance level was impossible to calculate due to the low number of unique sequences included in the analysis). No recombination was detected in *PePeTLR5*.

Discussion

In this study, we characterised *TLR4*, *TLR5* and *TLR7* in the grey partridge, a non-model species representing an important Galliform lineage. We were able to identify eleven nucleotide substitutions in *PePeTLR4*, eight in

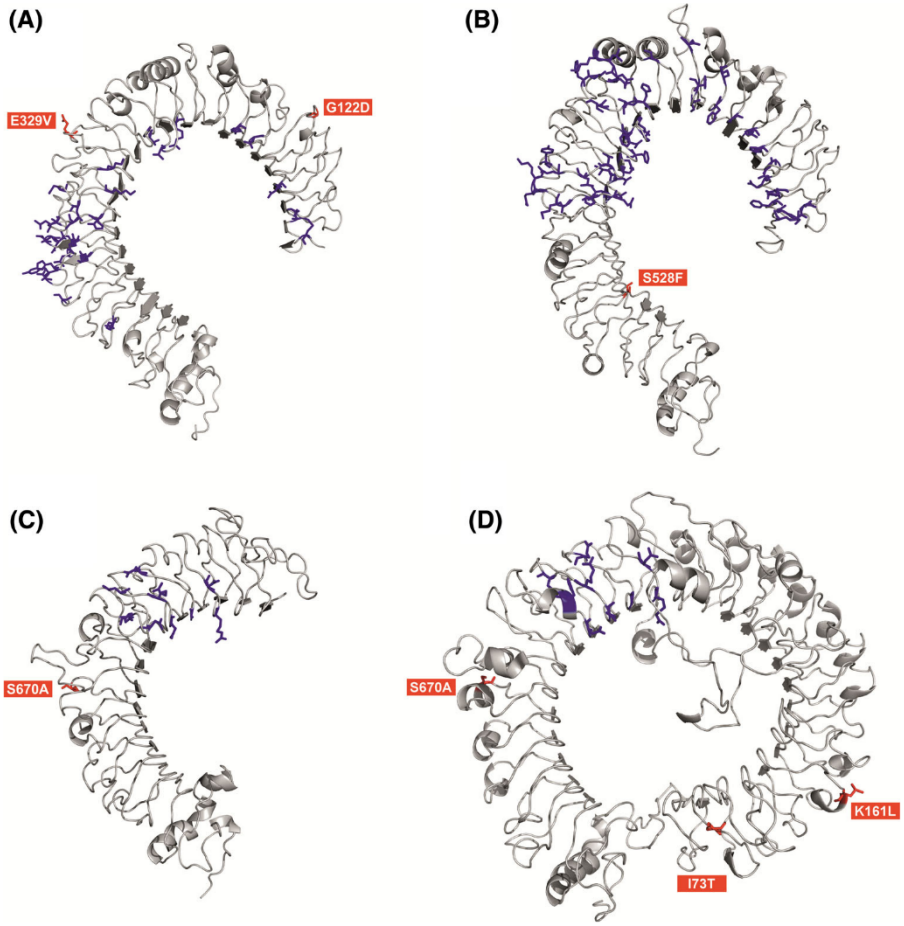


Fig. 2 (Color online) Projection of polymorphic sites on 3D models of PePeTLR ECDs: **a** TLR4, **b** TLR5, **c** TLR7-truncated form (see Ewald et al. 2008), **d** TLR7-full-length form. Polymorphic sites are highlighted by full-rectangles (*red*) while functionally important sites known from previous mammal and Þsh studies (Andersen-Nissen et al. 2007; Kim et al. 2007; Walsh et al. 2008; Park et al. 2009; Resman et al. 2009; Wei et al. 2009; Ohto et al. 2012; Yoon et al. 2012) are shown in *black* (*dark blue*)

PePeTLR5 and six in *PePeTLR7* in samples from a wild population, resulting in four, four and three amino acid replacements, respectively. Given their location and chemical features, most of these non-synonymous substitutions probably have a minor functional impact. Semi-quantitative analysis of gene expression revealed a unique pattern of tissue-speciÞc transcription in these genes, along with important conspeciÞc expression variability.

The grey partridge is a native European Galliform species, typical of agricultural landscapes. It has been intensively bred and re-introduced by both hunters and conservationists in many European countries (Vidus-Rosin et al. 2010; Buner et al. 2011). Most studies on this species have concentrated on population genetics (Andersen and Kahlert 2012) and behaviour (Beani and Dessifulgheri 1995; Svobodova et al. 2013), with relatively few immunological studies undertaken (but see Cucco et al. 2006, 2007; Vinkler et al. 2014a, b). Knowledge of the grey partridge immune system is, therefore, highly fragmentary; despite the fact that the grey partridge is susceptible to some human (e.g. *Campylobacter* or *Toxoplasma*; Sedlak et al. 2000; Horigan et al. 2014) and poultry (e.g. *Mycoplasma*; Vitula et al. 2011) pathogens and may be capable of transmitting diseases to humans and other species. Indeed, disease susceptibility is one of the likely causes of the recent population decline in the species (Tompkins et al. 2000, 2002; Kuijper et al. 2009). Detailed immunological research is needed in order to understand the interspeciÞc and intraspeciÞc immunological differences leading to variation in disease susceptibility. Here, we describe for the Þrst time the structure, expression and genetic variation of grey partridge PRRs, in this case three TLRs. Mapping of these genes in novel species, such as the grey partridge, is of growing importance as this is the only means of undertaking comparative analysis (see Vinkler et al. 2014a, b), and of obtaining information on general patterns of pathogen-mediated PRR evolution and speciÞc adaptations in individual vertebrate lineages.

Both the genetic structure of the *PePeTLRs* and the structure of the predicted receptor proteins were similar to

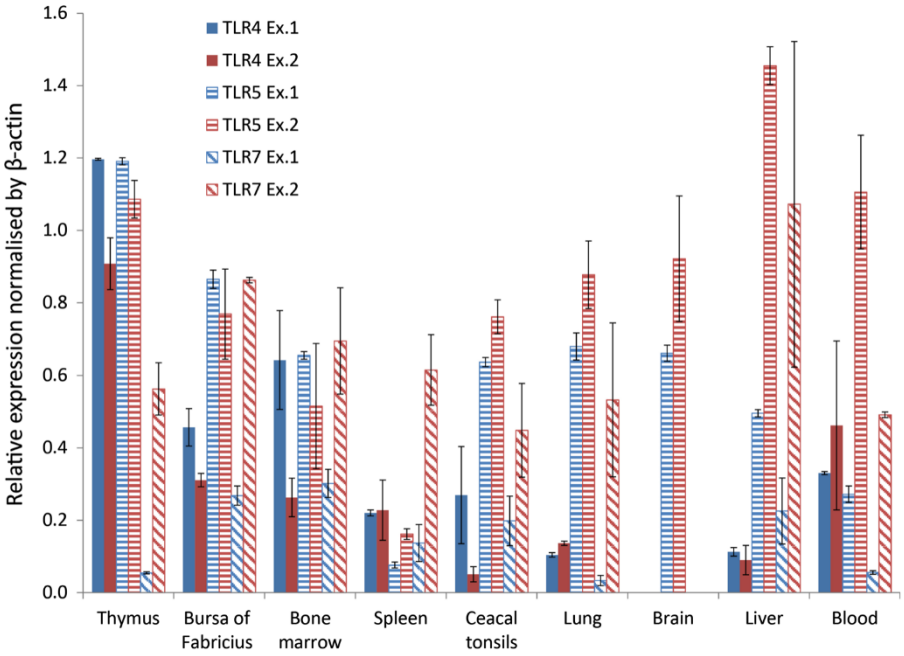


Fig. 3 (Color online) Tissue expression (Mean ± SD) of *TLR4* (*Þlled bars*), *TLR5* (*horizontally hatched bars*) and *TLR7* (*diagonally hatched bars*) in the grey partridge: semi-quantitative analysis performed on two individuals (Ex.1 and Ex.2; shown in *blue* and *red*, respectively [*black* and *grey*, respectively]). Expression data are normalised by tissue *β-actin* expression values

Table 2 *PePeTLR* genetic variation. Frequencies of individual putative SNPs (within 20 sequences obtained) and their effect on protein primary structure

Gene	Position of SNP	Sequence shift	SNP type	Domain	Frequency
PePeTLR4	50	CGG → CAG	aa replacement (R17Q)	Signal	3
PePeTLR4	54	G → A	Silent substitution	–	7
PePeTLR4	365	GGT → GAT	aa replacement (G122D)	LRR3	3
PePeTLR4	675	A → G	Silent substitution	–	1
PePeTLR4	986	GAG → GTG	aa replacement (E329V)	Non-LRR	1
PePeTLR4	1,119	C → T	Silent substitution	–	2
PePeTLR4	1,158	C → T	Silent substitution	–	2
PePeTLR4	1,698	T → C	Silent substitution	–	15
PePeTLR4	1,951	T → C	Silent substitution	–	2
PePeTLR4	2,174, 2,175	GAT → GTA	aa replacement (D725V)	TIR	6
PePeTLR5	−3 to −1	missing ATG	Initial methionine	Signal	15
PePeTLR5	14	CGG → CAG	aa replacement (R5Q)	Signal	15
PePeTLR5	1,583	TCT → TTT	aa replacement (S528F)	LRR20	3
PePeTLR5	2,175	AGA → AGT	aa replacement (R725S)	TIR	1
PePeTLR5	2,205	C → T	Silent substitution	–	2
PePeTLR5	2,358	G → A	Silent substitution	–	1
PePeTLR7	218, 219	ATC → ACT	aa replacement (I73T)	LRR1	5
PePeTLR7	279	A → G	Silent substitution	–	5
PePeTLR7	481, 482	AAA → TTA	aa replacement (K161L)	LRR4	18
PePeTLR7	2,008	GCT → TCT	aa replacement (A670S)	LRR21	18

those of the domestic chicken and other vertebrates (Temperley et al. 2008). With the exception of TLR5, PePeTLR length was identical to that of other Galliform species (Vinkler et al. 2014a, b). Further, the predicted ECD tertiary structures displayed a typical horseshoe shape in all three proteins, consistent with the 3D structure of other TLRs (Bell et al. 2003). The maximum likelihood phylogeny for all three *TLR* genes follows known species phylogenies within the Galloanserae lineage (Bao et al. 2010; Bonilla et al. 2010; Shen et al. 2010; Fig. 1), with grey partridge *TLRs* being closely related to those of pheasant and turkey. The only peculiarity noted was the

highly-supported clustering of turkey *TLR5* with junglefowl *TLR5* rather than that of pheasant and partridge. This may be the result of either distinct selection or trans-species polymorphism in turkey *TLR5*, or hidden errors in the published HQ436463.1 sequence (in their methods section, Gopinath et al. 2011 do not provide information on repeat sequencing to verify their sequencing results). We observed no sign of pseudogenisation in *TLR5* (also conÞrmed by expression analysis; see below); leading us to conclude that, unlike some passerine birds (Bainova et al. 2014), the *TLR5* gene is fully functional in all Galloanserae investigated to date.

All *TLR* genes were transcribed into mRNA for the majority of tissues investigated. Interestingly, we observed no *TLR4* and *TLR7* expression in grey partridge brain tissue. Rather than complete absence of expression, however, this may have been due to relatively minor levels of gene transcription, undetectable by the method utilised in this study. In contrast to our own results, Leveque et al. (2003) and Jia et al. (2012) observed some *TLR4* expression in chicken and duck brain, which would support this idea. Expression of *TLR4* mRNA was highest in the thymus, followed by bone marrow, bursa of Fabricius, blood, spleen and caecal tonsils, with lowest expression in lungs and liver. This is in agreement with the expression pattern reported by Vinkler et al. (2009) for *TLR4* in the zebra Þnch, and similar to that reported by Yilmaz et al. (2005) for the domestic chicken (though in that study the liver showed higher *TLR4* expression than the lungs). In contrast, Leveque et al. (2003) and Zhao et al. (2013) reported chicken and duck, respectively, as having relatively high expression of *TLR4* in the spleen and lungs (see Elfeil et al. 2012, Wang et al. 2011 and Jia et al. 2012 for comparisons to duck and goose studies that are in partial agreement with these two studies, and partial agreement with our study). Consistent with Gopinath et al. (2011) and Slawinska et al. (2013), who studied *TLR5* mRNA transcription in turkey and chicken, respectively, we detected relatively high expression of the *TLR5* gene in lungs and bursa of Fabricius. Unlike Gopinath et al. (2011), however, we also detected relatively high *TLR5* expression in the liver, and low expression in the spleen (similar to Slawinska et al. 2013). Furthermore, we detected a high *TLR5* expression in grey partridge thymus. Our results are in partial agreement with data recently reported for goose (Fang et al. 2012) and shelduck (*Tadorna tadorna*; Xiong et al. 2014), but contradict the expression pattern reported by Yilmaz et al. (2005) for the chicken, where no *TLR5* expression was recorded in blood, liver or spleen. Unlike any other study, including the present one, Ramasamy et al. (2012) recorded relatively high and constant *TLR4* and *TLR5* expression across all tissues investigated in the turkey. Constant high expression has also been reported for *TLR7* in both chicken (Yilmaz et al. 2005) and turkey (Ramasamy et al. 2012). In the grey partridge, however, *TLR7* transcription levels varied across both tissue type and individuals (see Fig. 3). Variation between species in the level of TLR expression is likely to be a key aspect inßuencing interspeciÞc immunological differences. IntraspeciÞc variation, however, may have an equally signiÞcant effect. Indeed, the potential importance of variation in intraspeciÞc TLR expression for disease resistance has already been documented in the domestic chicken (Leveque et al. 2003; Yilmaz et al. 2005; Slawinska et al. 2013). Unfortunately, given that both individuals in the present study were of the same age and housed similarly, our data cannot provide any evidence for the source of this conspeciÞc variation.

We detected 11 polymorphic positions (putative SNPs) in *PePeTLR4*, eight in *PePeTLR5* and six in *PePeTLR7*. In total, we identiÞed no more than Þve non-synonymous haplotypes in each gene, most of these variants being rare. The putative polymorphism we detected, however, appears to have only limited functional signiÞcance. The amino acid replacements R17Q in PePeTLR4 and R5Q in PePeTLR5, for example, do not appear to disrupt the signal peptide function as both substitutions can also be found at the interspeciÞc level within the Galliformes (possibly representing cases of trans-species polymorphism) and SignalP v. 4.0 analysis showed no difference in signal peptide site prediction. Nevertheless, the R17Q substitution in the TLR4 signal peptide and the D725V substitution in the TLR4 TIR domain (also known in other galliform species) produce important changes in residue chemical features and these may inßuence the behaviour of the resultant protein. Substitutions G122D and E329V also result in important alterations to residue chemical features (i.e. from non-polar to polar negative charges), though these are located at a wide topological distance from the TLR4 functional sites known in mammals (Park et al. 2009). Similarly, while substitution S528F at LRR20 in TLR5 may have a slight inßuence on protein structure, it is located outside the predicted ligand-binding region (Andersen-Nissen et al. 2007; Yoon et al. 2012). Most PePeTLR5 protein variants (aside from KC808007) were shown to lack the initial methionine, the CDS starting with methionine at position 2 of other avian species. Unlike in chicken and junglefowl (Downing et al. 2010), grey partridge *TLR7* displayed some coding variation. The I73T and L161K amino acid substitutions observed are unlikely to alter receptor function, however, as they are located in the N-terminal part of the protein that is cleaved off in endoplasmic reticulum before occurrence of ligand binding (Ewald et al. 2008). In our study, we were unable to verify cleavage of the TLR7 ECD in the endoplasmic reticulum experimentally; nevertheless, we believe that this can be assumed given the fact that the same type of cleavage takes

place (Ewald et al. 2008) in the related but distinct TLR9 protein (Roach et al. 2005). While substitution S670A does change the amino acid residuum features, once again it is located outside the predicted ligand-binding region (Wei et al. 2009). Compared to domestic chickens (Ruan et al. 2012a, b) and junglefowl (Downing et al. 2010), our results show lower levels of *PePeTLR* genetic sequence variation. In addition, the putative SNPs were located at different sites than those described in other species and distinct from the predicted functional sites. Overall, this leads us to assume that the genetic variation detected in the grey partridge has only limited importance.

The relatively low level of genetic variation we detected in grey partridge may indicate strong negative selection acting on the TLRs of this species. This view is consistent with the idea of conservativeness of TLR-ligand binding and low importance of naturally occurring variation in TLRs (Roach et al. 2005). A further possible explanation may be that the low genetic variation detected in this study is due to the relatively low sample size of the dataset examined. This is contradicted, however, by the fact that (1) the same alleles were detected repeatedly in the dataset, suggesting that there are few common *TLR* variants in the population, and (2) no putatively functional variation was detected. Finally, the low level of *TLR* population polymorphism observed may have resulted from local inbreeding and loss of genetic variation during a population bottleneck that impacted the study population. In the past seventy years, the European grey partridge has declined severely in numbers (see De Leo et al. 2004; Kuijper et al. 2009). Between the 1930s and 1990s, the population decreased from ca. 6 million individuals to less than 20 thousand in the Czech Republic (Šťastný et al. 2010), resulting in a highly fragmented population. The absence of functionally relevant SNPs may be an outcome of positive selection acting on *TLRs* prior to this bottleneck, resulting in the elimination of temporarily disadvantageous alleles during the bottleneck (for a meta-analysis documenting this phenomenon in MHC, see Sutton et al. 2011). We cannot completely exclude the possibility that the low number of individuals investigated prevented us detecting rare variants; nonetheless, it is our opinion that the bottleneck, together with positive selection, explains the general pattern of low TLR genetic variation observed in this study. This low genetic variability in immune-related genes may ultimately explain the relative susceptibility of the grey partridge population to infectious diseases (Sedlak et al. 2000; Vitula et al. 2011).

Conclusion

Our results indicate that grey partridge *TLR4*, *TLR5* and *TLR7* gene sequences are closely related to those of turkey and pheasant *TLRs*, which is consistent with the group phylogeny. All three *TLR* genes investigated appear to be fully functional, with expression pattern likely to vary between individuals. Within the population studied, genetic variation between the three genes was low and probably has only limited impact on receptor function. We found slightly higher variability in genes encoding bacterial-sensing TLRs (*TLR4* and *TLR5*) than in viral-sensing TLRs (*TLR7*). The generally low level of intraspeciÞc genetic variation in *PePeTLRs* compared to chickens and junglefowl may be a result of either negative selection or a population bottleneck, with further investigation required to fully explain this phenomenon.

Acknowledgments We would like to express our gratitude to Eva Holánová, Zdeňka Csibreiová, Miroslav Šálek, Jana Svobodová, Jitka Vinklerová, Dagmar Vinklerová and Kevin F. Roche for their help and advice. This study was supported by the Czech Science Foundation (Projects 206/08/1281, P505/10/1871 and P502/12/P179), the Ministry of Education, Youth and Sport of the Czech Republic within the NextGenProject (CZ.1.07/2.3.00/20.0303) and Institutional Research Support Grant SVV 260 087/2014. The authors contributed as follows: MV (25 %)—study design, main role in manuscript preparation; HB (25 %)—laboratory analysis, protein structure analysis; AB (25 %)—laboratory analysis; OT (10 %)—laboratory analysis; TA (5 %)—genetic sample collection; JB (10 %)—phylogenetic analysis. All authors contributed to the manuscript preparation.

References

Akira S, Uematsu S, Takeuchi O (2006) Pathogen recognition and innate immunity. Cell 124:783–801. doi:10.1016/j.cell.2006.02.015

Alcaide M, Edwards SV (2011) Molecular evolution of the Toll-like receptor multigene family in birds. Mol Biol Evol 28:1703–1715. doi:10.1093/molbev/msq351

Andersen LW, Kahlert J (2012) Genetic indications of translocated and stocked grey partridges (*Perdix perdix*): does the indigenous Danish grey partridge still exist? Biol J Lin Soc 105:694–710. doi:10.1111/j.1095-8312.2011.01833.x

Andersen-Nissen E, Smith KD, Bonneau R, Strong RK, Aderem A (2007) A conserved surface on Toll-like receptor 5 recognizes bacterial ßagellin. J Exp Med 204:393–403. doi:10.1084/jem.20061400

Bainova H, Kralova T, Bryjova A, Albrecht T, Bryja J et al (2014) First evidence of independent pseudogenization of Toll-like receptor 5 in passerine birds. Dev Comp Immunol 45:151–155. doi:10.1016/j.dci.2014.02.010

Bao XK, Liu NF, Qu JY, Wang XL, An B et al (2010) The phylogenetic position and speciation dynamics of the genus *Perdix* (Phasianidae, Galliformes). Mol Phylogenet Evol 56:840–847. doi:10.1016/j.ympev.2010.03.038

Beani L, Dessifulgheri F (1995) Mate choice in the grey partridge, *Perdix perdix*—role of physical and behavioral male traits. Anim Behav 49:347–356. doi:10.1006/anbe.1995.0047

Bell JK, Mullen GED, Leifer CA, Mazzoni A, Davies DR et al (2003) Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends Immunol 24:528–533. doi:10.1016/s1471-4906(03)00242-4

Bonilla AJ, Braun EL, Kimball RT (2010) Comparative molecular evolution and phylogenetic utility of 3′-UTRs and introns in Galliformes. Mol Phylogenet Evol 56:536–542. doi:10.1016/j.ympev.2010.04.006

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

Brownlie R, Zhu JZ, Allan B, Mutwiri GK, Babiuk LA et al (2009) Chicken TLR21 acts as a functional homologue to mammalian TLR9 in the recognition of CpG oligodeoxynucleotides. Mol Immunol 46:3163–3170. doi:10.1016/j.molimm.2009.06.002

Buner FD, Browne SJ, Aebischer NJ (2011) Experimental assessment of release methods for the re-establishment of a red-listed galliform, the grey partridge (*Perdix perdix*). Biol Conserv 144:593–601. doi:10.1016/j.biocon.2010.10.017

Cormican P, Lloyd AT, Downing T, Connell SJ, Bradley D et al (2009) The avian Toll-like receptor pathway—subtle differences amidst general conformity. Dev Comp Immunol 33:967–973. doi:10.1016/j.dci.2009.04.001

Cserzo M, Eisenhaber F, Eisenhaber B, Simon I (2004) TM or not TM: transmembrane protein prediction with low false positive rate using DAS-TMPlter. Bioinformatics 20:136–137. doi:10.1093/bioinformatics/btg394

Cucco M, Malacarne G, Ottonelli R, Patrone M (2006) Repeatability of cell-mediated and innate immunity, and other Þtness-related traits, in the Grey Partridge. Can J Zool Rev Can Zool 84:72–79

Cucco M, Guasco B, Malacarne G, Ottonelli R (2007) Effects of beta-carotene on adult immune condition and antibacterial activity in the eggs of the Grey Partridge, *Perdix perdix*. Comp Biochem Physiol A Mol Integr Physiol 147:1038–1046. doi:10.1016/j.cbpa.2007.03.014

Dalloul RA, Long JA, Zimin AV, Aslam L, Beal K et al (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. PLoS Biol. doi:10.1371/journal.pbio.1000475

De Leo GA, Focardi S, Gatto M, Cattadori IM (2004) The decline of the grey partridge in Europe: comparing demographies in traditional and modern agricultural landscapes. Ecol Model 177:313–335. doi:10.1016/j.ecolmodel.2003.11.017

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. doi:10.4049/jimmunol.0903092

Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. doi:10.1186/1471-2148-7-214

Elfeil WK, Abouelmaatti RR, Sun CJ, Han WY, Li XK et al (2012) IdentiÞcation, cloning, expression of a novel functional *Anas platyrhynchos* mRNA TLR4. J Anim Vet Adv 11:1727–1733. doi:10.3923/javaa.2012.1727.1733

Ewald SE, Lee BL, Lau L, Wickliffe KE, Shi G-P et al (2008) The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. Nature 456:658–662. doi:10.1038/nature07405

Fang Q, Pan ZM, Geng SZ, Kang XL, Huang JL et al (2012) Molecular cloning, characterization and expression of goose Toll-like receptor 5. Mol Immunol 52:117–124. doi:10.1016/j.molimm.2012.05.005

Fukui A, Inoue N, Matsumoto M, Nomura M, Yamada K et al (2001) Molecular cloning and functional characterization of chicken Toll-like receptors—a single chicken toll covers multiple molecular patterns. J Biol Chem 276:47143–47149

Gopinath VP, Biswas M, Raj GD, Raja A, Kumanan K et al (2011) Molecular cloning and tissue-speciÞc expression of Toll-like receptor 5 gene from turkeys. Avian Dis 55:480–485

Grueber CE, Wallis GP, King TM, Jamieson IG (2012) Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin. PLoS ONE. doi:10.1371/journal.pone.0045011

Grueber CE, Wallis GP, Jamieson IG (2014) Episodic positive selection in the evolution of avian Toll-like receptor innate immunity genes. PLoS ONE 9:e89632. doi:10.1371/journal.pone.0089632

Guindon S, Lethiec F, Duroux P, Gascuel O (2005) PHYML Online—a web server for fast maximum likelihood-based phylogenetic inference. Nucleic Acids Res 33:W557–W559. doi:10.1093/nar/gki352

Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL et al (2008) A phylogenomic study of birds reveals their evolutionary history. Science 320:1763–1768. doi:10.1126/science.1157704

Horigan V, Davies RH, Kelly LA, Mead GC, Irvine RM et al (2014) A qualitative risk assessment of the microbiological risks to consumers from the production and consumption of uneviscerated and eviscerated small game birds in the UK. Food Control 45:127–137. doi:10.1016/j.foodcont.2014.04.040

Huang YH, Temperley ND, Ren LM, Smith J, Li N et al (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. doi:10.1186/1471-2148-11-149

Iqbal M, Philbin VJ, Smith AL (2005) Expression patterns of chicken Toll-like receptor mRNA in tissues, immune cell subsets and cell lines. Vet Immunol Immunopathol 104:117–127. doi:10.1016/j.vetimm.2004.11.003

Iwasaki A, Medzhitov R (2010) Regulation of adaptive immunity by the innate immune system. Science 327:291–295. doi:10.1126/science.1183021

Jetz W, Thomas GH, Joy JB, Hartmann K, Mooers AO (2012) The global diversity of birds in space and time. Nature 491:444–448. doi:10.1038/nature11631

Jia H, Li G, Li J, Tian Y, Wang D et al (2012) Cloning, expression and bioinformatics analysis of the duck TLR 4 gene. Br Poult Sci 53:190–197. doi:10.1080/00071668.2012.674208

Jiang Q, Turner T, Sosa MX, Rakha A, Arnold S et al (2012) Rapid and efÞcient human mutation detection using a bench-top next-generation DNA sequencer. Hum Mutat 33:281–289. doi:10.1002/humu.21602

Keestra AM, van Putten JPM (2008) Unique properties of the chicken TLR4/MD-2 complex: selective lipopolysaccharide activation of the MyD88-dependent pathway. J Immunol 181:4354–4362

Keestra AM, de Zoete MR, van Aubel RAMH, van Putten JPM (2008) Functional characterization of chicken TLR5 reveals species-speciÞc recognition of ßagellin. Mol Immunol 45:1298–1307. doi:10.1016/j.molimm.2007.09.013

Keestra AM, de Zoete MR, Bouwman LI, Vaezirad MM, van Putten JPM (2013) Unique features of chicken toll-like receptors. Dev Comp Immunol 41:316–323. doi:10.1016/j.dci.2013.04.009

Kim HM, Park BS, Kim JI, Kim SE, Lee J et al (2007) Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist eritoran. Cell 130:906–917. doi:10.1016/j.cell.2007.08.002

Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) Automated phylogenetic detection of recombination using a genetic algorithm. Mol Biol Evol 23:1891–1901. doi:10.1093/molbev/msl051

Kuijper DPJ, Oosterveld E, Wymenga E (2009) Decline and potential recovery of the European grey partridge (*Perdix perdix*) population—a review. Eur J Wildl Res 55:455–463. doi:10.1007/s10344-009-0311-2

Letunic I, Doerks T, Bork P (2012) SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res 40:D302–D305. doi:10.1093/nar/gkr931

Leveque G, Forgetta V, Morroll S, Smith AL, Bumstead N et al (2003) Allelic variation in TLR4 is linked to susceptibility to *Salmonella enterica* serovar *typhimurium* infection in chickens. Infect Immun 71:1116–1124. doi:10.1128/IAI.71.3.1116-1124.2003

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. doi:10.1093/bioinformatics/btp187

Liu Y, Chang GB, Hu GS, Li Q, Xu Q et al (2011) Genetic variation at Exon2 of TLR4 gene and its association with resistant traits in chicken. Afr J Biotechnol 10:8260–8266

Lynn DJ, Lloyd AT, O'Farrelly C (2003) In silico identiÞcation of components of the Toll-like receptor (TLR) signaling pathway in clustered chicken expressed sequence tags (ESTs). Vet Immunol Immunopathol 93:177–184. doi:10.1016/S0165-2427(03)00058-8

MacDonald MRW, Xia JG, Smith AL, Magor KE (2008) The duck Toll like receptor 7: genomic organization, expression and function. Mol Immunol 45:2055–2061. doi:10.1016/j.molimm.2007.10.018

McGufÞn LJ (2008) The ModFOLD server for the quality assessment of protein structural models. Bioinformatics 24:586–587. doi:10.1093/bioinformatics/btn014

Offord V, Coffey TJ, Werling D (2010) LRRÞnder: a web application for the identiÞcation of leucine-rich repeats and an integrative Toll-like receptor database. Dev Comp Immunol 34:1035–1041. doi:10.1016/j.dci.2010.05.004

Ohto U, Fukase K, Miyake K, Shimizu T (2012) Structural basis of species-speciÞc endotoxin sensing by innate immune receptor TLR4/MD-2. Proc Natl Acad Sci USA 109:7421–7426. doi:10.1073/pnas.1201193109

Palsson-McDermott EM, O'Neill LAJ (2007) Building an immune system from nine domains. Biochem Soc Trans 35:1437–1444. doi:10.1042/BST0351437

Park BS, Song DH, Kim HM, Choi BS, Lee H et al (2009) The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191–1195. doi:10.1038/nature07830

Petersen TN, Brunak S, von Heijne G, Nielsen H (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods 8:785–786. doi:10.1038/nmeth.1701

Philbin VJ, Iqbal M, Boyd Y, Goodchild MJ, Beal RK et al (2005) IdentiÞcation and characterization of a functional, alternatively spliced Toll-like receptor 7 (TLR7) and genomic disruption of TLR8 in chickens. Immunology 114:507–521. doi:10.1111/j.1365-2567.2005.02125.x

Ramasamy KT, Verma P, Reddy MR, Murugesan S (2011) Molecular characterization of coding sequence and mRNA expression pattern of Toll-like receptor 15 in Japanese quail (*Coturnix japonica*) and indigenous chicken breeds (Aseel and Kadaknath). J Poult Sci 48:168–175. doi:10.2141/jpsa.011008

Ramasamy KT, Reddy MR, Verma PC, Murugesan S (2012) Expression analysis of turkey (*Meleagris gallopavo*) toll-like receptors and molecular characterization of avian speciÞc TLR15. Mol Biol Rep 39:8539–8549. doi:10.1007/s11033-012-1709-6

Resman N, Vasl J, Oblak A, Pristovsek P, Gioannini TL et al (2009) Essential roles of hydrophobic residues in both MD-2 and Toll-like receptor 4 in activation by endotoxin. J Biol Chem 284:15052–15060. doi:10.1074/jbc.M901429200

Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK et al (2005) The evolution of vertebrate toll-like receptors. Proc Natl Acad Sci USA 102:9577–9582. doi:10.1073/pnas.0502272102

Roy A, Kucukural A, Zhang Y (2010) I-TASSER: a uniÞed platform for automated protein structure and function prediction. Nat Protoc 5:725–738. doi:10.1038/nprot.2010.5

Ruan WK, Wu YH, An J, Zheng SJ (2012a) Polymorphisms of chicken toll-like receptors 4, 15, and 21 in different breeds. Poult Sci 91:2512–2516. doi:10.3382/ps.2012-02319

Ruan WK, Wu YH, Zheng SJJ (2012b) Different genetic patterns in avian Toll-like receptor (TLR)5 genes. Mol Biol Rep 39:3419–3426. doi:10.1007/s11033-011-1113-7

Sedlak K, Literak I, Vitula F, Benaak J (2000) High susceptibility of partridges (*Perdix perdix*) to toxoplasmosis compared with other gallinaceous birds. Avian Pathol 29:563–569

Shen YY, Liang L, Sun YB, Yue BS, Yang XJ et al (2010) A mitogenomic perspective on the ancient, rapid radiation in the Galliformes with an emphasis on the Phasianidae. BMC Evol Biol 10:132. doi:10.1186/1471-2148-10-132

Slawinska A, D'Andrea M, Pilla F, Bednarczyk M, Siwek M (2013) Expression proÞles of toll-like receptors 1, 2 and 5 in selected organs of commercial and indigenous chickens. J Appl Genetics 54:489–492. doi:10.1007/s13353-013-0161-1

Smith J, Speed D, Law AS, Glass EJ, Burt DW (2004) In-silico identiÞcation of chicken immune-related genes. Immunogenetics 56:122–133. doi:10.1007/s00251-004-0669-y

Šťastný K, Bejček V, Hudec K (2010) Atlas hnízdního rozšíření ptáků v České republice. Aventinum, Praha

Sutton JT, Nakagawa S, Robertson BC, Jamieson IG (2011) Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. Mol Ecol 20:4408–4420. doi:10.1111/j.1365-294X.2011.05292.x

Svobodova J, Gabrielova B, Synek P, Marsik P, Vanek T et al (2013) The health signalling of ornamental traits in the Grey Partridge (*Perdix perdix*). J Ornithol 154:717–725. doi:10.1007/s10336-013-0936-5

Temperley ND, Berlin S, Paton IR, GriÞn DK, Burt DW (2008) Evolution of the chicken Toll-like receptor gene family: a story of gene gain and gene loss. BMC Genom 9:62. doi:10.1186/1471-2164-9-62

Tompkins DM, Draycott RAH, Hudson PJ (2000) Field evidence for apparent competition mediated via the shared parasites of two gamebird species. Ecol Lett 3:10–14. doi:10.1046/j.1461-0248.2000.00117.x

Tompkins DM, Parish DMB, Hudson PJ (2002) Parasite-mediated competition among red-legged partridges and other lowland gamebirds. J Wildl Manage 66:445–450. doi:10.2307/3803177

Uno Y, Usui T, Fujimoto Y, Ito T, Yamaguchi T (2012) QuantiÞcation of interferon, interleukin, and Toll-like receptor 7 mRNA in quail splenocytes using real-time PCR. Poult Sci 91:2496–2501. doi:10.3382/ps.2012-02283

Vidus-Rosin A, Meriggi A, Pella F, Zaccaroni M (2010) Demographic parameters of reintroduced grey partridges in central Italy and the effect of weather. Eur J Wildl Res 56:369–375. doi:10.1007/s10344-009-0329-5

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

Vinkler M, Bryjova A, Albrecht T, Bryja J (2009) IdentiÞcation of the Þrst Toll-like receptor gene in passerine birds: TLR4 orthologue in zebra Þnch (*Taeniopygia guttata*). Tissue Antigens 74:32–41. doi:10.1111/j.1399-0039.2009.01273.x

Vinkler M, Bainova H, Bryja J (2014a) Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. Genet Sel Evol 46:72. doi:10.1186/s12711-014-0072-6

Vinkler M, Svobodova J, Gabrielova B, Bainova H, Bryjova A (2014b) Cytokine expression in phytohaemagglutinin-induced skin inßammation in a galliform bird. J Avian Biol 45:43–50. doi:10.1111/j.1600-048X.2011.05860.x

Vitula F, Peckova L, Bandouchova H, Pohanka M, Novotny L et al (2011) *Mycoplasma gallisepticum* infection in the grey partridge *Perdix perdix*: outbreak description, histopathology, biochemistry and antioxidant parameters. BMC Vet Res. doi:10.1186/1746-6148-7-34

Walsh C, Gangloff M, Monie T, Smyth T, Wei B et al (2008) Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa. J Immunol 181:1245–1254

Wang F, Lu LZ, Yuan H, Tian Y, Li JJ et al (2011) cDNA cloning, characterization and expression analysis of Toll-like receptor 4 gene in goose. Can J Anim Sci 91:371–377. doi:10.4141/cjas2011-002

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW et al (2010) The genome of a songbird. Nature 464:757–762. doi:10.1038/nature08819

Wei TD, Gong J, Jamitzky F, Heckl WM, Stark RW et al (2009) Homology modeling of human toll-like receptors TLR7, 8, and 9 ligand-binding domains. Protein Sci 18:1684–1691. doi:10.1002/pro.186

Wlasiuk G, Khan S, Switzer WM, Nachman MW (2009) A history of recurrent positive selection at the Toll-like receptor 5 in primates. Mol Biol Evol 26:937–949. doi:10.1093/molbev/msp018

Xiong D, Pan ZM, Kang XL, Wang J, Song L et al (2014) Molecular cloning and functional analysis of duck toll-like receptor 5. Res Vet Sci 97:43–45. doi:10.1016/j.rvsc.2014.05.010

Yilmaz A, Shen SX, Adelson DL, Xavier S, Zhu JJ (2005) IdentiPcation and sequence analysis of chicken toll-like receptors. Immunogenetics 56:743–753. doi:10.1007/s00251-004-0740-8

Yoon SI, Kurnasov O, Natarajan V, Hong MS, Gudkov AV et al (2012) Structural basis of TLR5-ßagellin recognition and signaling. Science 335:859–864. doi:10.1126/science.1215584

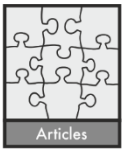
Zhang Y (2008) I-TASSER server for protein 3D structure prediction. BMC Bioinformatics. doi:10.1186/1471-2105-9-40

Zhao WM, Huang ZY, Chen Y, Zhang Y, Rong GH et al (2013) Molecular cloning and functional analysis of the duck TLR4 gene. Int J Mol Sci 14:18615–18628. doi:10.3390/ijms140918615

PAPER VI.

Vinkler M., Svobodová J., Gabrielová B., **Bainová H.** & Bryjová A. (2014): Cytokine expression in phytohaemagglutinin-induced skin inflammation in a galliform bird. Journal of Avian Biology 45(1):43–50. (IF_{2014}= 1.971)

In this study, we aimed to improve the understanding of the molecular mechanism triggering the response to frequently used immunological treatment - phytohaemagglutinin (PHA). Understanding of this mechanism is crucial to uncover the nature of ecological costs and benefits of investments into the immune response and thus to correctly test various ecological hypotheses. We examined the relationship between the magnitude of tissue swelling and cytokine expression (signalling molecules typically triggered by TLR activation). Our crucial finding was, that PHA skin-swelling test actually measures inflammatory processes (innate immunity), instead of T-cell proliferation (adaptive immunity) as was generally assumed previously. On the base of these findings, PHA-induced skin-swelling test seems to be a useful tool for ecoimmunologist to investigate the trade-offs associated with regulation of inflammatory responses linked to innate immune receptor variability in birds.



Journal of Avian Biology 45: 43–50, 2014
doi: 10.1111/j.1600-048X.2013.05860.x

Subject Editor: Ruedi Nager. Accepted 12 July 2013

Cytokine expression in phytohaemagglutinin-induced skin inflammation in a galliform bird

Michal Vinkler, Jana Svobodová, Barbora Gabrielová, Hana Bainová and Anna Bryjová

M. Vinkler (vinkler1@natur.cuni.cz), B. Gabrielová and H. Bainová, Dept of Zoology, Faculty of Science Charles Univ. in Prague, Viničná 7, CZ-128 44 Prague, Czech Republic. – J. Svobodová, Dept of Ecology, Faculty of Environmental Sciences, Czech Univ. of Life Sciences, Kamýcká 1176, CZ-165 21 Prague, Czech Republic. – A. Bryjová, Inst. of Vertebrate Biology, Academy of Sciences of the Czech Republic, v.v.i., Květná 8, CZ-603 65 Brno, Czech Republic.

Researchers interested in ecological immunology face substantial methodological problems: 1) most immunological approaches are difficult to perform in free-living animals, 2) in some of the applicable methods the immunological background of the test remains unclear. The latter is also true for the phytohaemagglutinin (PHA) skin-swelling test, a trait of cell-mediated immunity commonly measured in ecology. A lack of direct evidence documenting the immunological processes in the tissue limits our understanding of the mechanism triggering the response to PHA. Understanding of this mechanism is, nonetheless, crucial for us to uncover the nature of ecological costs and benefits of investments into the response. As knowledge of cytokine signalling in the tissue may clarify the response mechanism, in our study we investigated the association between the PHA-induced skin-swelling and tissue cytokine expression in males of grey partridge *Perdix perdix*. In PHA-challenged birds we assessed expression of nine cytokines (IL-1β, IL-2, IL-4, IL-6, IL-10, IL-12, IL-17, TGF-β, IFN-γ) in wing-web skin during an early stage of the immune response. We examined the relationship between the magnitude of tissue swelling and cytokine expression. Contrary to some earlier expectations we did not find any differential expression of T-cell growth factor, IL-2, in the tissue. Hence, T-cell proliferation at the time of the swelling measurement is unlikely. We detected differential expression in Th17 pro-inflammatory (IL-1β, IL-6) and anti-inflammatory (TGF-β) cytokines. The PHA-induced swelling response was only weakly linked to the expression of TGF-β. We also found relationships between the PHA-induced swelling response and phenotypic traits of the birds; the PHA swelling was positively associated with the extent of melanin-based breast ornamentation and negatively related to body size. Our results might suggest that variation in swelling is influenced by total numbers of responding cells rather than by differences in signalling. Moreover, we revealed significant correlations in expression of IL-1β, IL-6 and TGF-β. These findings are the first to show on the molecular level that the PHA skin-swelling test actually measures inflammation process which is part of innate immune defence and not the adaptive immune response (as assumed if the test was the reflection of T-cell proliferation).

Avian ecologists, ecological toxicologists as well as researchers in avian veterinary science often face a challenge of measuring the immune response in non-laboratory birds (Millet et al. 2007, Martin et al. 2011). Despite the widespread interest in studying the immune function in free-living birds, the methods currently used in avian ecological immunology are still coarse and interpretation of some of the tests remains unclear. As a result, pleading calls for elaboration of more detailed understanding of the methods and close cooperation between ecologists and immunologists are frequent in present literature (Martin et al. 2011, Pedersen and Babayan 2011). It has been repeatedly pointed out that there is no single method of measuring immunity in animals (Adamo 2004 and citations therein) because there is no single mechanism of immunity (or 'immunocompetence' as often termed in ecology; Vinkler and Albrecht 2011). There are several distinct immunological pathways which vary in terms of costs and benefits of investments (Blount et al. 2003a). The evolution of immunity is strictly dependent on the effectiveness of particular pathways in defence against certain parasites occurring in areas inhabited by individual populations (Schmid-Hempel 2011). We believe that limited understanding of the underlying immunological mechanisms may be partially responsible for frequently reported incongruence of results in ecological immunology (compare, e.g. conclusions of Møller 1998 and Martin et al. 2004 regarding immunity-investments of tropical birds). Therefore, avian ecologists interested in description of phenomena related to the immune function need to know (as precisely as possible) which types of the immune response they study, what cells are involved, what mediators of cellular signalling are expressed and what effector mechanisms are triggered.

One of the most common ecological tests to measure cell-mediated immunity in birds has been the phytohaemagglutinin (PHA) skin-swelling test (Smits et al. 1999, Kennedy and Nager 2006). This method is fast, efficient and simple, allowing collection of valuable data in a brief time period directly in the field where non-standard conditions preclude adoption of other, more sophisticated and rigorous immunological approaches. However, despite common usage of the PHA skin-swelling test in avian ecological studies, not much is currently known about the precise immunological mechanism triggering the swelling response induced by PHA (Kennedy and Nager 2006, Vinkler et al. 2010), and the significance of the test has been debated in the current literature (Owen and Clayton 2007, Vinkler and Albrecht 2011).

PHA is a plant lectin with specific immunomodulatory features produced by red kidney beans *Phaseolus vulgaris* (Rigas and Johnson 1964). Given the known mitogenic effect of the L subunit of PHA on lymphocytes (Leavitt et al. 1977) the PHA skin-swelling test was originally thought to evaluate T-cell-mediated immunocompetence and many important ecological and evolutionary issues (such as, e.g. signalling value of colourful ornaments or heritability of immune defence) have been discussed in this manner (Saino et al. 2003, Cucco et al. 2006, Bonato et al. 2009). Adaptive immunity, to which T cells belong, represents a complex of efficient immunological mechanisms that are precisely targeted to the structures of a particular parasite (Danilova 2006). This allows the immune system to limit self-damage on one hand (benefit from investment into adaptive immunity), but limits the generality of the response (narrow specificity represents a cost, e.g. in cases of epitope switching by parasites). Adaptive immunity is also rather slow in its action, taking approximately one week to develop after primary contact with the parasite. On the contrary, innate immunity comprises a set of diverse defence mechanisms that have speed and generally limited specificity in common. These benefits from investment into innate immunity are traded-off against lower efficiency of innate immunity to specialised parasites and wider self-destruction of the host in the course of defence. Which immune pathway is measured has therefore distinct implications for the ecological and evolutionary interpretation of results.

The interpretation of the PHA skin swelling test as a measurement of T-cell-mediated immunocompetence (i.e. measurement of adaptive immunity) has several weak points. Firstly, the term 'immunocompetence' itself is highly questionable when describing in vivo responsiveness to PHA (Vinkler and Albrecht 2011). Secondly, the specification 'T-cell-mediated' does not entirely fit the evidence currently available. A number of histological analyses have shown that lymphocytes as well as various granulocytes and monocytes are present in the dermal tissue shortly after PHA application (Goto et al. 1978, McCorkle et al. 1980, Martin et al. 2006, Vinkler et al. 2010, 2012) suggesting a complex and non-specific inflammatory response (Campbell and Ellis 2007). Moreover, the 6–24 h time period between the PHA application and the swelling measurement was found too short to allow any T-cell mitosis in mammals (Tokuyasu et al. 1968). The histology tells us only that the response is complex, combining innate and adaptive cell types (Kennedy and Nager 2006). This suggests that the test cannot show the capacity of T-cell proliferation. The intensity of the PHA-induced inflammatory immune response might be dependent on efficiency of the initial T-cell activation at the treatment site (Tella et al. 2008). The results published by Goto et al. (1978), however, show that some individuals responded to PHA even after thymectomy. This would not be possible if the response was based only on T-cell activation alone (see Licastro et al. 1993 for the way of PHA-induced T-cell activation). We are also currently not aware if T-cell growth factor, IL-2 (a cytokine essential for T-cell proliferation which is produced mainly by activated T cells; Lin and Leonard 2003), is up-regulated in the PHA-treated tissues. Hence, there is no unambiguous evidence showing that the PHA skin-swelling test actually measures the T-cell function. On the contrary, experiments with purified PHA subunits demonstrated that the erythroagglutinating subunit of PHA (PHA-E), which lacks any ability to activate T cells, triggers a stronger swelling response than the T-cell-binding subunit (PHA-L; Vinkler et al. 2010). Unfortunately, all these results only partly answer the question of interpretation of the PHA skin-swelling test in animal ecology. We are presently aware that the response to PHA is not fully dependent on T cells, but we still do not know the activation mechanism of the leukocyte infiltration into the treatment site that causes the swelling we measure. Does higher magnitude of the response reveal (at least partially) any investment into slow but efficient adaptive immunity (as proposed by Tella et al. 2008) or is it a measurement of actual activation of fast and broad-targeted innate immunity which is non-specific and self-destructive at the same time (Ashley et al. 2012)? What is the actual involvement of T cells in the course of the response? Not knowing the answers to these questions we cannot fully judge the value of the PHA skin-swelling test in solving ecological and evolutionary issues. To obtain the answers, we need to explore the direct immunological features of the response about which virtually nothing is currently known.

Most importantly, we are lacking any information on the immunological regulation of the inflammatory processes induced by PHA in the skin. Inflammation is one of the most essential immunological mechanisms responsible for pathogen clearance (van de Veerdonk et al. 2011). Therefore, investigation of factors governing inflammatory responsiveness to foreign antigens is an important task of the current immunoecological research. The initial phase of inflammation (the first few hours after recognition of infection, tissue damage or, in this case, administration of PHA) involves immunological signalling stimulating leukocyte infiltration into the injected site (Ashley et al. 2012). As inflammation is to some extent a self-damaging immunological process (Graham et al. 2005, Ashley et al. 2012), it may be potentially harmful to the organism's integrity and, hence, it requires careful regulation. This regulation is ensured by highly coordinated cytokine signalling. Investigation of cytokine signalling may, thus, inform us directly about the immunological processes in the tissue. These processes may be either pro-inflammatory (Th1- or Th17-directed pathway) or anti-inflammatory (Th2- or Treg-directed pathway). Pro-inflammatory cytokines such as interleukin-1β (IL-1β), IL-6, IL-12, IL-17 and interferon-γ (IFN-γ) are involved in boosting the inflammatory response (Thomson and Lotze 2003,

Kaiser and Staheli 2008). In order to avoid severe damage to the host and immunopathology, the immune system also produces negative regulators of the inflammation (anti-inflammatory cytokines) which include IL-4, IL-10 and Transforming Growth Factor-β (TGF-β). In later stages of the inflammatory immune response (ca 24 h after stimulation) the pro-inflammatory cytokine signalling is progressively down-regulated, which is then followed by slow reduction of cellular inflammatory activity in the tissue and initiation of wound healing (Ashley et al. 2012). Compared to mammalian immunity, avian immunity is (despite recent intensive effort) still poorly understood (Kaiser 2010). This is particularly true for many aspects of inflammation. In birds, we are lacking any information on skin-based cytokine expression during cutaneous inflammation and the immunobiology of avian dermal inflammation has never been rigorously explored. Full understanding of the informative value of the PHA skin-swelling test in avian studies, therefore, requires additional knowledge of the regulatory processes governing this inflammatory response in birds.

To gain insight into this problem, we investigated the association between the PHA-induced skin-swelling response and tissue cytokine expression in grey partridge *Perdix perdix*. In a sample of 25 captive males we assessed the skin expression profile of nine cytokines, IL-1β, IL-2, IL-4, IL-6, IL-10, IL-12 (by assessing expression of the IL-12α subunit, p35, from IL-12A gene), IL-17, TGF-β and IFN-γ, during the final stage of the early phase of the response, i.e. six hours after subcutaneous PHA treatment. At this time we may expect to detect expression of those cytokines that are involved in differentiation of immunological pathways and later (e.g. at the time of 24 h after the PHA treatment) direct the cellular activity in the tissue (infiltration, activation and suppression of particular cell types involved; MacHugh et al. 2012). Among pro-inflammatory cytokines IL-12 and IFN-γ are involved in Th1 immunity, while IL-1β, IL-6 and IL-17 are cytokines involved in Th17 pathway (Kaiser 2010, van de Veerdonk et al. 2011). As anti-inflammatory and regulatory cytokines were chosen IL-4 (representing Th2 pathway), IL-10 and TGF-β (the latter two expressed by Tregs; Kaiser 2010). By measuring IL-2 expression we tested if any increased T-cell proliferation may take place in the tissue. In subsequent analyses, we evaluated the impact of cytokine expression on the magnitude of the skin-swelling hypersensitivity reaction, simultaneously assessing the relationship to other phenotypic traits that were previously shown to be linked with the magnitude of the PHA-induced swelling response: body size (Tella et al. 2002), body mass (Møller and Petrie 2002) and ornamentation (Blount et al. 2003b, Jacquin et al. 2011). Sexual ornaments are assumed to serve as indicators of individual quality (Andersson 1994). The association between the PHA-induced swelling and ornamentation may, therefore, help us with the test's interpretation as it might suggest whether the test reveals investment into the immune response (highly ornamented healthier animals mount stronger responses, as reported e.g. by Jacquin et al. 2011) or rather the state of current immunity activation (poorly ornamented low quality individuals mount stronger responses, as recently suggested by Vinkler et al. 2012). Finally, we also explored correlations in expression among individual cytokines in which we detected high fold increase in levels of expression after the PHA treatment.

Methods

Animals and general experiment protocol

Twenty five grey partridge males in their first year of age were obtained from the breeding facility of the Univ. of Veterinary and Pharmaceutical Sciences Brno, Faculty of Veterinary Hygiene and Ecology in Jinačovice, Czech Republic and housed individually in separate cages (89 × 189 cm and 40 cm high). Each bird was marked by a metal ring with an individual number. All birds had permanent access to food (wheat) and water ad libitum. On 18 April 2009 (after a two-week acclimation period in which birds from outdoor aviaries were accustomed to separate indoor caging) all individuals were weighed and their tarsus length was measured (serving as an estimate of body size). Then, the PHA skin-swelling test was performed according to Smits et al. (1999). The thickness of the centre of left wing patagium (wing web) in each individual was measured three times with a pressure-sensitive digital thickness gauge with accuracy to 0.01 mm (Mitutoyo 547-313, Mitutoyo Corporation, Kanagawa, Japan). Following this measurement a dosage of 1 mg of PHA-P dissolved in 40 µl DPBS (product numbers L8754 and D5652, respectively; Sigma-Aldrich, St Louis, MO, USA) was injected subcutaneously into the wing web. PHA-P is in vivo a potent non-specific stimulator of inflammation (Vinkler et al. 2010). All birds were thereafter returned into their cages and left in semidarkness and undisturbed by any further manipulation. In all individuals the treatment was performed in the afternoon hours (15:45–17:30) to minimise temporal variation in our sample. The magnitude of the swelling response was measured after 6 ± 0.5 h three times with accuracy 0.01 mm (all thickness measurements were performed by the same person, MV, with a repeatability $r = 0.89$). This time scale allowed us to investigate the final stage of the early phase of the potential PHA-induced inflammation, i.e. the period when the type and the further direction of the immune response is being determined and particular signalling pathways are triggered. Six hours was also the period used for PHA skin-swelling test in several avian studies (for the usage of a 6 h period see e.g. Møller et al. 2003, Bonato et al. 2009, Vinkler et al. 2012). The PHA-induced skin-swelling response index was later calculated as the average tissue thickness 6 h after the treatment minus the average thickness before the PHA injection. Immediately after the second patagium-thickness measurement we collected a biopsy sample of the inflamed tissue by a sterile tissue corer (3 mm in diameter) and stored it in RNAlater (Qiagen, Hilden, Germany) at −80°C until total RNA extraction. As negative controls, biopsies out of untreated patagia (the opposite wing than to which the PHA-treatment was applied) were collected in ten individuals. Although the local PHA treatment might influence the lymphatic system and blood stream cells (e.g. similarly as found by Buchanan et al. 2003), according to the immunological evidence available it is unlikely that this should affect cellular activation in any other peripheral region which remains unstimulated.

Untreated patagia, therefore, represent suitable controls allowing pairwise comparison.

Finally, a standard digital image of the melanin-based breast plumage ornamentation (ventral part of the bird) of each male was taken with metric standard in a dark tent by scanner Epson Perfection V10 (Seiko Epson, Nagano, Japan). These images were later used to measure the ornament area using Adobe Photoshop CS3 (Adobe Systems, San Jose, CA). The scale was defined according to the ruler on the scanned images. To delimit the ornament area within its broad range precisely, we used an automatic selection tool with 20% colour tolerance. The area of melanin-based ornamentation was considered as a possible condition-dependent health indicator (Beani and Dessifulgheri 1995, Svobodová et al. 2013). The research was approved by the Ethical Committee of the Inst. of Vertebrate Biology, Academy of Sciences of the Czech Republic, v.v.i. (reference no. 147/2007) and was carried out in accordance with the current laws of the Czech Republic and the EU.

Molecular analysis

Total RNA was isolated from biopsy samples (tissue samples after the PHA-treatment as well as control samples) using High Pure RNA Tissue Kit (Roche Applied Science, Basel, Switzerland). From each sample 0.045 μg of the total RNA was reverse-transcribed into cDNA with Transcriptor First Strand cDNA Synthesis Kit (Roche Applied Science) and random hexamer primers according to the manufacturer's instructions. Then the expression of IL-1β, IL-2, IL-4, IL-6, IL-10, IL-12α, IL-17, TGF-β2, IFN-γ and B2M was quantified by relative quantification ΔΔCt method (Bustin 2000, Livak and Schmittgen 2001). We measured expression of B2M (β-2-microglobulin subunit of MHC class I) as an internal control in which no differential expression was expected. Real-time PCR was performed using LightCycler® 480 SYBR Green I Master (Roche Applied Science) with primers designed according to the partial sequences of grey partridge genes and the alignments of known homologous sequences of galliform birds published in GeneBank (Supplementary material Appendix 1, Table A1). The effectiveness of all primer pairs was tested on mixed cDNA isolated from several immunologically relevant tissues: blood, spleen and bone marrow. In cases of unsuccessful cytokine-specific real-time PCR in skin-tissue samples the amplification was repeated with other sets of primers (1–2) to avoid effect of splicing-variant polymorphism at primer-specific sites. Each real-time PCR reaction was performed three times in the LightCycler® 480 Detection System (Roche Applied Science) and the PCR efficiency was then calculated for all individual genes across all treatment and control samples. In all cases the PCR efficiency was $E > 1.72$. Relative and absolute quantification was performed in LightCycler® 480 SW 1.5 (Roche Applied Science) using the comparative Ct method (ΔCt) and Fit Points method, respectively. A housekeeping (reference) gene was selected from six candidate genes: subunit A of the succinate dehydrogenase complex (SDHA), TATA-box-binding protein (TBP), coenzyme A (CoA), 18S RNA, β-actin, and transcription elongation factor A2 (TCEA2) using geNorm analysis. GeNorm is a statistical algorithm to determine the most stable housekeeping genes from a set of tested candidate reference genes in a given sample panel based on their gene-specific inter-sample qPCR variability (Vandesompele et al. 2002). According to the geNorm analysis results, TBP was chosen as a single housekeeping gene with the best expression stability (M-value = 0.276; acceptable vales range between 0.25 and 0.7). Therefore, the relative quantity of particular cytokine cDNA was expressed as values relative to TBP. Quantification of gene expression relative to a housekeeping gene is a common approach that shows a relative change of expression per cell. In comparison, the absolute quantification should be less sensitive to changes in cellular composition of the tissue due to infiltration of novel cell types and, thus, more appropriate for analyses of inflammatory processes (Whelan et al. 2003). These two approaches to gene expression quantification are, however, only different representations of the same raw data. To select candidate genes for further analysis we estimated the fold change of cytokine expression in PHA-treated wings compared to control wings in 10 individuals. The fold change was calculated as $2^{-\Delta\Delta Ct}$ (standardised on TBP expression). Thus, up-regulation is indicated by values above 1 and down-regulation is indicated by values below 1. As a methodologically meaningful threshold of expression up-regulation we set the ten-fold mRNA increase (two-times higher than the maximum fold change of our B2M control, which was 5 × fold change). Only cytokines with expression above this threshold in at least one individual were included into the following analyses.

Statistics

Given the non-Gaussian distribution of the expression values, the cytokine expression differences between PHA-treated and control wings were tested using the paired Wilcoxon signed rank test. To test the relationship between the level of cytokine expression, ornament area, mass, body size, and the magnitude of the PHA-induced skin swelling response we used generalised linear models. In these models the distribution was approximated by the Gaussian distribution (this assumption was tested by the Shapiro–Wilk normality test). The models contained PHA skin-swelling index as a response variable and either relative or absolute cytokine expression (we included only cytokines with more than 10-fold increase in expression between treatment and control in at least one individual), body size, mass and ornament area as explanatory variables (no interactions were included due to our relatively small sample size). Minimal adequate models (MAMs), i.e. models with all terms significant (Crawley 2002), were obtained based on backward eliminations of non-significant terms ($p > 0.10$) in candidate models using likelihood ratio test assuming Chi square distribution of change in deviance. The order of term elimination was determined based on the minimal decrease in AIC values between the original model and its reduced forms. Marginally non-significant terms ($0.10 < p < 0.05$) were retained in the MAMs due to the relatively small sample size. The significance of a particular term adjusted to the effects of other terms was based on the change in deviance between the full and reduced models, distributed as F with degrees of freedom equal to the difference in the degrees of freedom between the models with and without the term in question. Correlations were tested

only in the up-regulated cytokines. This was performed by Pearson's product-moment correlation (to test for multiple correlations we adopted a Bonferroni correction). Repeatability was assessed according to Lessells and Boag (1987). All above mentioned statistical analyses were performed using R 2.12.1. software (<www.r-project.org>).

Results

In all our experimental individuals we detected the wing-web swelling response 6 h after the subcutaneous PHA treatment (min = 0.30 mm, max = 1.41 mm, mean ± SD = 0.80 ± 0.28). This was in contrast to the untreated controls, where no swelling was detected (min = –0.40 mm, max = 0.23 mm, mean ± SD = 0.00 ± 0.17). Expression of all investigated cytokines has been revealed in the collected tissue samples. IL-6 was, however, the only cytokine with significantly higher expression in the PHA-treated wing-webs then in control wing-webs (n = 10, V = 4, p = 0.014). Individual variability in the differences in relative expression between treatment and control sites (absolute mean difference ± SD) was highest in IL-1β (0.972 ± 2.374), IL-6 (2.259 ± 3.156) and TGF-β(−0.250 ± 0.775). The fold changes in expression of all measured cytokine genes are shown in Fig. 1. Only in three (IL-1β, IL-6 and TGF-β) of the nine tested cytokine genes we detected any signs of substantial expression up-regulation in at least one individual.

PHA-induced swelling was not associated with the relative expression of mRNA encoding the selected cytokine proteins (standardised to the expression of TBP) but was positively correlated with the ornament area (slope ± SE = 0.023 ± 0.008, $F_{1,22} = 7.87$, p = 0.010) and negatively correlated with body size (slope ± SE = −0.111 ± 0.026, $F_{1,22} = 16.88$, p < 0.001; MAM: $F_{2,22} = 10.67$, p < 0.001). However, analysis of absolute quantities of mRNA specific to individual genes (i.e. quantities standardised on absolute amount of total RNA and not on the TBP housekeeping gene) revealed that absolute tissue expression of TGF-β might be negatively correlated with the magnitude of the swelling response (slope ± SE = −6.277 ± 3.029, $F_{1,21} = 4.30$, p = 0.051; Fig. 2), although this association was marginally insignificant in the model that also contained ornament area (slope ± SE = 0.024 ± 0.008, $F_{1,21} = 9.96$, p = 0.005) and body size (slope ± SE = −0.098 ± 0.025, $F_{1,21} = 14.74$, p = 0.001; MAM: $F_{3,21} = 9.61$, p < 0.001). Examination of correlations in the relative expression of IL-1β, IL-6 and TGF-β identified significant relationships only between the expression of TGF-β and IL-1β (r = 0.49, $t_{23} = 2.69$, p = 0.013), and TGF-β and IL-6 (r = 0.70, $t_{23} = 4.63$, p < 0.001), both significant after Bonferroni correction for multiple tests (threshold α-level = 0.017). The correlation between IL-1β and IL-6 was not significant ($r = 0.31$, $t_{23} = 1.57$, p = 0.128).

Discussion

Our study revealed differential expression of both pro-inflammatory and anti-inflammatory cytokines in the initial phase of the PHA-induced immune response in grey partridge. However, expression of only a subset of the cytokines that are known to regulate inflammation (IL-1β, IL-6 and TGF-β) was substantially increased in the PHA-treated wing-webs. This cytokine composition indicates activation of Th17-mediated and regulatory pathways. It was proposed earlier that the PHA-induced swelling response would mirror T-cell proliferation (Tella et al. 2008). Although we measured the expression of the T-cell growth factor (IL-2) in our study we did not find any sign of up-regulation of expression in this cytokine. As predicted, no differential expression in

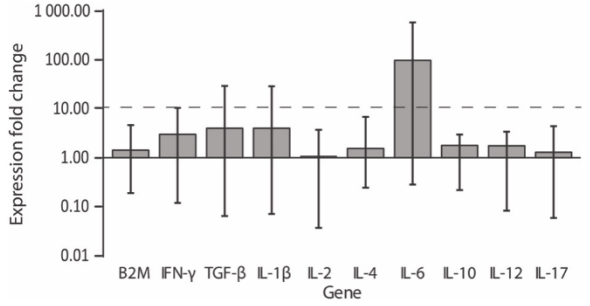


Figure 1. Fold change of cytokine expression in PHA-treated wings compared to control wings six hours after the PHA treatment (n = 10). The fold change was calculated as $2^{-\Delta\Delta Ct}$ (standardised on expression of TBP): up-regulation indicated by values above 1, down-regulation indicated by values below 1. We revealed the following fold changes in expression of our cytokine genes (mean fold change, minimum-maximum): IFN-γ 3×, 0.4× −10×; TGF-β 4×, 0.06× −29×; IL-1β 4×, 0.7× −28×; IL-2 1×, 0.03× −4×; IL-4 2×, 0.2× −7×, IL-6 96×, 0.3× −574×, IL-10 2×, 0.2× −3×, IL-12α 2×, 0.08× −3×, IL-17 1×, 0.06× −4×. In the figure, mean is represented by bar, differences between maximum and minimum by whiskers. Dashed line represents methodologically meaningful threshold of expression up-regulation (see main text for rationale) and genes with expression above this threshold in at least one individual were included into the statistical analyses.

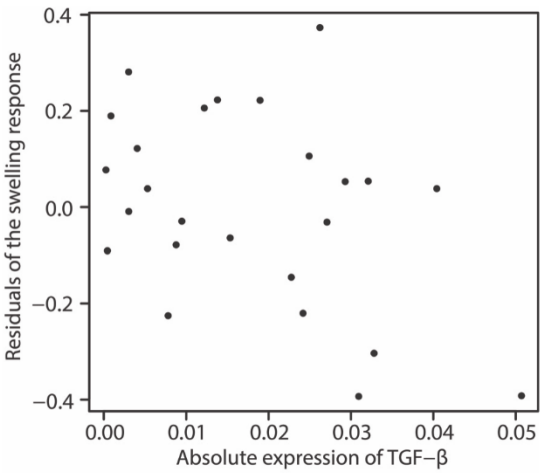


Figure 2. Association between the swelling response to PHA (measured as the increase in wing-web thickness in mm) and absolute expression of TGF-β($F_{1,21} = 4.30$, p = 0.051 in MAM containing also effect of size and ornamentation). The swelling response is plotted as residuals of wing-web thickness change after controlling for size and ornamentation.

B2M was found. Our data indicate that, at least at the final stage of the early phase of PHA-induced inflammation (6 h after treatment; i.e. in time when the type of the response commonly measured later at 24 h was determined), cytokine expression is only weakly associated with the metrically measured swelling response intensity. The negative relationship between thickness of the swelling and absolute tissue TGF-β mRNA quantity was only on the boundary of significance and, hence, needs to be treated with caution. The intensity of swelling was significantly negatively related to body size and significantly positively related to the melanin-based ornament area, which might suggest association of the response with body condition and/or performance in mate choice. Thirdly, our results corroborate the existence of relationships between expression of IL-1β, IL-6 and TGF-β.

Subcutaneous application of PHA causes rapid infiltration of various immune cells, including those involved in adaptive (lymphocytes) and innate (monocytes, basophils, neutrophils/heterophils) immunity, both in birds and mammals (Stadecker and Leskowit 1974, Goto et al. 1978, McCorkle et al. 1980, Martin et al. 2006, Vinkler et al. 2010, 2012). The Th17-character of the PHA-induced inflammatory response can be assumed based on the results of the present study. We detected high expression of IL-1β and IL-6, which are two general pro-inflammatory cytokines that are known to promote the Th17 immunological pathway (Kaiser 2010, van de Veerdonk et al. 2011).

Recent results in scarlet rosefinch *Carpodacus erythrinus* indicated that the quantity of tissue basophils (but not lymphocytes) may define the magnitude of the metrically measured skin-swelling response (Vinkler et al. 2012). These histological data suggest that contrary to some previous assumptions (Tella et al. 2008) T-cell proliferation does not cause the observed swelling response. This view is supported by three important facts. Firstly, the skin-based T cells are (at least in mammals) devoid of proliferative capacity in response to PHA (Hassan-Zahraee et al. 1998). Secondly, immunological studies in mammals show that the T-cell proliferation in the response to PHA starts after more than 24 h (Tokuyasu et al. 1968, Orteu et al. 1998), i.e. it cannot be reflected by swelling that is measured at 6 or 24 h. Thirdly, even purified PHA-E isolectin, that does not induce T-cell proliferation, stimulates an intensive skin-swelling response (Vinkler et al. 2010). The cytokine IL-2 is known to be produced by activated T cells (Lin and Leonard 2003) and to stimulate T-cell proliferation (Lillehoj et al. 2001). This cytokine was also predicted earlier to be up-regulated in the swelling (Tella et al. 2002). Our results show that in 8 out of 10 individuals the fold change in IL-2 expression was below 1 and, thus, consistent with down-regulation of expression of this cytokine. The fact that we did not find any clear up-regulation in expression of IL-2 in the present study corroborates the conclusion that T-cell proliferation is probably not an important mechanism causing the primary swelling response commonly measured by avian ecologists (though T-cell mitosis may be possibly present in some later phases of the response). This is one aspect important for discrimination between innate and adaptive immunity when interpreting the PHA skin-swelling test (as demanded in ecology ever since Kennedy and Nager 2006).

TGF-β is a commonly known negative regulator of the inflammation (Kaiser 2010). Taken together with the marginally non-significant negative association between TGF-β expression and PHA-induced swelling (Fig. 2), we assume that TGF-β negatively regulates also the inflammation underlying the swelling response to PHA. This inhibitory effect of TGF-β may result in reduced leukocyte infiltration into the inflamed skin tissue (Stadecker and Leskowit 1974, Vinkler et al. 2012). In general, however, our results show that there is only a limited association between the initial cytokine signalling triggered by the PHA treatment and the swelling response. As other studies indicated, the swelling response may be dependent on the numbers of cells that infiltrate into the tissue (Vinkler et al. 2012). We, therefore, suggest that by measuring the tissue swelling we estimate the responsiveness of cells to the cytokine signalling and not differences in the initial cellular signalling triggered in the tissue by the PHA treatment.

In the present study we have shown the positive association between expression of IL-6 and TGF-β. Furthermore, we have also found a positive correlation between IL-1β and TGF-β. This might possibly indicate that increased pro-inflammatory IL1- β/IL-6 signalling induces, in feedback, a stronger regulatory response mediated by TGF-β. The precise role of TGF-β at this stage of inflammation remains, nonetheless, unresolved, given the correlative nature of this study. Some caution is also needed given that the production of cytokines might be also regulated at the post-transcriptional level. Finally, further research would benefit from including more control genes that would enable statistical testing of the fold-change expression threshold.

Unlike cytokine expression, two phenotypic traits were significantly associated with the metrically measured magnitude of the PHA-induced swelling response. In our dataset we have found a strong negative relationship between PHA responsiveness and the animal size and a positive association between PHA responsiveness and the melanin-based ornament area. In the case of body size we assume that the relationship might be a mechanistic outcome of the cellular dispersal in the patagium tissue (see also Tella et al. 2002). For instance, larger birds might have a larger skin volume in the patagium region available for infiltration of the same leukocyte numbers as in the smaller individuals, making the swelling of the wing-web metrically thinner. Interestingly, in some other studies a relationship of an opposite direction was revealed (Møller and Petrie 2002, Vinkler et al. 2012), which makes the interpretation of the association unclear. Positive associations between melanin-based ornamentation and immune responsiveness to PHA similar to the one revealed in the present study have been found also in several other avian species (Gonzalez et al. 1999, Jacquin et al. 2011). Melanin-based ornaments are often associated with individual quality, condition and health (Andersson 1994). The positive relationship between melanin-based ornamentation and intensity of the PHA-induced swelling response might, hence, indicate that individuals in better health mount a stronger pro-inflammatory immune response (as assumed e.g. by Jacquin et al. 2011). In grey partridge the condition-dependence of the melanin-based ornament remains unknown. Some experimental data, however,

suggest that the melanin-pigmented ornament in this species may play a significant role in male–male competition (Beani and Dessifulgheri 1995, K. Kotasová, T. Albrecht. and M. Šálek pers. comm.). This is consistent with the results of our own recent study showing that males with larger melanin-based feather ornamentation show signs of elevated stress (Svobodová et al. 2013). Further investigation is needed to explore whether the test shows the investment into immunity rather than the state of current immunity activation (as suggested by Vinkler et al. 2012).

To conclude, in this study we have shown that at the end of the early phase of the PHA-induced inflammatory response tissue expression changes involve both pro-inflammatory and anti-inflammatory cytokines, but not cytokines responsible for T-cell proliferation. Our data indicate that while the metrical measurements in the PHA skin-swelling test may possibly evaluate intensity of cellular infiltration (shown in previous studies), they lack the ability to show differences in the underlying cytokine signalling. Hence, measurements of tissue cytokine expression may provide independent and potentially valuable information on the cellular activation. These findings contribute to the interpretation of results of the PHA skin-swelling test in immunological ecology which remains controversial (Kennedy and Nager 2006, Owen and Clayton 2007, Vinkler et al. 2010). The test is not a simple measurement of adaptive immunity but a measurement of complex innate pro-inflammatory responsiveness with ecological costs and benefits of innate inflammation (Ashley et al. 2012). The proper comprehension of immunological mechanisms adopted in ecological research is important for further development of ecological immunology as an up-to-date scientific field.

Acknowledgements – We are thankful to František Vitula and Tereza Chumlenová for their care about the experimental animals, to Marta Promerová and Zdeňka Csibreiová for their attentive help in laboratory and to Pete Kaiser and Lisa Rothwell for many stimulating ideas. We are also grateful to Josef Bryja, Jakub Kreisinger, Tomáš Albrecht, Jitka Vinklerová, Dagmar Vinklerová, Mark Gibson and to the Subject Editor Ruedi Nager for their valuable critical comments to the manuscript. This study was supported by the Czech Science Foundation (projects no. 206/08/1281, P505/10/1871 and P502/12/P179) and grants CIGA 20114217 and SVV-2013-267 201. The authors' contribution to this paper was as follows: MV (35%) – study design, PHA treatment, metrical measurements, statistics, main part in manuscript preparation, JS (10%) – assistance with animal manipulation, study design and statistics, HB (10%) – molecular analyses, BG (10%) – ornament analysis and AB (35%) – molecular analyses. All authors contributed by their comments to the manuscript preparation and there was no conflict of interests in this research.

References

Adamo, S. A. 2004. How should behavioural ecologists interpret measurements of immunity? – Anim. Behav. 68: 1443–1449.

Andersson, M. 1994. Sexual selection. – Princeton Univ. Press.

Ashley, N. T., Weil, Z. M. and Nelson, R. J. 2012. Inflammation: mechanisms, costs, and natural variation. – Annu Rev. Ecol. Evol. Syst. 43: 385–406.

Beani, L. and Dessifulgheri, F. 1995. Mate choice in the grey partridge, *Perdix perdix* – role of physical and behavioral male traits. – Anim. Behav. 49: 347–356.

Blount, J. D., Houston, D. C., Møller, A. P. and Wright, J. 2003a. Do individual branches of immune defence correlate? A comparative case study of scavenging and non-scavenging birds. – Oikos 102: 340–350.

Blount, J. D., Metcalfe, N. B., Birkhead, T. R. and Surai, P. F. 2003b. Carotenoid modulation of immune function and sexual attractiveness in zebra finches. – Science 300: 125–127.

Bonato, M., Evans, M. R., Hasselquist, D. and Cherry, M. I. 2009. Male coloration reveals different components of immunocompetence in ostriches, *Struthio camelus*. – Anim. Behav. 77: 1033–1039.

Buchanan, K. L., Evans, M. R. and Goldsmith, A. R. 2003. Testosterone, dominance signalling and immunosuppression in the house sparrow, *Passer domesticus*. – Behav. Ecol. Sociobiol. 55: 50–59.

Bustin, S. A. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. – J. Mol. Endocrinol. 25: 169–193.

Campbell, T. W. and Ellis, C. K. 2007. Avian and exotic animal hematology and cytology. – Blackwell.

Crawley, M. J. 2002. Statistical computing: an introduction to data analysis using S-Plus. – Wiley.

Cucco, M., Guasco, B., Malacarne, G. and Ottonelli, R. 2006. Effects of beta-carotene supplementation on chick growth, immune status and behaviour in the grey partridge, *Perdix perdix*. – Behav. Processes 73: 325–332.

Danilova, N. 2006. The evolution of immune mechanisms. – J. Exp. Zool. B 306: 496–520.

Gonzalez, G., Sorci, G. and De Lope, F. 1999. Seasonal variation in the relationship between cellular immune response and badge size in male house sparrows (*Passer domesticus*). – Behav. Ecol. Sociobiol. 46: 117–122.

Goto, N., Kodama, H., Okada, K. and Fujimoto, Y. 1978. Suppression of phytohemagglutinin skin response in thymectomized chickens. – Poult. Sci. 57: 246–250.

Graham, A. L., Allen, J. E. and Read, A. F. 2005. Evolutionary causes and consequences of immunopathology. – Annu. Rev. Ecol. Evol. Syst. 36: 373–397.

Hassan-Zahraee, M., Wu, J. P. and Gordon, J. 1998. Rapid synthesis of IFN-gamma by T cells in skin may play a pivotal role in the human skin immune system. – Int. Immunol. 10: 1599–1612.

Jacquin, L., Lenouvel, P., Haussy, C., Ducatez, S. and Gasparini, J. 2011. Melanin-based coloration is related to parasite intensity and cellular immune response in an urban free living bird: the feral pigeon *Columba livia*. – J. Avian Biol. 42: 11–15.

Kaiser, P. 2010. Advances in avian immunology – prospects for disease control: a review. – Avian Pathol. 39: 309–324.

Kaiser, P. and Staheli, P. 2008. Avian cytokines and chemokines. – In: Davison, F., Kaspers, B. and Schat, K. A. (eds), Avian immunology. Academic Press, pp. 203–222.

Kennedy, M. W. and Nager, R. G. 2006. The perils and prospects of using phytohaemagglutinin in evolutionary ecology. – Trends Ecol. Evol. 21: 653–655.

Leavitt, R. D., Felsted, R. L. and Bachur, N. R. 1977. Biological and biochemical properties of *Phaseolus vulgaris* isolectins. – J. Biol. Chem. 252: 2961–2966.

Lessells, C. M. and Boag, P. T. 1987. Unrepeatable repeatabilities – a common mistake. – Auk 104: 116–121.

Licastro, F., Davis, L. J. and Morini, M. C. 1993. Lectins and superantigens – membrane interactions of these compounds with T-lymphocytes affect immune-responses. – Int. J. Biochem. 25: 845–852.

Lillehoj, H. S., Min, W. G., Choi, K. D., Babu, U. S., Bumside, J., Miyamoto, T., Rosenthal, B. M. and Lillehoj, E. P. 2001. Molecular, cellular, and functional characterization of chicken

cytokines homologous to mammalian IL-15 and IL-2. – Vet. Immunol. Immunopathol. 82: 229–244.
Lin, J. X. and Leonard, W. J. 2003. Interleukin-2. – In: Thomson, A. W. and Lotze, M. T. (eds), The cytokine handbook. Academic Press, pp. 167–199.
Livak, K. J. and Schmittgen, T. D. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(T) (-Delta Delta C) method. – Methods 25: 402–408.
MacHugh, D. E., Taraktsoglou, M., Killick, K. E., Nalpas, N. C., Browne, J. A., Park, S. D. E., Hokamp, K., Gormley, E. and Magee, D. A. 2012. Pan-genomic analysis of bovine monocyte-derived macrophage gene expression in response to in vitro infection with *Mycobacterium avium* subspecies *paratuberculosis*. – Vet. Res. 43: 810–812.
Martin, L. B., Pless, M., Svoboda, J. and Wikelski, M. 2004. Immune activity in temperate and tropical house sparrows: a common-garden experiment. – Ecology 85: 2323–2331.
Martin, L. B., Han, P., Lewittes, J., Kuhlman, J. R., Klasing, K. C. and Wikelski, M. 2006. Phytohemagglutinin-induced skin swelling in birds: histological support for a classic immunoecological technique. – Funct. Ecol. 20: 290–299.
Martin, L. B., Hawley, D. M. and Ardia, D. R. 2011. An introduction to ecological immunology. – Funct. Ecol. 25: 1–4.
McCorkle, F., Olah, I. and Glick, B. 1980. Morphology of the phytohemagglutinin-induced cell response in the chickens wattle. – Poult. Sci. 59: 616–623.
Millet, S., Bennett, J., Lee, K. A., Hau, M. and Klasing, K. C. 2007. Quantifying and comparing constitutive immunity across avian species. – Dev. Comp. Immunol. 31: 188–201.
Møller, A. P. 1998. Evidence of larger impact of parasites on hosts in the tropics: investment in immune function within and outside the tropics. – Oikos 82: 265–270.
Møller, A. P. and Petrie, M. 2002. Condition dependence, multiple sexual signals, and immunocompetence in peacocks. – Behav. Ecol. 13: 248–253.
Møller, A. P., Erritzøe, J. and Saino, N. 2003. Seasonal changes in immune response and parasite impact on hosts. – Am. Nat. 161: 657–671.
Orteu, C. H., Poulter, L. W., Rustin, M. H. A., Sabin, C. A., Salmon, M. and Akbar, A. N. 1998. The role of apoptosis in the resolution of T cell-mediated cutaneous inflammation. – J. Immunol. 161: 1619–1629.
Owen, J. P. and Clayton, D. H. 2007. Where are the parasites in the PHA response? – Trends Ecol. Evol. 22: 228–229.
Pedersen, A. B. and Babayan, S. A. 2011. Wild immunology. – Mol. Ecol. 20: 872–880.
Rigas, D. A. and Johnson, E. A. 1964. Studies on phytohemagglutinin of *Phaseolus vulgaris* and its mitogenicity. – Ann. N. Y. Acad. Sci. 113: 800–818.

Supplementary material (Appendix JAB5860 at <www.oikosoffice.lu.se/appendix>) Appendix 1.

Saino, N., Ambrosini, R., Martinelli, R., Ninni, P. and Møller, A. P. 2003. Gape coloration reliably reflects immunocompetence of barn swallow (*Hirundo rustica*) nestlings. – Behav. Ecol. 14: 16–22.
Schmid-Hempel, P. 2011. Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics. – Oxford Univ. Press.
Smits, J. E., Bortolotti, G. R. and Tella, J. L. 1999. Simplifying the phytohaemagglutinin skin-testing technique in studies of avian immunocompetence. – Funct. Ecol. 13: 567–572.
Stadecker, M. J. and Leskowit, S. 1974. Cutaneous basophil response to mitogens. – J. Immunol. 113: 496–500.
Svobodová, J., Gabrielová, B., Synek, P., Maršík, P., Vaněk, T., Albrecht, T. and Vinkler, M. 2013. The health signalling of ornamental traits in the grey partridge (*Perdix perdix*). – J. Ornithol. 154: 717–725.
Tella, J. L., Scheuerlein, A. and Ricklefs, R. E. 2002. Is cell-mediated immunity related to the evolution of life-history strategies in birds? – Proc. R. Soc. B 269: 1059–1066.
Tella, J. L., Lemus, J. A., Carrete, M. and Blanco, G. 2008. The PHA test reflects acquired T-cell mediated immunocompetence in birds. – PLoS One 3: e3295.
Thomson, A. W. and Lotze, M. T. 2003. The cytokine handbook. – Academic Press.
Tokuyasu, K., Madden, S. C. and Zeldis, L. J. 1968. Fine atructural alterations of interphase nuclei of lymphocytes stimulated to growth activity in vitro. – J. Cell Biol. 39: 630–660.
van de Veerdonk, F. L., Netea, M. G., Dinarello, C. A. and Joosten, L. A. B. 2011. Inflammasome activation and IL-1 beta and IL-18 processing during infection. –Trends Immunol. 32: 110–116.
Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A. and Speleman, F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. – Genome Biol. 3: research 0034.
Vinkler, M. and Albrecht, T. 2011. Handling immunocompetence' in ecological studies: do we operate with confused terms? – J. Avian Biol. 42: 490–493.
Vinkler, M., Bainová, H. and Albrecht, T. 2010. Functional analysis of the skin-swelling response to phytohaemagglutinin. – Funct. Ecol. 24: 1081–1086.
Vinkler, M., Schnitzer, J., Munclinger, P. and Albrecht, T. 2012. Phytohaemagglutinin skin-swelling test in scarlet rosefinch males: low-quality birds respond more strongly. – Anim. Behav. 83: 17–23.
Whelan, J. A., Russell, N. B. and Whelan, M. A. 2003. A method for the absolute quantification of cDNA using real-time PCR. – J. Immunol. Methods 278: 261–269.