CHARLES UNIVERSITY IN PRAGUE

FACULTY OF SCIENCE

Study program Biology

Study field Zoology

Martin Těšický

Trans-species polymorphism in selected innate immunity genes in tits (Paridae family)

Mezidruhový polymorfismus vybraných genů vrozené imunity u sýkor (Paridae)

Master thesis

Supervisor: RNDr. Michal Vinkler, Ph.D.

Co-supervisor: Mgr. Hana Velová

RNDr. Radka Reifová, Ph.D.

Prague 2016

i. Abstract

Adaptation of host receptor system to optimal detection of infection-related structures is one of the key evolutionary challenges of immunity in host-pathogen interactions. Toll-like receptors (TLRs) are genetically variable molecules of vertebrate innate immunity that recognise danger signals, e.g. pathogenic molecules. Examination of genetic variation in *TLR*s may reveal mechanisms of host immunity adaptation to pathogenic pressure at molecular level. Trans-species polymorphism (TSP) is a phenomenon which assumes that several identical alleles or allelic lineages are inherited from ascendant to descendant species and these may be subsequently maintained over a long period of time in a polymorphic state. Whereas in adaptive immune genes the concept of TSP is well understood, little is presently known about TSP in innate immune genes such as *TLR*s. In this thesis I describe genetic polymorphism in functionally-relevant regions of TLR4 and TLR5 in 192 individuals representing 20 species Paridae family (tits, chickadees and titmice). These two receptors bind mainly bacterial ligands (TLR4 detects lipopolysaccharide and TLR5 detects flagelline), being among the first ones to trigger immune response to bacterial pathogens. To differentiate presumed TSP from gene flow among species, intron sequences of six autosomal neutral markers were sequenced. *TLR*s were variable on intra- and interspecific level in Paridae. Positive selection was detected in 14 amino acid residues in TLR4 and in 23 residues in TLR5. From these positively selected sites 4 positions in TLR4 and 14 positions in TLR5 were located in close proximity to predicted functionally important sites or being directly in the predicted binding sites. TSP was detected in both *TLR4* and *TLR5* genes in closely related species within genus level (American *Poecile*, *Cyanistes* and *Baeolophus*) assuming that no TSP was older than 4-8 millions of years. Given the extensive sharing of alleles in neutral markers and the recent divergence among these species we were not able to distinguish whether TSP identified in *TLR4* and *TLR5* is balanced or transient. Significant gene flow was detected within two pairs of closely related species assuming that at least some portion of shared polymorphism in *TLR4* and *TLR5* may originate from introgression. In this thesis I report for the first time TSP in *TLRs* and Pattern recognition receptors in general and provide evidence that TSP is a general evolutionary phenomenon in immune genes. Besides that, positively selected residues indentified in TLR4 ad TLR5 might have functional importance for binding properties of the TLRs and thus recognition of pathogens.

Key words: immune genes, innate immunity, introgression, selection, shared variability, TLR4, TLR5, trans-species polymorphism, TSP

ii. Abstrakt

Toll-like receptory jsou geneticky variabilní molekuly vrozené imunity obratlovců, které rozpoznávají tzv. struktury nebezpečí, např. struktury patogenů. Vyšetření genetické variability u *TLR*s může poodhalit obecné adaptace imunitního systému hostitelů proti tlaku parazitů na molekulární úrovni. Koncept mezidruhového polymorfismu (TSP) předpokládá, že několik identických alel či alelických linií je zděděno od společného předka druhů druhy dceřinými, u nichž následně mohou být dlouhodobě udržovány v polymorfním stavu. Zatímco u genů získané imunity je TSP dobře prostudován, naše znalosti o TSP v genech vrozené imunity, např. *TLR*s, u nichž bychom mohli TSP předpokládat, jsou nedostatečné. V této práci se proto zaměřuji na popis genetického polymorfismu ve funkčně významných oblastech TLR4 a TLR5 u 192 jedinců 20 druhů sýkor z čeledi sýkorovitých (Paridae). Tyto receptory vážou převážně bakteriální ligandy (TLR4 rozpoznává lipopolysacharid a TLR5 flagelin) a podílejí se tak na prvotní aktivaci imunity proti bakteriálním patogenům. Pro odlišení případného TSP od sdíleného polymorfismu způsobeného genovým tokem byly osekvenovány také introny šesti autosomálních neutrálních markerů. Ze získaných dat vyplývá *TLR*s jsou variabilní na vnitrodruhové a mezidruhové úrovni u sýkorovitých. Pozitivní selekce byla detekována na 14 aminokyselinových pozicích v *TLR4* a na 23 pozicích v *TLR5*. Z těchto selektovaných pozic se zároveň 4 pozice u *TLR4* a 14 pozic u *TLR5* nacházely v blízkosti predikovaných funkčně významných míst anebo byly přímo ve vazebných místech. TSP byl detekován jak v *TLR4,* tak v *TLR5* mezi blízce příbuznými druhy na úrovni rodu (konkrétně mezi americkými sýkorami rodu *Poecile* a dále pak v rodech *Cyanistes* a *Baeolophus*). Předpokládaná doba perzistence TSP tak nebyla vyšší než 4-8 milionů let. Nicméně vzhledem k nedávné divergenci mezi těmito druhy a rozsáhlému sdílení alel také u neutrálních markerů nebylo možno rozlišit, zda se u *TLR4* a *TLR5* jedná o balancovaný či transietní TSP. Výrazný genový tok byl detekován v rámci dvou dvojic blízce příbuzných druhů sýkor. To naznačuje, že minimálně část sdíleného polymorfismu v *TLR4* a *TLR5* by mohla pocházet z introgrese. V této práci jsem vůbec poprvé detekoval TSP u TLRs a u Pattern recognition receptorů, což naznačuje, že TSP je obecným evolučním jevem u imunitních genů. Identifikované pozitivně selektované pozice u TLR4 a TLR5 ležící v blízkosti vazebných míst by mohly ovlivňovat vazebné vlastnosti těchto receptorů a následné rozpoznání patogenů.

Klíčová slova: imunitní geny, introgrese, selekce, sdílená variabilita, *TLR4*, *TLR5*, trans-species polymorfismus, TSP, vrozená imunita

iii. Prohlášení

Tkáňové vzorky použité pro molekulárně-genetické analýzy v této práci pocházejí z Genetické kolekce Burkeho musea, University of Washington, USA, z Genetické banky Katedry zoologie, Přírodovědecké fakulty Univerzity Karlovy v Praze a z terénního sběru vzorků v České republice, který jsem samostatně prováděl v letech 2013 a 2014. Tato práce vzhledem k svému rozsahu vychází ze spolupráce celé řady spolupracovníků, z tohoto důvodu jsem se rozhodl v některých částech používat množné číslo. Přesto je můj podíl na předkládané práci hlavní. Podíl jednotlivých spolupracovníků na této práci je zmíněn v poděkování. Laboratorní část probíhala v genetické laboratoři Katedry zoologie, PřF UK a v molekulárně-genetických laboratořích Detašovaného pracoviště Studenec Ústavu biologie obratlovců AV ČR pod vedením Mgr. Hany Velové a Mgr. Anny Bryjové. Finální příprava sekvenačního běhu a sekvenování byly provedeny v laboratořích European Molecular Biology Laboratory v Heidelbergu ve spolupráci s Dr. Vladimírem Benešem. Výpočetně náročné analýzy byly provedeny za použití serveru Xukol Katedry zoologie, PřF UK.

Prohlašuji, že jsem závěrečnou práci zpracovával samostatně a že jsem uvedl všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze dne 14. 8. 2016

……..

Martin Těšický

iv. Content

1. General introduction

Parasitism is considered to be one of the most common ecological relationships around the world. Parasites therefore exert strong pervasive selection pressure on their host and try to overcome the diverse defensive mechanisms which were evolved just against parasites (Schmid-Hempel 2011). Thus, hosts and parasites have been constantly forced to adapt to one another. This relationship is one of the type of co-evolution that shapes natural and sexual selection. Co-evolution manifests as continuous arm race between hosts and parasites well-illustrated by the quotation which precisely depicts the *Red Queen hypothesis*: "It takes all the running you can do to keep in the same place" (van Valen, 1973). According to the *Red Queen hypothesis* this arm race can select on high polymorphism on both sides and then the arm race may be detectable even on molecular level. This is especially truth for immune genes whose products directly interact with pathogenic structures and therefore they are exposed to strong parasite-mediated selection. This phenomenon is probably the best understood in Major histocompatibility genes (MHC) (Piertney and Oliver, 2006). *MHC* genes are extremely variable genes of adaptive immunity in jawed vertebrates. They code proteins which bind short fragments of peptides of both endogenous and exogenous origin and play the central role in self and non-self recognition (Neefjes et al., 2011). Polymorphism in *MHC* is associated with resistance or susceptibility to diverse parasites and infectious diseases (Jeffery and Bangham, 2000; Trowsdale, 2011). Strong positive selection operating here manifests on molecular level by increased ratio of non-synonymous substitutions to synonymous ones (Nielsen, 2005). As a consequence, this diversifying selection leading to generating high polymorphism occurs particularly in positions which directly interact with parasitic structure, e.g. residues in peptide binding region. On the other hand, positions which determine anchoring molecules or the general shape of the molecule are functionally constrained being under the influence of purifying selection (Hughes and Yeager, 1998). Beside positive selection on emergence of new advantageous alleles, also ancestral alleles that are time-proven and well established can be used for speeding up the co-evolutionary arm race. In the case of strong long-lasting balancing selection, advantageous alleles or at least their allelic lineages may persist for millions or even tens of millions of years and can be passed through species boundaries as identical or nearly identical alleles (Klein et al., 1998). This phenomenon is termed trans-species polymorphism (TSP) and assumes that several allelic lineages are inherited from ascendant to descendant species and these may be subsequently maintained over a long period of time in a polymorphic state (Klein et al., 2007, 1998). Whereas in *MHC* genes, where concept of TSP was postulated and is well understood, little is presently known about TSP in innate immune genes, e.g. Pattern Recognition Receptors (PRRs) which also play a crucial role in pathogen recognition (Kawai

and Akira, 2010a; Takeuchi and Akira, 2010). Among them, the most known are Toll-like (TLRs) receptors that provide the first sensing of Pathogen associated molecular patterns (PAMPs) serving as a danger signal followed by triggering of early immune response. Simultaneously, TLRs also co-activate adaptive immunity (Kumar et al., 2009a; Uematsu and Akira, 2008). TLRs therefore represent one of the functional bridges between innate and adaptive immunity (Takeuchi and Akira, 2010). Although relatively evolutionary conserved, considerable nonsynonymous polymorphism in binding sites has been documented on both interspecific and intraspecific level (Alcaide and Edwards, 2011; Fornuskova et al., 2013; Vinkler et al., 2014) as well as association of particular *TLR* alleles with resistance or susceptibility to infectious diseases (Netea et al., 2012). Considering the direct physical association between PRRs and PAMPs in triggering the immune response (Lee and Min, 2007), in concordance with the *Red Queen hypothesis* we may predict strong evolutionary pressure maintaining balanced frequencies of PRR alleles. Therefore, TLRs would be good candidate genes for TSP oriented research.

1.1 Concept of trans-species polymorphism (TSP)

Trans-species polymorphism (TSP) is described as the occurrence of identical alleles or allelic lineages in similar species, excluding instances where the similarity arose by the convergence. These alleles are more similar in related species than alleles within species. TSP is generated by the passage of alleles from ancestral species to descendant species (Klein et al., 2007, 1998). TSP is, therefore, a special example of genetic polymorphism. Genetic polymorphism is a long-term occurrence of two or more genotypes in a population in frequencies that cannot be attributed to a recurrent mutation (King et al. 2006). Generally, we distinguish two forms of TSP – neutral TSP and balanced TSP. Although this distinguishing is rather virtual with no strict boundaries and sometimes it can hardly be done, it has an important consequence in term of adaptive value of such polymorphism. Neutral TSP (transient or also sometimes referred to as ancestral polymorphism) is a consequence of an extensive incomplete lineage sorting (ILs). It is frequent in closely related newly diverged species and as time passes it gradually disappears (Klein et al., 1998). Time persistence of neutral TSP is highly affected by the effective population size (N_e) and the speed of divergence (Klein et al., 1998). As a consequence, neutral TSP has a tendency to be widespread in a short term after speciation and/ or in adaptively radiated species in plenty of loci (Klein et al., 1998; Nagl et al., 1998; Samonte et al., 2007). This neutral TSP has also low coalescence and, therefore, is rather suitable for study of speciation, phylogeny and population demography within thousands up to millions of years (Klein et al., 1998; Samonte et al., 2007).

Considering evolution of immune system, host-pathogen interactions and adaptive variability in general, balanced long-lasting TSP is much more important (Klein et al., 2007). Balanced TSP in immune related genes is maintained by balancing selection and commonly persists for millions of years (Aguilar and Garza, 2007; Kamath and Getz, 2011; Li et al., 2011). Several mechanisms of balancing selection have been proposed to contribute to maintaining long-termed polymorphism: *heterozygote advantages hypothesis* (*overdominant selection*) (Hughes and Yeager, 1998; Jeffery and Bangham, 2000), *negative-frequency dependent selection* (Milinski, 2006; Yeager and Hughes, 1999) and *spatiotemporally fluctuating selection* (Meyer and Thomson, 2001; Spurgin and Richardson, 2010).

First, the hypothesis of *heterozygote advantages* proposes that individuals heterozygous in immune genes, e.g. in MHC loci are able to present a wider spectrum of antigenic peptides from a pathogen to T-cells than homozygotes. As a result, heterozygotes can challenge more parasites and have higher fitness and surveillance in comparison to both homozygotes. The benefits of heterozygosity in certain loci depend on particular alleles and on the degree of overlap among the repertoires of peptides that alleles can bind and present (Hughes and Yeager, 1998; Jeffery and Bangham, 2000; Wegner et al., 2004). Empiric evidence for heterozygote advantage has been reported for number of cases, reviewed in (Bernatchez and Landry, 2003; Hedrick, 2012; Penn et al., 2002). For example, outbred heterozygous individuals in *MHC* genes of *Oncorhynchus tshawytscha* have lower mortality after the experimental infection of IHNV (infectious hematopoietic necrosis virus) and are less susceptible to *Gyrodactylus* infection than homozygotes (Arkush et al., 2002). Second, the hypothesis of *frequency dependent selection* or more accurate *negative-frequency dependent selection* supposes that fitness of the host is dependent on the allele frequency in the population (Milinski, 2006). Parasites adapt to just the most common genotypes in the population, leaving out rare, the least infected genotypes (Spurgin and Richardson, 2010). The hypothesis assumes that rare alleles are favoured to increase in the frequency in a population up the specific equilibrium, but selected against when they become common since selected advantage of alleles negatively correlate with their frequency in the population. In a long term level the frequency of the alleles oscillates in populations and balanced polymorphism is maintained (Jeffery and Bangham, 2000). Empirical supports come from associations of MHC class I (*MHC I*) and MHC class II (*MHC II)* alleles of susceptibility to diseases (Jeffery and Bangham, 2000; Wegner et al., 2004). Third, *spatiotemporally fluctuating selection* assumes that selection pressure varies in different space and time (for this reason it is also called space-time selection) as a parasite abundance in different host subpopulations (Meyer and Thomson, 2001; Spurgin and Richardson, 2010). As a consequence, it creates a distinct selection pressure on different host populations.

Although there is still lacking empirical evidence in the nature for the importance of the mechanism, theoretical approaches suggest that *fluctuating selection* is admissible for harbouring *MHC* polymorphism (Hedrick, 2002).

In last paragraphs I have tried to describe the mechanism of balanced polymorphism driven by parasite mediated selection. Nevertheless, there is a gradual continuum among different hypotheses for the maintenance of polymorphism rather than a mutually exclusive model of selection. In addition, other mechanisms which are based on sexual selection (and hence linked with parasite mediated selection) contribute to harbouring polymorphism in immune genes (Meyer and Thomson, 2001; Penn, 2002). Therefore, they may contribute in some cases to long term maintaining of TSP.

Apart from TSP in immune genes (Chapter 1.2), TSP is well documented in *S*-genes of self-incompatibility loci preventing self-fertilization in Angiosperms, similar to *MHC* genes with highly variable and divergent alleles (Dwyer et al., 1991; Ioerger et al., 1990; Richman et al., 1995), in mating loci in fungi (Lukens et al., 1996), in ABO blood system in primates (Kermarrec et al., 1999; Martinko et al., 1993; Ségurel et al., 2013), in complementary sex determiner (*CSD* gene) influencing sex ratio in Hymenoptera (Heimpel and de Boer, 2008; Lechner et al., 2014) and in other proposed loci, as reviewed in (Klein et al., 1998).

1.2 TSP in immune genes

Presently, most articles dealing with TSP focus only on acquired immune genes, namely on genotyping *MHC* (Figure 1). MHC is an extremely polymorphic and highly dynamic multigene family encoding adaptive immunity receptors which play crucial role in immune defence against parasites in jawed vertebrates (Edwards and Hedrick, 1998; Hughes and Yeager, 1998). To simplify, classical MHC code glycoproteins in cytoplasmic membrane bind endogenous (MHC I) and exogenous oligopeptides (MHC class I and MHC II) originated from cell processing, and present them to T-cells. Owing to the importance of MHC in immune response and high variability in both intra- and interspecies levels, most studies dealing with TSP have been focusing traditionally on MHC genes, chiefly on their peptide binding region (PBR; summarized in Table 1). These PBRs directly physically interact with pathogenic molecules and are threfore exposed to the strong parasite-mediated selection (Edwards and Hedrick, 1998; Hughes and Yeager, 1998).

MHC class I consists of transmembrane α-chain composed of α_1, α_2, α_3 domain and β-microtubulin. PBR is coded by exon 2 (α_1-domain) and exon 3 (α_2-domain) and it binds oligopeptide fragments of approximately 8-11 amino acids in length originating from intracellular parasites inhabiting cytosolic milieu or endogenous peptides (Jeffery and Bangham, 2000; Neefjes et al., 2011). MHC II molecule consists of two non-covalently

associated α chains (α_1,α_2) and β chains (β_1,β_2) which are products of two different genes, in general termed as *MHC II A* and *MHC II B*. Their PBR is formed by N-terminal domains of these molecules - α_1 (exon 2) and β_1 (exon 2) (Hughes and Yeager, 1998; Jeffery and Bangham, 2000). It is more opened allowing binding of longer oligopeptides, approximately 15-35 amino acids in length which come from extracellular parasites or intracellular parasites inhabiting vesicular system (Hughes and Yeager, 1998; Neefjes et al., 2011).

TSP in MHC involves almost exclusively only exons coding PBR (exon 2 and exon 3 for MHC I and exon 2 for *MHC II B* and *MHC II A)* (Těšický and Vinkler, 2015). Beside that, TSP has been reported also in exons coding transmembrane chains but much less spreaded (Bos and Waldman, 2006). In MHC genes TSP has been described in number of taxa including mammals (Janova et al., 2009; Kriener et al., 2001; Zhou et al., 2005), reptiles (Glaberman and Caccone, 2008; Jaratlerdsiri et al., 2014; Stiebens et al., 2013) , amphibians (Bos and Waldman, 2006; Shu et al., 2013; Zhao et al., 2013) and fish (Aguilar and Garza, 2007; Kiryu et al., 2005; Ottova et al., 2005). Sharing identical or nearly identical alleles is common between closely related taxa in time scale of millions of years (MY), exceptionally up to tens of millions of years in mammals (Go et al., 2005; Kundu and Faulkes, 2007). In highly diverged allele lineages TSP may persists for tens of MY in mammals (Kriener et al., 2001), while in fish and reptiles the oldest allelic lineages are considered to be older than 100 MY (Stiebens et al., 2013; Wang et al., 2010). In birds TSP has been identified for the first time in genus *Acrocephalus* in MHC I (Richardson and Westerdahl, 2003), later in different taxa including e.g. *Luscinia* (Anmarkrud et al., 2010), Ardeidae (Li et al., 2011), Spheniscidae (Kikkawa et al., 2009) and recently in genus *Phoenicopterus* (Gillingham et al., 2016) or *Anthus* (Gonzalez-Quevedo et al., 2014).

Compared to MHC, little is currently known about TSP in innate immunity genes. Although TSP has been reported there only in several cases, it involves molecules from different families. Their common features could be that these molecules directly interact with pathogenic structure and their polymorphism is associated with resistance or susceptibility to diseases. TSP has been documented in Host defense peptides (HDPs), a diversified group of unrelated proteins possessing many functions. Regarding immunity, they play key role mainly in pathogen killing, e.g. by disruption of cytoplasmic membrane (Ganz, 2003). In HDPs TSP has been detected in avian β-defensins between *Parus major* a *Cyanistes caeruleus* (Hellgren and Sheldon, 2011), in cathelicidin in Gadidae (Halldórsdóttir and Árnason, 2015) and in *Drosophila* in six out of eleven investigated HDPs (Unckless and Lazzaro, 2016). In *TRIM5α* gene TSP appears in Ceratopogonidae in the domain which determines restriction specificity (Newman et al., 2006). This gene codes a viral restriction factor which interacts with viral capsid proteins in cytosol during retrovirus infection and thus prevents reverse

transcription (Johnson and Sawyer, 2009). Strong balancing selection leading to TSP has been also documented in *OAS1* gene (Ferguson et al., 2008) which is involved in an activation of latent endoribonuclease RNase L resulting in degradation of dsRNA and inhibition of viral replication during flavirus infections (Hovanessian and Justesen, 2007). TSP was reported between *Mus musculus* and *M. famulus* and it involved only the C-terminal domain of OAS1b which is responsible for the enzyme tetramerization and protein-protein binding. There were two highly diverged allelic lineages which provide resistance against different groups of flavivirus infections (Ferguson et al., 2008).

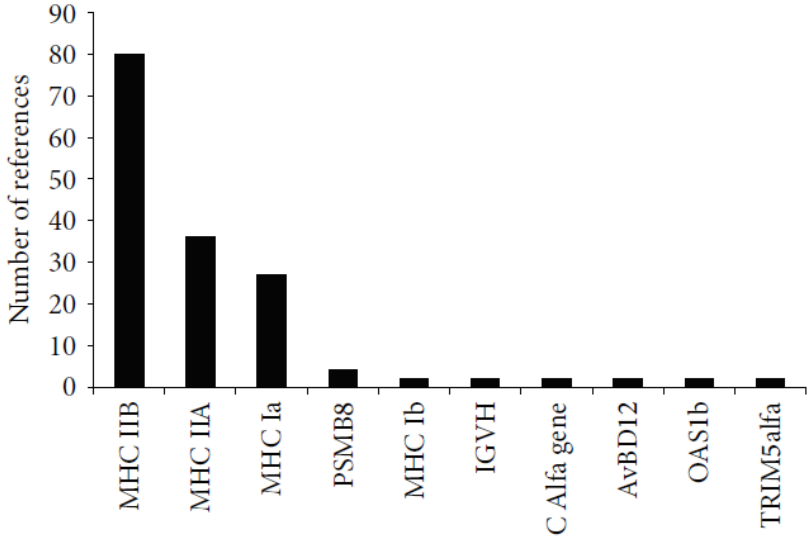


Figure 1: Number of published research articles dealing with TSP in vertebrate immune genes available on Web of Science.

Final update 19 March 2015 (adopted from Těšický and Vinkler (2015), wherein also see for more details).

1.3 Evolutionary mechanisms explaining the origin of shared variability: distinguishing TSP from other TSP-like patterns

Several evolutionary patterns have been reported which can be applied to explain the existence of shared polymorphism as an occurrence of identical or similar alleles in related taxa. Besides TSP, which is discussed in detail in Chapter 1.1, there are also convergent evolution and introgression (see Figure 2). Despite the fact that most studies have been conducted on *MHC* and I will focus on a description of these mechanisms just in *MHC*, the conclusions may also provide a more general insight to mechanisms employed in other immune genes.

First, convergent evolution is termed as "a process whereby organisms independently evolve similar traits as a result of adaptation to similar environments or ecological niches"

(Klein et al., 2007). Although convergence may be quite common in *MHC* genes, mainly among more distant species usually with estimated divergence in order of tens of MY, its detection has been difficult to demonstrate (Hughes and Yeager, 1998). Several cases of convergence have been reported, for example in exon 2 of the *MHC II DRB* gene between New World monkeys and human (Kriener et al., 2000), in exon 2 of the *MHC II DRB, DQA, DQB* and *DPB* between New World monkeys and Old World monkeys (Kriener et al., 2001) and among different Placental mammals(Hughes and Yeager, 1998) . In the case of convergence, new motifs should evolve independently in several different evolutionary lineages and the most recent common ancestor (MRCA) of the lineages should not have these motifs (Segurel et al., 2012). As the situation in MRCA is often unknown, different strategies sometimes have to be applied. Convergent evolution usually operates in short functionally important motifs, for example in *MHC* genes in PBR site rather than non-peptide binding region (non-PBR) (Hughes and Yeager, 1998). By comparison of the topology of phylogenetic trees constructed separately on PBR sequence and non-PBR sequence of *MHC II* exon 2 (coding β_1- domain) and in the case of the discrepancy between them it may be a signal for convergent evolution (Musolf et al., 2004). The same may be true for a discrepancy of phylogenetic trees between exon 2 and flanking introns (Klein et al., 1998; Kriener et al., 2000; Kundu and Faulkes, 2007), and exon 3 (coding trans-membrane α_2- domain), since introns and exon 3 usually have not been a subject of convergent evolution (Kriener et al., 2000). Another way how to elucidate convergence is based on idea that synonymous and non-synonymous sites in the same region should have different evolutionary history if convergence occurred (Li et al., 2011)

Second, hybridization with subsequent introgression may be underestimated confounding factor of TSP. The importance of hybridization in zoology used to be minimized which resulted in that there have been only few convincing examples of adaptive introgression in animals (Hedrick, 2013). Contrary to convergence as the source of the same adaptive variants for distant species, introgression occurs mainly in evolutionary young, radiated or closely related species with incomplete reproductive isolation mechanisms (RIM). Mixing alleles of the both trans-specific and hybrid origin, which are barely distinguishable, have been reported in adaptive radiated species, in Darwin finches (Sato et al., 2011; Vincek et al., 1997) or cichlid fish of *Haplochromis* species flock of East Africa crater lakes (Samonte et al., 2007). In addition, hybridization can occur in diverged taxa, and almost one tenth of bird species may hybridize (Grant and Grant 1992). How common is hybridization linked adaptive introgression of immune genes in nature? There is still lack of evidence. Existence of identical *MHC* alleles shared among species which diverged several millions of years ago (MYA) might suggest introgression as a prospective mechanism. For instance in exon 2 in *MHC DRB*-like gene in penguins (Bollmer et al., 2007), in different *MHC II* loci in cetaceans (Xu et al., 2009)

in *MHC II B* in trout (Aguilar and Garza, 2007), in *MHC II* in newts (Nadachowska-Brzyska et al., 2012) and in *DAB* genes in cyprinid fish (Seifertová and Šimková, 2011), where identical or nearly identical alleles of exon 2 were shared among species and simultaneously hybridization was detected. What is interesting is that in the last case sympatric populations of two species of *Chodrostoma* fish share more "trans-specific" alleles of genes in comparison to allopatric populations. It may implicate that at least some alleles might result from introgression rather than TSP. It was suggested that species living in sympatry can be expcted to face similar parasite exposure and then the adaptive introgression of resitance alleles could be advantagous (Wegner and Eizaguirre, 2012). If the sequence closely linked to the adaptive variant is indicative of another species, then this should indicate that the variant is the result of adaptive introgression. Moreover, distinguishing between adaptive introgression and long-term retention of polymorphism (TSP) in balanced loci is difficult (Hedrick, 2013; Wegner and Eizaguirre, 2012). In the case of TSP, we would expect that haplotype blocks should be smaller than under gene introgression, which should cause linkage disequilibrium in a larger genomic region around the *MHC* (Hedrick, 2013; Wegner and Eizaguirre, 2012). These authors suggest to combine large data sets with applying genomic methods, i.e. highly dense SNPs chips, RAD sequencing or also to use NGS methods. Adaptive introgression should preferentially concern immune genes in comparison with neutral loci (Grossen et al., 2014; Nadachowska-Brzyska et al., 2012). To conclude, more effort should be given into distinguishing other TSP-like patterns from true TSP in immune genes.

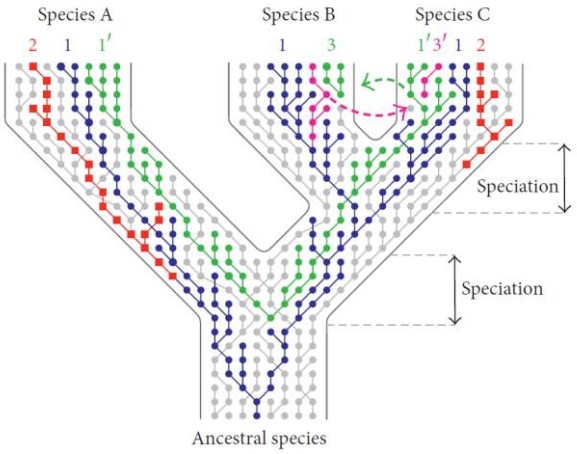


Figure 2: Mechanisms explaining polymorphism shared between taxa.

The three proposed mechanisms are depicted in alleles' genealogy: (1) trans-species polymorphism, TSP (incomplete lineage sorting; allelic lineages predate speciation and are passed to descendent species), (2) convergence (allelic lineages evolve similar features independently in separate lineages), and (3) introgression (allelic lineages are horizontally transferred either from recipient species to donor species or in both directions). Each row depicts a gene pool of one generation, each circle/square an allele of specific features. Different colours highlight individual allelic lineages, where interconnecting lines mark antecedent-descendent relationships. Green and purple dashed arrows represent directions of introgression (adopted from Těšický and Vinkler (2015).

1.4 Pattern recognition receptors (PRRs)

Pattern recognition receptors (PRRs) are evolutionary conserved molecules of innate immunity in vertebrates which recognize danger signals (alarmins) of both exogenous (Pathogen associated molecular patterns, PAMPs also known as Microbe associated molecular patterns, MAMPs) and endogenous origin (Damages associated molecular patterns, DAMPs) (Akira et al., 2006; Kawai and Akira, 2010b). PRRs are germ-line encoded, non-clonal and constitutively expressed. They include various families of receptors present not only in vertebrates, but also in invertebrates and plants (Nurnberger et al., 2004; Zipfel and Felix, 2005). However, in the following text I will focus only on vertebrate PRRs. PRRs include five different types of receptors: Toll-like receptors (TLRs), Nucleotide oligomerization binding domain receptors (NOD-like receptors, NLRs), Retinoic acid inducible gene (RIG-I like receptors, RLRs), C-type lectin receptors (CLRs) and pentraxins (Mogensen and H, 2009; Takeuchi and Akira, 2010). Whereas some authors term exogenous ligands by PAMPs (Kawai and Akira, 2010b), I would prefer the term Microbe Associated Molecular Patterns (MAMPs) over PAMPs since not all ligands that are recognized via PRRs are derived from pathogens (Bianchi, 2007). MAMPs are characteristic evolutionary conserved microbial components necessary for the survival of microbes whose expression cannot be blocked, e.g. flagellin from flagellum of bacteria, LPS (lipopolysaccharide) from cytoplasm membrane of Gram-negative bacteria, viral single stranded and double stranded RNA (ssRNA a dsRNA), CpG motif of viral and DNA, GPI anchor and others (Kawai and Akira, 2011; Kumar et al., 2009a; Mogensen, 2009).

Moreover, stimulation of immune system via PRRs by microbial ligands like those derived from commensal microorganism from gut microbiota are instrumental for good function of immune system (Chu and Mazmanian, 2013). Under stress or non-infectious inflammatory conditions, endogenous ligands can be released from cytosol to extracellular space which can lead to their denaturation (after change from reductive to oxidative environment) and as a result they may become immunogenic. This is true e.g. for High Mobility Group Box-1 (HMGB1), Heat shock proteins (HSP), RNA, DNA, S100 (Bianchi, 2007; Mogensen and H, 2009). Other molecules like phosphatidylserine are immunogenic only after they move from inner to outer cytoplasm membrane as happens during cell death. These molecules called Damage Associated Molecular Patterns can initiate and propagate non-infectious inflammatory response or perpetuate immune response during infectious inflammation as a result of tissue injury and cell lysis (Bianchi, 2007; Mogensen and H, 2009; Takeuchi and Akira, 2010).

PRRs are expressed mainly in immune cells, e.g. especially on antigen presenting cells (APC): monocytes, macrophages, dendritic cells (DC), but also non-immune epithelial cell

(Kumar et al., 2009a; Mogensen and H, 2009). After binding to MAMPs or DAMPs, a downstream signalling cascade triggers gene expression of pro-inflammatory signalling molecules such as specific cytokines, chemokines or lymphokines that mediate the pro-inflammatory immune response (Kawai and Akira, 2011, 2010b). Besides initiating and triggering primary immune response and co-activation of adaptive immune response, PRRs are also involved in later phases of infection mediating immune response. Hence, they make a bridge between innate and adaptive immunity in vertebrates (Iwasaki and Medzhitov, 2004; Kawai and Akira, 2011).

1.5 Toll-like receptors (TLRs)

The most well known and best understood family of PRRs are Toll-like receptors (Kawai and Akira, 2011; Kumar et al., 2009a). They were named according to the Toll receptor of fruit fly that playes an important role for the establishment of dorzo-ventral axis in developing embryo (Medzhitov et al., 1997). In case of suppressor mutation, young fruit flies tend to move in confused directions (the German researchers who discovered this pattern are said to exclaim "Das war ja Toll!" and so the newly discovered receptors were named as Toll-like receptors (Hansson and Edfeldt, 2005)). Later in ontogeny Drosophila´s TLRs detect fungal MAMPs - manans and protect against fungal infection. First mammalian TLR homolog was identified in 1997 (Medzhitov et al., 1997) and since then TLRs have been reported also in other vertebrates, tunicates, urochordates, crustaceans (in sensu stricto) and insects (Vinkler and Albrecht, 2009). Some TLRs are common for all taxa, other are rather specific for particular groups. In vertebrates, more than 12 TLRs belonging to six development groups (TLR1, TLR3, TLR4, TLR5, TLR7 and TLR11 (Roach et al., 2005) have been identified so far.

TLRs are transmembrane glycoproteins with a horseshoe-like shaped structure expressed either into cytoplasm membrane or endosomal membrane (I Botos et al., 2011). Similarly to other PRRs, they exist also in soluble form in cytoplasm, e.g. TLR2 in human (LeBouder et al., 2003). The localization of TLRs is important for the detection of the ligands. Nucleic acids of viruses and bacteria are recognized mainly by TLR3, TLR7, TLR8 and TLR9 which are anchored in endosomal membrane. On the contrary TLRs sensing ligands in extracellular space (LPS, peptidoglycan, flagellin) are located in outer layer of cytoplasm membrane (Kawai and Akira, 2010b; Kumar et al., 2009a). Typical TLR consists of N-terminal binding ectodomain, transmembrane hydrophobic alpha helix and C-terminal signalling domain. Extracytosolic or extraendosomal N-terminal domain is composed of approximately 16-28 leucine-rich repeats region (LRRs) (Istvan Botos et al., 2011) . Each LRR contains around 20-30 amino acids in a well conserved motif LxxLxLxxN (Kumar et al., 2009a). To be able to bind broader spectrum of ligands TLRs dimerize: some TLRs form homodimers

(e.g. TLR3 and TLR4), where other heterodimerize (e.g. TLR1/ TLR2 or TLR1/ TLR6). Besides that, correctors are also involved in ligand binding and form an initiation complex, e.g. MD2 and CD14 in TLR4 (Kawai and Akira, 2010b; Kumar et al., 2009a) (Chapter 1.5.1). Main function of transmembrane domain is anchored whole protein via hydrophobic interactions in a membrane. C-terminal domain is known as Toll/IL-1 receptor domain (TIR domain) according to homology with IL-1 receptor. This domain associates with adaptor proteins and it is essential to triggering downstream signalization cascade (I Botos et al., 2011; Kawai and Akira, 2010b).

When MAMPs or DAMPs are bound to extracellular domain of TLR directly or via specific corrector molecules, TLR molecule approches each other and they form homdimers or heterodimers. Then TIR domain recruits specific adaptor protein also containing the TIR domain depending on the type of TLR (MyD88, TIRAP, TRIF, TRAM) (Akira and Kiyoshi, 2004). The cascade follows either by MyD88-dependent pathway for all TLRs except TLR3 or by TRIF-dependent pathway for TLR3 and alternatively also for TLR4. Then activated transcriptional factors, e.g. NF-κB drive gene expression of pro-inflammatory cytokines, chemokines, viral interferons type I which activate humoral immunity of adaptive immune response especially to Th1 or Th17 types (Kawai and Akira, 2011; Kumar et al., 2009a)

From TLR family I decided to study TLR4 and TLR5 since they are well characterized in birds from previous study (Vinkler et al., 2014, 2009), the both recongnized bacterial ligands and play important role in innate immunity as well as they mediate immune response in later phases (Kawai and Akira, 2011). Furthermore, they are well variable on both intra- and inter-specific level in natural populations and their polymorphism is associted with resistance or suspectibility to disseses in human and also in animals (Chapter 1.5.3.).

1.5.1 Structure and function of TLR4

TLR4 is located in outer layer of cytoplasm membrane and its ectodomain binds ligands in extracellular space (Kumar et al., 2009a; Park et al., 2009). It detects a broad spectrum of ligands ranging from bacterial MAMPs (lipopolysaccharides of gram-negative bacteria), fungal (mannans and glucuronoxylomannans), protozoal (GPI anchors and glycoinositolphospholipids) and viral (F-protein of respiratory syncytial virus (RSV) and envelope protein of mouse mammary tumour virus (MMTV)) to DAMPs (e.g. heat shock proteins) (Miller et al., 2005; Uematsu and Akira, 2008). According to X-ray crystallographic structure described in human and mouse (Kim et al., 2007; Park et al., 2009). TLR4 has conventional structure as other TLRs: N-terminal exodomain with LRR, transmembrane domain and signalization TIR domain. N-terminal exodomain consists of three sub-domains:

N-terminal, central and C-terminal part. Whereas in mammals signalization via TLR4 is well described for LPS induced response (Lu et al., 2008) a cascade from binding LPS to expression pro-inflammatory response, in birds it remains to be revealed in details; though, regarding to the evolutionary conservativity in TLR we might predict similar complexed pathway. First, released and partially denatured LPS is captured by soluble Lipopolysaccharide-binding protein (LBP) and transported to the host cell. Here, the complex LPS-LBP is recognized by CD14 (Cluster of Differentiation 14) (Park et al., 2009). Then LPS is directly transferred to MD2molecule which forms a big hydrophobic pocket. Simultaneously, with LPS binding to MD-2 the whole complex is bound to TLR4 which is followed by dimerization of TLR4-MD-2-LPS complexes (Park et al., 2009). MD-2 dimerization binding sites are located in the concave surface of the terminal N- and central subdomains. Binding of ligand causes that the receptor brings two TIR domains into close proximity (Lu et al., 2008; Park et al., 2009). They recruit adaptor proteins, the most often MyD88 to them, and thus downstream signalization pathway is initiated (Lu et al., 2008). Terminating by an activation of the transcriptional factor NF-κB leads to gene expression of pro-inflammatory mediators such as interleukin-1-beta (IL1 β), IL-6, IL-12, IL-18, tumour necrosis factor- α (TNFα), INF-γ, chemokines (CCL2, CXCL8) and others (Kogut et al., 2005). To simplify, typical Th1 response ends up with an activation of macrophages and their oxidative burst accompanied with nitric oxide (NO) and reactive oxygen species (ROS) and INF- γ production. Of course, gene expression of molecules mentioned above is not driven exclusively by TLR4, but several bacterial sensing receptors might co-activate their expression as well via NF-κB and other transcriptional factors (e.g. Vazquez-Torres et al., 2004) and immune response is therefore regulated by an integration of different signals. of Apart from initiation of primary immune response, TLR4 co-activates also adaptive immunity (e.g. Vazquez-Torres et al., 2004).

1.5.2 Structure and function of TLR5

As TLR4, TLR5 also forms horseshoe-like structure with binding N-terminal exodomain, integral transmembrane domain and signalization TIR domain (Park et al., 2009). It binds flagellin from flagellated Gram-negative and Gram-positive bacteria directly in extracellular space (Hayashi et al., 2001). TLR5 is highly expressed in gut, mainly in lamina propria dendritic cells (DC), where it controls gut microbiota composition (Botos et al., 2011). Contrary to binding of extracullar flagellin, similar function in cytosol is fulfilled by NOD-receptor Ipaf, which binds also virulence factor of bacteria, and both receptors can have synergic effect on pro-inflammatory immune response (Miao et al., 2007). Since flagellin is recognized in conserved site situated in D1-domain, a part which is normally buried in native polymerized fibber, flagellin must first depolymerize to flagellin monomers

to be uncovered for binding to TLR5. Besides that, flagellin monomers can be sloughed from intact flagella (Miao et al., 2007). Mutational analysis in human and mouse has shown that flagellin-binding sites are on TLR5 located in central part of the exodomain – in highly conserved concavity formed by β sheets on one face of the LRR structure (Andersen-Nissen et al., 2007). Similarly to TLR4, TLR5 also forms hodimers but probably before ligand binding. Then MyD88 dependent signalization pathway is initiated leading to production of pro-inflammatory molecules such as IL-6, IL-12, TNF- α (Miao et al., 2007). Dimerization sites lie also in central part of exodomian – most probably on lateral patch (Andersen-Nisssen et al., 2007). However, later analysis of binding sites in *Danio rerio* showed that a lot of functionally important residues do not overap with those predicted for mammals (Yoon et al., 2013). Apart from an activation of mainly pro-inflamantory Th1 response, flagellin activates also TLR5 present on natural $CD4^+/CD25^+$ T-regulatory cells leading to increased suppressive activity, suggesting that TLR5 (flagellin) has a complex role in bridging innate immunity and adaptive immunity (Steiner, 2007).

1.5.3 Evolutionary perspective of polymorphism in TLRs

In TLRs our current knowledge about the importance of polymorphism at intra- and interspecies level for susceptibility or resistance to diseases are still limited, in particular in comparison with adaptive immunity genes (MHC) (Vinkler and Albrecht, 2011). In view of the fact that TLRs are directly exposed to parasites' molecules, their importance could be comparable with MHC (Vinkler and Albrecht, 2009). As pointed out by Acevedo-Whitehouse and Cunningham (2006), more than half of variation explaining resistance to diseases in immune genes cannot be attributed to MHC. There is a need to look for other candidate genes. Therefore, polymorphism in TLRs might explain substantial portion of resistance to diverse spectrum of diseases (Vinkler and Albrecht, 2009).

Specific alleles of TLRs are associated with susceptibility or resistance to infectious and autoimmune disease. In human particular *TLR4* alleles are associated with susceptibility to malaria (Eriksson et al., 2014), RSV (Puthothu et al., 2006) or infections caused by Gram-negative bacteria (Jana et al., 2016). In TLR2 susceptibility to leprosy, tuberculosis, staphylococci infections or resistance to Lyme disease were reported (reviewed in Mogensen, 2009)). Particular residues of TLR4 are also linked to suspectibility to Salmonella infection in chicken (Leveque et al., 2003). However, there is still lack of studies focused on assessing polymorphism in natural populations. In contrast to human or domestic and laboratory animals on which most studies regarding TLR polymorhism were performed (Abel et al., 2002; Mucha et al., 2009; Swiderek et al., 2006), wild animals are considered to be exposed

to much stronger parasite-mediated selection and therefore information about polymorhism from non-natural populations are of limited value for evolutionary and ecological research.

TLRs are evolutionary well conserved molecules in their tertiary and quaternary structure in vertebrates (Roach et al., 2005). Most of sites are therefore functionally constrained under the influence of purifying selection (Fornůsková et al., 2013; Grueber et al., 2014; Wlasiuk and Nachman, 2010). However, this is not completely truth for regions with binding properties. Whereas particular positions for binding ligands are well conserved even across phylogenetically distant taxa (Vinkler et al., 2014), other positions are under the influence of positive (diversifying) selection with frequent non-synonymous polymorphism (e.g substitutions which alter polarity or charge). Positive selection going on TLRs has been identified in different lineages of taxa in birds (Alcaide and Edwards, 2011; Vinkler et al., 2014), primates (Wlasiuk and Nachman, 2010), bats (Escalera-Zamudio et al., 2015) or rodents (Fornůsková et al., 2013). Bacterial sensing TLRs also appear to be more variable, probably reflecting higher structural variation of ligands in comparison to RNA or DNA sensing TLRs (Vinkler et al. 2014). Besides positive selection shaping variability of TLRs, genetic drift also plays a crucial role and island bird populations were identified to suffer by bottle-neck in TLRs showing that genetic-drift may prevail over selection, as shown in genus *Anthus* (Gonzalez-Quevedo et al., 2015) or in differently threatened New Zealand birds (Grueber et al., 2015).

There is still lack of studies which reported associations of TLRs in wild animals with resistance or susceptibility to pathogens. In rodents individuals carrying specific *TLR2* ($TLR2_{c2}$) haplotype in *Myodes glareolus* were almost three times less likely to be *Borrelia* infected compared to animals carrying other haplotypes (Tschirren et al., 2013). Furthermore, neutrality tests also confirmed that *TLR2* is under the influence of positive selection. In birds it has been recently reported that amino acid substitution Q549R in *TLR4* is associated with different responsiveness in skin-swelling after injection of LPS from *Escherichia coli* and *Salmonella enterica* in chicks of *Parus major* (Vinklerová 2013). Moreover, this substitution is also associated with plumage characteristics in adults (Bainová 2011). Given the fact that this substitution is also asscoiated with the width of the black melanin-based stripe and yellow carotenoid-based breast colouration in both sexes, it shows the influence of innate immunity on ornamental signalling and its role in sexual selection (Bainová 2011).

1.6 Paridae

Paridae is a family of small, conspicuous songbirds, well-conserved in their morphology. They are arboreal, familial (particularly occurring in flocks after breeding season) with diversified

vocal repertoires and feeding behaviour (e.g. food storing) (Cramp et al. 1993). An exception from the general tit-like appearance could only be represented by the *Pseudopodoces humilis* inhabiting Tibetan plateau which was misclassified as Corvids (James et al., 2003; Qu et al., 2013), *Melanochlora sultanea* and *Sylviparus modestus*. Members of the Paridae family are widely distributed in Northern Hemisphere and in Africa. English common names of the species do not correspond with phylogeny of the family. New World tits are called chickadees (genus *Poecile*) and titmice (genus *Baeolophus*). In Eurasia (Palearctic and Indomalaya) and in Africa (Afrotropic) members of the family are jointly named as tits. To simplify terminology and interpretation of the results, I decided to use the name tits for all members of Paridae family.

1.6.1 Phylogeny and biogeography

Because of their overall morphological similarities most of these species used to be originally assigned to the genus *Parus* (Gill et al., 2005; Slikas et al., 1996), while the other 8-9 genera included less species (Gill et al., 2005). However, molecular phylogenetics contributed to disentangling relatedness of the species and presently the family Paridae is considered to include approximately 55 species in 14 genera*: Cephalophyrus*, *Sylviparus*, *Melanochora*, *Pardaliparus*, *Periparus*, *Baeolophus*, *Lophophanes*, *Sittiparus*, *Poecile*, *Cyanistes*, *Pseudopodoces*, *Parus*, *Machlolophus*, *Melaniparus* (Johansson et al., 2013). Taxonomical status of some species and subspecies remains to be resolved and some authors consider particular subspecies as independent species while others do not, or it is rather a matter of debate underlying species concept definitions (e.g Päckert and Martens, 2008). For instance, even though the great tit species complex including *P. m. major, P. m. minor, P. m. bokharensis, P. m. cinereus* has genetic distances in mitochondrial control region (CR) well comparable with the genus *Poecile* (Kvist et al., 2003), Parus major is commonly considered as one species with several subspecies (Kvist et al., 2001; Päckert et al., 2005) rather than a complex of true species. Similar situation with uncertain species status is also the case of *Poecile palustris*/*montanus* complex, as indicated by Johansson et al. (2013). The same authors also call for extensive study to clarify relationships and species boundaries.

Old World Species

The sister group of Paridae is Remizidae and the common ancestor of both families is assumed to had inhabited tropical Africa and China (Tietze and Borthakur, 2012). However, tits likely originated in China where there is also the highest species diversity (Tietze and Borthakur, 2012). According to the most recent and the most comprehensive molecular phylogeny of tits (Johansson et al., 2013) based on one nuclear intron and one mitochondrial gene, the most

basal is *Cephalopyrus flamiceps.* It was separated from the rest of Paridae in China where *Sylviparus* was split off from the remaining larger-sized Parids towards Southeast Asia, followed also by larger body-sized *Melanochlora sultanea* (Tietze and Borthakur, 2012). Then tits evolved in a mediate-sized forms and radiated in China (or in Indohimalayan area) and in other Old World areas. The lineage of *Pseudopodoces* and formerly *Parus* (now also with African genera) radiated early in China and Afrotropics (Tietze and Borthakur, 2012). Within afrotropical monofyletic lineage of formerly *Parus*, genera *Melaniparus* and *Machlolophus* radiated in Africa in plenty of species. Sister lineage to this Africal tits includes *Parus major* complex with species occupying almost whole Eurasia (Tietze and Borthakur, 2012). The estimated time of divergence between *Parus major* and *Melaniparus afer* is 9-12 MYA (Packert et al., 2007). The crown group of this clade diversified in China evolving into several extant East and Southeast Asian species, e.g. well known *Parus monticolus* resembling *Parus major* in the appearance. The *Poecile* originated and diversified first in China, from where three species spread out to Europe and one lineage colonized North America followed by rapid radiation there (Gill et al. 2005). The Chinese most recent common ancestor (MRCA) of *Periparus* tit split approximately 11 MYA (Packert et al., 2007) into two lineages: relatively recently diversified Southeast lineage and Sino-Himalayan lineage, the latter *with Periparus ater* which colonized almost whole Eurasia as well as North Africa establishing many subspecies or species with uncertain species status (Pentzold et al., 2013). Genus *Cyanistes* diversified in Western Palearctic in three extant species around 3 millions of years ago (MYA) (Packert et al., 2007). It has been shown that *C. caerulus* is paraphyletic taxon with the European lineage (sister to *C. cyanus*) and Afrocanarian lineages which inhabit North Africa and Canarian islands. Canarian islands were colonized at lest twice in spite of general less dispersal status of tits. *Cyanistes* had to disperse from Canarian islands after the climate change. They had to overcome more than 100 km from African coasts to recolonize the islands when climate improved substantially (Gohli et al., 2015).

New World species

Tits colonised North America in two independent colonization events in the Late Tertiary (Gill et al., 2005). Regarding the fact that they are rather small or middle-sized arboreal species, no trans-ocean migration (trans-Atlantic or trans-Indian ocean) is expected to occur as in the case of some *Turdus* species (Voelker et al., 2009). Therefore, the colonization of North America probably occurred in the time of land bridges (Tietze and Borthakur, 2012). It was hypothesised that the common ancestors of modern titmice (*Baeolophus*) colonized North America ~ 4 MYA from the presumed sister group of Old Word species *Lophophanes* successively splitting off from four extant *Baelophus* species. However, phylogenetic

relationship among *Poecile*, *Lophophanes* and *Baeolophus* and therefore the Old Word ancestor of *Baeolophus* is still matter of debate. The ancestor of all North American chickadees (genus *Poecile*) colonized North America ~ 3.5 MYA from sister species to both *Poecile montanus* and *Poecile palustris*. Tietze and Borthakur (2012) suggested that supposed time of colonization is older, before Bering strait opened (5,5 MYA). According to another estimates based on different rate of molecular clock, split in *Poecile* between Eurasian lineage including *P. montanus* and *P. palustris* and American lineage occurred approximately 8 MYA (Packert et al., 2007). Considering the placement of *P. cinctus* into North American *Poecile* clade, it was hypothesised that *P. cinctus* colonized Eurasia back from North America in Pleistocene (Gill et al., 2005; Tietze and Borthakur, 2012). This species lives in a wide boreal area ranging from East Asia to Northern Europe and small population in Alaska. The monophyly of North American *Poecile* is strongly supported (Gill et al., 2005), though relationships among New World *Poecile* are complicated. Diversification of New World chickadees occurred relatively fast and the species might have relative high population size so that extensive ancestral polymorphism complicates the phylogeny (Harris et al., 2014). Besides that, hybridisation also might have played a significant role in the past (Curry, 2005; Gill et al., 1999), alongside with areal constriction and expansion in glacial cycles (Burg et al., 2006; Lovette, 2005). Only two clades within this group receive strong support; the "brown-capped clade", containing *Poecile cinctus*, *Poecile hudsonicus*, and *Poecile rufescens*, and the second group with *Poecile atricapillus* and *Poecile gambeli.* Relationships between these two taxa and *Poecile carolinensis* and *Poecile sclateri* remain unresolved (Johansson et al., 2013).

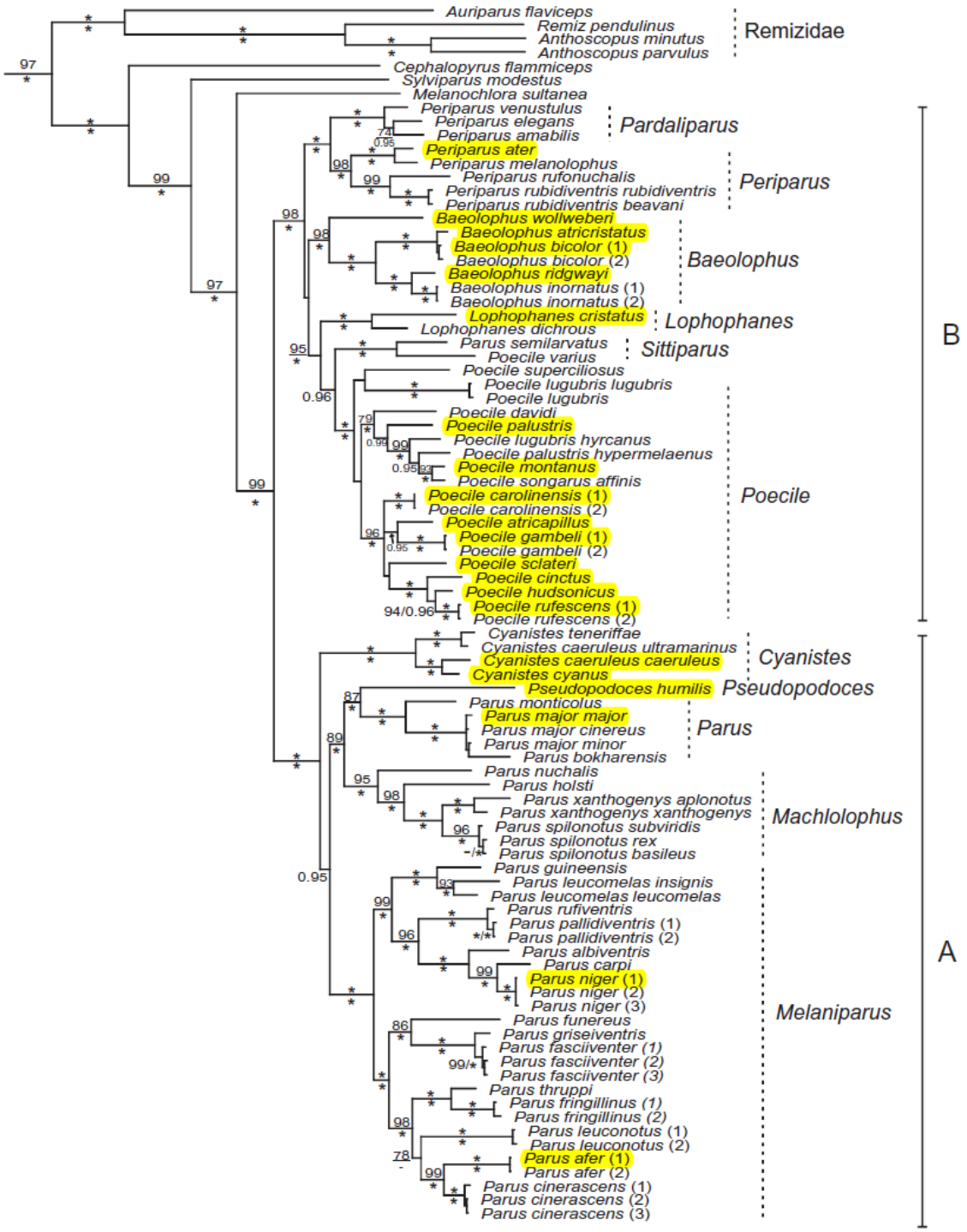


Figure 3: The phylogeny of tits from Johansson et al., (2013) based on two nuclear introns and one mitochondrial genes with highlighted species included in this study

1.6.2 Hybridization in Paridae

For a very long time hybridization has been considered as very rare or unlikely in animals leading to chromosomal imbalance and infertility of hybrids (Hedrick, 2013). Thus, zoologist

did not take hybridization into account as a plausible evolutionary mechanism for the origin of new species. However, later it has been shown that more than 10% of avian species hybridize (Grant and Grant, 1992). In addition to that, many hybrids can remain unnoticed since as a rule of thumb, distinctness of hybrids seems to be a function of the plumage differences between the hybridizing species (Randler, 2004). Despite the fact that hybridization often occurs, it does not necessarily result in the gene flow between species due to an existence of post-zygotic reproductive barriers, e.g. hybrids may be sterile or less viable (Dowling and Secor, 1997). Nevertheless, very rare hybridization events occuring between low fitess hybrid and one of the parental species may be enough to pass advantageous alleles (Hedrick, 2013).

Hybridization in Paridae is quite common phenomenon, particularly among closely related species within genera. Curry et all. (2007) and Mallet (2005) supposed that hybridization occurs among 25% of Paridae. The hybridization may also occur between higher taxonomical units (above the genus level) but little is known about viability or fertility of such hybrids, such reports may appear to be even anecdotal (Randler, 2004). One can distinguish very rare events among species leading to reporting hybrids rather as rarities, such as between *Parus major* and *Cyanistes caeruleus,* which occasionally form mixed pairs barely leading to hybrid offspring (Samplonius and Both, 2014), and in other cases (summarized in Table 1). All these interbreedings are very rare and reported only once or a few times maybe with exception of *P. montanus* and *P. palustris*, where frequency of hybrids can be underestimated because of very similar appearance of both species (Curry, 2005). In spite of the fact that hybridization is relatively common both between closely related species and between more diverged species, it remains to be resolved if it does result in gene flows since little is currently known about sterility such hybrids.

On the other hand, there are also several pairs of species which extensively hybridize in the contact hybrid zones. One of them, the tension contact zone between parapatrically distributed *Poecile atricapillus* and *P. carolinensis* is well studied in the USA (Bronson et al., 2005; C. M. Curry and Patten, 2014; Reudink et al., 2007). *P. atricapillus* inhabits most of Canada and northern half of the United States coming in contact in a narrow contact zone extending from Texas to New Jersey with *P. carolinensis* spread from the higher elevation of the Appalachian Mountains to southwestern Virginia. Contact zone has been recently moving northward being driven probably by behavioural male dominance of smaller *P. carolinensis* over *P. atricapillus* and mate choice by females (Bronson et al., 2003). In the centre of the hybrid zone hybrid pairs have lower breeding success in terms of less hatched eggs and lower fitness of offspring (Curry, 2005), and sex ratio is biased to males

which is in concordance with Haldane rules (Curry, 2005). The offspring of hybrid origin were viable but less fertile in subsequent generations (Curry, 2005).

Occasional mixing of *P. atricapillus* (lower altitudes with mixed forests) and *P. gambeli* (higher altitudes with dry coniferous forests) at an altitudinal interface represents an example of mosaic hybrid zone (Grava et al., 2012). However, latest research showed that hybridization of both species might be forced by forestry which makes mosaics of coniferous and deciduous forest in Canada where both closely related species live in sympatry. Besides that, it was suggested that hybridization may result from males of the *P. gambeli* having lower expression of a preferred trait (here dominance in behaviour) than the *P. atricapillus* (Grava et al., 2012). In spite of the fact that nestlings of hybrid origin have been genetically reported in mitochondrial DNA (mtDNA) as mountain chickadee and mixed in microsatellite loci, it remains to resolve if they are fertile (Grava et al., 2012). Hybridization occurs also between other pairs of North American *Poecile*, e.g. *P. gambeli* and *P. carolinensis*, *Poecile cinctus* and *Poecile hudsonicus* or after secondary contact between *Poecile cinctus* and *P. montanus* (review in Graves, 2008), but little endeavour has been paid to assess how common these hybridization events are (Curry, 2005). Apart from chickadees, hybridization in a hybrid zone is common also for recently diverged titmice where *Baeolophus atricristatus* and *B. bicolor* hybridize extensively within a narrow zone in Texas and southwestern Oklahoma. In Texas, hybridization has been occurring for several thousand years, while evidence suggests that the southwestern Oklahoma contact is more recent, stemming in the last century (Curry and Patten, 2014).

Whereas in North America most of the species live rather in allopatry or in parapatry than in sympatry with small overlapping areas where hybridization takes place, in Europe up to six or seven species live in sympatry without frequent hybridization (Dhondt, 2014). It may mean that mechanisms to prevent hybridization within a contact zone have had less opportunity to evolve in North America in comparison with European tits which diverged relatively long time ago (Gill et al., 2005; Päckert et al., 2007) and thus they are well ecologically adapted for different niches (Curry, 2005). An exception from that could be the hybridization between *Cyanistes caeruleus* and *C. cyanus* which has been known particularly from North-western part of European Russia and from Belarus where hybrids with plumage characteristics ranking from almost pure *Cyanistes caeruleus* to pure *C. cyanus* have been identified (Ławicki, 2012). These hybrids were even named Pleske's Tit (Cramp et al. 1993). Molecular analysis of blue tit species complex has shown that *Cyanistes caeruleus* in traditional point of view is parafyletic with afrocanarian lineage and basal Euroasian lineage where C. cyanus is sister to *Cyanistes caeruleus* in Europe (Gohli et al., 2015; Salzburger et al., 2002).

Table 1: Reported hybrids in tits

Species 1	Species 2	Within genera	Frequency	References	Note
Baeolophus atricristauts	***Baeolophus bicolor***	**yes**	**common**	**[1], [2], [13]**	**two narrow hybrid zones exist in Texas and Oklahoma, hybrids genetically reported**
Baeolophus bicolor	*Poecile gambeli*	no	rarity	[1], [2]	
Baeolophus bicolor	*Poecile atricapillus*	no	rarity	[1], [2]	records older than 1900
Baeolophus inornatus	***Baeolophus rigdwai***	**yes**	**common**	**[1], [2], [12]**	**contact zone in California, hybrids genetically reported**
Cyanistes caerulus	***Cyanistes cyanus***	**yes**	**common**	**[1], [2], [11]**	**hybrid zone, probably fertile hybrids, suspected F2 hybrids occasionally caught, hybrid individuals called Pleske's tit**
Cyanistes caerulus	*Parus major*	no	rarity	[1], [2]	occasionally forming mixed pairs, offspring are probably rather from extra-pair copulation
Cyanistes caerulus	*Poecile palustris*	yes	rarity	[2]	one report older than 1900
Parus major comlex		**yes**	**common**	**[1], [2], [10]**	**occurring in several contact zones between different subspecies (e.g. *P. m. cinereus* x *P. m. bokharensis*, *P. m. cinereus* x *P. m. minor*, *P. m. minor* x *P. m. major*)**
Periparus ater	*Parus major*	no	rarity	[2], [3]	
Periparus ater	Lophophanes cristatus	no	rarity	[2], [3]	one report
Periparus ater	*Poecile montanus*	no	rarity	[1], [2]	
Periparus ater	*Poecile palustris*	no	rarity	[2], [3]	
Poecile atricapillus	***Poecile caroliensis***	**yes**	**common**	**[1], [2], [8], [9]**	**very intensively studied, long contact zone, hybrids fertile, but with lower fitness**
Poecile atricapillus	***Poecile gambeli***	**yes**	**less common**	**[1], [2], [7]**	**mosaic hybrid zones, hybridization occur over a broad geographic region, hybrids genetically reported**
Poecile atricapillus	*Poecile hudsonicus*	yes	occurring	[1], [6]	genetically confirmed, probably F1 hybrids
Poecile caroliensis	*Poecile gambeli*	yes	probably occurring	[1], [2]	
Poecile cinctus	*Poecile hudsonicus*	yes	probably occurring	[1], [2]	on the basis of similar appearance extensive hybridization is expected
Poecile cinctus	*Poecile montanus*	yes	rare	[1], [2], [3],	mixed pairs not rare
Poecile montanus	*Parus major*	no	rarity	[2]	one report older than 1900
Poecile montanus	*Cyanistes cyanus*	yes	rarity	[2]	
Poecile montanus	*Poecile palustris*	yes	probably very rare	[1], [2]	
Poecile montanus	Lophophanes cristatus	yes	rarity	[1], [2]	
Poecile palustris	*Parus major*	no	rarity	[1], [2]	
Poecile palustris	Lophophanes cristatus	no	rarity	[1], [2]	one record older than 1900

Species in which hybridization is more common are higlighted in bold. References: [1] McCarthy, (2006); [2] http://www.bird-hybrids.com/ [10/8/2016]; [3] Gosler and Clement (2007); [4] Randler, (2002), [5] Jarvinen, (1987); [6] (Lait et al., 2012); [7] Grava et al., (2012); [8] Curry, (2005); [9] Reudink et al., (2007); [10] Kvist et al., (2003); [11] Ławicki, (2012); [12] Source et al., (2004); [13] Curry and Patten, (2014)

2 Aims and hypotheses

1. **To describe intra- and interspecific polymorphism in *TLR4*, *TLR5* and in six neutral autosomal markers in 20 tit species**

   TLR4 and *TLR5* have higher sequence variation on both inter- and intraspecific level compared to neutral sequences.

2. **To identify signatures of positive selection in *TLR4* and *TLR5* genes**

   TLRs are immune genes which directly interact with parasitic and pathogenic ligands. Based on assumptions of the *Red Queen hypothesis* (van Valen, 1973) we may expect positive selection operating on particular residues in these genes. The positive selection manifests on molecular level by increased ratio of non-synonymous to synonymous substitutions (d*N*/d*S* ratio).

3. **To identify positively selected residues which may affect binding properties of the TLRs**

   We hypothesise that selected positions might have a functional importance for binding ligands since they may be located in close proximity to functionally important binding sites. Based on knowledge of functionally important sites predicted for mammals in TLR4 (Kim et al., 2007; Park et al., 2009) and for both mammals (Andersen-Nissen et al., 2007) and fish in TLR5 (Yoon et al., 2013) we may suppose that some positively selected residues lie in close proximity to functional binding sites.

4. **To investigate TSP in TLR4 and TLR5 genes and to distinguish it from other mechanisms leading to shared polymorphism**

   Several mechanisms (inherited polymorphism leading to TSP, introgression, convergence) may be responsible for shared polymorphism in immune genes (Hedrick, 2013). We hypothesise that TSP may be the most common phenomenon explaining shared polymorphism in related species. TSP and shared polymorphism in general should be more common in TLRs than in neutral markers since TLRs may be under the influence of positive and balancing selection.

5. **To detect gene flow and introgression**

Hybridization in Paradiae is common and may involve up 25% tit species (Curry, 2005). However, less is known about viability such hybrids and potential introgression. We hypothesise that gene flow (introgrssion) occurs in closely related species and therefore introgression may be responsible for the origin of shared polymorphism in species that hybridize.

3 Methods

3.1 Tissue samples

Dataset of 192 individuals of 20 tit species from Palearctic, Neoarctic and South Africa were gathered. We included approximately 10 individuals per species where possible (Table 2). Only in great tit (*Parus major*) there were 25 individuals since this species inhabits a large area spanning from Western Europe to East Asia with many subspecies. The samples included in the dataset were selected considering the following criteria: to have representatives throughout tit phylogeny (Figure 3), to have representative sampling across their whole area of distribution and only nonrelated individuals were chosen. Since it would have been difficult to personally collect these samples in the field, most samples were gained from genetic banks.

172 genetic samples were gained from Genetic Resources Collection (GRC) at the Burke Museum of Natural History and Culture, University of Washington (http://www.burkemuseum.org/research-and-collections/genetic-resources), mainly non-European species from North America, South Africa, but also Eurasian species with sampling outside Europe and 5 samples were gained from Genetic bank of the Department of Zoology, Charles University in Prague (https://www.natur.cuni.cz/biology/zoology/geneticka-banka). Besides that, other 12 individuals of the six European species were caught into mist nests according to standard protocol given by the Czech Ringing Centre of National Museum in Prague in different parts of the Czech Republic in post breeding season (July, August, September) in 2013 and 2014. After capture small volume of blood (approximately 100 µl) was taken by jugular venipuncture, the samples were stored in ethanol in freezer in -20°C.

Table 2: The list including investigated species with their sample size, area of distribution and the locality, where these individuals were sampled

Scientific name	English name	Abbreviation	Number of individuals	Distribution	Locality
Baeolophus atricristatus	Black-crested titmouse	BaAt	3	South USA, Mexico	USA - Texas
Baeolophus bicolor	Tufted titmouse	BaBi	10	Southeast USA	USA - Virginia, North Carolina
Baeolophus ridgwayi	Juniper titmouse	BaRi	6	Southwest USA, Mexico	USA - Arizona, Nevada, New Mexico
Baeolophus wollweberi	Bridled titmouse	BaWo	6	Southwest USA, Mexico	USA - Arizona
Cyanistes caeruleus	Eurasian blue tit	CyCa	12	West Eurasia	CZ, GER, LITH, RUS
Cyanistes cyanus	Azure tit	CyCy	10	Eurasia	MGL, RUS
Parus major	Great tit	PaMa	25	Eurasia	CZ, GER, LITH, NOR, RUS, KZ
Melaniparus niger	Southern black tit	MeNi	2	South Africa	JAR
Melaniparus afer	Grey tit	MeAf	5	South Africa	JAR
Periparus ater	Coal tit	PeAt	12	Eurasia	CZ, MGL, RUS
Poecile atricapillus	Black-capped chickadee	PoAt	9	USA, Canada	USA
Poecile carolinensis	Carolina chickadee	PoCa	10	Southwest USA	USA - North Carolina,Luisiana
Poecile cinctus	Siberian ti	PoCi	10	Eurasia	RUS
Poecile gambeli	Mountain chickadee	PoGa	9	USA	USA
Poecile hudsonicus	Boreal chickadee	PoHu	10	South USA, Canada	USA - Alaska, Newfounland, Washington
Poecile montanus	Willow tit	PoMo	14	Eurasia	CZ, LITH, MGL, RUS
Poecile rufescens	Chestnut-backed chickadee	PoRu	10	West USA	USA - Alaska, Oregon, Washington
Poecile sclateri	Mexican chickadee	PoSc	6	Mexico	MEX, USA - Arizona
Poecile palustris	Marsh tit	PoPa	12	Eurasia	CZ, RUS
Lophophanes cristatus	European crested tit	LoCr	12	West Eurasia	CZ, RUS

3.2 Molecular-genetics analysis

3.2.1 DNA extraction, primer design, PCR optimization

DNA was extracted from blood and other tissues (muscle, bone, skin) by using Quiagen DNeasy Blood & Tissue Kit and Quiagen DNeasy 96 Blood & Tissue Kit (spin column and plate kit, Quiagen 2006). These samples were stored in ethanol and freezed in -20°C. Primers were designed using Oligoanylzer web tool (version 3.1, http://eu.idtdna.com/calc/analyzer) (Owczarzy et al., 2008) and synthesised by Generi Biotech company (http://www.generi-biotech.com/homepage-generi-biotech/). For both *TRL4* and *TLR5* genes primers were designed to cover the whole ligand binding regions, i.e. partial exon 3 in *TLR4* and partial exon 1 in *TLR5* (Bainová 2011, Andersen-Nissen et al., 2007; Fitzgerald et al., 2004). For autosomal neutral markers we used six primer sets adopted from avian reference genomic markers set (Backström et al., 2008). However, these markers were adjusted according to reference genomic sequence of *Pseudopodoces humilis* (NCBI *Pseudopodoces humilis* annotation Release 101 and Table 3). These primers were located either on different macrochromozomes or in long distance from each other allowing free recombination. They were designed into more conservative exon regions, which were surrounded by more variable intron sequences. As a result, the intron sequences were thus mainly sequenced (in total 450-490 bp).

During Polymerase chain reaction (PCR) condition optimization different types of polymerase were used: FastStart Taq DNA polymerase (Roche), HotStart Taq DNA polymerase (Quiagen) and HotStart Taq plus DNA polymerase (Quiagen). Successful PCR amplification was checked by agarose gel electrophoresis with Goldview as a DNA-labelling dye (mostly 1,5% agarose gel, 100 V for 20 min). For *TLR4* and *TLR5*, we first sequenced whole exon 1 and exon 3 sequences applying the primer sets from previous research (Bainova, 2011; Bainova et al., 2014) in four phylogenetically distant tit species (*Parus major*, *Cyanistes caeruleus*, *Poecile palustris* and *Periparus ater*) by using Sanger sequencing. Amplified PCR products were purified using Exo-CIP PCR clean-up protocol (0.05 µl Exo, 0.1 µl CIP and 1 µl ddH_2O per one reactions) and then labelled with sequencing primers using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosciences). Subsequently, these PCR products were sequenced using ABI 3730xl Genetic Analyzer (Applied Biosystems) at the External Research Facility Studenec of Institute of Vertebrate Biology, The Czech Academy of Science. These pilot sequences were analysed in Seqscape version 2.5 (Applied Biosystems) and BioEdit Alignment Editor version 7.2.5 (Hall, 1999). Based on the knowledge of these *TLR4* and *TLR5* sequences and other available sequences (in total, one sequence per species of following species was used: *Parus major*, *Cyanistes caeruleus*, *Poecile palustris*, *Poecile montanus*, *Lophophanes cristatus*, *Periparus ater*, *Psedopodeces humilis* and *Taenopygia guttata*) we then designed more specific primers for PCR product

of ca. 470-500 bp in length which evenly covered the binding sites with overlapping parts (ca. 150 -180 bp).

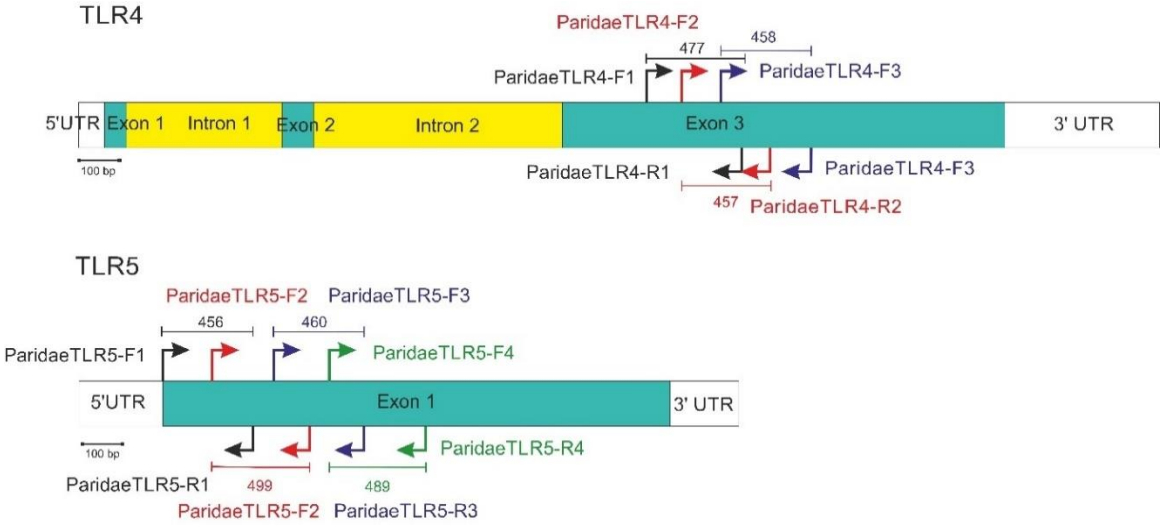


Figure 4: Schematic exon-intron structure of the *TLR4* and *TLR5* gene with highlighted sequenced range and primer positions

Exons are highlighted in blue, introns in yellow and UTRs in white. The arrows show primer positions in the sequenced exons. The numbering here is according to ground tit sequences.

Table 3: Autosomal neutral markers

The numbering of loci is adopted from Backström et al. (2008), length of sequenced region is according to ground tit, both chromosome number and marker chromosome position (total length of marker) are according to Zebra finch due to insufficient annotation of Ground tit genome (NCBI *Pseudopodoces humilis* annotation Release 101). Ground tit sequences used for numbering (from GenBank): *DLD* GI539359180, *CHMP5* GI539359184, *TIAL* GI224381693, *MMAA* GI539359160, *DDB1* GI539359131 and *UCHL3* GI212551129. Zebra finch (ENSEMBL *Taeniopygia guttata* Release 3.2.4) contigs used for the chromosome position: (from ENSEMBL): DLD ENSTGUG00000003229, CHMP5 ENSTGUG00000008197, TIAL1 ENSTGUG00000011148, MMAA ENSTGUG00000002671, DDB1 ENSTGUG00000006414 and UCHL3 ENSTGUG00000012534.

Locus	Gene abbrev.	Gene	Length (bp)	Chromosome	Chromosome location
12884	*DLD*	dihydrolipoyl dehydrogenase, mitochondrial	493	1	13,941,333-13,954,207
27356	*UCH-L3*	ubiquitin carboxyl-terminal hydrolase isozyme L3	489	1	69,466,452-69,508,074
21491	*CHMP5*	putative SNF7 domain containing 2 variant 1	470	2	90,756,181-90,774,404
16214	*MMAA*	methylmalonic aciduria type A protein, mitochondrial	452	1	12,394,737-12,399,922
8352	*DDB1*	DNA damage-binding protein 1	485	5	6,699,379-6,711,034
15439	*TIAL1*	nucleolysin TIAR	461	6	30,596,342-30,606,479

Table 4: Summary of primers for *TLR4* and *TLR5* used for PCR amplification

Position of primers are given according to ground tit *TLR4* and *TLR5* coding sequences. The sequences were obtained from GenBank database: *TLR4* - GI539359149, *TLR5* - GI539359169.

Gene	Primer name	Primer sequence (5´-3´)	start	end	length
TLR4	ParidaeTLR4-F1	CAGGTCCGCTTTTGAGAACTTC	711	732	22
TLR4	ParidaeTLR4-R1	GCTGAAGGTGAGTCTATTCTC	1168	1188	21
TLR4	ParidaeTLR4-F2	GTCTTAATCTGCTTCAGGGAG	874	894	21
TLR4	ParidaeTLR4-R2	CCAAATAAAGTTGTGTGCTG	1312	1331	20
TLR4	ParidaeTLR4-F3	GTGCTCCGTATTACCAAGAAC	1081	1101	21
TLR4	ParidaeTLR4-R3	GCTTGAAATATCCAAGGTGTGG	1518	1539	22
TLR5	ParidaeTLR5-F1	ATGATGTTGTGCCATCAGCTCCTC	1	24	24
TLR5	ParidaeTLR5-R1	CCAATTCTTCTAATGACCTC	438	457	20
TLR5	ParidaeTLR5-F2	CTGTTACCATAGGAAAAGGAGCG	254	276	23
TLR5	ParidaeTLR5-R2	GGCTGTAGAGAGATACTGG	735	753	19
TLR5	ParidaeTLR5-F3	CCAATCTTACCAGCTTCCAAGG	569	590	22
TLR5	ParidaeTLR5-R3	GAGAGTTTTTAGGTTGCCCAAGCC	1006	1029	24
TLR5	ParidaeTLR5-F4	GCAGGACTAGGAAGAAGTAATC	853	874	22
TLR5	ParidaeTLR5-R4	GGAAAAGAATATACAGGTCACC	1321	1342	22

Table 5: Summary of primers for neutral markers used for PCR amplification

Position of primers are given according to the ground tit neutral markers sequences (already specified in Table 3).

Gene	Primer name	Primer sequence (5´-3´)	start	end	length
DLD	ParidaeDLD-F	AGATGATGGAACAGAAGAG	9911	9929	19
DLD	ParidaeDLD-R	GCTATGAGTATGTTCTTTG	10385	10403	19
UCH-L3	ParidaeUCH-L3-F	GCTTGTGGAACAATTGGG	13316	13333	18
UCH-L3	ParidaeUCH-L3-R	TATTTGGCCCTCTCTTCAGG	13785	13804	20
CHMP5	ParidaeCHMP5-F	AGTCGTAGCTATGGAACACC	7584	7603	20
CHMP5	ParidaeCHMP5-R	GTAGGAATTGTCTTCATCAGC	8033	8053	21
MMAA	ParidaeMMAA-F	GCATACATCAGGCCATCTCC	4688	4707	20
MMAA	ParidaeMMAA-R	TCAACCATATCAGCCACAGC	5120	5139	20
DDB1	ParidaeDDB1-F	CATGGTGTATCCCGAGGA	8783	8800	18
DDB1	ParidaeDDB1-R	TGGCTAACAGCTTCCCGTTG	9248	9267	20
TIAL1	ParidaeTIAL1-F	GCTATTGTACACATGGGAG	2641	2660	19
TIAL1	ParidaeTIAL1-R	GCAATTCCTCCACAGTACAC	2203	2222	20

3.2.2 Next Generation Sequencing (MiSeq Illumina)

Considering the high number of individuals and the expected variability in *TLRs* genes, we applied Next Generation Sequencing Miseq Illumina platform to avoid the need of cloning. It allows us to sequence all PCR products from all samples in one sequencing run.

Due to MiSeq Illumina chemistry all PCR products had to be no longer than 500 bp. As a consequnce, PCR products ranging from 450 to 500 bp were designed. Since the sequenced binding regions were much longer for both *TLRs*, final *TLRs* sequences were composed of three independent PCR products in *TLR4* and of four PCR products in *TLR5*. In contrast, each neutral marker was covered by only one PCR product. To reduce time and budget for preparation of sequencing library, we performed multiplex PCRs wherein we co-amplified several independent PCR products in several multiplex reactions. In total, for 13 PCR products we set up four independent multiplex reactions (Table 7) with following criteria: Only one PCR product per one gene was admisible for each multiplex.To avoid primer heterodimerization from different primer sets in the same multiplex reaction we evaluated the possibility of forming heterodimers by using Oligoanalyzer web tool (version 3.1, http://eu.idtdna.com/calc/analyzer) (Owczarzy et al., 2008) for all primer combinations. Success of PCR amplification (whether all PCR products were amplified and if any unspecific products occured) was checked in each multiplex by three independent ways: first, by melting curve analysis using LightCycler 480 (Roche) with DNA Binding Dye EvaGreen, where we compared melting curves of each multiplex set with "pooled" multiplex, i. e. all PCR products from one multiplex were amplified in independent PCRs and afterwards these PCR products were pooled together and the melting cuves were examined. Second, simultaneously by gel electrophoresis in 4% agarose gel running 24 hours where we loaded amplified multiplex PCR products and counted number of occurring bands. Since the number and length of our PCR products differ in each multiplex (Table 7), we supposed to count the number of occurring bands and reveal potential unspecifities. Whereas in melting curve analysis we confirmed a successful amplification in all four multiplex reactions, on gel electrophoresis the exact number of bands with amplified PCR products was barely distinguishable. In spite of the fact that we confirmed successful amplification of all PCR products by melting curve analysis to be sure that there were no unspecific products, we applied also Sanger sequencing. Final preparation of Miseq sequenced library consisted of two independent PCRs. In the first PCR, multiplex PCR was performed in 20 cycles (prior optimization in order to minimize the number of cycles was done) with specific MiSeq primers followed by purification using HighPrep™ PCR reagent (Macbio Genomics). These MiSeq specific primers were designed as prolonged previously optimized primers (Table 4 and Table 5) by identical 30 bp adaptor seqeunces which were different for forward: CTCTTTCCCTACACGACGCTCTTCCGATCT and for reverse primers: CTGGAGTTCAGACGTGTGCTCTTCCGATCT). The second PCR was perfomed in next 15 cycles with purified PCR products from the first PCR used here as templates and with specific indices and sequencing primers. I will describe both PCR steps

in more details in the following paragraphs (additionally, they are also summarized inTable 6).

The first PCR was perfomed by using QIAGEN Multiplex PCR *Plus* Kit (Quiagen Germany). It was done in the volume of 12 µl for each multiplex reaction, where for multiplexes 1-3 (see Table 7) 6 µl master mix, 1.44 µl 0.2 µM primers for each PCR product, 3.36 µl RNA free water and 1.2 µl gDNA were added into reaction mix. For multiplex 4 (see also Table 7) 6 µl master mix, 1.92 µl 0.2 µM primers for each PCR product, 2.88 µl RNA free water and 1.2 µl gDNA were added into reaction mix. Multiplex PCRs ran in termocyclers in 20 cycles with following parameters: initial denaturation in 95°C for 5 minutes, then in each cycle denaturation in 94°C for 30 seconds, primer annealing in 55°C for 75 seconds, extension in 72°C for 30 seconds and after that the final extension was done in 68°C for 10 minutes.

The second PCR was performed by using PCR Using NEBNext® High-Fidelity 2X PCR Master Mix (M0541) (Bioo Scietific) chemistry. It was performed in volume of 15 µl for each multiplex reaction (Table 5), where 5 µl ddH_2O, 1 µl PCR product, 7.5 µl master mix, 1 µl barcode primers which labelled each individual by a unique index (NEXTflex™ 16S V1-V3 Amplicon-Seq Kit) were added into reaction mix. The PCR ran in termocyclers in 15 cycles with following parameters: initial denaturation in 98°C for 5 minutes, then in each cycle denaturation in 98°C for 45 seconds, primer annealing in 65°C for 20 seconds, extension in 72°C for 30 seconds and after that final extension was done in 68°C for 10 minutes. The comparison of PCR conditions for both PCRs done for MiSeq run is further shown in Table 6. The second PCR and final run preparation was done by Hana Velová in European Molecular Biology Laboratory in Heidelberg where also final sequencing on MiSeq ILLUMINA platform was perfomed in collaboration with Dr. Vladimír Beneš.

Table 6: PCR conditions used for amplification in the first and the second PCR reaction

Different kits were used for each PCR. For first PCR QIAGEN Multiplex PCR Plus Kit and for the second PCR PCR Using NEBNext® High-Fidelity 2X PCR Master Mix (M0541); Bioo Scietific were used.

Step	First PCR	Second PCR
Initial PCR activation	95°C/ 5 min	98°C/ 5 min
Denaturation	94°C/ 30 sec	98°C/ 45 sec
Annealing	55°C/ 75 sec	65°C/ 20 sec
Extension	72°C/ 30 sec	72°C/ 30 sec
Number of cycles	20	15
Final extension	68°C/ 10 min	72°C/ 3 min

Table 7: Specific MiSeq Illumina primers used for PCR amplification, the composition of multiplex PCR reactions and basic properties of sequnced PCR products

Specific MiSeq Illumins primers are prolonged primers from Table 4 and Table 5 by specific 30 nucleotide Ilumina sequences: CTCTTTCCCTACACGACGCTCTTCCGATCT for forward and for reverse CTGGAGTTCAGACGTGTGCTCTTCCGATCT. The length in bp (Δ prod.) and CG content of PCR products are shown.

Multiplex	Gene	Primer name	Δ prod. [bp]	CG content [%]
1	TLR5	MiSeq-ParidaeTLR5-F1	456	45.5
	TLR5	MiSeq-ParidaeTLR5-R1		
1	TIAL1	MiSeq-ParidaeTIAL1-F	461	41.9
	TIAL1	MiSeq-ParidaeTIAL1-R		
1	TLR4	MiSeq-ParidaeTLR4-F1	477	47.9
	TLR4	MiSeq-ParidaeTLR4-R1		
2	MMAA	MiSeq-ParidaeMMAA-F	452	41.4
	MMAA	MiSeq-ParidaeMMAA-R		
2	TLR4	MiSeq-ParidaeTLR4-F2	457	42.4
	TLR4	MiSeq-ParidaeTLR4-R2		
2	TLR5	MiSeq-ParidaeTLR5-F3	460	36.2
	TLR5	MiSeq-ParidaeTLR5-R3		
3	UCH-L3	MiSeq-ParidaeUCH-L3-F	489	31.5
	UCH-L3	MiSeq-ParidaeUCH-L3-R		
3	DLD	MiSeq-ParidaeDLD-F	493	34.1
	DLD	MiSeq-ParidaeDLD-R		
3	TLR5	MiSeq-ParidaeTLR5-F2	499	41.8
	TLR5	MiSeq-ParidaeTLR5-R2		
4	TLR4	MiSeq-ParidaeTLR4-F3	458	40.5
	TLR4	MiSeq-ParidaeTLR4-R3		
4	CHMP5	MiSeq-ParidaeCHMP5-F	470	36
	CHMP5	MiSeq-ParidaeCHMP5-R		
4	DDB1	MiSeq-ParidaeDDB1-F	485	50.9
	DDB1	MiSeq-ParidaeDDB1-R		
4	TLR5	MiSeq-ParidaeTLR5-F4	489	36.5
	TLR5	MiSeq-ParidaeTLR5-R4		

3.3 Sequence data filtering in UNIX and in Geneious

At the beginning of sequence analysis in UNIX, raw sequences from MiSeq run (with already trimmed out barcode sequence) were first grouped to gene clusters and then to subclusters according to particular PCR products. After that only two most abundant sequences per PCR product (separately for forward and reverse sequences) per barcode were filtered out. Simultaneously, the quality of the filtered sequences based on Phred quality score was checked and the values of most of the sequences were over 30. Then MiSeq primer sequences

were trimmed from both forward and reverse sequence. Then in program Geneious all sequences were manually checked and only two most abundant PCR products (for heterozygote) or one (for homozygote) were selected. To distinguish true heterozygote alleles from incorrect alleles, as a rule of thumb less abundant alleles should not differ in their abundance (number of reads) more than by 1/3. The lower abundance of true alleles might be around this threshold particularly in cases where SNP in primer binding sites occured. Exceptionally, incorrect (chimeric) sequences had higher abundance in comparison with true alleles, however, these cases were revealed based on multiple sequence alignment (MSA). Only PCR products having at least 9-10 reads per individual were treated in subsequent analysis.

3.4 Allele composition and assessing genetic polymorphism

TLR4 and *TLR5* alleles were manually put together into contigs from three and four independent PCR products respectively (meaning in total six and eight forward and reverse sequences per one allele) according to SNPs in overlapping parts in program Geneious (version 9.0.5.). In the case of too low variability or too short overlapping parts which did not allow us compose the whole allele (contig) in some cases, we made a consensus sequence from all PCR products in Geneious. After that, MSA were performed for both resolved and unresolved alleles for each gene in Geneious. Alleles were furthered reconstructed by using PHASE algorithm implemented in program PHASE 2.1 (Stephens and Donnelly, 2003; Stephens et al., 2001). The analysis ran for each species separately with both resolved and unresolved alleles with following parameters: run = 5x with different seeds, burn-in = 1000, number of iterations = 10000 and model with recombination. The consistency of independent PHASE runs was checked in Geneoius, all runs were consistent. In contrast to *TLR4* and *TLR5*, neutral markers were composed of only one PCR product, i.e. from one forward and one reverse sequence; therefore, the number of non-decoded alleles was lower but they must be treated by PHASE algorithm as well. Interspecific and intraspecific single nucleotide polymorphism was also identified in Geneious. These SNPs positions were further visualised for *TLR4* and *TLR5* in FaBox web tool, version 1.35 (www.birc.au.dk/sofware/fabox) (Villesen, 2007) and basic population genetics characteristics were calculated for each gene (Chapter 3.5).

3.5 Population genetics characteristics for *TLR4*, *TLR5* and neutral markers

For both neutral markers and *TLRs* basic population genetics characteristics were calculated for each species in program DnaSP 5 (Librado and Rozas, 2009; Rozas, 2009). Prior the analysis, all INDELs mutations within species were excluded. Those parameters included

sequenced length, number of nucleotide haplotypes, nucleotide diversity per site (π), proportion of segregating sites per site (Waterson' s θ), divergence to outgroup as an average number of nucleotide differences per base (D_{xy}) and divergence to outgroup as an average total number of nucleotide substitutions (K). As an outgroup sequences of zebra finch (*Taeniopygia guttata*) were chosen. Those sequences were obtained from GenBank (*TLR4* – GI224381674, *TLR5* – GI224381689, *DDB1* – GI224381692, *DLD* – GI224381677, *CHMP5* – GI224381690, *MAMA* – GI224381690, *TIAL* – GI224381693 and *UCHLP3* – GI224381666). By applying four gamete tests for detecting recombination (Hudson and Kaplan, 1985) recombination parameter (R) and number of recombination events (R_m) were estimated. To find out if theses loci evolve under neutrality or under the influence of recent positive or balancing selection Tajima's D test (Tajima, 1989) and Fu and Li's test (Fu and Li, 1993) were performed. Tajima's D test is based on the comparison of π and θ where under neutrality both estimates are equal and thus Tajima's D is 0. Negative Tajima's D (excess of rare mutations, $\pi<\theta$) can indicate positive or negative selection acting on these loci selection or recent selective sweeps. Positive Tajima's D (excess of mutation with intermediate frequency, $\pi>\theta$) may indicate balancing selection (Tajima, 1989). Fu and Li test is based on the similar expectations but additionally it takes genealogy of alleles into account and compare the numbers of mutations in both internal and external branches of phylogenetic tree with the expectation of neutrality (Fu and Li, 1993).

3.6 Protein structure modelling

Three-dimensional structures of both *TLR4* and *TLR5* partial ectodomain with binding sites were modelled by using homolog modelling implemented in I-TASSER (http://zhanglab.ccmb.med.umich.edu/I-TASSER/) (Roy et al., 2010; Yang et al., 2015; Zhang, 2008). 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 (Roy et al., 2010). From top five predicted models the best models were selected by C-score a confidence score for estimating the quality of predicted models by I-TASSER. It is an integrative score based on the significance of threading template alignments and convergence parameters of the structure assembly simulations. C-score usually fall in the range of (-5; 2), where the higher value, the better model with higher confidence. For our selected models C-score ranged in the interval (-0,68; 0,59).

Only one sequence per species was selected for the protein modelling making in total 20 sequences for *TLR4* and 20 sequences for *TLR5*. To choose characteristic sequence from each species, the sequence with the highest frequency in population was included

(Supplement 6 for *TLR4* and Supplement 8 for *TLR5*, respectively). These datasets are consistent with those for analysis of the surface charge and for selection and recombination analyses on interspecies level.

3.7 Detection of recombination and positive selection on interspecies level

For both recombination and selection analyses the same datasets of 20 sequences were used (one sequence per species; see Supplement 5 and Supplement 7).

The recombination breakpoints were estimated in *TLR4* and *TLR5* gene by Rapid Screening for Recombination Using a Single Break Point (further SBP analysis) and Genetic Algorithm Recombination Detection (GARD; Pond et al., 2006) web tool on Adaptive evolution server (www.datamonkey.org) (Kosakovsky Pond and Frost, 2005).

The signature of long-term positive selection was estimated by applying four different methods which are based on the comparison of *dN/dS* ratio across the whole sequence (the number of nonsynonymous substitutions per nonsynonymous sites to the number of synonymous substitutions per synonymous sites). Three of these methods were done by using Adaptive evolution server (Kosakovsky et al., 2005): Random Effect Likelihood (REL analysis), A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection (FUBAR) (Murrell et al., 2013) and Mixed Effects Model of Evolution (MEME) (Murrell et al., 2012). Besides that, Phylogenetic Analysis by Maximum Likelihood (PAML) for inferring of positive selection implemented in program PAML, version 4.8 (Yang, 2007) was applied. The computation of PAML was performed on the computational server Xukol of the Department of Zoology, Faulty of Sciences, Charles University in Prague. For all these tests the positively selected sites were considered to be statistically significant at $p < 0.1$.

Finally, positively selected amino acid positions were visualised in three-dimensional structural protein model of TLR4 and TLR5 in program PyMOL, version 1.8 (Schrödinger, LLC, 2015). Here we highlighted also other variable sites and ligand-binding sites known in mammals for TLR4 and in mammals and fish in TLR5 (Supplement 9).

3.8 Analysis of evolutionary conservative and non-conservative sites in TLR4 and TLR5 (ConSurf)

The evolutionary conservativism of amino acid positions in *TLR4* and *TLR5* genes was predicted using the ConSurf tool (http://consurf.tau.ac.il/2016//overview.php) (Ashkenazy et al., 2010). We assumed that functionally important sites for binding pathogen might be the least conserved. For ConSurf analysis all obtained *TLR4* (380 sequences) and *TLR5* (368 sequences) sequences were used and 3D models of great tit TLR4 and TLR5 modelled by I-TASSER were included into the analysis (see Supplement 6 and Supplement 8). Bayesian

computational method was chosen for generation of the phylogenetic trees of sequences and the "best model of amino acid substitution" was chosen as a default setting. Based on MSA, 3D structure and phylogenetic tree ConSurf calculates the "Amino Acid Conservation Score" for each residue. This score is further normalized and thus the average scores for all residues are zero and the standard deviation is one. This score is used as a relative measure of evolutionary conservation for each amino acid position, where the lowest (negative) value represents the most conserved positions, whereas the highest (positive) value is achieved for the least conserved positions (Ashkenazy et al., 2010). Finally, amino acid conservatism for each position was visualised by colour gradient on the 3D models of the biding region in great tit TLR4 and TLR5 in program FirstGlance in Jmol, version 2.51 (http://bioinformatics.org/firstglance/fgij/). This colour gradient was automatically derived from conservation score.

3.9 Haplotype networks and phylogenetic trees

For *TLR4* and *TLR5* both nucleotide and amino acid haplotype networks were constructed by median neighbour-joining method in program Network 5, version 5 (Bandelt et al., 1999) whereas for neutral markers only nucleotide-based haplotype networks were constructed. For preparation of input file to Network 5, FASTA sequences were converted into rdf format in program DNA alignment. Compared to *TLR*s, neutral markers contained INDELs mutations, however, they were not excluded from the analysis. The haplotype networks were further visualized and edited in program Network Publisher, version 2 (http://www.fluxus-engineering.com/nwpub.htm). Further neighbour-joining networks were also constructed for *TLR4* and *TLR5* using Splits Tree4 (version 4. 14.2) (Huson and Bryant, 2006) Unlike the first method, these haplotype networks depict phylogenetic relationship more precisely, i.e. species-specific vs species-nonspecific (trans-specific) clustering. Species can also share not only identical nucleotide or amino acid sequences, but also more diverged allelic lineages. In contrast to haplotype networks constructed in Network, Splits Tree4 networks take evolutionary distances among species into account and they depict them by the length of branches. Phylogenetic trees of *TLR4* and *TLR5* were constructed in MEGA software, version 6.06 (Tamura et al., 2013) by Maximum likelihood method (ML) with bootstrap value 1000 and general time reversible model.

3.10 Analysis of molecular variance (AMOVA)

Analysis of molecular variance (AMOVA) was calculated to address the questions if the detected shared polymorphism results from extensive ILs and, therefore, TSP is rather transient, or if TSP is maintained in a long term as balanced polymorphism. AMOVA was calculated in program GenAlEx 6.5 (Peakall and Smouse, 2012) based on the PHIPT values

with 999 permutations. The level of variability explained among species was compared for each gene.

We hypothesise that balancing selection acting on immune genes would lead to higher intra-species polymorphism (highly divergent trans-specific allele can be maintained within different species) and lower interspecies differences (TSP alleles can be maintained as identical or nearly identical alleles among species). In the case of strong balanced selection going on *TLR4* and *TLR5*, we might observe lower interspecies differences in the *TLRs* in comparison to neutral markers. On the other side, if the shared polymorphism is rather transient TSP resulting from ILs, the proportion of shared variability in neutral markers and *TLR*s could be approximately similar.

3.11 Electrostatic surface charge analysis

Analysis of electrostatic surface of both TLR4 and TLR5 binding regions (for the list of sequences see Supplement 6 and Supplement 8) was performed in Protein Interaction Property Similarity Analysis (PIPSA) by using webPIPSA, http://pipsa.h-its.org/pipsa/pipsa-index.jsp (Gabdoulline et al., 2007; Richter et al., 2008). For this analysis the initial three_dimensional structural models created by I-TASSER (Chapter 3.6) in Protein Databank format (PDB format) were used as an input. To ensure that all models are superimposed, the structural alignment was performed on webPIPSA using the default setting with an option "optimize sup2pdbs" where all input PDB files are considered to be templates for modelling. Electrostatic potential of all structures was then calculated by Adaptive Poisson-Boltzmann Solver (APBS) in standard environment (T = 300 K, ion strength = 50 nM). The program calculates the potential in complete surface skin which is defined by using probe of radius 2 Å. Hodgkin similarity indices of the protein electrostatic potentials as well as average electrostatic potential differences (the difference in electrostatic potentials of two proteins given in $kcal.mol^{-1}.e^{-1}$ divided by the number of grid points in the comparison region where the two protein skins overlap) were calculated. The similarity indices are here expressed in the interval from -1 (anti-correlated potential), through 0 (uncorellated) to +1 (identical potentials). These values are further converted into distances expressed by $\sqrt{2-2\mathrm{SI}}$ where SI means the respective similarity index. After the conversion, the final values range from 0 (identical values) to 2 (anticorrelated potentials) (Richter et al., 2008). Using R program on the webpage, the conversed values were subsequently automatically visualized in distance matrix presented here as a colour heat map which scores a degree of similarities in surface electrostatic charge among all species by different colour and by a dendrogram (epogram), which grouped species by their electrostatic charge. Partial electrostatic surface

charge of both TLR4 and TLR5 was further visualised in 3D models for each species in program Jmol, version 13 (http://jmol.sourceforge.net/).

In contrast to *TLR4* gene, the sequenced region of *TLR5* gene contains short signal peptide sequence at the beginning of exon 3 which is important for its localisation in plasma membrane (*TLR4* was sequenced from the number 238 of amino acid position of coding region, whereas *TLR5* gene was sequenced from the beginning of coding region). Localisation of signal peptide for each species was screened by SignalP 4.1 server: http://www.cbs.dtu.dk/services/SignalP/ (Nielsen Henrik, Jacob Engelbrecht, Soren Brunak, 1997). Based on the prediction of the cleavage sites, preceding PIPSA analysis signal peptide sequences were manually trimmed out from PDB files in Notepad ++.

3.12 Isolation with migration model for more than two populations

To reveal and quantify the potential gene flow among species as a source of potential shared variability, model Isolation with migration for more than to populations (IMa2) was applied to our data. This coalescence model allows us to estimate the following demographic parameters: time of divergence between population (t), effective population size (N_e) and gene flow among populations (m or $2Nm$). The calculation is based on Markov chain Monte-Carlo simulations and it enables to include up to 10 populations unlike Isolation with migration model (IM). The calculation is based on Markov chain Monte-Carlo simulations and it enables to include up 10 populations (Hey, 2010) as it differs from previous Isolation with migrtion model (IM) (Hey and Nielsen, 2004). Based on the known hybridization in Paridae (Chapter 1.6.2) and tit phylogeny (Harris et al., 2014; Johansson et al., 2013) the three IMa2 model were designed. IMa2 is based on the several assumptions: selective neutrality, no recombination within loci, free recombination between loci and data that "fits" to selected mutation model (Hey, 2010) . First, tests for neutrality (Tajima's D and Fu and Li's test) were performed and recombination was screened by four gametic test (Hudson and Kaplan, 1985) in DNASp for each locus. Considering the results of recombination estimates all loci were treated with IMgC program (http://hammerlab.biosci.arizona.edu/imgconline.html) which filters and extracts recombination-free blocks of sequences or even whole alleles if it is necessary to maximize DNA sequence rich content (Woerner et al., 2007). All INDELs were excluded from each dataset and we applied infinite site model (Kimura, 1969). We ran the program three times with different random seeds up to 4 millions of simulations with following parameters: -q10 -m5 -t10 -b 100000 -l 1.0 -hfg -hn40 -ha0.975 -hb0.75 -p2567 -s2749. The computations were performed on the linux server (Xukol) possessed by the Department of Zoology, Faculty of Science, Charles University in Prague.

To transform population demographic parameters from the relative values to biologically more relevant quantity, the parameters of effective population size in number of individuals (N_e), migaration rate per year ($2Nm$), population migration rate (m) and divergence time in years (t) were calculated from neutral mutation rate determined for each locus (assessed as the divergence to outgroup). The neutral mutation rate was calculated from the formula $D = 2\mu t$ where D is the estimated D_{xy} (Nei 1987). Great tit was choosen as an outgroup for all models. First, to estimate time of divergence among species in the datasets, we calculated divergence to outgroup (D_{xy} values) in DnaSp for cytochrome b sequences obtained from GenBank (http://www.ncbi.nlm.nih.gov/) and published by Gill et al (2005). Based on the calibrated molecular clock for tits from Päckert et al. (2007), the estimated average substitution rate 1.2% of sequence divergence per MY was used for the calculating the time of divergence to outgroup. Second, D_{xy} values beween the common ancestor of all species in the model and the great tit were calculated from our sequenation data (pairwise comparison of all individuals per species and markers). From these D_{xy} values and the time of divergence mutation rates were determined for all loci independently in each model. Finally, overall mutation rate was determined as geometric mean of mutaton rates for each locus (obtained as an aritmetic mean for all species in the model).

Figure 5: Three compiled IMa2 models

The phylogeny relationships are adopted from Johansson et al. (2013) for model 1 and model 3 and from Harris et al. (2014) for model 2. The numbering is in concordance with numbering of populations in the models, where numbers higher than number of species labelled ancestral populations.

Model	Species	No. markers	Phylogeny
model 1	*Cyanistes caeruleus* (0) and *C. cyanus* (1)	5	(0,1):2
model 2	*P. atricapillus* (0), *P. caroliensis* (1) and *P. gambeli* (2)	6	((0,1):3,2):4
model 3	*P. rufescens* (0), *P.hudsonicus* (1), *P. cinctus* (2) and *P. scaleteri* (3)	6	(((0,1):4,2):5,3):6

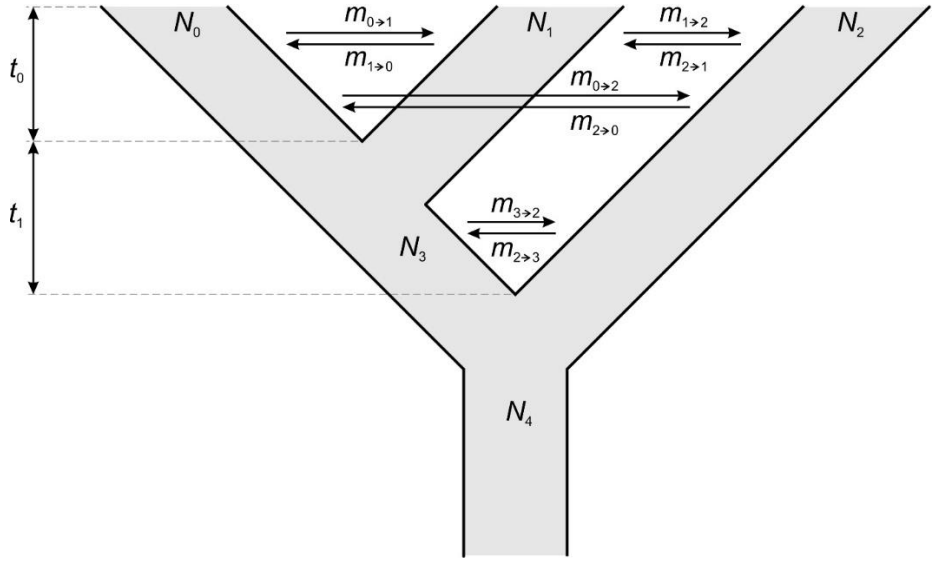


Figure 6: Model isolation with migration for three populations (species)

The model has 15 parameters including effective population size (N_e), time of divergence in population (t) and the gene flow among populations (m) according to Hey (2010). The proper (non-coalescent) direction of the gene flow is indicated by the arrow. However, IM is a coalescence-based model where the coalescence moves backward in time. Therefore, the direction of migrations estimated as outputs from the model are in opposite direction than showed in this figure.

3.13 Ethical note

The research and field sampling in the Czech Republic were approved by Prague Municipality Department of Environmental Protection (S-MHMP-1061728/2010/OOP-V-790/R-235/Bu), by the Institute of Vertebrate Biology of the Czech Academy of Science within grant project of the Czech Science Foundation (project GACR P505/10/1871) and by the Ethical committee of theFaculty of Science, Charles University in Prague (22003/ENU/16-1009/630/16) within collecting samples for Genetic Bank of the Department of Zoology, Charles University in Prague.

4 Results

4.1 General information on Illumina MiSeq run and sequences

We obtained 2 569 625 553 bp from our run in total, the mean average coverage being 2186 reads per PCR product. The conservative capacity of one MiSeq run is estimated to be approximately 5 Gbp (Beneš in personal communication, 2014). After filtering only the most abundant sequences, the mean amplicon coverage was 671 sequences per amplicon per barcode. The mean amplicon coverage differed substantially both within the same gene and among different genes (Figure 7) and multiplexes (Figure 8). At least one sequence was obtained from 190 species altogether. The final numbers of sequences which were used for the subsequent molecular genetics analyses are summarized for each gene in Table 8.

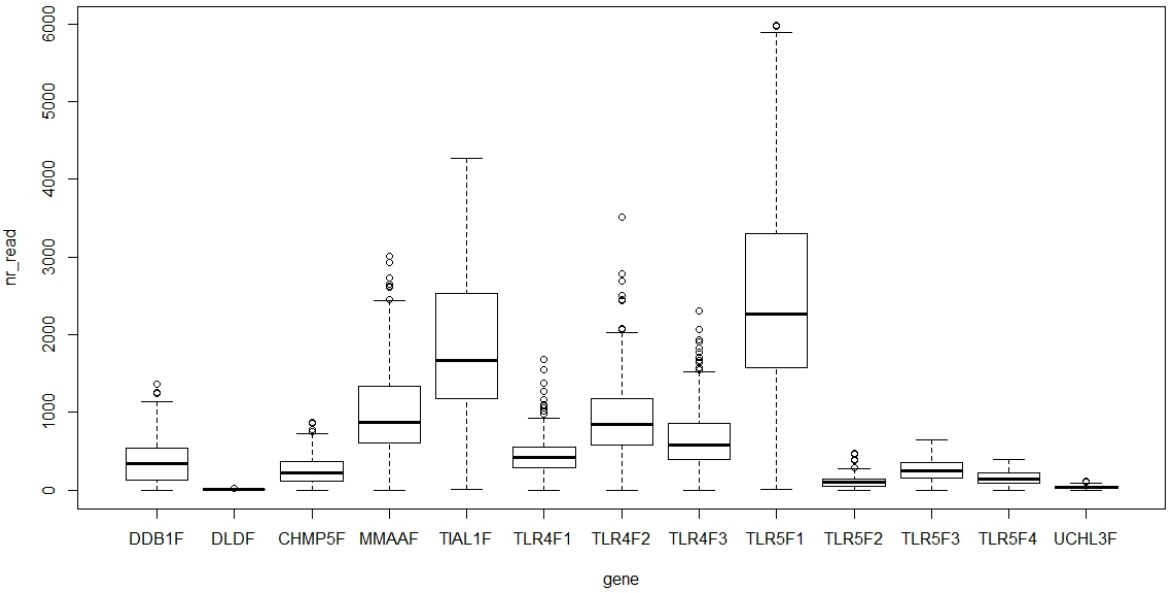


Figure 7: Mean amplicon coverage in both neutral markers and *TLR*s after filtering the most abundant sequences per gene per barcode

The box plot is visualized in R-program, R Core Team (2014).

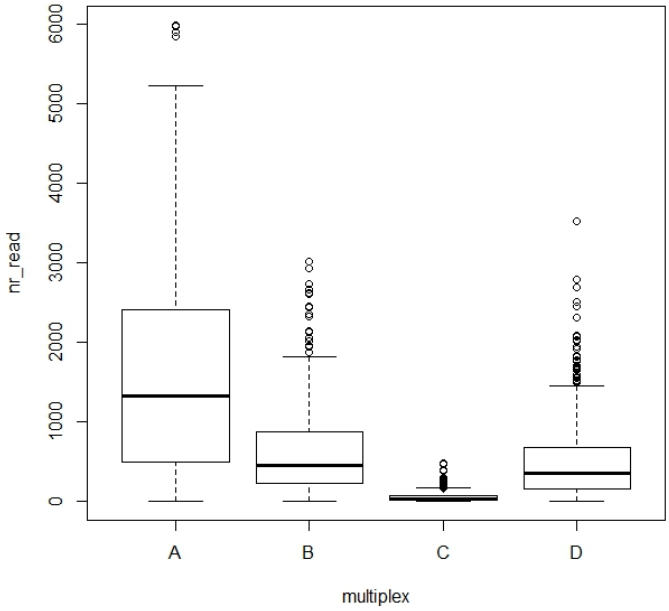


Figure 8: Mean amplicon coverage in both neutral markers and *TLRs* after filtering the most abundant sequnces in different multiplexes

The box plot is visualized in R-program, R Core Team (2014).

Table 8: The final number of sequences used for subsequent molecular genetics analysis (after "phasing process")

Species	*DDB1*	*DLD*	*CHMP 5*	*MMAA*	*TIAL*	*UCHLP 3*	*TLR4*	*TLR5*
Baeolophus atricristatus	6	6	6	6	6	6	6	6
Baeolophus bicolor	18	12	18	8	20	14	20	16
Baeolophus ridgwayi	12	2	12	12	12	12	12	12
Baeolophus wollweberi	12	6	12	8	12	10	12	12
Cyanistes caeruleus	/	18	24	24	24	22	24	24
Cyanistes cyanus	/	20	18	20	20	20	20	20
Lophophanes cristatus	22	16	22	22	22	22	22	22
Melaniparus afer	4	4	4	4	4	4	4	4
Melaniparus niger	10	8	10	10	10	10	10	10
Parus major	50	50	50	50	50	50	50	48
Periparus ater	22	12	22	22	22	22	22	20
Poecile atricapillus	18	14	18	18	18	18	18	18
Poecile carolinensis	18	6	18	18	18	16	18	16
Poecile cinctus	20	18	20	20	20	20	20	16
Poecile gambeli	18	12	18	18	18	18	18	18
Poecile hudsonicus	20	12	20	20	20	18	20	20
Poecile montanus	28	22	28	28	28	28	28	28
Poecile palustris	12	18	24	24	24	24	24	24
Poecile rufescens	20	20	20	20	20	20	20	18
Poecile sclateri	12	8	10	12	12	12	12	12

4.2 Polymorphism in TLRs and neutral markers

Basic population genetics characteristics are summarized for neutral markers (Supplement 1) and for *TLR4* and *TLR5* (Table 9). By using Tajima's D and Fu and Li's D statistics no signature of prevailing selection or population demographic change was detected, since these test characteristics were non-significant for most cases of both neutral markers and *TLR*s. Moreover, negative Tajima's D statistics was significant for *DDB1* in *P. atricapillus* and marginally significant for DDB1 in *P. hudsonicus*, for *TIAL* in *B. wollweberi*, *P. gambeli, P. montanus* and for *UCHLP3* in *P. ater* and *P. gambeli*. Positive Tajima's D were only marginally significant for *DDB1* in *M. niger* and and for *TLR5* in *C. caeruleus*. We may assume that aforementioned loci in this particular species may be under the direct influence of recent positive/ negative selection (with negative Tajima's) or be influenced by the selection indirectly via hitch-hiking. Positive Tajima's D may indicate maintaining of polymorphism either directly by balancing selection or indirectly via hitch-hiking.

Afterwards we calculated mean π and Tajima's D value separately for the neutral markers and *TLR*s (Table 10) and performed a pairwise comparison by paired t-test. We hypothesised that positive selection acting on *TLR*s would have led to lower π for *TLR*s and more negative Tajima´s D value in *TLR*s compared to neutral markers. On the other side, balancing selection acting on *TLR*s would have led to higher π and more positive Tajima´s D in *TLR*s in comparison with neutral markers. However, a pairwise comparison of average π and Tajima's D values by paired t-test showed that these characteristics did not statistically differ between the neutral markers and the *TLR*s (p=0.221461 for π and p=0.884798 for Tajima´s D).

.

Table 9: Basic population genetics characteristics, Tajima's D and Fu and Li's D and recombination estimates for *TLR4* (Table A) and *TLR5* (Table B)

Number of haploid sequences (N), number of unique nucleotide haplotypes (N_2), number of segregating sites (S), number of mutations (n), nucleotide diversity per site (π), proportion of polymorphic sites per site ($\boldsymbol{\theta}$), estimate of recombination parameter (R), minimal number of recombination events (R_m), divergence to outgroup (zebra finch) - average number of nucleotide substitutions (K), divergence to outgroup – average number of nucleotide substitution per base (D_{xy}). Tajima's D, Fu and Li's D statistic, R and R_m are not defined if there is no polymorphism within species. Significant Tajima's D and Fu and Li' s D values (p<0,05) are labelled by three asterisks ***, marginally significant values (p>0,05 and p<0.1) are labelled by one asterisk *. The legend shown here is identical for all tables.

A)

TLR4	Species	Length	*N*	N_2	*S*	*n*	π	θ	Tajima's D	Fu and Li's D	R_m	*R*	*K*	D_{xy}
	Baeolophus atricristatus	829	6	5	7	7	0.00410	0.00370	0.63465	0.71980	0	0.0385	73.500	0.08661
	Baeolophus bicolor	829	20	16	20	21	0.00503	0.00714	-1.12879	-1.21889	3	0.1120	70.950	0.08559
	Baeolophus ridgwayi	829	12	10	9	9	0.00274	0.00359	-0.96364	-1.19243	1	/	74.333	0.08967
	Baeolophus wollweberi	829	12	4	4	4	0.00110	0.00160	-1.10317	-1.28584	0	0.0121	69.333	0.08309
	Cyanistes caeruleus	829	24	16	16	16	0.00465	0.00517	-0.35967	-0.87467	4	0.1485	68.167	0.08223
	Cyanistes cyanus	829	20	20	6	6	0.0025	0.00204	0.072070	0.547727	1	0.0056	67.500	0.08017
	Lophophanes cristatus	829	22	4	3	3	0.00077	0.00099	-0.58648	-1.30921	0	0.7403	71.733	0.08658
	Melaniparus afer	829	4	1	0	0	0	0	/	/	0	/	66.000	0.07961
	Melaniparus niger	829	10	3	4	4	0.00161	0.00171	-0.21888	-0.33833	0	0.0114	65.600	0.07913
	Parus major	829	50	11	12	12	0.00274	0.00323	-0.45227	-1.22570	3	0.0091	68.540	0.08268
	Periparus ater	829	22	21	17	17	0.00693	0.00570	0.76619	-0.01205	7	0.1534	72.619	0.08760
	Poecile atricapillus	829	18	17	23	24	0.00590	0.00842	-1.18447	-1.21901	4	/	72.444	0.08739
	Poecile carolinensis	829	18	13	23	24	0.00569	0.00842	-1.28083	-0.78457	5	0.0641	72.167	0.08705
	Poecile cinctus	829	20	9	11	11	0.00265	0.00374	-1.03012	-1.20487	1	0.0374	74.150	0.08945
	Poecile gambeli	829	18	13	16	16	0.00407	0.00561	-1.04930	-1.26346	2	0.0536	73.611	0.08880
	Poecile hudsonicus	829	20	17	23	23	0.00629	0.00782	-0.75476	-0.74928	5	0.7669	72.900	0.08794
	Poecile montanus	829	28	23	16	16	0.00503	0.00496	0.04616	-0.26897	5	0.1147	74.786	0.09021
	Poecile palustris	829	24	10	8	9	0.00205	0.00291	-0.97162	-0.81094	2	0.1126	73.750	0.08896
	Poecile rufescens	829	20	10	8	8	0.00283	0.00272	0.13812	-0.35425	3	/	73.600	0.08878
	Poecile sclateri	829	12	6	8	8	0.00274	0.00320	-0.56737	-0.53139	1	0.5290	72.583	0.08756

B)

TLR5	**Species**	**Length**	***N***	N_2	***S***	***n***	π	θ	**Tajima's D**	**Fu and Li's D**	R_m	***R***	***K***	D_{xy}
	Baeolophus atricristatus	1342	6	3	5	5	0.00144	0.00163	-0.65543	-0.79148	0	0.0003	124.667	0.09331
	Baeolophus bicolor	1342	16	12	14	14	0.00248	0.00314	-0.81643	-1.16203	2	0.0332	123.813	0.09143
	Baeolophus ridgwayi	1342	18	10	18	12	0.00500	0.00444	0.54963	0.82754	2	0.0383	125.750	0.09412
	Baeolophus wollweberi	1342	12	7	6	6	0.00165	0.00148	0.43244	0.10129	2	0.0461	125.167	0.09286
	Cyanistes caeruleus	1342	24	16	13	14	0.00234	0.00279	-0.57128	-0.80569	2	0.1245	125.417	0.09270
	Cyanistes cyanus	1342	20	2	2	2	0.00078	0.00042	1.98958*	0.86615	0	0.0034	125.100	0.09325
	Lophophanes cristatus	1342	22	2	2	2	0.00034	0.00041	-0.037070	-0.62931	0	/	123.273	0.09227
	Melaniparus afer	1342	4	2	1	1	0.00037	0.00041	-0.61237	-0.61237	0	/	121.750	0.09113
	Melaniparus niger	1342	10	7	10	10	0.00222	0.00263	-0.69853	-0.76777	0	/	122.400	0.09162
	Parus major	1342	48	19	23	23	0.00279	0.00388	-0.91685	-0.87144	6	/	122.872	0.09197
	Periparus ater	1342	20	14	27	27	0.00345	0.00567	-1.52343	-1.83875	1	0.0302	131.550	0.09847
	Poecile atricapillus	1342	18	12	22	22	0.00365	0.00470	-0.92221	-1.46195	3	0.1111	120.833	0.09044
	Poecile carolinensis	1342	16	17	24	24	0.00576	0.00539	0.28226	0.14754	6	0.0036	122.000	0.09113
	Poecile cinctus	1342	16	2	1	1	0.00007	0.00021	-1.16439	-1.53959	0	/	120.950	0.09053
	Poecile gambeli	1342	18	16	20	21	0.00468	0.00455	0.10843	0.84757	7	0.0835	119.500	0.08945
	Poecile hudsonicus	1342	20	7	6	6	0.00113	0.00126	-0.33057	-0.15415	0	0.0369	119.600	0.08952
	Poecile montanus	1342	28	15	22	22	0.00386	0.00421	-0.29695	-0.88032	5	0.0177	118.893	0.08899
	Poecile palustris	1342	24	6	5	5	0.00082	0.00100	-0.52186	1.16632	0	/	116.208	0.08698
	Poecile rufescens	1342	18	13	10	10	0.00201	0.00217	-0.25764	-0.42276	2	/	120.389	0.09011
	Poecile sclateri	1342	12	11	25	26	0.00462	0.00642	-1.25127	-1.28182	3	0.0042	120.167	0.08763

Table 10: Arithmetic mean of estimated nucleotide diversity and Tajima's D for both neutral markers and TLRs

	π		Tajima's D	
Species	**neutral markers**	***TLRs***	**neutral markers**	***TLRs***
Baeolophus atricristatus	0.00197	0.00144	-0.704	-0.328
Baeolophus bicolor	0.00399	0.00222	-0.032	-0.959
Baeolophus ridgwayi	0.00258	0.00335	-0.334	0.619
Baeolophus wollweberi	0.00291	0.00172	-0.268	0.002
Cyanistes caeruleus	0.00324	0.00117	0.072	-0.286
Cyanistes cyanus	0.00079	0.00039	-0.384	0.995
Lophophanes cristatus	0.00083	0.00026	-0.776	-0.600
Melaniparus afer	0.00121	0.00037	0.167	-0.306
Parus major	0.00383	0.00325	0.202	0.537
Periparus ater	0.00461	0.00397	-0.890	-0.457
Melaniparus niger	0.00641	0.00501	-0.004	-1.256
Poecile atricapillus	0.00463	0.00229	-0.634	-1.388
Poecile carolinensis	0.00573	0.00534	0.146	0.746
Poecile cinctus	0.00141	0.00101	0.149	-0.430
Poecile gambeli	0.00248	0.00550	-0.867	-0.173
Poecile hudsonicus	0.00164	0.00088	-0.827	-1.027
Poecile montanus	0.00355	0.00296	-0.503	-0.822
Poecile palustris	0.00294	0.00406	-0.539	-0.770
Poecile rufescens	0.00129	0.00140	-0.673	-0.513
Poecile sclateri	0.00416	0.00628	0.093	-0.692
Average	**0.00301**	**0.00264**	**-0.330**	**-0.355**

4.3 Detection of recombination in TLR4 and TLR5

The degree of recombination was evaluated by SBP analysis and GARD method on the dataset of 20 sequences (1 individual per species; see Supplement 5 and Supplement 7) in both *TLR4* and *TLR5* gene. In *TLR4* one recombination breakpoint was revealed by SBP in amino acid position 471 with model average support 100% (AIC = 22.40) and with model average support 99.60 % for cAIC = 10.99. In *TLR5* one recombination breakpoint was revealed by SBP in amino acid position 1076 with model average support 99.75% (AIC = 10.27) and with model average support 90.61 % for cAIC = 2.82. GARD analysis identified no breakpoint either in *TLR4* or in *TLR5*.

4.4 Detection of recurrent positive selection in TLR4 and TLR5

4.4.1 Positive selection in TLR4

By applying four different methods for detecting recurrent positive selection based on dN/dS ratio (REL, PAML, FUBAR and MEME), the signature of positive selection was revealed in 14 amino acid positions (251, 270, 272, 279, 281, 308, 320, 331, 337, 351, 374, 397,

404, 427) on interspecies level. From these 14 selected sites only 3 selected sites (270, 272, 397) were identified by at least three methods. The comparison of all methods is shown in Table 11. The selected positions were further visualized in TLR4 three-dimensional structural protein model in PyMol (Figure 9). Afterward, the position of selected amino acid residue was compared to the predicted mammal binding sites (LPS/ MD2/ TLR4) in 3D model (based on the visual inspection). Since the position of binding sites may differ between birds and mammals, identification of positive selected sites in birds in close proximity to mammal binding may reveal functionally important sites in birds. From the total of 14 identified positively selected sites four positions (320, 374, 397 and 420) are located in the close proximity to mammal binding sites (Figure 9 and see Supplement 9 for the list of predicted binding sites). Subsequently, the level of amino acid conservatism was compared in positively selected sites (Table 12). From overall 14 selected sites there were non-conservative substitutions in 13 sites, out of which a change of charge occurred in 11 sites.

Table 11: Identification of positive selection in *TLR4* gene on interspecies level by using different selection methods: REL, PAML, FUBAR and MEME

Numbering is according to translated great tit CDs sequence, aa stands for amino acid. The residues located in the close proximity of mammal binding sites are surrounded by parentheses ( ).

aa position	REL	PAML	FUBAR	MEME	SUBSTITUTION
251	x				Q/R
270	x	x		x	K/D
272	x		x	x	L/V/Q
279	x				L/V
281	x	x			T/A/I
308	x				N/D
(320)	x				D/S/N
331		x			A/V/E
337	x	x			Q/H
351	x				L/R
(374)	x	x			S/R
(397)	x	x	x	x	S/G/R
404	x				Q/R
(427)	x				A/T

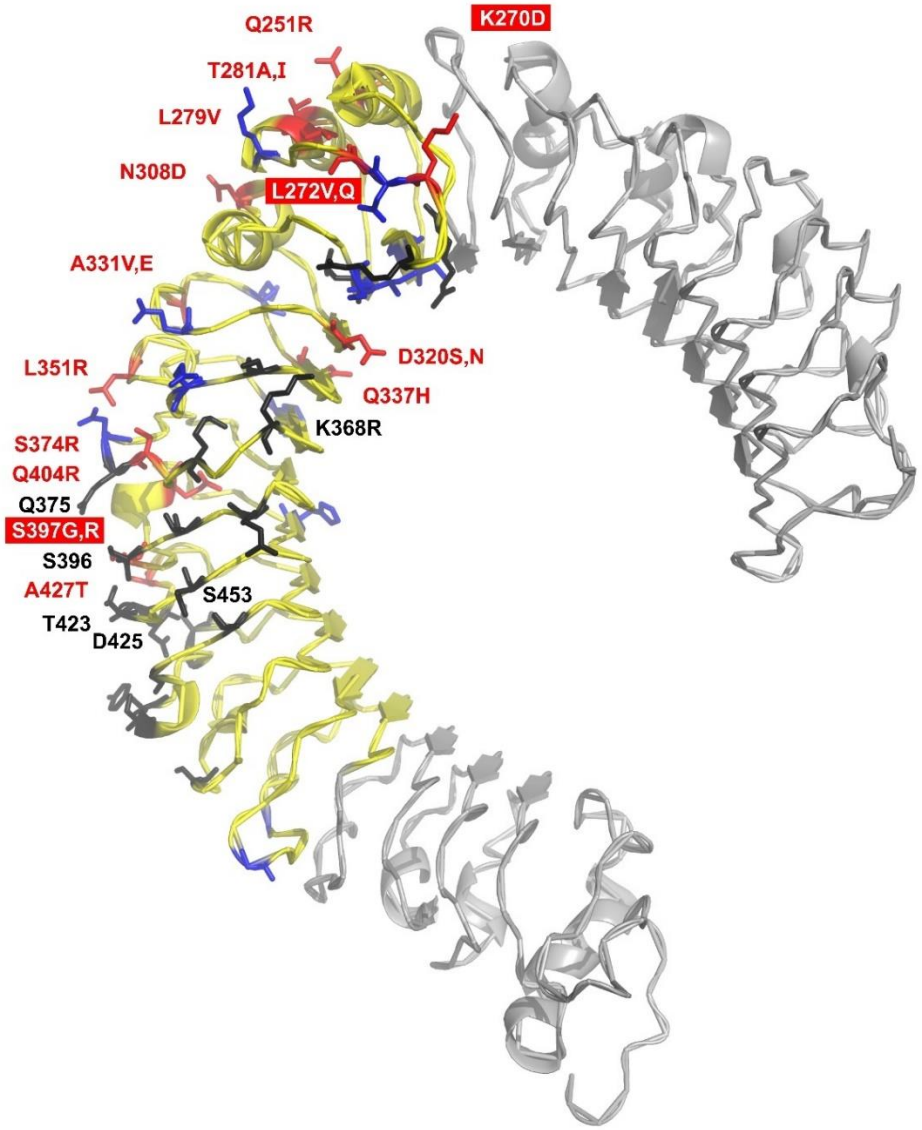


Figure 9: Three-dimensional structural model of great tit TLR4 ectodomain with highlighted positively selected sites, mammal binding sites and variable sites

The model is based on interspecies comparison of 20 tit species. The sequenced region ranging from 238 to 513 aa (according to great tit´s numbering) is highlighted in yellow. Positively selected aa residues with the substitutions are highlighted in red: positions identified on the consensus of at least three methods are in red full-filled boxes, selected sites detected by less than three methods are red. Functionally important mammal sites (for LPS binding, MD2 binding sites and homodimerisation sites) are black. Only mammal binding sites located in the close proximity of positively selected sites are labelled. Non-labelled variable positions are blue.

Table 12: Amino acid substitutions in positively selected sites in *TLR4* gene with the basic chemical properties of substituted aa

aa positions are numbered according to translated great tit *TLR4* sequence. The order of substitutions in the first column is consistent with the one in the structural model (the first substituted bases are according to great tit). In the second column, the presumed polarity (direction) of substitutions is according to tit phylogeny (Ulf S. Johansson et al., 2013) and should reflect the idea of maximal parsimony of evolution. The physiochemical properties of amino acids are adopted and simplified from (Zamyatnin, 1984). Type of conservatism is shown: N – non-conservative substitution and C – conservative substitution. Species which shared particular substitution are in parentheses and they are labelled either by an abbreviation of the scientific name or by the latine name of the genus in cases where all species within the genus share this substitution.

aa position	substitution	polarity	charge	size	type of change	species
Q251R	Gln -> Arg	both polar	uncharged -> positively charged	tiny -> large	N	most **Q**, **R** (BaRi)
K270D	Asp -> Lys	both polar	negatively charged -> positively charged	small -> large	N	most **D**, **K** (PaMa and *Baeolophus*)
L272V,Q	Leu-> Gln	nonpolar-> polar	positively charged -> uncharged	large -> large	N	most **L**, **V** (LoCr), **Q** (BaWo)
	Leu -> Val	both hydrophobic	both uncharged	large -> small	C	
L279V	Val -> Leu	both hydrophobic	both uncharged	small -> large	C	most **L**, **V** (BaRi, BaAt)
T281A,I	Thr -> Asp	polar -> nonpolar	both uncharged	small -> tiny	N	most **T**, **I** (*Baeolophus*), **A** (PoCa)
	Thr -> Ile	polar -> nonpolar	both uncharged	small -> large	N	
N308D	Asn -> Asp	both polar	uncharged -> negatively charged	small -> small	N	all **N**, **D** (BaWo, LoCr)
D320S,N	Asp -> Ser	both polar	negatively charged -> uncharged	small -> tiny	N	**D** (*Baeolophus*, *Cyanistes*, *Melaniparus*, PaMa, PeAt), **N** (*Poecile* except PoCa), **S** (PoCa)
	Asp -> Asn	both polar	negatively charged -> uncharged	small -> small	N	
A331V,E	Ala -> Val	both nonpolar	both uncharged	tiny -> small	C	**A** (*Baeolophus*, LoCr, PeAt, PaMa, PoCa, PoAt, PoRu, PoSc, PoHu, PoGa), **E** (PoMo, PoPa, PoCi), **V** (PaNi, PaAf, PeAt)
	Ala -> Glu	hydrophobic -> polar	uncharged -> negatively charged	tiny -> large	N	
Q337H	Gln -> His	both polar	uncharged -> positively charged	large -> large	N	**Q** (*Baeolophus*, PaMa, PeAt, PoAt, PoCa, PoGa), **H** (*Melaniparus*, PoCi, PoHu, PoMo PoPa, PoRu)
L351R	Leu -> Arg	nonpolar-> polar	uncharged -> positively charged	large -> large	N	most **L**, **R** (PaNi)
S374R	Ser -> Arg	both polar	uncharged -> positively charged	tiny -> large	N	**S** (BaWo, *Cyanistes*, LoCr, PeAt, PaMa), **R** (BaRi, BaAt, BaBi, *Poecile*)
S397G,R	Ser -> Gly	polar -> nonpolar	both uncharged	tiny -> tiny	N	**G** (*Poecile* except PoPa, BaRi, BaAt, CyCy), **S** (BaRi, CyCa, LoCr, PaMa, MeNi, MeAf), **R** (BaWo), **S** (PoPa)
	Ser -> Arg	both polar	uncharged -> positively charged	tiny -> large	N	
Q404R	Gln -> Arg	both polar	uncharged -> positively charged	large -> large	N	most **Q**, **R** (LoCr)
A427T	Ala -> Tyr	nonpolar -> polar	both uncharged	tiny -> small	N	most **G**, **T** (*Baeolophus* and PaMa)

4.4.2 Positive selection in TLR5

By applying four different methods for detecting positive selection (REL, PAML, FUBAR and MEME) the signature of positive selection was revealed in 23 amino acid positions (24, 30, 33, 35, 56, 64, 80, 84, 101, 106, 156, 158, 181, 183, 209, 213, 225, 237, 261, 294, 379, 386, 439) on interspecies level (Table 13Table 13). From these selected sites only 9 selected sites (24, 56, 80, 84, 106, 209, 213, 225, 237) were identified by at least three methods. Similarly to *TLR4*, the positively selected sites were further visualized in TLR5 3D structural protein model in PyMol (Figure 10) and their position was compared to the predicted fish and mammal binding sites (flageline/TLR5) in 3D model. From the total of 23 identified positively selected sites 14 positions (33, 35, 56, 80, 106, 156, 181, 183, 209, 213, 237, 261, 294, 379) were located in the close proximity to mammal binding sites (Figure 10 and see Supplement 9 for the list of predicted binding sites). Subsequently, the level of amino acid conservatism was compared in positively selected sites (Table 14). From overall 23 selected sites there were non-conservative substitutions in 19 sites, out of which a change of charge occurred in 11 sites.

Table 13: Identification of positive selection in *TLR5* gene on interspecies level using different selection methods: REL, PAML, FUBAR and MEME

Numbering is according to translated CDs sequence of great tit. The residues located in the close proximity of mammal binding sites are surrounded by parentheses ( ).

aa position	REL	PAML	FUBAR	MEME	SUBSTITUTION
24	x	x	x		R/S/K/G
30	x				Q/R
(33)		x			I/M/V/T
(35)	x		x		L/F
(56)	x	x	x	x	Y/F/H
64		x			V/A
(80)	x	x	x		T/A/S
84	x	x	x		H/R/P
101	x	x			I/T/V
(106)	x	x	x		G/D/N
(156)	x	x			A/I/T
158		x			E/Q
(181)	x	x			F/S/L
(183)		x			N/K
(209)	x	x	x		N/H
(213)	x	x		x	T/M/A
225	x	x	x	x	F/L
(237)	x	x	x		N/E/D
(261)		x			I/T/M/A
(294)	x				F/L/V
(379)		x			D/Y
386	x				V/I
439		x			D/N

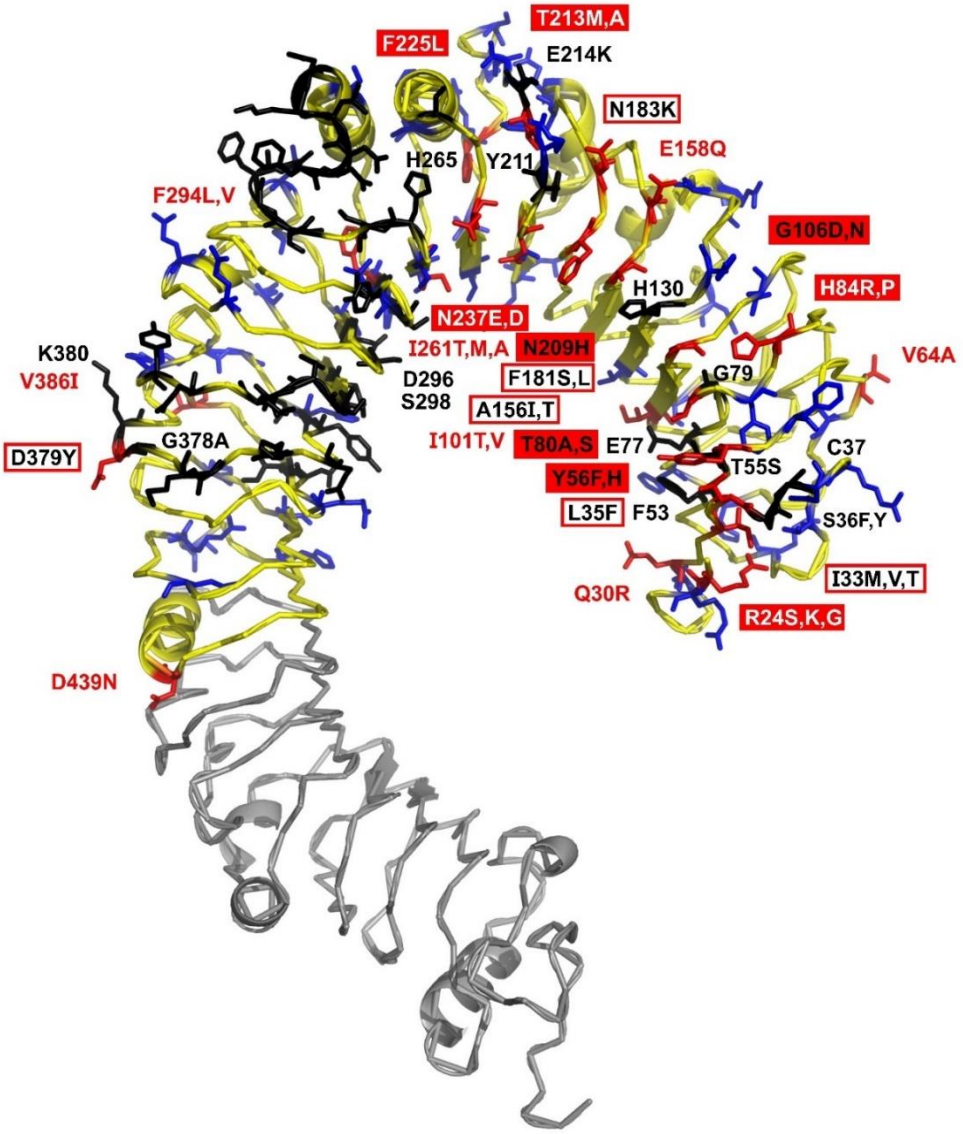


Figure 10: Three-dimensional structural model of great tit TLR5 ectodomain with highlighted positively selected sites, mammal binding sites and variable sites

The model is based on interspecies comparison of 20 tit species. The sequenced region ranging from 1 to 747 aa (great tit´s numbering) is highlighted yellow. Selected aa positions with substitutions are highlighted in red: positions identified on the consensus of at least three selection methods are in red full-filled boxes, selected sites detected by less than three methods are red. Functionally important mammal and fish binding sites are black. Only mammal binding sites located in the close proximity of the selected sites are labelled. Mammal and fish binding sites which were identified to be under positive selection based on the consensus of at least three methods have black label and red full-filled boxes, binding sites under positive selection identified by less than three selection methods have black text with red frames. Non-labelled variable positions are blue.

Table 14: Amino acid substitutions in positively selected sites in *TLR5* gene with the basic chemical properties of substituted aa

aa positions are numbered according to translated great tit *TLR5* sequnce. The order of substitutions in the first column is consistent with the one in the structural model (the first substituted bases are according to great tit). In the second column, the presumed polarity (direction) of substitutions is according to tit phylogeny (Ulf S. Johansson et al., 2013) and should reflect the idea of maximal parsimony of evolution. The physiochemical properties of amino acids are adopted and simplified from (Zamyatnin, 1984). Type of conservatism is shown: N – non-conservative substitution and C – conservative substitution. Species which shared particular substitution are in parentheses and they are labelled either by an abbreviation of the scientific name or by the latine genus name in cases where all species within genus share this substitution.

aa position	substitution	polarity	charge	size	type of change	species
R24S,K,G	Arg -> Gly	polar -> nonpolar	positively charged -> uncharged	large -> tiny	N	**S** (PeAt, *Cyanistes*), **R** (BaWo, *Poecile*, *Melaniparus*, PaMa), **G** (BaAt, BaBi, BaRi), **Y** (LoCr)
	Arg -> Ser	both polar	positively charged -> uncharged	large -> tiny	N	
	Arg -> Lys	both polar	both positively charged	both large	C	
Q30R	Gln -> Arg	both polar	uncharged -> positively charged	both large	N	most **Q**, **R** (CyCa, CyCy, PoAt, PoCa)
I33M,V,T	Ile -> Thr	nonpolar -> polar	both uncharged	large -> small	N	**I** (BaWo, LoCr, *Melaniparus*, PaMa, PoAt, PoPa), **M** (PeAt, PoCa, PoCi, PoGa, PoHu, PoMo PoRu, PoSc), **T** (BaAt, BaRi, BaBi), **V** (*Cyanistes*)
	Ile -> Met	both nonpolar	both uncharged	both uncharged	C	
	Ile -> Val	both nonpolar	both uncharged	large -> small	C	
L35F	Phe -> Leu	both nonpolar	both uncharged	both large	C	most **F**, **L** (CyCa, CyCy, *Melaniparus*, PaMa)
Y56F,H	Tyr -> Phe	both nonpolar	both uncharged	both large	C	most **F, Y** (BaAt, BaBi, BaRi, PaMa, *Melaniparus*), **H** (BaWo)
	Hist -> Tyr	polar -> nonpolar	positively charged -> uncharged	both large	N	
V64A	Ala -> Val	both nonpolar	both uncharged	small -> tiny	C	most **A**, **V** (*Cyanistes*, PeAt, PaMa)
T80A,S	Ala -> Thr	nonpolar -> polar	both uncharged	small -> small	N	most **T**, **S** (BaRi, BaBi, BaAt), **A** (PeAt)
	Thr -> Ser	both polar	both uncharged	small-> tiny	C	
H84R,P	Arg -> Hist	both polar	positively charged	both large	C	**R** (*Baeolophus*, *Cyanistes*, PeAt, LoCr, MeNi), **H** (*Poecile*, PaMa), **P** (MeAf)
	Hist -> Pro	polar -> special	postively charged -> uncharged	large -> small	N	
I101T,V	Ile -> Val	both nonpolar	both uncharged	large -> small	C	most **I**, **V** (BaWo), **T** (LoCr)
	Ile -> Thr	nonpolar -> polar	both uncharged	large -> small	N	
G106D,N	Asp -> Gly	polar -> nonpolar	negatively charged -> uncharged	small -> tiny	N	most **D**, **G** (PaMa, PeAt, PoPa), **N** (BaWo)
	Asp -> Asn	both polar	negatively charged -> uncharged	both small	N	
A156I,T	Ala -> Ile	nonpolar	both uncharged	tiny -> large	probably C	most **A**, **I** (*Melaniparus*), **T** (LoCr)
	Ala -> Thr	nonpolar -> polar	both uncharged	tiny -> small	N	

aa position	substitution	polarity	charge	size	type of change	species
E158Q	Glu -> Gln	polar	negatively charged -> uncharged	both large	N	most **E**, **Q** (*Cyanistes*, PaMa)
F181S,L	Phe -> Ser	nonpolar -> polar	both nocharged	large -> tiny	N	most **F**, **S** (*Cyanistes*, PoMo), **L** (*Melaniparus*)
	Phe -> Leu	nonpolar	both uncharged	both large	C	
N183K	Asp -> Lys	polar	uncharged -> positively charged	large -> small	N	most **K**, **N** (*Cyanistes*, PaMa)
N209H	His -> Asp	polar	postively charged -> uncharged	large -> small	N	most **H**, **N** (PaMa, PeAt, PoCi)
T213M,A	Thr -> Met	polar -> nonpolar	both uncharged	small -> large	N	most **T**, **A** (BaBi, BaAt, BaRi), **M** (PoAt, PoCa)
	Thr -> Ala	polar -> nonpolar	both uncharged	small -> tiny	N	
F225L	Leu -> Phe	nonpolar	both uncharged	both large	C	most **F**, **L** (*Baeolophus*, LoCr, PoCa, PoCi, PoHu, PoPa, PoRu)
N237E,D	Glu -> Asn	polar	negatively charged -> uncharged	large -> small	N	**E** (PeAt, LoCr, *Poecile*), **N** (*Cyanistes*, *Melaniparus*, PaMa), **D** (*Baeolophus*)
	Glu -> Asp	polar	negatively charged	large -> small	C	
I261TM,A	Ile -> Ala	nonpolar	both uncharged	tiny -> larged	probably C	**I** (PeAt, *Cyanistes*, *Melaniparus*, PaMa) **T** (BaBi, BaRi), **A** (BaWo), **M** (BaWo, LoCr, *Poecile*)
	Ile -> Met	nonpolar	both uncharged	both large	C	
	Ile -> Thr	nonpolar -> polar	both uncharged	large -> small	N	
F294L,V	Phe -> Leu	both nonpolar	both uncharged	both large	C	**F (***Baeolophus*, *Cyanistes*,PeAt, PaMa, MeNi), **L** (LoCr, *Poecille*), **V** (MeAf)
	Phe -> Val	both nonpolar	both uncharged	large -> small	C	
D379Y	Asp -> Tyr	polar -> nonpolar	negatively charged -> uncharged	small -> large	N	most **D**, **Y** (PeAt, PoCi, PoGa, PoHu, PoMo, PoRu, PoSc)
V386I	Val -> Ile	both nonpolar	uncharged	small -> charged	C	most **V**, **I** (BaRi, *Cyanistes*)
D439N	Asp -> Asn	both polar	negatively charged -> uncharged	small	N	most **D**, **N** (PeAt, PoHu, PoRu, PoMo)

4.5 Analysis of evolutionary conservative and non-conservative sites in TLR4 and TLR5 (ConSurf)

ConSurf analysis measures a degree of evolutionary conservatism of amino acid substitutions for each amino acid position. The amino acid conservatism is visualized in a 3D model of *TLR4* and *TLR5* sequenced region by colour gradient (Figure 11 and Figure 12, respectively). To simplify the interpretation of results, the program categorized the degree of conservatism based on quartiles of the conservation score into 8-grade numbering scale, where 1 labels the most variable (the most non-conservative) sites through 5 with average conservatism to the most conservative sites labelled by 9. The most non-conservative sites from ConSurf analysis for *TLR4* and *TLR5* gene are listed in Table 15 and Table 16, respectively.

In *TLR4* gene 26 amino acid residues were ranked into the most variable sites (category 1), which is in total 9.4 % from all 276 aa in sequenced region. According to our prior assumptions, most of the identified non-conservative sites are in concordance with those sites, which were identified also by the tests of positive selection (10 sites from total 12 selected sites on interspecies level are considered to be the most non-conservative as well). The non-conservative sites which were also identified by at least one selection method are positions 251, 279, 281, 320, 331, 337, 351, 427, and positions 272 and 397 were identified based on the consensus of at least three selection methods. The amino acid position 397 has also the highest conservation score indicating that it is the most variable position in *TLR4* we detected. However, apart from positively selected sites, other variable sites which are located in close proximity to mammal binding sites can also have functional importance for binding MAMPs. Besides 3 selected sites lying in close proximity to mammal binding sites which also belong to the category of most non-consevative ones, another 6 non-conservative sites (262, 267, 294, 317, 369, 376) were located near mammal functional sites. Additionally, site 368 is directly the predicted binding site (see Table 15 and Supplement 9 for description of binding sites).

In TLR5, 44 amino acid residues fell into the category of the most non-conserved sites, i. e. 9.8% of the sequenced region counting 447 amino acids in total. Similarly, in *TLR5* from overall 23 identified sites being under the influence of positive selection 22 sites were considered to be the most variable as well (in category 1). The non-conservative sites on which selection was detected by at least one method were positions 30, 33, 64, 101, 156, 158, 181, 183, 209 ,261, 294, 379, 386 and 439, and sites where selection was revealed based on the consensus of at least three methods are positions 24, 56, 80, 84, 106, 213, 225 and 237. From the total number of selected sites, 15 sites identified as non-conservative were situated either in close proximity of predicted mammal and fish binding sites or they were binding sites themselves. Furthermore, ConSurf analysis identified other 5 non-conservative sites

(82, 132, 318, 328, 409) which lie in the close proximity of the functionally important sites. In addition to that, 5 other non-conservative sites (36, 53, 301, 376,390) are binding sites themselves.

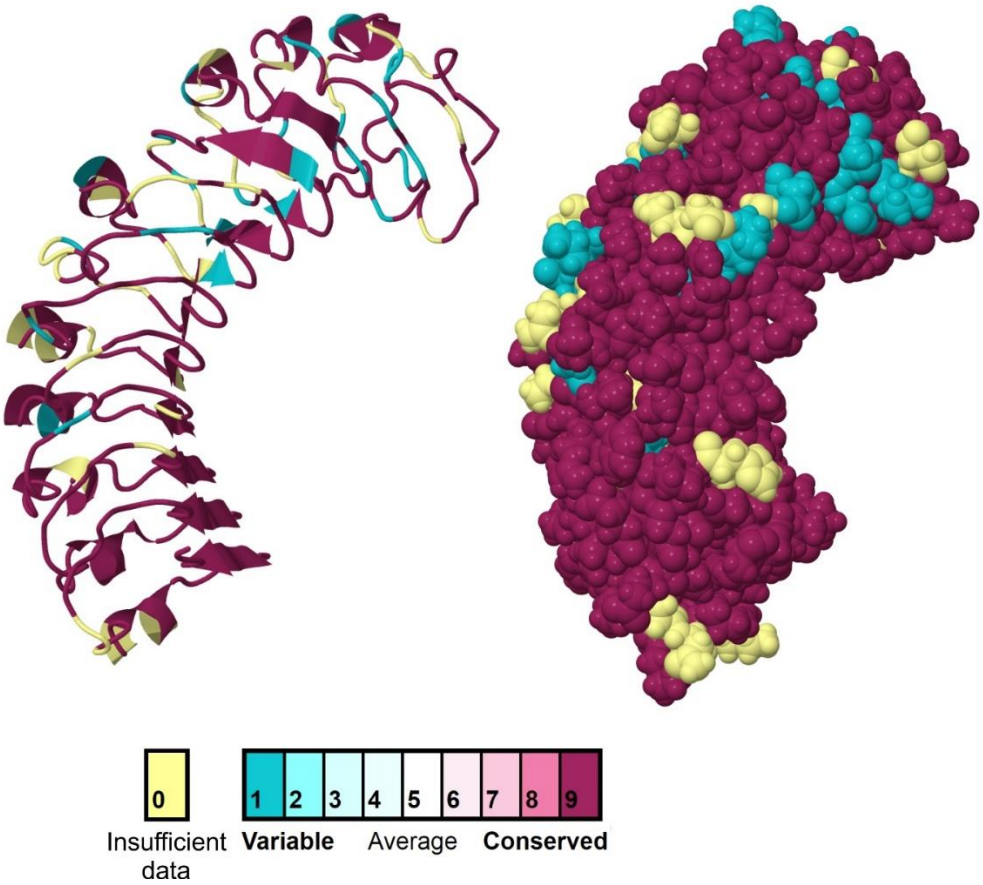


Figure 11: Three-dimensional model of great tit TLR4 binding region modeled by I-TASSER where conservative and non-conservative sites identified by ConSurf analysis are highlighted

The degree of conservatism is showed in colour gradient ranging from pale blue (the most variable site) to dark purple (the most conservative). Insufficient data below confidence cut-off are highlighted in yellow.

Table 15: List of the most non-conservative sites of *TLR4* gene identified by ConSurf analysis

Only the most variable sites are listed based on the conservation score. Numbering is according to translated CDs great tit *TLR4* sequence. Positively selected sites identified by at least one selection method are surrounded by parentheses (), sites identified by at least three selection methods are marked by square brackets [], sites lying in close proximity of predicted mammal binding sites are labelled by an asterisk * and predicted mammal binding site are labelled by a superscript[X]. For complete results please see Supplement 10.

Residue	PaMa sequence	substitutions	conservation score	colour
246	M	R,M	3.336	1
(251)	Q	R,Q	1.017	1
262*	I	T,V,M,I	2.146	1
267*	R	W,R,K	3.735	1
[272]	L	L,Q,V	1.7	1
276	K	K,E	2.051	1
(279)	L	F,V,L	1.592	1
(281)	T	A,T,I	4.829	1
285	Q	Q,R	1.019	1
294*	I	I,S	3.906	1
301	D	D,H	1.061	1
311	G	G,R,S	2.913	1
317*	R	H,R	1.837	1
(320)*	D	D,S,N	1.574	1
325	E	E,K	2.051	1
(331)	A	A,V,E	5.002	1
334	K	K,E,Q	2.129	1
(337)	Q	Q,H	1.751	1
(351)	L	R,W,L	2.871	1
363	R	S,H,C,R	5.274	1
364	I	V,I	1.417	1
368*[X]	K	R,N,K	1.193	1
369*	R	R,G,K	1.071	1
376*	N	N,K	2.133	1
[397]*	S	T,S,R,G	5.399	1
(427)*	T	G,T,A	4.261	1
442	L	L,V	1.698	1

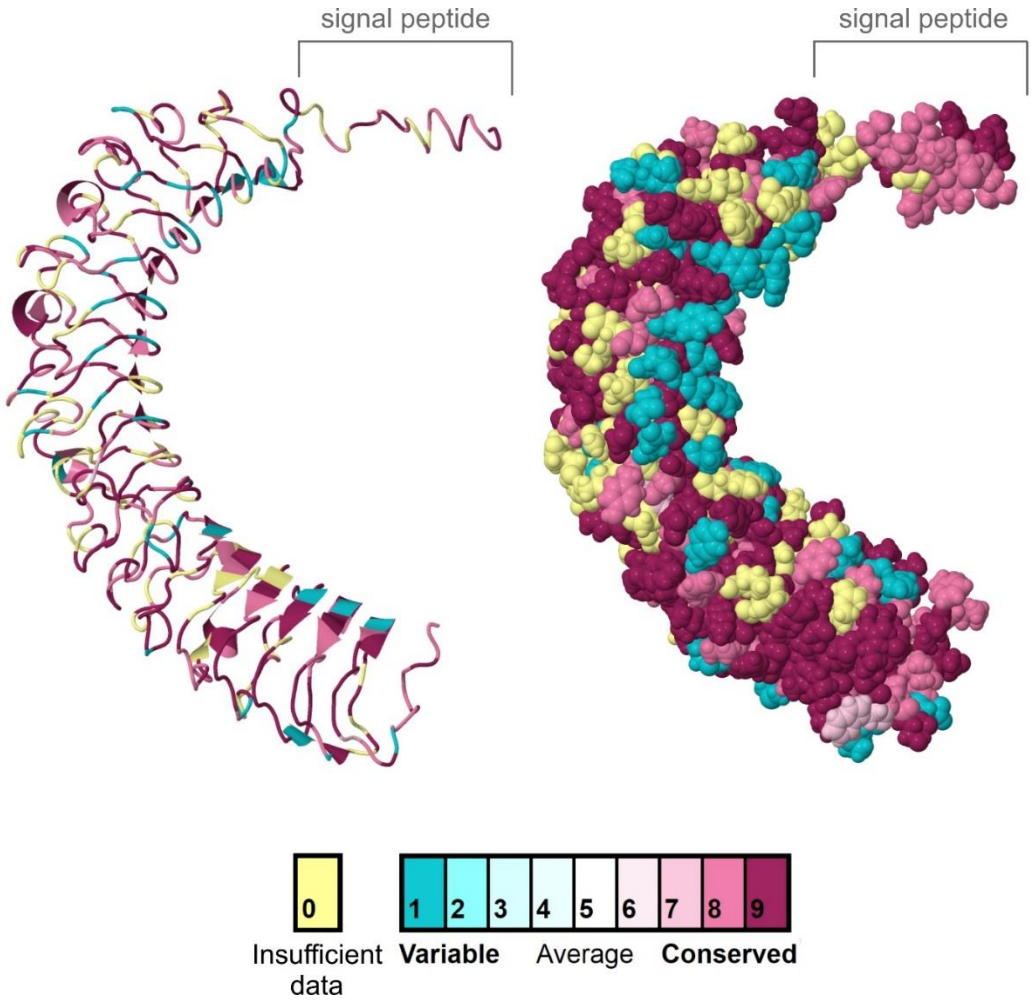


Figure 12: Three-dimensional model of great tit TLR5 binding region and signal peptide modeled by I-TASSER with highlighted conservative and non-conservative sites identified by ConSurf analysis

The degree of conservativism is showed in colour gradient ranging from pale blue (the most variable site) to dark purple (the most conservative). Insufficient data below confidence cut-off are highlighted in yellow.

Table 16: List of non-conservative sites of *TLR5* gene identified by ConSurf analysis

Only the most variable sites are listed based on the conservation score. Numbering is according to translated CDs *TLR5* great tit sequence. Positively selected sites identified by at least one selection method are surrounded by parentheses (), sites identified by at least three selection methods are marked by square brackets [], sites lying in close proximity of predicted mammal binding sites are labelled by an asterisk * and the sites which are directly predicted mammal binding sites are labelled by a superscript X. For complete results please see Supplement 11.

.

Residue	PaMa sequence	substitutions	conservation score	colour
[24]	R	S,K,R,G	4.448	1
29	D	D,N,H	1.159	1
(30)	Q	Q,R	3.829	1
(33)*X	I	T,V,M,I	5.165	1
36*X	S	S,F,Y	2.93	1
53*X	F	F,L	2.978	1
[56]*X	Y	H,Y,F	3.94	1
63	T	T,N,A	1.606	1
(64)	V	V,A,E	4.119	1
[80]*X	T	A,S,T	1.642	1
82*	F	Y,F	1.868	1
[84]*	H	S,P,H,R	4.754	1
100	R	C,R,H	3.045	1
(101)	I	I,V,T	1.753	1
[106]*X	G	G,N,D	3.997	1
132*	Y	Y,C	1.936	1
145	D	D,N	1.168	1
147	R	G,R	1.208	1
(156)*X	A	A,I,T	1.027	1
(158)	Q	Q,E	1.294	1
(181)*X	F	L,F,S	3.693	1
(183)*X	N	N,K	1.252	1
187	F	L,F	1.861	1
(209)*X	N	Q,H,N	4.879	1
[213]*	T	A,M,T	1.661	1
[225]	F	L,F	4.783	1
[237]*	N	N,E,D	1.297	1
249	S	S,C,F	1.578	1
(261)*	I	T,A,I,M,V	2.674	1
(294) *	F	V,L,F	2.965	1
301*X	Y	F,Y	1.934	1
314	G	V,S,G	1.703	1
318*	S	T,S,L	2.226	1
328*	Q	R,K,Q	1.184	1
376*X	M	I,M	1.008	1
(379)*X	D	Y,D	4.075	1
(386)	V	V,I	3.475	1
390*X	I	I,T,K	1.72	1
400	R	R,K	1.195	1
406	H	H,R	1.085	1
409*	S	F,S	1.053	1
418	M	M,I,T	1.003	1
429	H	L,Y,H	2.156	1
(439)	D	D,N	5.041	1

4.6 Evolutionary relationships in TLR4, TLR5 and neutral markers and shared variability in Paridae

Haplotype networks for *TLR*s and neutral markers calculated in program Network are shown in figures Figure 13, Figure 17 and Supplement 2. The topology of some genes appear to be tangled or reticulated with a lot of crosstalks, as in the case of *TLR4* and less in *TLR5*. For this reason, to gain a better insight into tangled evolution of *TLR4* and *TLR5* gene in the genus of *Poecile*, independent nucleotide(Figure 14 and Figure 18) and amino acid haplotype (Figure 16 and Figure 20) networks were constructed for these loci in Networks also nucleotide haplotype networks SplitsTree (Supplement 3 and Supplement 4). The haplotype networks of neutral markers differ in their topology and proportion of shared variability, but a lot of variability is shared among closely related chickadees in general.

Shared alleles and haplotypes of *TLR4* and *TLR5* genes summarized from the haplotype networks and the phylogenetic trees are listed in Table 17. However, the topologies of both trees are poorly supported regarding the low evolutionary distances among sequences. As apparent from the Figure 17 and-Figure 20, *TLR5* appears to evolve in more independent manner (species-specific) since the shared variability is only restricted to pairs of closely related species. By contrast, *TLR4* alleles are occasionally shared by up to three or four species in chickadees (compare Figure 21 and Figure 22). In total, shared variability in *TLR4*, *TLR5* as well as neutral markers is restricted only to genus level (in *Cyanistes*, *Poecile* and *Baeolophus* genera), and in addition to that most frequently only between closely related species. Eurasian *Poecile* are usually well separated in both neutral markers and *TLR*s from American chickadees (except *P. cinctus* in *TLR5* gene) whose evolution appears to be more complicated and reticulated.

Table 17: The overview of shared variability in *TLR4* and *TLR5* genes in Paridae

Shared identical nucleotide (nt) and amino acid (aa) haplotypes are denoted from the haplotype networks. Shared nucleotide allelic lineages are derived from the ML phylogenetic trees and include both shared identical alleles and more diversified allelic lineages. Species sharing alleles are put into parentheses () and are labelled by abbreviation of the first two letters of their scientific name (for a complete list of species see Table 2).

Level of shared variability	***TLR4***	***TLR5***
as nt identical alleles	(PoGa, PoSc), (PoCi, PoMo)	(CyCa, CyCy), (PoHu, PoRu), (PoAt, PoCa)
as aa identical haplotypes	(CyCa, CyCy), (PoCi, PoMo), (PoRu, PoHu), (PoAt, PoCa, PoGa, PoSc)	(CyCa, CyCy), (PoHu, PoRu), (PoGa, PoSc), (PoAt, PoCa), (BaBi, PoAt)
as nucleotide allelic lineages	(CyCa, CyCy), (BaBi, BaAt), (PoCi, PoMo) (PoHu, PoRu, PoCa), (PoCa, PoAt, PoSc, PoGa)	(CyCa, CyCy), (BaBi, BaAt), (PoHu, PoRu), (PoGa, PoSc), (PoCa, PoAt)

Figure 13: Nucleotide haplotype network of *TLR4* in Paridae

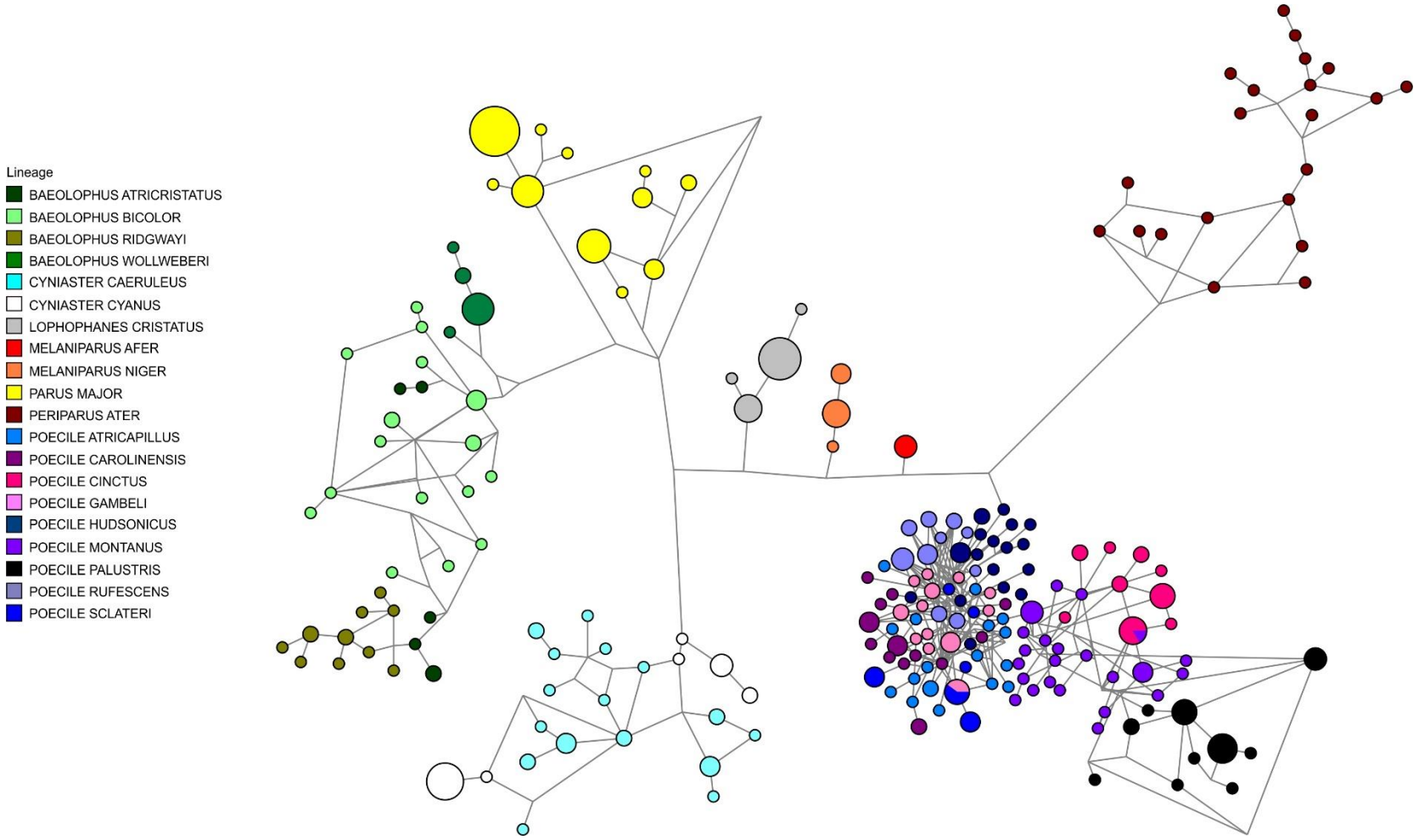


Figure 14: Nucleotide haplotype network of *TLR4* in *Poecile*

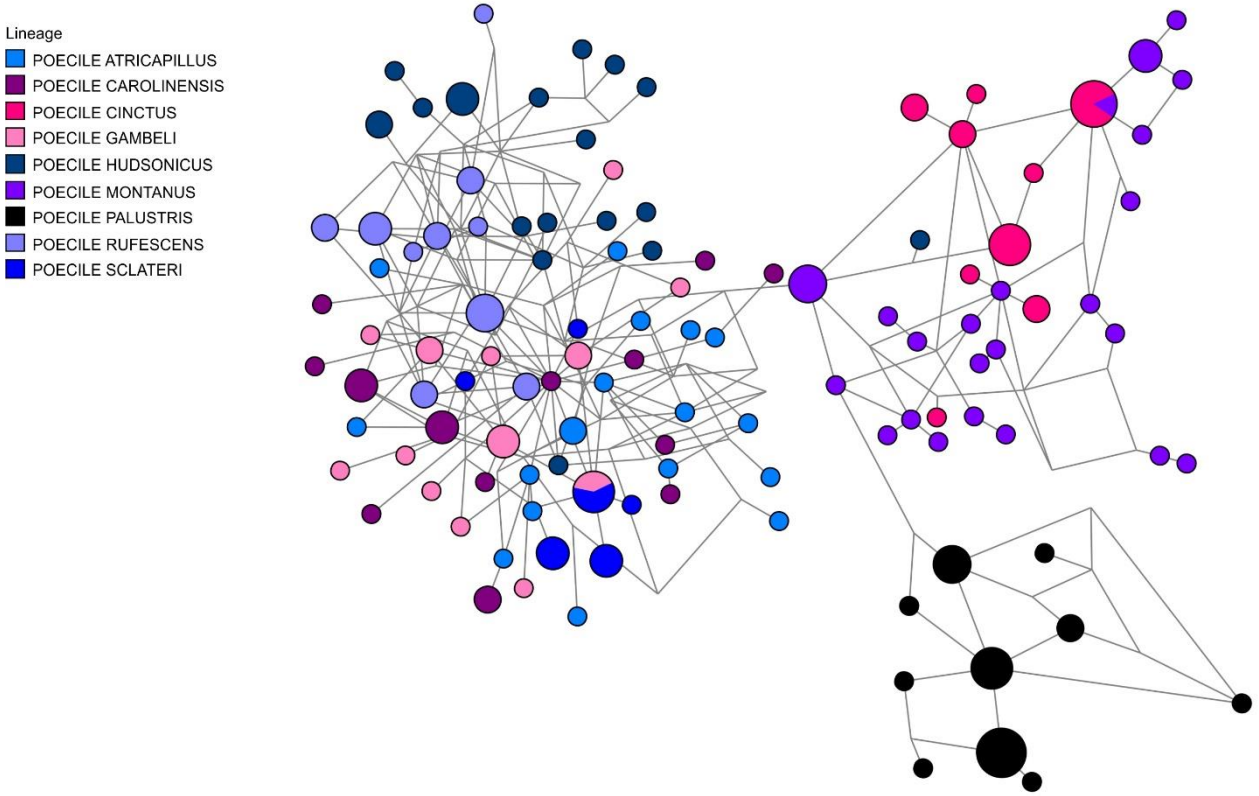


Figure 15: Amino acid haplotype network of *TLR4* in Paridae

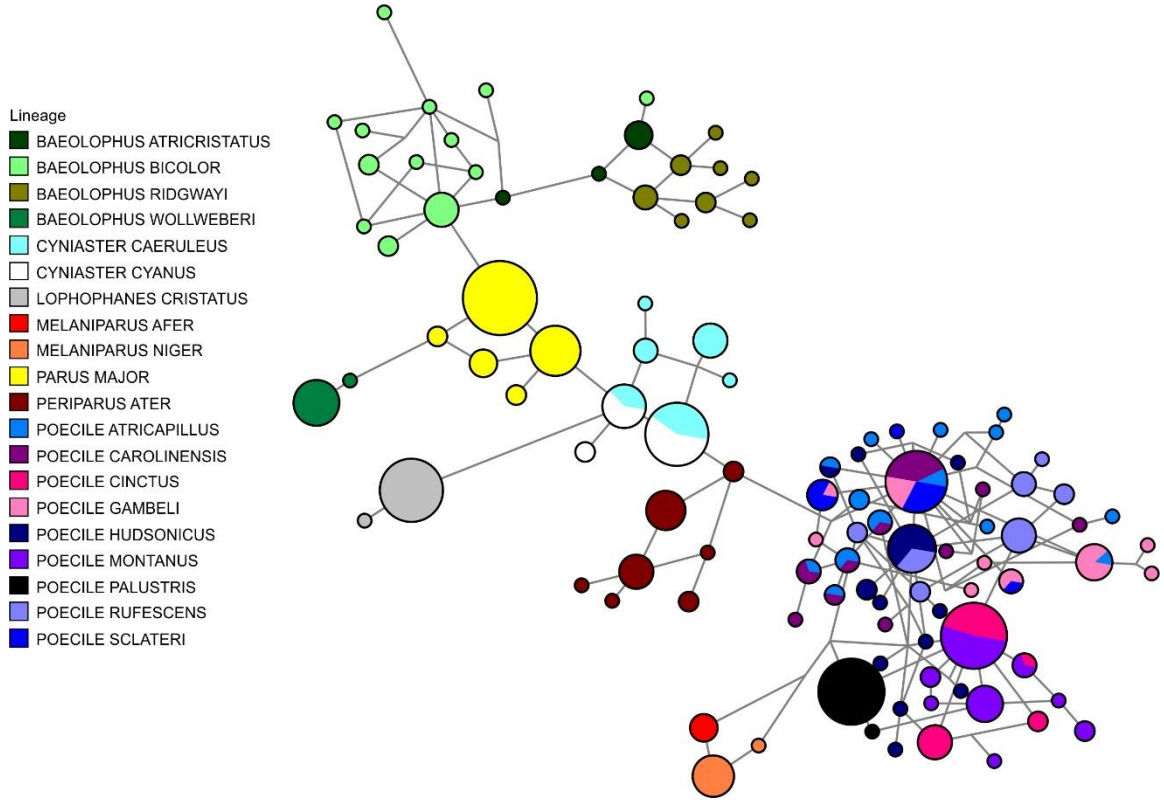


Figure 16: Amino acid haplotype network of *TLR4* in *Poecile*

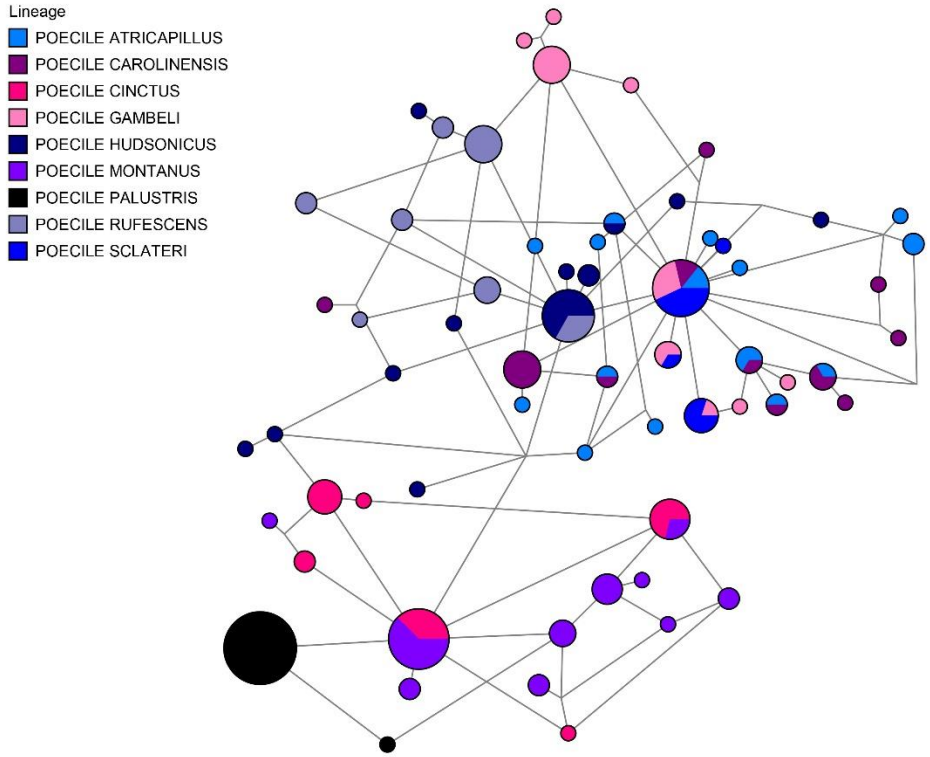


Figure 17: Nucleotide haplotype network of *TLR5* in Paridae

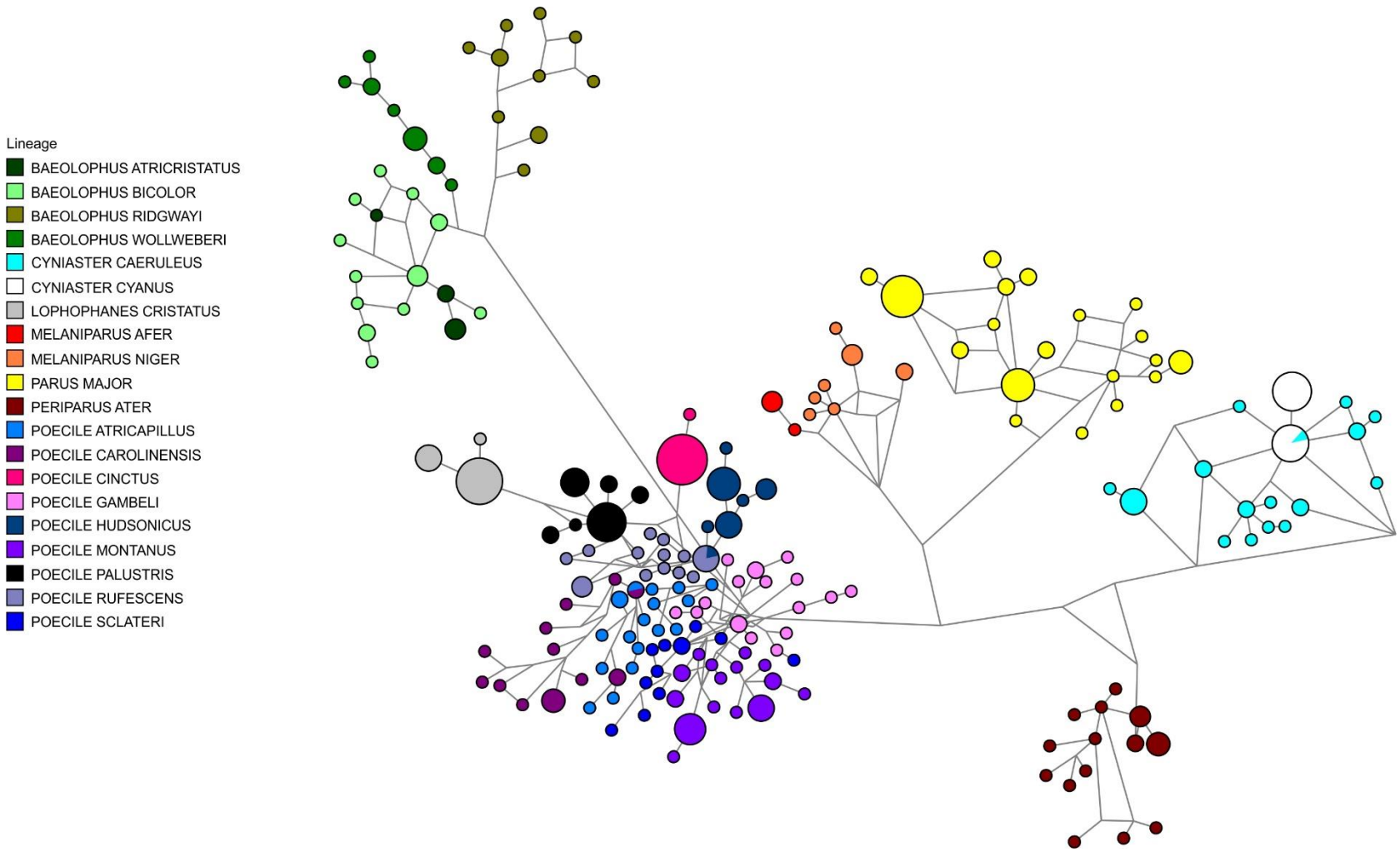


Figure 18: Nucleotide haplotype network of *TLR5* in *Poecile*

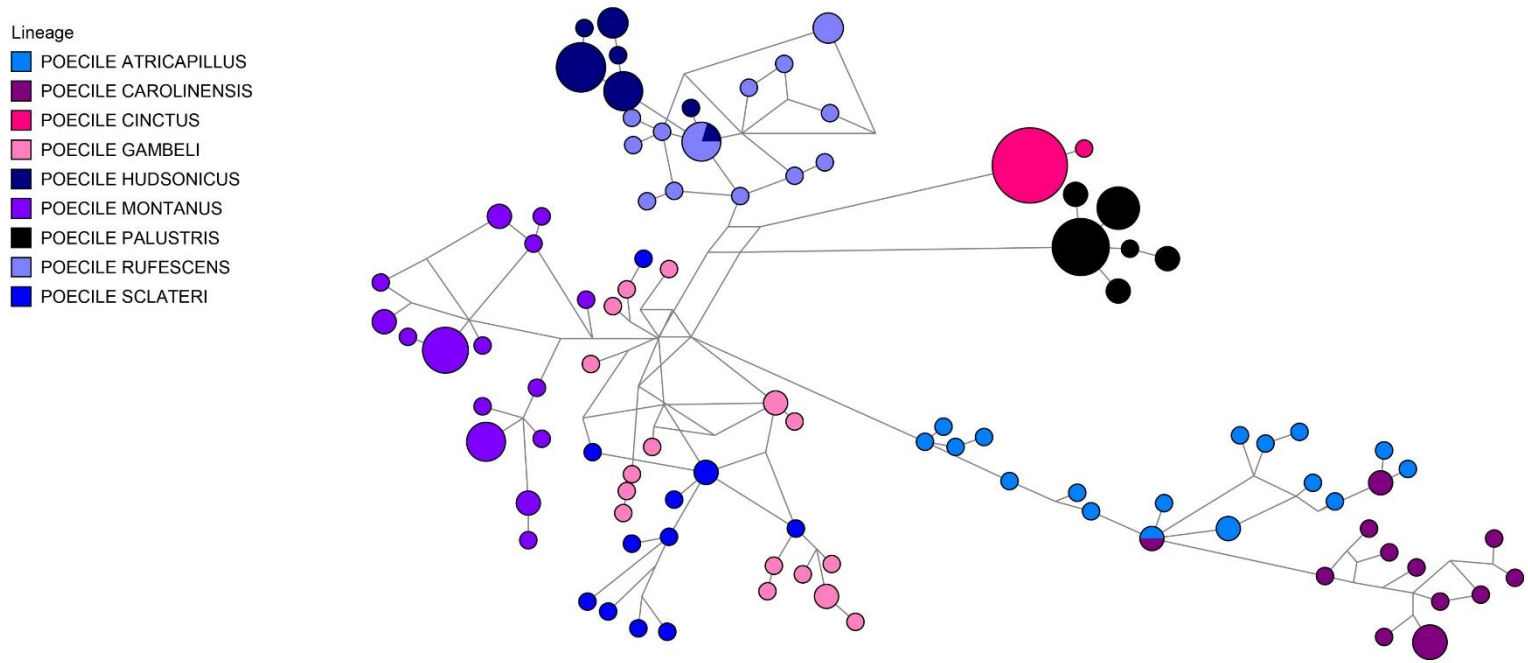


Figure 19: Amino acid haplotype network of *TLR5* in Paridae

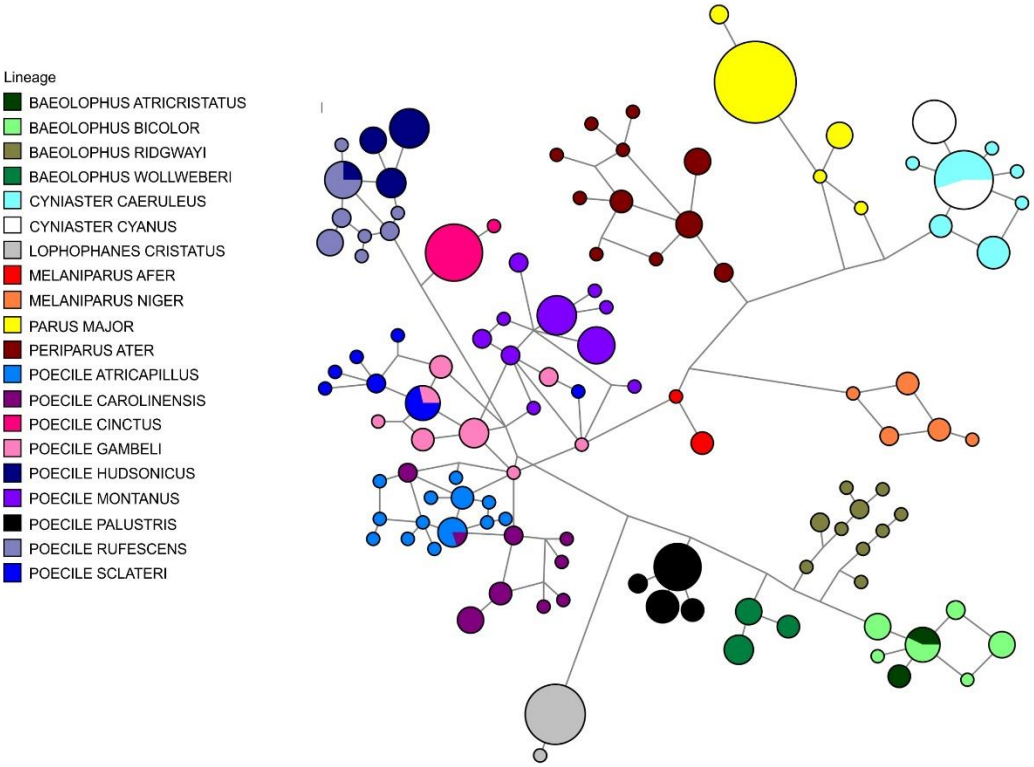


Figure 20: Amino acid haplotype network of *TLR5* in *Poecile*

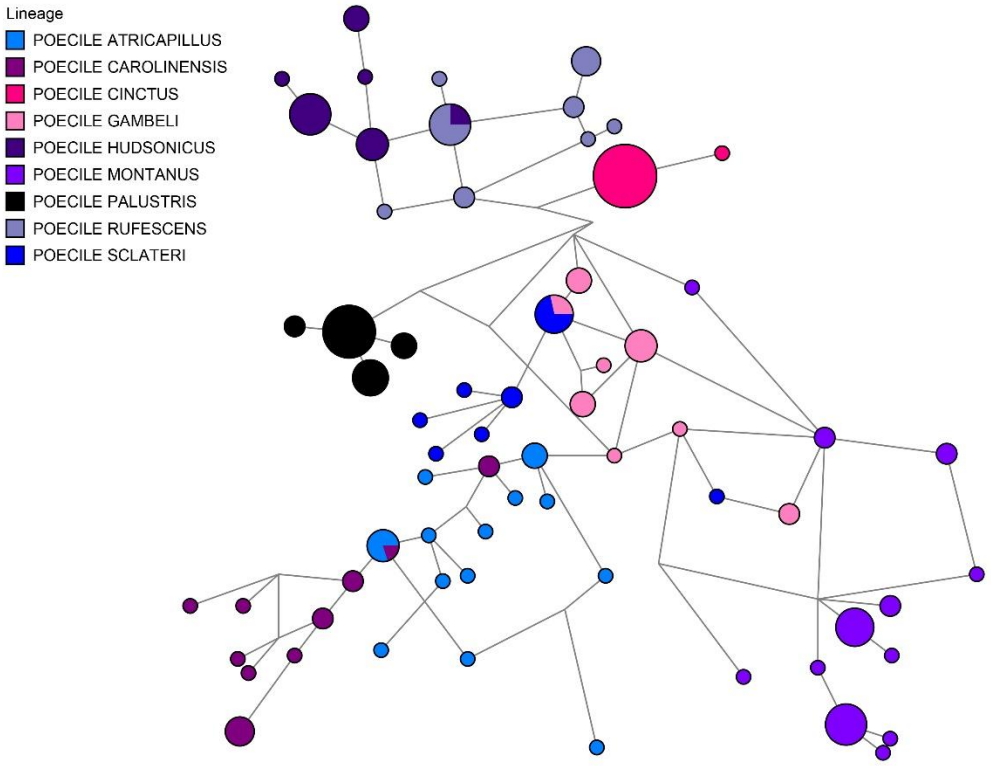


Figure 21: Maximum likelihood tree of *TLR4* in Paridae

The branches without shared alleles among species are condensed. The bootstrap values only above 70 are shown.

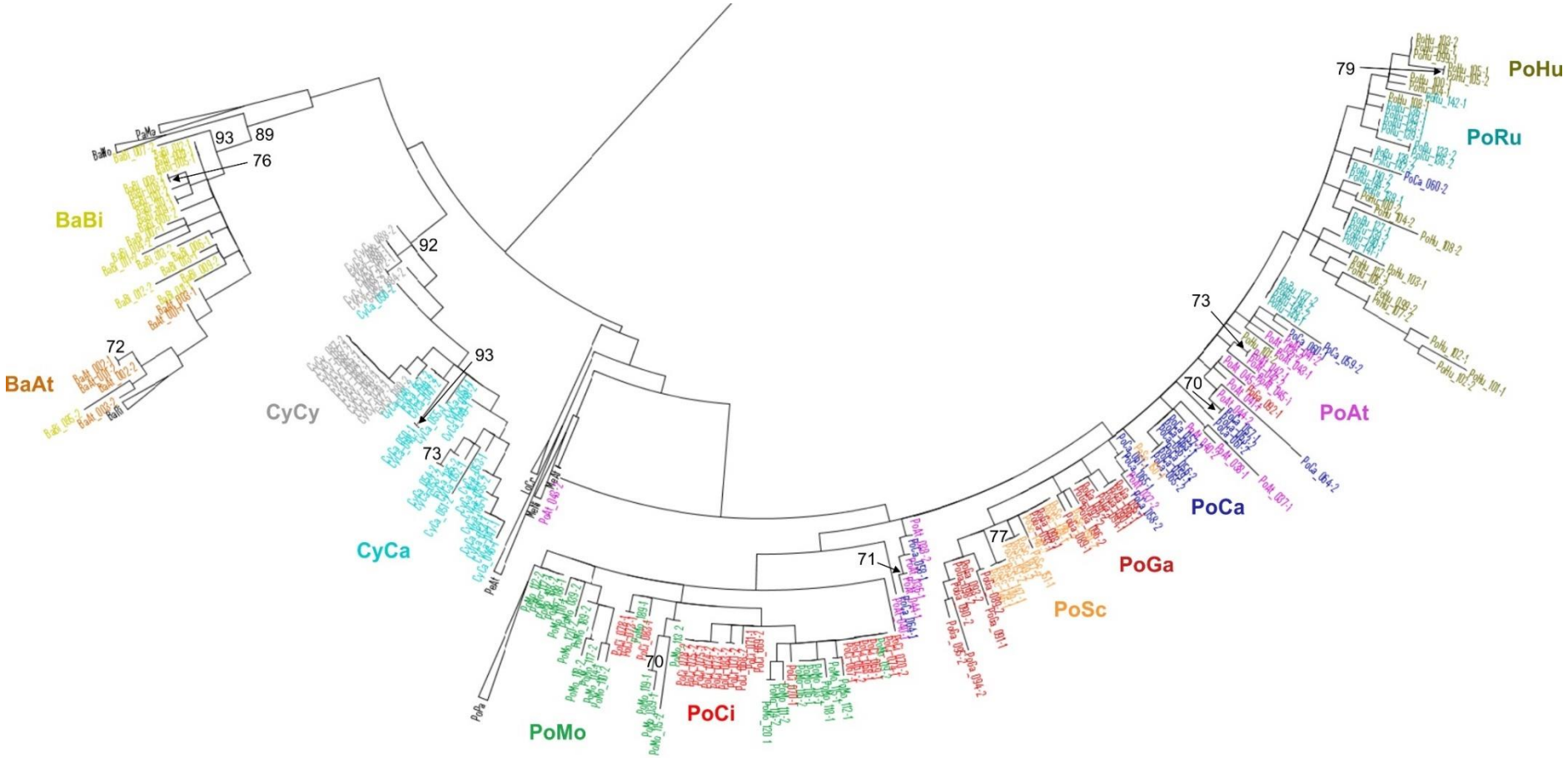


Figure 22: Maximum likelihood tree of *TLR5* in Paridae

The branches without shared alleles among species are condensed. The bootstrap values only above 70 are shown.

4.7 AMOVA

On the basis of haplotype networks (Figure 13-Figure 20) and phylogenetic tree (Figure 3) we picked species which shared *TLR4* and *TLR5* alleles and divided them into two groups. The first group included all 10 species sharing *TLRs* nucleotide alleles (*B. atrcicristatus, B. bicolor, C. caeruleus, C. cyanus, P. atricapillus, P. carolinensis, P. gambeli, P. hudsonicus, P. scalteri* and *P. rufescens*), and the second group was defined as a subset of the first group containing only chickadees (*P. atricapillus, P. carolinensis, P. gambeli, P. hudsonicus, P. scalteri* and *P. rufescens*). Then AMOVA was perfomed for each locus separately and the results are graphically presented in Figure 23, Figure 24 and in summary AMOVA tables Table 18 and Table 19. Although not statistically tested, no obvious differences are apparent between *TLR*s and neutral markers in terms of proportion of explained variability within the species and among species.

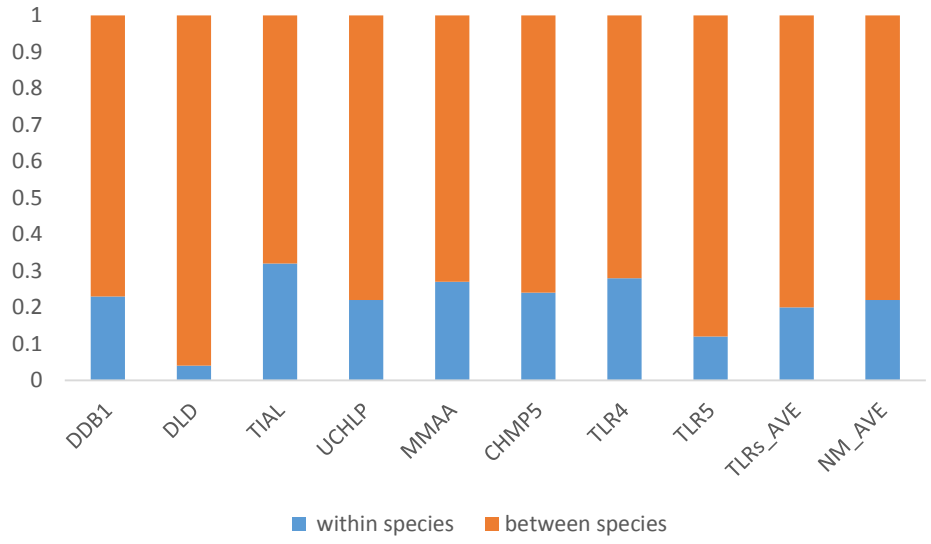


Figure 23: Proportion of variability explained by AMOVA on intraspecies and interspecies level for 10 selected species which shared alleles

TLRs_AVE and NM_AVE mean average value for TLRs and neutral markers, respectively.

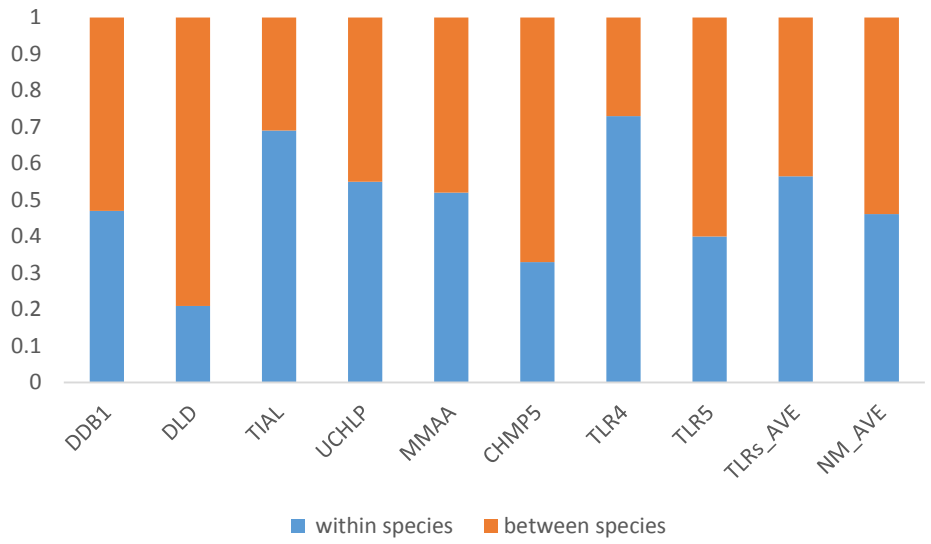


Figure 24: Proportion of variability explained by AMOVA on intraspecies and interspecies level for 6 selected *Poecile* species which shared alleles

TLRs_AVE and NM_AVE mean average value for *TLR*s and neutral markers, respectively.

Table 18: Summary AMOVA tables for 10 tested species

Degree of freedom (df), sum of squares (SS), estimated variability explained by the model (Est. Var.) and estimated variability explained by the model in % (Est. Var. %)

Gene	Structure	df	SS	Est. Var.	Est. Var. %	PhiPT	P value
TLR4	**among spec.**	9	426.173	2.654	72%		
	within spec.	166	171.383	1.032	28%		
	total	175	597.557	3.686	100%	0.72	<0.001
TLR5	**among spec.**	9	1012.229	6.695	88%		
	within spec.	158	150.075	0.950	12%		
	total	167	1162.304	7.645	100%	0.876	<0.001
DDB1	**among spec.**	7	254.074	2.218	77%		
	within spec.	122	80.372	0.659	23%		
	total	129	334.446	2.876	100%	0.771	<0.001
DLD	**among spec.**	9	803.949	7.522	96%		
	within spec.	111	36.745	0.331	4%		
	total	120	840.694	7.853	100%	0.958	<0.001
TIAL	**among spec.**	9	319.668	2.016	68%		
	within spec.	163	156.436	0.96	32%		
	total	172	476.104	2.976	100%	0.678	<0.001
UCHLP3	**among spec.**	9	451.486	3.147	78%		
	within spec.	148	127.722	0.863	22%		
	total	157	579.209	4.01	100%	0.785	<0.001
MMAA	**among spec.**	9	405.928	2.572	73%		
	within spec.	163	158.442	0.972	27%		
	total	172	564.37	3.544	100%	0.480	<0.001
CHMP5	**among spec.**	9	162.818	1.054	76%		
	within spec.	160	54.606	0.341	24%		
	total	169	217.424	1.395	100%	0.755	<0.001

Table 19: Summary AMOVA tables for 6 tested species of genus *Poecile*

Degree of freedom (df), sum of squares (SS), estimated variability explained by the model (Est. Var) and estimated variability explained by the model in % (Est. Var. %).

Gene	Structure	df	SS	Est. Var.	Est. Var. %	PhiPT	P value
TLR4	**among spec.**	5	72.942	0.719	27%		
	within spec.	100	194.633	1.946	73%		
	total	105	267.575	2.665	100%	0.270	<0.001
TLR5	**among spec.**	5	312.353	3.553	60%		
	within spec.	96	224.372	2.337	40%		
	total	101	536.725	5.890	100%	0.603	<0.001
DDB1	**among spec.**	5	77.500	0.839	53%		
	within spec.	100	74.094	0.741	47%		
	total	105	151.594	1.580	100%	0.531	<0.001
DLD	**among spec.**	5	70.205	1.298	79%		
	within spec.	60	20.917	0.349	21%		
	total	65	91.121	1.647	100%	0.788	<0.001
TIAL	**among spec.**	5	39.321	0.396	31%		
	within spec.	100	89.689	0.897	69%		
	total	105	129.009	1.293	100%	0.306	<0.001
UCHLP3	**among spec.**	5	75.772	0.883	45%		
	within spec.	90	97.707	1.086	55%		
	total	95	173.479	1.969	100%	0.449	<0.001
MMAA	**among spec.**	5	80.549	0.863	48%		
	within spec.	100	93.611	0.936	52%		
	total	105	174.160	1.799	100%	0.480	<0.001
CHMP5	**among spec.**	5	104.351	1.180	67%		
	within spec.	98	56.467	0.576	33%		
	total	103	160.817	1.756	100%	0.672	<0.001

4.8 Isolation with migration model for more than to populations

Three isolation with migration models were applied: model 1 - between *C. caeruleus* and *C. cyanus* (Table 20 and Figure 25), model 2 - among *P. atricapillus*, *P.carolinensis* and *P. gambeli* (Table 21 and Figure 26) and *P. cinctus*, *P. hudsonicus*, *P.scalteri* and *P. rufescens* (Table 22 and Figure 27). All three runs for each dataset were convergent and we achieved adequate chain mixing as indicated by effective sample size values (ESS) in first two models (all ESS were always higher than 500). However,several ESS were lower than 30 in the third model indicating a mixing problem. Moreover, the estimates of time divergence were less reliable in all models as indicated by trend-line plots with plotted posterior probability distribution (e.g. in model 1 and model 3). This holds also for other estimates of parameters which had often relatively "wide" confidence interval. It is probably caused by too low number of neutral markers, especially apparent for model 3 with four species. Therefore, these results, from model 3 in particular, must be interpreted with caution.

Significant gene flow was detected only between *C. Cyanistes* and *C. caeruleus* and between *P. atricapillus* and *P. carolinensis*. Although occurring in both directions, the gene flow appears to be asymmetric - prevailing in direction from *P. atricapillus* to *P. carolinensis* (2Nm=1.937000) rather than in the opposite direction (2*Nm* = 0.529500). On the contrary, between *C. caeruleus* and *C. cyanus* the gene flow occurs preferentially in the direction from *C. caeruleus* (2*Nm* = 0.288500), while the flow is negligible in the opposite direction (2*Nm* = 0.005631). The latter species also differ more in their effective population size which is estimated to be much higher in *C. caeruleus* (N_e=470 602) than in *C. cyanus* (N_e = 116 309).

4.8.1 Model 1: *Cyanistes caeruleus* and *Cyanistes cyanus*

Table 20: Maximum-likelihood estimates (MLE) and 95% highest posterior density (HPD) intervals of demographic parameters for model 1

Parameter	MLE	HPD95Lo	HPD95Hi
N_0	470602	259458	814160
N_1	116309	41155	270194
N_2	130623	0	3362213
t_0	1395703	851737	14307745
$2N_0m_0>1$	0.005631	0.000000	1.199
$2N_1m_1>0$	0.288500	0.068530	0.719
$m_0>1$	1.746E-09	0	0
$m_1>0$	1.034E-06	1.01E-07	0

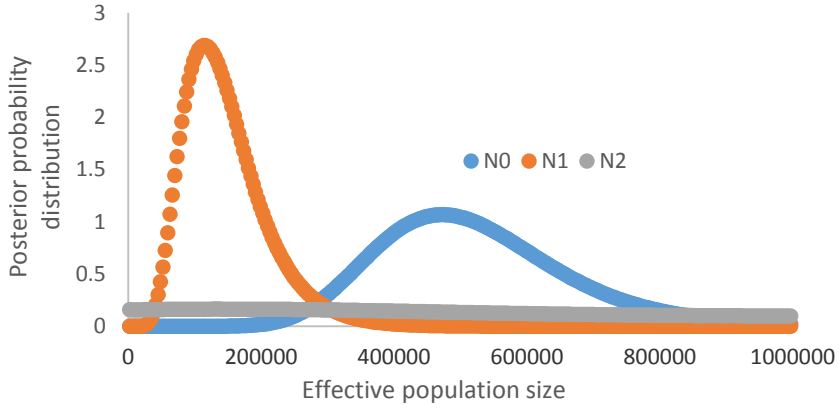
N, effective population size for *C. caeruleus* (N_0), for *C. cyanus* (N_1) and the ancestral population (N_2)

t, time of divergence in MY between *C. cyanus* to *C. caeruleus* (t_0)

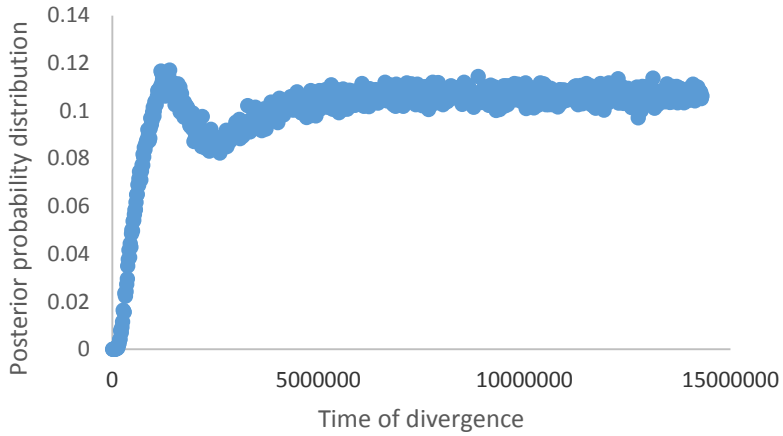
2*Nm*, population migration rate from *C. cyanus* to *C. caeruleus* ($2N_0m_0>1$) and from *C. caeruleus* to *C. cyanus* ($2N_1m_1>0$)

m, migration rate per year from *C. cyanus* to *C. caeruleus* ($m_2>1$) and from *C. caeruleus* to *C. cyanus* ($m_1>0$).

A)



B)



C)

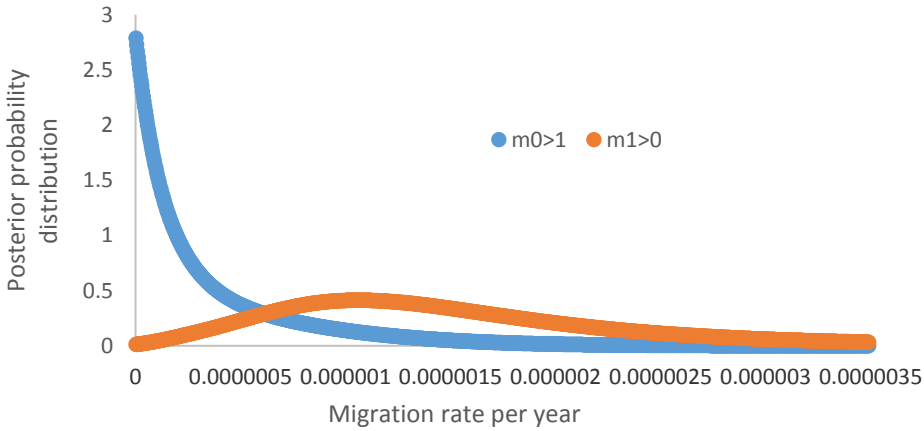


Figure 25: The marginal posterior probability distributions for the demographic parameters of the IMa2 for model 1

A) Effective population size is shown for *C. caeruleus* (N_0), *C. cyanus* (N_1) and ancestral population (N_2); B) Time of divergence in MY between *C. caeulus* and *C. cyanus*; C) Migration rates per year from *C. caeruleus* to *C. cyanus* (m_1>0) and from *C. cyanus* to *C. caeruleus* (m_0>1).

4.8.2 Model 2: *P. atricapillus*, *P. carolinensis* and *P. gambeli*

Table 21: Maximum-likelihood estimates (MLE) and 95% highest posterior density (HPD) intervals of demographic parameters for model 2

Parameter	MLE	HPD95Lo	HPD95Hi
N_0	460061	191509	948739
N_1	746224	349999	1543076
N_2	636161	363206	1094021
N_3	2201	0	3731118
N_4	46226	0	4114136
t_0	1998735	942135	3284265
t_1	2491815	1840245	17601193
$2N_0m_0>1$	0.529500	0	1.763
$2N_0m_0>2$	0.001006	0	0.220
$2N_1m_1>0$	1.937000	0	5.285
$2N_1m_1>2$	0.002844	0	0.299
$2N_2m_2>0$	0.002894	0	0.171
$2N_2m_2>1$	0.002994	0	0.195
$2N_2m_2>3$	0.008219	0	4.808
$2N_3m_3>2$	0.012490	0	14.930
$m_0>1$	1.661E-07	0	2.3754E-06
$m_1>0$	1.341E-06	4.046E-07	2.8268E-06
$m_0>2$	1.42E-09	0	2.3708E-07
$m_2>0$	1.42E-09	0	1.2919E-07
$m_1>2$	1.42E-09	0	1.803E-07
$m_2>1$	1.42E-09	0	1.4338E-07
$m_2>3$	1.42E-09	0	2.8376E-06
$m_3>2$	1.562E-08	0	2.8376E-06

N, effective population size for *P. atricapillus* (N_0), *P. carolinensisP. carolinensis* (N_1), *P. gambeli* (N_2) and ancestral populations between *P. atricapillus* and *P. gambeli* (N_3) and common ancestor for all species (N_4).

t, time of divergence in MY between *P. atricapillus* and *P. carolinensis* (t_0) and between *P. gambeli* and the common ancestor of *P. atricapillus* and *P. carolinensis* (t_1).

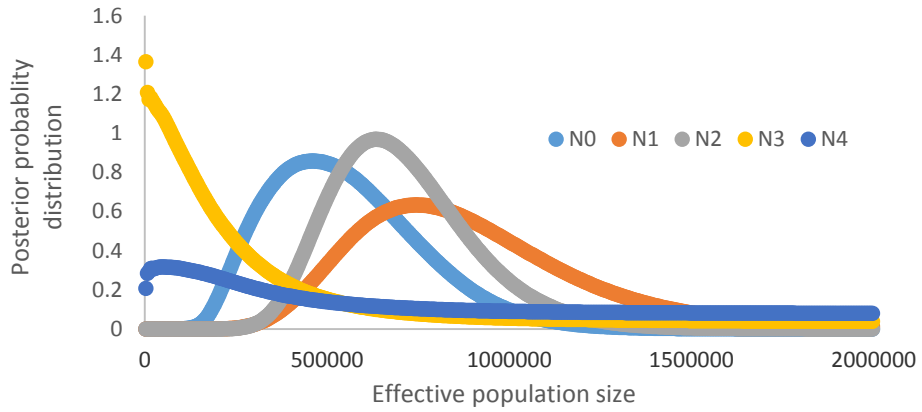
2*Nm*, population migration rate from *P. atricapillus* to *P. carolinensis* ($2N_1m_1>0$) and from *P. carolinensis* to *P. atricapillus* ($2N_0m_0>1$), *P. gambeli* to *P. atricapillus* ($2N_0m_0>2$), from *P. atricapillus* to *P. gambeli* ($2N_2m_2>0$), from *Poecile gambeli* to *P. carolinensis* ($2N_1m_1>2$), from *P. carolinensis* to *P. gambeli* $2N_2m_2>1$), from the common ancestor of *P. atricapillus* and *P. carolinensis* to *P. gambeli* ($2N_2m_2>3$) and from the common ancestor of *P. atricapillus* and *P. carolinensis* to *P. gambeli* ($2N_3m_3>2$).

m, migration rate per year from *P. atricapillus* to *P. carolinensis* ($m_1>0$) and from *P. carolinensis* to *P. atricapillus* ($m_0>1$), *P. gambeli* to *P. atricapillus* ($m_0>2$), from *P. atricapillus* to *P. gambeli* ($m_2>0$), from *Poecile gambeli* to *P. carolinensis* ($m_1>2$), from *P. carolinensis* to *P. gambeli* ($m_2>1$), from the common ancestor of *P. atricapillus* and *P. carolinensis* to *P. gambeli* ($m_2>3$) and from the common ancestor of *P. atricapillus* and *P. carolinensis* to *P. gambeli* ($m_3>2$).

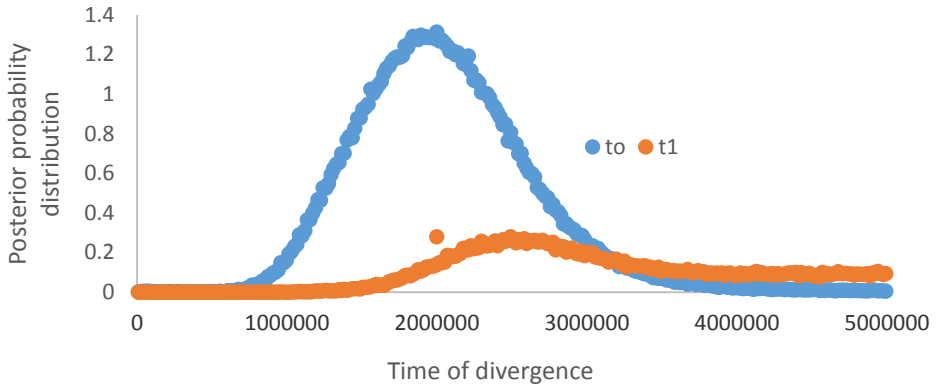
Figure 26: The marginal posterior probability distributions for the demographic parameters of the IMa2 for model 3

A) Effective population size is shown for *P. atricapillus* (N_0), *P. carolinensis* (N_1), *P. gambeli* (N_2), for the ancestral populations between *P. atricapillus* and *P. gambeli* (N_3) and common ancestor for all species (N_4); B) Time of divergence in MY between *P. atricapillus* and *P. carolinensis* (t_0) and *P. gambeli* and the common ancestor of *P. atricapillus* and *P. carolinensis* (t_1); C) Migration rates per year from *P. atricapillus* to *P. carolinensis* (m_1>0) and from *P. carolinensis* to *P. carolinensis* and (m_0>1). D) Migration rates per year from *P. gambeli* to *P. atricapillus* (m_0>2), from *P. atricapillus* to *P. gambeli* (m_2>0), from *Poecile gambeli* to *P. carolinensis* (m_1>2), from*P. carolinensis* to *P. gambeli* (m_2>1), from the common ancestor of *P. atricapillus* and *P. carolinensis* to *P. gambeli* (m_2>3) and from common ancestor of *P. atricapillus* and *P. carolinensis* to *P. gambeli* (m_3>2).

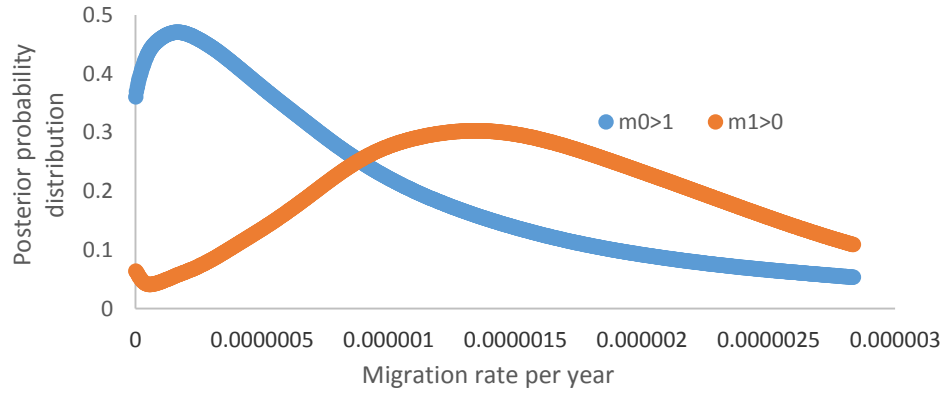
A)



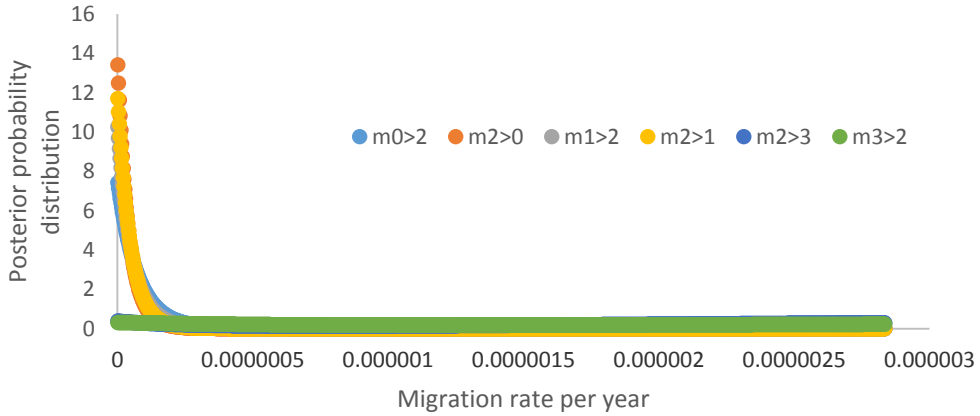
B)



C)



D)



4.8.3 Model 3: *P. rufescens*, *P. hudsonicus*, *P. cinctus* and *P. scalteri*

Table 22: Maximum-likelihood estimates (MLE) and 95% highest posterior density (HPD) intervals of demographic parameters for model 3

Parameter	MLE	HPD95Lo	HPD95Hi	Parameter	MLE	HPD95Lo	HPD95Hi
N_0	189943	48954	2692486	$m_0>1$	1.59588E-09	0	3.18984E-06
N_1	303517	99867	973211	$m_1>0$	3.18984E-06	0	3.18984E-06
N_2	41122	13707	123365	$m_0>2$	1.59588E-09	0	3.14388E-07
N_3	589410	327015	1024124	$m_2>0$	1.59588E-09	0	1.30351E-06
N_4	21540	0	3448340	$m_0>3$	1.59588E-09	0	1.77142E-07
N_5	5875	0	3702903	$m_5>0$	1.59588E-09	0	2.37786E-07
N_6	620740	21540	3730317	$m_1>2$	1.59588E-09	0	3.87798E-07
t_0	211483	54829	4488130	$m_2>1$	1.59588E-09	0	1.39671E-06
t_1	4535126	885094	12916100	$m_1>3$	1.59588E-09	0	2.34594E-07
t_2	15610545	4347141	15657541	$m_3>1$	1.59588E-09	0	2.44169E-07
$2N_0m_0>1$	0.012490	0	5.810	$m_2>3$	1.59588E-09	0	5.09085E-07
$2N_0m_0>2$	0.006506	0	0.475	$m_3>2$	1.59588E-09	0	1.45225E-07
$2N_0m_0>3$	0.005531	0	0.304	$m_2>4$	9.43482E-07	0	2.97982E-06
$2N_1m_1>0$	0.012490	0	4.010	$m_4>2$	1.59588E-09	0	3.18984E-06
$2N_1m_1>2$	0.003506	0	0.340	$m_3>4$	1.59588E-09	0	3.18984E-06
$2N_1m_1>3$	0.003231	0	0.210	$m_4>3$	1.59588E-09	0	3.18984E-06
$2N_2m_2>0$	0.000544	0	0.131	$m_3>5$	1.59588E-09	0	3.18984E-06
$2N_2m_2>1$	0.000606	0	0.136	$m_5>3$	3.18984E-06	0	3.18984E-06
$2N_2m_2>3$	0.000294	0	0.052				
$2N_2m_2>4$	0.078930	0	0.449				
$2N_3m_3>0$	0.003631	0	0.301				
$2N_3m_3>1$	0.003331	0	0.310				
$2N_3m_3>2$	0.001044	0	0.187				
$2N_3m_3>4$	0.011270	0	4.293				
$2N_3m_3>5$	0.009994	0	4.567				
$2N_4m_4>2$	0.012490	0	15.400				
$2N_4m_4>3$	0.012490	0	13.280				
$2N_5m_5>3$	0.012490	0	17.430				

N, effective population size is shown for *P. rufescens* (N_0), *P. hudsonicus* (N_1), *P. cinctus* (N_2), *P. scalteri* (N_3), for the ancestral populations between *P. rufescens* and *P. hudsonicus* (N_4), for the common ancestor of *P. rufescens*, *P. hudsonicus*, *P. cinctus* (N_5) and for the common ancestor of all species (N_6)

t, time of divergence in MY between *P. rufescens* and *P. hudsonicus* (t_0), *P. cinctus* and the common ancestor of *P. rufescens*, *P. hudsonicus* (t_1) and *P. scalteri* and the common ancestor of *P. rufescens*, *P. hudsonicus*, *P. cinctus* (t_2)

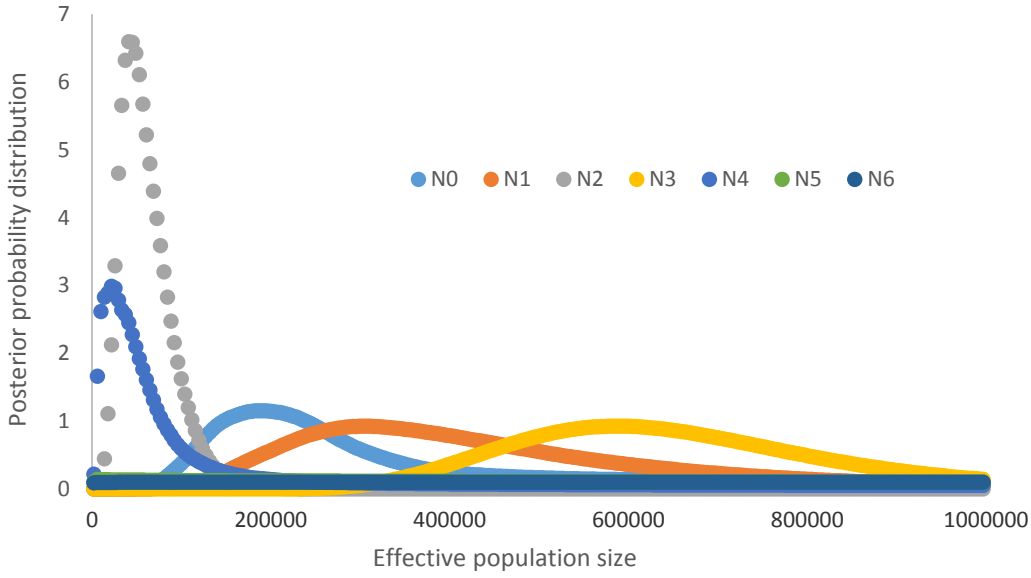
$2Nm$, population migration rate from *P. hudsonicus* to *P. rufescens* ($2N_0m_0>1$), from *P. rufescens* to *P. hudsonicus* ($2N_1m_1>0$), from *P. cinctus* to *P. rufescens* to ($2N_0m_0>2$), from *P. rufescens* to *P. cinctus* ($2N_2m_2>0$), from *P. rufescens* to *P. scalteri* ($2N_0m_0>3$) and from *P. scalteri* to *P. rufescens* ($2N_3m_3>0$), from *P. cinctus* to *P. hudsonicus* ($2N_1m_1>2$), from *P. hudsonicus* to *P. cinctus* ($2N_2m_2>1$), from P.scalteri to *P. rufescens* ($2N_1m_1>3$), from *P. rufescens* to *P. scalteri* ($2N_3m_3>1$), *P. scalteri* to *P. cinctus* ($2N_2m_2>3$), from *P. cinctus* to *P. scalteri* ($2N_3m_3>2$), for migration between ancestral populations it follows the same logic as for numbering of N_e.

m, migration rate per year follows the same rule as for numbering of $2Nm$.

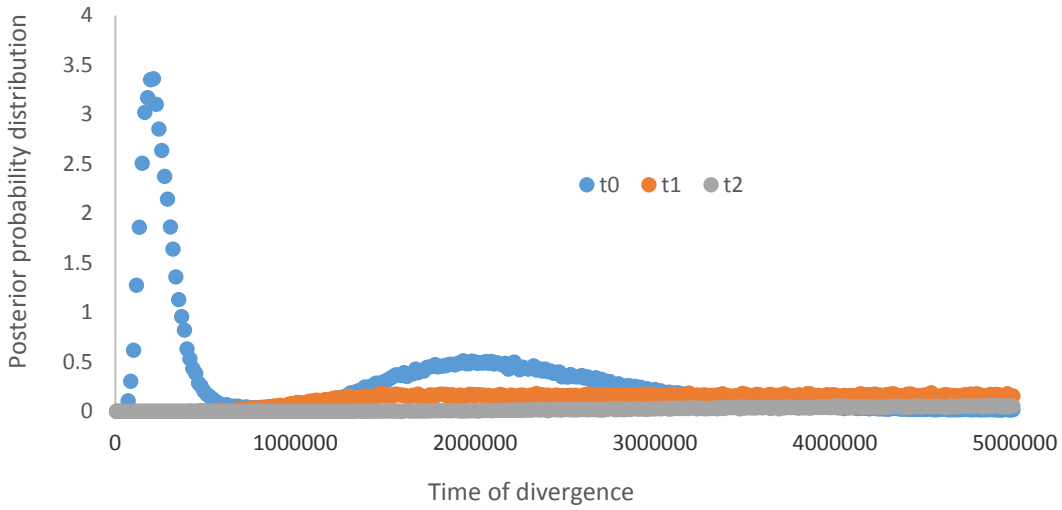
Figure 27: The marginal posterior probability distributions for the demographic parameters of the IMa2 model for model 3

A) Effective population size is shown for *P. rufescens* (N_0), *P. hudsonicus* (N_1), *P. cinctus* (N_2), *P. scalteri* (N_3), for the ancestral populations between *P. rufescens* and *P. hudsonicus* (N_4), for the common ancestor of *P. rufescens*, *P. hudsonicus*, *P. cinctus* (N_5) and for the common ancestor of all species (N_6) *P. rufescens, P. hudsonicus, P. cinctus, P. scalteri*; B) Time of divergence in MY between *P. rufescens* and *P. hudsonicus* (t_0), *P. cinctus* and the common ancestor of *P. rufescens*, *P. hudsonicus* (t_1) and *P. scalteri* and the common ancestor of *P. rufescens*, *P. hudsonicus*, *P. cinctus* (t_2); C) Migration rates per year from *P. hudsonicus* to *P. rufescens* ($m_0>1$), from *P. rufescens* to *P. hudsonicus* ($m_1>0$), from *P. cinctus* to *P. rufescens* to ($m_0>2$), from *P. rufescens* to *P. cinctus* ($m_2>0$), from *P. rufescens* to *P. scalteri* ($m_0>3$) and from *P. scalteri* to *P. rufescens* ($m_3<0$), from *P. cinctus* to *P. hudsonicus* ($m_1>2$), from *P. hudsonicus* to *P. cinctus* ($m_2>1$), from *P. scalteri* to *P. rufescens* ($m_1>3$), from *P. rufescens* to *P. scalteri* ($m_3>1$), *P. scalteri* to *P. cinctus* ($m_2>3$), from *P. cinctus* to *P. scalteri* ($m_3>2$), for migration between ancestral populations it follows the same logic as for numbering of N_e.

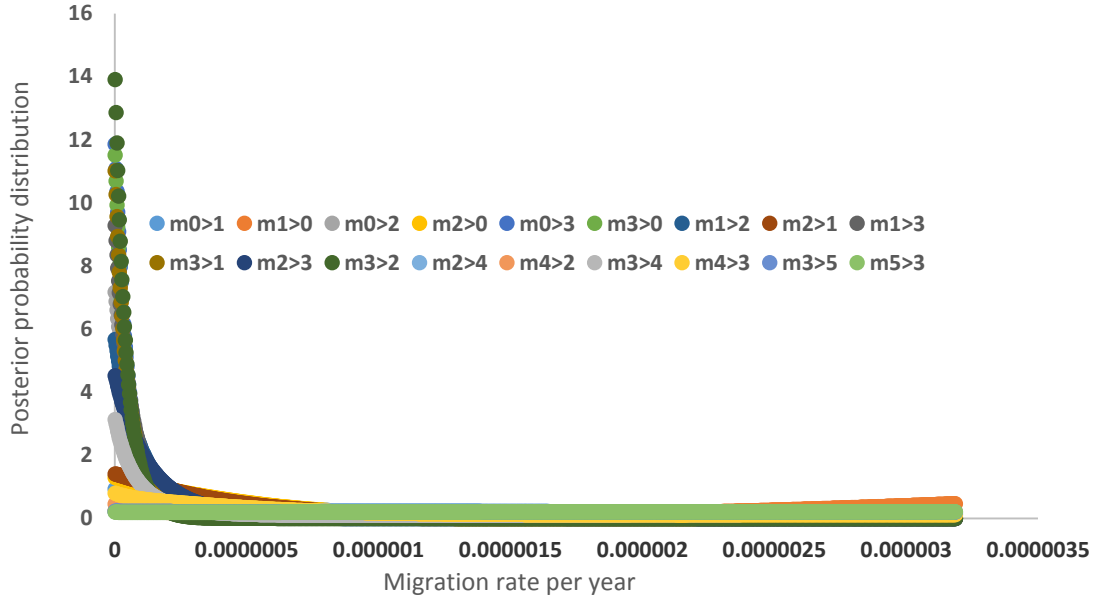
A)



B)



C)



4.9 Electrostatic surface charge analysis

The results from PIPSA analysis are visualized in the electrostatic matrix and epograms for both TLR4 (Figure 28 and Figure 29) and TLR5 (Figure 31 and Figure 32) and as partial electrostatic on 3D model from I-TASSER for TLR4 (Figure 30) and TLR5 (Figure 33) Although the highest similarities in surface electrostatics exist mostly in closely related species, e.g. *Melaniparus* in TLR4 or *Cyanistes* in TLR5, these species are more dissimilar in TLR5 and TLR4. *P. gambeli* and *P. scalteri* share identical alleles in *TLR5* on amino acid level resulting in identity of elestrotatic surface charge for these particular alleles.

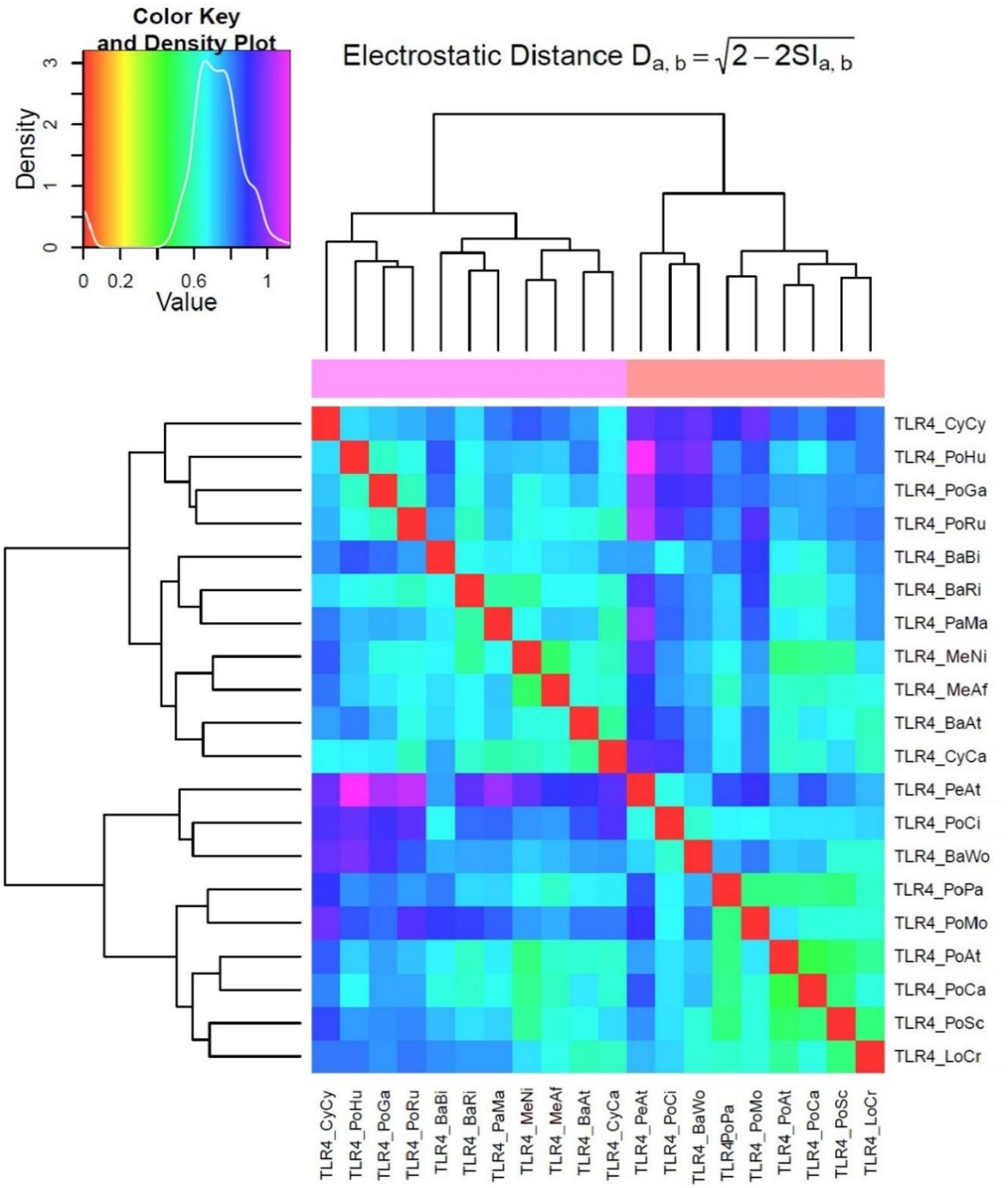


Figure 28: Electrostatic distance matrix (heat map) of overall surface charge in TLR4 from PIPSA analysis

The pairwise comparison is done for all species. The degree of similarities is shown in colour gradient ranging from red for the most similar (positively correlated) to violet for the most dissimilar (anti-correlated).

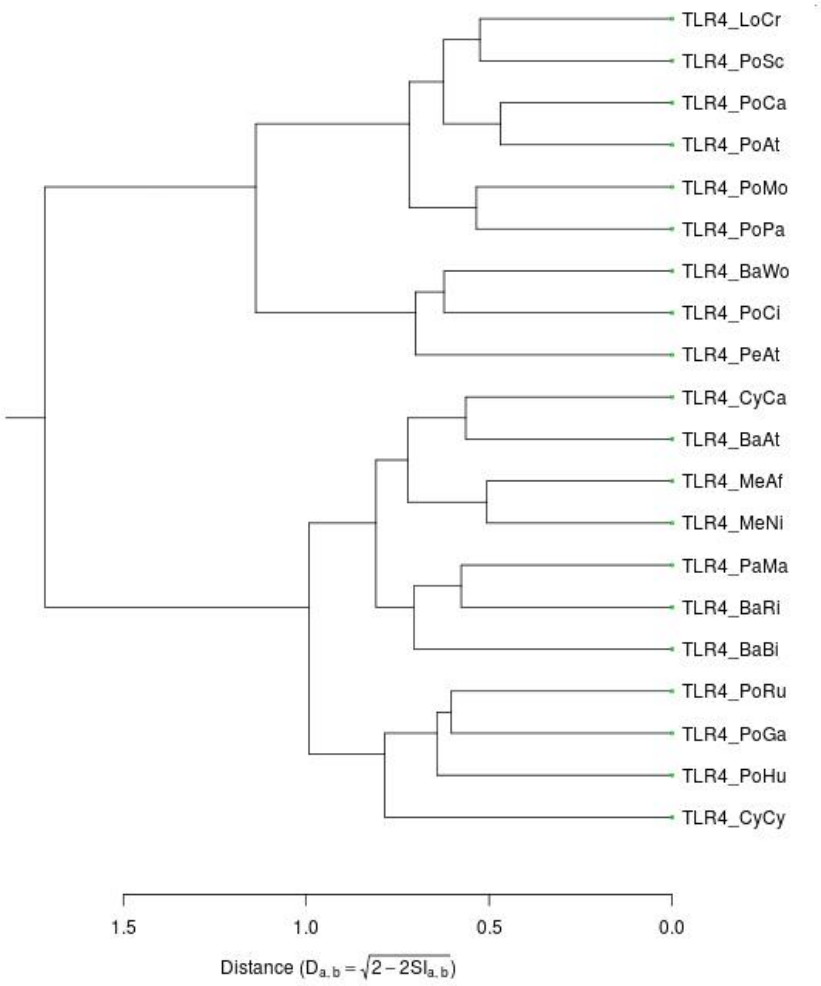


Figure 29: The cluster dendrogram (epogram) of electrostatic surface charge from PIPSA for TLR4

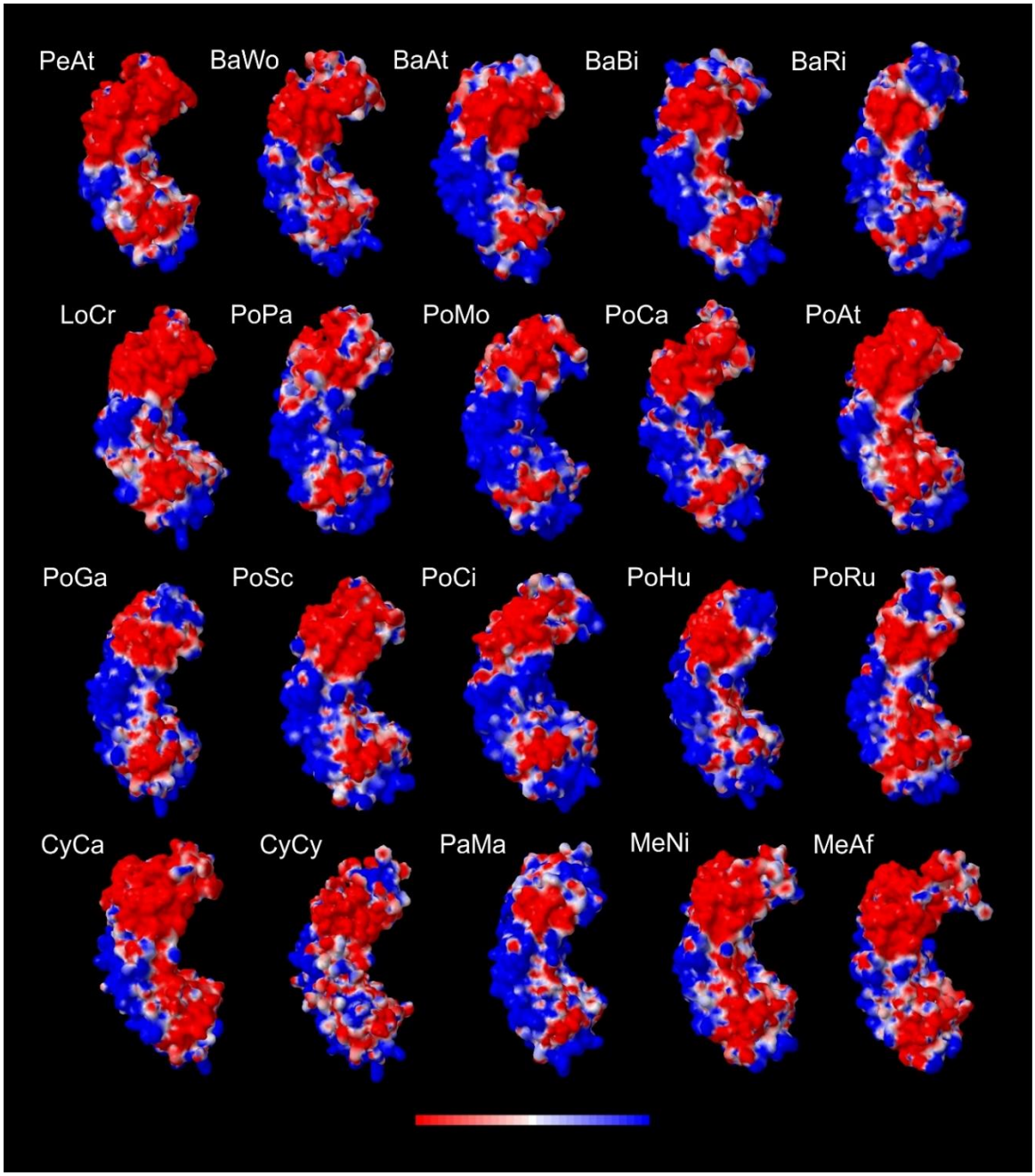


Figure 30: Visualisation of partial electrostatic surface charge of TLR4 in different species

Partial electrostatic charge is shown in colour gradient ranging from red for the most negative charge to blue for the most positive charge.

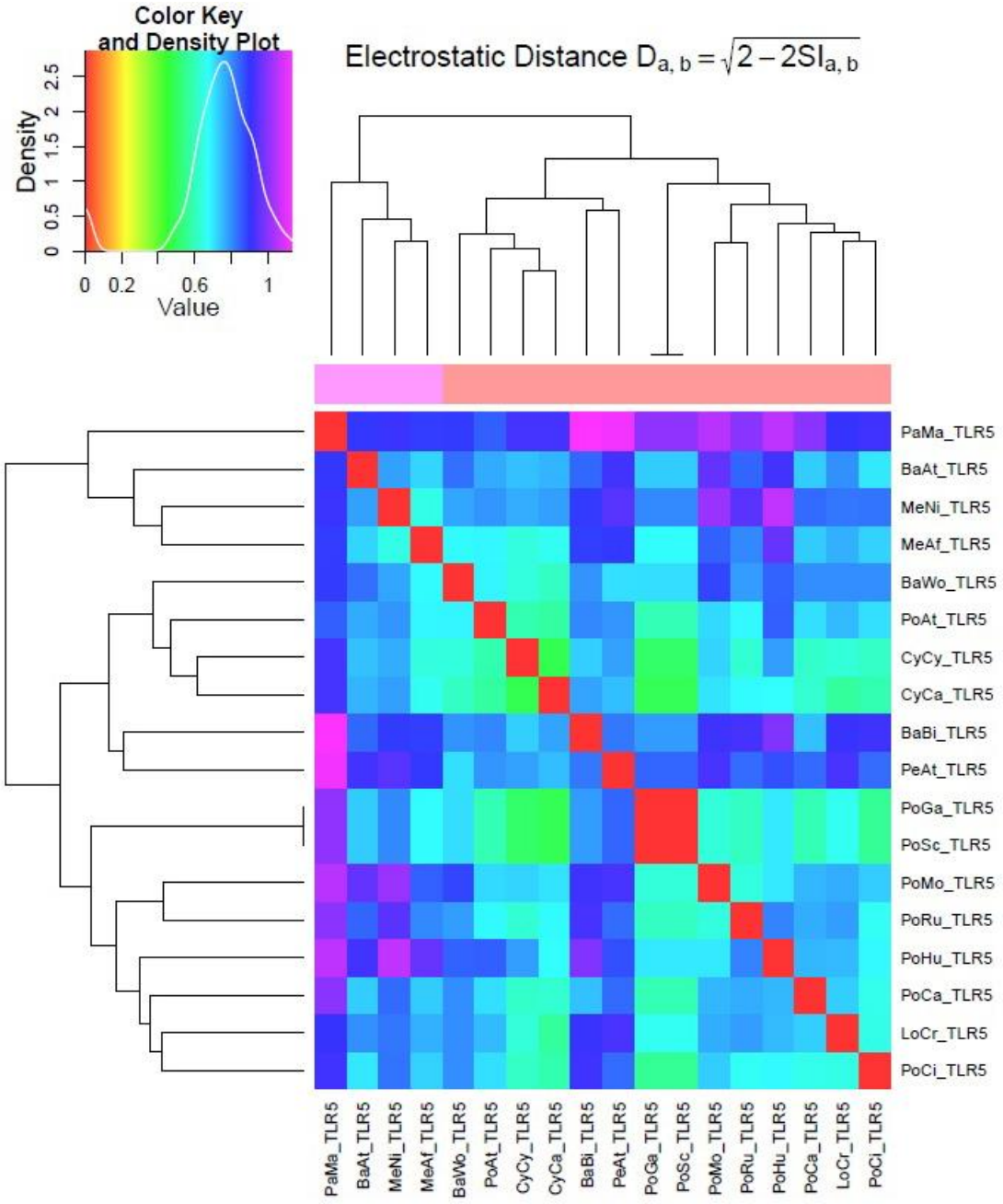


Figure 31: Electrostatic distance matrix of overall surface charge in TLR5 from PIPSA analysis

The pairwise comparison is done for all species. The degree of similarities is shown in colour gradient ranging from red for the most similar (positively correlated) to violet for the most dissimilar (anti-correlated).

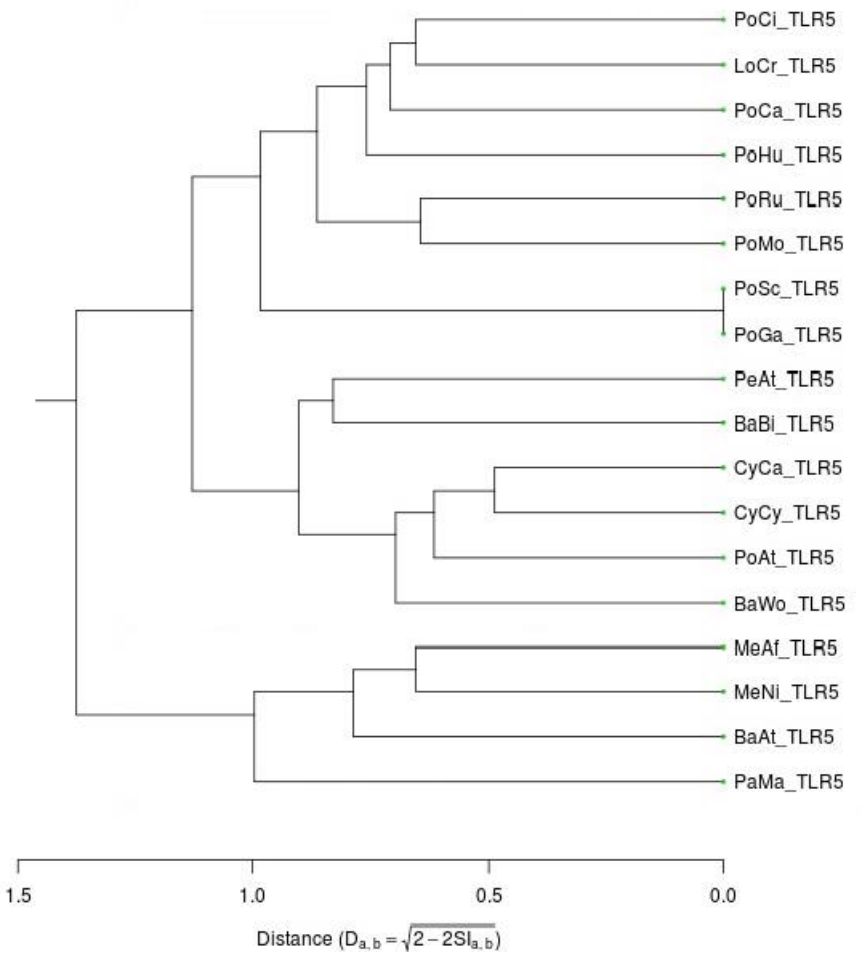


Figure 32: The cluster dendrogram (epogram) of electrostatic surface charge from PIPSA for TLR5

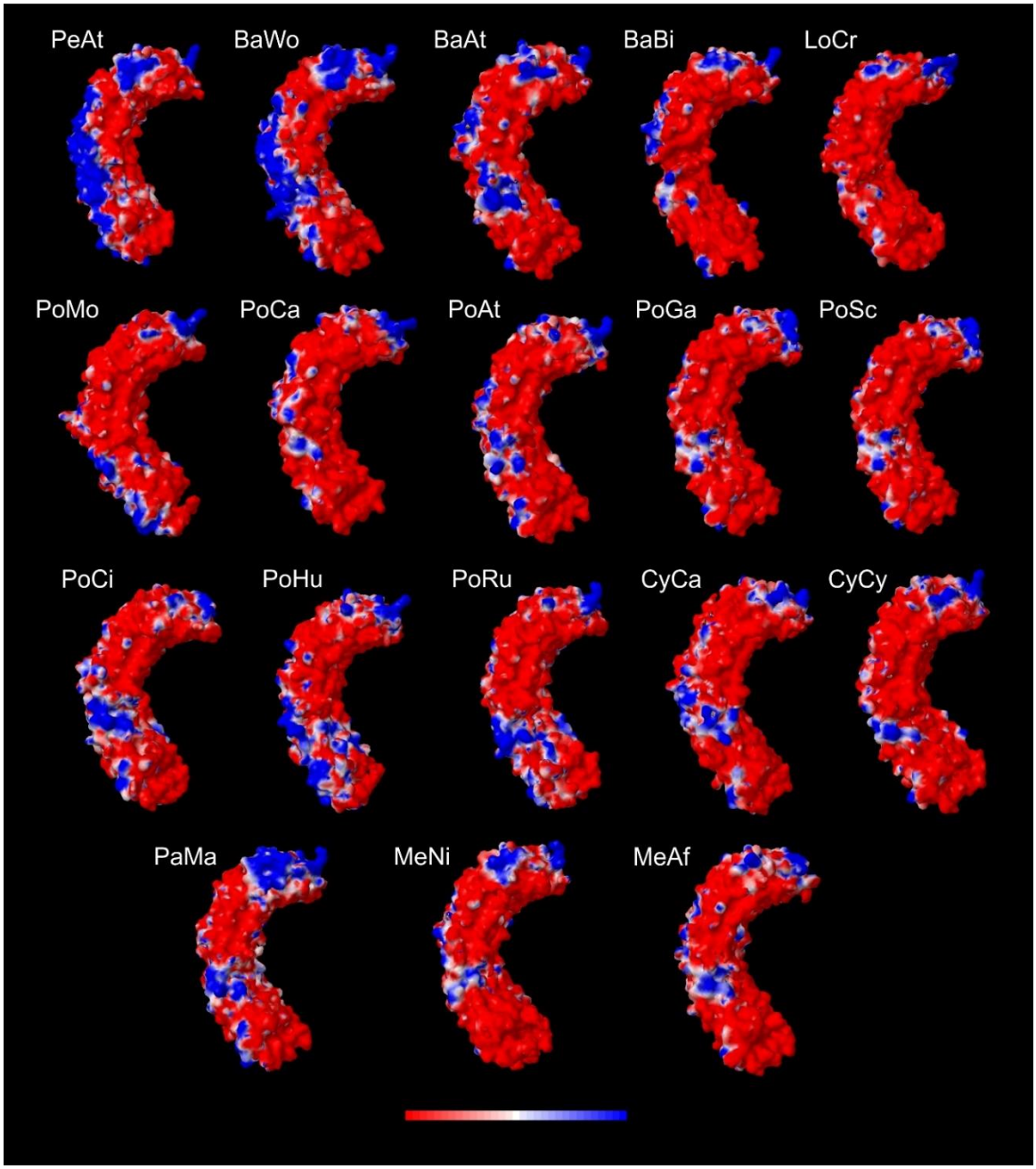


Figure 33: Visualisation of partial electrostatic surface charge of TLR5 in different species

Partial electrostatic charge is shown in colour gradient ranging from red for the most negative charge to blue for the most positive charge.

5 Discussion

In this study we sequenced the ligand-binding and dimerization regions of *TLR4* and *TLR5* and six neutral autosomal markers in 192 individuals representing 20 species of tits, chickadees and titmice (Paridae family). First, I described genetic polymorphism and calculated basic population genetic characteristics for *TLR4*, *TLR5* and for the neutral markers. Both *TLRs* and neutral markers were variable on intra- and interspecific level thought differencies exist mainly among different neutral markers. To reveal phylogenetic relationship among alleles, both nucleotide and amino acid networks as well as phylogenetic trees were constructed for the *TLR*s. For neutral markers nucleotide haplotype networks were constructed. Sharing of alleles both in the *TLR*s and neutral markers was detected only in closely related species within genus level. I confirmed positive selection acting on the *TLR* genes as a necessary assumption for the putative TSP. Most of these sites were also considered to be the most non-conservative by ConSurf analysis and with non-conservative substitutions occurred in different lineages. Surface electrostatic charge analysis of TLR4 and TLR5 molecules was done to seek for functional efects of TLR sequence variation. The results show reasonable levels of variation in the protein surface charges, suggesting also variation in ligand-binding features of tits. Based on these results, I further focused on identifying the origin of shared variability. To distinguish putative balanced polymorphism from neutral ones, population structure analysis (AMOVA) was applied. However, I did not show any consistent differences in the variability explained among species in *TLR*s and neutral markers. Therefore, in concordance with other results we did not see any indication for strong pervasive balancing selection in *TLR*s. Afterwards potential gene flow among several closely related species was evaluated by using IMa2. Although gene flow was detected only between two closely related species, our results indicate that introgression could be as plausible evolutionary scenario for explaining the shared variability in *TLR4* and *TLR5 as TSP*.

5.1 Polymorphism in TLR4, TLR5 and neutral markers

TLRs were variable on both intra- and interspecific level comparably with polymorphism identified in other studies, e.g. Murinae (Fornuskova et al., 2013), Galloanseres (Vinkler et al., 2014), Anthidae (Gonzalez-Quevedo et al., 2015) or New Zealand birds (Grueber et al., 2014). Although noticeable differences among in number of nucleotide haplotypes and segregating sites exist among different species, their mean nucleotide diversity, did not statistically differ from neutral markers (p=0.221461). This is interesting since it shows that TLRs as coding genes in which major part of the molecule is under negative selection exhibit similar overall variation as neutral sequences that are free to mutations. Hence, while TLRs are functionally

constrained strong diversifying selection occurs in several particular residues that are likely responsible for ligand binding (Fornuskova et al., 2013).

5.2 Detection of positive selection, evolutionary non-conservative sites and recombination in TLR4 and TLR5

Positive (diversifying) selection was revealed in both *TLR4* and *TLR5* genes on interspecies level. 14 residues in *TLR4* and 23 residues in *TLR5* in total were detected by at least one selection method. From the total number of positively selected sites in *TLR4*, four positions (320, 374, 397,427) lie in close proximity to mammal predicted binding sites. Simultaneously, all these positions have non-conservative substitutions with changes in charge or polarity (Table 12) and they were also classified by ConSurf analysis under the category with the most non-conservative residues (Table 15). Therefore, they may influence binding properties of TLR4. This is truth especially for the position S397G,R, which was detected by all four selection methods and which has also the highest conservation score (considered to be the most non-conservative in *TLR4*).

From the total of 23 identified positively selected sites in *TLR5* 14 positions (33, 35, 56, 80, 106, 156, 181, 183, 209, 213, 237, 261, 294, 379) were located either in the close proximity to the predicted functional binding sites or were directly in the binding sites. Besides that, excluding position 35 all these positions were considered to be the most non-conservative by ConSurf analysis (Table 16), as well as with non-conservative substitutions in different lineages (Table 14). Furthermore, from these 14 positions 10 residues 33, 35, 56, 80, 156, 106, 181, 183, 209, 379 were directly predicted binding sites for flagelline (Yoon et al., 2013). From these 10 flagelline binding sites 8 positions (33, 35, 56, 80, 106, 183, 209, 379) were variable also in Galloaneseres usually with similar substitutions suggesting convergent evolution in passerines and galliform birds (see Supplement 9). Moreover, not only positively selected sites may play role in ligand binding but also other evolutionary non-conservative sites indentified by ConSurf analysis located in close proximity to predicted binding sites in both TLR4 and TLR5 might influence binding properties (Table 15 and Table 16).

I did not detect significantly recent positive selection using Tajima's D test and Fu and Li' s test on intraspecies level, although on interspecies level the positive selection was detected. The discrepancy between the selection methods applied when seeking selection on intraspecies and intraspecies can be explained by the fact, that Tajima's D test and Fu and Li' s test reveal recent selective sweeps and operate with allele's frequency in population (Nielsen, 2005), while positive selection identified by MEME, PAML, FUBAR and REL show much older diversifing selection (between lineages of species). On the other hand, some of these tests might also overestimate number of identified residues, e.g. REL

might be sensitive to false positive signal, and therefore only positions identified by several methods are typically reliable (Wlasiuk and Nachman, 2010).

To conclude, in TLR4 and TLR5 we identified several functionally important binding sites based on the consensus of different methods: tests of positive selection, ConSurf analysis and by comparing conservative and non-conservative substitutions in different tit lineages and final evalvulation of their locations on 3D models. Given the consensus of aforementioned attitudes, these sites may affect binding properties of TLR4 and TLR5 and thus recognition of pathogens. Nevertheless, it is important to note that there is no crystallographic structure of bird TLR4 and TLR5 and our predictions are, therefore, derived by using both homolog modelling and known fuctionally important sites of mammal *TLR4* (Kim et al., 2007; Park et al., 2009) and of mammal (Andersen-Nissen et al., 2007) and fish (Yoon et al., 2013) *TLR5*. Especially position of flagelline binding sites differs between mammals and fish.

The recombination breakpoints were detected only in aa position 471 in *TLR4* and 1039 in *TLR5*. Despite the fact that only one recombination breakpoint occurred in these genes, it could have potential impact on maintaining TSP. We hypothesise that recombination may disrupt TSP maintained in the long term and it would be interesting to compare TSP in sequences after filtering out recombination. Instead of whole alleles as defined here, shorter recombinations blocks might be favour in the long-term scale.

5.3 TSP in Paridae

We have identified sharing of alleles between tit species in both *TLR4* and *TLR5*. Although several evolutionary mechanisms have been proposed to explain shared polymorphism (e.g. introgression and convergence; see Chapter 1.3), the shared variability identified here is in most cases TSP. As far as we know, this is the very first evidence for TSP in TLR family and in PRRs in general. As TSP oriented research has been stereotypically focused mainly on MHC genes for a long time (Těšický and Vinkler, 2015), this is another case of TSP identified in innate immunity genes and another piece of evidence that TSP is a general evolutionary phenomenon explaining the origin of shared beneficious variability. However, TSP identified here was restricted only to closely related species: no variability was shared among different genera. This pattern is in contrast to TSP in MHC genes, where TSP alleles are shared as identical alleles or diverged allelic lineages above species level, e.g. among genera (e.g. Bryja et al., 2006; Kriener et al., 2001; Walsh and Friesen, 2003) or even among families (e.g. Go et al., 2005; Sin et al., 2012; Xu et al., 2009, 2008). Therefore, such TSP persists in the order of up to tens millions of years in different taxa (reviewed in Těšický and Vinkler, 2015). Shared polymorphism identified in TLRs here was mainly among American *Poecile*, *Baelophhus atricristatus* and *B. bicolor* and *Cyanistes caeruleus* and *C. cyanus*. The estimates

of the time of persistence of TSP rely on molecular clock. Estimated time of divergence in Paridae significantly differs in connection with the methods used (Gill et al., 2005; Päckert et al., 2007) and in addition to that, there is also high variation in the speed of molecular clock even among closely related tit species (Päckert et al., 2007). Given the estimated time divergence in American Poecile – between 4 MYA (Gill et al., 2005) and 8 MYA (Packert et al., 2007), between *C. caeruleus* and *C. cyanus* 2.5-3 MYA (Illera et al., 2011; Packert et al., 2007) and between *B. bicolor* and *B. atricristatus* >0.25 MYA (Johnson and Cicero, 2004), TSP in TLRs persists in tits no logner than 8 MY, more consevatively 4 MY. Combined with results of Tajima´s D test and Fu and Li's D test in *TLR*s we did not find any signature of strong balancing selection acting on these innate immunity genes. Similarly, average nucleotide diversity and Tajima's D of the *TLR*s did not statistically differ from neutral markers. Identification of convincing balancing selection by these tests, however, often fails in natural populations. On the other hand, shared polymorphism can be considered as a piece of evidence supporting balancing selection operating on these loci in a long-term scale (Klein et al., 1998; Takahata, 1993) while Tajima's D and Fu and Li's D tests are based on comparison of the allele frequency in population and they are, therefore, more sensitive to footprints of very recent selection (Nielsen, 2005). Distinguishing of long-term maintained polymorphism and thus likely functionally important TSP from transient TSP resulting from ILs is a great challenge in current evolutionary genetics (Hedrick, 2013). To attempt to distinguish balanced polymorphism from transient ones originating from ILs we hypothesised that balanced polymorphism should increase intraspecies differentiation and consequenty decrease interspecies differentiation when compared to neutral markers. We tested the level of intraspecies and interspecies differentiation among all species that share the *TLR*s alleles and then separately among American *Poecile* species. However, AMOVA analysis did not show any consistent differences in the variability explained within species and among species in *TLR*s and neutral markers (Table 18 and Table 19) . There was relatively high variability explained by the species entity among neutral genes. Nevertheless, in recently diverged species there is often extensive mixing of alleles of both "neutral" genes and functionally important genes (Klein et al., 1998; Nagl et al., 1998; Samonte et al., 2007). Regarding the relatively recent divergence among our species with detected TSP and high percentage of shared neutral (ancestral) polymorphism (Supplement 2), it is more likely that most of the observed shared polymorphism (TSP) may result rather from extensive incomplete lineage sorting rather than from strong pervasive balancing selection. Especially in closely related American *Poecile* an extensive ancestral polymorphism is known, and along with occasional introgression both phenomenons complicate phylogeny reconstruction. As a result, there is even disconcordance in species tree inferred from autosomal neutral

markers and from mitochondrial genes (Harris et al., 2014). It would be interesting to test the hypothesis of higher population structuring on intraspecies level for MHC genes (where extensive TSP is supposed) by using whole genome sequencing population data of several closely related species. Such data are becoming more and more available and it could be the way how to distinguish balanced polymorphism from transient ones.

Regarding the pleiotropy of immune system, other types of PRRs can fulfil a similar function as TLR4 or TLR5. They might also bind same ligands as the TLRs, e.g. intracellular NLRs bind flagellated bacteria just as TLR5 (Miao et al., 2007). It has been shown recently that *TLR5* pseudogenized independently at least seven times in passerines and other *TLR5* gene loss occurred in different bird orders (Bainova et al., 2014). We might thus hypothesize that individual *TLR*s and *TLR5* in particular can be under different selection pressures, e.g. weaker selection for maintained (balancing selection) and rather diversifying (positive) selection may result in lower persitance of TSP in *TLR*s. There also seems to be a difference between *TLR4* and *TLR5*. *TLR4* binding broader spectrum of ligands (Kumar et al., 2009b) has also more shared alleles (up to four species which share the same trans-specific alelles) and its evolution appears to be more reticulated (compare Figure 13 and Figure 17) , while *TLR5* appeared to be more diversified. It shared alleles only between pairs of closely related species (Table 17) and along with its higher nucleotide diversity it seems to evolve rather in a species-specific manner in Paridae (Figure 17 and Figure 22).

5.4 Gene flow and introgression in Paridae

Adaptive introgression as a source of beneficious variability used to be an overlooked topic and many immunogenetic studies do not take introgression into account (Hedrick, 2013; Nadachowska-Brzyska et al., 2012). To evaluate how common such variability is and how this variability in immune genes contributes to the protection against pathogens and diseases is a great challenge for evolutionary biologists. Our original idea was to distinguish introgression from haplotype networks using neutral markers. However, regarding the high persistence of ancestral polymorphism among species which simultaneously hybridize, this was not possible. Therefore, we applied IMa2 model to evalulate the gene flow between hybridizing species. Although the number of neutral markers included in this study was relatively low and thus the results from the IM model must be interpreted with caution, we found evidence for gene flow between *P. atricapillus* and *P. carolinensis* and *Cyanistes caeruleus* and *C. cyanus*. The estiamated level of the gene flow is similar to the level which was identified in other closely related passerine taxons, e.g. from *Luscinia megarhynchos* to *Luscinia luscinia* and vice versa ($2Nm = 0.118426$; $2Nm = 0.325948$, respcetively) (Storchova et al., 2010), from *Acrocephallus scirpaceus* to *A. palustris*

(2*Nm* = 0.238) (Reifova et al., 2016) or from *Ficedula albicollis* to *Ficedula hypoleuca* and vice versa (2NM=0.538; 2NM=0.123, respectively) (Nater et al., 2015). This is not surprising since *C. caeruleus* and *C. cyanus* as well as *P. atricapillus* and *P. carolinensis* are closely related and probably frequently hybridize in nature. In addition to that, tension zone between *P. atricapillus* and *P. carolinensis* is considered to be a textbook example of hybrid zone (Curry et al., 2007; Curry, 2005). While hybrids between *P. atricapillus* and *P. carolinensis* have been genetically reported to be fertile with lower fitness (Bronson et al., 2003), hybrids between *C. cyanus* a *C. caeruleus* have not been studied so far. However, in view of the fact that birds with intermediating plumage characteristic varry in their appearance from "pure *C. caeruleus*" to " pure *C. cyanus*" (Ławicki, 2012), some of them are probably F2 hybrids, suggesting the hybridization probably leads to fertile hybrids. Our results therefore show that hybridization followed by introgresion is plausible scenario for explaining of shared polymorphism.

We did not include *Baeolophus* genus into IM analysis due to insufficient number of individuals but considering the well known hybrid zones between *B. atricristatus* and *B.bicolor* (C. Curry and Patten, 2014) and their recent split in Pleistocene we may hypothesise that some shared alleles could be also from introgression. Similarly, hybrids between *P. montanus* and *P. cinctus* have been occasionally reported and even though we detected that they share *TLR4* alleles, they were not included in IM model since there are not closely related (Johansson et al., 2013). Despite the known occurence of occasionally genetically reported hybrids between *P. atricapillus* and *P. gambeli* and their casual mixing in mosaic hybrid zone (Grava et al., 2012) we did not detect any gene flow between these chickadees. Similarly no gene flow was identified between other species which also hybridize, e.g. *P. cinctus* and *P. hudsonicus*. However, our results are consistent with Rebecca B. Harris et al. (2014) who have detected the gene flow by using 40 neutral and mitochondrial markers from seven American *Poecile* species only between *P. atricapillus* and *P. carolinensis*.

To conclude, we are not able to distinguish between the origin of the shared polymorphism in species which are closely related and simultaneously hybridize. On the other hand, considering the substantial introgression detected, at least some portion of allele sharing may be of hybrid origin (particularly between *P. atricapillus* and *P. carolinensis*, *C. caeruleus* and *C. cyanus* and presumably also between *B. bicolor* and *B. atricristatus*). Introgression of *TLR* alleles between other species cannot be excluded, but based on the literature search and our results it seems to be unlikely. Moreover, higher percentage adaptive variability is probably inherited from common ancestors (as TSP) rather than introduced by introgression. More neutral markers covering different chromozomes in high density or in better case the whole genome sequencing and sampling in different

distances from hybrid zone would be appropriate to answer this question.

5.5 Evaluating convergence and surface charge analysis

Convergence has not been tested properly since shared variability was restricted only to closely related species (within genus level) whereas documented examples of convergence in immune genes involved more diverged taxa (Chapter 1.3). However, given a few identical amino acid substitutions which probably occurred independently in different lineages (Table 12 and Table 14) and the incongruences between the gene tree of both *TLR5* and *TLR4* with the epograms showing surface charge clustering, convergence might occur in some cases but probably between less closely related species. Moreover, convergence might involve only particular positions e.g., in the form of variation in specific aa features, such as the aa charge. Convergence was most probably not responsible for origin of the shared polymorphism in the whole nucleotide or amino acid sequences. The results of the electrostatic surface charge analysis show reasonable levels of variation in the protein surface charges, suggesting also variation in ligand-binding features of tits.

6 Summary

Balanced TSP is an evolutionary phenomenon explaining the origin of shared polymorphism as a passage of alleles from ancestral species to descendant species and their subsequent long-termed maintenance in related species. Although traditionally well studied in *MHC* genes, little endeavour has been paid to genes outside *MHC*, particularly innate immunity genes. In this thesis I try to expand our knowledge about TSP and look for putative TSP in *TLR4* and *TLR5* genes. I aimed at distinguishing TSP from other evolutionary phenomena which explain shared polymorphism among species (introgression and converegence). Toll-like receptors (TLRs) are molecules of vertebrate innate immunity that recognise danger signals (alarmins) of both exogenous (MAMPs) and endogenous (DAMPs) origin. These two receptors bind mainly bacterial ligands (TLR4 detects lipopolysaccharide and TLR5 detects flagellin), being among the first ones to trigger immune response to bacterial pathogens. The main aims of this thesis were therefore: 1) to describe polymorphism in *TLR4*, *TLR5* and selected neutral markers, 2) to test for signatures of positive selection in *TLR4* and *TLR5*, 3) to identify positively selected residues which may affect binding properties of TLR4 and TLR5, 4) to investigate TSP in *TLR4* and *TLR5* genes and to distinguish it from other mechanisms leading to shared polymorphism and 5) to detect gene flow and introgression.

We sequenced the whole ligand-binding and dimerization regions of *TLR4* and *TLR5* and autosomal neutral markers in 192 individuals representing 20 species of tits, chickadees and titmice (Paridae family). *TLR*s were variable at both intra- and interspecies level. TSP was identified in both *TLR4* and *TLR5* genes. As far as I know, this is the first identification of TSP in *TLR*s and in *PRR*s in general. Sharing of alleles was, however, restricted only to closely related species within a genus (in American *Poecile*, *Cyanistes* and *Baeolophus*). TSP appears to persist in *TLR4* and *TLR5* no logner than 8 MY, more conservatively 4 MY, which is less than in *MHC* genes in which alleles commonly persist above genus level in the order of up to tens of millions of years in different taxa. Due to recent divergence of species with TSP identified in *TLR*s and high ancestral shared polymorphism (transient TSP) in neutral markers we are not able to distinguish whether TSP in *TLR4* and *TLR5* identified here involves balanced polymorphism. Significant gene flow was detected only from *Cyanistes caeruleus* to *Cyanistes cyanus* and between *Poecile atricapillus* and *Poecile carolinensis* in both directions. I was not able to identify the origin of the shared polymorphism in species which are closely related and simultaneously hybridize. However, considering the detected substantial introgression, at least some portion of shared alleles may come from introgression. Convergence has not been properly tested since shared variability documented here was restricted only to genus level. Based on the incongruences between gene tree in both *TLR4* and *TLR5* and the epograms from the surface charge clustering and the

facts that several identical substitutions probably occurred independently in different lineages, convergence might occur but only in few amino acid residues and above genus level. Signature of positive (diversifying) selection was detected in both *TLR4* and *TLR5* on interspecies level. In *TLR4,* from overall 14 detected positively selected residues 4 positions (320, 374, 397,427) were located in close proximity to predicted binding sites. In *TLR5,* from overall 23 positively selected sites 14 positions (33, 35, 56, 80, 106, 156, 181, 183, 209, 213, 237, 261, 294, 379) lied in close proximity to predicted functionally important sites or were directly predicted binding sites. Furthermore, most of these sites were also identified to be the most non-conservative by ConSurf, and where non-conservative substations (e.g. change in size, polarity) occurred in different tit lineages. Therefore, these positions identified here in both *TLR4* and *TLR5* may influence binding properties of TLR4 and TLR5 molecules and thus recognition of pathogens and parasites.

To conclude, our results are well consistent with literature showing that TLRs are variable in both intra- and interspecies level with particular residues being under positive (diversifying selection) in free-living animals. TSP identified here is another evidence that TSP is probably more common evolutionary phenomenon explaining beneficious shared polymorphism. More effort should be paid to distinguish balanced polymorphism and transient polymorphism and the origin of shared variability in species which are closely related and simultaneously hybridize, which would enable us to answer if the hybridization and adaptive introgression are influential sourceprovisioning beneficious polymorphisms in immune genes and therefore also of resistance against pathogens.

7 Acknowledgment

First of all, I wish to acknowledge my supervisor Michal Vinkler for his initial idea to study evolutionary immunology and TSP, valuable advices, suggestions and comments to the thesis. My special thanks belong to my first co-supervisor Hanka Velová who supervised my molecular-genetics analysis and taught me from my first steps in laboratory to sequence analysis. I am also thankful to Hanka for her helpful comments to the manuscript, for her help with sequencing design, PCR optimization, final preparation of sequencing run which Hanka performed in laboratories of European Molecular Bioinformatics Laboratory (EMBL), the first filtering of raw sequence data and providing me PAML selection analysis. I would like to thank a lot to my second co-supervisor Radka Reifová for valuable advices, suggestions and giving me an idea how to proceed neutral marker data as well as helpful comments to the manuscript. I am grateful to Anna Bryjová from External Research Facility of Institute of Vertebrate Biology, AS CR for her restless help and advices which saved me many hours of work in the laboratory. I would like to thank Vladimír Beneš from European Molecular Biology Laboratory (EMBL) in Heidelberg for sequencing design and data sequencing on MiSeq Illumina platform. I am grateful to Pavel Munclinger for teaching me AMOVA analysis. I would like to thank to all members of our Evolutionary and Ecology Immunology Group for helpful advices and friendly atmosphere, especially to Zuzka Bainová for her advices and providing me with data from melting curve analysis. My gratitude belongs to all my family for lifelong support and attentive patience. My very great thanks belong last but not least to Lenka, not only for helpful comments to the manuscript but mainly for her continual support.

This work was supported by Grant Agency of Charles University in Prague (project nr. 540214) and Grant Agency of the Czech Republic (projects nr. P505/10/1871 and P506/15-11782S).

8 List of tables

9 List of figures

10 Abbreviations

aa	amino acid
AMOVA	Analysis of molecular variance
APBS	Adaptive Poisson-Boltzmann Solver
bp	base pairs
CLRs	C-lectin receptors
DAMPs	Damaged associated molecular patterns
CR	Control region
ESS	Effective size values
FUBAR	A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection
GARD	Genetic Algorithm Recombination Detection
HDPs	Host defense peptides
HMGB1	High Mobility Group Box-1
HSP	Heat shock proteins
ILs	Incomplete lineage sorting
IM	Isolation with migration
IMa2	Isolation with migration model for more than two populations
LPS	Lipopolysaccharide
LRRs	Leucine-rich repeats
MAMPs	Microbe associated molecular patterns
MEME	Mixed Effects Model of Evolution
MHC	Major histocompatibility complex
MHC I	Major histocompatibility complex class I
MHC II	Major histocompatibility complex class II
ML	Maximum likelihood method
MRCA	Most recent common ancestor
MSA	Multiple sequence alignement
mtDNA	mitochondrial DNA
MY	millions of years
MYA	millions of years ago
NLRs	Nucleotide oligomerization domain receptors (NOD-like recptors)
NO	Nitric oxide

PAML	Phylogenetic Analysis by Maximum Likelihood
PAMPs	Pathogen associated molecular patterns
PBR	Peptide binding region
PCR	Polymerase chain reaction
PDB	Protein Data Bank
PIPSA	Protein Interaction Property Similarity Analysis
REL	Random Effect Likelihood
RIM	Reproductive isolation mechanisms
ROS	Reactive oxygen species
RLRs	Retinotic acid receptors (RIG-like receptors)
SBP	Rapid Screening for Recombination Using a Single Break Point
TLRs	Toll-like receptors
TSP	Trans-species polymorphism

11 References

Abel, B., Thieblemont, N., Quesniaux, V.J.F., Brown, N., Mpagi, J., Miyake, K., Bihl, F., Ryffel, B., 2002. Toll-like receptor 4 expression is required to control chronic Mycobacterium tuberculosis infection in mice. J. Immunol. 169, 3155–3162. doi:10.4049/jimmunol.169.6.3155

Acevedo-Whitehouse, K., Cunningham, A.A., 2006. Is MHC enough for understanding wildlife immunogenetics? Trends Ecol. Evol. 21, 433–438. doi:10.1016/j.tree.2006.05.010

Aguilar, A., Garza, J.C., 2007. Patterns of historical balancing selection on the salmonid major histocompatibility complex class II beta gene. J. Mol. Evol. 65, 34–43. doi:10.1007/s00239-006-0222-8

Akira, S., Kiyoshi, T., 2004. Toll-like receptor signalling. Nat. Rev. Immunol. 4. doi:10.1038/nri1391

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, S. V., 2011. Molecular evolution of the toll-like receptor multigene family in birds. Mol. Biol. Evol. 28, 1703–1715. doi:10.1093/molbev/msq351

Andersen-Nissen, E., Smith, K.D., Bonneau, R., Strong, R.K., Aderem, A., 2007. A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. J. Exp. Med. 204, 393–403. doi:10.1084/jem.20061400

Anmarkrud, J.A., Johnsen, A., Bachmann, L., Lifjeld, J.T., 2010. Ancestral polymorphism in exon 2 of bluethroat (Luscinia svecica) MHC class II B genes. J. Evol. Biol. 23, 1206–1217. doi:10.1111/j.1420-9101.2010.01999.x

Arkush, K.D., Giese, A.R., Mendonca, H.L., McBride, A.M., Marty, G.D., Hedrick, P.W., 2002. Resistance to three pathogens in the endangered winter-run chinook salmon (Oncorhynchus tshawytscha): effects of inbreeding and major histocompatibility complex genotypes. Can. J. Fish. Aquat. Sci. 59, 966–975. doi:10.1139/f02-066

Ashkenazy, H., Erez, E., Martz, E., Pupko, T., Ben-Tal, N., 2010. ConSurf 2010: Calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. 38, 529–533. doi:10.1093/nar/gkq399

Backström, N., Fagerberg, S., Ellegren, H., 2008. Genomics of natural bird populations: A gene-based set of reference markers evenly spread across the avian genome. Mol. Ecol. 17, 964–980. doi:10.1111/j.1365-294X.2007.03551.x

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. doi:10.1016/j.dci.2014.02.010

Bandelt, H.J., Forster, P., Röhl, A., 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16, 37–48. doi:10.1093/oxfordjournals.molbev.a026036

Bernatchez, L., Landry, C., 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J. Evol. Biol. 16, 363–377. doi:10.1046/j.1420-9101.2003.00531.x

Bianchi, M.E., 2007. DAMPs, PAMPs and alarmins: all we need to know about danger. J. Leukoc. Biol. 81, 1–5. doi:10.1189/jlb.0306164

Bollmer, J.L., Vargas, F.H., Parker, P.G., 2007. Low MHC variation in the endangered

Galapagos penguin (Spheniscus mendiculus). Immunogenetics 59, 593–602. doi:10.1007/s00251-007-0221-y

Bos, D.H., Waldman, B., 2006. Evolution by recombination and transspecies polymorphism in the MHC class I gene of Xenopus laevis. Mol. Biol. Evol. 23, 137–143. doi:10.1093/molbev/msj016

Botos, I., Segal, D.M., Davies, D.R., 2011. The structural biology of Toll-like receptors. Structure 19, 447–459. doi:10.1016/j.str.2011.02.004

Bronson, C.L., Grubb, T.C., Sattler, G.D., Braun, M.J., 2003. Mate preference: a possible causal mechanism for a moving hybrid zone. Anim. Behav. 65, 489–500. doi:10.1006/anbe.2003.2103

Bryja, J., Galan, M., Charbonnel, N., Cosson, J.F., 2006. Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58, 191–202. doi:10.1007/s00251-006-0085-6

Burg, T.M., Gaston, A.J., Winker, K., Friesen, V.L., 2006. Effects of Pleistocene glaciations on population structure of North American chestnut-backed chickadees. Mol. Ecol. 15, 2409–2419. doi:10.1111/j.1365-294X.2006.02957.x

Cagliani, R., Fumagalli, M., Biasin, M., Piacentini, L., Riva, S., Pozzoli, U., Bonaglia, M.C., Bresolin, N., Clerici, M., Sironi, M., 2010. Long-term balancing selection maintains trans-specific polymorphisms in the human TRIM5 gene. Hum. Genet. 128, 577–588. doi:10.1007/s00439-010-0884-6

Curry, C.M., Patten, M.A., 2014. Current and Historical Extent of Phenotypic Variation in the Tufted and Black-crested Titmouse (Paridae) Hybrid Zone in the Southern Great Plains. Am. Midl. Nat. 171, 271–300.

Curry, R., 2005. Hybridization in Chickadees: Much To Learn From Familiar Birds. Auk 122, 747. doi:10.1642/0004-8038(2005)122[0747:HICMTL]2.0.CO;2

Curry, R.L., Rossano, L.M., Reudink, M.W., 2007. Behavioral aspects of chickadee hybridization. Ecol. Behav. Chickadees Titmice An Integr. Approach 95–110. doi:10.1093/acprof:oso/9780198569992.003.0008

Dhondt, A.A., 2014. What drives diffences between North American and Eurasian tit studies?, In: Ecology and behavior of chickadees and titmice: an integrated approach. doi:10.1093/acprof

Dowling, T., Secor, C., 1997. The Role of Hybridization and Introgression in the Diversification of Animals Author (s): Thomas E . Dowling and Carol L . Secor Source : Annual Review of Ecology and Systematics , Vol . 28 (1997), pp . 593-619 Published by : Annual Reviews Stable URL. Annu. Rev. Ecol. Syst. 28, 593–619.

Dwyer, K.G., Balent, M.A., Nasrallah, J.B., Nasrallah, M.E., 1991. DNA-SEQUENCES OF SELF-INCOMPATIBILITY GENES FROM BRASSICA-CAMPESTRIS AND BRASSICA-OLERACEA - POLYMORPHISM PREDATING SPECIATION. Plant Mol. Biol. 16, 481–486. doi:10.1007/bf00024000

Edwards, S. V, Hedrick, P.W., 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol. Evol. 13, 305–311. doi:10.1016/s0169-5347(98)01416-5

Eriksson, E., Sampaio, G., Schofield, L., 2014. Toll-Like Receptors and Malaria – Sensing and Susceptibility. J. Trop. Dis. 02, 1–7. doi:10.4172/2329-891X.1000126

Escalera-Zamudio, M., Zepeda-Mendoza, M.L., Loza-Rubio, E., Rojas-Anaya, E., Méndez-Ojeda, M.L., Arias, C.F., Greenwood, A.D., 2015. The evolution of bat nucleic acid-sensing Toll-like receptors. Mol. Ecol. 24, 5899–5909. doi:10.1111/mec.13431

Ferguson, W., Dvora, S., Gallo, J., Orth, A., Boissinot, S., 2008. Long-term balancing selection at the West Nile virus resistance gene, Oas1b, maintains transspecific polymorphisms in the house mouse. Mol. Biol. Evol. 25, 1609–1618. doi:10.1093/molbev/msn106

Fitzgerald, K.A., Rowe, D.C., Golenbock, D.T., 2004. Endotoxin recognition and signal transduction by the TLR4/MD-2 complex. Microbes Infect. 6, 1361–1367. doi:10.1016/j.micinf.2004.08.015

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, 17. doi:10.1186/1471-2148-13-194

Fu, Y.X., Li, W.H., 1993. Statistical tests of neutrality of mutations. Genetics 133, 693–709. doi:evolution

Gabdoulline, R.R., Stein, M., Wade, R.C., 2007. qPIPSA: relating enzymatic kinetic parameters and interaction fields. BMC Bioinformatics 8, 373. doi:10.1186/1471-2105-8-373

Ganz, T., 2003. Defensins: Antimicrobial peptides of innate immunity. Nat. Rev. Immunol. 3, 710–720. doi:10.1038/nri1180

Gill, F.B., Slikas, B., Agro, D., 1999. Speciation in North American chickadees: II. Geography of mtDNA haplotypes in Poecile carolinensis. Auk 116, 274–277.

Gill, F.B., Slikas, B., Sheldon, F.H., 2005. Phylogeny of titmice (Paridae): II. Species relationships based on sequences of the mitochondrial cytochrome-b gene. Auk 122, 121–143. doi:10.1642/0004-8038(2005)122[0121:potpis]2.0.co;2

Gillingham, M.A.F., Courtiol, A., Teixeira, M., Galan, M., Bechet, A., Cezilly, F., 2016. Evidence of gene orthology and trans-species polymorphism, but not of parallel evolution, despite high levels of concerted evolution in the major histocompatibility complex of flamingo species. J. Evol. Biol. 29, 438–454. doi:10.1111/jeb.12798

Glaberman, S., Caccone, A., 2008. Species-specific evolution of class I MHC genes in iguanas (Order : Squamata; subfamily : Iguaninae). Immunogenetics 60, 371–382. doi:10.1007/s00251-008-0298-y

Go, Y., Rakotoarisoa, G., Kawamoto, Y., Shima, T., Koyama, N., Randrianjafy, A., Mora, R., Hirai, H., 2005. Characterization and evolution of major histocompatibility complex class II genes in the aye-aye, Daubentonia madagascariensis. Primates 46, 135–139. doi:10.1007/s10329-004-0101-0

Gohli, J., Leder, E.H., Garcia-del-Rey, E., Johannessen, L.E., Johnsen, A., Laskemoen, T., Popp, M., Lifjeld, J.T., 2015. The evolutionary history of Afrocanarian blue tits inferred from genomewide SNPs. Mol. Ecol. 24, 180–191. doi:10.1111/mec.13008

Gonzalez-Quevedo, C., Phillips, K.P., Spurgin, L.G., Richardson, D.S., 2014. 454 screening of individual MHC variation in an endemic island passerine. Immunogenetics 67, 149–162. doi:10.1007/s00251-014-0822-1

Gonzalez-Quevedo, C., Spurgin, L.G., Illera, J.C., Richardson, D.S., 2015. Drift, not selection,

shapes toll-like receptor variation among oceanic island populations. Mol. Ecol. 24, 5852–5863. doi:10.1111/mec.13437

Grant, P.R., Grant, B.R., 1992. HYBRIDIZATION OF BIRD SPECIES. Science (80-.). 256, 193–197. doi:10.1126/science.256.5054.193

Grava, A., Grava, T., Didier, R., Lait, L.A., Dosso, J., Koran, E., Burg, T.M., Otter, K.A., 2012. Interspecific dominance relationships and hybridization between black-capped and mountain chickadees. Behav. Ecol. 23, 566–572. doi:10.1093/beheco/arr229

Graves, G.R., 2008. Handbook of Avian Hybrids of the World, The Wilson Journal of Ornithology. doi:10.1676/0043-5643(2008)120[233:HOAHOT]2.0.CO;2

Grossen, C., Keller, L., Biebach, I., Croll, D., Int Goat Genome, C., 2014. Introgression from Domestic Goat Generated Variation at the Major Histocompatibility Complex of Alpine Ibex. Plos Genet. 10, 16. doi:10.1371/journal.pgen.1004438

Grueber, C.E., Knafler, G.J., King, T.M., Senior, A.M., Grosser, S., Robertson, B., Weston, K.A., Brekke, P., Harris, C.L.W., Jamieson, I.G., 2015. Toll-like receptor diversity in 10 threatened bird species: relationship with microsatellite heterozygosity. Conserv. Genet. 16, 595–611. doi:10.1007/s10592-014-0685-x

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, 9. doi:10.1371/journal.pone.0089632

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp. Ser. doi:citeulike-article-id:691774

Halldórsdóttir, K., Árnason, E., 2015. Trans-species polymorphism at antimicrobial innate immunity cathelicidin genes of Atlantic cod and related species. PeerJ 3, e976. doi:10.7717/peerj.976

Hansson, G.K., Edfeldt, K., 2005. Toll to be paid at the gateway to the vessel wall. Arterioscler. Thromb. Vasc. Biol. 25, 1085–1087. doi:10.1161/01.ATV.0000168894.43759.47

Harris, R.B., Carling, M.D., Lovette, I.J., 2014. The influence of sampling design on species tree inference: A new relationship for the new world chickadees (Aves: Poecile). Evolution (N. Y). 68, 501–513. doi:10.1111/evo.12280

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

Hedrick, P.W., 2013. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol. Ecol. 22, 4606–4618. doi:10.1111/mec.12415

Hedrick, P.W., 2012. What is the evidence for heterozygote advantage selection? Trends Ecol. Evol. 27, 698–704. doi:10.1016/j.tree.2012.08.012

Hedrick, P.W., 2002. Pathogen resistance and genetic variation at MHC loci. Evolution (N. Y). 56, 1902–1908.

Heimpel, G.E., de Boer, J.G., 2008. Sex determination in the Hymenoptera, in: Annual Review of Entomology, Annual Review of Entomology. Annual Reviews, [Heimpel, George E.] Univ Minnesota, Dept Entomol, St Paul, MN 55108 USA. [de Boer, Jetske G.] Univ

Groningen, Ctr Biol, Ctr Ecol & Evolut Studies, NL-9751 NN Haren, Netherlands. Heimpel, GE (reprint author), Univ Minnesota, Dept Entomol, St Paul, MN 5510, pp. 209–230. doi:10.1146/annurev.ento.53.103106.093441

Hellgren, O., Sheldon, B.C., 2011. Locus-specific protocol for nine different innate immune genes (antimicrobial peptides: beta-defensins) across passerine bird species reveals within-species coding variation and a case of trans-species polymorphisms. Mol. Ecol. Resour. 11, 686–692. doi:10.1111/j.1755-0998.2011.02995.x

Hey, J., 2010. Isolation with migration models for more than two populations. Mol. Biol. Evol. 27, 905–920. doi:10.1093/molbev/msp296

Hey, J., Nielsen, R., 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D. persimilis. Genetics 167, 747–760. doi:10.1534/genetics.103.024182

Hovanessian, A.G., Justesen, J., 2007. The human 2 ’-5 ' oligoadenylate synthetase family: Unique interferon-inducible enzymes catalyzing 2 '-5 ' instead of 3 '-5 ' phosphodiester bond formation. Biochimie 89, 779–788. doi:10.1016/j.biochi.2007.02.003

Hudson, R.R., Kaplan, N.L., 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. Genetics 111, 147–164.

Hughes, A.L., Yeager, M., 1998. Natural selection at major histocompatibility complex loci of vertebrates. Annu. Rev. Genet. 32, 415–435. doi:10.1146/annurev.genet.32.1.415

Huson, D.H., Bryant, D., 2006. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. doi:10.1093/molbev/msj030

Chu, H., Mazmanian, S.K., 2013. Innate immune recognition of the microbiota promotes host-microbial symbiosis. Nat. Immunol. 14, 668–675. doi:10.1038/ni.2635

Illera, J.C., Koivula, K., Broggi, J., Packert, M., Martens, J., Kvist, L., 2011. A multi-gene approach reveals a complex evolutionary history in the Cyanistes species group. Mol. Ecol. 20, 4123–4139. doi:10.1111/j.1365-294X.2011.05259.x

Ioerger, T.R., Clark, A.G., Kao, T.H., 1990. POLYMORPHISM AT THE SELF-INCOMPATIBILITY LOCUS IN SOLANACEAE PREDATES SPECIATION. Proc. Natl. Acad. Sci. U. S. A. 87, 9732–9735. doi:10.1073/pnas.87.24.9732

Iwasaki, A., Medzhitov, R., 2004. Toll-like receptor control of the adaptive immune responses. Nat. Immunol. 5, 987–95. doi:10.1038/ni1112

James, H.F., Ericson, P.G.P., Slikas, B., Lei, F.M., Gill, F.B., Olson, S.L., 2003. Pseudopodoces humilis, a misclassified terrestrial tit (Paridae) of the Tibetan Plateau: evolutionary consequences of shifting adaptive zones. Ibis (Lond. 1859). 145, 185–202. doi:10.1046/j.1474-919X.2003.00170.x

Jana, N., Vidhu, A., Raini, D., Zhang, L., Saluja, A., Meng, J., Lisa, K., Santanu, B., Sabita, R., 2016. Differential effects of gram-positive and gram-negative bacterial products on morphine induced inhibition of phagocytosis. Sci. Rep. 6, 21094. doi:10.1038/srep21094

Janova, E., Matiasovic, J., Vahala, J., Vodicka, R., Van Dyk, E., Horin, P., 2009. Polymorphism and selection in the major histocompatibility complex DRA and DQA genes in the family Equidae. Immunogenetics 61, 513–527. doi:10.1007/s00251-009-0380-0

Jaratlerdsiri, W., Isberg, S.R., Higgins, D.P., Miles, L.G., Gongora, J., 2014. Selection and Trans-Species Polymorphism of Major Histocompatibility Complex Class II Genes in the Order

Crocodylia. PLoS One 9, 13. doi:10.1371/journal.pone.0087534

Jarvinen, A., 1987. Asuccessful mixed breeding between Parus cinctus and Parus montanus in Finnish Lapland. Tiedonantoja.

Jeffery, K.J.M., Bangham, C.R.M., 2000. Do infectious diseases drive MHC diversity? Microbes Infect. 2, 1335–1341. doi:10.1016/s1286-4579(00)01287-9

Johansson, U.S., Ekman, J., Bowie, R.C.K., Halvarsson, P., Ohlson, J.I., Price, T.D., Ericson, P.G.P., 2013. A complete multilocus species phylogeny of the tits and chickadees (Aves: Paridae). Mol. Phylogenet. Evol. 69, 852–860. doi:10.1016/j.ympev.2013.06.019

Johnson, N.K., Cicero, C., 2004. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. Evolution 58, 1122–1130. doi:10.1111/j.0014-3820.2004.tb00445.x

Johnson, W.E., Sawyer, S.L., 2009. Molecular evolution of the antiretroviral TRIM5 gene. Immunogenetics 61, 163–176. doi:10.1007/s00251-009-0358-y

Kamath, P.L., Getz, W.M., 2011. Adaptive molecular evolution of the Major Histocompatibility Complex genes, DRA and DQA, in the genus Equus. Bmc Evol. Biol. 11, 16. doi:10.1186/1471-2148-11-128

Kawai, T., Akira, S., 2011. Toll-like Receptors and Their Crosstalk with Other Innate Receptors in Infection and Immunity. Immunity 34, 637–650. doi:10.1016/j.immuni.2011.05.006

Kawai, T., Akira, S., 2010a. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat. Immunol. 11, 373–384. doi:10.1038/ni.1863

Kermarrec, N., Roubinet, F., Apoil, P.A., Blancher, A., 1999. Comparison of allele O sequences of the human and non-human primate ABO system. Immunogenetics 49, 517–526. doi:10.1007/s002510050529

Kikkawa, E.F., Tsuda, T.T., Sumiyama, D., Naruse, T.K., Fukuda, M., Kurita, M., Wilson, R.P., LeMaho, Y., Miller, G.D., Tsuda, M., Murata, K., Kulski, J.K., Inoko, H., 2009. Trans-species polymorphism of the Mhc class II DRB-like gene in banded penguins (genus Spheniscus). Immunogenetics 61, 341–352. doi:10.1007/s00251-009-0363-1

Kim, H.M., Park, B.S., Kim, J.I., Kim, S.E., Lee, J., Oh, S.C., Enkhbayar, P., Matsushima, N., Lee, H., Yoo, O.J., Lee, J.O., 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

Kimura, M., 1969. The number of heterozygous nucleotide sites maintained in a finite population due to steady flux of mutations. Genetics 61, 893–903.

Kiryu, I., Dijkstra, J.M., Sarder, R.I., Fujiwara, A., Yoshiura, Y., Ototake, M., 2005. New MHC class Ia domain lineages in rainbow trout (Oncorhynchus mykiss) which are shared with other fish species. Fish Shellfish Immunol. 18, 243–254. doi:10.1016/j.fsi.2004.07.007

Klein, J., Sato, A., Nagl, S., O'HUigin, C., 1998. Molecular trans-species polymorphism. Annu. Rev. Ecol. Syst. 29, 1–+. doi:10.1146/annurev.ecolsys.29.1.1

Klein, J., Sato, A., Nikolaidis, N., 2007. MHC, TSP, and the origin of species: From immunogenetics to evolutionary genetics, in: Annual Review of Genetics, Annual Review of Genetics. Annual Reviews, [Klein, Jan Nikolaidis, Nikolas] Penn State Univ,

Dept Biol, University Pk, PA 16801 USA. [Sato, Akie] Tsurumi Univ Sch Dent Med, Yokohama, Kanagawa 2308501, Japan. Klein, J (reprint author), Penn State Univ, Dept Biol, University Pk, PA 16801 USA. sato-a, pp. 281–304. doi:10.1146/annurev.genet.41.110306.130137

Kogut, M.H., He, H., Kaiser, P., 2005. Lipopolysaccharide binding protein/CD14/ TLR4-dependent recognition of salmonella LPS induces the functional activation of chicken heterophils and up-regulation of pro-inflammatory cytokine and chemokine gene expression in these cells. Anim. Biotechnol. 16, 165–181. doi:10.1080/10495390500264896

Kosakovsky Pond, S.L., Frost, S.D.W., 2005. Datamonkey: Rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21, 2531–2533. doi:10.1093/bioinformatics/bti320

Kosakovsky, P., Sergei, L., Frost, S.D.W.S.D.W., 2005. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. Mol. Biol. Evol. 22, 1208–1222. doi:10.1093/molbev/msi105

Kriener, K., O'HUigin, C., Klein, J., 2001. Independent origin of functional MHC class II genes in humans and New World monkeys. Hum. Immunol. 62, 1–14. doi:10.1016/s0198-8859(00)00233-0

Kriener, K., O'HUigin, C., Tichy, H., Klein, J., 2000. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. Immunogenetics 51, 169–178. doi:10.1007/s002510050028

Kumar, H., Kawai, T., Akira, S., 2009a. Toll-like receptors and innate immunity. Biochem. Biophys. Res. Commun. 388, 621–625. doi:10.1016/j.bbrc.2009.08.062

Kundu, S., Faulkes, C.G., 2007. A tangled history: patterns of major histocompatibility complex evolution in the African mole-rats (Family : Bathyergidae). Biol. J. Linn. Soc. 91, 493–503. doi:10.1111/j.1095-8312.2007.00814.x

Kvist, L., Martens, J., Ahola, A., Orell, M., 2001. Phylogeography of a Palaearctic sedentary passerine, the willow tit (Parus montanus). J. Evol. Biol. 14, 930–941. doi:10.1046/j.1420-9101.2001.00354.x

Kvist, L., Martens, J., Higuchi, H., Nazarenko, A.A., Valchuk, O.P., Orell, M., 2003. Evolution and genetic structure of the great tit (Parus major) complex. Proc. R. Soc. B-Biological Sci. 270, 1447–1454. doi:10.1098/rspb.2002.2321

Lait, L.A., Lauff, R.F., Burg, T.M., 2012. Genetic evidence supports Boreal Chickadee (Poecile hudsonicus) × Black-capped Chickadee (Poecile atricapillus) hybridization in Atlantic Canada. Can. Field-Naturalist 126, 143–147.

Ławicki, L., 2012. Azure Tits and hybrids Azure x European Blue Tit in Europe Azure Tits and hybrids Azure x European Blue Tit in Europe. Dutch Bird. 34, 219–231.

LeBouder, E., Rey-Nores, J.E., Rushmere, N.K., Grigorov, M., Lawn, S.D., Affolter, M., Griffin, G.E., Ferrara, P., Schiffrin, E.J., Morgan, B.P., Labeta, M.O., 2003. Soluble forms of toll-like receptor (TLR)2 capable of modulating TLR2 signaling are present in human plasma and breast milk. J. Immunol. 171, 6680–6689.

Lee, M.S., Min, Y.J., 2007. Signaling pathways downstream of pattern-recognition receptors and their cross talk, in: Annual Review of Biochemistry, Annual Review of Biochemistry. Annual Reviews, Yonsei Univ, Dept Biochem, Seoul 120749, South Korea. Lee, MS (reprint author), Yonsei Univ, Dept Biochem, Seoul 120749, South Korea. mycongsup.lee@gmail.com yjkim@yonsei.ac.lcr, pp. 447–480.

doi:10.1146/annurev.biochem.76.060605.122847

Lechner, S., Ferretti, L., Schoning, C., Kinuthia, W., Willemsen, D., Hasselmann, M., 2014. Nucleotide Variability at Its Limit? Insights into the Number and Evolutionary Dynamics of the Sex-Determining Specificities of the Honey Bee Apis mellifera. Mol. Biol. Evol. 31, 272–287. doi:10.1093/molbev/mst207

Leveque, G., Forgetta, V., Morroll, S., Adrian, L., Bumstead, N., Barrow, P., Morgan, K., Malo, D., Smith, A.L., 2003. Allelic Variation in TLR4 Is Linked to Susceptibility to Salmonella enterica Serovar Typhimurium Infection in Chickens 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

Li, L., Zhou, X.P., Chen, X.L., 2011. Characterization and Evolution of Li, L., Zhou, X.P., Chen, X.L., 2011. Characterization and Evolution of MHC Class II B Genes in Ardeid Birds. J. Mol. Evol. 72, 474–483. doi:10.1007/s00239-011-9446-3MHC Class II B Genes in Ardeid Birds. J. Mol. Evol. 72, 474–483. doi:10.1007/s00239-011-9446-3

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

Lovette, I.J., 2005. Glacial cycles and the tempo of avian speciation. Trends Ecol. Evol. 20, 57–59. doi:10.1016/j.tree.2004.11.011

Lu, Y.C., Yeh, W.C., Ohashi, P.S., 2008. LPS/TLR4 signal transduction pathway. Cytokine 42, 145–151. doi:10.1016/j.cyto.2008.01.006

Lukens, L., Yicun, H., May, G., 1996. Correlation of genetic and physical maps at the A mating-type locus of Coprinus cinereus. Genetics 144, 1471–1477.

Mallet, J., 2005. Hybridization as an invasion of the genome. Trends Ecol. Evol. 20, 229–237. doi:10.1016/j.tree.2005.02.010

Martinko, J.M., Vincek, V., Klein, D., Klein, J., 1993. PRIMATE ABO GLYCOSYLTRANSFERASES - EVIDENCE FOR TRANSSPECIES EVOLUTION. Immunogenetics 37, 274–278.

McCarthy, E.M., 2006. Handbook of avian hybrids of the world. Oxford University Press.

Medzhitov, R., PrestonHurlburt, P., Janeway, C.A., 1997. A human homologue of the Drosophila Toll protein signals activation of adaptive immunity. Nature 388, 394–397.

Meyer, D., Thomson, G., 2001. How selection shapes variation of the human major histocompatibility complex: a review. Ann. Hum. Genet. 65, 1–26. doi:10.1046/j.1469-1809.2001.6510001.x

Miao, E.A., Andersen-Nissen, E., Warren, S.E., Aderem, A., 2007. TLR5 and Ipaf: Dual sensors of bacterial flagellin in the innate immune system. Semin. Immunopathol. 29, 275–288. doi:10.1007/s00281-007-0078-z

Milinski, M., 2006. The major histocompatibility complex, sexual selection, and mate choice, in: Annual Review of Ecology Evolution and Systematics, Annual Review of Ecology Evolution and Systematics. Annual Reviews, Max Planck Inst Limnol, Dept Evolutionary Ecol, D-24306 Plon, Germany. Milinski, M (reprint author), Max Planck Inst Limnol, Dept Evolutionary Ecol, D-24306 Plon, Germany. milinski@mpil-plocn.mpg.de, pp. 159–186. doi:10.1146/annurev.ecolsys.37.091305.110242

Miller, S.I., Ernst, R.K., Bader, M.W., 2005. LPS, TLR4 and infectious disease diversity. Nat. Rev. Microbiol. 3, 36–46. doi:10.1038/nrmicro1068

Mogensen, H, 2009. Pathogen Recognition and Inflammatory Signaling in Innate Immune Defenses. Clin. Microbiol. Rev. 22, 240–+. doi:10.1128/cmr.00046-08

Mucha, R., Bhide, M.R., Chakurkar, E.B., Novak, M., Mikula, I., 2009. Toll-like receptors TLR1, TLR2 and TLR4 gene mutations and natural resistance to Mycobacterium avium subsp. paratuberculosis infection in cattle. Vet. Immunol. Immunopathol. 128, 381–388. doi:10.1016/j.vetimm.2008.12.007

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. Mol. Biol. Evol. 30, 1196–1205. doi: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 Genet. 8. doi:10.1371/journal.pgen.1002764

Musolf, K., Meyer-Lucht, Y., Sommer, S., 2004. Evolution of MHC-DRB class II polymorphism in the genus Apodemus and a comparison of DRB sequences within the family Muridae (Mammalia : Rodentia). Immunogenetics 56, 420–426. doi:10.1007/s00251-004-0715-9

Nadachowska-Brzyska, K., Zielinski, P., Radwan, J., Babik, W., 2012. Interspecific hybridization increases MHC class II diversity in two sister species of newts. Mol. Ecol. 21, 887–906. doi:10.1111/j.1365-294X.2011.05347.x

Nagl, S., Tichy, H., Mayer, W.E., Takahata, N., Klein, J., 1998. Persistence of neutral polymorphisms in Lake Victoria cichlid fish. Proc. Natl. Acad. Sci. U. S. A. 95, 14238–14243. doi:10.1073/pnas.95.24.14238

Nater, A., Burri, R., Kawakami, T., Smeds, L., Ellegren, H., 2015. Resolving evolutionary relationships in closely related species with whole-genome sequencing data. Syst. Biol. 64, 1000–1017. doi:10.1093/sysbio/syv045

Neefjes, J., Jongsma, M.L.M., Paul, P., Bakke, O., 2011. Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat. Rev. Immunol. 11, 823–836. doi:10.1038/nri3084

Netea, M.G., Wijmenga, C., O'Neill, L.A.J., 2012. Genetic variation in Toll-like receptors and disease susceptibility. Nat Immunol 13, 535–542.

Newman, R.M., Hall, L., Connole, M., Chen, G.L., Sato, S., Yuste, E., Diehl, W., Hunter, E., Kaur, A., Miller, G.M., Johnson, W.E., 2006. Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5 alpha. Proc. Natl. Acad. Sci. U. S. A. 103, 19134–19139. doi:10.1073/pnas.0605838103

Nielsen Henrik, Jacob Engelbrecht, Soren Brunak, G. von H., 1997. Identi cation of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng. 10, 1–6.

Nielsen, R., 2005. Molecular signatures of natural selection. Annu. Rev. Genet. 39, 197–218. doi:doi: 10.1146/annurev.genet.39.073003.112420

Nurnberger, T., Brunner, F., Kemmerling, B., Piater, L., 2004. Innate immunity in plants and animals: striking similarities and obvious differences. Immunol. Rev. 198, 249–266. doi:10.1111/j.0105-2896.2004.0119.x

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, 7421–6. doi:10.1073/pnas.1201193109

Ottova, E., Simkova, A., Martin, J.F., de Bellocq, J.G., Gelnar, M., Allienne, J.F., Morand, S., 2005. Evolution and trans-species polymorphism of MHC class II beta genes in cyprinid fish. Fish Shellfish Immunol. 18, 199–222. doi:10.1016/j.fsi.2004.07.004

Owczarzy, R., Tataurov, A. V., Wu, Y., Manthey, J.A., McQuisten, K.A., Almabrazi, H.G., Pedersen, K.F., Lin, Y., Garretson, J., McEntaggart, N.O., Sailor, C.A., Dawson, R.B., Peek, A.S., 2008. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res. 36, 163–169. doi:10.1093/nar/gkn198

Päckert, M., Martens, J., 2008. Taxonomic pitfalls in tits – comments on the Paridae chapter of the Handbook G R EAT TITS AN D T URKE S TAN TITS – SPECI E S S TAT U S COAL T I T S AND S P OT-WI NGE D T I T S – S P E CI E S S TAT US 829–831.

Päckert, M., Martens, J., Eck, S., Nazarenko, A.A., Valchuk, O.P., Petri, B., Veith, M., 2005. The great tit (Parus major) – a misclassified ring species major sector minor sector cinereus sector contact zones (I – V) 153–174.

Päckert, M., Martens, J., Thomas, D., Dietzen, C., Wink, M., Kvist, L., 2007. Calibration of a molecular clock in tits (Paridae)— Do nucleotide substitution rates of mitochondrial genes deviate from the 2 % rule ? 44, 1–14. doi:10.1016/j.ympev.2007.03.006

Packert, M., Martens, J., Tietze, D.T., Dietzen, C., Wink, M., Kvist, L., 2007. Calibration of a molecular clock in tits (Paridae) - Do nucleotide substitution rates of mitochondrial genes deviate from the 2% rule? Mol. Phylogenet. Evol. 44, 1–14. doi:10.1016/j.ympev.2007.03.006

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, 1191–5. doi:10.1038/nature07830

Peakall, R., Smouse, P.E., 2012. GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. Bioinformatics 28, 2537–2539. doi:10.1093/bioinformatics/bts460

Penn, D.J., 2002. The scent of genetic compatibility: Sexual selection and the major histocompatibility complex. Ethology 108, 1–21. doi:10.1046/j.1439-0310.2002.00768.x

Penn, D.J., Damjanovich, K., Potts, W.K., 2002. MHC heterozygosity confers a selective advantage against multiple-strain infections. Proc. Natl. Acad. Sci. U. S. A. 99, 11260–11264. doi:10.1073/pnas.162006499

Pentzold, S., Tritsch, C., Martens, J., Tietze, D.T., Giacalone, G., Lo Valvo, M., Nazarenko, A.A., Kvist, L., Packert, M., 2013. Where is the line? Phylogeography and secondary contact of western Palearctic coal tits (Periparus ater: Aves, Passeriformes, Paridae). Zool. Anz. 252, 367–382. doi:10.1016/j.jcz.2012.10.003

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

Pond, S.L.K., Posada, D., Gravenor, M.B., Woelk, C.H., Frost, S.D.W., 2006. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23, 1891–1901. doi:10.1093/molbev/msl051

Puthothu, B., Forster, J., Heinzmann, A., Krueger, M., 2006. TLR-4 and CD14 polymorphisms in respiratory syncytial virus associated disease. Dis Markers 22, 303–308.

Qu, Y., Zhao, H., Han, N., Zhou, G., Song, G., Gao, B., Tian, S., Zhang, J., Zhang, R., Meng, X., Zhang, Y., Zhang, Y., Zhu, X., Wang, W., Lambert, D., Ericson, P.G.P., Subramanian, S.,

Yeung, C., Zhu, H., Jiang, Z., Li, R., Lei, F., 2013. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. Nat. Commun. 4. doi:10.1038/ncomms3071

Randler, C., 2004. Frequency of bird hybrids: does detectability make all the difference? J. Ornithol. 145, 123–128. doi:10.1007/s10336-004-0022-0

Randler, C., 2002. Avian hybridization, mixed pairing and female choice. Anim. Behav. 63, 103–119. doi:10.1006/anbe.2001.1884

Reifova, R., Majerova, V., Reif, J., Ahola, M., Lindholm, A., Procházka, P., 2016. Patterns of gene flow and selection across multiple species of Acrocephalus warblers : footprints of parallel selection on the Z chromosome Patterns of gene flow and selection across multiple species of Acrocephalus warblers : footprints of parallel selec. Bmc Evol. Biol. doi:10.1186/s12862-016-0692-2

Reudink, M.W., Mech, S.G., Mullen, S.P., Curry, R.L., Curr, R.L., 2007. BLACK-CAPPED CHICKADEE (POECILE ATRICAPILLUS) AND CAROLINA CHICKADEE (P . CAROLINENSIS) IN SOUTHEASTERN PENNSYLVANIA.

Richardson, D.S., Westerdahl, H., 2003. MHC diversity in two Acrocephalus species: the outbred Great reed warbler and the inbred Seychelles warbler. Mol. Ecol. 12, 3523–3529. doi:10.1046/j.1365-294X.2003.02005.x

Richman, A.D., Kao, T.H., Schaeffer, S.W., Uyenoyama, M.K., 1995. S-ALLELE SEQUENCE DIVERSITY IN NATURAL-POPULATIONS OF SOLANUM CAROLINENSE (HORSENETTLE). Heredity (Edinb). 75, 405–415. doi:10.1038/hdy.1995.153

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 Res. 36, 276–280. doi:10.1093/nar/gkn181

Roach, J.C., Glusman, G., Rowen, L., Kaur, A., Purcell, M.K., Smith, K.D., Hood, L.E., Aderem, A., 2005. The evolution of vertebrate Toll-like receptors. Proc. Natl. Acad. Sci. U. S. A. 102, 9577–9582. doi:10.1073/pnas.0502272102

Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738. doi:10.1038/nprot.2010.5

Rozas, J., 2009. DNA Sequence Polymorphism Analysis Using DnaSP, in: Posada, D. (Ed.), Bioinformatics for DNA Sequence Analysis. Humana Press, Totowa, NJ, pp. 337–350. doi:10.1007/978-1-59745-251-9_17

Salzburger, W., Martens, J., Sturmbauer, C., 2002. Paraphyly of the Blue Tit (Parus caeruleus) suggested from cytochrome b sequences 24, 19–25.

Samonte, I.E., Satta, Y., Sato, A., Tichy, H., Takahata, N., Klein, J., 2007. Gene flow between species of Lake Victoria haplochromine fishes. Mol. Biol. Evol. 24, 2069–2080. doi:10.1093/molbev/msm138

Samplonius, J.M., Both, C., 2014. A Case of a Three Species Mixed Brood after Two Interspecific Nest Takeovers A case of a three species mixed brood after two interspecific nest takeovers. Ardea 102, 105–107. doi:10.5253/078.102.0113

Sato, A., Tichy, H., Grant, P.R., Grant, B.R., Sato, T., O'HUigin, C., 2011. Spectrum of MHC Class II Variability in Darwin's Finches and Their Close Relatives. Mol. Biol. Evol. 28, 1943–1956. doi:10.1093/molbev/msr015

Ségurel, L., Thompson, E.E., Flutre, T., Lovstad, J., Venkat, A., Susan, W., Moyse, J., Ross, S., Gamble, K., Sella, G., 2013. Correction for Segurel et al., The ABO blood group is a trans-species polymorphism in primates. Proc. Natl. Acad. Sci. 110, 6607–6607. doi:10.1073/pnas.1304029110

Seifertová, M., Šimková, A., 2011. Structure, diversity and evolutionary patterns of expressed MHC class IIB genes in chub (Squalius cephalus), a cyprinid fish species from Europe. Immunogenetics 63, 167–181. doi:10.1007/s00251-010-0495-3

Shu, Y.L., Hong, P., Yang, Y.W., Wu, H.L., 2013. An Endemic Frog Harbors Multiple Expression Loci With Different Patterns of Variation in the MHC Class II B Gene. J. Exp. Zool. Part B-Molecular Dev. Evol. 320, 501–510. doi:10.1002/jez.b.22525

Schrödinger, LLC, 2015. The {PyMOL} Molecular Graphics System, Version~1.8.

Sin, Y.W., Dugdale, H.L., Newman, C., Macdonald, D.W., Burke, T., 2012. Evolution of MHC class I genes in the European badger (Meles meles). Ecol. Evol. 2, 1644–1662. doi:10.1002/ece3.285

Slikas, B., Sheldon, F.H., Gill, F.B., 1996. Phylogeny of titmice (Paridae) .1. Estimate of relationships among subgenera based on DNA-DNA hybridization. J. Avian Biol. 27, 70–82. doi:10.2307/3676963

Source, C.C., By, P., Society, T., Doi, E., 2004. BARRIERS TO SYMPATRY BETWEEN AVIAN SIBLING SPECIES (PARIDAE : BAEOLOPHUS) IN LOCAL SECONDARY CONTACT Author (s): Carla Cicero Published By : The Society for the Study of Evolution (PARIDAE : BAEOLOPHUS) IN LOCAL SECONDARY CONTACT 1 58, 1573–1587.

Spurgin, L.G., Richardson, D.S., 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. R. Soc. B-Biological Sci. 277, 979–988. doi:10.1098/rspb.2009.2084

Steiner, T.S., 2007. How flagellin and toll-like receptor 5 contribute to enteric infection. Infect. Immun. 75, 545–552. doi:10.1128/IAI.01506-06

Stephens, M., Donnelly, P., 2003. Report A Comparison of Bayesian Methods for Haplotype Reconstruction from Population Genotype Data. Am. J. Hum. Genet 73, 1162–1169. doi:10.1086/379378

Stephens, M., Smith, N.J., Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68, 978–989. doi:10.1086/319501

Stiebens, V.A., Merino, S.E., Chain, F.J.J., Eizaguirre, C., 2013. Evolution of MHC class I genes in the endangered loggerhead sea turtle (Caretta caretta) revealed by 454 amplicon sequencing. Bmc Evol. Biol. 13, 11. doi:10.1186/1471-2148-13-95

Storchova, Z., Landova, E., Frynta, D., 2010. Why some tits store food and others do not: evaluation of ecological factors. J. Ethol. 28, 207–219. doi:10.1007/s10164-009-0200-x

Swiderek, W., Bhide, M., Gruszczynska, J., Soltis, K., Witkowska, D., Mikula, I., 2006. Toll-like receptor gene polymorphism and its relationship with somatic cell concentration and natural bacterial infections of the mammary gland in sheep. Folia Microbiol. (Praha). 51, 647–652. doi:10.1007/BF02931633

Tajima, F., 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123, 585–595. doi:PMC1203831

Takahata, N., 1993. ALLELIC GENEALOGY AND HUMAN-EVOLUTION. Mol. Biol. Evol. 10, 2–22.

Takeuchi, O., Akira, S., 2010. Pattern Recognition Receptors and Inflammation. Cell 140, 805–820. doi:10.1016/j.cell.2010.01.022

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. doi:10.1093/molbev/mst197

Těšický, M., Vinkler, M., 2015. Trans-Species Polymorphism in Immune Genes: General Pattern or MHC-Restricted Phenomenon? J. Immunol. Res. 2015, 838035. doi:10.1155/2015/838035

Tietze, D.T., Borthakur, U., 2012. Historical biogeography of tits (Aves: Paridae, Remizidae). Org. Divers. Evol. 12, 433–444. doi:10.1007/s13127-012-0101-7

Trowsdale, J., 2011. The MHC, disease and selection. Immunol. Lett. 137, 1–8. doi:10.1016/j.imlet.2011.01.002

Tschirren, B., Andersson, M., Scherman, K., Westerdahl, H., Mittl, P.R., Raberg, L., 2013. Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. Proc Biol Sci 280, 20130364. doi:10.1098/rspb.2013.0364

Uematsu, S., Akira, S., 2008. Toll-Like Receptors (TLRs) and Their Ligands, in: Bauer, S., Hartmann, G. (Eds.), Toll-Like Receptors (TLRs) and Innate Immunity. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–20. doi:10.1007/978-3-540-72167-3_1

Unckless, R.L., Lazzaro, B.P., 2016. The potential for adaptive maintenance of diversity in insect antimicrobial peptides. Philos. Trans. R. Soc. b in review. doi:10.1098/not

van Valen, L., 1973. A new evolutionary law. Evol. Theory. doi:10.1038/344864a0

Vazquez-Torres, A., Vallance, B.A., Bergman, M.A., Finlay, B.B., Cookson, B.T., Jones-Carson, J., Fang, F.C., 2004. Toll-like receptor 4 dependence of innate and adaptive immunity to Salmonella: importance of the Kupffer cell network. J. Immunol. (Baltimore, Md 1950) 172, 6202–6208. doi:10.4049/jimmunol.172.10.6202

Villesen, P., 2007. FaBox: An online toolbox for FASTA sequences. Mol. Ecol. Notes 7, 965–968. doi:10.1111/j.1471-8286.2007.01821.x

Vincek, V., Ohuigin, C., Satta, Y., Takahata, N., Boag, P.T., Grant, P.R., Grant, B.R., Klein, J., 1997. How large was the founding population of Darwin's finches? Proc. R. Soc. B-Biological Sci. 264, 111–118.

Vinkler, M., Albrecht, T., 2011. Phylogeny, longevity and evolution of adaptive immunity. Folia Zool. 60, 277–282.

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., Bainova, H., Bryja, J., 2014. Protein evolution of Toll-like receptors 4, 5 and 7 within Galloanserae birds. Genet. Sel. Evol. 46, 12. doi:10.1186/s12711-014-0072-6

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. doi:10.1111/j.1399-0039.2009.01273.x

Voelker, G., Rohwer, S., Outlaw, D.C., Bowie, R.C.K., 2009. Repeated trans-Atlantic dispersal catalysed a global songbird radiation. Glob. Ecol. Biogeogr. 18, 41–49. doi:10.1111/j.1466-8238.2008.00423.x

Walsh, C., Gangloff, M., Monie, T., Smyth, T., Wei, B., McKinley, T.J., 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. doi:181/2/1245 [pii]

Walsh, H.E., Friesen, V.L., 2003. A comparison of intraspecific patterns of DNA sequence variation in mitochondrial DNA, alpha-enolase, and MHC class II B loci in auklets (Charadriiformes : Alcidae). J. Mol. Evol. 57, 681–693. doi:10.1007/s00239-003-2518-2

Wang, D.Q., Zhong, L., Wei, Q.W., Gan, X.N., He, S.P., 2010. Evolution of MHC class I genes in two ancient fish, paddlefish (Polyodon spathula) and Chinese sturgeon (Acipenser sinensis). Febs Lett. 584, 3331–3339. doi:10.1016/j.febslet.2010.05.065

Wegner, K.M., Eizaguirre, C., 2012. New(t)s and views from hybridizing MHC genes: introgression rather than trans-species polymorphism may shape allelic repertoires. Mol. Ecol. 21, 779–781. doi:10.1111/j.1365-294X.2011.05401.x

Wegner, K.M., Kalbe, M., Schaschl, H., Reusch, T.B.H., 2004. Parasites and individual major histocompatibility complex diversity - an optimal choice? Microbes Infect. 6, 1110–1116. doi:10.1016/j.micinf.2004.05.025

Wlasiuk, G., Nachman, M.W., 2010. Adaptation and constraint at toll-like receptors in primates. Mol. Biol. Evol. 27, 2172–2186. doi:10.1093/molbev/msq104

Woerner, A.E., Cox, M.P., Hammer, M.F., 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics 23, 1851–1853. doi:10.1093/bioinformatics/btm253

Xu, S.X., Chen, B.Y., Zhou, K.Y., Yang, G., 2008. High similarity at three MHC loci between the baiji and finless porpoise: Trans-species or convergent evolution? Mol. Phylogenet. Evol. 47, 36–44. doi:10.1016/j.ympev.2007.05.026

Xu, S.X., Ren, W.H., Li, S.Z., Wei, F.W., Zhou, K.Y., Yang, G., 2009. Sequence polymorphism and evolution of three cetacean MHC genes. J. Mol. Evol. 69, 260–275. doi:10.1007/s00239-009-9272-z

Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat Meth 12, 7–8. doi:10.1038/nmeth.3213\rhttp://www.nature.com/nmeth/journal/v12/n1/abs/nmeth.3213.html#supplementary-information

Yang, Z., 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24, 1586–1591. doi:10.1093/molbev/msm088

Yeager, M., Hughes, A.L., 1999. Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunol. Rev. 167, 45–58. doi:10.1111/j.1600-065X.1999.tb01381.x

Yoon, S., Kurnasov, O., Natarajan, V., Hong, M., Gudkov, A., Osterman, A., Wilson, I., 2013. Structural basis of TLR5-flagellin recognition and signaling. Science (80-.). 335, 859–864. doi:10.1126/science.1215584.Structural

Zamyatnin, a. a., 1984. Amino acid, peptide, and protein volume in solution. Annu. Rev. Biophys. Bioeng. 13, 145–165. doi:10.1146/annurev.bb.13.060184.001045

Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinformatics 9, 40. doi:10.1186/1471-2105-9-40

Zhao, M.A., Wang, Y.Z., Shen, H., Li, C.L., Chen, C., Luo, Z.H., Wu, H., 2013. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. Bmc Evol. Biol. 13, 13. doi:10.1186/1471-2148-13-113

Zhou, H., Hickford, J.G.H., Fang, Q., 2005. Polymorphism of the DQA2 gene in goats. J. Anim. Sci. 83, 963–968.

Zipfel, C., Felix, G., 2005. Plants and animals: a different taste for microbes? Curr. Opin. Plant Biol. 8, 353–360. doi:10.1016/j.pbi.2005.05.004

Books and Theses

Bainová, H., 2011. The Influence of Toll-like Receptor 4 Polymorphism on Condition and Ornamentation in Great Tit. Master's Thesis. Charles University in Prague.

Cramp, S., Perrins, C.M., Brooks, D.J. (Eds.), 1993. The Birds of the Western Palearctic, Volume 7: Old World Flycatchers to Shrikes. Oxford University Press, Oxford.

Gosler A.G. & Clement, P. 2007. Family Paridae (Tits and Chick-adees). In del Hoyo, J., Elliott, A. & Christie, D.A. (eds) Hand-book of the Birds of the World: Vol. 12: 662–750. Barcelona:Lynx Edicions

King, R.C., Mulligan, P., Stansfield, W., 2013. A dictionary of genetics. Oxford University Press.

McCarthy, E.M., 2006. Handbook of avian hybrids of the world. Oxford University Press.

Nei, M. 1987. Molecular evolutionary genetics. Columbia University Press, New York.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Schmid-Hempel, P., 2011. Evolutionary parasitology : the integrated study of infections, immunology, ecology, and genetics. Oxford University Press.

Vinklerová, J., 2013. Impact of Toll-like Receptor 4 Polymorphism on Pro-inflammatory Responsiveness in Great Tit. Master's Thesis. Charles University in Prague.

12 Supplement

Supplement 1: Basic population genetics characteristics, Tajima's D and Fu and Li's D and recombination estimates for neutral markers (Table A-F)

The length of the sequences is after excluding INDELs mutations within species. Number of haploid sequences (N), number of unique nucleotide haplotypes (N_2), number of segregating sites (S), number of mutations (n), nucleotide diversity per site (π), proportion of polymorphic sites per site (θ), estimate of recombination parameter (R), minimal number of recombination events (R_m), divergence to outgroup (zebra finch) - average number of nucleotide substitutions (K), divergence to outgroup – average number of nucleotide substitution per base (D_{xy}). Tajima's D, Fu and Li's D statistic, R and R_m are not defined if there is no polymorphism within species. Significant Tajima's D and Fu and Li' s D values (p<0,05) are labelled by three asterisks ***, marginally significant values (p>0,05 and p<0.1) are labelled by one asterisk *. The legend shown here is identical for all tables.

A)

DDB1	Species	Length	*N*	N_2	*S*	*n*	π	θ	Tajima's D	Fu and Li's D	R_m	*R*	*K*	D_{xy}
	Baeolophus atricristatus	485	6	1	0	0	0	0	/	/	/	/	34.000	0.07039
	Baeolophus bicolor	485	18	4	5	5	0.00195	0.00300	-1.10169	0.42002	0	0	34.056	0.07051
	Baeolophus ridgwayi	485	12	2	2	2	0.00169	0.00137	0.68788	0.97295	0	0.0012	35.500	0.07265
	Baeolophus wollweberi	486	12	3	3	3	0.00178	0.00204	-0.42854	-0.93419	0	0.0093	33.583	0.06939
	Cyanistes caeruleus	/	/	/	/	/	/	/	/	/	/	/	/	/
	Cyanistes cyanus	/	/	/	/	/	/	/	/	/	/	/	/	/
	Lophophanes cristatus	503	22	2	1	1	0.00018	0.00055	-1.16240	-1.57469	0	/	22.045	0.04701
	Melaniparus afer	503	4	1	0	0	0	0	/	/	/	/	28.000	0.05882
	Melaniparus niger	484	10	3	4	4	0.00427	0.00292	1.77236*	1.23914	0	0.0151	31.500	0.06535
	Parus major	480	50	12	11	11	0.00515	0.00512	0.002096	0.29066	2	0.0798	34.920	0.07305
	Periparus ater	486	22	13	15	16	0.00657	0.00903	-0.98843	-0.46949	1	0.0748	36.318	0.07519
	Poecile atricapillus	485	18	5	4	4	0.00092	0.00240	-1.85306***	-2.52547***	0	/	32.222	0.06671
	Poecile carolinensis	485	18	5	6	6	0.00492	0.00360	1.21025	1.25898	1	0.0140	31.611	0.06545
	Poecile cinctus	485	20	3	3	3	0.00194	0.00174	0.30478	1.00649	0	0.0145	31.150	0.06449
	Poecile gambeli	485	18	10	12	12	0.00631	0.00719	-0.45514	-0.91136	1	0.0376	33.333	0.06901
	Poecile hudsonicus	485	20	3	3	3	0.00062	0.00174	-1.72331*	-2.38573*	0	0	31.000	0.06439
	Poecile montanus	485	28	7	7	7	0.00205	0.00371	-1.34753	-1.46193	0	0.0153	32.429	0.06714
	Poecile palustris	484	12	6	14	14	0.00729	0.00958	-1.01752	-0.27901	1	0.0101	31.833	0.06604
	Poecile rufescens	485	20	3	2	2	0.00078	0.00116	-0.76857	0.86615	0	/	31.000	0.06418
	Poecile sclateri	485	12	8	12	12	0.00794	0.00819	-0.13225	-0.22543	2	0.0971	32.583	0.06746

B)

DLD	Species	Length	*N*	N_2	*S*	*n*	π	θ	Tajima's D	Fu and Li's D	R_m	*R*	*K*	D_{xy}
	Baeolophus atricristatus	495	6	1	0	0	0	0	/	/	/	/	38.000	0.07755
	Baeolophus bicolor	495	12	2	1	1	0.00098	0.00067	1.06589	0.75202	0	/	38.333	0.07823
	Baeolophus ridgwayi	495	2	1	0	0	0	0	/	/	/	/	40.000	0.08163
	Baeolophus wollweberi	495	6	1	0	0	0	0	/	/	/	/	38.889	0.08239
	Cyanistes caeruleus	482	18	5	6	6	0.00301	0.00362	-0.55384	0.57735	0	0.0094	43.050	0.08786
	Cyanistes cyanus	495	20	2	1	1	0.00020	0.00057	-1.16439	-1.53959	0	/	43.050	0.08786
	Lophophanes cristatus	491	16	3	2	2	0.00107	0.00123	-0.33010	-0.50381	0	/	42.750	0.08796
	Melaniparus afer	489	4	2	1	1	0.00136	0.00112	1.63299	1.632299	0	/	41.500	0.08557
	Melaniparus niger	492	8	4	3	3	0.00247	0.00235	0.20364	0.30073	0	/	42.000	0.08650
	Parus major	495	50	6	6	6	0.00256	0.00271	-0.13950	0.31528	0	0.2874	39.740	0.08110
	Periparus ater	495	12	4	5	5	0.00367	0.00334	0.36176	0.56268	1	0.0101	41.083	0.08384
	Poecile atricapillus	496	14	0	0	0	0	0	/	/	/	/	43.000	0.08758
	Poecile carolinensis	496	6	4	6	6	0.00605	0.00530	0.81086	1.05892	0	0.0915	44.167	0.08995
	Poecile cinctus	496	18	1	0	0	0	0	/	/	/	/	43.000	0.08758
	Poecile gambeli	496	12	3	2	2	0.00180	0.00134	1.02214	0.97295	0	/	43.583	0.08876
	Poecile hudsonicus	496	12	1	0	0	0	0	/	/	/	/	42.000	0.08554
	Poecile montanus	496	22	5	4	4	0.00180	0.00221	-0.52596	0.14251	0	0.0592	42.545	0.08665
	Poecile palustris	496	18	2	1	1	0.00022	0.00059	-1.16467	-1.49949	0	/	39.944	0.08135
	Poecile rufescens	496	20	1	0	0	0	0	/	/	/	/	42.000	0.08554
	Poecile sclateri	494	8	3	3	3	0.00318	0.00234	1.47376	1.23376	0	0.0352	42.750	0.08582

C)

CHMP5	Species	Length	*N*	N_2	*S*	*n*	π	θ	Tajima's D	Fu and Li's D	R_m	*R*	*K*	D_{xy}
	Baeolophus atricristatus	472	6	2	1	1	0.00071	0.00093	-0.93302	-0.95015	0	/	46.833	0.10101
	Baeolophus bicolor	471	18	4	6	6	0.00334	0.00370	-0.31945	-0.78589	0	0.0242	44.167	0.09581
	Baeolophus ridgwayi	471	12	5	6	6	0.00399	0.00422	-0.20740	-0.50357	0	0.2530	46.167	0.10014
	Baeolophus wollweberi	472	12	5	3	3	0.00173	0.00210	-0.57864	-0.93419	0	/	46.333	0.10029
	Cyanistes caeruleus	462	24	5	5	5	0.00251	0.00290	-0.38855	0.33154	0	0.1017	39.667	0.08776
	Cyanistes cyanus	462	18	4	5	5	0.00300	0.00315	-0.14819	1.19899	0	0.0026	39.778	0.08800
	Lophophanes cristatus	472	22	3	3	3	0.00075	0.00174	-1.47087	-1.30921	0	0	42.091	0.09111
	Melaniparus afer	471	4	2	1	1	0.00106	0.00116	-0.61237	-0.61237	0	0	40.750	0.08785
	Melaniparus niger	472	10	1	0	0	0	0	/	/	/	/	39.000	0.08442
	Parus major	472	50	15	16	16	0.00416	0.00757	-1.39443	-1.03251	1	0.0066	41.000	0.08874
	Periparus ater	460	22	15	15	16	0.012111	0.00954	0.97846	0.20032	1	0.0347	44.273	0.09838
	Poecile atricapillus	472	18	8	6	6	0.00407	0.00370	0.33437	-0.10427	0	0	41.556	0.08995
	Poecile carolinensis	470	18	8	9	9	0.00560	0.00557	0.02352	-0.61358	2	0.0221	41.500	0.09022
	Poecile cinctus	472	20	2	1	1	0.00107	0.00060	1.43024	0.64952	0	/	41.000	0.08874
	Poecile gambeli	465	18	5	4	4	0.00150	0.00250	-1.19565	-1.61330	0	/	39.389	0.08657
	Poecile hudsonicus	472	20	4	3	3	0.00101	0.00179	-1.15810	-0.12425	0	/	40.250	0.08712
	Poecile montanus	472	28	7	6	6	0.00305	0.00327	-0.18839	-0.30596	1	2.6624	42.393	0.09176
	Poecile palustris	473	24	2	1	1	0.00018	0.00057	-1.15933	-1.60583	0	/	41.958	0.09082
	Poecile rufescens	472	20	2	1	1	0.00021	0.00060	-1.16439	-1.53959	0	/	40.050	0.08669
	Poecile sclateri	472	10	9	9	9	0.00452	0.00674	-1.44250	-1.81276	0	0.0121	41.800	0.09048

D)

MMAA	**Species**	**Length**	***N***	***N_2***	***S***	***n***	***π***	***θ***	**Tajima D**	**Fu and Li´s D**	***Rm***	***R_m***	***K***	***D_{xy}***
	Baeolophus atricristatus	452	6	2	1	1	0.00074	0.00097	-0.93302	-0.95015	0	/	43.833	0.09719
	Baeolophus bicolor	452	8	9	8	8	0.00454	0.00515	-0.40883	-0.29418	2	0.1388	44.278	0.09818
	Baeolophus ridgwayi	452	14	4	6	6	0.00357	0.00417	-0.51624	0.02019	0	0	43.786	0.09529
	Baeolophus wollweberi	452	8	5	8	8	0.00811	0.00586	1.53387	1.38342***	0	0.0073	44.333	0.0983
	Cyanistes caeruleus	451	24	11	13	13	0.0084	0.00772	0.30393	0.25917	3	0.1113	46.333	0.10296
	Cyanistes cyanus	451	20	3	2	2	0.00134	0.00125	0.1727	-0.59347	0	0	45	0.1
	Lophophanes cristatus	452	22	4	4	4	0.00133	0.00243	-1.26827	-0.81047	0	0	42.227	0.09363
	Melaniparus afer	452	4	2	1	1	0.00111	0.00121	-0.61237	-0.61237	0	/	50.25	0.11142
	Melaniparus niger	452	10	6	8	8	0.00629	0.00626	0.02526	0.06382	0	0.3415	47.4	0.1051
	Parus major	439	50	11	12	12	0.00417	0.0061	-0.93532	-0.1439	0	0.0186	43.612	0.09957
	Periparus ater	452	22	14	21	23	0.00935	0.01396	-1.24316	-0.58666	2	0.1055	42.864	0.09504
	Poecile atricapillus	449	18	10	13	13	0.00853	0.00842	0.04984	0.37045	2	0.144	44.944	0.10032
	Poecile carolinensis	447	18	14	15	15	0.0086	0.00976	-0.45031	-0.12532	2	1.536	43	0.09641
	Poecile cinctus	452	20	3	2	2	0.00044	0.00125	-1.51284	-2.05308*	0	0	45.1	0.1
	Poecile gambeli	448	18	6	5	5	0.00187	0.00324	-1.34363	-1.13794	0	/	44	0.0975
	Poecile hudsonicus	452	20	5	4	4	0.00257	0.00249	0.09161	0.17445	0	0.2202	45.55	0.10145
	Poecile montanus	452	28	11	11	11	0.00827	0.00625	1.05357	1.43895***	2	0.1233	43.071	0.0955
	Poecile palustris	450	24	1	0	0	0	0	/	/	0	/	45	0.10022
	Poecile rufescens	452	20	5	4	4	0.00126	0.00249	-1.43544	-1.69308	0	/	45	0.10022
	Poecile sclateri	450	12	5	6	6	0.00465	0.00442	0.19977	0.70614	1	0.05121	44.833	0.09752

E)

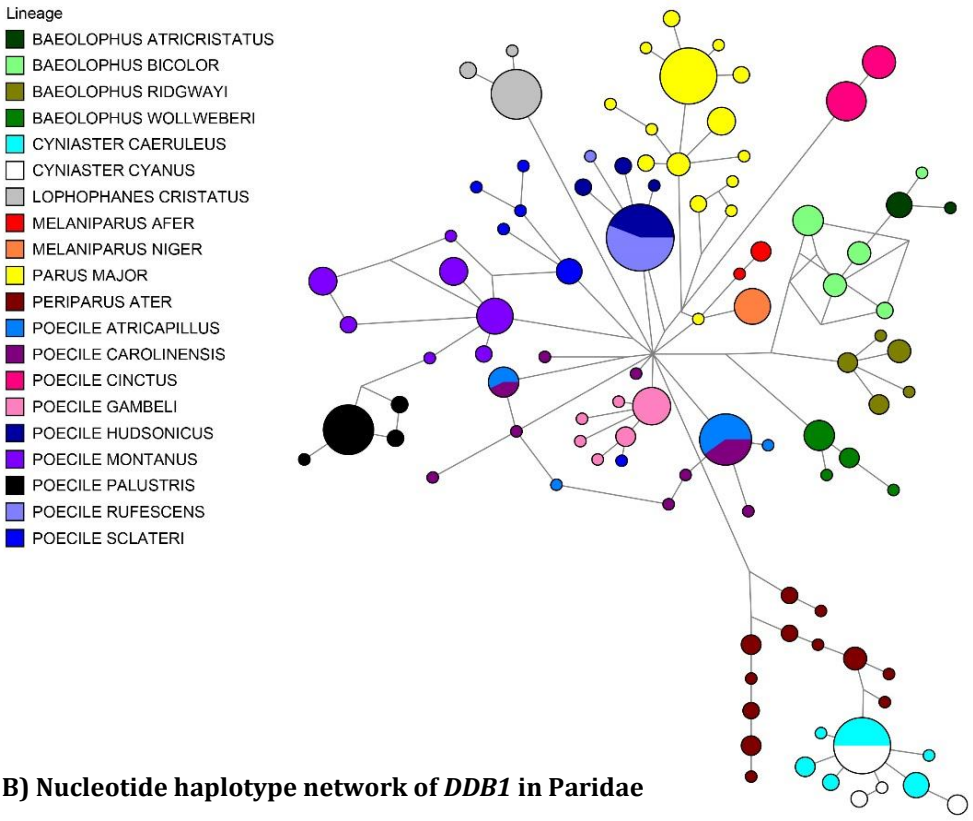
TIAL	**Species**	**Length**	***N***	***N_2***	***S***	***n***	***π***	***θ***	**Tajima's D**	**Fu and Li's D**	***R_m***	***R***	***K***	***D_{xy}***
	Baeolophus atricristatus	457	6	5	5	5	0.00438	0.00525	-1.12397	-1.12397	0	0	36.600	0.08133
	Baeolophus bicolor	460	20	10	14	14	0.00986	0.00858	0.54793	0.41700	4	0.0346	37.500	0.08179
	Baeolophus ridgwayi	461	12	3	2	2	0.00072	0.00144	-1.45138	-1.72038	0	/	35.167	0.07746
	Baeolophus wollweberi	461	12	3	3	3	0.00108	0.00215	-1.62929*	-1.95374*	0	0	38.083	0.83880
	Cyanistes caeruleus	461	24	5	4	4	0.00293	0.00232	0.71565	0.11422	0	0.0204	32.875	0.07241
	Cyanistes cyanus	461	20	2	1	1	0.00022	0.00061	-1.16439	-1.53959	0	/	33.950	0.07478
	Lophophanes cristatus	460	22	2	1	1	0.00038	0.00060	-0.64112	0.63504	0	/	34.909	0.07706
	Melaniparus afer	461	4	3	2	2	0.00253	0.00237	0.59158	0.59158	0	/	33.750	0.07434
	Melaniparus niger	461	10	6	5	5	0.00366	0.00383	-0.17819	-0.02396	0	/	33.200	0.07313
	Parus major	445	50	5	5	5	0.00694	0.00251	-1.27145	0.13389	0	0	32.960	0.07525
	Periparus ater	461	22	8	7	7	0.00399	0.00417	-0.13403	-0.63526	2	0.8435	36.955	0.08140
	Poecile atricapillus	457	18	11	12	13	0.00506	0.00827	-1.44662	-1.47833	2	0.1432	34.333	0.07630
	Poecile carolinensis	461	18	8	10	10	0.00413	0.00631	-1.24530	-0.88148	1	0.0759	34.389	0.07575
	Poecile cinctus	461	20	0	0	0	0	0	/	/	/	/	35.000	0.07700
	Poecile gambeli	461	18	8	7	7	0.00230	0.00441	-1.62793*	-1.72671	0	/	34.389	0.07575
	Poecile hudsonicus	461	20	6	5	5	0.00229	0.00306	-076304	-0.41302	0	/	35.850	0.07896
	Poecile montanus	461	28	6	6	6	0.00138	0.00329	-1.65814*	-1.88589	0	0.0283	35.223	0.07761
	Poecile palustris	461	24	0	0	0	0	0	/	/	/	/	35.000	0.07709
	Poecile rufescens	459	20	6	5	5	0.00266	0.00307	-0.40881	-1.21271	0	0.0361	35.800	0.07785
	Poecile sclateri	461	12	5	4	4	0.00345	0.00287	0.70723	0.36794	0	0.1226	35.083	0.07728

F)

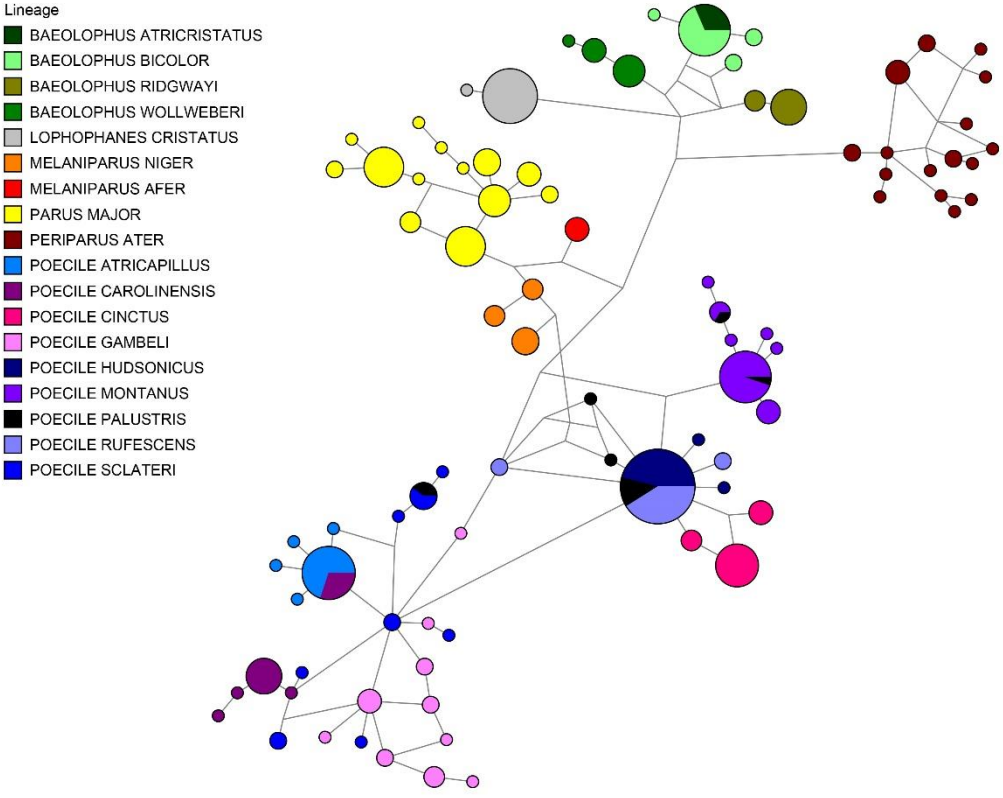
UCHLP3	Species	Length	*N*	N_2	*S*	*n*	π	θ	Tajima's D	Fu and Li's D	R_m	*R*	*K*	D_{xy}
	Baeolophus atricristatus	487	6	3	3	3	0.00205	0.00270	-1.23311	-1.26013	0	0.0374	29.500	0.06427
	Baeolophus bicolor	487	14	6	5	5	0.00325	0.00323	0.02237	-0.02235	0	0.0327	30.071	0.06395
	Baeolophus ridgwayi	488	12	4	5	5	0.00292	0.00339	-0.51530	-0.013525	0	0.0250	34.333	0.07309
	Baeolophus wollweberi	487	10	4	3	3	0.00187	0.00218	-0.50669	0.174464	0	/	29.500	0.06427
	Cyanistes caeruleus	488	22	6	5	5	0.00260	0.00257	0.35747	0.21365	0	3.2094	30.227	0.06571
	Cyanistes cyanus	488	20	1	0	0	0	0	/	/	/	/	30.000	0.06522
	Lophophanes cristatus	474	22	3	1	2	0.00127	0.00116	0.21923	0.85062	0	/	25.500	0.05717
	Melaniparus afer	488	4	0	0	0	0	0	/	/	/	/	30.000	0.06522
	Parus major	488	50	8	7	7	0.00245	0.00323	-0.60863	0.46305	0	/	28.560	0.06209
	Periparus ater	488	22	10	15	15	0.00468	0.00843	-1.60274*	-1.65727	0	0.0228	30.445	0.06622
	Melaniparus niger	488	10	4	3	3	0.00278	0.00217	1.00120	1.15417	0	0.1694	28.500	0.06196
	Poecile atricapillus	480	18	7	10	10	0.00456	0.00606	-0.88909	-0.88148	2	0.0188	31.667	0.07006
	Poecile carolinensis	479	16	6	7	7	0.00506	0.00440	0.52688	1.31791	2	0.0121	30.563	0.06770
	Poecile cinctus	488	20	4	3	3	0.00217	0.00173	0.67051	1.00649	1	0.0027	29.150	0.06337
	Poecile gambeli	488	18	4	4	4	0.00111	0.00238	-1.60021*	-1.61330	0	0.0138	32.056	0.06969
	Poecile hudsonicus	488	18	5	5	5	0.00171	0.00298	-1.34363	-1.13794	0	0.0101	29.333	0.00637
	Poecile montanus	481	28	10	10	10	0.00476	0.00545	-0.34889	0.35088	0	/	26.214	0.05787
	Poecile palustris	476	24	8	7	7	0.00407	0.00394	0.10652	-0.022894	1	0.0925	29.708	0.06472
	Poecile rufescens	488	20	4	3	3	0.00156	0.00173	-0.26042	-0.12425	0	0.0051	29.150	0.06337
	Poecile sclateri	488	12	2	2	2	0.00124	0.00136	-0.24805	0.97295	0	0	30.333	0.65940

Supplement 2: Nucleotide haplotype networks for neutral markers (Figure A-E)

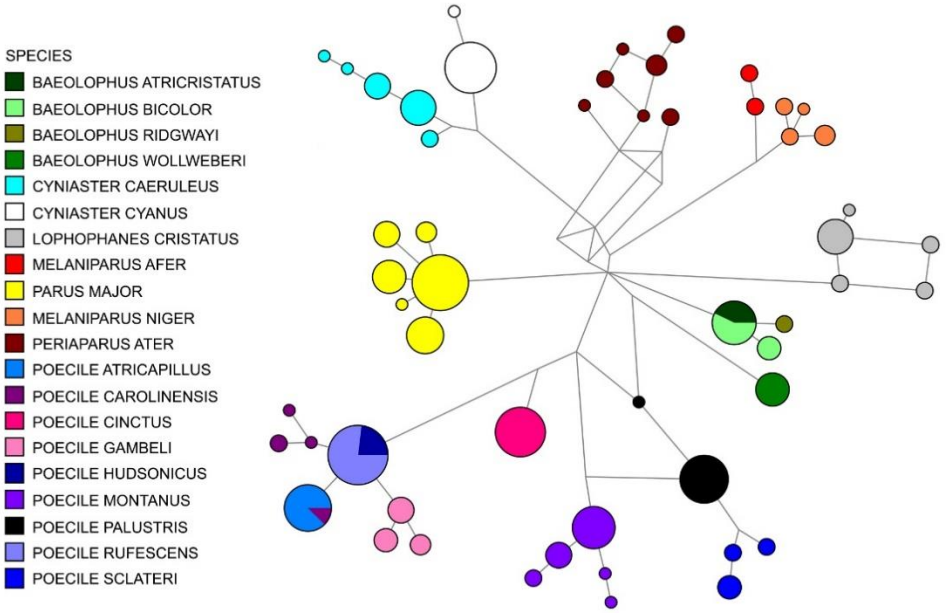
A) Nucleotide haplotype network of *CHMP5* in Paridae



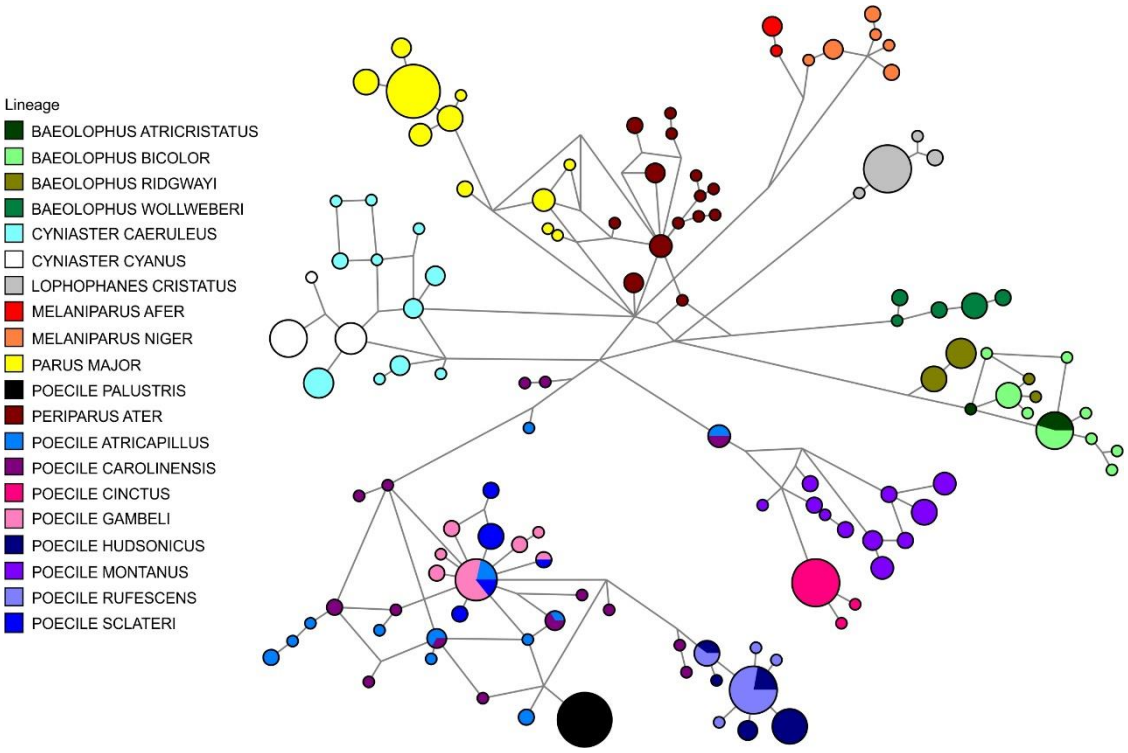
B) Nucleotide haplotype network of *DDB1* in Paridae



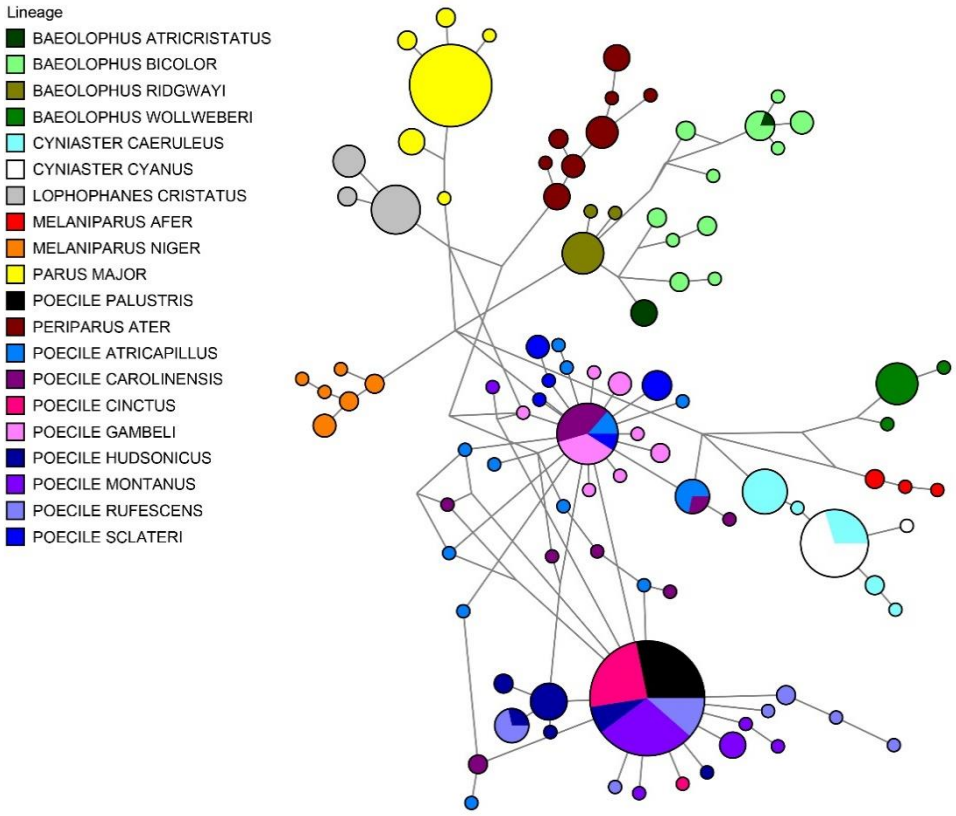
C) Nucleotide haplotype network of *DLD* in Paridae



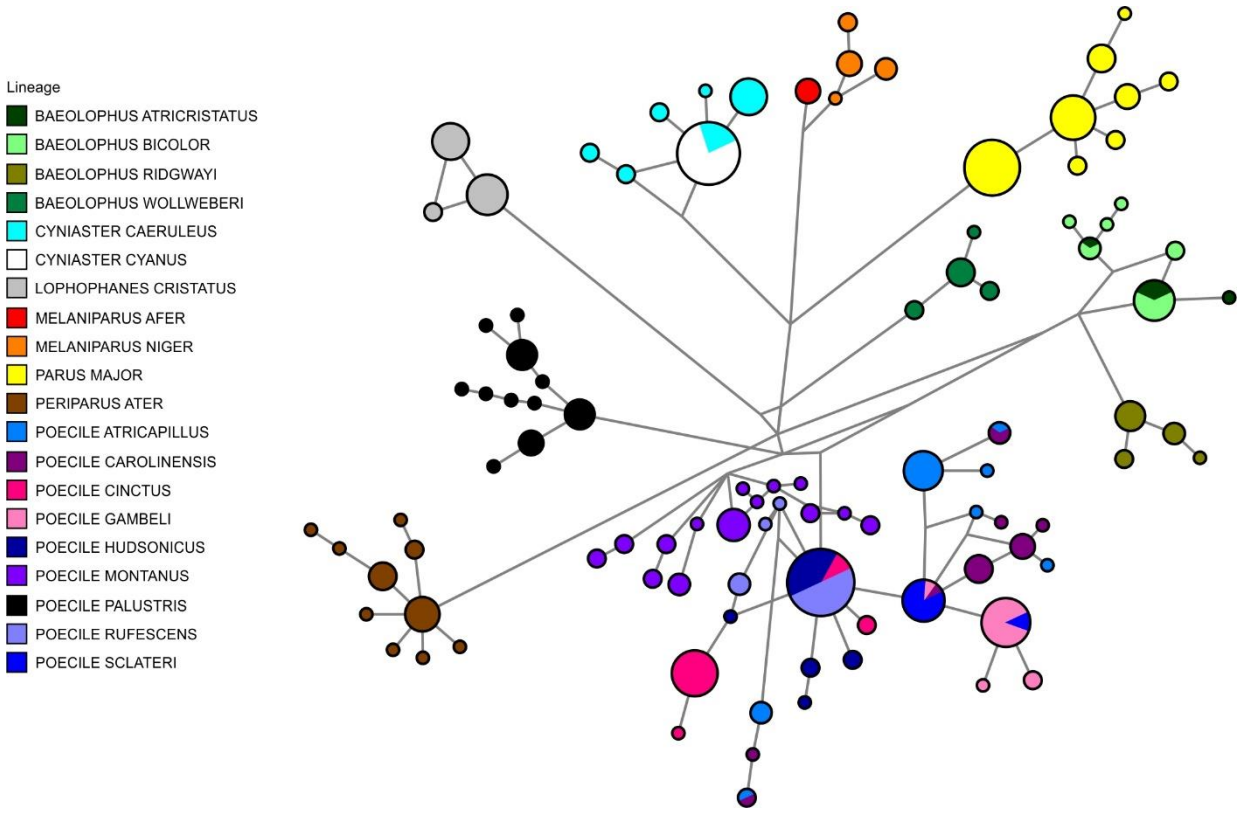
D) Nucleotide haplotype network of *MAMA* in Paridae



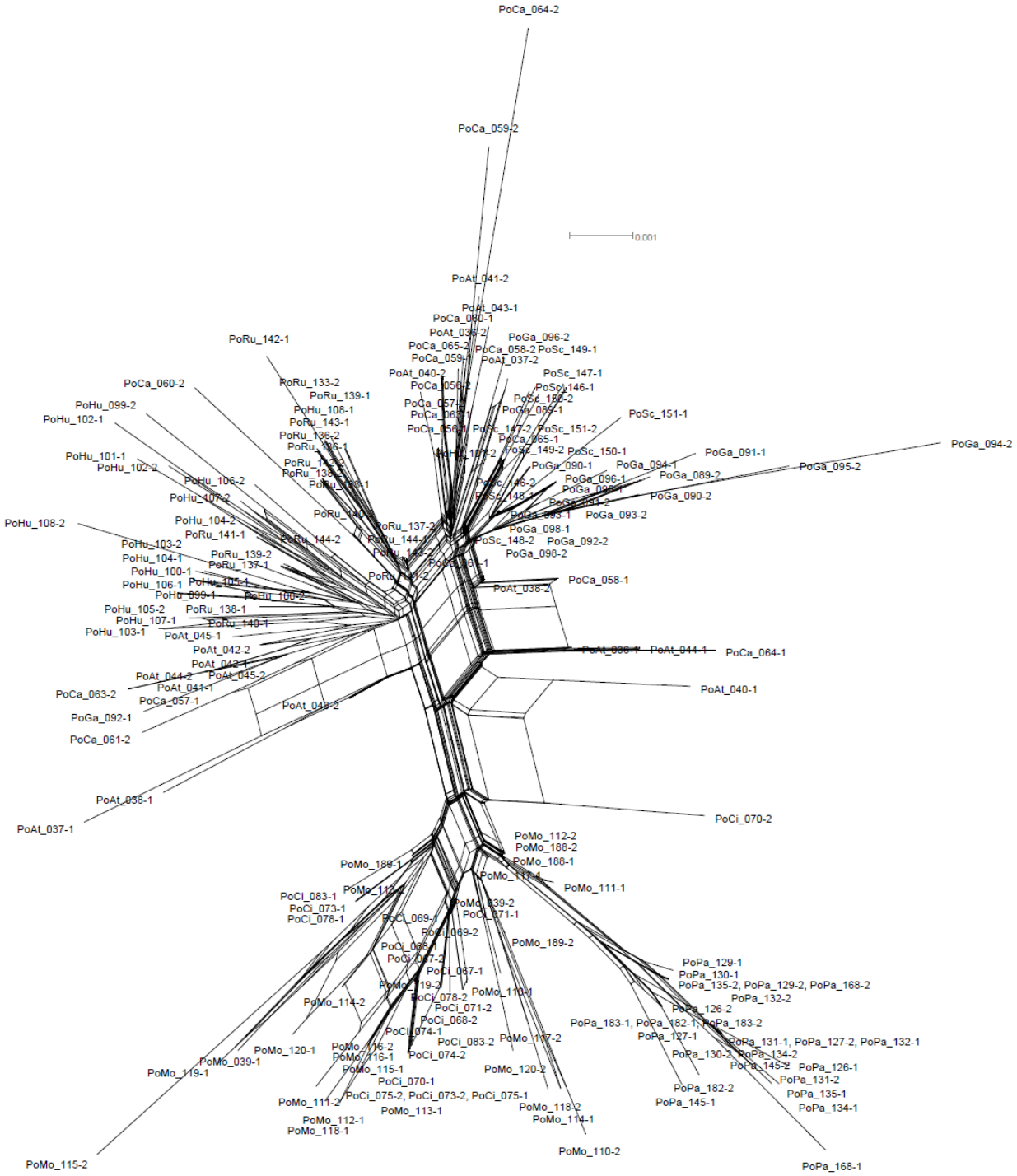
E) Nucleotide haplotype network of *TIAL* in Paridae



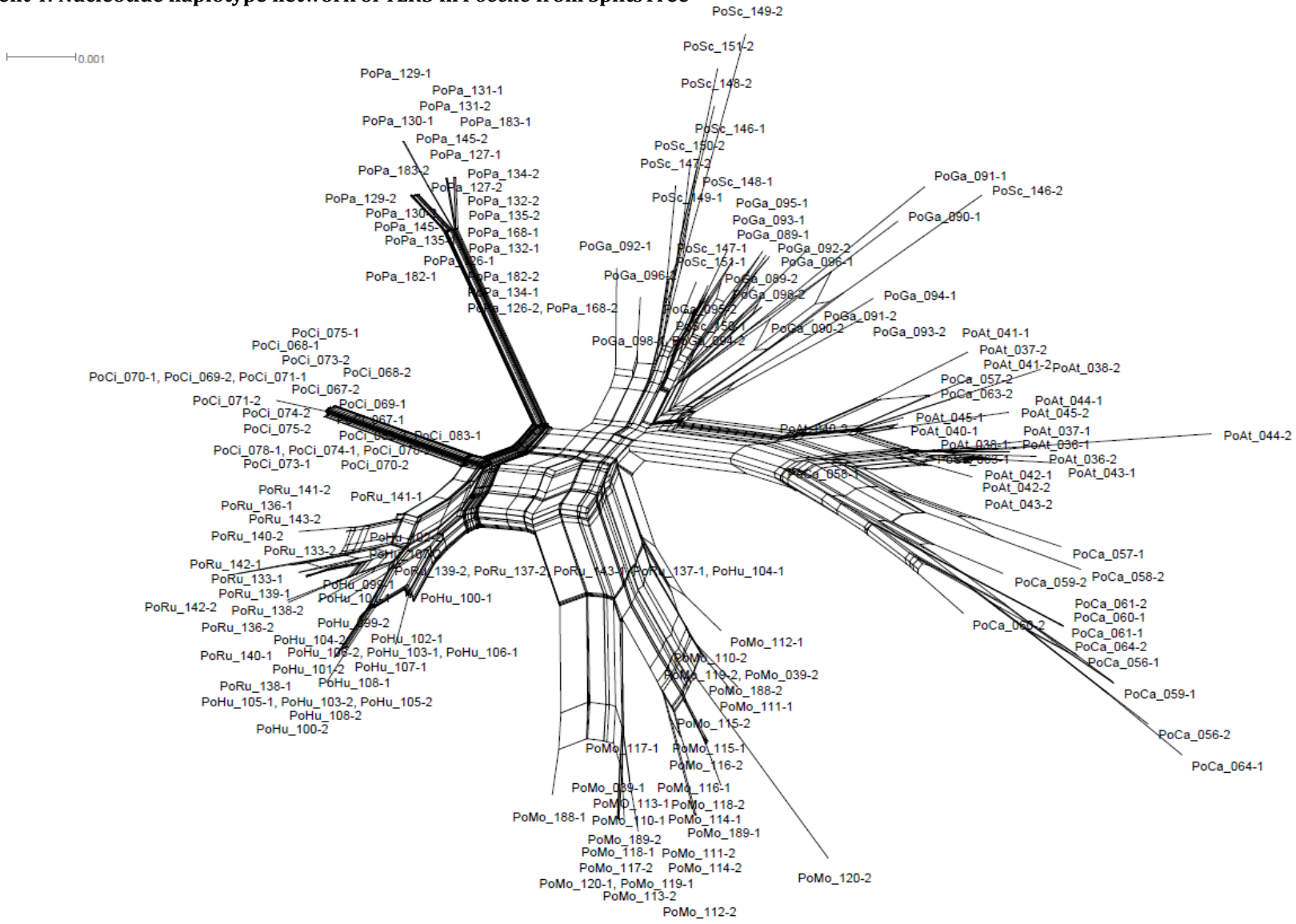
F) Nucleotide haplotype network of *UCHLP3* in Paridae



Supplement 3: Nucleotide haplotype network of *TLR4* in *Poecile* from SplitsTree



Supplement 4: Nucleotide haplotype network of *TLR5* in *Poecile* from SplitsTree



Supplement 5: Extracted nucleotide variable sites in *TLR4* for 20 representative sequences

This dataset used for selection and recombination analysis, I-TASSER modelling and PIPSA analysis. The sequence with the highest frequency in population for each species was chosen. Numbering shown here is from the beginning of the sequenced region (exon 3) for Paridae (starting at nucleotide position 714 of great tit CDs). The extraction of variable sites was done by FABOX web tool (Villesen, 2007).

```
                      1111111111111222222222223333333333344444444555555555566677777778
                    3477990001123337779123445668880112234456911667799022455666945812467 80
                    3145792344540150342124782271290891501219217257828002335478288116174 14
BaAtTLR4_BC001-1 TATAAGCCTTGGATATCCTACTGATGCCGAGCATGTTGGAAGTTACGTCACCTTGTCACTCTACCGGAC
BaBiTLR4_BC005-2 ..........C.C.....................................................TC.........
BaRiTLR4_BC175-2 .G.............................C............A......C.....C.........
BaWoTLR4_BC020-1 ........A..C.C..G.CG......A....TG........CA...C...........TC.........
CyCaTLR4_BC047-1 ....GTA....C.C..G.......C.A...........A..T....A.T........GTC.CGT..AG.
CyCyTLR4_BC072-2 ....GTA....C.C..GT......C.A..............T......T........GTC.CG.T.AG.
PaAfTLR4_BC025-1 ....GTA....C.C..G.......C.AT..C................A........CTGTC.....A...
PaNiTLR4_BC014-1 ....GTA....C.C..G.......C.AT..C.....G.........A..........GTC.....A..T
LoCrTLR4_BC177-1 ..C.GTAG...C.CG.G..GAC..C.A.......C......T....A.TG.......G.C.........
PaMaTLR4_BC161-1 .........CAC.C..G.......C.A..............T.C..A.T.........TCT......G.
PeAtTLR4_BC030-1 ...TGTA..C.C.C..G....C....AT.C.....C.....T..GT........A.TGTC.........
PoAtTLR4_BC037-1 ....GTA..C.C.CG.G....CA.C.A.A..........CG..........TT....TGTC.........
PoCaTLR4_BC056-1 ....GTA..C.CGCG.G....CA.C.A.A.....................TT....TGTC.........
PoCiTLR4_BC067-1 ....GTA..C.C.CG.G....CAGCAAA..C...................TTC....GTC.........
PoGaTLR4_BC092-2 ....GTA..C.C.CGGG....CA.C.A.A.....................TT....TGTC.........
PoHuTLR4_BC099-1 ....GTA..C.C.C..G....CA.C.A.A.C......C............TTC...TGTC.........
PoMoTLR4_BC112-2 ....GTA..C.C.CG.G....CA.CAAA..C...................TT.....GTC.........
PoPaTLR4_BC127-1 G...GTA..C.C.CG.G....CA.CAAA..C...............AC..TT.....GTC.........
PoRuTLR4_BC137-2 ....GTA..C.C.CG.G....CA.C.A.A.C...................TT....TGTC.........
PoScTLR4_BC149-2 ....GTA..C.C.CG.G....CA...A.T.....................TT....TGTC.........
```

Supplement 6: Extracted amino acid variable sites in *TLR4* for 20 representative sequences

This dataset (same as in Supplement 5) is used for selection and recombination analysis, I-TASSER modelling and PIPSA analysis. Numbering shown here is from the beginning of the sequenced region (exon 3) for Paridae (starting at amino acid position 238 of great tit translated TLR4 sequence.
The extraction of variable sites was done by FABOX web tool (Villesen, 2007).

```
                                  11111111111122
                 12333344557889900012333667956
                 45345924781384707943179074061
BaAtTLR4_BC001-1 QIKNLEVIISNDEAKQSELRKRNGQHTGS
BaBiTLR4_BC005-2 ......LT.....................
BaRiTLR4_BC175-2 R.....................S.....
BaWoTLR4_BC020-1 ....Q.LT.CD.....G....SKR.....
CyCaTLR4_BC047-1 ..DK..LT.C...........S.S..A.G
CyCyTLR4_BC072-2 ..DK..LT.C...........S....A.G
PaAfTLR4_BC025-1 ..DK..LT.C...V.H.......S..AD.
PaNiTLR4_BC014-1 ..DK..LT.C...V.H..R....S..AD.
LoCrTLR4_BC177-1 .TDKV.LT.CD......Q...S.SR.A..
PaMaTLR4_BC161-1 .....KLT.C...........S.S....G
PeAtTLR4_BC030-1 ..DK..LT.C...VQ......S....A..
PoAtTLR4_BC037-1 ..DK..LT.C.N.......SR....YA..
PoCaTLR4_BC056-1 ..DK..LA.C.N.............YA..
PoCiTLR4_BC067-1 ..DK..LT.C.SKE.H.........YA..
PoGaTLR4_BC092-2 ..DK..LTSC.N.............YA..
PoHuTLR4_BC099-1 ..DK..LT.C.N...H.........YA..
PoMoTLR4_BC112-2 ..DK..LT.C.NKE.H.........YA..
PoPaTLR4_BC127-1 ..DK..LT.C.NKE.H.......S.YA..
PoRuTLR4_BC137-2 ..DK..LT.C.N...H.........YA..
PoScTLR4_BC149-2 ..DK..LT.C.N.............YA..
```

Supplement 7: Extracted nucleotide variable sites in *TLR5* for 20 representative sequences

This dataset shown here is used for selection and recombination analysis, I-TASSER modelling and PIPSA analysis. The sequence with the highest frequency in population for each species was chosen. Numbering shown here is from the beginning of the sequenced region (exon 1) for Paridae (starting at nucleotide position 1 of great tit CDs.) The extraction of variable sites was done by FABOX web tool (Villesen, 2007).

```
                                  1111111111111112222222223333333333333333444444555555556666666666666666666666667777777777777888888888999999999990000000000000011111112222222222222223333
                        5557778899900112466677999911234578000111222367789901366744456669000011123333344556677889011244577888800133488800233455688912334678888992335688011244556678889113
                        1590125978957394146734123836285165123067157663845968367223951454135957950578903094956155917246916123757847502926759023313470054673567678356912078658346735670450
BaAtTLR5_BC001-1 TGGGGGAAACATTGCGTGTATTCGAGGCGTTGACATCTGATGTGCGCGATTAGGCGTCGCCTCATAATTGTCGAGCAGCCATTAATATGCCGCCTAAGCAATATTACTCAATCGTGTCTTACCAACCTAGACACAGGGGTAGAGTCTGTACAACATCTG
BaBiTLR5_BC010-2 ................................................G.....................................................A.............................G.............................C.
BaRiTLR5_BC015-2 .........G............A........G............A.....GA................................C...........A.G.............................G.........A.G..A.............C.
BaWoTLR5_BC019-1 ...A..G..T.......CC.......A..A....G...A.......TTG..G.....................A.....................AC.ATG..........TC........G.......G...........G..A.......C.....C.
CyCaTLR5_BC048-1 ...A.TGGGTGA.AG..C.TA.T.....TA.AG.....................G...CC.T.GC.G....CA....A...G.GC.T...CA.....ACCAT...G...T.......G....C.......CGA.TG...C.A.G..AC....G..CT..TC.
CyCyTLR5_BC079-2 ...A.TGGGTGA.AG..C.TA.T.....TA.AG.....................G...CC.T.GC.G....CA....A...G.GC.T...CA.....ACCAT.......T...........C.......CGA.TG...C.A.G..AC....G..CT..TC.
LoCrTLR5_BC177-1 .AAAA.G..T..A....C.T............G..C..............G.A........G..GG...G.A.A...G........C.AT...ACCATG.....T..A..............G...GA.T........G..AC....G.......C.
MeAfTLR5_BC025-2 ...A..C..T.A.A...C...C.......A.CG...T.....C........G.AT..A..........C.....A...G.GC.T...CA..A..ACCAT.....C..G...........C.......CGA.TG.......G..A..C..GTG..T..C.
MeNiTLR5_BC123-1 ...A..G..T.A.A...C...........A..G.................G.AT..G..........C.....A...G.GC.T..GCA.....ACCAT..CG...............C.G.....CGA.TG......CG..A..C..G.G..T..C.
PaMaTLR5_BC162-2 ...A..G..T.ACAG.AC....T.....TA.AGT...C.G..........G...C..T..C.G....C..A.GA...G.GC.TG..CA.....ACCAT...G......C...A.....C......TCGAGTG.......GA.A.....G.....C.C.
PeAtTLR5_BC031-1 C..A.CG..TGA...A...T..T....G.G..G......G.A.......C.G...C..TT.C.....CC..A..A..AG.GC.T...C.A..GTACCAT.T.G.........T..A.........G..GA.TG....T..G..A...ACG...T..TCA
PoAtTLR5_BC042-1 ...A..GG.T.A.....C.T....T....A.AG...........A......G..........G..........ATG.G...CT...C.AT...ACCATG...A....A..................GA.T.TG.....G..ACT.A.G.......C.
PoCaTLR5_BC060-1 ...A..GG.TGA.....C.T....T....A.AG.................GG..........G..........ATG.G........C.AT...ACCATG...A....A.......A.G....G....G..T.TG.....G..ACT...G.........
PoCiTLR5_BC073-1 ...A..G..TGA.....C.T....TA...AAAG.................G..........GG......A..A...G........C.AT...ACCATG...A....A.............G.....GA.T.TG..T..G..ACT...G.......C.
PoGaTLR5_BC098-1 ...A..G..TGA.....C.T....T....A.AG.................G..........G..........A...G....T...C.AT...ACCATG...A....A..................GA.T.TGT.T..G..ACT.A.G.......C.
PoHuTLR5_BC100-1 ...A..G..TGA.....C.T....T....AAAG..........C.......G..........G..........A...G........C.AT...ACCATG...A....A.............G.....GA.T.TG..T.AG..ACT...G.......CA
PoMoTLR5_BC111-1 ...A..G..TGA.....C.T....T....A.AG.......C..........G....C....C.G..........A...G....T...C.AT...ACCATG...A....A........T..........GA.T.TG..T..G.GACT...G.......CA
PoPaTLR5_BC126-1 ...A..G..T.A.....C.T....T....AAAG..C...G.A.........G..........G..........A...G........C.AT...ACCATG...A....A..................GA.T.TG.....G..ACT...G.......C.
PoRuTLR5_BC137-1 ...A..G..TGA.....C.T....T....AAAG..........C.......G..........G..........A...GT.......C.AT...ACCATG...A....A.............G.....GA.T.TG..T..G..ACT...G.......CA
PoScTLR5_BC150-1 ...A..G..TGA.....C.T....T....A.AG..........C.......G..........G..........A...G....T...C.AT...ACCATG...A....A..................GA.T.TGT.T..G..ACT.A.G.........
```

Supplement 8: Extracted amino acid variable sites in *TLR5* for 20 representative sequences

This dataset (same as in Supplement 7) is used for selection and recombination analysis, I-TASSER modelling and PIPSA analysis. Numbering shown here is from the beginning of the sequenced region (exon 1) for Paridae (starting at amino acid position 1 of great tit translated CDs. The extraction of variable sites was done by FABOX web tool (Villesen, 2007).

```
                                       1111111111111112222222222222222222222222222233333333333333333334444444
                 122233333444555667788800022334455888800000111112222233445566678990111233456667778990012223
                 904903568027568452402416925260568137912679234570256927193913932471248829462366896046983599
BaAtTLR5_BC001-1 VCGNQTFFSTVNSYFATHLSFRIDVSVYSKDAEFKFQLYGSHKAEHANSLKTLDSSNYASSFSFIYSGSQQNNYHSHMGDVIQRSIRIHD
BaBiTLR5_BC010-2 .............................E............................T...............................
BaRiTLR5_BC015-2 ..........................I..EN...........................T.....................I.RH......
BaWoTLR5_BC019-1 ..RD.I......TH.....T..VN...C.E.............T............K.M......F...R............RH......
CyCaTLR5_BC048-1 ..SDRVL.NR..TFYV..FT.H.......E..QSNL...D...T.D.SPF..SN..KSI...F..........H.G..A.I.RH.M.LY.
CyCyTLR5_BC079-2 ..SDRVL.NR..TFYV..FT.H.......E..QSNL...D...T.D.SPF..SN..KSI...F..........H.G..A.I.RH.M.LY.
LoCrTLR5_BC177-1 MYKD.I.Y....TF........T......E.T....E.C.A..T.D......SE..KSM....L........D.........RH.M....
MeAfTLR5_BC025-2 ..RH.IL.N...T......T.P.......E.I.L.........T.D.SPF..SNN.KSI....V.........H.G......RH.MG.L.
MeNiTLR5_BC123-1 ..RD.IL.N...T......T.........E.I.L.........T.D.SPF.ASN..KSI.P.........E..H.G.....TRH.MG.L.
PaMaTLR5_BC162-2 ..RD.ILSNR.KT..V..FT.H.G.....E..Q.N......NRT.D.SPFE.SN..KSI....L.N.......HRG......RH.M....
PeAtTLR5_BC031-1 ..SD.ML...I..F.V.Q.A...GI....E...........N.TKD.SPF..SE.FKSIC.......S.......G...Y..RH.T..YN
PoAtTLR5_BC042-1 ..RDRIL.....TF..S..T.H...R...E.............M.D...F..SE..KSM..Y.L............C.....RHFM....
PoCaTLR5_BC060-1 ..RDRML.....TF..S..T.H......AE.............M.D.....SE..KSM..Y.L....T.......C.....RHFM....
PoCiTLR5_BC073-1 ..RD.ML.....TF..S..TYH.......E......V...N.T.D......SE..KSM..Y.L.......K....C..Y..RHFM....
PoGaTLR5_BC098-1 ..RD.ML.....TF..S..T.H.......E.............T.D...F..SE..KSM..Y.L............CI.Y..RHFM....
PoHuTLR5_BC100-1 ..RD.ML.....TF..S..TYH.......E.............T.D......SE..KSM..Y.L.......K....C..Y.KRHFM...N
PoMoTLR5_BC111-1 ..RD.ML.....TF..S..T.H.......E...S.........T.D...F..SE..KSM..Y.L....L.......C..Y..RHFM...N
PoPaTLR5_BC126-1 ..RD.IL.....TF..S..TYHTGI....E.............T.D......SE..KSM..Y.L............C.....RHFM....
PoRuTLR5_BC137-1 ..RD.ML.....TF..S..TYH.......E.............T.DV.....SE..KSM..Y.L.......K....C..Y..RHFM...N
PoScTLR5_BC150-1 ..RD.ML.....TF..S..T.H.......E.............T.D...F..SE..KSM..Y.L............CI.Y..RHFM....
```

Supplement 9: Predicted binding sites of TLR4 and TLR5 and their comparison between Paridae and Galloanseres

TLR	Site	Residue function	Ref.	aa	Paridae conservatism	Galloaserae conservatism	HoSaTLR	MuMuTLR	DaReTLR
TLR4	238	MD-2 dimerization	1	R	uniformly	uniformly R	R234	R233	—
TLR4	267	LPS and MD-2 binding	1	R	uniformly	Galliformes T or S (MeGa), Aseriformes R	R264	K263	—
TLR4	293	MD-2 dimerization	1	L	uniformly	uniformly V	R289	R288	—
TLR4	344	LPS binding	2	S	uniformly	uniformly K or R (MeGa)	K341	Q339	—
TLR4	368	LPS binding	2	K/R	mostly K, R (PoAt)	uniformly K	K362	K360	—
TLR4	371	TLR dimerization	2	K	uniformly	uniformly N	N365	I363	—
TLR4	375	lipid IVa recognition	3	Q	uniformly	uniformly Q	E369	K367	—
TLR4	392	LPS binding	2,4	R	uniformly	uniformly R	G384	A382	—
TLR4	394	TLR dimerization	2	T	uniformly	uniformly S	S386	S384	—
TLR4	396	LPS binding	2	S	uniformly	Galliformes uniformly L, Aseriformes T	K388	S386	—
TLR4	419	TLR dimerization	2	N	uniformly	uniformly D	V411	A409	—
TLR4	423	LPS binding	3	T	uniformly	uniformly T	S415	S413	—
TLR4	424	MD-2 dimerization	2	G	uniformly	uniformly G	S416	A414	—
TLR4	425	MD-2 dimerization	2	D	uniformly	D or E (order *Gallus*)	N417	N415	—
TLR4	427	MD-2 dimerization	3	A	uniformly	Galliformes uniformly A, Aseriformes T	L419	M417	—
TLR4	424	TLR dimerization	2	G	uniformly	uniformly K	N433	T431	—
TLR4	444	LPS binding	2	G	uniformly	Galliformes uniformly H, AnAn D, AnPl N	Q436	R434	—
TLR4	447	MD-2 dimerization	2	S	uniformly	uniformly T	E439	E437	—
TLR4	448	LPS and MD-2 binding	2	Y	uniformly	uniformly Y	F440	F438	—
TLR4	452	LPS and MD-2 binding	2,3	L	uniformly	uniformly L	L444	L442	—
TLR4	453	MD-2 dimerization	3	S	uniformly	Galliformes uniformly L, Anseriformes S	S445	S443	—
TLR4	471	LPS and MD-2 binding	2	S	uniformly	uniformly S	F463	F461	—
TLR5	**33**	**FLA binding**	5	**I/V/ M/T**	I (BaWo, LoCr, *Melaniparus*, PaMa, PoAt, PoPa), M (PeAt, PoCa, PoCi, PoGa, PoHu, PoMo PoRu, PoSc), T (BaAt, BaRi, BaBi), V (*Cyanistes*)	Galliformes uniformly M, AnAn V, AnPl&TaTa M	F32	F32	I33
TLR5	**35**	**FLA binding**	5	**L/F**	mostly F, L (CyCa, CyCy, Melaniparus, PaMa)	Galliformes mostly N (NuMe S), AnAn Y, AnPl&TaTa N	R34	R34	I35

TLR	Site	Residue function	Ref.	aa	Paridae conservatism	Galloaserae conservatism	HoSaTLR	MuMuTLR	DaReTLR
TLR5	36	FLA binding	5	S/F/Y	mostly F, Y (LoCr), S (PaMa)	mostly S (PhCo F)	F35	G35	I35
TLR5	37	FLA binding	5	C	uniformly	uniformly C	C36	C36	R37
TLR5	53	FLA binding	5	F	uniformly	Galliformes uniformly F, Anseriformes uniformly L	L52	L53	D53
TLR5	55	FLA binding	5	T/S	mostly T, S (BaAt, BaBi, BaRi, PeAt)	Galliformes uniformly T, AnAn S, AnPl&TaTa N	S54	S55	S55
TLR5	**56**	**FLA binding**	5	**Y/F/H**	mostly F, L (CyCa, CyCy, Melaniparus, PaMa)	Galliformes uniformly Y, Anseriformes uniformly F	F55	F56	L56
TLR5	77	FLA binding	5	E	uniformly	uniformly E	E76	E77	K77
TLR5	79	FLA binding	5	G	uniformly	uniformly G	G78	G79	E79
TLR5	**80**	**FLA binding**	5	**T/A/S**	mostly T, S (BaRi, BaBi, BaAt), PeAt (A)	mostly T (AnPl&TaTa S)	S79	T80	Q80
TLR5	**106**	**FLA binding**	5	**G/D/N**	mostly D, G (PaMa, PeAt, PoPa), N (BaWo)	Galliformes uniformly F, AnAn Y, AnPl&TaTa Q	S104	Q105	Y105
TLR5	130	FLA binding	5	H	uniformly	mostly Q (NuMe R, AnPl&TaTa H)	F128	S129	Q129
TLR5	**156**	**FLA binding**	5	**A/I/T**	mostly A, I (Melaniparus), T (LoCr)	uniformly G	K154	G155	D155
TLR5	**181**	**FLA binding**	5	**F/L/S**	mostly F, S (Cyanistes, PoMo), L (Melaniparus)	uniformly F	S179	F180	F180
TLR5	**183**	**FLA binding**	5	**N/K**	mostly K, N (Cyanistes, PaMa)	Galliformes uniformly K, AnAn A, AnPl&TaTa D	Q181	Q182	K182
TLR5	**209**	**FLA binding**	5	**N/H**	mostly H, N (PaMa, PeAt, PoCi)	GaGa,GaLa,MeGa&NuMe T,PePe,PhCo&AnAn S, AnPl&TaTa Y	S207	K208	T208
TLR5	211	FLA binding	5	Y	uniformly	uniformly Y	Y209	F210	Q210
TLR5	214	FLA binding	5	E/K	mostly E, K (PeAt)	mostly D, AnPl&TaTa N	V212	V213	N213
TLR5	gap	FLA binding	5			position missing in Amniotes	—	—	Y215
TLR5	241	FLA binding	5	S/N	mostly S, N (PaNi)	mostly S (NuMe N)	T239	T240	K242
TLR5	gap	FLA binding	5			position missing in Amniotes	—	—	N265
TLR5	265	FLA binding	5	H	uniformly	uniformly H	H263	H264	Y267
TLR5	266	FLA binding	5	I	uniformly	mostly T (PhCo I)	I264	I265	N268
TLR5	268	FLA binding	5	G	uniformly	uniformly G	G266	G267	G270
TLR5	269	FLA binding	5,6	S/P	mostly S, P (MeNi)	uniformly S	A267	P268	S271
TLR5	270	FLA binding	5	G	uniformly	uniformly G	G268	G269	S272
TLR5	271	FLA binding, TLR dimerization	5	F	uniformly	uniformly F	F269	F270	F273

TLR	Site	Residue function	Ref.	aa	Paridae conservatism	Galloaserae conservatism	HoSaTLR	MuMuTLR	DaReTLR
TLR5	272	Pred. FLA binding	6	G	uniformly	uniformly G	G270	G271	G274
TLR5	273	FLA binding	5	F/Y	mostly F, Y (*Poecille*)	mostly F (AnPl&TaTa Y)	F271	F272	H275
TLR5	274	FLA binding	5	D	uniformly	uniformly N	H272	Q273	T276
TLR5	275	FLA binding	5	N	uniformly	uniformly N	N273	N274	N277
TLR5	276	FLA binding	5	L	uniformly	uniformly L	I274	I275	F278
TLR5	277	FLA binding	5	K	uniformly	uniformly K	K275	R276	K279
TLR5	296	Pred. FLA binding	6	D	uniformly	uniformly D	D294	D295	D298
TLR5	298	Pred. FLA binding	6	S	uniformly	uniformly S	S296	S297	S300
TLR5	301	FLA binding	5	Y/F	mostly Y, F (BaWo)	Galliformes mostly F (NuMe Y), Anseriformes uniformly Y	F299	F300	K303
TLR5	320	Pred. FLA binding	6	N	uniformly	uniformly N	N318	N319	T322
TLR5	322	Pred. FLA binding	6	S	uniformly	Galliformes uniformly F, AnAn S	A320	A321	A324
TLR5	344	Pred. FLA binding	6	N/D	mostly N, D (LoCr)	uniformly N	N342	N343	N346
TLR5	346	Pred. FLA binding	6	S	uniformly	uniformly S	S344	S345	S348
TLR5	347	TLR dimerization	5	S	uniformly	uniformly S	Y345	Y346	Q349
TLR5	348	TLR dimerization	5	N	uniformly	uniformly N	N346	N347	N350
TLR5	349	TLR dimerization	5	L	uniformly	uniformly L	L347	L348	F351
TLR5	352	FLA binding	5	E	uniformly	uniformly E	E350	E351	S354
TLR5	354	FLA binding	5	Y	uniformly	uniformly Y	Y352	Y353	D356
TLR5	367	Pred. FLA binding	6	I	uniformly	uniformly I	I365	V366	I369
TLR5	368	Pred. FLA binding	6	Y	uniformly	uniformly D	D366	D367	D370
TLR5	371	TLR dimerization	5	Q	uniformly	uniformly Q	K369	R370	Y373
TLR5	373	TLR dimerization	5	H	uniformly	uniformly H	H371	H372	H375
TLR5	375	TLR dimerization	5	G	uniformly	uniformly G	A373	G374	R377
TLR5	376	FLA binding	5	M/I	mostly M, I (PoGa, PoSc)	uniformly M	I374	I375	A378
TLR5	378	FLA binding	5	G/A	mostly G, A (*Cyanistes*)	mostly G (TaTa D)	Q376	Q377	G380
TLR5	**379**	**FLA binding**	5	**D/Y**	mostly D, Y (PeAt, PoCi, PoGa, PoHu, PoMo, PoRu, PoSc)	GaGa,GaLa,MeGa&NuMe E,PePe,PhCo Q, Anseriformes uniformly Q	D377	D378	D381
TLR5	380	FLA binding	5	K	uniformly	uniformly K	Q378	Q379	Q382
TLR5	391	Pred. FLA binding	6	I	uniformly	uniformly I	L389	L390	L393
TLR5	392	Pred. FLA binding	6	D	uniformly	uniformly D, AnAn N	D390	D391	N394

Numbering of amino acid sites is according to great tit translated *TLR4* and *TLR5* sequences. Function of particular residues and their polymorphism are shown for different taxa. *Homo sapiens* (HoSa), *Mus musculus* (MuMu), *Danio rerio* (DaRu) References are cited as follows: 1 -Kim et al., (2007); 2- Park et al., (2009); 3- Ohto et al., (2012); 4- Walsh et al., (2008);5- Yoon et al., (2013); 6 -Andersen-Nissen et al., (2007). Adjusted from Vinkler et al., (2014).

Supplement 10: Full list of conservative and non-conservative amino acid sites for *TLR4* gene identified by ConSurf

Legend: normalized conservativity score (score), confidence interval (conf. int.), colour in three-dimensional model (colour), colour confidence interval (colour conf. int.) and substitutions. Position below the confidence cut-off are labelled by an asterisk *. Numbering of residues is according to translated great tit *TLR4* sequence.

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
238	R	-0.427	-0.564,-0.461	9	9,9	R
239	S	-0.459	-0.564,-0.488	9	9,9	S
240	A	-0.446	-0.564,-0.461	9	9,9	A
241	F	-0.373	-0.564,-0.381	9	9,9	F
242	E	-0.418	-0.564,-0.425	9	9,9	E
243	N	-0.448	-0.564,-0.461	9	9,9	N
244	F	-0.373	-0.564,-0.381	9	9,9	F
245	M	0.215	-0.381, 0.574	3*	9,1	I,M
246	M	3.336	1.499, 5.400	1	1,1	R,M
247	Q	-0.429	-0.564,-0.461	9	9,9	Q
248	T	-0.452	-0.564,-0.461	9	9,9	T
249	S	-0.459	-0.564,-0.488	9	9,9	S
250	L	-0.379	-0.564,-0.425	9	9,9	L
251	Q	1.017	0.111, 1.499	1	4,1	R,Q
252	G	-0.375	-0.564,-0.381	9	9,9	G
253	L	-0.379	-0.564,-0.425	9	9,9	L
254	A	0.807	-0.041, 1.499	1*	5,1	T,A
255	G	-0.375	-0.564,-0.381	9	9,9	G
256	L	-0.379	-0.564,-0.425	9	9,9	L
257	Q	-0.429	-0.564,-0.461	9	9,9	Q
258	V	-0.445	-0.564,-0.461	9	9,9	V
259	S	0.105	-0.381, 0.309	4*	9,2	G,S
260	R	-0.427	-0.564,-0.461	9	9,9	R
261	L	-0.379	-0.564,-0.425	9	9,9	L
262	I	2.146	0.574, 2.448	1	1,1	T,V,M,I
263	V	0.187	-0.381, 0.574	3*	9,1	A,V
264	G	-0.375	-0.564,-0.381	9	9,9	G
265	E	-0.418	-0.564,-0.425	9	9,9	E
266	F	-0.373	-0.564,-0.381	9	9,9	F
267	R	3.735	1.499, 5.400	1	1,1	W,R,K
268	D	-0.429	-0.564,-0.461	9	9,9	D
269	S	-0.459	-0.564,-0.488	9	9,9	S
270	K	0.342	-0.324, 0.574	2*	8,1	D,K
271	N	0.314	-0.324, 0.574	2*	8,1	K,N
272	L	1.7	0.309, 2.448	1	2,1	L,Q,V
273	Q	-0.429	-0.564,-0.461	9	9,9	Q
274	D	-0.429	-0.564,-0.461	9	9,9	D
275	F	-0.373	-0.564,-0.381	9	9,9	F
276	K	2.051	0.574, 2.448	1	1,1	K,E
277	R	-0.427	-0.564,-0.461	9	9,9	R
278	G	-0.375	-0.564,-0.381	9	9,9	G

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
279	L	1.592	0.309, 2.448	1	2,1	F,V,L
280	L	-0.379	-0.564,-0.425	9	9,9	L
281	T	4.829	2.448, 5.400	1	1,1	A,T,I
282	G	-0.375	-0.564,-0.381	9	9,9	G
283	L	-0.379	-0.564,-0.425	9	9,9	L
284	C	-0.358	-0.564,-0.381	9	9,9	C
285	Q	1.019	0.111, 1.499	1	4,1	Q,R
286	V	-0.445	-0.564,-0.461	9	9,9	V
287	Q	-0.429	-0.564,-0.461	9	9,9	Q
288	M	0.216	-0.381, 0.574	3*	9,1	M,I
289	E	-0.418	-0.564,-0.425	9	9,9	E
290	E	-0.418	-0.564,-0.425	9	9,9	E
291	F	-0.373	-0.564,-0.381	9	9,9	F
292	V	-0.445	-0.564,-0.461	9	9,9	V
293	L	-0.379	-0.564,-0.425	9	9,9	L
294	I	3.906	1.499, 5.400	1	1,1	I,S
295	C	0.554	-0.251, 0.945	1*	7,1	C,S
296	F	-0.373	-0.564,-0.381	9	9,9	F
297	R	-0.427	-0.564,-0.461	9	9,9	R
298	E	-0.418	-0.564,-0.425	9	9,9	E
299	F	-0.373	-0.564,-0.381	9	9,9	F
300	E	-0.418	-0.564,-0.425	9	9,9	E
301	D	1.061	0.111, 1.499	1	4,1	D,H
302	D	-0.429	-0.564,-0.461	9	9,9	D
303	T	-0.452	-0.564,-0.461	9	9,9	T
304	D	-0.429	-0.564,-0.461	9	9,9	D
305	T	-0.452	-0.564,-0.461	9	9,9	T
306	L	-0.379	-0.564,-0.425	9	9,9	L
307	F	-0.373	-0.564,-0.381	9	9,9	F
308	N	0.791	-0.041, 1.499	1*	5,1	N,D
309	C	-0.358	-0.564,-0.381	9	9,9	C
310	I	-0.448	-0.564,-0.461	9	9,9	I
311	G	2.913	0.945, 5.400	1	1,1	G,R,S
312	N	-0.448	-0.564,-0.461	9	9,9	N
313	V	0.814	-0.041, 1.499	1*	5,1	V,I
314	S	0.687	-0.041, 0.945	1*	5,1	P,S
315	T	-0.452	-0.564,-0.461	9	9,9	T
316	V	0.822	-0.041, 1.499	1*	5,1	I,V
317	R	1.837	0.574, 2.448	1	1,1	H,R
318	L	-0.379	-0.564,-0.425	9	9,9	L
319	V	-0.445	-0.564,-0.461	9	9,9	V
320	D	1.574	0.309, 2.448	1	2,1	D,S,N
321	L	-0.379	-0.564,-0.425	9	9,9	L
322	G	-0.375	-0.564,-0.381	9	9,9	G
323	L	-0.379	-0.564,-0.425	9	9,9	L
324	E	-0.418	-0.564,-0.425	9	9,9	E
325	E	2.051	0.574, 2.448	1	1,1	E,K

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
326	I	-0.448	-0.564,-0.461	9	9,9	I
327	S	-0.459	-0.564,-0.488	9	9,9	S
328	Q	-0.429	-0.564,-0.461	9	9,9	Q
329	V	-0.445	-0.564,-0.461	9	9,9	V
330	P	-0.389	-0.564,-0.425	9	9,9	P
331	A	5.002	2.448, 5.400	1	1,1	A,V,E
332	R	0.303	-0.324, 0.574	2*	8,1	R,G
333	S	-0.459	-0.564,-0.488	9	9,9	S
334	K	2.129	0.574, 2.448	1	1,1	K,E,Q
335	V	0.189	-0.381, 0.574	3*	9,1	M,V
336	K	-0.413	-0.564,-0.425	9	9,9	K
337	Q	1.751	0.574, 2.448	1	1,1	Q,H
338	L	-0.379	-0.564,-0.425	9	9,9	L
339	E	-0.418	-0.564,-0.425	9	9,9	E
340	C	-0.358	-0.564,-0.381	9	9,9	C
341	K	-0.413	-0.564,-0.425	9	9,9	K
342	K	-0.413	-0.564,-0.425	9	9,9	K
343	C	-0.358	-0.564,-0.381	9	9,9	C
344	S	0.111	-0.381, 0.309	4*	9,2	S,G
345	F	-0.373	-0.564,-0.381	9	9,9	F
346	E	0.358	-0.324, 0.574	1*	8,1	E,Q
347	D	0.293	-0.324, 0.574	2*	8,1	G,D
348	V	-0.445	-0.564,-0.461	9	9,9	V
349	P	-0.389	-0.564,-0.425	9	9,9	P
350	A	-0.446	-0.564,-0.461	9	9,9	A
351	L	2.871	0.945, 5.400	1	1,1	R,W,L
352	K	-0.413	-0.564,-0.425	9	9,9	K
353	L	-0.379	-0.564,-0.425	9	9,9	L
354	S	-0.459	-0.564,-0.488	9	9,9	S
355	L	0.619	-0.251, 0.945	1*	7,1	F,L
356	F	-0.373	-0.564,-0.381	9	9,9	F
357	K	-0.413	-0.564,-0.425	9	9,9	K
358	E	-0.418	-0.564,-0.425	9	9,9	E
359	L	0.619	-0.251, 0.945	1*	7,1	L,V
360	R	0.303	-0.324, 0.574	2*	8,1	R,S
361	V	-0.445	-0.564,-0.461	9	9,9	V
362	L	0.62	-0.251, 0.945	1*	7,1	F,L
363	R	5.274	2.448, 5.400	1	1,1	S,H,C,R
364	I	1.417	0.309, 2.448	1	2,1	V,I
365	T	-0.452	-0.564,-0.461	9	9,9	T
366	K	-0.413	-0.564,-0.425	9	9,9	K
367	N	-0.448	-0.564,-0.461	9	9,9	N
368	K	1.193	0.111, 1.499	1	4,1	R,N,K
369	R	1.071	0.111, 1.499	1	4,1	R,G,K
370	L	-0.379	-0.564,-0.425	9	9,9	L
371	K	-0.413	-0.564,-0.425	9	9,9	K
372	N	-0.448	-0.564,-0.461	9	9,9	N

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
373	F	-0.373	-0.564,-0.381	9	9,9	F
374	S	0.924	-0.041, 1.499	1*	5,1	S,R
375	Q	0.293	-0.324, 0.574	2*	8,1	Q,E
376	N	2.133	0.574, 2.448	1	1,1	N,K
377	F	-0.373	-0.564,-0.381	9	9,9	F
378	E	0.358	-0.324, 0.574	1*	8,1	E,K
379	G	-0.375	-0.564,-0.381	9	9,9	G
380	L	-0.379	-0.564,-0.425	9	9,9	L
381	T	0.758	-0.041, 1.499	1*	5,1	T,S,P
382	N	0.78	-0.041, 1.499	1*	5,1	K,D,N
383	L	-0.379	-0.564,-0.425	9	9,9	L
384	E	-0.418	-0.564,-0.425	9	9,9	E
385	V	-0.445	-0.564,-0.461	9	9,9	V
386	I	-0.448	-0.564,-0.461	9	9,9	I
387	D	-0.429	-0.564,-0.461	9	9,9	D
388	L	-0.379	-0.564,-0.425	9	9,9	L
389	S	-0.459	-0.564,-0.488	9	9,9	S
390	E	-0.418	-0.564,-0.425	9	9,9	E
391	N	-0.448	-0.564,-0.461	9	9,9	N
392	R	-0.427	-0.564,-0.461	9	9,9	R
393	L	-0.379	-0.564,-0.425	9	9,9	L
394	T	-0.452	-0.564,-0.461	9	9,9	T
395	F	-0.373	-0.564,-0.381	9	9,9	F
396	S	-0.459	-0.564,-0.488	9	9,9	S
397	S	5.399	2.448, 5.400	1	1,1	T,S,R,G
398	C	-0.358	-0.564,-0.381	9	9,9	C
399	C	-0.358	-0.564,-0.381	9	9,9	C
400	S	-0.459	-0.564,-0.488	9	9,9	S
401	P	0.55	-0.251, 0.945	1*	7,1	R,P
402	Q	-0.429	-0.564,-0.461	9	9,9	Q
403	F	-0.373	-0.564,-0.381	9	9,9	F
404	Q	0.297	-0.324, 0.574	2*	8,1	R,Q
405	N	-0.448	-0.564,-0.461	9	9,9	N
406	C	0.757	-0.251, 1.499	1*	7,1	S,C
407	P	-0.389	-0.564,-0.425	9	9,9	P
408	N	-0.448	-0.564,-0.461	9	9,9	N
409	L	-0.379	-0.564,-0.425	9	9,9	L
410	K	-0.413	-0.564,-0.425	9	9,9	K
411	H	0.413	-0.324, 0.945	1*	8,1	Y,H
412	L	-0.379	-0.564,-0.425	9	9,9	L
413	N	-0.448	-0.564,-0.461	9	9,9	N
414	L	-0.379	-0.564,-0.425	9	9,9	L
415	S	-0.459	-0.564,-0.488	9	9,9	S
416	F	-0.373	-0.564,-0.381	9	9,9	F
417	N	-0.448	-0.564,-0.461	9	9,9	N
418	S	-0.459	-0.564,-0.488	9	9,9	S
419	N	0.171	-0.381, 0.574	3*	9,1	Y,N

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
420	I	-0.448	-0.564,-0.461	9	9,9	I
421	R	-0.427	-0.564,-0.461	9	9,9	R
422	L	-0.379	-0.564,-0.425	9	9,9	L
423	T	-0.452	-0.564,-0.461	9	9,9	T
424	G	-0.375	-0.564,-0.381	9	9,9	G
425	D	-0.429	-0.564,-0.461	9	9,9	D
426	F	-0.373	-0.564,-0.381	9	9,9	F
427	T	4.261	1.499, 5.400	1	1,1	G,T,A
428	N	-0.448	-0.564,-0.461	9	9,9	N
429	V	-0.445	-0.564,-0.461	9	9,9	V
430	K	-0.413	-0.564,-0.425	9	9,9	K
431	N	-0.448	-0.564,-0.461	9	9,9	N
432	L	-0.379	-0.564,-0.425	9	9,9	L
433	L	-0.379	-0.564,-0.425	9	9,9	L
434	Y	-0.367	-0.564,-0.381	9	9,9	Y
435	L	-0.379	-0.564,-0.425	9	9,9	L
436	D	-0.429	-0.564,-0.461	9	9,9	D
437	L	0.62	-0.251, 0.945	1*	7,1	L,F
438	Q	-0.429	-0.564,-0.461	9	9,9	Q
439	H	-0.437	-0.564,-0.461	9	9,9	H
440	T	-0.452	-0.564,-0.461	9	9,9	T
441	T	-0.452	-0.564,-0.461	9	9,9	T
442	L	1.698	0.309, 2.448	1	2,1	L,V
443	F	-0.373	-0.564,-0.381	9	9,9	F
444	G	-0.375	-0.564,-0.381	9	9,9	G
445	P	-0.389	-0.564,-0.425	9	9,9	P
446	G	-0.375	-0.564,-0.381	9	9,9	G
447	S	-0.459	-0.564,-0.488	9	9,9	S
448	Y	-0.367	-0.564,-0.381	9	9,9	Y
449	P	-0.389	-0.564,-0.425	9	9,9	P
450	V	0.187	-0.381, 0.574	3*	9,1	V,A
451	F	-0.373	-0.564,-0.381	9	9,9	F
452	L	-0.379	-0.564,-0.425	9	9,9	L
453	S	-0.459	-0.564,-0.488	9	9,9	S
454	L	-0.379	-0.564,-0.425	9	9,9	L
455	Q	-0.429	-0.564,-0.461	9	9,9	Q
456	K	-0.413	-0.564,-0.425	9	9,9	K
457	L	-0.379	-0.564,-0.425	9	9,9	L
458	I	-0.448	-0.564,-0.461	9	9,9	I
459	Y	-0.367	-0.564,-0.381	9	9,9	Y
460	L	-0.379	-0.564,-0.425	9	9,9	L
461	D	-0.429	-0.564,-0.461	9	9,9	D
462	I	-0.448	-0.564,-0.461	9	9,9	I
463	S	-0.459	-0.564,-0.488	9	9,9	S
464	H	0.923	-0.041, 1.499	1*	5,1	H,P,Y
465	T	-0.452	-0.564,-0.461	9	9,9	T
466	K	-0.413	-0.564,-0.425	9	9,9	K

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
467	T	-0.452	-0.564,-0.461	9	9,9	T
468	E	-0.418	-0.564,-0.425	9	9,9	E
469	V	-0.445	-0.564,-0.461	9	9,9	V
470	K	-0.413	-0.564,-0.425	9	9,9	K
471	S	-0.459	-0.564,-0.488	9	9,9	S
472	Q	-0.429	-0.564,-0.461	9	9,9	Q
473	C	-0.358	-0.564,-0.381	9	9,9	C
474	T	-0.452	-0.564,-0.461	9	9,9	T
475	F	-0.373	-0.564,-0.381	9	9,9	F
476	C	-0.358	-0.564,-0.381	9	9,9	C
477	G	-0.375	-0.564,-0.381	9	9,9	G
478	L	-0.379	-0.564,-0.425	9	9,9	L
479	N	-0.448	-0.564,-0.461	9	9,9	N
480	S	-0.459	-0.564,-0.488	9	9,9	S
481	L	-0.379	-0.564,-0.425	9	9,9	L
482	Q	-0.429	-0.564,-0.461	9	9,9	Q
483	V	-0.445	-0.564,-0.461	9	9,9	V
484	L	-0.379	-0.564,-0.425	9	9,9	L
485	K	-0.413	-0.564,-0.425	9	9,9	K
486	M	-0.441	-0.564,-0.461	9	9,9	M
487	A	-0.446	-0.564,-0.461	9	9,9	A
488	G	-0.375	-0.564,-0.381	9	9,9	G
489	N	-0.448	-0.564,-0.461	9	9,9	N
490	S	-0.459	-0.564,-0.488	9	9,9	S
491	F	-0.373	-0.564,-0.381	9	9,9	F
492	E	-0.418	-0.564,-0.425	9	9,9	E
493	G	0.64	-0.251, 0.945	1*	7,1	G,D
494	N	0.168	-0.381, 0.574	3*	9,1	S,N
495	K	-0.413	-0.564,-0.425	9	9,9	K
496	L	-0.379	-0.564,-0.425	9	9,9	L
497	A	-0.446	-0.564,-0.461	9	9,9	A
498	G	0.71	-0.041, 0.945	1*	5,1	S,G
499	N	-0.448	-0.564,-0.461	9	9,9	N
500	F	-0.373	-0.564,-0.381	9	9,9	F
501	K	0.39	-0.324, 0.945	1*	8,1	K,Q
502	N	-0.448	-0.564,-0.461	9	9,9	N
503	L	-0.379	-0.564,-0.425	9	9,9	L
504	S	-0.459	-0.564,-0.488	9	9,9	S
505	H	-0.437	-0.564,-0.461	9	9,9	H
506	L	-0.379	-0.564,-0.425	9	9,9	L
507	H	-0.437	-0.564,-0.461	9	9,9	H
508	T	-0.452	-0.564,-0.461	9	9,9	T
509	L	-0.379	-0.564,-0.425	9	9,9	L
510	D	-0.429	-0.564,-0.461	9	9,9	D
511	I	-0.448	-0.564,-0.461	9	9,9	I
512	S	-0.459	-0.564,-0.488	9	9,9	S
513	S	-0.459	-0.564,-0.488	9	9,9	S

Supplement 11: Full list of conservative and non-conservative amino acid sites for *TLR5* gene identified by ConSurf

Legend: normalized conservativity score (score), confidence interval (conf. int), colour in three-dimensional model (colour), colour confidence interval (colour conf. int.) and substitutions. Numbering of residues is according to translated great tit *TLR5* sequence. Position below the confidence cut-off are labelled by an asterisk *.

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
1	M	-0.522	-0.777,-0.479	9	9,9	M
2	M	-0.522	-0.777,-0.479	9	9,9	M
3	L	-0.417	-0.772,-0.349	8	9,8	L
4	C	-0.385	-0.772,-0.269	8	9,7	C
5	H	-0.515	-0.777,-0.479	9	9,9	H
6	Q	-0.500	-0.777,-0.419	9	9,8	Q
7	L	-0.417	-0.772,-0.349	8	9,8	L
8	L	-0.417	-0.772,-0.349	8	9,8	L
9	L	-0.417	-0.772,-0.349	8	9,8	L
10	V	0.205	-0.479, 0.626	3*	9,1	I,V
11	F	-0.408	-0.772,-0.349	8	9,8	F
12	G	0.684	-0.349, 1.290	1*	8,1	G,S
13	L	-0.417	-0.772,-0.349	8	9,8	L
14	S	-0.555	-0.777,-0.531	9	9,9	S
15	L	-0.417	-0.772,-0.349	8	9,8	L
16	A	-0.530	-0.777,-0.479	9	9,9	A
17	S	0.108	-0.479, 0.401	4*	9,2	S,R
18	G	-0.411	-0.772,-0.349	8	9,8	G
19	V	0.981	-0.065, 1.823	1*	6,1	M,V
20	C	0.795	-0.269, 1.290	1*	7,1	C,Y
21	A	-0.530	-0.777,-0.479	9	9,9	A
22	S	-0.555	-0.777,-0.531	9	9,9	S
23	R	-0.497	-0.777,-0.419	9	9,8	R
24	R	4.448	1.823, 5.198	1	1,1	S,K,R,G
25	C	-0.385	-0.772,-0.269	8	9,7	C
26	Y	-0.398	-0.772,-0.349	8	9,8	Y
27	S	-0.555	-0.777,-0.531	9	9,9	S
28	E	-0.482	-0.777,-0.419	9	9,8	E
29	D	1.159	0.063, 1.823	1	4,1	D,N,H
30	Q	3.829	1.823, 5.198	1	1,1	Q,R
31	V	-0.529	-0.777,-0.479	9	9,9	V
32	S	-0.555	-0.777,-0.531	9	9,9	S
33	I	5.165	2.692, 5.198	1	1,1	T,V,M,I
34	Y	-0.398	-0.772,-0.349	8	9,8	Y
35	L	0.701	-0.349, 1.290	1*	8,1	F,L
36	S	2.930	0.913, 5.198	1	1,1	S,F,Y
37	C	0.792	-0.269, 1.290	1*	7,1	C,Y
38	N	0.850	-0.065, 1.290	1*	6,1	N,S
39	L	-0.417	-0.772,-0.349	8	9,8	L
40	R	0.188	-0.479, 0.626	3*	9,1	R,T
41	D	-0.500	-0.777,-0.419	9	9,8	D

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
42	V	0.206	-0.479, 0.626	3*	9,1	V,I
43	P	-0.433	-0.772,-0.349	9	9,8	P
44	P	0.573	-0.349, 0.913	1*	8,1	S,P
45	V	-0.529	-0.777,-0.479	9	9,9	V
46	P	-0.433	-0.772,-0.349	9	9,8	P
47	K	0.194	-0.479, 0.626	3*	9,1	K,N
48	D	-0.500	-0.777,-0.419	9	9,8	D
49	T	-0.542	-0.777,-0.479	9	9,9	T
50	V	-0.529	-0.777,-0.479	9	9,9	V
51	K	-0.473	-0.777,-0.419	9	9,8	K
52	L	-0.417	-0.772,-0.349	8	9,8	L
53	F	2.978	0.913, 5.198	1	1,1	F,L
54	L	-0.417	-0.772,-0.349	8	9,8	L
55	T	0.864	-0.065, 1.290	1*	6,1	S,T
56	Y	3.940	1.823, 5.198	1	1,1	H,Y,F
57	N	-0.535	-0.777,-0.479	9	9,9	N
58	F	0.705	-0.349, 1.290	1*	8,1	F,Y
59	I	-0.534	-0.777,-0.479	9	9,9	I
60	R	-0.497	-0.777,-0.419	9	9,8	R
61	Q	-0.500	-0.777,-0.419	9	9,8	Q
62	V	-0.529	-0.777,-0.479	9	9,9	V
63	T	1.606	0.401, 2.692	1	2,1	T,N,A
64	V	4.119	1.823, 5.198	1	1,1	V,A,E
65	T	0.831	-0.065, 1.290	1*	6,1	S,T,I
66	S	-0.555	-0.777,-0.531	9	9,9	S
67	F	-0.408	-0.772,-0.349	8	9,8	F
68	P	-0.433	-0.772,-0.349	9	9,8	P
69	L	-0.417	-0.772,-0.349	8	9,8	L
70	L	-0.417	-0.772,-0.349	8	9,8	L
71	E	-0.482	-0.777,-0.419	9	9,8	E
72	H	0.271	-0.479, 0.626	3*	9,1	H,Q
73	L	-0.417	-0.772,-0.349	8	9,8	L
74	F	0.683	-0.349, 1.290	1*	8,1	L,F
75	L	-0.417	-0.772,-0.349	8	9,8	L
76	L	-0.417	-0.772,-0.349	8	9,8	L
77	E	-0.482	-0.777,-0.419	9	9,8	E
78	L	-0.417	-0.772,-0.349	8	9,8	L
79	G	-0.411	-0.772,-0.349	8	9,8	G
80	T	1.642	0.401, 2.692	1	2,1	A,S,T
81	Q	-0.500	-0.777,-0.419	9	9,8	Q
82	F	1.868	0.401, 2.692	1	2,1	Y,F
83	V	-0.529	-0.777,-0.479	9	9,9	V
84	H	4.754	2.692, 5.198	1	1,1	S,P,H,R
85	P	-0.433	-0.772,-0.349	9	9,8	P
86	V	0.952	-0.065, 1.290	1*	6,1	I,V
87	T	0.862	-0.065, 1.290	1*	6,1	A,I,T
88	I	-0.534	-0.777,-0.479	9	9,9	I

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
89	G	-0.411	-0.772,-0.349	8	9,8	G
90	K	-0.473	-0.777,-0.419	9	9,8	K
91	G	-0.411	-0.772,-0.349	8	9,8	G
92	A	-0.530	-0.777,-0.479	9	9,9	A
93	F	-0.408	-0.772,-0.349	8	9,8	F
94	R	-0.497	-0.777,-0.419	9	9,8	R
95	N	-0.535	-0.777,-0.479	9	9,9	N
96	L	-0.417	-0.772,-0.349	8	9,8	L
97	P	-0.433	-0.772,-0.349	9	9,8	P
98	N	0.183	-0.479, 0.626	4*	9,1	K,N
99	L	-0.417	-0.772,-0.349	8	9,8	L
100	R	3.045	1.290, 5.198	1	1,1	C,R,H
101	I	1.753	0.401, 2.692	1	2,1	I,V,T
102	L	-0.417	-0.772,-0.349	8	9,8	L
103	D	-0.500	-0.777,-0.419	9	9,8	D
104	L	-0.417	-0.772,-0.349	8	9,8	L
105	G	-0.411	-0.772,-0.349	8	9,8	G
106	G	3.997	1.823, 5.198	1	1,1	G,N,D
107	N	-0.535	-0.777,-0.479	9	9,9	N
108	K	-0.473	-0.777,-0.419	9	9,8	K
109	V	0.961	-0.065, 1.290	1*	6,1	I,V
110	L	-0.417	-0.772,-0.349	8	9,8	L
111	Q	-0.500	-0.777,-0.419	9	9,8	Q
112	L	-0.417	-0.772,-0.349	8	9,8	L
113	D	-0.500	-0.777,-0.419	9	9,8	D
114	L	-0.417	-0.772,-0.349	8	9,8	L
115	D	-0.500	-0.777,-0.419	9	9,8	D
116	A	-0.530	-0.777,-0.479	9	9,9	A
117	F	-0.408	-0.772,-0.349	8	9,8	F
118	V	-0.529	-0.777,-0.479	9	9,9	V
119	G	0.685	-0.349, 1.290	1*	8,1	D,G
120	L	-0.417	-0.772,-0.349	8	9,8	L
121	P	-0.433	-0.772,-0.349	9	9,8	P
122	S	0.118	-0.479, 0.401	4*	9,2	S,R
123	L	-0.417	-0.772,-0.349	8	9,8	L
124	T	-0.542	-0.777,-0.479	9	9,9	T
125	V	0.207	-0.479, 0.626	3*	9,1	I,V
126	L	-0.417	-0.772,-0.349	8	9,8	L
127	R	-0.497	-0.777,-0.419	9	9,8	R
128	L	-0.417	-0.772,-0.349	8	9,8	L
129	F	-0.408	-0.772,-0.349	8	9,8	F
130	H	-0.515	-0.777,-0.479	9	9,9	H
131	N	-0.535	-0.777,-0.479	9	9,9	N
132	Y	1.936	0.401, 2.692	1	2,1	Y,C
133	L	-0.417	-0.772,-0.349	8	9,8	L
134	G	-0.411	-0.772,-0.349	8	9,8	G
135	D	0.323	-0.419, 0.626	2*	8,1	D,N

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
136	S	0.106	-0.479, 0.401	4*	9,2	A,S
137	I	-0.534	-0.777,-0.479	9	9,9	I
138	L	-0.417	-0.772,-0.349	8	9,8	L
139	E	-0.482	-0.777,-0.419	9	9,8	E
140	E	0.397	-0.419, 0.913	2*	8,1	K,E
141	R	-0.497	-0.777,-0.419	9	9,8	R
142	Y	-0.398	-0.772,-0.349	8	9,8	Y
143	F	-0.408	-0.772,-0.349	8	9,8	F
144	Q	0.322	-0.419, 0.626	2*	8,1	Q,E
145	D	1.168	0.063, 1.823	1	4,1	D,N
146	L	-0.417	-0.772,-0.349	8	9,8	L
147	R	1.208	0.063, 1.823	1	4,1	G,R
148	S	-0.555	-0.777,-0.531	9	9,9	S
149	L	-0.417	-0.772,-0.349	8	9,8	L
150	E	-0.482	-0.777,-0.419	9	9,8	E
151	E	-0.482	-0.777,-0.419	9	9,8	E
152	L	-0.417	-0.772,-0.349	8	9,8	L
153	D	-0.500	-0.777,-0.419	9	9,8	D
154	L	-0.417	-0.772,-0.349	8	9,8	L
155	S	-0.555	-0.777,-0.531	9	9,9	S
156	A	1.027	0.063, 1.823	1	4,1	A,I,T
157	N	-0.535	-0.777,-0.479	9	9,9	N
158	Q	1.294	0.063, 1.823	1	4,1	Q,E
159	I	0.188	-0.479, 0.626	3*	9,1	V,I
160	T	0.157	-0.479, 0.626	4*	9,1	K,T
161	K	-0.473	-0.777,-0.419	9	9,8	K
162	L	-0.417	-0.772,-0.349	8	9,8	L
163	H	-0.515	-0.777,-0.479	9	9,9	H
164	P	-0.433	-0.772,-0.349	9	9,8	P
165	H	-0.515	-0.777,-0.479	9	9,9	H
166	P	-0.433	-0.772,-0.349	9	9,8	P
167	L	-0.417	-0.772,-0.349	8	9,8	L
168	F	-0.408	-0.772,-0.349	8	9,8	F
169	Y	-0.398	-0.772,-0.349	8	9,8	Y
170	N	0.911	-0.065, 1.290	1*	6,1	N,K
171	L	-0.417	-0.772,-0.349	8	9,8	L
172	T	-0.542	-0.777,-0.479	9	9,9	T
173	A	-0.530	-0.777,-0.479	9	9,9	A
174	L	-0.417	-0.772,-0.349	8	9,8	L
175	K	-0.473	-0.777,-0.419	9	9,8	K
176	S	0.106	-0.479, 0.401	4*	9,2	S,N
177	V	-0.529	-0.777,-0.479	9	9,9	V
178	N	-0.535	-0.777,-0.479	9	9,9	N
179	L	-0.417	-0.772,-0.349	8	9,8	L
180	K	-0.473	-0.777,-0.419	9	9,8	K
181	F	3.693	1.823, 5.198	1	1,1	L,F,S
182	N	-0.535	-0.777,-0.479	9	9,9	N

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
183	N	1.252	0.063, 1.823	1	4,1	N,K
184	I	-0.534	-0.777,-0.479	9	9,9	I
185	S	-0.555	-0.777,-0.531	9	9,9	S
186	N	0.907	-0.065, 1.290	1*	6,1	N,S
187	F	1.861	0.401, 2.692	1	2,1	L,F
188	C	-0.385	-0.772,-0.269	8	9,7	C
189	Q	0.326	-0.419, 0.626	2*	8,1	E,Q
190	T	-0.542	-0.777,-0.479	9	9,9	T
191	N	-0.535	-0.777,-0.479	9	9,9	N
192	L	-0.417	-0.772,-0.349	8	9,8	L
193	T	-0.542	-0.777,-0.479	9	9,9	T
194	S	-0.555	-0.777,-0.531	9	9,9	S
195	F	-0.408	-0.772,-0.349	8	9,8	F
196	Q	-0.500	-0.777,-0.419	9	9,8	Q
197	G	-0.411	-0.772,-0.349	8	9,8	G
198	K	-0.473	-0.777,-0.419	9	9,8	K
199	H	-0.515	-0.777,-0.479	9	9,9	H
200	F	-0.408	-0.772,-0.349	8	9,8	F
201	L	0.661	-0.349, 1.290	1*	8,1	V,L
202	Y	0.745	-0.269, 1.290	1*	7,1	Y,C
203	F	-0.408	-0.772,-0.349	8	9,8	F
204	N	-0.535	-0.777,-0.479	9	9,9	N
205	L	0.662	-0.349, 1.290	1*	8,1	F,L
206	G	0.679	-0.349, 1.290	1*	8,1	D,G
207	S	0.113	-0.479, 0.401	4*	9,2	S,A
208	N	-0.535	-0.777,-0.479	9	9,9	N
209	N	4.879	2.692, 5.198	1	1,1	Q,H,N
210	L	0.662	-0.349, 1.290	1*	8,1	F,L
211	Y	-0.398	-0.772,-0.349	8	9,8	Y
212	R	0.433	-0.419, 0.913	1*	8,1	K,R
213	T	1.661	0.401, 2.692	1	2,1	A,M,T
214	E	0.409	-0.419, 0.913	2*	8,1	E,K
215	D	0.323	-0.419, 0.626	2*	8,1	H,D
216	V	0.205	-0.479, 0.626	3*	9,1	A,V
217	A	0.947	-0.065, 1.290	1*	6,1	A,V
218	W	-0.275	-0.772,-0.175	7	9,6	W
219	A	-0.530	-0.777,-0.479	9	9,9	A
220	S	0.159	-0.479, 0.626	4*	9,1	S,N
221	C	-0.385	-0.772,-0.269	8	9,7	C
222	P	0.243	-0.479, 0.626	3*	9,1	S,P
223	N	-0.535	-0.777,-0.479	9	9,9	N
224	P	-0.433	-0.772,-0.349	9	9,8	P
225	F	4.783	2.692, 5.198	1	1,1	L,F
226	E	0.437	-0.419, 0.913	1*	8,1	E,K
227	D	-0.500	-0.777,-0.419	9	9,8	D
228	I	-0.534	-0.777,-0.479	9	9,9	I
229	T	0.160	-0.479, 0.626	4*	9,1	A,T

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
230	F	-0.408	-0.772,-0.349	8	9,8	F
231	S	-0.555	-0.777,-0.531	9	9,9	S
232	S	0.170	-0.479, 0.626	4*	9,1	L,S
233	L	-0.417	-0.772,-0.349	8	9,8	L
234	D	-0.500	-0.777,-0.419	9	9,8	D
235	L	-0.417	-0.772,-0.349	8	9,8	L
236	S	-0.555	-0.777,-0.531	9	9,9	S
237	N	1.297	0.063, 1.823	1	4,1	N,E,D
238	N	-0.535	-0.777,-0.479	9	9,9	N
239	G	-0.411	-0.772,-0.349	8	9,8	G
240	W	-0.275	-0.772,-0.175	7	9,6	W
241	S	0.105	-0.531, 0.401	4*	9,2	N,S
242	T	-0.542	-0.777,-0.479	9	9,9	T
243	E	-0.482	-0.777,-0.419	9	9,8	E
244	R	-0.497	-0.777,-0.419	9	9,8	R
245	V	-0.529	-0.777,-0.479	9	9,9	V
246	Q	0.321	-0.419, 0.626	2*	8,1	H,Q
247	Y	-0.398	-0.772,-0.349	8	9,8	Y
248	L	0.662	-0.349, 1.290	1*	8,1	F,L
249	S	1.578	0.401, 2.692	1	2,1	S,C,F
250	T	-0.542	-0.777,-0.479	9	9,9	T
251	A	-0.530	-0.777,-0.479	9	9,9	A
252	I	-0.534	-0.777,-0.479	9	9,9	I
253	K	0.405	-0.419, 0.913	2*	8,1	K,N
254	G	-0.411	-0.772,-0.349	8	9,8	G
255	T	-0.542	-0.777,-0.479	9	9,9	T
256	Q	-0.500	-0.777,-0.419	9	9,8	Q
257	I	-0.534	-0.777,-0.479	9	9,9	I
258	S	0.774	-0.065, 1.290	1*	6,1	G,R,S
259	S	0.906	-0.065, 1.290	1*	6,1	Y,S
260	L	-0.417	-0.772,-0.349	8	9,8	L
261	I	2.674	0.913, 5.198	1	1,1	T,A,I,M,V
262	F	-0.408	-0.772,-0.349	8	9,8	F
263	S	0.147	-0.479, 0.401	4*	9,2	C,S
264	T	-0.542	-0.777,-0.479	9	9,9	T
265	H	-0.515	-0.777,-0.479	9	9,9	H
266	I	0.189	-0.479, 0.626	3*	9,1	T,I
267	M	-0.522	-0.777,-0.479	9	9,9	M
268	G	-0.411	-0.772,-0.349	8	9,8	G
269	S	0.116	-0.479, 0.401	4*	9,2	P,S
270	G	-0.411	-0.772,-0.349	8	9,8	G
271	F	-0.408	-0.772,-0.349	8	9,8	F
272	G	-0.411	-0.772,-0.349	8	9,8	G
273	F	0.717	-0.349, 1.290	1*	8,1	F,Y
274	D	-0.500	-0.777,-0.419	9	9,8	D
275	N	-0.535	-0.777,-0.479	9	9,9	N
276	L	-0.417	-0.772,-0.349	8	9,8	L

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
277	K	-0.473	-0.777,-0.419	9	9,8	K
278	N	-0.535	-0.777,-0.479	9	9,9	N
279	P	-0.433	-0.772,-0.349	9	9,8	P
280	D	-0.500	-0.777,-0.419	9	9,8	D
281	I	-0.534	-0.777,-0.479	9	9,9	I
282	S	0.130	-0.479, 0.401	4*	9,2	S,F
283	T	-0.542	-0.777,-0.479	9	9,9	T
284	F	-0.408	-0.772,-0.349	8	9,8	F
285	A	0.200	-0.479, 0.626	3*	9,1	T,A
286	G	-0.411	-0.772,-0.349	8	9,8	G
287	L	-0.417	-0.772,-0.349	8	9,8	L
288	G	-0.411	-0.772,-0.349	8	9,8	G
289	R	0.335	-0.419, 0.626	2*	8,1	K,R
290	S	-0.555	-0.777,-0.531	9	9,9	S
291	N	-0.535	-0.777,-0.479	9	9,9	N
292	L	-0.417	-0.772,-0.349	8	9,8	L
293	N	-0.535	-0.777,-0.479	9	9,9	N
294	F	2.965	0.913, 5.198	1	1,1	V,L,F
295	F	-0.408	-0.772,-0.349	8	9,8	F
296	D	-0.500	-0.777,-0.419	9	9,8	D
297	L	0.210	-0.479, 0.626	3*	9,1	L,I
298	S	-0.555	-0.777,-0.531	9	9,9	S
299	H	-0.515	-0.777,-0.479	9	9,9	H
300	G	-0.411	-0.772,-0.349	8	9,8	G
301	Y	1.934	0.401, 2.692	1	2,1	F,Y
302	I	-0.534	-0.777,-0.479	9	9,9	I
303	F	-0.408	-0.772,-0.349	8	9,8	F
304	S	-0.555	-0.777,-0.531	9	9,9	S
305	L	-0.417	-0.772,-0.349	8	9,8	L
306	N	-0.535	-0.777,-0.479	9	9,9	N
307	S	-0.555	-0.777,-0.531	9	9,9	S
308	L	-0.417	-0.772,-0.349	8	9,8	L
309	I	-0.534	-0.777,-0.479	9	9,9	I
310	F	-0.408	-0.772,-0.349	8	9,8	F
311	Q	-0.500	-0.777,-0.419	9	9,8	Q
312	N	0.113	-0.479, 0.401	4*	9,2	S,N
313	L	-0.417	-0.772,-0.349	8	9,8	L
314	G	1.703	0.216, 2.692	1	3,1	V,S,G
315	N	-0.535	-0.777,-0.479	9	9,9	N
316	L	-0.417	-0.772,-0.349	8	9,8	L
317	E	-0.482	-0.777,-0.419	9	9,8	E
318	S	2.226	0.913, 2.692	1	1,1	T,S,L
319	L	-0.417	-0.772,-0.349	8	9,8	L
320	N	-0.535	-0.777,-0.479	9	9,9	N
321	L	-0.417	-0.772,-0.349	8	9,8	L
322	S	-0.555	-0.777,-0.531	9	9,9	S
323	K	0.431	-0.419, 0.913	2*	8,1	Q,K

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
324	N	-0.535	-0.777,-0.479	9	9,9	N
325	K	-0.473	-0.777,-0.419	9	9,8	K
326	I	-0.534	-0.777,-0.479	9	9,9	I
327	N	-0.535	-0.777,-0.479	9	9,9	N
328	Q	1.184	0.063, 1.823	1	4,1	R,K,Q
329	I	-0.534	-0.777,-0.479	9	9,9	I
330	Q	-0.500	-0.777,-0.419	9	9,8	Q
331	R	0.336	-0.419, 0.626	2*	8,1	K,R
332	Q	0.325	-0.419, 0.626	2*	8,1	E,Q
333	A	-0.530	-0.777,-0.479	9	9,9	A
334	F	-0.408	-0.772,-0.349	8	9,8	F
335	F	0.698	-0.349, 1.290	1*	8,1	F,Y
336	G	-0.411	-0.772,-0.349	8	9,8	G
337	L	-0.417	-0.772,-0.349	8	9,8	L
338	G	-0.411	-0.772,-0.349	8	9,8	G
339	N	0.932	-0.065, 1.290	1*	6,1	K,N
340	L	-0.417	-0.772,-0.349	8	9,8	L
341	K	0.430	-0.419, 0.913	2*	8,1	K,R
342	T	-0.542	-0.777,-0.479	9	9,9	T
343	L	-0.417	-0.772,-0.349	8	9,8	L
344	N	0.917	-0.065, 1.290	1*	6,1	D,S,N
345	L	0.663	-0.349, 1.290	1*	8,1	L,I
346	S	-0.555	-0.777,-0.531	9	9,9	S
347	S	0.105	-0.479, 0.401	4*	9,2	N,S
348	N	-0.535	-0.777,-0.479	9	9,9	N
349	L	-0.417	-0.772,-0.349	8	9,8	L
350	L	-0.417	-0.772,-0.349	8	9,8	L
351	G	-0.411	-0.772,-0.349	8	9,8	G
352	E	-0.482	-0.777,-0.419	9	9,8	E
353	L	-0.417	-0.772,-0.349	8	9,8	L
354	Y	-0.398	-0.772,-0.349	8	9,8	Y
355	D	-0.500	-0.777,-0.419	9	9,8	D
356	H	0.621	-0.349, 1.290	1*	8,1	Y,H
357	T	-0.542	-0.777,-0.479	9	9,9	T
358	F	-0.408	-0.772,-0.349	8	9,8	F
359	E	-0.482	-0.777,-0.419	9	9,8	E
360	G	-0.411	-0.772,-0.349	8	9,8	G
361	L	-0.417	-0.772,-0.349	8	9,8	L
362	R	0.263	-0.479, 0.626	3*	9,1	R,H
363	G	0.261	-0.479, 0.626	3*	9,1	G,S
364	V	0.205	-0.479, 0.626	3*	9,1	I,V
365	M	-0.522	-0.777,-0.479	9	9,9	M
366	H	0.423	-0.419, 0.913	2*	8,1	H,C
367	I	-0.534	-0.777,-0.479	9	9,9	I
368	Y	-0.398	-0.772,-0.349	8	9,8	Y
369	L	-0.417	-0.772,-0.349	8	9,8	L
370	Q	-0.500	-0.777,-0.419	9	9,8	Q

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
371	Q	-0.500	-0.777,-0.419	9	9,8	Q
372	N	-0.535	-0.777,-0.479	9	9,9	N
373	H	-0.515	-0.777,-0.479	9	9,9	H
374	I	-0.534	-0.777,-0.479	9	9,9	I
375	G	-0.411	-0.772,-0.349	8	9,8	G
376	M	1.008	0.063, 1.823	1	4,1	I,M
377	I	-0.534	-0.777,-0.479	9	9,9	I
378	G	0.660	-0.349, 1.290	1*	8,1	G,A
379	D	4.075	1.823, 5.198	1	1,1	Y,D
380	K	0.433	-0.419, 0.913	1*	8,1	E,K
381	S	-0.555	-0.777,-0.531	9	9,9	S
382	F	-0.408	-0.772,-0.349	8	9,8	F
383	R	0.338	-0.419, 0.626	2*	8,1	W,R
384	Q	-0.500	-0.777,-0.419	9	9,8	Q
385	L	-0.417	-0.772,-0.349	8	9,8	L
386	V	3.475	1.290, 5.198	1	1,1	V,I
387	N	-0.535	-0.777,-0.479	9	9,9	N
388	L	-0.417	-0.772,-0.349	8	9,8	L
389	K	-0.473	-0.777,-0.419	9	9,8	K
390	I	1.720	0.401, 2.692	1	2,1	I,T,K
391	I	-0.534	-0.777,-0.479	9	9,9	I
392	D	-0.500	-0.777,-0.419	9	9,8	D
393	L	-0.417	-0.772,-0.349	8	9,8	L
394	R	0.336	-0.419, 0.626	2*	8,1	R,Q
395	D	-0.500	-0.777,-0.419	9	9,8	D
396	N	-0.535	-0.777,-0.479	9	9,9	N
397	A	-0.530	-0.777,-0.479	9	9,9	A
398	I	-0.534	-0.777,-0.479	9	9,9	I
399	K	-0.473	-0.777,-0.419	9	9,8	K
400	R	1.195	0.063, 1.823	1	4,1	R,K
401	L	0.663	-0.349, 1.290	1*	8,1	L,V
402	P	-0.433	-0.772,-0.349	9	9,8	P
403	S	-0.555	-0.777,-0.531	9	9,9	S
404	F	-0.408	-0.772,-0.349	8	9,8	F
405	P	-0.433	-0.772,-0.349	9	9,8	P
406	H	1.085	0.063, 1.823	1	4,1	H,R
407	L	-0.417	-0.772,-0.349	8	9,8	L
408	T	-0.542	-0.777,-0.479	9	9,9	T
409	S	1.053	0.063, 1.823	1	4,1	F,S
410	A	-0.530	-0.777,-0.479	9	9,9	A
411	F	-0.408	-0.772,-0.349	8	9,8	F
412	L	-0.417	-0.772,-0.349	8	9,8	L
413	G	-0.411	-0.772,-0.349	8	9,8	G
414	D	-0.500	-0.777,-0.419	9	9,8	D
415	N	-0.535	-0.777,-0.479	9	9,9	N
416	K	-0.473	-0.777,-0.419	9	9,8	K
417	L	-0.417	-0.772,-0.349	8	9,8	L

Residue	PaMa seq.	score	conf. int.	colour	colour conf. int.	substitutions
418	M	1.003	0.063, 1.823	1	4,1	M,I,T
419	S	-0.555	-0.777,-0.531	9	9,9	S
420	V	-0.529	-0.777,-0.479	9	9,9	V
421	A	0.200	-0.479, 0.626	3*	9,1	A,S
422	D	-0.500	-0.777,-0.419	9	9,8	D
423	R	0.355	-0.419, 0.913	2*	8,1	G,R
424	A	-0.530	-0.777,-0.479	9	9,9	A
425	I	0.215	-0.479, 0.626	3*	9,1	L,I
426	T	-0.542	-0.777,-0.479	9	9,9	T
427	A	-0.530	-0.777,-0.479	9	9,9	A
428	T	-0.542	-0.777,-0.479	9	9,9	T
429	H	2.156	0.626, 2.692	1	1,1	L,Y,H
430	L	-0.417	-0.772,-0.349	8	9,8	L
431	E	-0.482	-0.777,-0.419	9	9,8	E
432	L	-0.417	-0.772,-0.349	8	9,8	L
433	E	-0.482	-0.777,-0.419	9	9,8	E
434	R	-0.497	-0.777,-0.419	9	9,8	R
435	N	-0.535	-0.777,-0.479	9	9,9	N
436	W	-0.275	-0.772,-0.175	7	9,6	W
437	L	-0.417	-0.772,-0.349	8	9,8	L
438	S	-0.555	-0.777,-0.531	9	9,9	S
439	D	5.041	2.692, 5.198	1	1,1	D,N
440	L	-0.417	-0.772,-0.349	8	9,8	L
441	G	-0.411	-0.772,-0.349	8	9,8	G
442	D	-0.500	-0.777,-0.419	9	9,8	D
443	L	-0.417	-0.772,-0.349	8	9,8	L
444	Y	-0.398	-0.772,-0.349	8	9,8	Y
445	I	-0.534	-0.777,-0.479	9	9,9	I
446	L	-0.417	-0.772,-0.349	8	9,8	L
447	F	-0.408	-0.772,-0.349	8	9,8	F