Genes of innate immunity and their significance in evolutionary ecology of free livings rodents

Alena Fornuskova

▶ **To cite this version:**

Alena Fornuskova. Genes of innate immunity and their significance in evolutionary ecology of free livings rodents. Populations and Evolution [q-bio.PE]. Université Montpellier II - Sciences et Techniques du Languedoc; Masarykova univerzita (Brno, République tchèque), 2013. English. NNT : 2013MON20103 . tel-01021258

HAL Id: tel-01021258

https://theses.hal.science/tel-01021258

Submitted on 9 Jul 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE MONTPELLIER II
SCIENCES ET TECHNIQUES DU LANGUEDOC
FACULTE DES SCIENCES

and

MASARYK UNIVERSITY, BRNO
FACULTY OF SCIENCE

THESIS

To obtain doctoral degree

Formation doctorale: Biologie de l'évolution et écologie
Ecole Doctorale: Systèmes Intégrés en Biologie, Agronomie, Géosciences, Hydrosciences, Environnement, SIBAGHE

Presented and defended publicly

AUTHOR: Alena Fornuskova

GENES OF INNATE IMMUNITY AND THEIR SIGNIFICANCE IN EVOLUTIONARY ECOLOGY OF FREE LIVING RODENTS

Toll-like receptor polymorphisms in rodents

Thesis supervised by **Dr. Jean-François Cosson/Dr. Josef Bryja**

Date and place of defence: the 19 December 2013 in CBGP

COMMITTEE:

M. Pierre Boursot, Director of research CNRS, Université Montpellier 2, Montpellier, FR	**Examiner**
M. Petr Hořín, Professor, University of Veterinary and Pharmaceutical Sciences, Brno, CZ	**Reviewer**
M. Dirk Werling, Professor, The Royal Veterinary College, University of London, UK	**Reviewer**
M. Jan Zima, Professor, Institute of Vertebrate Biology AS CR, Brno, CZ	**Examiner**
M. Josef Bryja, Assoc. professor, Institute of Vertebrate Biology AS CR and Masaryk University, Brno, CZ	**Supervisor**

ABSTRACT

Appropriate recognition of parasites is crucial for effective immune response, ensuring activation of adequate defence mechanisms. In vertebrates, it has frequently been demonstrated that genes encoding proteins involved in pathogen recognition by an adaptive immune system are often subject to intense selection pressures. On the contrary, much less information has been provided on the evolution of recognition mechanisms of innate immunity. The aim of this thesis is to describe the pattern of natural variation of innate immunity genes involved in pathogen recognition in rodents and to analyze the mechanisms of their evolution. We used murine rodents (subfamily Murinae) as a principal model group because they often live in our close proximity and thus are potential reservoirs of various pathogens dangerous to humans.

First, we studied the intraspecific variability of five bacterial sensing Toll-like receptors (TLR1, TLR2, TLR4, TLR5, and TLR6) in inbred strains derived from two subspecies of the house mouse (*M. m. musculus*, hereafter abbreviated as Mmm and *Mus musculus domesticus*, Mmd). Wild-derived inbred strains are suitable tools for studying variation of immunity genes because they provide information about alleles that occur in natural populations, and at the same time they occur at homozygous state. The most significant results include the findings of a stop codon in exon 2 of the *Tlr5* gene in one Mmm strain and no variability in *Tlr4* of Mmd. The results also provide the set of diagnostic SNPs for each gene allowing future studies of introgression of immunity genes across the house mouse hybrid zone and their possible role in the speciation process.

Following these results we decided to check whether the absence of *Tlr4* polymorphism in Mmd reflects the pattern found in natural populations, or whether it is a consequence of insufficient sampling or subsequent breeding. We therefore sequenced *Tlr4* in both subspecies across a large part of the Western Palaearctic region (in total 39 Mmm and 62 Mmd individuals), then we compared these results with variability on mitochondrial DNA (cytochrome b). The result confirmed our prediction that observed variability in Mmd is strongly reduced also in free-living populations (compared to Mmm), probably due to strong purifying selection by pathogens with which they met during the westward colonization. However, the influence of random evolutionary processes (e.g. drift during bottlenecks) cannot be excluded based on our data. At the intraspecific level, we could not find any sign of positive selection. Our results revealed also species specific variants of *Tlr4* and an important role of recombination in Mmm during evolutionary history.

The last part of my dissertation is devoted to interspecific comparison of two receptors, TLR4 and TLR7. These two TLRs differ in the exposure and the ligands detection. TLR4 is an extracellular receptor detecting mainly bacterial ligands (especially lipopolysaccharides), while TLR7 is located inside the cell and detects ssRNA viruses. The aim of this part of the thesis was to describe variability of both receptors at the interspecific level and to reveal selection forces acting on TLRs in longer evolutionary time scale. In total we analyzed 23 rodent species of the subfamily Murinae in Europe, Asia and Africa. Our results suggest that purifying selection has been a dominant force in evolution of the *Tlr4* and *Tlr*7 genes, but we also demonstrated that episodic diversifying selection has shaped the present species-specific variation in rodent *Tlr*s. Sites under positive selection were concentrated mainly in the extracellular domain of both receptors, which is responsible for ligand binding. The comparison between two TLRs lead us to the conclusion that the intracellular TLR7 is under much stronger negative selection pressure, presumably due to its interaction with viral nucleic acids, which are similar to those of the host and even small changes in TLR7 conformation could cause autoimmunity.

Key words: Toll-like receptors, receptors of innate immunity, Pattern Recognition Receptors, selection, evolution, natural selection, genetic polymorphism, phylogeny, host-parasite coevolution, genetic diversity.

RÉSUMÉ

Une reconnaissance appropriée des parasites est essentielle pour une réponse immunitaire efficace, assurant l'activation adéquate des mécanismes de défense immunitaire. Chez les vertébrés, il a été fréquemment démontré que les gènes codant pour les récepteurs de l'immunité adaptative impliqués dans la reconnaissance des agents pathogènes sont souvent soumis à une intense pression sélective. En revanche, beaucoup moins d'études se sont intéressées à la sélection agissant sur les récepteurs de l'immunité innée. Le but de cette thèse est de décrire la variabilité naturelle des gènes de l'immunité innée impliqués dans la détection des agents pathogènes chez les rongeurs et d'analyser les mécanismes responsables de leur évolution. Ce travail s'est focalisé principalement sur les rongeurs de la sousfamille des Murinae de part leur présence fréquente à proximité des populations humaines et de leur rôle potentiel en tant que réservoirs d'agents pathogènes dangereux pour l'Homme.

Tout d´abord nous avons étudié la variabilité intraspécifique de cinq *Toll-like* récepteurs ciblant les bactéries (TLR1, TLR2, TLR4, TLR5 et TLR6) pour des lignées consanguines de souris domestiques issues d'une population sauvage de deux sous-espèces : *Mus musculus domesticus* (Mmd) et *Mus musculus musculus* (Mmm). Les souches consanguines constituent un outil adapté à l'étude de la variabilité des gènes immunitaires car elles confèrent une information sur les allèles présents dans les populations naturelles tout en bénéficiant de génotypes homozygotes. Les résultats les plus significatifs concernent la découverte d'un codon stop dans l'exon 2 du *Tlr5* chez une lignée de Mmm et l'absence de variabilité du *Tlr4* chez Mmd. Ces résultats ont également permis de constituer un jeu de SNPs diagnostics utilisable pour de futures études afin de mieux comprendre le rôle de l'introgression de ces gènes immunitaires dans les mécanismes de spéciation de la zone hybride de la souris domestique.

A la suite de ces résultats, nous avons décidé de vérifier si l'absence de polymorphisme du *Tlr4* chez Mmd reflète une absence de variabilité dans les populations naturelles, ou si il s'agit plutôt d'un effet de l'échantillonnage ou des croisements ultérieurs. Nous avons donc séquencé le gène *Tlr4* pour les deux sous-espèces provenant de la région du Paléarctique Occidentale (au total 39 Mmm et 62 Mmd) puis nous avons comparé ces résultats avec la variabilité génétique d'un gène mitochondrial (cytochrome b). Nous avons confirmé notre prédiction : la variabilité de *Tlr4* chez Mmd est fortement réduite par rapport à Mmm, probablement à cause d'agents pathogènes ayant exercé une sélection purifiante chez

Mmd durant la colonisation vers l'ouest. Cependant, l'influence de mécanismes évolutifs neutres, tel que la dérive consécutive à un goulot d'étranglement démographique, ne peut être exclue sur la base de nos données. De plus, nos résultats ont montrés que les deux sous-especes présentent des variants différents de *Tlr4*, et que la recombinaison a joué un rôle important dans le maintien de la variabilité chez Mmm.

La dernière partie de cette thèse a été consacrée à la comparaison interspécifique de deux récepteurs : TLR4 et TLR7. Ces deux TLRs se différencient à la fois par leur localisation et leur capacité de détection. TLR4 est un TLR extracellulaire reconnaissant principalement les ligands bactériens, essentiellement les lipopolysaccharides, tandis que TLR7 est localisé dans la cellule et détecte les virus à ARN simple brin. L'objectif était de décrire la variabilité inter-spécifique de chaque récepteur et de révéler les mécanismes de sélection s'exerçant sur ces gènes au cours de leur évolution sur une échelle de temps plus importante. Nous avons analysé 23 espèces de Murinae provenant d'Europe, d'Asie et d'Afrique. Nos résultats suggèrent que la sélection purifiante est la force principale ayant agit sur l'évolution des gènes TLR4 et TLR7. Cependant, nous avons également mis en évidence des épisodes de sélection diversifiante qui ont pu être à l'origine des variations intra-spécifiques de TLRs observée aujourd'hui chez les rongeurs. Des sites sous sélection positive sont principalement concentrés dans les domaines extracellulaires des deux récepteurs, domaines responsables de la reconnaissance des agents pathogènes. Enfin, la comparaison entre ces deux TLRs montre que le TLR7 localisé dans le compartiment intracellulaire est soumis à une sélection négative plus forte. Cette sélection peut s'expliquer en raison des interactions du TLR7 avec les acides nucléiques viraux qui peuvent être similaires à ceux de l'hôte. Ainsi, un changement même faible dans la conformation du TLR7 pourrait provoquer des réactions auto-immunes chez l'hôte.

Mots clés: *Toll-like receptors*, récepteurs de l'immunité innée, *Pattern Recognition* récepteurs, sélection, évolution, sélection naturelle, polymorphisme génétique, phylogénétique, co-évolution hôte/parasite, diversité génétique.

ABSTRAKT

Včasné rozpoznání patogenů je zásadní pro efektivní imunitní odpověď, zajišťující aktivaci adekvátních obranných mechanismů. U obratlovců je obecně zdokumentováno, že molekuly adaptivní imunity, které se účastní rozpoznání patogenů, jsou často předmětem intenzivních selekčních tlaků. Naopak mnohem méně údajů je známo o selekci působící na receptory vrozené imunity. Cílem této práce je popsat variabilitu genů vrozené imunity, které se účastní detekce patogenů u volně žijících hlodavců a odhalit, které selekční síly na ně během evoluce působily. Práce je zaměřena na podčeleď Murinae, jelikož často žijí v bezprostřední blízkosti lidí a jsou potencionálními nositeli různých pro člověka nebezpečných patogenů. Nejprve jsme studovali vnitrodruhovou variabilitu pěti anti-bakteriálních Toll-like receptorů (TLR1, TLR2, TLR4, TLR5 a TLR6) u inbredních linií myši domácí, které byly odvozeny z populací volně žijících zvířat. Tyto kmeny jsou vhodným nástrojem pro studování variability imunitních genů, jelikož nám poskytují informaci o alelách, které se v přirozené populaci vyskytují, a zároveň je možné vhodně využít homozygotních genotypů u inbredních jedinců. Výsledkem této části bylo popsání polymorfismu pěti genů TLR u dvou poddruhů myši domácí (*Mus musculus musculus*, Mmm a *M. m. domesticus*, Mmd). K nejzajímavějším výsledkům patřilo zjištění stop kodonu v exonu 2 u *Tlr5* a nulová variabilita *Tlr4* u Mmd. Tato část práce zároveň poskytla sadu diagnostických SNPs, které mohou být využity ke studiu introgrese *Tlr* genů přes hybridní zónu myši domácí a ke stanovení vlivu jejich polymorfismu na speciační procesy.

Na tento výsledek jsme navázali další studií, kdy jsme se rozhodli ověřit, zda nízká variabilita *Tlr4* u Mmd je reálným odrazem variability přírodních populací, či se jedná o nepřirozenou odchylku způsobenou nedostatečným vzorkováním či následným inbredním křížením. Osekvenovali jsme tedy *Tlr4* u obou poddruhů z velké části jejich výskytu v západním Palearktu (celkem 39 Mmm a 62 Mmd) a srovnali variabilitu *Tlr4* se znakem na mitochondriální DNA. Hlavním výsledkem bylo potvrzení hypotézy, že variabilita u Mmd je oproti Mmm opravdu radikálně snížena, pravděpodobně vlivem silné selekce způsobené patogeny, se kterými se během kolonizační cesty tento poddruh potkal. Vliv náhodných evolučních mechanismů (např. genetický drift při opakovaných sníženích efektivní velikosti populace) však nemůže být na základě našich dat zcela vyloučen. Dále jsme zjistili, že oba poddruhy mají zcela odlišné *Tlr4* varianty, a že u Mmm je variabilita *Tlr4* udržována i mechanismem rekombinace.

Poslední část disertace je věnována mezidruhovému srovnání dvou receptorů (TLR4 a TLR7). Ty se liší jednak lokalizací v buňce a jednak schopností detekce odlišných ligandů. TLR4 patří mezi extracelulární receptory detekující hlavně bakteriální ligandy (především lipopolysacharidy), zatímco TLR7 je lokalizován na vnitrobuněčných membránách a detekuje ssRNA virů. Cílem mezidruhového srovnání bylo popsat variabilitu obou zmíněných receptorů a působení selekce během evoluce obou TLR. Celkem bylo pro analýzu použito 23 druhů hlodavců z podčeledi Murinae z Evropy, Asie a Afriky. Přestože se obě molekuly vyvíjejí především pod vlivem purifikující selekce, která eliminuje negativní mutace a udržuje funkčnost receptorů, některé jejich části vykazovaly znaky pozitivní selekce. Místa pod pozitivní selekcí byla koncentrována především v extracelulární části obou receptorů, která je zodpovědná za rozpoznání patogenů. Následným srovnáním obou TLR jsme zjistili, že intracelulární TLR7 je pod mnohem větším negativním selekčním tlakem, neboť rozpoznává virové nukleové kyseliny, které jsou značně podobné hostitelským molekulám, a jakákoliv i drobná změna by mohla vést k autoimunitním poruchám.

Klíčová slova: *Toll-like* receptory, receptory vrozené imunity, *Pattern-recognition* receptory, selekce, evoluce, přírodní selekce, polymorfismus, fylogeneze, koevoluce mezi hostitelem a patogenem, genetická variabilita.

The thesis was prepared in the laboratories of:

CENTRE DE BIOLOGIE ET GESTION DES POPULATIONS

(UMR INRA/IRD/CIRAD/MontpellierSupAgro)

Campus International de Baillarguet

CS 30016

34988 Montfferrier sur Lez cedex

France

&

INSTITUTE OF VERTEBRATE BIOLOGY

Research Facility Studenec

Academy of Sciences of the Czech Republic

Studenec 122

675 02 Koněšín

Czech Republic

FUNDING

This thesis was supported by the French National Agency for Research projects CERoPath (grant number 00121 0505, 07 BDIV 012) http://www.ceropath.org/ and BioDivHealthSEA (grant number ANR 11 CPEL 002), and the Czech Science Foundation (grant number 206/08/0640). The thesis was partly funded by a three year French government fellowship and the fellowship from Masaryk University. My thesis was also supported by the project (MUNI/A/0937/2012) of Institute of Botany and Zoology of Masaryk University which provided me the generous financial support. Travelling expanses were partially funded by project of Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Next-generation technologies in evolutionary genetics (CZ.1.07/2.3./20.0303) and by bilateral project BARRANDE (grant number MEB021130/24504WM).

ACKNOWLEDGEMENTS

Firstly, I would like to thank sincerely to my official thesis supervisors Josef Bryja and Jean-François Cosson, for their excellent mentoring, advices, and understanding over the past years. I appreciate the time they provide me and also their enormous patience. I would like to thank also to my "shadow" supervisors Jaroslav Piálek and Miloš Macholán for all long hours of discussion and big personal support during my whole stay in Studenec. I really admire their infinite enthusiasm (often infectious) and "active service" during weekends, holidays and late evenings in final phase of my thesis writing. Many thanks go also to Nathalie Charbonnel for all instrumental meetings and working coffee breaks. All these people gave me great advices when needed and I can easily say that once I would like to become such brilliant scientists as they are. I would like to thank also to other members of my thesis committee, Dr Pierrick Labbé and Dr Nicolas Bierne for their constructive advices and stimulant suggestions during the regular committee meetings. My special thanks belong to the French government, whose scholarship for Ph.D. students "en co-tutelle" has allowed my Ph.D. study in Montpellier and to Dr Jaroslav Piálek and his project (project 206-08-0640) funded by the Grant Agency of the Academy of Sciences of the Czech Republic from which I was partially paid. I thank also to Institute of Botany and Zoology of Masaryk University for the generous financial support from project MUNI/A/0937/2012. Because this thesis is the result of a great team work I tried to avoid using "I" in the text and replaced it by "we".

Thanks to all my colleagues and dear friends from Studenec, especially to lab and mouse team Aňa Bryjová, Michal Vinkler, Lucka Vlčková, Zuzka Bainová, Iva Martincová, Dáša Čížková, Tania Aghová, Terka Králová, Ludovít Ďureje, Hanka Konvičková, Jana Piálková, Wasimuddin, Start J. E. Baird and Joëlle Goüy de Bellocq, for making the research real, successful and often also more fun. Same thanks goes also to all my French colleagues and friends from CBGP, especially to girls and my very good friends Pascaline Dumas, Joséphine Piffaretti, Marie Pagès, Anne Laure Clamens, Caroline Tatard and Laure Sauné for their personal support and lot of funny moments. Than I would like to thank also to masculine part of CBGP, to phylogeny master Gael Kergoat, IT God Alex Dehne Garcia, fitness coach Laurent Soldati, IT genius Sylvain Piry, rat hunter Yannick Chaval and "brother in beer" Fabian Codamine. Special thanks go to my friend and „snowboard mate" Max Galan for his moral and technical support and precious friendship.

And last but not least I would like to thank to my family for all their support over the three years. I would like to especially thank to my mother, for her support and optimism. My

big thanks go also to my “big” brother who took care about whole family during my absence. My biggest thanks go to my boyfriend and great scientist Olda, who encouraged me every day and helped me to solve many problems and crazy hypothesis and again and again we spend our evenings in discussions about *Tlr*s and their evolution. He was my principal moral support when I lost my hope or motivation. He taught me how to be patient and tolerant and thanks to him I become better person and also better scientist I guess. I thank also to Mr Skype that gave us the possibility to be together and bridge the distance of 1600km and to my computer for smashing service during whole Ph.D and especially during last weeks.

Finally I would like to dedicate this work to my father who unfortunately died on 5th December 2009. Thanks to him I am where I am and I am who I am. He was the person, who pushed me to study foreign languages especially French and English and who introduced me France as beautiful and magic country full of colours, fragrance and divine tastes. He opened me the world and awaked my passion for travelling and studing. Thanks to him I could become independent and strong person. He gave me wisdom advices but let me also do my own mistakes, because he knew that experiences are untransmittable. He supported me every time even when I choose science as my main life path and he used to say to me “fly low”!

PREFACE

Organisation of the thesis in "co-tutelle"

The thesis was from the beginning designed as the thesis under double co-supervision performed officially at the University of Montpellier 2 (Montpellier, France) and Masaryk University (Brno, Czech Republic). I have started my thesis at the Masaryk University in September 2009. I spent the first year by writing the project and application for the Grant of French government and by preparing my first year in France. For next three years (from 2010 to 2013) I have spent seven months (from 1st March till the end of September) at the Research Facility Studenec of Institute of Vertebrate Biology of the Academy of Sciences of the Czech Republic ("IVB") and for the rest of the year (or better during winter) I have "migrated" southward to the Centre de Biology pour la Gestion des Populations, Montferriez-sur-Lez, France (CBGP). If you check the photos bellow you probably figure out my intention....

Photos: on the left, mouse breeding facility at IVB in Studenec in winter 2010, on the right, CBGP in winter 2010.

In CBGP, I had the opportunity to use the immense sample collections of rodents from Southeast Asia, which are also important vectors of emerging diseases (data collected mostly within the project CERoPath). At IVB, I was directly involved in projects using one of the best model systems for speciation studies, i.e. the European house mouse hybrid zone, which has been a subject of a long-term study at the institute. The material collected in the field during previous projects was complemented by a huge sample collection of both mouse subspecies from a large part of the Western Palaearctic area as well as by cca 20 inbred strains derived from wild populations. We have decided to focus on interspecific level of *Tlr* variation during my stays in CBGP and on intraspecific level during my work in Studenec.

LIST OF TABLES AND FIGURES

Chapter 1 and 2: General Introduction and Material and methods

Fig. 1, Overview of localization of signalling PRRs and their MAMPs
Fig. 2, VDJ recombination
Fig. 3, Function TCR and BCR
Fig. 4, Frequency-dependent selection
Fig. 5, Balancing selection through negative frequency dependent selection.
Fig. 6, Biochemical properties of amino acids
Fig. 7, Simplified description of the TLR4 structure
Fig. 8, Crystal structure of TLR ligand-binding domains with its ligands
Fig. 9, Schematic of mammalian TLR signaling pathways
Fig. 10, Evolution of major vertebrate
Fig. 11, Overview of composition
Fig. 12, Overview of localization and ligands of TLRs
Fig. 13, Phylogeny of the subfamily Murinae
Fig. 14, Phylogeny of rats
Fig. 15, Phylogenetic relationships of the Indochinese Rattini
Fig. 16, Maximum likelihood tree reconstructed with a molecular clock hypothesis
Fig. 17, Phylogenetic relationships among 10 mouse taxa
Fig. 18, Colonization routes of the house mouse subspecies starting from northern parts of Indian subcontinent
Fig. 19, The course of the *M. m. musculus/M. m. domesticus* hybrid zone in Europe
Fig. 20, Taxonomic structure of animals used for experiments by humans. by Daniel Engber

Chapter 3.1: Polymorphism of bacterial-sensing TLRs in wild-derived and classical laboratory strains of Mus musculus

Fig. 1, Origin localities of wild-derived strains kept in Research Facility Studenec
Fig. 2, Description of non-synonymous polymorphic sites (SNPs) and their position within TLRs
Fig. 3, Haplotype network based on nucleic acids

Table 1, Summary of strains
Table 2, Overview of sequenced TLRs
Table 3, Overview of sites and regions important for ligand binding and dimerization
Table 4, Genetic diversity of *Tlr*s in two subspecies of the house mouse
Table 5, Physicochemical properties of the amino acids involved in non-synonymous substitutions in ligand binding regions

Chapter 3.2: Analysis of variability at intraspecific level in wild populations of Mus musculus

Fig. 1. Distribution of samples analyzed in this study
Fig. 2, Overview of *Tlr4* non-synonymous substitutions in Mmd and Mmm
Fig. 3, Ribbon diagram of the TLR4 ECD 3D structure
Fig. 4, Haplotype network and haplogroup distribution of *Tlr4* (a) and *mt-Cytb* (b)

Table 1, Genetic diversity of *Tlr4* and *mt-Cytb* in two house mouse subspecies
Table 2, Description of LBR variants
Table 3, Physicochemical properties of the amino acids involved in non-synonymous substitutions of *Tlr4*

Table 4, Selection tested by REL in both subspecies together

Supplementary materials

Fig. S1, Phylogeny based on Bayesian inference (MrBayes v.3.1) *Tlr4* (a) and *mt-Cytb* (b)
Fig. S2, Haplogroup definition *Tlr4* (a) and *mt-Cytb* (b)
Fig. S3, Evidence of recombination between HG-Im and HG-IIm of Mmm

Table S1, Summary of sampled specimens, identification of haplotypes and NCBI GenBank accession numbers.
Table S2, Binding sites between TLR4/LPS/MD-2

Chapter 3.3: Analysis of variability at interspecific level of wild rodents

Fig. 1, Comparison of phylogenetic trees based on *Tlr*s and neutral markers
Fig. 2, Distribution of sites under selection identified by SLAC and MEME
Fig. 3, Sites under positive selection identified in evolutionary lineages by MEME
Fig. 4, Mapping of evolutionary conservation of amino acid positions in a protein molecule based on the phylogenetic relations between homologous sequences

Table 1, Estimates of sequence diversity and average codon-based evolutionary divergence over all sequence pairs for the exon 3 and particular domains of *Tlr4* and *Tlr7* genes
Table 2, Positively (MEME and SLAC-PS) and negatively (SLAC-NS) selected sites detected for the exon 3 of *Tlr4* and *Tlr7* at $p < 0.05$

Supplementary materials

Fig. S1, Protein structure of TLR4 (a, c) and TLR7 (b, d) identified by SMART
Fig. S2, Phylogeny based on the exon 3 of *Tlr4* (a) and *Tlr7* (b) gene reconstructed by Bayesian inference method in MrBayes
Fig. S3, Phylogeny based on the exon 3 of *Tlr4* (a) and *Tlr7* (b) gene reconstructed by maximum likelihood method in RAxML
Fig. S4, Test of congruence between the presumably neutral and *Tlr* phylogenies (*Tlr*4 (a), *Tlr*7 (b))
Fig. S5, Superimposition of structures, tree clustering diagrams based on linkage distance, LBR_{TLR4} (a) and LBR_{TLR7} (b)
Fig. S6, Analysis of LBR amino acid sequence charge at pH 7 (LRRFinder) for LBR_{TLR4} (a) and LBR_{TLR7} (b)

Table S1, Summary of sampled specimens and identification of haplotypes
Table S2, Primer description
Table S3, Residues binding to LPS in TLR4 based on knowledge of 3D-crystalography in human predicted by Park et al. 2009
Table S4, Potential residues binding ssRNA predicted by Wei et al. 2009

Discussion

Fig. 21, Hypothesis about the origin of reduced *Tlr4* variability in Mmd.
Fig. 22, Hierarchical model outlining the evolutionary dynamics and biological relevance of the various families of PRRs

LIST OF ABREVIATIONS

APC, anitgen presenting cells;
BCRs, B cell receptors;
CARD, caspase activating and recruitment domain;
CLRs, C-type lectin receptors;
CLS, classical laboratory strains;
CpG DNA, unmethylated cytosine-guanine dinucleotide sequences
CpG, cytosine phosphate guanosine;
CRDs, carbohydrate recognition domains;
CTD, C-terminal domain;
CTLD, C-type lectin-like domain;
DCs, dendritic cells;
DC-SIGN, dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin;
dN, number of non-synonymous substitutions per non-synonymous site;
dS, number of synonymous substitutions per synonymous site;
dsRNA, double-stranded RNA;
ECD, N-terminal horseshoe-like extracellular domain, or extracellular domain;
FDS, Frequency-dependent selection;
ICD, C-terminal intracellular domain;
IFN, interferon;
IL, interleukin;
IL-1β, Interleukin-1β;
IL-6, Interleukin-6;
ILS, incomplete lineage sorting;
IRAK, interleukin-1-receptor-associated kinase;
IRF, interferon-regulated factor;
LBR, ligand binding region;
LGP2, laboratory of genetics and physiology 2;
LP, lipoproteins;
LPS, lipopolysaccharide;
LRRs, leucine rich repeats;
LS, laboratory strains;
MAL, MyD88 adaptor-like protein;
MAMPs, microbe associated molecular patterns;
MBL, mannose-binding lectin;
MDA5, melanoma differentiation associated factor 5;
MD-2, lymphocyte antigen 96;
MGL, Macrophage galactose-type C-type lectin;
MHC, major histocompatibility complex;
Mmd, *M. m. domesticus*;
Mmm, *M. m. musculus*;
MR, mannose receptor;
MyD88, Myeloid differentiation factor 88;
Myr, Million years ago;
NACHT, domain present in NAIP, CIITA, HET-E, TP-1;
NALP, NACHT-, LRR-, and pyrin-domain containing proteins;
NAP1, NAK-associated protein 1;
NFDS, negative frequency dependent selection;

NF-κB, Nuclear factor κB;
NLRs, NOD-like receptors;
NOD, nucleotide-binding oligomerization domain;
NPCs, neural progenitor cells;
nsSNPs, non-synonymous single nucleotide polymorphisms;
ORs, opsonin receptors;
PAMPs, pathogen-associated molecular patterns;
PG, peptidoglycan;
polyI:C, polyinosine-deoxycytidylic acid;
PRRs, pattern recognition receptors;
RIG-I, retionic acid-inducible gene I;
RLRs, RIG-I like receptors;
SNPs, single nucleotide polymorphisms;
SR , scavenger receptors;
ssRNA, single stranded viral RNA;
TCRs, T cell receptors;
TIR, Toll/interleukin-1 receptor;
TIRAP, TIR-associated protein;
TLR, Toll-like receptor;
TMD, single transmembrane helix or transmembrane domain;
TNFα, tumour necrosis factor α;
TRAF, TNF-receptor-associated factor;
TRAF6, tumor necrosis factor receptor associated factor 6;
TRAM, TRIF-related adaptor protein;
TRIF, TIR domain containing adapter-inducing interferon β
WDS, wild derived strains
WT, wild type

TLR – used for proteins

Tlr – used for genes

TABLE OF CONTENTS

1 GENERAL INTRODUCTION 20

1.1 Immune system and recognition of antigens 20

1.1.1 Recognition receptors of innate immunity 21

1.1.2 Antigen recognition in adaptive immunity 25

1.2 Evolutionary processes affecting evolution of immune receptors 27

1.2.1 Selection imposed by pathogens: adaptive evolution 29

1.2.2 Stochastic evolutionary processes 32

1.2.3 Polymorphism and the effect of non-synonymous substitutions on protein functions 33

1.2.4 Effect of non-synonymous substitutions on the function of pattern recognition receptors 35

1.3 Toll-like receptors – a general overview 36

1.3.1 A brief historical survey: from Toll to Toll-like receptors 36

1.3.2 Structure of TLRs 36

1.3.3 Signalling of TLRs 39

1.3.4 Origin and function of the TLR family 40

1.3.5 Variability and polymorphism of TLRs 44

1.3.6 Evolutionary forces acting on TLRs 45

1.4 Thesis aims 47

2 MATERIAL AND METHODS 48

2.1 Rodents and rodent-born infectious emergent diseases 48

2.1.1 Origin and radiation of the tribe Rattini in the Southeast Asia 49

2.1.2 Tribe Murini and evolution of house mice (Mus musculus) 53

2.1.3 Rodent-borne diseases, emergence risk for humans and rodents as model species 57

2.2 Analysis of natural selection 60

2.2.1 Analysis of selection at the intraspecific level 60

2.2.2 Analysis of selection at the intraspecific or population level 61

3 RESULTS 62

3.1 Polymorphism of bacterial-sensing TLRs in wild-derived and classical laboratory strains of *Mus musculus* 63

3.1.1 Introduction 64

3.1.2 Material and Methods 64

3.1.3 Laboratory techniques 65
3.1.4 Data analysis 66
3.1.5 Results 68
3.1.6 Discussion 71
3.2 Analysis of variability at intraspecific level in wild populations of *Mus musculus* 73
3.2.1 Introduction 75
3.2.2 Materials and methods 79
3.2.3 Results 82
3.2.4 Discussion 89
3.3 Analysis of variability at interspecific level of wild rodents 104
3.3.1 Introduction 106
3.3.2 Materials and methods 109
3.3.3 Results 115
3.3.4 Discussion 122
4 GENERAL DISCUSSION 142
4.1 Selection forces acting on TLRs in free living populations: intra- vs. inter-specific level 144
4.2 The role of recombination: instrument of stochastic processes or selection 147
4.3 Selection forces acting on TLRs: bacterial sensing vs. viral sensing 148
4.4 Selection forces acting on TLRs: ECD vs. ICD 151
4.5 TLRS IN SPECIATION RESEARCH - FUTURE PROSPECTS 152
4.6 CONCLUSION 153
5 REFFERENCES 154
6 ANNEX 179
6.1 PRIMERS 179
6.2 PCR PROTOCOLS 180
6.3 CURRICULUM VITAE 181
6.4 ACCEPTED ARTICLE 185

1 GENERAL INTRODUCTION

1.1 Immune system and recognition of antigens

Ability of immune system to distinguish between self and non-self molecules is fundamental for fitness, i.e. for the survival of organisms and for their reproductive success. Evolution created a wide spectrum of defence mechanisms allowing organisms to deal with non-invited visitors (e.g. Danilova 2006). In the first line of self/non-self discrimination there are many receptors able to detect foreign molecules expressed by invading microbes considered as non-self (Danilova 2006). In this work, pathogens are referred to all infectious agents (e.g. diverse pathogenic viruses, unicellular or multicellular organisms), against which a host has to intervene.

In jawed vertebrates we can traditionally categorize recognition molecules into two major subdivisions corresponding to main branches of the immune system, i.e. innate and adaptive immunity (Medzhitov 2007). Yet, according to recent publications we have to take this division with caution because there are many cells (e.g. γδ T cells, CD8αα T cells, B1 B cells, MZ B cells and natural killer T cells) whose classifications into a respective branch is not definite, because they evince patterns of both innate and adaptive immunity. Therefore they should be assumed rather as a bridge between both immunity branches and the immune system should be then considered as extremes of a continuum (Getz 2005; Borghesi and Milcarek 2007; Sun and Lanier 2009; Criscitiello and de Figueiredo 2013). However in this work I will stick to the traditional division into two branches mentioned above.

The function of innate and adaptive immune responses is closely related to creating a sophisticated network, although the two branches are fundamentally different. The first major difference is in the reaction time. Innate immunity receptors trigger an immediate answer to a pathogen invasion, while adaptive immunity needs more time to be manifested. The second main difference is their recognition specificity. Receptors of innate immunity provide non-specific responses, i.e. they detect similar molecules shared by groups of related pathogens that are essential for their survival and are different from molecules present in host cells. Recognition mechanisms of adaptive immunity are specific, what means that they are able to distinguish between individual forms or groups of microbes according to their antigens. Antigens are usually large molecules (proteins or polysaccharides) that can be found on the surface of microbes or they can be released through secretion into the extracellular fluid (e.g.

toxins) (Medzhitov and Janeway 1999). Finally, the third contrast is in their reaction for repeatedly encountered invaders (Schenten and Medzhitov 2011). The receptors of adaptive immune system exhibit an immunological memory (anamnestic response). They "remember" that they already met invading pathogens and during a subsequent exposure they improve their response and react more quickly and appropriately to the same pathogen. The innate immune system exhibits no memory response at all and repeated exposure to the same antigen does not lead to a qualitative or quantitative improvement of the following response. A receptor function can be temporarily upregulated as a result of exposure to pathogens, but the components of the innate immune system do not change permanently during individual's lifetime. In the next part I shall present a short overview of main actors involving in antigen recognition of innate and adaptive immunity.

1.1.1 Recognition receptors of innate immunity

Evolutionary older are receptors of innate immunity which probably arise 700 million years ago (mya) as a result of the first interactions with pathogens and before the separation of protostomes and deuterostomes (Kimbrell and Beutler 2001; Danilova 2006; Bosch 2013). In some forms they are present in all organisms from plants to vertebrates and provide an early and immediate host response (Akira et al. 2006). Altogether these germline-encoded receptors are called pattern recognition receptors (PRRs). PRRs behave as economical inspection, which detect essential and in general conserved microbial components called microbe associated molecular patterns (MAMPs) or similarly used equivalent pathogen associated molecular patterns (PAMPs) (Janeway 1989; Medzhitov and Janeway 2002). PAMPs are different from host structures (to fulfil the assumption about self and non-self discrimination), but in contrast they are often common for diverse microbes and therefore the number of PRRs can be relatively low (Villaseñor-Cardoso and Ortega 2011). There are about 10^2 PRRs known today which are able to recognize around 10^3 of PAMPs, however the number is probably not definitive. PAMPs include, for example, lipopolysaccharides (LPSs) from cell wall of gram-negative bacteria, lipoproteins (LPs), phosphorylcholines, peptidoglycan (PG), lipotechoic acids from gram-positive cell wall, mannose (a terminal sugar common in microbial glycolipids and glycoproteins), bacterial and viral nucleic acids such as unmethylated cytosine-guanine dinucleotide sequences (CpG DNA), bacterial flagellin and pilin, the amino acid N-formylomethionin found in bacterial proteins, double-stranded

(dsRNA) and single stranded viral RNAs (ssRNA), glycolipids and zymosan from fungal cell walls. Injured, infected or transformed host cells are often also considered as PAMPs by PRRs (Gordon 2004). It is also not surprising that each microbe is composed of several different PAMPs and therefore it is detected by multiple PRRs. As a result, several PRRs detecting the same PAMPs can overlap in their function. PRRs are primarily present on the surface of dendritic, endothelial, and mucosal cells, lymphocytes and macrophages. Other types of PRRs react within phagolysosomes of phagocytes or in the cytosol. Therefore we can divide PRRs according to their function and localization into two groups: endocytic PRRs, mediating non-opsonic absorption of microbes, and signalling PRRs (Areschoug and Gordon 2008). Some of PRRs can be classed into both groups (e.g several delegates of C-type lectin receptors; CLRs).

Endocytic pattern recognition receptors

Endocytic pattern-recognition receptors are found on the surface of phagocytes and are responsible for the attachment of phagocytes to microbes and their subsequent destruction.

The first example of these receptors is the **mannose receptor** (MR). MRs belongs to a subfamily of CLRs and bind mannose rich glycans and fucose groups on microbial glycoproteins and glycolipids. MRs interact with gram positive and negative bacteria, fungal pathogens and envelope protein gp120 of human immunodeficiency virus (HIV) (Fraser et al. 1998; Lai et al. 2009). MRs participate also in the complement pathway (Medzhitov and Janeway 2002).

Other receptors present on the surface of phagocytes are **scavenger receptors** (SR). They bind components of bacterial cell walls such as LPSs, peptidoglycan or teichoic acids and dsRNA (Gough and Gordon 2000; Peiser et al. 2002). SRs are transmembrane receptors, which are composed of various domains (e.g. collagenous, cysteine-rich, C-type-lectin or other domains) (Peiser et al. 2002). They mediate non-opsonic phagocytosis and cooperate with Toll-like receptors (TLR), modulating the inflammatory response to TLR agonists (Areschoug and Gordon 2009).

The last type of main endocytic PRRs are **opsonin receptors** (ORs). These receptors bind microbes to phagocytes and include mainly mannose-binding lecitin, C-reactive protein (CRP) and complement pathway proteins, such as C3b and C4b etc.

Signalling pattern recognition receptors

Signalling pattern recognition receptors promotes secretion of intracellular regulatory molecules such as inflammatory cytokines and chemokines. These cytokines trigger innate inflammation, fever and phagocytosis. These signalling receptors can be divided into distinct families according to shared structural domains, ligand specificity, cellular distribution and downstream signalling pathways. Their common pattern is activation of adaptive immune response and they are often considered as a bridge between innate and adaptive immunity. Crosstalk between different families and redundancy of their function was described by several studies (Opitz et al. 2009; Loo and Gale Jr. 2011; Vasseur et al. 2011; Vasseur et al. 2012).

The first type of these receptors are **NOD-like receptors** (nucleotide-binding oligomerization domain) (NLRs). This large family contains more than 20 different NLRs (Carneiro et al. 2008; Vasseur et al. 2012) which share the cytostolic location and structural composition of three domains: ligand-sensing leucine rich repeats (LRRs), the NACHT domain (the NACHT domain has been named after NAIP, CIITA, HET-E and TP1), responsible for oligomerization and an effector domain (for example caspase recruitment domain, CARD) (Vasseur et al. 2012). NLRs are involved in intracellular recognition of peptidoglycans components, dsDNA and LPS, however, the full range of NOD PAMPs is still unknown. Among the best studied NLRs are NOD1 and NOD2 cytostolic proteins which bind muramyl dipeptides from bacterial cell walls and NALP (NACHT-, LRR-, and pyrin-domain containing proteins) subfamily sensing bacterial RNA and PG (Balamayooran et al. 2010).

Intracellular recognition of viral proteins, especially RNA helicases spread by viruses, is provided by **RIG-I like receptors** (RLRs). Up to date there are three known RLRs members: RIG-I (retionic acid-inducible gene I), MDA5 (melanoma differentiation associated factor 5) and LGP2 (laboratory of genetics and physiology 2). These receptors block viral replication via induction of interferons type I (IFN-α and IFN-β). Three principal domains were described as follows: CARDs involved in signalling, central DExD/H box RNA helicases domain with the capacity to hydrolyze ATP and interact with viral RNA and C-terminal domain (CTD) which is involved in autoregulation (Vasseur et al. 2011).

Extracellular receptors involved in antifungal immunity are **C-type lectin receptors** (CLRs) which have been already mentioned in the previous PRRs groups. CLRs are a large subfamily composed of 17 groups based on their phylogeny and domain organization

(Vazquez-Mendoza et al. 2013). Characteristic structures of CLRs are carbohydrate recognition domains (CRDs) or C-type lectin-like domain (CTLD). Macrophage galactose-type C-type lectin (MGL), dendritic cell-specific intercellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN), the mannose receptor (MR) including mannose-binding lectin (MBL) and Dectin-1(which can be classed as well in previous group, because of it ability to bind and phagocyte yeast and fungal-derived zymosan particles) are the most famous receptors of this group (Medzhitov 2007; Vazquez-Mendoza et al. 2013).

The most important and explored PRRs are **Toll-like receptors** (TLRs), a class of membrane proteins that play a fundamental role in pathogen recognition and activation of innate and adaptive immunity. Their fame came with finding that they are able to detect wide range of bacterial, fungal and viral PAMPs and they are in the front line of host defence against microbes (Akira et al. 2006; Medzhitov 2007; Beutler 2009). Common pattern of TLRs is their structural organization and downstream signalling through TIR domain. In humans, 10 different TLRs have been described. According to their localization and ligand detection we can divide them into two main groups: cellular bacterial-sensing on the cell surface and endosomal viral-sensing. Detailed description of TLRs and their signalling can be found in Chapter 1.3.

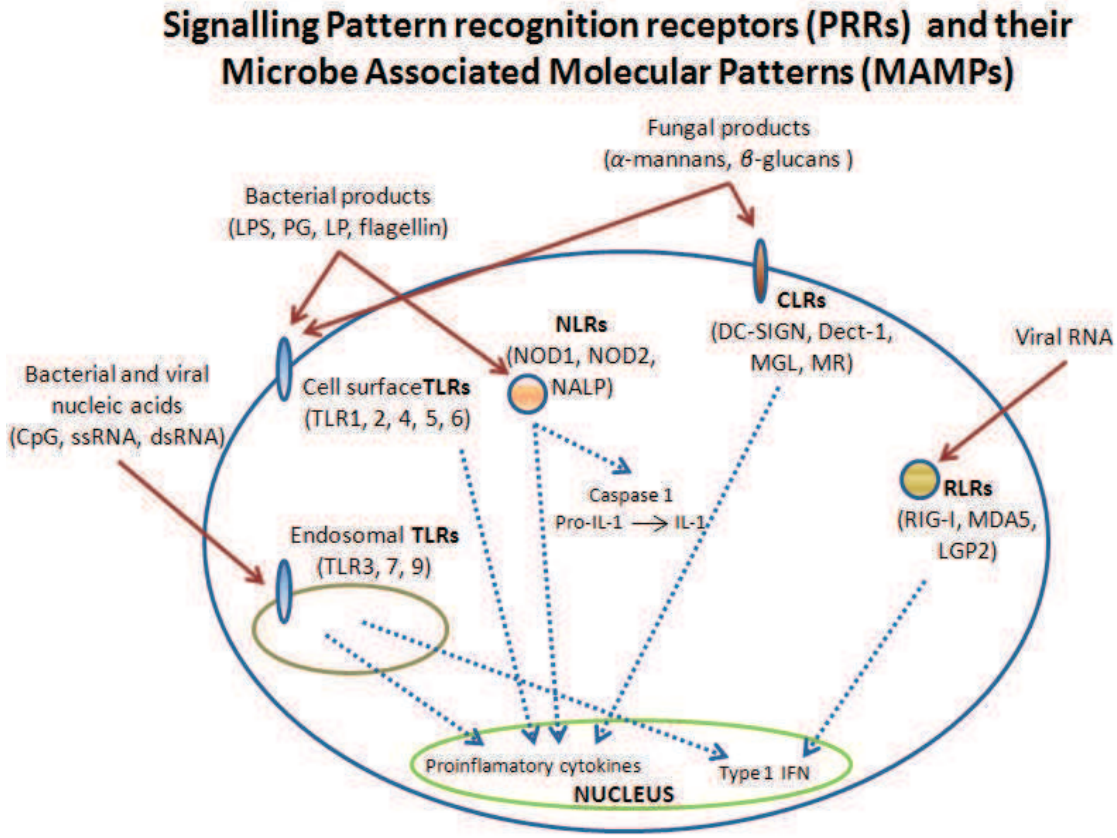


Fig. 1, Overview of localization of signalling PRRs and their MAMPs

1.1.2 Antigen recognition in adaptive immunity

The adaptive immune system evolved 500 mya in jawed vertebrates (Danilova 2006; Leulier and Lemaitre 2008; Flajnik and Kasahara 2010). The origin of adaptive immunity is still discussed, but the most probable seems the theory about transposons invasion (Travis 2009). There are two key players of antigen recognition, B lymphocytes (or B cells) and T lymphocytes (or T cells), which are highly specialized and adaptable. Each T and B lymphocyte contains protein molecules developed by somatic hypermutation, gene conversion or clonal gene rearrangements (assembled from gene segments: V-variable, D-diverse and J-joining) known as somatic or V(D)J recombination (Borghesi and Milcarek 2007; Medzhitov 2007). This mechanism enables a small number of genes to produce a huge number of different antigen receptors, which are then uniquely expressed on each individual lymphocyte. More than 10^8 combinations of receptors are generated, although some are removed due to self reactivity, while the others are improved in the host over time (Mogensen 2009). Structurally unique receptors allow pathogen-specific recognition of a vast number of different antigens. The area of the antigen that binds to the antigen receptor is known as the epitope. An antigen has usually multiple epitopes, which are specific for distinct receptors, to which they will bind, exclusively. Bellow I describe two major groups of antigen-specific receptors.

B cell receptors (BCRs) are membrane-bound immunoglobulin involved in humoral immunity against extracellular pathogens and toxins. They are presented on the surface of B cells and recognize mostly bacterial components outside the cell (Schwensow et al. 2011). BCRs attack undigested antigens and break them into small peptides, which bind with major histocompatibility complex (MHC) molecules, and then with T cell receptors through the process known as immunological synapse.

T cell receptors (TCRs) are important in cell-mediated immunity. They can be found on the surface of T cells and are responsible for recognizing antigens bound to major histocompatibility complex (MHC) class I or class II molecules inside of infected cells. Appropriate TCRs reactivity to self MHC molecules is strictly controlled in the thymus and only correctly responding TCRs are released into the periphery (Jameson and Bevan 1998). Binding of an antigen activates signal transduction pathways which leads to cell proliferation, differentiation and secretion of cytokines and growth factors (Choudhuri et al. 2005). MHC molecules have to fight against a high number of pathogens in the environment; therefore they

must be able to present a wide range of peptides and are considered as the most polymorphic genes in vertebrate genomes. Mechanisms as recombination, codominant expression and polygenic locus are used to achieve such variability.

The activation of TCRs and BCRs is initiated and modulated by the signals from innate immune receptors. Both arms of the vertebrate immune system therefore create the complex and interconnected system of defence against microbe invasion.

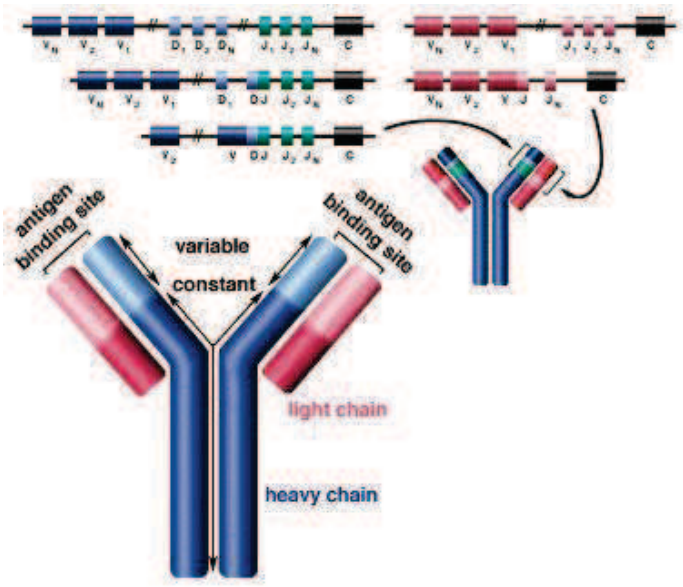


Fig. 2, VDJ recombination, an antibody is composed of two identical light and two identical heavy chains, and the genes specifying them are found in the 'V' (Variable) region and the 'C' (Constant) region. In the heavy-chain 'V' region there are three segments; V, D and J, which recombine randomly, in a process called VDJ recombination, to produce a unique variable domain of each individual B cell. Similar rearrangements occur for light-chain 'V' region except there are only two segments involved; V and J.
Adopted from http://www.shaltech.com/about-lymphoma/

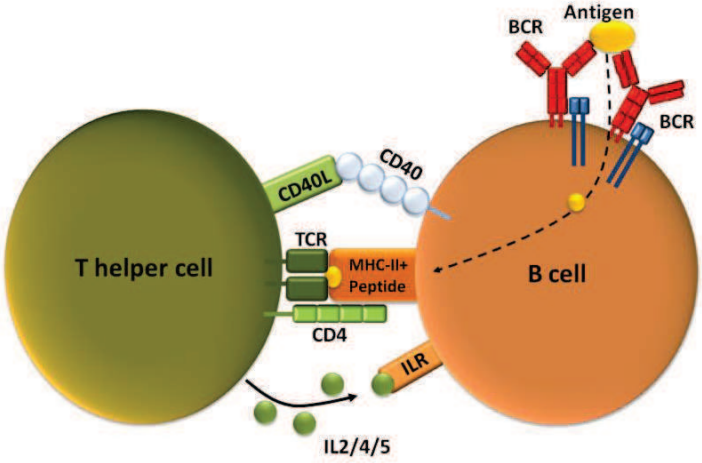


Fig. 3, Function TCR and BCR, the effector T Helper cells activate specific B cells through a phenomenon known as an immunological synapse (BCR=B cell receptor, TCR =T cell receptor, IL = interleukin, and ILR = interleukin receptor). Activated B cells differentiate into plasma cells that subsequently produce antibodies which assist in clearing the host of the pathogen. CD4, CD40L are glycoproteins expressed on the surface of T helper cells, while CD40 is expressed on the surface of B cells and other anitgen presenting cells (APC).
Adopted from http://www.shaltech.com/about-lymphoma/

1.2 Evolutionary processes affecting evolution of immune receptors

Genetic diversity is important for survival and adaptability of species in changing environments. To decide which evolutionary forces shaped variation in specific genes is a difficult task because populations and species have to adapt to the abiotic environment (e.g. temperature, sunlight, pollution), to other species with which they interact (e.g. prey, predators, competitors, parasites) or more often to the combination of both. Moreover we should not underestimate neutral or non-adaptive evolutionary forces such as mutation, recombination and random genetic drift (Andrews 2010; Honnay 2013). Understanding processes which drive evolution of immune receptors is a fascinating research subject and big challenge.

According to the Red Queen hypothesis (the famous "running as fast as you can to stay in the same place"), firstly proposed by Leigh Van Valen (1973), organisms are running the arms race with other biological "partners" such as predators, source competitors at the intra- or interspecific levels or parasites. According to generally accepted theories the adaptation imposed by pathogens belongs among the most dynamic – continuous adaptive changes (Lederberg 1999; Zimmer 2001).

Parasites co-evolving with their hosts are viewed there as a key factor modulating different life traits of their hosts (e.g. population genetic structure, demographical changes, mating system, sexual dimorphism etc.) (Sheldon 1998). Generally microbes have an advantage during the arms race due to their shorter generation time and high mutation rates, which enhance genetic novelty and evolutionary potential, and so give better opportunity to update their invasion strategies tricking immune system of their hosts (Meyer 1991). Some bacteria generate surface proteins that bind to antibodies, rendering them ineffective; examples include *Streptococcus pneumoniae* (protein G), *Staphylococcus aureus* (protein A), and *Peptostreptococcus magnus* (protein L). Others are able to knock out or kill phagocytes. Some microbes can eventually mimic host cells and then block the interferon (IFN) production pathway or construct protective capsules (*Mycobacterium*) (Fortune et al. 2004). Another example is the ability of *Mycobacterium leprae* to suppress cell-mediated immunity or to play with immune system "hide-and-seek" inside host cells to avoid their detection (Maizels et al. 2004). Extracellular pathogens often alter their antigens (surface proteins), so we can imagine that it is like a thief who is escaping police by changing the coat and wig all the time and police is still one step behind (e.g. 84 known serotypes of *Streptococcus pneumonia*, glycoprotein of trypanosomes, surface flagellin of *Salmonella typhimurium,* pilin

protein of *Neisseria gonorrhoeae*, or LPS from *H. pylori* and *P. gingivalis*) (Andersen-Nissen et al. 2005). The same strategy, i.e. antigenic variation and antigenic diversity, is used also by protozoan parasites (e.g. *Plasmodium falciparum* causing malaria) (Reeder and Brown 1996).

Experts and prominent specialists in evasion strategies are viruses. Their short generation times and relatively high mutation rates give them huge advantage in red queen running with host defence mechanism and allows them quick adaptation to changing environment of their host. The best known example is influenza virus with constant replacement of its surface envelope proteins. Cytomegalovirus, another cunning invader, is able to evade the host defence system by expressing its MHC class I homologues thus pretending to be part of its body (Reyburn et al. 1997). Viruses can become also invisible for the immune system when they enter the latency state during which they are inactive (i.e. without replication). In this state, avirus particle does not cause a disease, but also does not produce any viral peptides which normally attract attention of different immune receptors. Viruses in the latent form can not be eliminated by the immune system and hence are sources of potential recurrent illnesses. Most famous for such a strategy are Herpes simplex viruses (Bowie and Unterholzner 2008). Another viral strategy is inhibition of immune response. Paramyxovirus for example can inhibit the type I IFN response which is included in the RIG-I signalling cascade. The most insolent viruses even use PRRs as entry ports (Yamada et al. 2005).

At first glance the only fair players seem to be parasitic worms, which evolve slower and therefore give a putative chance to the immune system to adapt. However, they are even bigger (literally) swindlers than their fast evolving cousins. During many millions of years of close coexistence with their hosts (parasites are often species specific) they have had time to evolve a sophisticated weapon arsenal designed to evade and modulate the host immune system (Wakelin 1996; Zaccone et al. 2006). Due to their generally bigger bodies it is very difficult for the immune system to eliminate them. There is a known pack of chameleons (e.g. coating with host proteins by schistosomes or filarial nematodes), the squad of chemical terrorists causing immunosuppression (e.g. hookworms producing a protein which binds the ß integrin CR3 and inhibits neutrophil extravasation or immunosuppression made by *Burgia* spp. or *Nocardia brasiliensis*) and also a group of nomads which avoids local inflammatory reactions by migrations through the host body (hookworms, *Brugia* spp., *Wuchereria bancrofti*, or microfilariae of *Onchocerca volvulus*) (Pearce and Sher 1987; Wakelin 1996; MacDonald et al. 2002). On the other hand long-term coevolution sometimes leads to a relatively stable relationship such as commensalism or mutualism. Moreover, according to the

hygiene hypothesis absolute elimination and subsequent absence of parasites can lead to autoimmune disorders (Zaccone et al. 2006).

As a consequence of co-evolution with microorganisms the immunity genes belong to the fastest evolving genes (Nielsen et al. 2005; Barreiro et al. 2009; Barreiro and Quintana-Murci 2010; Quach et al. 2013; Quintana-Murci and Clark 2013). Since immune receptors are chief gatekeepers protecting against entrance of microbes, we suggest that antagonistic host-parasite interactions are the principal force shaping their evolution.

1.2.1 Selection imposed by pathogens: adaptive evolution

The model which explains the evolution of immune receptors in the light of host-parasite co-evolution is called the matching alleles model (Frank 1993; Agrawal and Lively 2003). This model depicts the host-pathogen interaction as the process of reciprocal adaptive genetic change. In the context of microbe and pathogen it means receptors evolve to perfectly match the specific parasite structures (as the lock and key). In other words changes in gene frequencies resulting from selection acting on one population (species) create selective pressure for changes in gene frequencies in another population (species). This type of selection, called frequency-dependent selection, signifies that relative fitness of a genotype depends on its frequency (Carius 2001) (Fig. 4).

Frequency-dependent selection (FDS) can be positive or negative, Positive FDS favours the most numerous allele or genotype which thus increases its frequency and rapidly tends to fixation (Fig. 4c and d). Therefore, in essence the mode of this selection is directional and can be detected by important non-synonymous amino acid changes in different groups or lineages (Quintana-Murci and Clark 2013). Nevertheless positive FDS is less probable in the scope of immune genes.

In contrast, co-evolution determined by negative frequency dependent selection (NFDS) maintains high genetic diversity by favouring rare allelic variants (Takahata and Nei 1990; Stevens 2001) (Fig. 4a and b). In the context of the host-parasite interaction the mechanism of NFDS can be described as follows. The host immune system is adapted to tackle the most common parasite genotype and hence less common genotypes are favoured by natural selection. Rare gentoypes increase in frequency and subsequently become common therefore the cycle goes on (Fig. 5). NFDS is the type of balancing selection, which was already described in MHC genes (Takahata and Nei 1990; Bernatchez and Landry 2003;

Garrigan and Hedrick 2003; Aguilar et al. 2004; Bryja et al. 2006; Piertney and Oliver 2006; Smith et al. 2011). Balancing selection maintains genetic variation and leads to excess of polymorphism and excess of intermediate-frequency alleles. Besides NFDS, balancing selection can act as antagonistic and cyclic selection, selection in variable environment or through overdominance, i.e. heterozygote advantage where heterozygous genotypes confer higher fitness that homozygotes, for example, because they allow to recognize a wider variety of parasites (Doherty and Zinkernagel 1975).

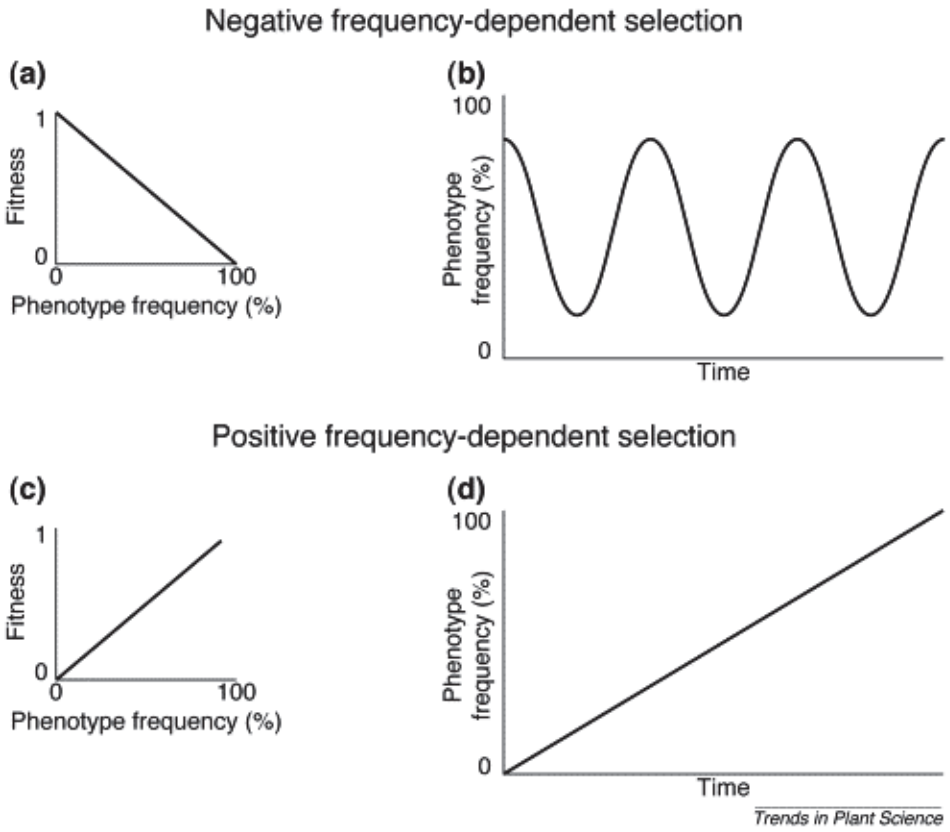


Fig. 4, Frequency-dependent selection. (a) Negative frequency-dependent selection. The rarer a phenotype, the higher its fitness. As a rare phenotype becomes more common, its fitness will decline, leading to a decrease in its frequency. (b) Thus, in negative frequency-dependent selection, the frequency of a phenotype will vary over time. (c) Positive frequency-dependent selection. The more common a phenotype is, the higher is its fitness. This means that over time, the more common phenotype is favored by selection and eventually becomes fixed in the population (d). Therefore, reduced variation is expected there. Schemes (a) and (c) represent simplification of relationships between fitness and genotype or fenotype, because these relationships are not neceserray linear. Adopted from Roy and Widmer (1999). The agreement for reuse of this picture in my disertation was provided by Elsevier using Copyright Clearance Center ("CCC"). Copyright © 2003, Elsevier

Except for various sorts of balancing selection typical Darwinian selection leads to loss of variation since harmful alleles are eliminated from the population by negative slection whereas beneficial alleles are fixed by positive selection. Both positive and negative selection can be detected in a genome on the basis of a decrease of polymorphism in the vicinity of a

selected locus. As the advantegous mutation is driven to fixation so are driven also tightly linked ("hitchhiking") neutral or even slightly deleterious neighbouring DNA regions. Therefore, the presence of strong linkage disequilibrium can be used to identify sites which have recently been under selection. In addition, this process, called selective sweep, will result in overrepresentation of the region around the favoured locus in the population. In other words, besides apparent sequence homogeneity around the positively selected locus, a signature of a selective sweep is an excess of rare variants of this region in the population. A similar reduction of genetic variation can result from elimination of a harmful mutation by negative selection. The tighter the linkage with the counterselected locus, the greater the reduction in polymorphism. By contrast to the selective sweep, the effect of a harmful allele (called background selection) is not an excess of rare polymorphisms since such the allele causes one chromosome (or part of the chromosome) to merely drop out of the population (Hartl and Clark 1997). Genes affected by negative selection generally have very important functions. It was shown that this is a prevalent type of selection acting in the human genome (Quintana-Murci and Clark 2013). However, both positive and negative selections often act at the same time in different parts of the same genomic region, with varying strength and direction.

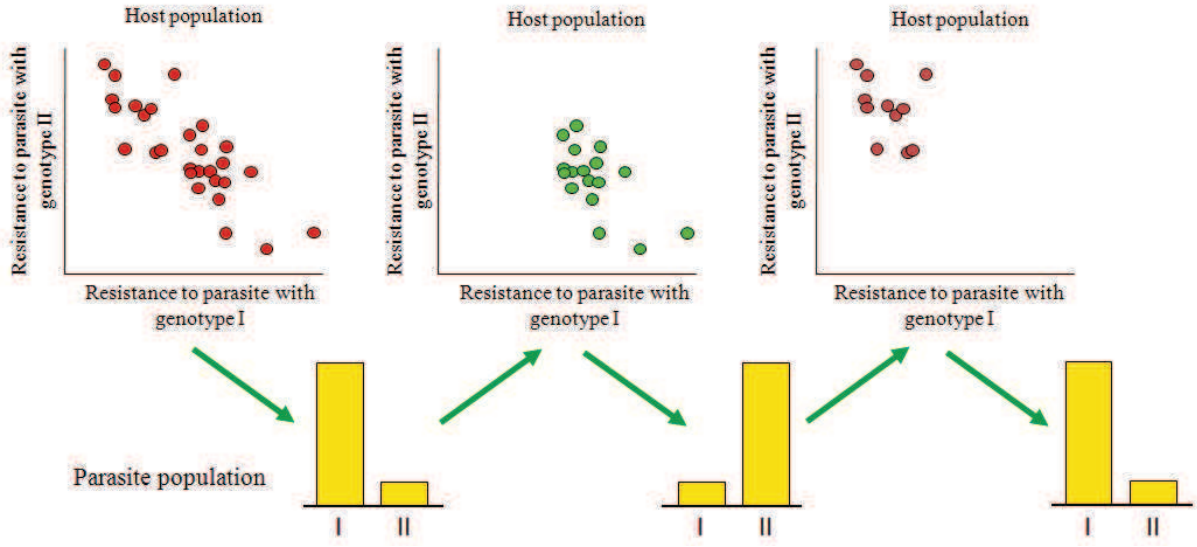


Fig. 5, Balancing selection through negative frequency dependent selection. Hosts resistant to parasite genotype I are likely to be susceptible to parasite genotype II and vice versa. Since the parasite population evolution is closely associated with the evolution of its host's population, the high frequency of parasite genotype I (left) will result in selection favouring hosts more resistant to this genotype (centre) yet susceptible to parasite genotype II. As a result, the latter genotype will prevail in the parasite population, which, in turn, will lead to selection in favour of the respective host resistance etc. Inspired by Freeman and Herron (2007).

We should keep in the mind that pathogen distribution and host specificity may vary in space and time (Hedrick 2002; Hedrick 2004). Local adaptation is a general phenomenon

found in most host-parasite relationships. That means that parasites in a particular area and time infect hosts from that area more efficiently than they infect hosts from another geographically distinct population. Local adaptation to pathogens (for example malaria) presented by different haplotype structure of TLR4 at distinct localities was shown for example in human populations from India, European-Americans and African-Americans (Mukherjee et al. 2009; Netea et al. 2012).

1.2.2 Stochastic evolutionary processes

Although immunity genes are usually targeted by selection (Barreiro and Quintana-Murci 2010; Fumagalli et al. 2011; McTaggart et al. 2012), stochastic processes might also play an important role during their evolution (Grueber et al. 2013), however, their contribution is still difficult to assess. Selectively neutral evolutionary processes are brought about by random changes in a species´ gene pool that are neither advantageous nor disadvantageous for individual organisms nor they are connected to an increase or decrease of mean fitness of the respective population. Among principal stochastic mechanisms are neutral genetic drift (including bottlenecks and founder events) and neutral mutations (Futuyma 2005).

We can imagine genetic drift as a random process of gamete sampling. All genotypes have therefore the same chance to contribute to the next generation. The result of this random sampling is a shift in allele frequencies between subsequent generations. In most extreme situations one allele can either disappear from the gene pool or be fixed owing to genetic drift. Neutral mutations are mutations which do not affect the fitness of individuals. They are invisible for selection and therefore their fate depends on genetic drift. However we should note that the impact of genetic drift is not limited to neutral mutations. Due to genetic drift even advantageous mutations are eventually lost whereas some weakly deleterious mutations may become fixed. This is often the case of small effective population sizes (N_e) where drift can overcome the effect of selection (Nei and Tajima 1981). In such cases populations can pass through an adaptive valley towards higher adaptive peaks and thus follow a new evolutionary trajectory (Eyre-Walker and Keightley 2007). Such mechanism was described, for example, in hominids which have N_e around 10,000 to 30,000 and about 30% of non-synonymous mutations are effectively neutral while *Drosophila* which has N_e about 10^6, the proportion of neutral non-synonymous mutations is less than 16% (Eyre-Walker and

Keightley 2007). In contrast to this finding, the number of non-synonymous mutations fixed by positive selection is close to zero in hominids and in *Drosophila* the proportion is about 50% (Eyre-Walker and Keightley 2007). Comparison between closely related species with similar biologies such as the *Mus musculus* (e.g. *M. m. musculus*, *M. m.m domesticus* and *M. m. castaneus*) confirmed that differences in N_e influence the proportion of amino acid substitutions that have been fixed by selection (Phifer-Rixey et al. 2012).

Genetic drift seems to play an important role during fixation of duplicated gene copies. The relaxed selection permits one copy to accumulate various mutations (neutral, advantageous or deleterious). They can act as a source of sequence variants which can be transferred to the functional gene in the novel combination (e.g. by gene conversion) or they can manifested themselves as advantageous in changing environment (Kimura 1991; Lynch 2006; Chen et al. 2007).

1.2.3 Polymorphism and the effect of non-synonymous substitutions on protein functions

The principal prerequisite enabling to participate in the Red Queen dynamics is genetic polymorphism. In previous chapter I have already described how this variability can be maintained and that evolution of living organisms is the consequence of both adaptive and neutral evolutionary processes. So how we can define genetic polymorphism (variation) and what is its origin?

Nucleotide sequences are composed of four nucleotide bases (A, C, G, T). These bases form 64 triplets (codons), and most of them encode for 20 amino acids (three of them are stop codons). One amino acid then can be logically coded by multiple codons. Nucleotide substitutions generated randomly and continuously during evolution, belong among the most important sources of new genetic variants. They can be either synonymous (silent), which do not change amino acids, and function of protein or non-synonymous, which replace amino acids and might have an important functional impact. Even if the probability of synonymous substitutions to arise by random mutagenesis is nearly twice lower compared to non-synonymous ones, they occur almost always at a much higher rate than non-synonymous substitutions since they are not eliminated by selection. However, the proportion between both types can significantly vary between different genes and/or genome regions.

According to their effects on fitness non-synonymous substitutions can be deleterious, favourable, or selectively neutral. It was shown that approximately half of the known disease-causing mutations result from amino acid substitutions and in humans 25-30% of non-synonymous substitutions were predicted to negatively affect protein function (Ng and Henikoff 2006).

Effect of (point) substitutions on protein function depends largely on biochemical properties of amino acids (Fig. 6). These exchanges can lead to differences in protein conformation what can results in the disruption of interactions with other structures (e.g. ligands or receptor partners) or influence the stability of the protein. Important factors are charge, hydrophilicity or hydrophobicity, size and functional groups (Villaseñor-Cardoso and Ortega 2011). For example, a substitution of Isoleucine (I) for Leucine (L) may not bring about significant changes in a receptor function as both amino acids have similar biochemical properties: nonpolar and neutral). By contrast, replacement of Arginine (R) by Alanine (A) at position 441 of the receptor-binding domain of the SARS coronavirus spike protein interrupts the binding activity (He et al. 2006). Alanine and Arginine are amino acids with completely different qualities (A is small, nonpolar and neutral while R is polar and positive) and therefore we can expect more serious impact on protein function (Fig. 6). Substitutions of amino acids that have similar biochemical properties and are therefore less likely to affect the function of a protein occur much more frequently than more radical changes.

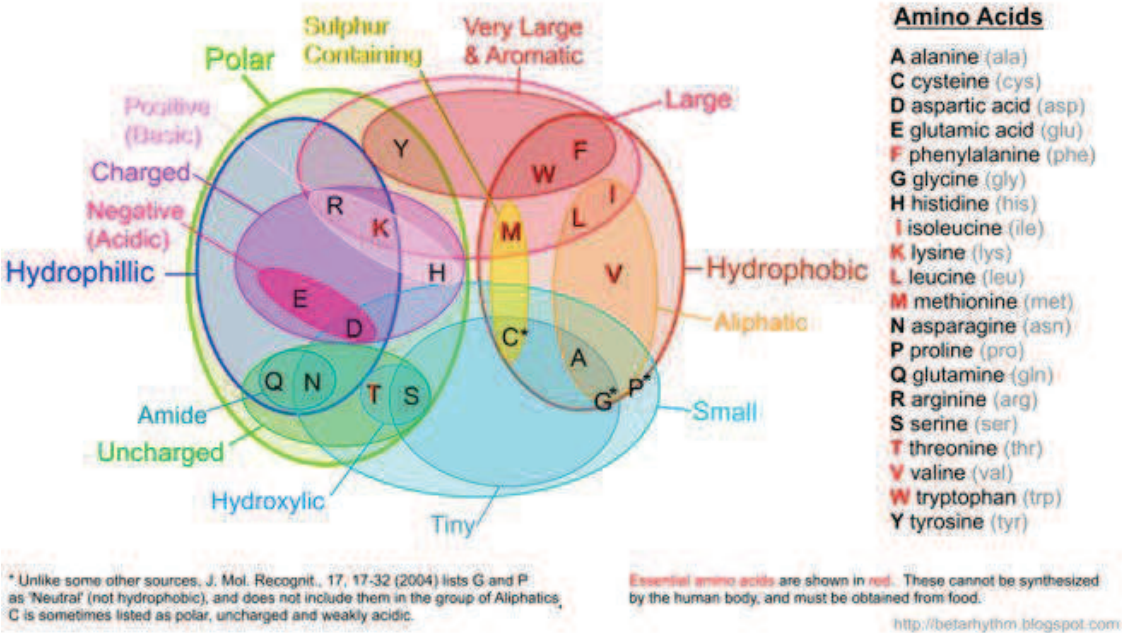


Fig. 6, Biochemical properties of amino acids, adopted from http://nanobiologynotes.blogspot.cz/.

1.2.4 *Effect of non-synonymous substitutions on the function of pattern recognition receptors*

Study of differences in immune response caused by the receptor variation is important for comprehension of the immune system function. Assuming the matching alleles concept mentioned above, the arms-race between parasites and immune receptors could be observed in their DNA sequences in the form of non-synonymous single nucleotide polymorphisms (SNPs). Polymorphism is defined as an existence of different alleles or versions of gene in populations (Villaseñor-Cardoso and Ortega 2011). It was supposed that the PRRs are rather less variable then, for example, highly polymorphic MHC molecules ensuring adaptive immunity. This "high" conservatism of PRRs is related to the fact that parasites cannot easily change their essential structural motifs (PAMPs) that are targeted by the PRRs (Roach et al. 2005; Leulier and Lamaitre 2008; Kang and Lee 2011). However, we saw previously that pathogens are able to change their PAMPs and evade host immune detection. Many studies since discovery of PRRs linked their polymorphism to resistance or increased susceptibility to infectious diseases, e.g. mouse NOD variability associated with susceptibility to Crohn's disease (Hugot et al. 2001), human RLR polymorphism associated with resistance to type 1 diabetes caused by RNA viruses (Loo and Gale 2011) and human TLR4 associated with resistance to Legionnaires disease (Hawn et al. 2005). Similarly, the variability in human NOD2 and TLR4 was linked to Crohn´s disease and tuberculosis susceptibility (Brand et al. 2005; Austin et al. 2008; Kanaan et al. 2012; Zaki et al. 2012). And polymorphism of MBL was for example associated with diverse susceptibility to malaria (Boldt et al. 2009). Many other examples of SNPs having important clinical effect were described for TLRs and will be discussed in chapter 3.5. To conclude this part, polymorphism of immune receptors observed in recent populations is most probably the result of an arm race during co-evolution with microbes; however, non-adaptive evolutionary processes such as genetic drift may also play an important role under certain conditions (e.g. low effective population size, or weak selection). And we saw that even simple substitutions in PRRs sequences can significantly influence the potential of pathogen recognition and the intensity of the immune response. In the following chapter I will focus mainly on Toll-like receptors, which are the main subject of my thesis. However, to get more general objective insight to the problematic of TLRs the knowledge of other types of PRRs will be also discussed.

1.3 Toll-like receptors – a general overview

1.3.1 A brief historical survey: from Toll to Toll-like receptors

Discovery of Toll-like receptors was preceded by the description of Toll genes in *Drosophila* in the mid-80´s. The function of Toll proteins was originally associated with dorso-ventral axis formation in the early stage of *Drosophila* development (Anderson and Nüsslein-Volhard 1984). This finding of Chritstiane Nüsslein-Volhard and Eric Wieschaus was rewarded by the Nobel Prize in Physiology and Medicine in 1995. Soon after this discovery Toll (meaning fantastic, mad or amazing in German slang) genes were connected with immunity in *Drosophila* flies and protection against fungal infection (Lemaitre et al. 1996). Not surprisingly immunologists began to look after a family of proteins with similar immune mechanisms also in mammals (Janeway 1989; Medzhitov et al. 1997; Medzhitov and Janeway 1999). In 1997 the first human orthologous receptor, called Toll-like receptor (TLR4) was identified in humans by homologue sequencing with *Drosophila Toll* genes (Medzhitov et al. 1997). Five years later ten human TLRs were already described (Medzhitov et al. 1997; Janeway and Medzhitov 2002). Since the publication of the famous article by Charles Janeway (Janway 1989) revealing the missing element between innate and adaptive immunity, PRRs with TLRs in the lead stirred up attention of immunologists. After first associations of polymorphism with several infectious diseases in humans have been discovered, TLRs became very popular (Arbour et al. 2000; Lorenz et al. 2000). Today, the TLRs are known to play an important role in the great variety of diseases including autoimmune disorders and therefore the TLR discovery is considered as the most important in immunology during the past 25 years (Janssens and Beyaert 2003; O'Neill 2004).

1.3.2 Structure of TLRs

In spite of their long evolutionary history since their emergence in animals, TLRs share structural and functional similarities from *Drosophila* to humans (Roach et al. 2005; Takeda and Akira 2005; Bassett and Rich). TLRs are relatively large proteins, approximately of 780-1100 amino acids in length. In mammals, TLRs are type I transmembrane

glycoproteins composed of three major domains. The N-terminal horseshoe-like extracellular domain (ECD) is responsible for detection of pathogens (i.e. PAMPs). ECD contains varying numbers (16-28 residues) of repeated motifs that are rich in leucin (LRRs = leucine rich repeats). Detailed description of LRRs in distinct TLRs is in Matsushima et al. (2007). The part of ECD directly responsible for ligand detection is called ligand binding region (LBR). LBR is the most variable part of TLRs. ECD is followed by a single transmembrane helix (TMD) and the C-terminal intracellular domain (ICD) that involves the so-called Toll-Interleukin-1 receptor (TIR) domain (Fig. 7). The TIR domain is homologous to the IL-1R signalling family, which is present from plants to mammals and which is responsible for transmitting signals to cells (Akira et al. 2006; Botos et al. 2011).

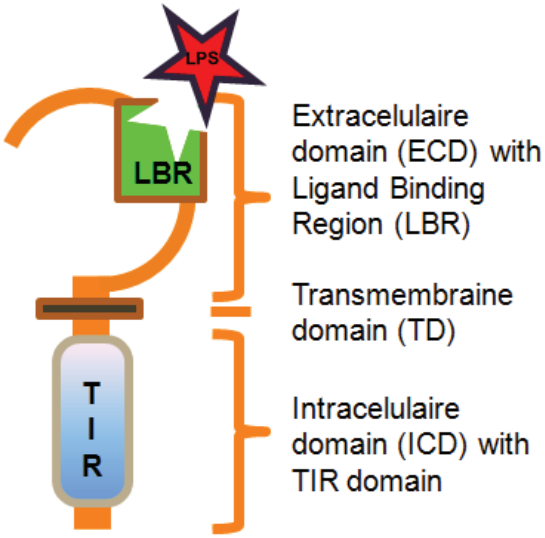


Fig. 7, Simplified description of the TLR4 structure. LBR, ligand binding region, ECD, extracellular domain, TD, transmembrane domain, ICD, intracellular domain, TIR, Toll-Interleukin-1 receptor domain, LPS, lipopolysaccharides of gram negative bacteria.

Until now, four crystal structures of human and/or mouse TLR/TLR-ligand complexes have been described: TLR1/TLR2-lipopeptide (Pam3CSK4), TLR3-dsRNA, TLR4/TLR4/MD-2-LPS and TLR2/TLR6-lipopetide (Pam2CSK4) (Bell et al. 2005; Jin et al. 2007; Kang et al. 2009; Park et al. 2009) (Fig. 8). These studies represent first insight into the molecular basis of TLR recognition and give us a new study field and possibility to analyse relations between TLRs, their partners and ligands. Homology modelling therefore provides us a great tool to explore TLRs, for which the crystal structure, binding sites or even ligands are still unknown.

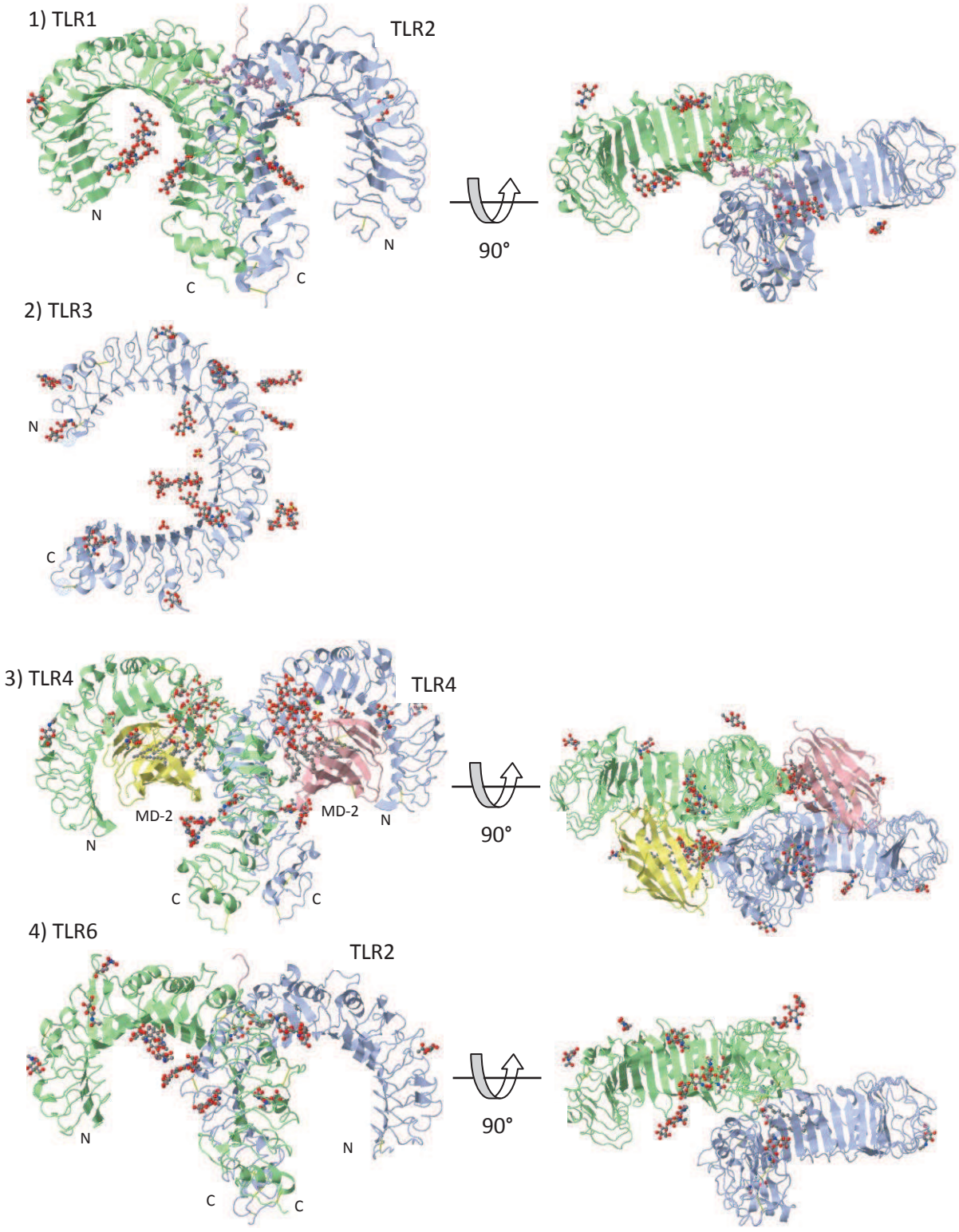


Fig. 8, Crystal structure of TLR ligand-binding domains with its ligands (ligands are represented by small colour dots, detail list of displayed ligands can be found in corresponding publications). Crystal structures (PDB numbers are in parentheses) were adopted from RCSB PDB Protein Data Bank, http://www.rcsb.org/pdb/explore.do?structureId=2z64, 1) dimer TLR1 and TLR2 (2Z7X) (Jin et al. 2007), 2) TLR3 (2A0Z) (Bell et al. 2005), 3) TLR4with MD-2 (3FXI) (Park et al. 2009), 4) TLR2 and TLR6 heterodimer (3A79), (Kang et al. 2009). Visualization and deign corrections were made by myself using First Glance in jMOL v.1.951 (http://bioinformatics.org/firstglance/fgij/index.htm).

1.3.3 Signalling of TLRs

Binding of PAMPs induces ECD dimerization of Toll-like receptors and activates signal transduction pathways (Fig. 9) (Gay et al. 2006). There are two main signalling pathways. The first one, known as MyD88-dependent, is via the TIR-containing adaptors of the myeloid differentiation primary response protein 88 (MyD88). MyD88 is used by all TLRs with exception of TLR3. The second pathway (MyD88-independent) is through TIR domain containing adapter-inducing interferon β (TRIF) – this one is used only by TLR3 and TLR4. Signalling includes activation of the nuclear factor (NF)-κB and interferon regulatory factor-3. Two other TIR domain adaptor proteins that promote signalling are the MyD88 adaptor like protein (MAL) (Fitzgerald et al., 2001; Horng et al., 2001) and TRIF related adaptor protein (TRAM) (Yamamoto et al., 2003). In general, the signalling cascade results in releasing inflammatory cytokines (interleukin-1, TNF-α and interleukin-12), chemokines (interleukin-8, MCP-1, RANTES) and interferons. These cytokines trigger innate immune defences such as inflammation, fever, and phagocytosis in order to provide an immediate response against the invading microorganism. Chemokines are a group of cytokines enabling migration of leukocytes from blood to tissues at the site of inflammation. The cytokines, in turn, bind to cytokine receptors on other defence cells. TLRs are mainly expressed in antigen-presenting cells such as the dendritic cells (DCs), macrophages and B cells and their signalling induce maturation of DCs, which is responsible for initiation of adaptive immune responses (Akira et al. 2001; van den Berg et al. 2004; Pasare and Medzhitov 2004a; Tipping 2006; Kawai and Akira 2010; Michallet et al. 2013).

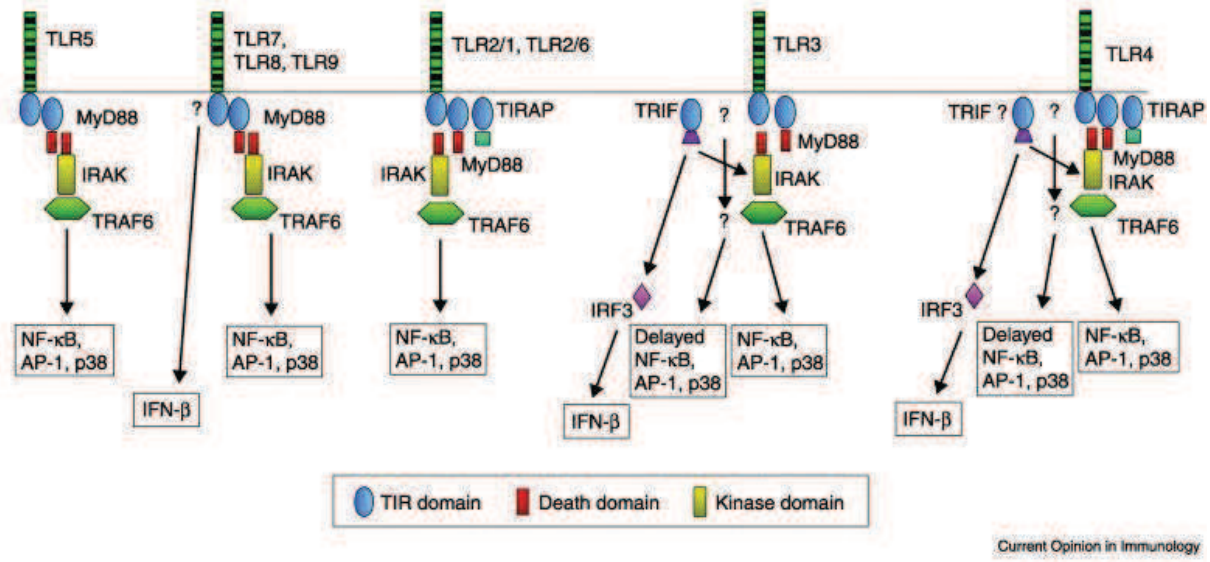


Fig. 9, Schematic of mammalian TLR signaling pathways. All TLRs are thought to signal through a MyD88–IRAK–TRAF6 pathway to induce NF-κB and MAP kinases. The MyD88-dependent pathway downstream of

TLR4 and TLR2 also requires TIRAP. TRIF interacts with TLR3 and induces IFN-β by activating IRF3. Ligands for TLR7, TLR8, TLR9 and TLR4 also induce IFN-β, although it is unclear whether TRIF is involved in these pathways. The TLR3 and TLR4 pathways can induce NF-κB and MAP kinases in the absence of MyD88 with delayed kinetics. Question marks indicate the possible presence of additional signaling molecules. Adopted from (Kopp and Medzhitov 2003). The agreement for reuse of this picture in my disertation was provided by Elsevier using Copyright Clearance Center ("CCC"). Copyright © 2003, Elsevier

1.3.4 Origin and function of the TLR family

TLRs are an ancient family which has probably arose before the Cambrian period (Roach et al. 2005). We can therefore find TLR homologues in many organisms from plants to invertebrates (insects, cnidarians, nematodes, and crustaceans) to vertebrates and mammals (reviewed in Leulier and Lemaitre 2008; Vinkler and Albrecht 2009). During the last 700 million years the TLR family experienced multiple gene duplications and sequence divergence and today we classify the vertebrate TLRs into six major groups or families: TLR1, TLR3, TLR4, TLR5, TLR7 and TLR11 (Roach et al. 2005; Temperley et al. 2008; Huang et al. 2011) (Fig. 10). A half of these families involve only one gene; however, in TLR1, TLR7 and TLR11 groups we can find multiple paralogues. Family TLR1 contains TLR1, TLR2, TLR6 and TLR10, in birds known as TLR1La, TLR1Lb, TLR2a, TLR2b. Family TLR7 include TLR7, TLR8 and TLR9 and family TLR11 cover TLR11, TLR12, TLR13, TLR21 and TLR23. One should be also aware that each species has its own spectrum of TLRs. For example, there are 17 distinct TLRs in bony fish, while only 11 TLRs were described in the fugu fish, 10 different TLRs were described in chicken and in the sea urchin (*Strongylocentrotus purpuratus*), as many as 222 different TLRs were revealed (Huang et al. 2008; Hughes and Piontkivska 2008; Rebl et al. 2010). The number of TLRs differs also across mammalian species. To date 10 and 12 functional TLRs have been described for humans and rats and mice, respectively (Dembic 2005). TLR10 which exists in humans is not functional in the mouse genome due to a retrovirus insertion; while mice have, in addition to humans, TLR11, TLR12 and TLR13 (Fig. 11, these three TLRs were lost in humans). The role of TLR11 was associated with detection of uropathogenic bacterial components (Zhang et al. 2004). However the function and ligands of TLR12 and TLR13 are still not well identified (Kawai and Akira 2010). Detailed overview of TLR number in distinct groups can be found in Leulier and Lemaitre (2008).

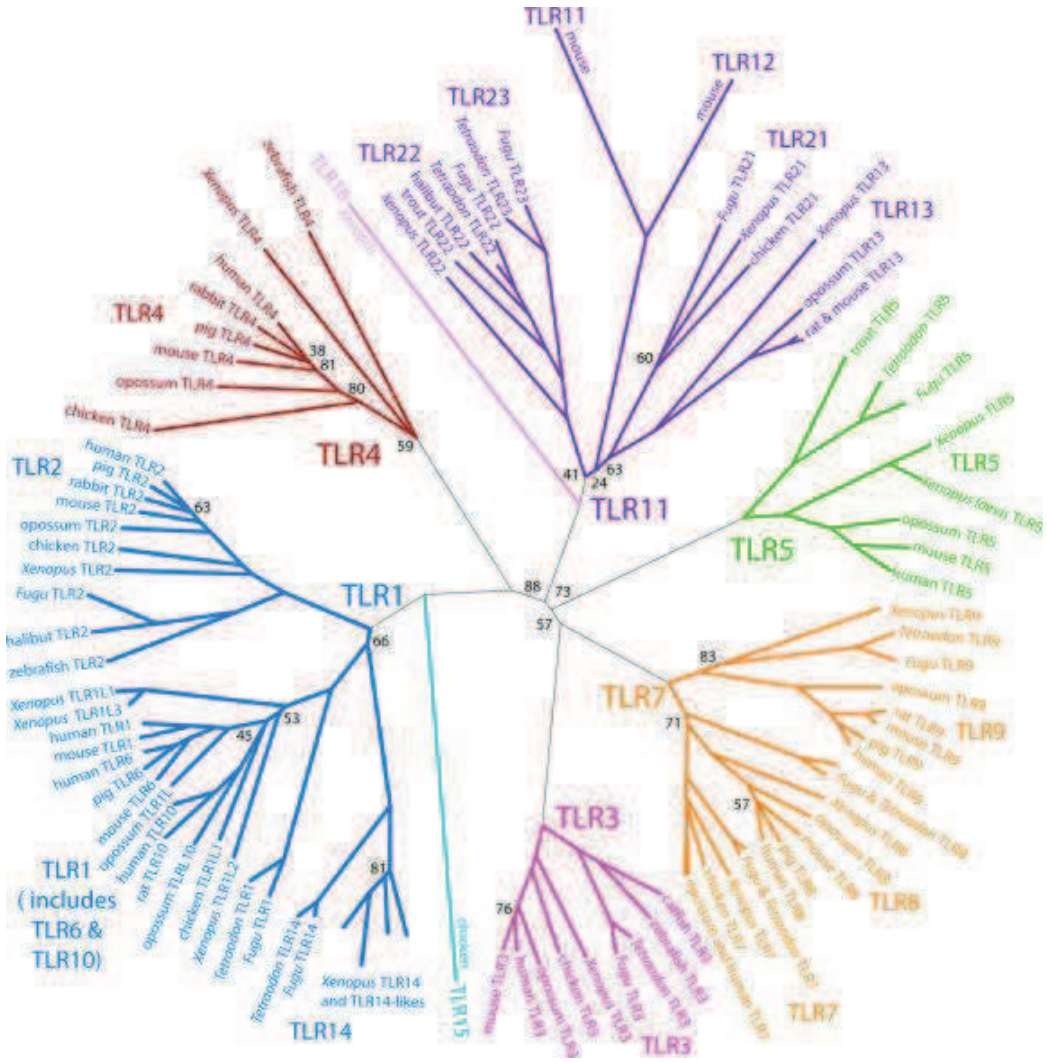


Fig. 10, Evolution of major vertebrate TLR families Adopted from (Roach et al. 2005) with permission from the National Academy of Sciences, USA. Copyright © 2005 National Academy of Sciences, U.S.A.

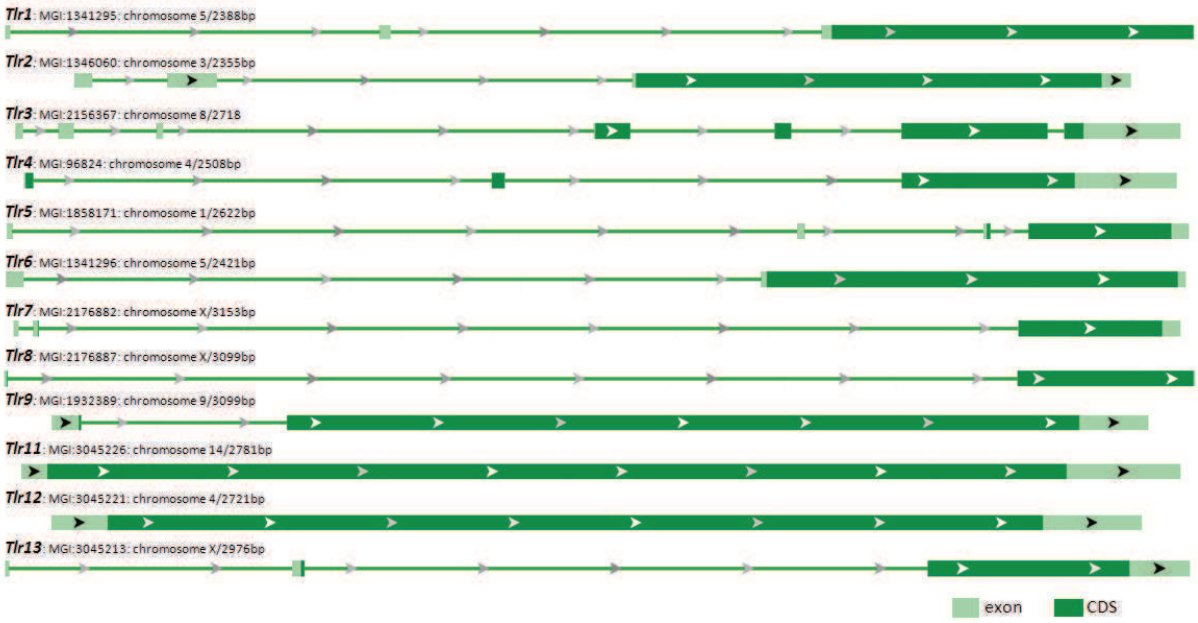


Fig.11, Overview of composition, localization and length of 12 *Mus musculus Tlr*s, CDS, coding region, number behind *Tlr* name indicates Mouse Genome Informatics identification.

TLRs often cooperate together and form heterodimers. This fusion results in dimerization of TIR domains what is necessary for downstream signalling (Manavalan et al. 2011). Different TLRs are responsible for recognition of different PAMPs (Roach et al. 2005). It was also documented that same TLR in different species may recognize different ligands or can significantly differ in immune response. Intraspecific variability of commensal flora in the gut-system seems as one of possible explanations (Werling et al. 2009). Generally two subclasses of TLRs are distinguished according to the ligands that are targeted. Here I describe ten main mammal TLRs of both subclasses (Fig. 12).

The first subclass (TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10) can detect predominantly bacterial components and in this work I will call them "bacterial-sensing". They are expressed in the cell membrane. TLR2 forms heterodimers with TLR1 and TLR6 (Ozinsky et al. 2000). As a consequence of this cooperation it is able to detect wide spectrum of different ligands. Heterodimers with TLR1 detect triacyl lipopeptides, while collaboration with TLR6 results in detection of diacyl lipopeptides (Akira et al. 2006). TLR1 and 6 has very similar length and organization of domains because in mammals they arise by gene duplication after the divergence of placental mammals and marsupialis and before the radiation of the eutherians (cca 100Mya) (Kruithof et al. 2007). Therefore they are located in tandem at the same chromosome 5 (Huang et al. 2011). TLR2 recognizes also peptidoglycan, bacterial lipoproteins, porins from *Neisseria* spp., phenol soluble factor from *Staphylococcus*, glycosylphosphotidylinositol lipid from *Trypanosoma cruzi*, zymosan from yeast cell walls and LPS dissimilar to LPS of gram-negative bacteria (e.g. LPS of *Prophyromonas gingivitis* and *Leptospira interrogans*) (Janeway and Medzhitov 2002; Akira et al. 2006). TLR10 functions as homodimer or heterodimers with TLR1 or TLR2 (Hasan et al. 2005). Its ligands seem to be the same as for TLR2/1 and TLR2/6 heterodimers (Govindaraj et al. 2010).

The repertoire of TLR4 ligands is impressive and comprises mannan of *Candida albicans* (Tada et al. 2002), LPS of gram-negative bacteria (Poltorak et al. 1998) or, for example, flavolipin of *Flavobacterium meningosepticum* (Gay and Gangloff 2007). While classified among bacterial sensing TLRs, TLR4 binds also some viral components, e.g. viral fusion proteins of respiratory syncitial virus (RSV) and vesicular stomatitis virus (VSV) (Kurt-Jones et al. 2000; Georgel et al. 2007). Other endogenous ligands detected by TLR4 are fibrinogen, fibronectin, hyaluronic acid, heparin sulphate, beta-defensins or heat-shock proteins (Okamura et al. 2001; Akira et al. 2006). Binding of LPS occurs after dimerization of TLR4 and binding MD-2 (lymphocyte antigen 96). MD-2 binds to ECD of TLR4 and to the hydrophobic portion of LPS (Botos et al. 2011). TLR5 homodimers recognize flagellin, the

essential component of bacterial flagella (Hayashi et al. 2001). TLR5 binds flagellin monomers using conserved regions at the N- and C-terminal ends. Complementary positions at flagellin sequence are highly conservative due to maintaining the correct function. Any mutations at those places would lead to defect flagella. Therefore flagellin becomes an ideal ligand detected by immune receptors (Andersen-Nissen et al. 2005; Botos et al. 2011).

The second subclass of TLRs (TLR3, TLR7, TLR8 and TLR9) is expressed within endosomal/lysosomal compartments and the endoplasmatic reticulum (ER) and can target mainly viral components (ssRNA, dsRNA, CpG DNA). Therefore, I will call them "viral-sensing". TLR3 functions as a homodimer and detects dsRNA which is produced by most viruses at some stage in their life cycles. TLR3 is therefore a very important receptor during viral infection. Binding to dsRNA at least 46 bp long occurs only at pH 6.5 and bellow (Leonard et al. 2008). The limited length of dsRNA is probably the mechanism how to avoid self reactivity, because shorter sequences of dsRNA (miRNA or tRNA hairpins) occur normally in host cells (Botos et al. 2011). TLR7 and 8 detects single-stranded RNA (ssRNA). Together with TLR9 they belong among the largest TLRs with more than 1000 amino acids thanks to insertions in LRR 2, 5 and 8 and many structurally undefined stretches between LRRs 14 and 15. C-terminal fragment of ECD till LRR 14 is cleaved in endolysosome and is not involved in ligand binding (Ewald et al. 2008; Wei et al. 2009; Botos et al. 2011). TLR9 recognizes unmethylated CpG motifs present in viruses and bacterial DNA (bacteria do not poses methylation enzymes) after bacterial lysis (Häcker et al. 1998; Krieg 2000).

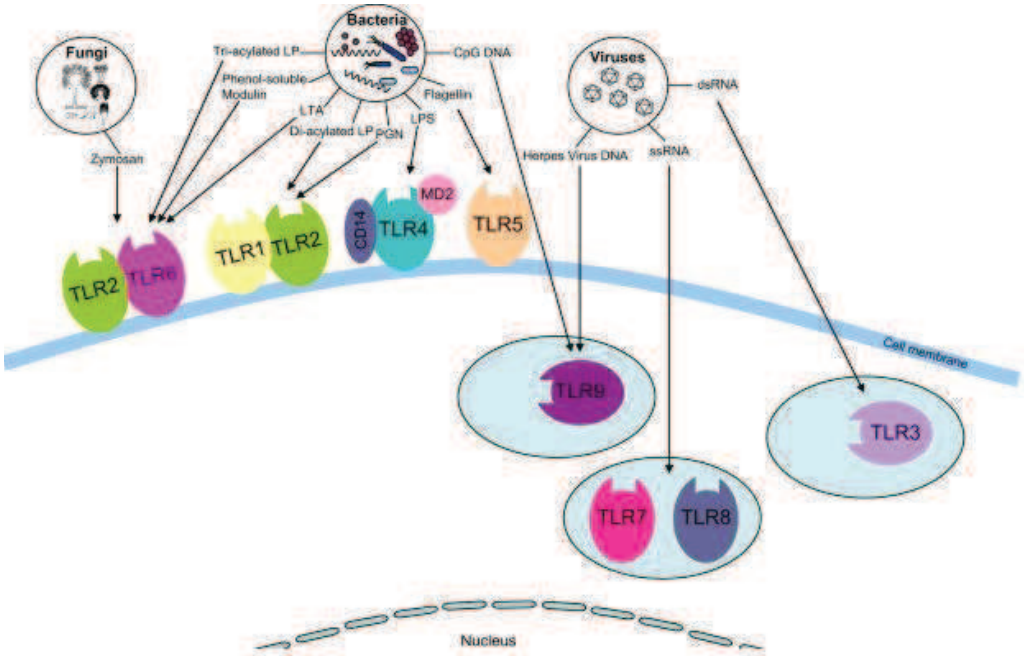


Fig. 12, Overview of localization and ligands of TLRs, adopted from (Mempel et al. 2007), Copyright © 2007, Elsevier.

1.3.5 Variability and polymorphism of TLRs

Changes in immune response caused by non-synonymous SNPs in TLR genes have been documented mainly in humans and domestic or laboratory animals. For example substitution in intracellular domain (ICD) of TLR2 was associated with susceptibility to lepromatous leprosy (Kang and Chae 2001) and substitution R753Q was also associated with susceptibility to tuberculosis (Ogus et al. 2004). TLR4 polymorphism was associated with resistance to Legionnaires disease (Hawn et al. 2005). Two non-synonymous SNPs (D299G and T399I) in the ECD of TLR4 have been associated with an increased susceptibility to infections caused by the Gram-negative bacteria *Brucella spp*, a respiratory syncitial virus, and the parasite *Plasmodium falciparum* causing malaria (Tal et al. 2004; Mockenhaupt et al. 2006; Rezazadeh et al. 2006; Ferwerda et al. 2007). TLRs variability was described also between laboratory mouse strains, chicken, cattle or pigs with possible functional consequences (Smirnova et al. 2000; Leveque et al. 2003; Seabury et al. 2010; Fisher et al. 2011; Raja et al. 2011; Malik 2011; Shinkai et al. 2012). Other examples of association between individual substitutions and diseases can be found in following reviews: Schröder and Schumann 2005; Pandey and Agrawal 2006; Villaseñor-Cardoso and Ortega 2011; Netea et al. 2012. However, present knowledge about TLR polymorphism in free-living populations of vertebrates is rather scarce (Tschirren et al. 2011; Tschirren et al. 2012) and actually only one study reports real measurable impact of observed polymorphism to animal fitness (Tschirren et al. 2013). Tschirren et al. (2013) studied functional significance of naturally occurring polymorphism at TLR2 in wild bank voles (*Myodes glareolus*) and infection with *Borrelia afzelii* (Tschirren et al. 2013). They found that different variants of TLR2 were associated with infection caused by *B. afzelii* and that protective variants increased in frequency relatively recently.

Evasion microbe strategies concerning TLR signalling

Several microbe evasion strategies were described also in relation with TLRs. For example flagellated bacteria *Campylobacter jejuni*, *Helicobacter pylori* and *Bartonella baciliformis*, can produce modified flagellin which can not be recognized by most frequent variant of TLR5 (Andersen-Nissen et al. 2005). Manipulation (inhibition of signalling) of

TLRs by viruses was documented in studies of Vaccinia virus (O'Neill 2004; Bowie and Haga 2005). The variability of LPS, which is the main ligand for TLR4, was associated with differences of virulence among bacterial strains and even intrastrains (e.g. *Pseudomonas aeruginosa* or *Yersinia pestis*) (Day and Marrceau-Day 1982; Ray et al. 1991; Knirel et al. 2006; Montminy et al. 2006). The variability of LPS can affects the adhesion properties of the microorganism to the cells of its host and induces release of inflammatory mediators. Modifications of LPS play an important role in the infection process, evasion of immune response of the host, and serotypification of Gram–negative bacteria (Robinson et al. 2008). Deatiled overview of variations in lipid A which is important component of LPS and its effects on recognition by TLR4 is described in Maeshima and Fernandez (2013). All these strategies can be therefore important impulses for TLRs to maintain polymorphism allowing detecting also modified PAMPs.

1.3.6 Evolutionary forces acting on TLRs

Because of their important function in recognizing conserved pathogen structures, TLRs have been considered as conservative genes evolving mostly under purifying (i.e. negative) selection (Roach et al. 2005). However distinct TLRs might differ in their biological relevance (in the whole spectrum from essential to redundant) and therefore in direction and strength of selective pressures (Innan and Kondrashov 2010; Huang et al. 2011). Soon after the first studies at population level (see Roach et al. 2005) have been performed, various mechanisms were described to play the role in the evolution of TLRs besides negative selection. For example, Ferrer-Admettla and his team have revealed adaptive balancing selection acting on human TLR1 and TLR6 genes and signal of positive selection in TLR9 and TLR10 in Europeans, while in TLR4 no evidence of positive selection was found (Ferrer-Admetlla et al. 2008). Evidences of positive selection were described also in all mammal TLRs, at both intraspecific and interspecific level (Park et al. 2010; Wlasiuk and Nachman 2010; Areal et al. 2011; Tschirren et al. 2011). Intraspecific positive directional selection, as well as signals of positive selection and purifying selection were found to affect TLRs also in birds (Downing et al. 2010; Alcaide and Edwards 2011; Grueber et al. 2012). On the contrary, in bird populations affected by strong bottleneck the dominant force influencing evolution of TLRs seem to be neutral genetic drift (Grueber et al. 2012; Grueber et al. 2013). Genetic drift also shaped genetic history of human TLR4 during population expansion (lost of

advantageous mutation Gly299 against Malaria in Indians from Amazon jungle) (Netea et al. 2012). Gene conversion have been recently reported between paralogues of TLR1 gene family in birds and mammals and TLR6 and TLR10 in platypus (Huang et al. 2011; Mikami et al. 2012). Results of Huang et al. (2011) also suggest the important role of gene conversion during maintaining the conservation of domains responsible for dimerization, transmembrane and TIR domain in TLR1 gene family.

All mentioned examples, which came from very recent studies, bring us the first insight into problematic of TLR evolution. They showed us that diverse evolutionary processes play an important role during TLR evolution and that they differ at interspecific and intraspecific level. Therefore it is necessary to observe these interesting receptors from different perspectives.

1.4 Thesis aims

Toll-like receptors (TLRs) are among the most important and at this time the best studied receptors of innate immunity. However, the majority of knowledge came from studies based on humans and domestic or laboratory animals. We know that results raised from such studies can differ from the situation occurring in free-living populations (Abolins et al. 2011; Pedersen and Babayan 2011). Therefore study of TLRs in natural populations is of the main interest for immunologists, evolutionary biologists and epidemiologists seeking to understand and foresee emergence of zoonoses. The general objective of this thesis is to study evolutionary forces shaping variability of TLRs in free-living populations. As a biological model I chose mostly commensal murine rodents, which 1) are widespread clinical models in biomedicine and 2) are principal vectors in transmission of emerging diseases and thereafter considered as an actual threat to humans during ongoing global changes.

To decide which evolutionary processes shaped rodent TLRs I applied various approaches and I tried to study *Tlr* genes from different perspectives. The first important aspect influencing the evolution of genes in general is simply the history of species. To analyze how *Tlr* genes have been shaped by neutral or non-adaptive processes I called up also genes, which reliably reflect the neutral evolutionary processes. Both types of genes were subsequently analyzed by phylogenetic and population-genetic methods. Second, in view of the fact that TLRs belong to important immunity receptors and their function is linked with microbe detection I have looked for the signs of selection at the molecular level by comparing the pattern of synonymous and non-synonymous substitutions and by analysing protein structures of focussed molecules. The evolution of TLRs was finally discussed at both the interspecific and intraspecific levels, because the most relevant evolutionary processes are not necessarily the same at microevolutionary and macroevolutionary scales. From the macroevolutionary perspective we are able to observe and detect direction of selection over broader timescale, while microevolutionary perspective permits us to study recent local adaptations to pathogenic environments. These two different scales were chosen to get a complex and objective insight into TLR evolution.

2 MATERIAL AND METHODS

2.1 Rodents and rodent-born infectious emergent diseases

"We live in a time where there exists a virtual viral superhighway, bringing people into contact with pathogens that affect our adaptation. The present pattern reflects an evolutionary trend that can be traced to the beginning of primary food production. ***The scale has changed!*** *The rates of emerging disease and their impact can now affect large segments of the world population at an ever increasing rate, and we need to be increasingly aware of the implications for today s human populations around the globe."*

George J. Armelagos at al. 1996 (Professor of Anthropology at the Emory University in Atlanta, Georgia, whose work has provided invaluable contributions to the theoretical and methodological understanding of human disease).

Curiously I will start this chapter with humans. The agricultural revolution which happened 10,000 years ago has resulted in important change in human life style (Armelagos et al. 1991). People have started to live at the same place for a long time and their dwellings have started to provide irresistible shelter for unwanted illegal lodgers – rodents (Ewald 1994). Especially representatives of two murine rodent genera (*Rattus* and *Mus*) have become frequent visitors in human pantries. Since this period, the evolutionary history of mice and rats was closely linked to humans, what is documented in several reports about common migration routes (Tollenaere et al. 2010; Jones et al. 2011; Jones et al. 2012; Jones et al. 2013; Konečný et al. 2013). Murine rodents are, not only thanks to humans, widespread animals and today we can found them on all continents at relatively high densities except of Antarctica. The Old World murine rodents (i.e. original Old World rats and mice which belong to the subfamily Murinae) include over 570 species, 125 genera and 29 genus divisions (Musser and Carleton 2005; Steppan et al. 2005) (see Fig. 13 for current phylogeny of murine rodents). Due to recent diversification of most murine lineages, phylogenetic relationships are not completely resolved and a lot of taxonomic work is necessary.

Fig. 13, Phylogeny of the subfamily Murinae. The tree is based on dataset of Rowe et al. (2008) (phylogeny was based on GHR, BRCA1, RAG1, BDR, IRBP, AP5, and 4 mitochondrial genes), Steppan et al. (2004) (GHR, BRCA1, RAG1, c-myc), Lecompte et al., (2008) (GHR, IRBP, *mt-Cytb*), Jansa and Weksler (2004) (IRBP), Watts and Baverstock (1995) (DNA-DNA hybridization). Red arrows depict the two groups analyzed in this dissertation. Page copyright © 2005 http://tolweb.org/Murinae/16536/2008.09.08 *in* The Tree of Life Web Project, http://tolweb.org/, The figure was adopted from http://tolweb.org/Murinae/16536,

This group originated most probably Southeast Asia, what is supported by biogeographic patterns and fossils, especially dental morphology, which reflects the ecology of the area (Patnaik 2011). The first murine fossil is the genus *Antemus* (*Antemus chinijiensis*) was recorded from the Middle Miocene (~14 Mya), from Siwalik fossil beds of Northwestern India and Pakistan (Steppan et al. 2005; Patnaik 2011). Subsequently two genera (*Progonomys* and *Karnimata*) appeared, giving rise to two lineages: *Mus* and relatives and *Rattus* and relatives, respectively. Their most recent common ancestor lived about 12.5 Mya after the initial radiation of the subfamily (Fig. 14) (Benton and Donoghue 2007). Subsequent rapid diversification may be associated with the expansion of murines from Southeast Asia into Africa, Australia, and Eurasia (Steppan et al. 2005). Further I will focus on two murine lineages, the tribe Rattini and the house mouse (*Mus musculus*).

2.1.1 Origin and radiation of the tribe Rattini in the Southeast Asia

The tribe Rattini covers 35 genera including, in total, 167 species (Musser and Carleton 2005). The following five groups are generally accepted: *Crunomys*, *Dacnomys* consisting of *Leopoldamys* and *Niviventer*, *Maxomys*, *Miromys*, and *Rattus*. The latter group encompasses genera *Rattus, Berylmys, Tarsomys, Limnomys, Abditomys, Bullimus, Bandicota, Sundamys*, nevertheless some phylogenetic relationships are still not resolved, especially in the genus *Rattus* (Verneau et al. 1998; Musser and Carleton 2005; Pagès et al. 2010). This genus underwent two episodes of intense speciation, the first about 2.7 Mya, and the second

about 1.2 Mya, the latter possibly still being ongoing (Fig. 14, Verneau et al. 1998). Recently Marie Pagès has published a study based on extensive sampling and sequencing of two mitochondrial (cytb, *COI*) and one nuclear (the first exon of *IRBP*) genes in which she shed light on the taxonomy of Rattini, difficult to resolve due to recent radiation and probably also hybridization between some species (Fig. 15) (Pagès et al. 2010; Pagès et al. 2013). Especially difficult relationships are within *Rattus tanezumi* which revealed two divergent mitochondrial lineages while this pattern was not confirmed by nuclear sequences (Pagès et al. 2013). The two mtDNA lineages are referred to as clades *R. tanezumi R2* and *R3* in this thesis. *Rattus sakeratensis* is another problematic species which was previously classified with the *R. losea* lineage from central and northern Thailand and the Vientiane Plain in Laos (referred to as *Rattus losea*-like by Pagès et al. (2010)). This lineage was recently distinguished from true *R. losea*, which is restricted to Cambodia, Vietnam, China and Taiwan (Aplin et al. 2011). Monophyly of the genus *Rattus* has also been questioned because of the uncertain position of *Bandicota* (Pagès et al. 2010; Pagès et al. 2013).

Today the genus *Rattus* encompasses around 66 species (Musser and Carleton 2005). Most of them are non-commensal species living in fields and forests. Only two species, the black rat (*R. rattus*) and Norway rat (*R. norvegicus*) are exclusively commensal and tend to settle along human migration routes (Tollenaere et al. 2010). The Rattini includes many vectors of dangerous pathogens infectious to human such as Hantavirus, found, for example, in *R. rattus* (Gou virus), *R. norvegicus* (Soeul virus), *R. tanezumi* (Serang virus), and *Bandicota indica* (Thailand virus) (Hugot et al. 2006; Plyusnina et al. 2009; Blasdell et al. 2011), *Leptospira*, which was found in *R. norvegicus*, *R. exulans*, *R. rattus*, *Bandicota indica*, and *B. savilei* (Kositanont et al. 2003), or *Orientia tsutsugamushi*, the causative agent of scrub typhus which was found in all species within of *Rattus* sensu lato (represented by *Maxomys*, *Dacnomys* and *Rattus* divisions) (Badenhorst et al. 2012). Therefore they represent serious menace to human helth and detailed study of their phylogenetic relationships, potential pathogen transmission and variation of immunity genes of different are of the major priority.

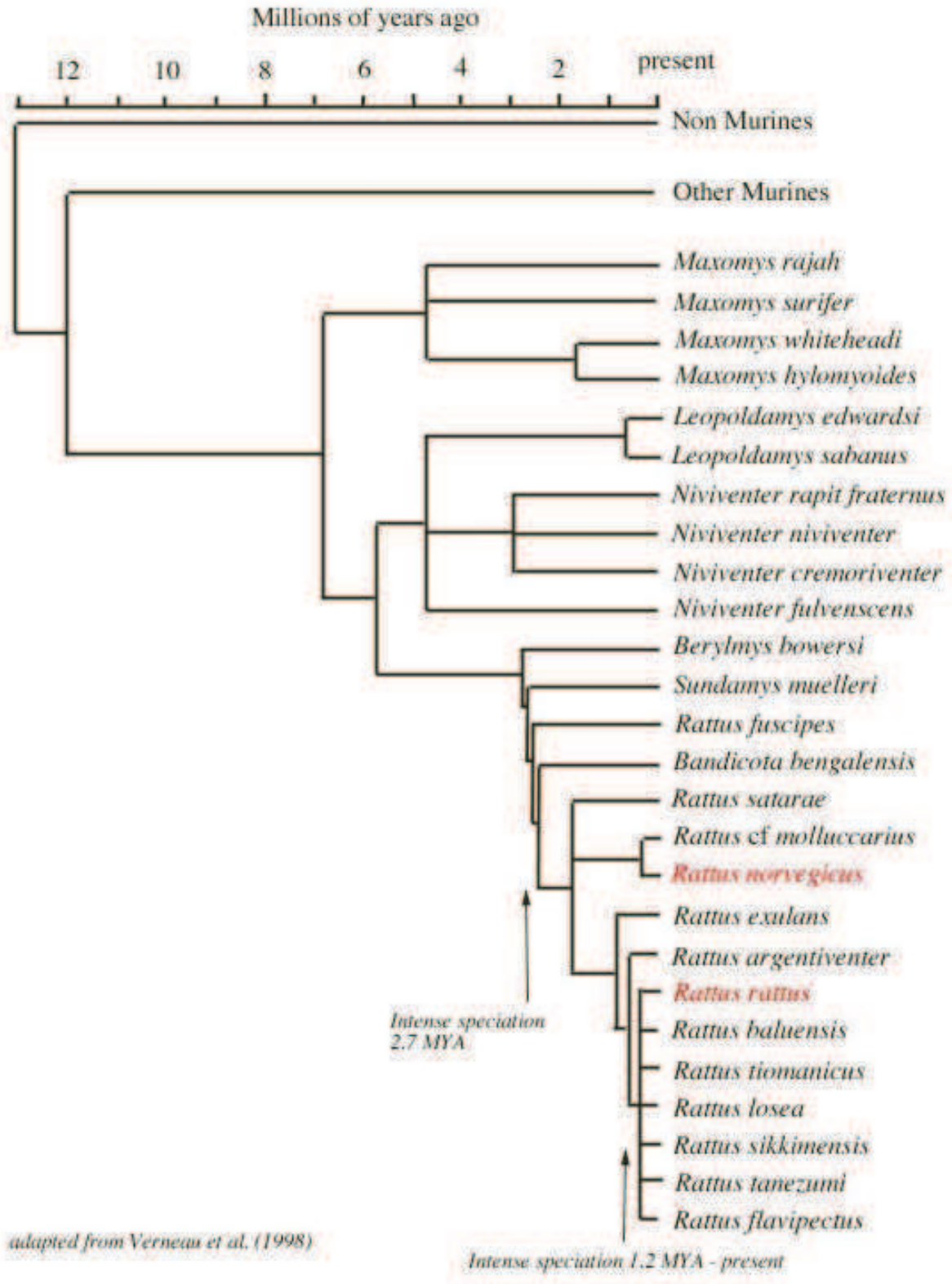


Fig. 14, Phylogeny of rats. *Rattus* underwent two intense periods of speciation. One about 2.7 million years ago (Mya), another about 1.2 Mya which may still be ongoing (Verneau et al. 1998). With the permission from the National Academy of Sciences, USA. Copyright © 1998 National Academy of Sciences, U.S.A.

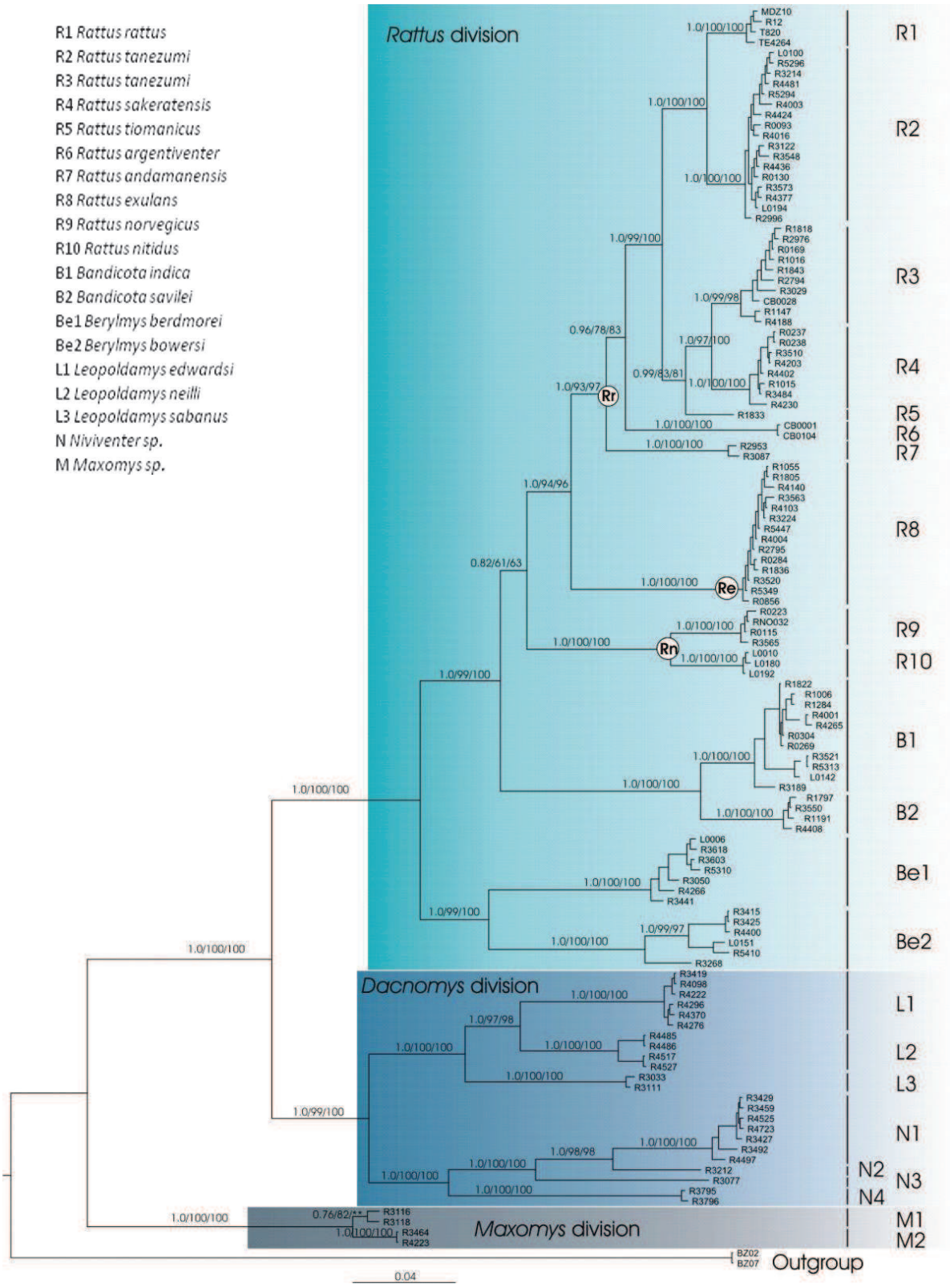


Fig.15, Phylogenetic relationships of the Indochinese Rattini based on combined analysis of cytb, COI and IRBP genes, inferred using Bayesian inference (Pagès et al. 2010). Numbers above the branches are, in turn, posterior probabilities/unpartitioned maximum-likelihood (ML) bootstrap support/partitioned ML bootstrap support. Support values are not shown for very short branches; "**" means the respective clade is not supported by the partitioned ML analysis. Rr stands for *Rattus rattus* species group, Re for *Rattus exulans* species group, Rn for *Rattus norvegicus* species group, following Musser and Carleton's denominations (Musser and Carleton 2005). Adopted from (Pagès et al. 2010).

2.1.2 *Tribe Murini and evolution of house mice (*Mus musculus*)*

The genus *Mus* (Linnaeus, 1758) is geographically widespread group of murine rodents which encompasses about 41 species (Suzuki et al. 2004; Aplin et al. 2011). This genus most probably originated in Northwestern India (Suzuki et al. 2004; Chevret et al. 2005; Duvaux et al. 2011; Macholán et al. 2012). Four subgenera were described: *Pyromys* and *Coelomys* from Southeastern Asia, African *Nannomys*, and the worlwide subgenus *Mus* (Fig. 16) (Marshall 1976; Suzuki et al. 2004; Chevret et al. 2005; Aplin et al. 2011; Macholán et al. 2012).

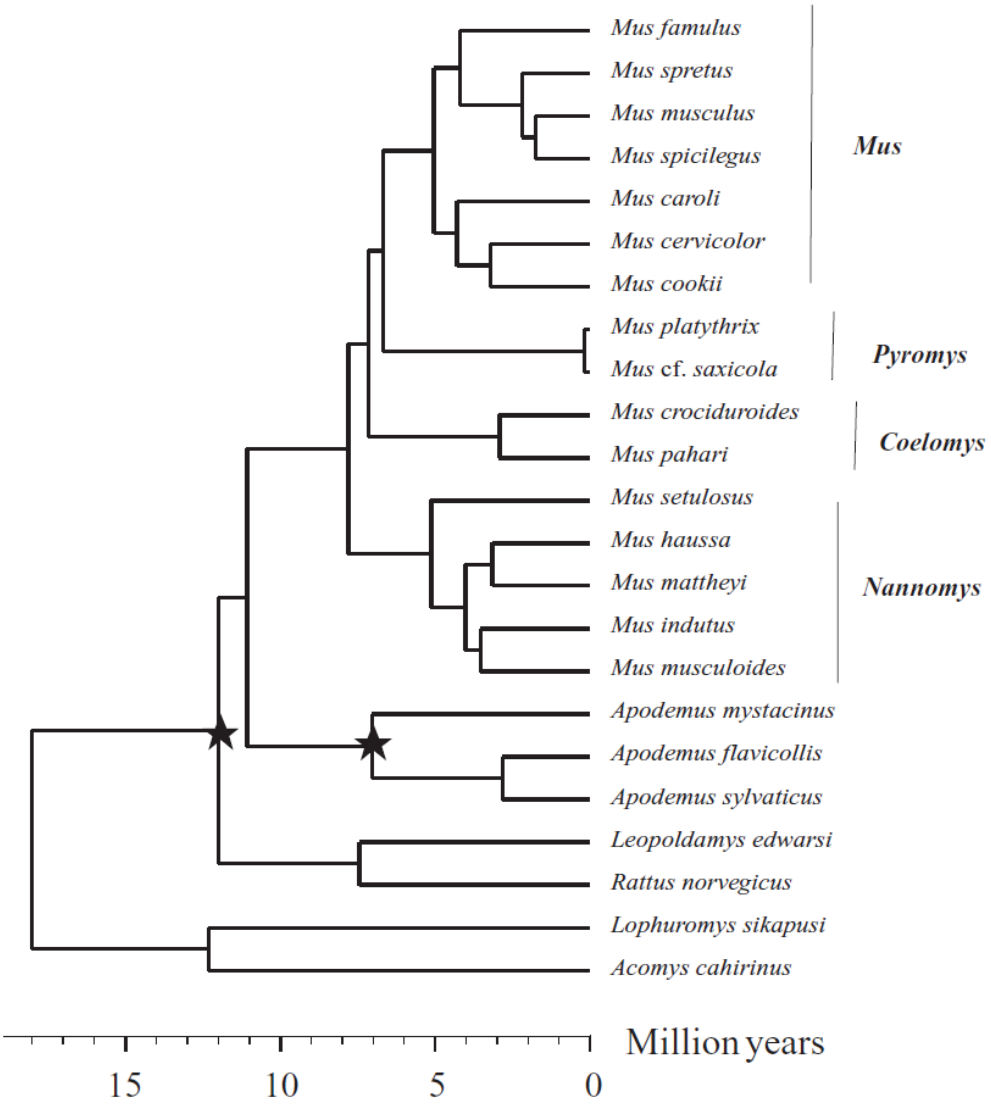


Fig. 16, Maximum likelihood tree reconstructed with a molecular clock hypothesis. Subgeneric names within *Mus* are indicated on the right. The divergence time estimates derived from the two calibration points (stars) are similar. Adopted from Chevret et al. (2005).

The latter includes about 30 species according to Suzuki and Aplin (2012). Phylogenetic relationships within this clade are insufficiently resolved, however, there is a well defined Western Palaearctic group comprising non-commensal species *M. spretus*, *M. scpicilegus*, *M. cypriacus*, and *M. macedonicus*, and one commensal species *M. musculus* (Fig. 17) and can be found in eastern and south-eastern Europe and the Middle East, while *Mus spretus* occupies habitats around the western Mediterranean (Boursot et al. 1993; Bonhomme et al. 2011). Asian clade of subgenus Mus includes *Mus cervicolor* in south-east Asia, *M. cookii* in Nepal, India, Barma and Thailand and *M. caroli* in Indonesia and some of the western Pacific island (Lundrigan et al. 2002). *Mus musculus* originated probably in northern India or northeastern Pakistan and have diversified in geographic isolation during the Pleistocene into several subspecies including *M. m. domesticus*, *M. m. musculus* and *M. m. castaneus* (Boursot et al. 1993; Boursot et al. 1996; Duvaux et al. 2011; Auffray 2012; Cap. 1 in Macholán et al. 2012). *M. m. castaneus*, which has recently been described as polytypic subspecies (Rajabi-Maham et al. 2012), has expanded to southern India, Southeast Asia and Japan, and to the Indo-Pacific. In Japan another "subspecies", *Mus m. molossinus*, was described. However, molecular analyses suggested this form to be a hybrid lineage of *M. m. castaneus* and *M. m. musculus* (Yonekawa et al. 1988; Nunome et al. 2010). This is important with respect to the origin of classical laboratory strains contributed, to some extent, by Japanese fancy mice derived from *M. m. molossinus* (Takada et al. 2013).

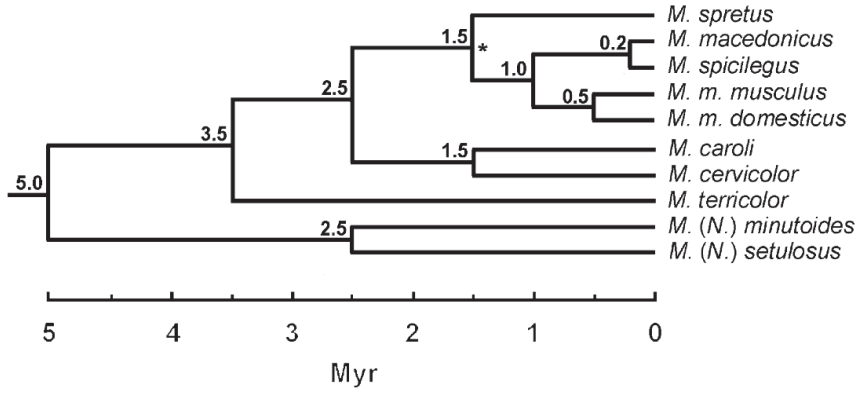


Fig. 17, Phylogenetic relationships among 10 mouse taxa with approximate times of divergence (in millions of years) indicated (numbers above branches). Adopted from Macholán (2008) Copyright © 2008, Elsevier.

Two other house mouse subspecies, *M. m. musculus* (Mmm) and *M. m. domesticus* (Mmd), split around 0.5 Mya (Fig. 17) (Boursot et al. 1993; Geraldes et al. 2008; Macholán et al. 2012) and thereafter they have expanded to their current ranges. Although according to Duvaux et al. (2011), the two subspecies at least once came into contact during the range expansions (probably in Transcaucasia or Iran around 4000 BC), allowing them to exchange

beneficial mutations, they remained for most of the time in allopatry (Duvaux et al. 2011). Given their distinct migration routes (Fig. 18), the subspecies may have been exposed to distinct pathogenic environments for appreciable periods. It seems that commensalism of Mmd arose in the Eastern Mediterranean region known as the Levant (Fig. 18) during the Neolithic Revolution 12,000 years ago (Cucchi et al. 2012; Cucchi et al. 2013). The Western Mediterranean and western/north-western Europe were colonized later, during the Bronze and Iron Age, respectively (Cucchi et al. 2005; Bonhomme et al. 2011). In contrast with Mmd the expansion of Mmm is much less understood because of insufficient archaeological data and sampling. Mmm presence in south-eastern Europe was documented by unique fossil record from Romania around 3600 BC. Subsequent expansion to the west was probably linked with cooper trade routes. Recent analysis has also suggested two geographically distinct subgroups (South and North) in Mmm based on nuclear markers (Nunome et al. 2010). Expansion of *Mus musculus* through geographically distinct regions allow as to study pattern of local adaptations.

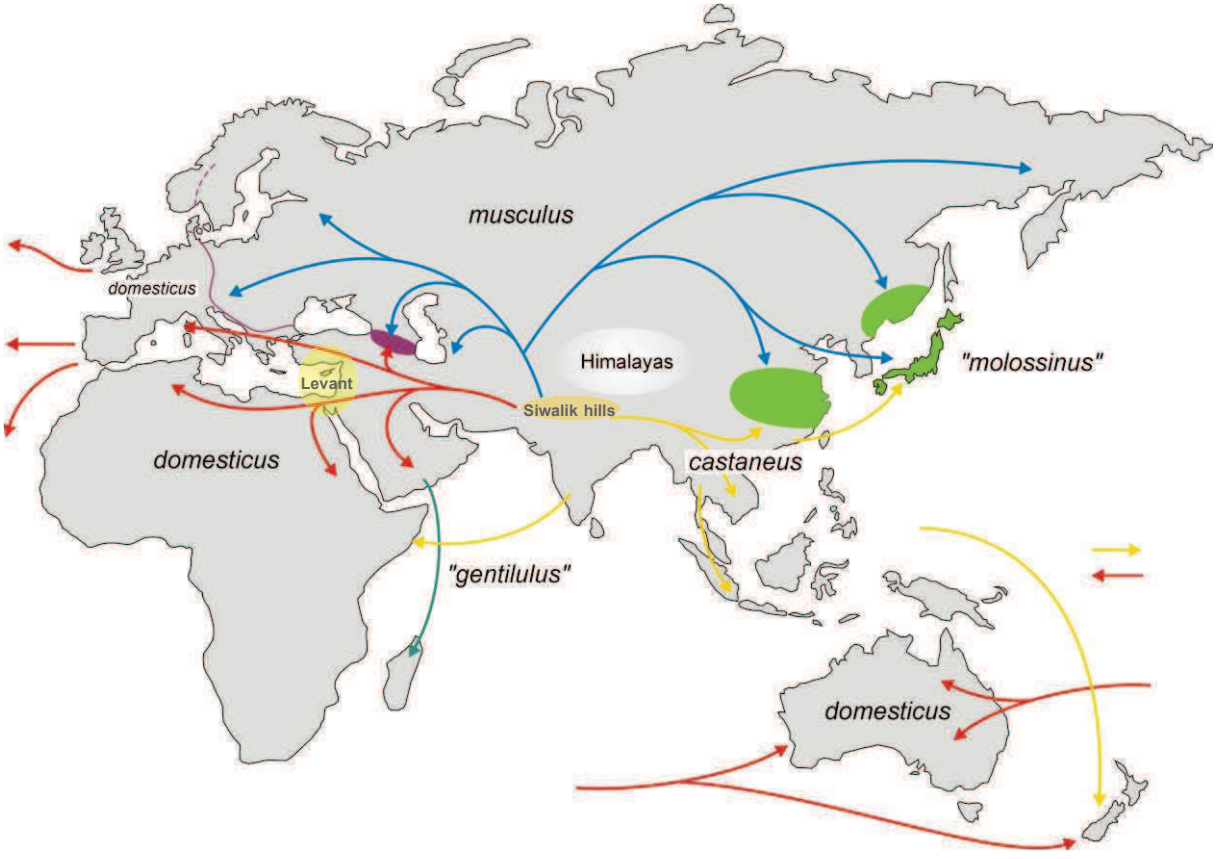


Fig. 18, Colonization routes of the house mouse subspecies starting from northern parts of Indian subcontinent. Arrows indicate movements of the *castaneus* (yellow), *domesticus* (red), and *musculus* (blue) subspecies; *gentilulus* (green) refers to an mtDNA lineage found in Yemen and Madagascar. Mice of the Americas, Australia, and Africa south of the Sahara have been imported by humans. Hybridization between *M. m. musculus* and *M. m. castaneus* has resulted in the formation of a hybrid taxon (*M. m. molossinus*) in Japan, China, and the Russian Far East (green). Hybrid zones between *M. m. musculus* and *M. m. domesticus* in Europe and Transcaucasia are depicted in purple. Adopted from Macholán (2013).

Upon secondary contact Mmm and Mmd established about 20 km wide and over 2500 km long hybrid zone in Europe (European house mouse hybrid zone, HMHZ), stretched from Scandinavia to the Black Sea in Bulgaria (Fig. 17) (see Baird and Macholán 2012 for review). The HMHZ can be considered as a dynamic system and its maintaining is traditionally explained by gene flow into the zone on the one hand and selection against hybrids on the other hand (Key 1968; Bazykin 1969; Barton and Hewitt 1985). Therefore, the HMHZ provides us an exceptional geographic region which allows us to study processes of speciation in the "live broadcast" and to test several conventional hypotheses about hybrid survival.

In general it is suggested that hybrids will have lower fitness due to accumulation of incompatibilities as explicated by the Dobzhansky-Muller model. Searching for genes that confer reproductive barriers between species was firstly logically focused on sex chromosomes (Payseur et al. 2004). It was shown that clines representing gene passage across a hybrid zone are much steeper for sex chromosomes than for autosomes (Tucker et al. 1992; Payseur et al. 2004). However introgression was documented in both chromosomes (Macholán et al. 2008) and searching for genes responsible for speciations continue (Teeter et al. 2008).

Another traditional hypothesis says that the main force driving speciation processes are parasites (reviewed in Karvonen and Seehausen 2012). According to this hypothesis hybrids should suffer from higher parasitic load and subsequent infection due to broken imunity defense mechanism. Such pattern was described also on earlier studies of the HMHZ (Sage et al. 1986; Moulia et al. 1991). Neverthless a recent study on more comprehensive material by Baird et al. (2012) discovered that hybrids have significantly reduced diversity and load of helminths. In direct contradiction to the prevailing paradigm Baird et al. (2012) proposed a new, "vicariant Red Queen", model to explain their results. According to this model immunity genes tracking parasites are more likely to escape Dobzhansky-Muller incompatibilities by generating hybrid variants untargeted by parasites.

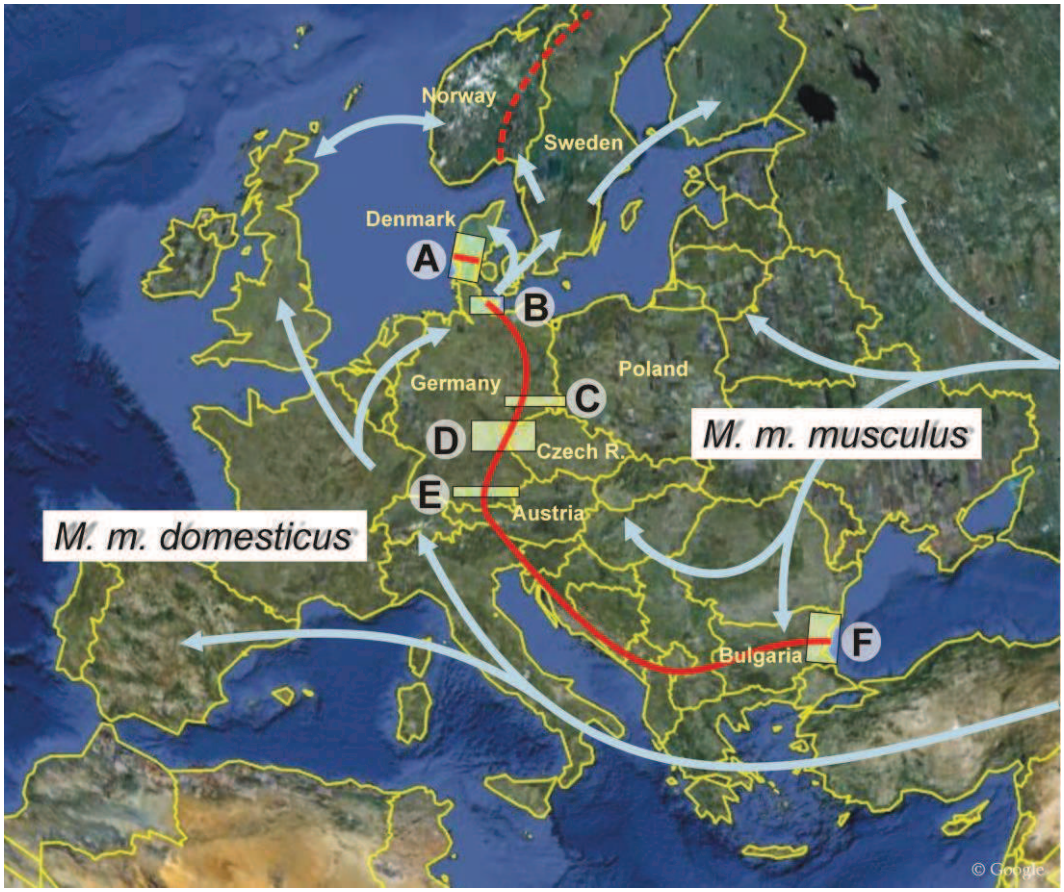


Fig. 19, The course of the *M. m. musculus/M. m. domesticus* hybrid zone in Europe with hitherto studied transects in Denmark (A), northern (B) and eastern Germany (C), northeastern Bavaria (Germany) and Czech Republic (D), southern Germany (E), and Bulgaria (F). The position of the zone in Norway is tentative only (dashed line). As shown by the light-blue arrows, *domesticus* mice colonized the continent from the Middle East through Asia Minor and along the Mediterranean, whereas *musculus* mice followed the route north of the Black Sea. This scenario assumes that the hybrid zone is older in southern parts than in northern parts, with the most recent contact in Denmark and Norway. Note that colonization of Scandinavia by a few hybrid individuals from East Holstein has led to a mosaic genome of Scandinavian mice with nuclear *M. m. musculus* genome and mtDNA of *M. m. domesticus*. Adopted from (Macholán 2013).

2.1.3 Rodent-borne diseases, emergence risk for humans and rodents as model species

Murine rodents are potential reservoirs of many pathogens causing dangerous human diseases, for example, bubonic plaque, salmonellosis, murine typhus, leptospirosis, rat-bite fever or hantavirus pulmonary syndrome (Schönrich et al. 2008; Meerburg et al. 2009). Today due to geographic expansion of humans and their relatively dense settlements, people and synanthropic rodents live in contact and share the same space and often also the same resources (food, water). Therefore rodents represent serious risks for humans because of possibility of the spread of infectious diseases. Climatic changes, antibiotic resistance, and interactions of social, demographic and environmental changes can reinforce the risk of

transmission of rodent zoonotic emerging diseases (Polley and Thompson 2009; Thompson 2013). Consequently the study of non-synonymous SNPs and their associations with pathogens is of the prime importance for assessing potential hazards (Jungi et al. 2011). However, to be honest, people might also represent similar jeopardy for rodents. First, transmission of pathogens is not unidirectional. And second, the Neolithic ancestors of present-day mice and rats surely could not anticipate that their descendants would have to pay once a luxury tax. When people became looking for a model for various medical and research experiments mice and rats were of the first choice. Soon they have become the most common and useful laboratory species (Fig. 20).

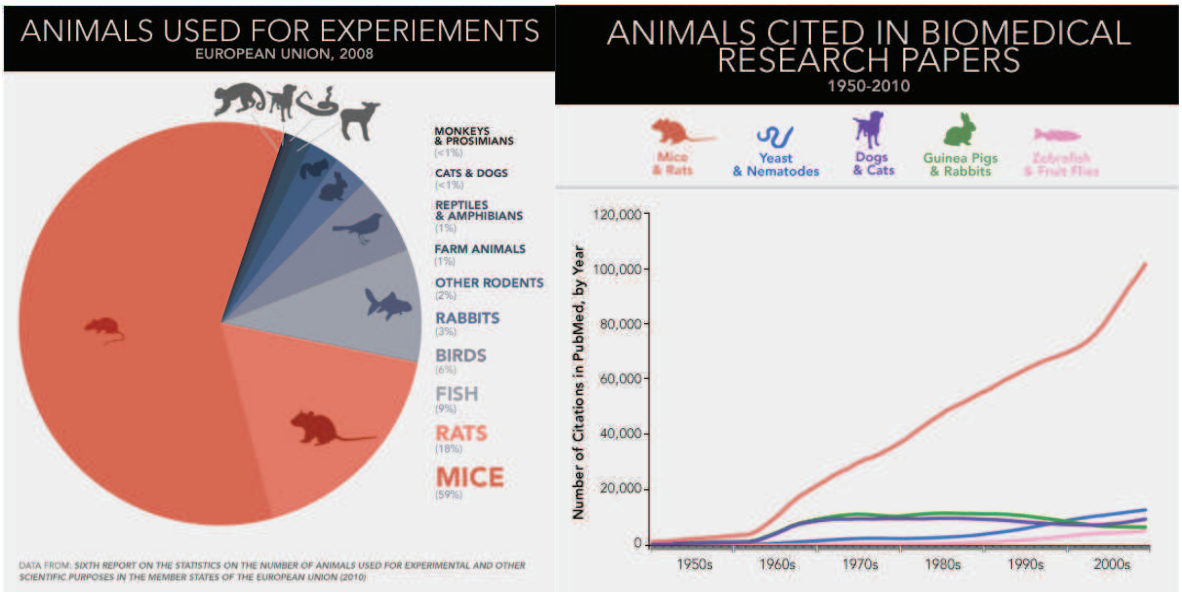


Fig. 20, Taxonomic structure of animals used for experiments by humans. Adopted from Daniel Engber. http://www.slate.com/articles/health_and_science/the_mouse_trap/2011/11/lab_mice_are_they_limiting_our_understanding_of_human_disease_.html

A great advantage of inbred laboratory strains is repeatability of various experiments allowing to understand lots of biological mechanisms from basic to more complex one. However, some recent studies have shown that laboratory strains of unclear origin are not the most optimal models for medical research, because their genetic background is influenced by inbreeding and unnatural conditions in captivity (Salcedo et al. 2007; Stephan et al. 2007; Abolins et al. 2011). Some scientists warn about monocultures in biomedicine (meaning the use of laboratory mice as a principal and unique model) and appeal for using also other vertebrates (Pedersen and Babayan 2011). This also means that for established laboratory models it is very important to focus on free-living populations, where the genetic diversity created by long-time evolution is still preserved (Guénet 1998; Guénet and Bonhomme 2003;

Babayan et al. 2011; Pedersen and Babayan 2011). The choice of our model species was not random, because by choosing wild *Mus* and *Rattus* species we profit from a huge amount of knowledge derived from laboratory strains of rodents (Poltorak et al. 1998; Smirnova et al. 2000; Stephan et al. 2007) and, at the same time, we can get an important insight into variation in the wild (Pedersen and Babayan 2011).

2.2 Analysis of natural selection

To characterize the intensity, direction or type of natural selection which operates on genes involved in immune defense several approaches are available. These tools differ according to the analyzed levels. To search for ancient imprint of selection we can use interspecies neutrality tests whereas for testing for recent selection activities (occurred no longer than $4N_e$ generations ago) intraspecies neutrality tests are the best choice (Barreiro and Quintana-Murci 2010).

2.2.1 Analysis of selection at the intraspecific level

The most standard evaluation of selection is uses codon-based analyses (known as *dN/dS*, detailed later). Comparing protein-coding sequences between species is usefull for searching for rapidly evolving codons or lineages. The null hypothesis of neutrality suggests that the ratio of fixed differences between species to polymorphisms within species should be the same for nonsynonymous and synonymous mutations (Kimura 1991). Rejection of the null hypothesis and deviations from neutrality are then interpreted as the influence of natural selection. The codon-based analyses tend to estimate the ratio between *dN* (number of non-synonymous substitutions per non-synonymous site, also Ka) and *dS* (number of synonymous substitutions per synonymous site, also Ks). The ratio $dN/dS > 1$ is interpreted as the evidence of strong positive selection, while $dN/dS < 1$ is considered as the proof of negative selection. This ratio informs us about accelerated evolution of certain regions of the studied proteins. The McDonald-Kreitman test is another neutrality test which use information from both polymorphic and divergence data and therefore can detect past as well as recent selection (comparison of at least two species is necessary).

Another approach consists of analysis of co-evolutionary relationships. By comparison of species trees with gene trees we can potentially reveal discrepancies between both types of phylogenies which can indicate departures from neutrality. However, results of such analysis should be interpreted with caution because incongruence between a gene genealogy and the species phylogeny can also result from incomplete lineage sorting (ILS) and/or hybridization. Then discrepancies in terminal branches (e.g. due to a recent radiation) are difficult to resolve because both scenarios (adaptive selection and ILS) are plausible, while deeper differences

are more likely to be caused by adaptive selection. For example, ILS was estimated to be responsible for 25% of differences between gene genealogies and species phylogenies for humans, chimpanzees and gorillas (Hobolth et al. 2011). Different types of selection can influence ILS in different ways even if ILS is mostly consequence of stochastic processes. Negative and positive directional selection result in smaller amount of ILS, while balancing selection produce higher amount of ILS (Charlesworth et al. 1993; Hobolth et al. 2011).

2.2.2 Analysis of selection at the intraspecific or population level

On the intraspecific or population level we can use tools of population genetics, such as estimates of nucleotide variability between populations, analysis of population expansion, population differentiation (F_{ST} statistic) etc. Several statistics permit to asses departures from a neutral model of evolution in the distribution of allele frequencies. The most famous is Tajima's D which compares the number of segregating sites with the mean pairwise difference between sequences (Tajima 1989). A negative Tajima's D signifies an excess of low frequency polymorphisms, indicating population size expansion or purifying selection, while positive Tajima's D signifies low levels of both low and high frequency polymorphisms, indicating a decrease in population size and/or balancing selection (e.g. population after bottleneck) (Städler et al. 2009). Other tests which evaluate departures from neutrality based on allele frequencies are Fu and Li's D and F test, and Fay and Wu's H test. As for Tajima's D significantly negative values indicate an excess of low-frequency variants, which can resut from population expansion, weak negative or positive selection. On the other hand significantly positive values represent an excess of intermediate-frequency alleles, which can result from population bottlenecks or and balancing selection. Fay and Wu's H statistic tests for an excess of high-frequency derived mutations, which is the hallmark of posiive selection. However, deep knowledge of demographic history of studied populations is necessary to interpret these statistics.

3 RESULTS

Chapter 3.1: Polymorphism of bacterial-sensing TLRs in wild-derived and classical laboratory strains of *Mus musculus*

Chapter 3.2: Analysis of variability of *Tlr4* at intraspecific level in wild populations of *Mus musculus*

Chapter 3.3: Interspecific level of polymorphism in *Tlr4* and *Tlr7* in wild rodents

3.1 Polymorphism of bacterial-sensing TLRs in wild-derived and classical laboratory strains of *Mus musculus*

Annotation

Classical laboratory strains (CLS) of the house mouse are widely used models in immunogenetics, despite the fact that their genetic polymorphism is limited, and genome as well as gene expression were influenced by artificial laboratory conditions. In comparison to CLS, recently developed wild-derived strains (WDS) reflect reality observed in free-living populations and offer higher level of polymorphism. In the first chapter of this dissertation, I describe structure and variability of five bacterial-sensing TLRs (TLR1, TLR2, TLR4, TLR5 and TLR6) in 20 WDS and compare them with CLS. Obtained data can be used as a basis for planning specific and well-controlled experiments focussed on mechanisms of recognition of pathogenic bacteria by mammalian hosts.

3.1.1 Introduction

As I have already mentioned, classical house mouse laboratory strains (CLSs) are widely used model in immunogenetics, despite the fact that genetic polymorphism is limited and gene expression can be biased by artificial laboratory conditions (Yang et al. 2011). Consequently, one can suspect that they can differ from free-living animals in observed clinical manifestations conditioned by variation among CLSs (e.g. reported variation in susceptibility to monocytotropic rickettsia, *Ehrlichia risticii* (G- bacteria) between ten mouse strains (Williams and Timoney 1994) or observed defective tumoricidal capacity of A/J and P/J LS (Boraschi and Meltzer 1979a; Boraschi and Meltzer 1979b; Boraschi and Meltzer 1979c). In this context we decided to study the variability in wild derived strains (WDS) of the house mouse, *Mus musculus.* In comparison to laboratory strains, WDS reflect reality observed in wild and offer higher level of natural polymorphism. It was also shown that WDS are often resistant while CLSs are more susceptible with defect immune response which is maintain generation after generation in laboratory sterile conditions (Guénet and Bonhomme 2003). The other advantage in comparison with CLSs is known and well documented origin of WDS and even if CLSs are often considered as Mmd, their genome is mosaic and mixture of other subspecies of *Mus musculus* (Yang et al. 2011). The interspecific crosses between WDS allow us to study for example speciation genes, because it brings together products of genes separated by divergent evolution and reveal deleterious gene combinations (Guénet and Bonhomme 2003). Good examples are genes responsible for hybrid sterility, because for example cross between *M. musculus* and *Mus spretus* results always in sterile males, while backcross of F1 hybrid female with male from parental species result in fertile males. In addition, homozygous genomes of WDS represents natural variants occurring in free-living populations, which can be very useful for separating heterozygous sequences sampled from natural populations (Guénet and Bonhomme 2003; Stephan et al. 2007; Piálek et al. 2008).

3.1.2 Material and Methods

We focused on five bacterial-sensing TLRs (TLR1, TLR2, TLR4, TLR5 and TLR6), which are important sensors of bacterial components such as LPS, LP, lipopeptides, teichoic acid and flagellin. They form dimers on the cellular surface (either heterodimers: TLR2/1,

TLR2/6, TLR4/MD-2 or homodimers: TLR4/TLR4, TLR5/TLR5) to recognize wider spectrum of molecules and trigger the pro-inflammatory signalling pathway. In this contribution we therefore focused on description of allelic polymorphism in selected TLRs in following strains of the house mice derived from wild populations of *M. m. musculus* (Mmm) (BULS, BUSNA, PWD, SLINT, STUF, STUP, STUS), *M. m. domesticus* (Mmd) (STLT, STRA, STRB, SCHEST, SCHUNT, SFEL, SIN, SIT, SPOS, STAIL, SUV, SCHEFE) and their natural hybrids (SPLY, SMIL) (Piálek et al. 2008; Vyskocilová et al. 2009) (Fig. 1, Table 1). To complete our data we added also three classical laboratory strains (CLSs) with overwhelming domesticus background: C57BL/6J, A/J and C3Ha. Assignement of WDS and CLS to subspecies were based on hybrid index (HI). HI was computed for all strains based on five X-linked loci (for more details see Ďureje et al. 2012). The scale of HI is set from minimum 0 to maximum 1, where 0 mean Mmd and 1 Mmm.

3.1.3 Laboratory techniques

We sequenced female and male of each WDS (including CLSs) to control for heterozygosis. DNA was extracted from rodent tissue (biopsy from ear or necropsy from liver) using the DNeasy Blood & Tissue Kit (Qiagen AB, Hilden, Germany). PCR primers and conditions as well as sequencing protocols can be found in Annex 1. For all genes we sequenced complete coding region (CDS) (Table 2, Fig. 2). Only successfully sequenced and complete sequences were analyzed. Standard Sanger sequencing of PCR products was performed using Big Dye Terminators chemistry (Applied Biosystems) and capillary electrophoresis was performed at ABI 3130 Genetic Analyzer (Applied Biosystems). Sequences were manually checked and aligned in SEQSCAPE v.2.5 (Applied Biosystems) and BIOEDIT v.7.1.3 (Hall 1999). We sequenced complete CDS of all five TLRs (incomplete sequences were not included in analysis). Description of sequenced regions can be found in Table 2.

3.1.4 Data analysis

Number of nucleotide haplotypes (*hN*) and amino acid variants (*hA*) for all *Tlr*s genes were enumerated using Fabox DNA collapser (Villesen 2007). Indices of genetic variability such as sequence polymorphism (π), average number of nucleotide differences (k), and number of polymorphic sites (S) were computed in DnaSP v.5.10 (Librado and Rozas 2009). For the analysis genetic variability, CLSs were excluded. Relationships of individual haplotypes based on nucleotide sequences were computed in Network v. 4.6.1.1. by median joining network (Bandelt et al. 1999). Visualization and description of main domains (ECD, TMD and ICD with TIR) was done in SMART (Schultz et al. 1998; Letunic et al. 2011). Position of LBR was set according to already published results for all five bacterial TLRs in human and/or in mice (Kang and Chae 2001; Mizel et al. 2003; Andersen-Nissen et al. 2005; Jin et al. 2007; Kim et al. 2007; Park et al. 2010; Ohto et al. 2012). List of published ligand binding sites and regions important for dimerization are described in Table 3.

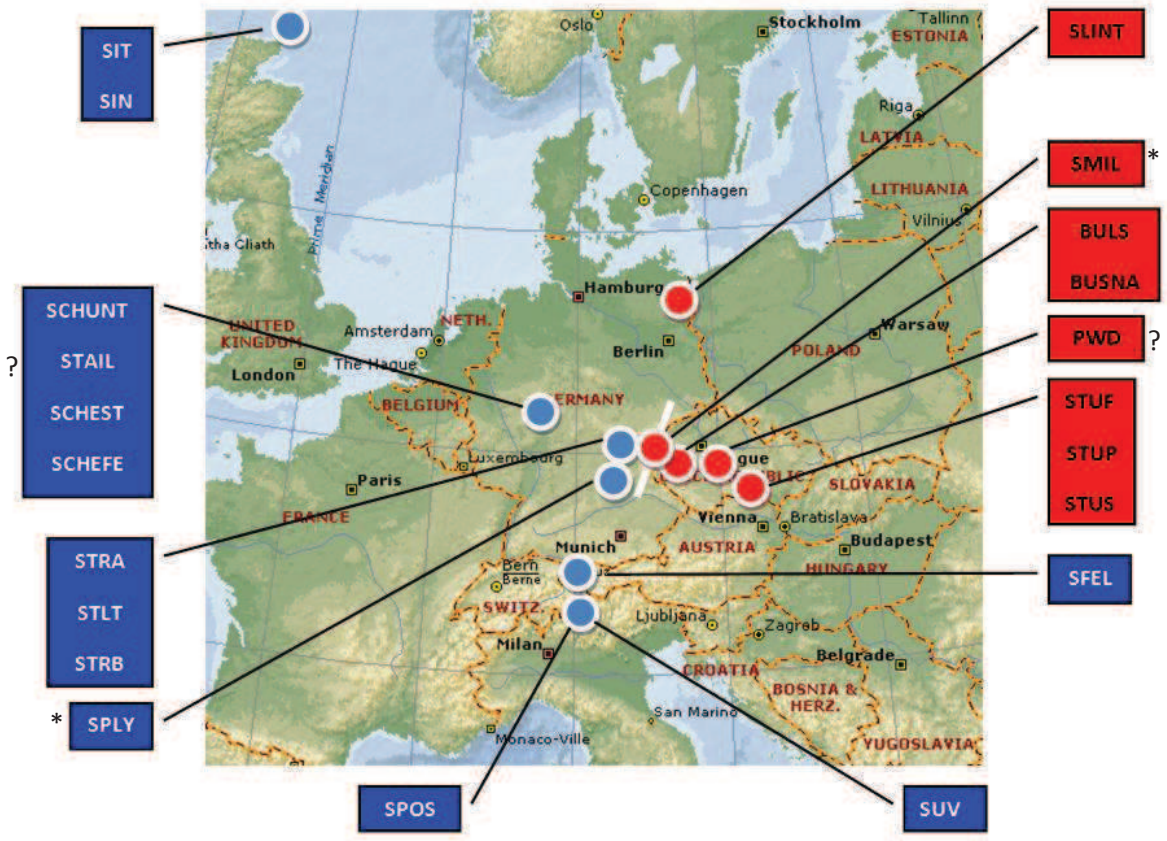


Fig. 1, Origin localities of wild-derived strains kept in the Research Facility Studenec. Strains derived from *Mus musculus domesticus* are in blue, those from *Mus musculus musculus* are in red. * SMIL and SPLY strains arised directly from the hybrid zone in western Czech Republic and Bavaria (approximate position marked by white line) as hybrids between both subspecies. Detailed geographic position of all localities can be found in Table 1. ?, indicates strains with potential introgression from other subspecies according to our results.

ID	Genome	Locality	Country	Sex	G	HI	Latitude	Longitude
STUF-WDS	Mmm	Studenec	CZ	F,M	31	1	49° 12' 00"	16° 04' 00"
STUP-WDS	Mmm	Studenec	CZ	F,M	30	1	49° 12' 00"	16° 04' 00"
STUS-WDS	Mmm	Studenec	CZ	F,M	22	1	49° 12' 00"	16° 04' 00"
BULS-WDS	Mmm	Buškovice	CZ	F,M	25	1	50° 13' 18"	13° 22' 27"
BUSNA-WDS	Mmm	Buškovice	CZ	F,M	30	1	50° 13' 18"	13° 22' 27"
PWD-WDS	Mmm	Kunratice	CZ	F,M	x+12	1	50° 00' 47"	14° 29' 07"
SMIL-hWDS	Mmm	Milhostov	CZ	F,M	3	1	50° 09' 33"	12° 27' 30"
SLINT-WDS	Mmm	Lindhorst	DE	F,M	1	1	53° 26' 34"	13° 46' 06"
STRA-WDS	Mmd	Straas	DE	F,M	29	0	50° 10' 53"	11° 45' 44"
STRB-WDS	Mmd	Straas	DE	F,M	26	0	50° 10' 53"	11° 45' 44"
STLT-WDS	Mmd	Straas	DE	F,M	19	0	50° 10' 53"	11° 45' 44"
SCHUNT-WDS	Mmd	Schweben	DE	F,M	7	0	50° 26' 00"	09° 35' 00"
SCHEST-WDS	Mmd	Schweben	DE	F,M	7	0	50° 26' 10"	09° 35' 10"
SIN-WDS	Mmd	Scar, Sanday Isl., Orkneys	UK	F,M	5	0	59° 18' 00"	-02° 33' 00"
SIT-WDS	Mmd	Scar, Sanday Isl., Orkneys	UK	F,M	6	0	59° 18' 00"	-02° 33' 00"
SFEL-WDS	Mmd	Feldkirch	AT	F,M	2	0	47° 15' 42"	09° 35' 10"
SPOS-WDS	Mmd	Migiondo	IT	F,M	3	0	46° 19' 23"	10° 18' 17"
SUV-WDS	Mmd	Sernio	IT	F,M	3	0	46° 13' 26"	10° 12' 20"
SCHEFE-WDS	Mmd	Schweben	DE	F,M	7	0	50° 26' 10"	09° 35' 10"
STAIL-WDS	Mmd	Schweben	DE	F,M	9	0	50° 26' 00"	09° 35' 00"
SPLY-hWDS	Mmd	Plössen	DE	F,M	2	0	49° 51' 18"	11° 47' 10"
C57BL/6J-CLS	Mmd	classical laboratory strain	lab	F,M	-	0.2	-	-
A/J-CLS	Mmd	classical laboratory strain	lab	F,M	-	0	-	-
C3Ha-CLS	Mmd	classical laboratory strain	lab	F,M	-	0	-	-

Table 1, Summary of strains, **WDS** - wild derived strain; **hWDS** - hybrid wild derived strain; **CLS** – classical laboratory strain, **ID** - identification of specimens; **Genome** – classification of mice based on hybrid index (for details see Methods) to either subspecies; **Mmd** - *Mus musculus domesticus*; **Mmm** - *Mus musculus musculus*; **Country codes**: **AT** - Austria, **CZ** - Czech Republic, **DE** - Germany, **IT** - Italy, **SK** - Slovakia, **UK** - United Kingdom; **G** – number of generation under brother x sister mating, G0 stands for wild mice, x+ indicates that the strain was kept for unknown number (x) of generations in different laboratory; **HI** - hybrid index based on five X-linked loci (Ďureje et al. 2012), 0 - pure Mmd, 1-pure Mmm.

		TLR1*		**TLR2***		**TLR4**		**TLR5**		**TLR6***	
total lenght	*(bp)*	7866		5337		14987		21257		6940	
mRNA	*(bp)*	2567		2874		3847		3211		2600	
cds	*(bp)*	2388		2355		2508		2622		2421	
cds on exons	*(bp)*	3rd		3rd		1st(90), 2nd (167) and 3rd (2251)		3rd (38) and 4th (2584)		2nd	
protein	*(AA)*	795	H/A	784	H/A	835	H/A	873	H/A	806	H/A
STUF-WDS	Mmm	comp	1/1	comp	1/1	comp	1/1	comp	1/1	comp	1/1
STUP-WDS	Mmm	comp	2/1	comp	2/2	comp	2/2	comp	1/1	comp	2/2
STUS-WDS	Mmm	comp	1/1	comp	3/3	comp	3/1	comp	1/1	comp	1/1
BULS-WDS	Mmm	comp	3/2	incomp	-	comp	4/2	comp	2/2	comp	3/1
BUSNA-WDS	Mmm	comp	3/2	comp	1/1	comp	4/2	comp	3/3	comp	3/1
PWD-h?WDS	Mmm	comp	4/3	incomp	-	comp	4/2	comp	4/4	comp	2/2
SMIL-hWDS	Mmm	comp	1/1	comp	4/4	comp	4/2	comp	5/5	comp	1/1
SLINT-WDS	Mmm	comp	1/1	incomp	-	comp	5/3	comp	6/1	comp	1-3/1
STRA-WDS	Mmd	comp	5/4	comp	4/4	comp	6/4	comp	5/5	comp	4/3
STRB-WDS	Mmd	comp	6/5	comp	5/5	comp	6/4	comp	7/4	comp	5/4
STLT-WDS	Mmd	comp	5/4	comp	4/4	comp	6/4	comp	5/5	comp	4/3
SCHUNT-WDS	Mmd	comp	7/6	comp	6/4	comp	6/4	comp	8/4	comp	6/5
SCHEST-WDS	Mmd	comp	5/4	comp	6/4	comp	6/4	comp	8/4	comp	4/3
SIN-WDS	Mmd	comp	7/6	comp	6/4	comp	6/4	comp	7/4	comp	7/5
SIT-WDS	Mmd	comp	7/6	comp	7/6	comp	6/4	comp	7/4	comp	7/5
SFEL-WDS	Mmd	comp	5/4	comp	6/4	comp	6/4	comp	9/5	comp	4/3
SPOS-WDS	Mmd	comp	5/4	comp	6/4	comp	6/4	comp	7/4	comp	4/3
SUV-WDS	Mmd	comp	5/4	comp	6/4	comp	6/4	comp	7/4	comp	8/3
SCHEFE-WDS	Mmd	comp	5/4	comp	4/4	comp	6/4	comp	8/4	comp	4/3
STAIL-h?WDS	Mmd	comp	7/6	comp	8/7	comp	6/4	comp	7/4	comp	6/5
SPLY-WDS	Mmd	comp	8/4	comp	9/8	comp	6/4	comp	7/4	comp	5/4
C57BL/6J-CLS	Mmd	comp	6/5	comp	10/8	comp	7/5	comp	10/6	comp	5/4
A/J-CLS	Mmd	comp	6/5	comp	11/6	comp	8/6	comp	4/4	comp	5/4
C3Ha-CLS	Mmd	comp	8/4	comp	11/6	comp	9/7	comp	11/7	comp	5/4

Table 2, Overview of sequenced TLRs, cds, coding region, bp, base pairs, com, complete cds, incomp, incomplete cds, not included in analysis, H, assignment to nucleotide haplotypes, A, assignment to amino acid alleles, 1-3, indicate heterozygote haplotypes (concerning SLINT and its TLR6), WDS, wild derived strains, hWDS, hybrid wild derived strain,?, indicates potential hybrid strains according to our results, CLS, classical laboratory strain. Numbers of nucleotide haplotypes (H) correspond to haplotype networks on Fig. 3 and

numbers of amino acids alleles (A) correspond to overview of polymorphic sites in Fig. 2. *, TLR with * were sequenced by Z. Bainová, as well as first and second exon of TLR4 and third exon of TLR5.

TLRs	Ligand binding region or sites	Dimerization region or sites	Publications
TLR1h	**258-399 (F312, G313, Q316)**	**310-385**	(Jin et al. 2007)
TLR2m	**266-355 (D327, F325, F349)**	**318-404 (E375)**	(Jin et al. 2007; Kang et al. 2009)
TLR4m	**248-469 (K263, R434, F438, F461)**	**D41, D83, E134, H158, R233, R288**	(Kim et al. 2007; Park et al. 2009; Ohto et al. 2012)
TLR5h	**174-401 or 386-407* (in mouse 386-406)***	**-**	(Mizel et al. 2003; Andersen-Nissen et al. 2005; Smith et al. 2012)
TLR6m	**F317, L318,F319, P342, V347**	**311-390 (K313, D340)**	(Kang et al. 2009)

Table 3, Overview of sites and regions important for ligand binding and dimerization in mouse (m) or human (h). Sites in brackets signify important sites for ligand binding or dimerization.

3.1.5 Results

Number of nucleotide haplotypes of all *Tlr*s, except *Tlr4*, varied from 3 to 5 in both subspecies and almost the same pattern appeared when we translated sequences to proteins (*hA* = 2-4). The only exception was *Tlr4*, where we found 5 (3) nucleotide (protein) sequences in strains derived from Mmm, but all 13 Mmd strains had identical *Tlr4* sequence (Table 2 and 4). Nucleotide diversity varied from 0 ($Tlr4_{\mathrm{Mmd}}$) to 0.002 ($Tlr4_{\mathrm{Mmm}}$). If this one outlier ($Tlr4_{\mathrm{Mmd}}$) was excluded, the polymorphism of the remaining *Tlr*s was comparable among subspecies (Table 4). Hybrids and deviants according to Fig. 3 were excluded from these analyses (e.g. PWD for *Tlr1*, SMIL and STAIL for *Tlr2* and SMIL, BULS - strain with STOP codon and PWD for *Tlr5*).

All but one nucleotide sequences translated to functional proteins. The only exception was the gene for *Tlr5* in the strain BULS (Mmm, Buškovice, CZ), where the deletion of one nucleotide at position 19 in exon 3 (length of CDS in exon 3 is 38bp) resulted in the stop codon. Important polymorphism was documented also at protein level. In total we revealed 59 non-synonymous SNPs in all five TLRs. Most of these SNPs (76%) were located in the ECD. In ICD we found 22% of SNPs and in TMD there was only 1 SNP in TLR4 (Fig. 2). Five SNPs were placed in LBRs and seven in TIR domains (Fig. 2). Amino acid substitutions in LBRs were in two cases rather functionally neutral, because they shared same biochemical properties (TLR1, S324N and TLR4 V254I, see Table 5). Three others amino acid substitutions differed in their properties (TLR4 R340S, TLR4 D462N and TLR5 D240G), what might indicate the potential functional implications, i.e. differences in the binding of ligands (Table 5).

	TLR1		TLR2		TLR4		TLR5		TLR6	
	Mmd	Mmm*	Mmd*	Mmm*	Mmd	Mmm	Mmd	Mmm*	Mmd	Mmm
hN	4	3	5	3	**1**	5	4	3	5	3
hA	3	2	4	3	**1**	3	4	2	3	2
π±S.D.	0.0018 ±0.0004	0.0008 ±0.0003	0.0006 ±0.0002	0.0006 ±0.0002	**0**	0.0020 ±0.0006	0.0008 ±0.0001	0.0013 ±0.0007	0.0017 ±0.0002	0.0009 ±0.0003
k	4.256	1.810	1.455	1.500	**0**	5.286	2.205	3.400	4.000	2.111
S	13	5	5	3	**0**	12	5	8	10	5
Hd±S.D.	0.654 ±0.106	0.667 ±0.139	0.727 ±0.113	0.833 ±0.222	**1**	0.786 ±0.151	0.679 ±0.112	0.700 ±0.218	0.769 ±0.099	0.722 ±0.097

Table 4, Genetic diversity of *Tlr*s in two subspecies of the house mouse; 13 Mmd, *Mus musculus domesticus*; 8 Mmm, *Mus musculus musculus*; *hN*, number of nucleotide haplotypes; *hA*, number of amino acid variants; *π*, nucleotide diversity; *k*, average number of nucleotide differences; *S*, number of polymorphic sites; *Hd*, haplotype diversity; *S.D.,* standard deviation. Analysis of variability does not include classical laboratory strains, * Hybrid strains were excluded from these analyses (e.g. PWD for *Tlr1*, SMIL and STAIL for *Tlr2* and SMIL, BULS - strain with STOP codon and PWD for *Tlr5*). For details see Fig. 3.

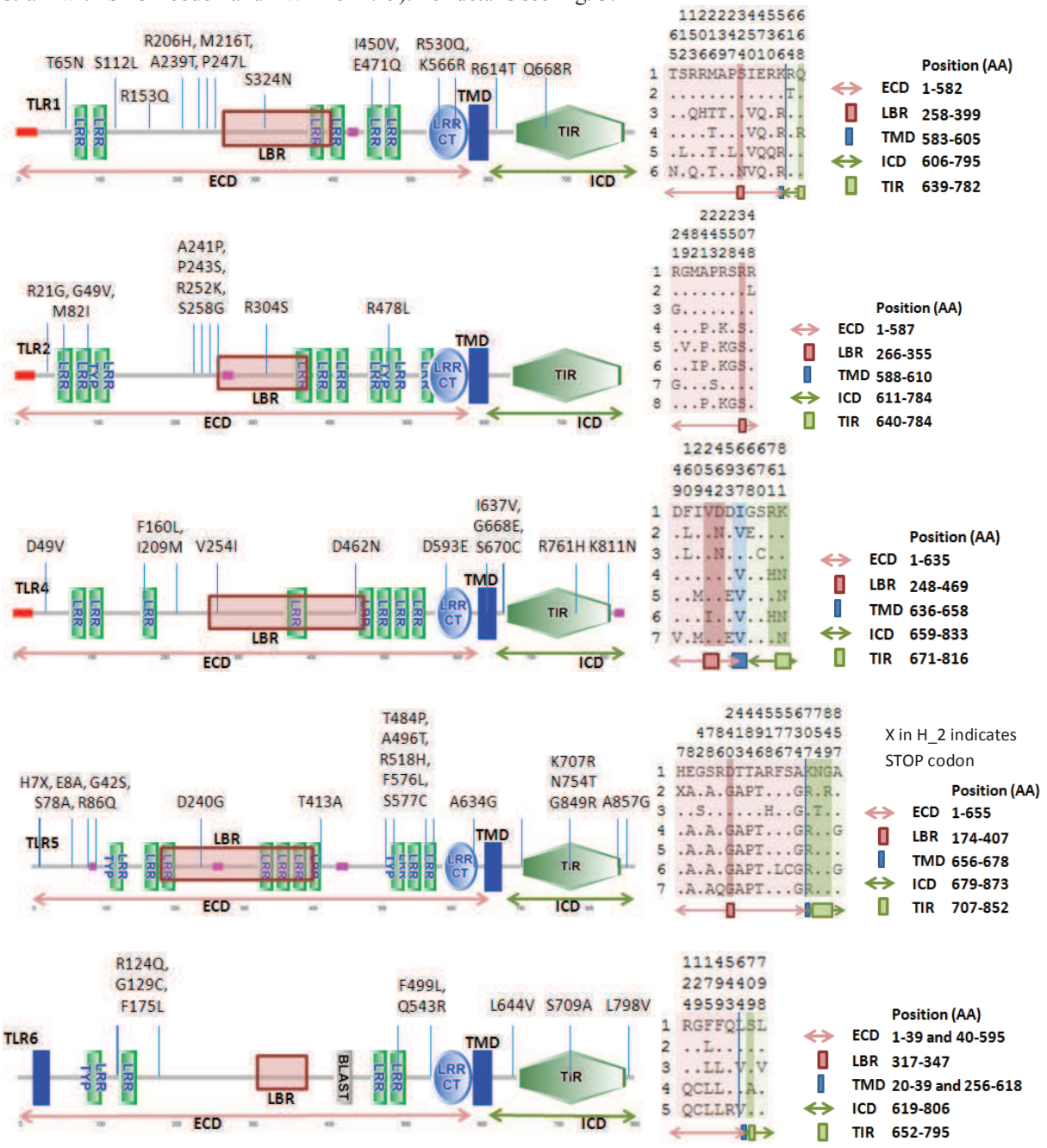


Fig. 2, Description of non-synonymous polymorphic sites (SNPs) and their position within TLRs. Domain architecture analysis was performed in SMART (http://smart.embl-heidelberg.de/). Haplotypes numbers in tables correspond to amino acid haplotypes (A) described in Table 2.

TLR and position	aa1	Properties	aa2	Properties
TLR1-342	S, Ser	SM, P, NEU	N, Asn	SM, P, NEU
TLR2-304	R, Arg	P, **POS**	S, Ser	SM, P, **NEU**
TLR4-254	V, Val	NP, NEU	I, Ile	NP, NEU
TLR4-462	D, Asp	SM, P, **NEG**	N, Asn	SM, P, **NEU**
TLR5-240	D, Asp	SM, **P, NEG**	G, Gly	SM, **NP, NEU**

Table 5, Physicochemical properties of the amino acids involved in non-synonymous substitutions in ligand binding regions; SM, small; NP, non-polar; P, polar; NEU, neutral; POS, positively charged; NEG, negatively charged.

Haplotype network analysis separated in most cases Mmd and Mmm in all studied *Tlr*s. CLS were usually clustered with Mmd, confirming the previous data showing their predominantly Mmd origin (Piálek et al. 2008). However, there are few interesting exceptions. In *Tlr1* Mmm strain PWD had the haplotype H_4, clustering inside Mmd haplogroup. In *Tlr2* hybrid strain SMIL (H_4) clustered with Mmd. Mmd strain STAIL (H_8) clustered by contrast with Mmm. In *Tlr5* haplotype hybrid strain SMIL (H_5) again joined same strains of Mmd like in the case of *Tlr2* and PWD (H_4) strain clusters with Mmd and had the same haplotype as CLS.

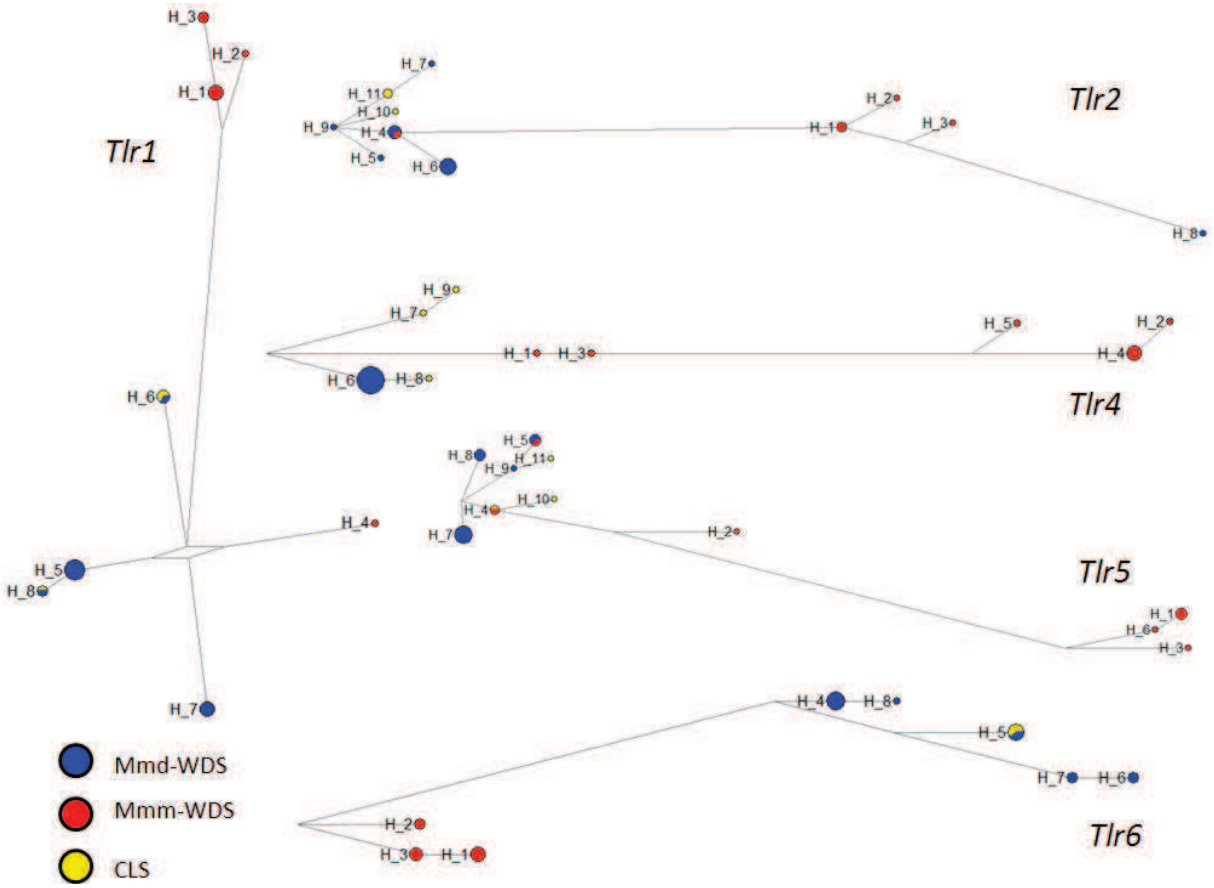


Fig. 3, Haplotype network based on nucleic acids (H), assignment of individual haplotypes to wild-derived strains (WDS) or classical laboratory strain (CLS) can be found in Table 2. WDS of Mmm, *Mus musculus*

musculus, are represented by red circles, WDS of Mmd, *Mus musculus domesticus* are represented by blue circles, CLS, classical laboratory strains are represented by yellow circles.

3.1.6 Discussion

Overall our data revealed unexpectedly high allelic variability among inbred strains derived from wild populations (i.e. WDS). Therefore we suggest that WDS are convenient source of polymorphism and potentially useful biomedical model for subsequent studies. In CLS the variability is also maintained, however, they differ from WDS analyzed in this study. Variants found in CLS are often unique (e.g. see haplotype network of *Tlr2*, *Tlr4* and *Tlr5,* Fig. 3), and even if they always cluster with Mmd strains, it is not confirmed that they really represent natural variation (and not the results of artificial selection in captivity). Predominant domesticus origin of CLS is not surprising and is was well documented by previous studies (Beck et al. 2000; Smirnova et al. 2000; Yang et al. 2011).

In detailed analyses we found that two strains SMIL and PWD, which are predominantly Mmm strains has parts of the genome introgressed from Mmd and in contrast one strain of Mmd STAIL has parts of the genome introgressed from Mmm. Strain SMIL has its origin in the hybrid zone and therefore it is not surprising. Thorough genomic exploration of PWD strain brought up that this strain is on autosomes formed originally by 93% of Mmm and 7% of Mmd genomes (Yang et al. 2011). When we checked the presumable position of *Tlr5* (chromosome 1, 182954788..182976044) in Mouse Phylogeny Viewer (http://msub.csbio.unc.edu/), we found that there is a massive introgression from Mmd into Mmm (Wang et al. 2012). However, similar approach used in PWD did not clarify observed introgression from Mmd in *Tlr1* (chromosome 5, 64924834...64932699). Last mentioned strain STAIL is a harder nut, because even if it has the same origin as the Mmd strain SCHUNT, introgression from Mmm showed up in *Tlr2* (but not in SCHUNT). Therefore next analysis of indigenous specimens from which this strain originated, is necessary.

Other interesting result appeared when we analyzed *Tlr5*. In its exon 3 we found stop codon in the strain BULS. Interestingly, the stop codon at the same place was found also in the strain SMON, derived in the Research Facility Studenec from *Mus spretus* population sampled in Montferrier-sur-Lez, France (Z. Bainová et al., unpublished results). Another WDS from the hybrid zone, SOTT (Ottmannsreuth, Gremany), which has been added to the collection of WDS recently (and also is not analyzed in this part of dissertation) was

heterozygous for the stop codon at the same position of the gene. These three findings indicate relaxed selective pressure on *Tlr5* and further experimental research on the function of *Tlr*5, i.e. presumably flagellin detection, will be required to assess how much this protein is necessary in house mouse innate immunity. TLR5 pseudogenes were already described in human (Hawn et al. 2003), birds (Alcaide and Edwards 2011, Králová et al., unpublished data) or in domestic livestock (Smith et al. 2012). There are two possible explanations of frequent functional failures of *Tlr5*. The first suppose that flagellated bacteria do not generate strong selective pressure to host immune system. However, it was experimentaly shown that *Tlr5*$^{-/-}$ mice were more susceptible to *Escherichia coli* infection (Andersen-Nissen et al. 2007), so this possibility is less probable. Alternative explanation expexts redundant function of TLR5 with other PRRs, for example NOD-like receptor family CARD domain-containing 4 (Franchi et al. 2006).

We have found that variability of TLR4 (which is principally responsible for LPS detection) is drastically reduced in WDS of Mmd. All 13 WDS derived from Mmd harboured completely same haplotype and even synonymous substitutions were not found. In the next step, it should be tested whether the reduced variability of Mmd-WDS is not only artefact of inbreeding in captivity and ascertainment bias.

In contrast, CLS preserved higher level of polymorphism of TLR4. Each of three analysed CLS had its own haplotype translated into unique protein variants. However if we can observe this pattern also in wild another question concerning artificial variability of CLS arise. Specific haplotypes of CLS were found and discussed also in study of Smirnova et al. (2000). Since several studies have reported defective tumoricidal capacity of A/J and P/J CLS (Boraschi and Meltzer 1979a; Boraschi and Meltzer 1979b; Boraschi and Meltzer 1979c), we suggest that observed differences between haplotypes from wild and laboratories are caused probably by relaxed selection in captivity, rather than by insufficient sampling of free-living populations.

3.2 Analysis of variability at intraspecific level in wild populations of *Mus musculus*

Annotation

TLR4 is one of the most important TLRs and because of its crucial function in sensing of bacterial infections it is recently intensively studied also in non-model organisms. In the previous chapter, I showed that variability of TLR4 is drastically reduced in strains derived from *M. m. domesticus* (Mmd) in comparison to those from *M. m. musculus* (Mmm). However, this can be theoretically an artefact of intensive random drift and inbreeding during development of wild-derived strains. Here, in the second chapter of the dissertation results, I analyze the variability of *Tlr4* gene in the house mice from free-living populations collected in large area of Western Palaearctic. The results suggest that different evolutionary forces have influenced the recent variation of *Tlr4* in two subspecies of the house mouse. In Mmd, the functional parts of TLR4 responsible for recognition of ligands are monomorphic in whole studied area, while functional polymorphism is relatively well maintained in Mmm.

CONTRASTING PATTERNS OF POLYMORPHISM AND SELECTION IN BACTERIAL-SENSING TOLL-LIKE RECEPTOR 4 IN TWO HOUSE MOUSE SUBSPECIES

Fornuskova, A.[1,2,3]*, Bryja, J.[1,2], Vinkler M.[4], Macholán M. [5], Pialek, J [1]

[1] Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, Brno, Czech Republic; [2] Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic; [3] Centre de Biologie pour la Gestion des Populations (CBGP), Institut National de la Recherche Agronomique (INRA), Campus International de Baillarguet, Montferrier-sur-Lez, France; [4] Department of Zoology, Faculty of Science, Charles University in Prague, Czech Republic, [5]Laboratory of Mammalian Evolutionary Genetics, Institute of Animal Physiology and Genetics, Academy of Sciences of the Czech Republic, Brno, Czech Republic

* Corresponding author: Alena Fornůsková, Research Facility Studenec; Institute of Vertebrate Biology, Academy of Sciences of the Czech Republic, 675 02 Studenec 122, Czech Republic; e-mail: fornuskova@ivb.cz

First submission Journal of Evolutionary Biology, November 2013
Resubmitted in Ecology and Evolution, December 2013

Abstract

Detailed investigation of variation in genes involved in pathogen recognition is crucial for understanding co-evolutionary processes between parasites and their hosts. Triggering immediate innate response to invading microbes, Toll-like receptors (TLRs) belong presently among the best studied receptors of vertebrate immunity. TLRs exhibit remarkable interspecific variation and also intraspecific polymorphism is well documented. In humans and laboratory mice several studies have recently shown that single amino acid substitution may significantly alter receptor function. Unfortunately, data concerning polymorphism in free-living species are still surprisingly scarce. In this study we analyzed the polymorphism of Toll-like receptor 4 (*Tlr4*) over the Palaearctic range of house mouse (*Mus musculus*). Our results reveal contrasting evolutionary patterns between the two recently (0.5mya) diverged house mouse subspecies: *M. m. domesticus* (Mmd) and *M. m. musculus* (Mmm). Comparison with cytochrome *b* indicates strong directional selection in Mmd *Tlr4*. Throughout the whole Mmd Western Palaearctic region a single variant of the ligand binding region is spread, encoded mainly by one dominant haplotype (71% of Mmd). In contrast, *Tlr4* in Mmm is much more polymorphic with several haplotypes at intermediate frequencies. Moreover, we also found clear signals of recombination between two principal haplogroups in Mmm. Though on intrasubspecific level we did not detect any signal of positive selection, we identified sites under positive selection when comparing between the two subspecies. Our

results suggest that observed differences in *Tlr4* diversity may be attributed to contrasting parasite-mediated selection acting in the two subspecies.

Key words: Arms race, host-pathogen interaction, parasite-mediated selection, pattern recognition receptors, PRRs, adaptive evolution, directional selection, MAMPs, *Mus musculus*

3.2.1 Introduction

Selective forces imposed by parasites can affect various traits of their hosts, including population dynamics, life histories, mating systems, sexual dimorphism etc. (Schmid-Hempel 2011). The detrimental effects of parasites are countered by function of immune system which in vertebrates comprises both innate and acquired immunity (Danilova 2006). Study of evolution in immune-related genes is, therefore, of paramount importance for comprehension of dynamics in parasite-host relationships (see e.g. Woolhouse et al. 2002; Carlton 2003). Despite the complexity of the immune system, most studies in free-living vertebrates have focused on genes involved in acquired immunity, namely the major histocompatibility complex (MHC) (e.g. Milinski 2006). However, mapping and association studies have revealed that at least half of the genetic variation responsible for resistance to various infections is attributable to non-MHC genes (Acevedo-Whitehouse and Cunningham 2006). Most of these genes seem to be associated with innate immunity and there is an increasing evidence that variation in these genes may have a fundamental effect on the host fitness in free-living populations (e.g. Turner et al. 2011; Tschirren et al. 2013).

Innate immunity receptors that directly detect and bind to parasite structures (pathogen-associated molecular patterns, PAMP), the pattern-recognition receptors (PRR), stand in the first line of immune defence (Medzhitov & Janeway 2002; Akira et al. 2006). Their fast and effective functioning is thus crucial for host survival (O'Neill 2004; Akira et al. 2006). Among PRRs, the Toll-like receptors (TLR) have been shown to be particularly important (Akira et al. 2001). These receptors form a group of membrane-bound, non-catalytic proteins present in most immune cells, especially in macrophages. Distinct PAMPs (e.g. lipopolysaccharides and lipoproteins in bacterial cell walls, zymosan of yeast, bacterial

flagellin or viral nucleic acids) are recognized by distinct TLRs and the set of TLR types varies substantially among vertebrate lineages (Janssens and Beyaert, 2003; Akira et al. 2006; Vinkler and Albrecht 2009; Kawai and Akira 2010). The potential action of TLRs in the context of host-parasite interactions in free-living organisms is increasingly drawing attention of evolutionary biologists and immunologists (Medzhitov et al. 1997; Pasare and Medzhitov 2004; Takeda and Akira 2005; Vinkler and Albrecht 2009). Contradicting the previous assumption of evolutionary conservatism of these receptors, evolution-focused immunogenetic investigations yielded a clear evidence that at the interspecific level diversifying selection has significantly increased diversity of orthologous *Tlr* genes, mainly in the ligand-binding region (LBR, Poltorak et al. 1998; Smirnova et al. 2000; Downing et al. 2010; Park et al. 2010; Wlasiuk and Nachman, 2010; Areal et al. 2011; Tschirren et al. 2011; Fornuskova et al. 2013).

Information regarding the structure and variation of TLRs in free-living rodents is still relatively scarce. Interspecific comparisons of European and Asian rodents confirmed purifying selection as a prevalent evolutionary force shaping these genes (namely *Tlr2, 4,* and *7*), probably due to functional constraints posing on the receptor molecules (Tschirren et al. 2011; Fornuskova et al. 2013). However, signatures of positive selection have also been revealed in all studied genes (mainly in the extracellular domain, ECD, containing LBRs responsible for pathogen recognition, see below), with a more intense signal in the bacterial-sensing *Tlr2* and *Tlr4* than in the viral-sensing *Tlr7* gene (Tschirren et al. 2011; Fornuskova et al. 2013). Following study of Tschirren et al. (2012) showed that TLRs are polymorphic even within species and that intraspecific variation may strikingly differ even between two sympatric species of rodents inhabiting the same environment. In one of these species, the bank vole (*Myodes glareolus*), a particular group of alleles was shown to be significantly associated with resistance to *Borrelia* infection, suggesting an on-going evolution in the receptor (Tschirren et al. 2013). These results illustrate the urgent need for further research focused on polymorphism in PRRs at the intraspecific level, since the genetic variability in PRRs might represent an important missing element for understanding the effects of a host genotype on individual fitness.

TLR4 is one of the most essential bacterial-sensing PRRs, binding, among others, bacterial endotoxins (i.e. lipopolysaccharides, LPS) as ligands (Poltorak et al. 1998). At the interspecific level, this cell-surface receptor has the highest number of positively selected sites among all mammalian TLRs (Areal et al. 2011). Most of these sites are localized in the ECD which is responsible for LPS binding (Poltorak et al. 1998; Kim et al. 2007; Vinkler et al.

2009; Fornuskova et al. 2013). This domain consists of several leucine-rich repeat motifs and includes the LBR which is in direct physical contact with PAMP structures. The ECD is followed by the transmembrane domain (TMD), anchoring the receptor into the cell membrane, and the intracellular domain (ICD). The ICD comprises the Toll/interleukin-1 (TIR) domain responsible for signal transduction and cell activation triggering the immune responses (Werling et al. 2009; Botos et al. 2011).

As a key model in biomedical and evolutionary research, the house mouse (*Mus musculus*) yields wealth of data compiled over decades of intensive research (Fox et al. 2000; Macholán et al. 2012). The importance of this animal as a model species is underscored by the availability of a vast array of laboratory strains. Genetic research in laboratory mice enabled identification of the *Tlr4* gene function and assessment of the level of its polymorphism among strains (Poltorak et al. 1998; Smirnova et al. 2000; Stephan et al. 2007). However, artificial genetic variation occurring in "classical" laboratory strains (CLS) (Yang et al. 2011) hampers understanding of variation present in wild mice displaying much wider ranges of immunoresponsivity (Piálek et al. 2008; Abolins et al. 2011; Babayan et al. 2011; Pedersen and Babayan 2011; Riley and Viney 2011). Our own study of five bacterial-sensing TLRs in 13 strains derived from one of the house mouse subspecies, *Mus musculus domesticus* (see below), from a broad geographic region (wild-derived strains, WDS) revealed drastic reduction in *Tlr4* variation (we found only a single haplotype by sequencing a complete coding region; A. Fornůsková et al., unpublished data), while polymorphism among CLSs and *M. m. musculus* WDSs was well-maintained. Moreover, all alleles found in CLSs differed from the single allele observed in the *M. m. domesticus*-WDSs. These data show striking difference between CLSs and WDSs. By studying natural variation of the *Tlr4* gene in free-living house mice we aimed to test whether the pattern we observed was an artefact arising from controlled laboratory breeding or if it reflects polymorphism in the wild.

Several house mouse subspecies have been described. Divergence of house mice is usually located to northern India and/or Pakistan and dated to about 0.5 million years ago (Boursot et al. 1993; Suzuki et al. 2004; Geraldes et al. 2008; Macholán et al. 2012). Two subspecies, *M. m. musculus* (Mmm) and *M. m. domesticus* (Mmd), have colonized Europe where they met along a secondary hybrid zone running across the continent (Boursot et al. 1993; Macholán et al. 2007; Bonhomme et al. 2011; Cucchi et al. 2012, 2013). Although the two subspecies might come into contact at least ones during the expansions and contractions of their ranges (Duvaux et al. 2011), allowing them to exchange beneficial mutations, they remained for most of the colonization time in allopatry. Since their westward expansions

followed different routes (Mmm north of the Black Sea, Mmd through the Middle East and Mediterranean region), the two subspecies may have been exposed to diverse pathogenic environments leaving distinct genetic footprints in PRR genes, including *Tlr4*. A recent study of the gastrointestinal tract microbiota in western European mouse populations showed geography to be the most significant factor explaining the composition of bacterial communities in this species (Linnenbrink et al. 2013). Even though gastrointestinal bacteria may have not necessarily been the pathogenic agents selecting for immunological divergence in the two subspecies, we may expect similar geographic or subspecies-specific variation also among other pathogens. Genetic differences between non-bacterial parasites of the two house mouse subspecies and the lack of their significant introgression in the hybrid zone have been described recently (Kváč et al. 2013).

In this study we have analyzed free-living specimens of the two European *Mus musculus* subspecies across a wide geographic range to answer the question if the distinct recent evolutionary histories of the subspecies have left any footprints in *Tlr4* variation. Based on preliminary data from CLS and WDSs, we expect significant differences between two house mouse subspecies. These potentially contrasting patterns could be explained either by different selection forces mediated by pathogens or simply by differences in demographic histories of the taxa (e.g. population expansions or bottlenecks). Given scarcity of data on pathogen background in the sampled regions we tested the two plausible explanations by analysing also the mitochondrial cytochrome b (*mt-Cytb*) gene widely used as a selectively neutral marker for assessing demographic histories of species and populations. Whereas similar patterns observed in both *mt-Cytb* and *Tlr4* would support the effect of demographic changes, distinct patters in the two genes would suggest the effect of selection on *Tlr4*. By genotyping *Tlr4* and *mt-Cytb* in 44 Mmd and 30 Mmm sampled across the Western Palaearctic region we document (1) a subspecies-specific distribution of genetic variation, (2) different selection patterns operating on *Tlr4* gene in the two subspecies, and (3) important role of recombination increasing the polymorphism of the *Tlr4* gene.

3.2.2 Materials and methods

Sampling

We sampled 28 and 42 populations (1-2 individuals per site) of free-living Mmm and Mmd, respectively, scattered across the Western Palaearctic region (with exception of two localities from central Asia; Fig. 1, Table S1). In addition, we included also mice of three CLSs of predominantly Mmd origin (C3Ha, A/J, C57BL/6J; see Yang et al. (2011) for their genomic composition), 15 WDSs of the Mmd origin, and nine WDSs of the Mmm origin (Piálek et al. 2008; Vyskocilová et al. 2009; for the origin of WDSs, see Table S1). In comparison with laboratory strains WDSs encompass more natural polymorphism and, at the same time, the homozygote variants are useful for distinguishing heterozygote sequences of natural populations (Guénet and Bonhomme, 2003; Stephan et al. 2007; Piálek et al. 2008).

Assignment of specimens to subspecies

Assigning each analyzed individual to one of the two subspecies was based on the combination of X-linked and mtDNA diagnostic markers proven to display low levels of introgression across the European house mouse hybrid zone (Ďureje et al. 2012). The first set of markers consisted of five X-linked SINE and/or LINE insertions chosen to be distributed along the whole chromosome: *X332*, *X65*, *X347*, *Btk*, and *Syap1* (Macholán et al. 2011). For each individual, a hybrid index (HI) was calculated as the mean frequency of Mmm-specific alleles over all five loci (10 for a female and 5 for a male). While the majority of mice displayed HI = 1.00 (Mmm) or HI = 0.00 (Mmd), 13 individuals were not fixed for all Mmm or all Mmd alleles (Table S1). This may be due to introgression and/or incomplete lineage sorting of the X-linked markers. Note that also C57BL/6, i.e. one of the most "classical" laboratory strains of predominately Mmd origin (Yang et al. 2011), harbours Mmm alleles (Table S1). However, regardless underlying causes, in all these cases admixture was negligible, allowing reliable subspecific identification.

The mitochondrial marker was a *Bam*HI restriction site in the *Nd1* gene shown to discriminate between the subspecies (Božíková et al. 2005). Mice were assigned to Mmm when the site was absent and to Mmd when the site was present. All 62 mice (wild, CLS, and WDS) assigned on the basis of the HI to Mmd also carried the *Bam*HI restriction site. Of the

remaining 39 mice assigned to Mmm according to the HI, three individuals (two wild individuals from Lindhorst and Lauchhammer in Germany, and one from a still not fully inbredized WDS established from Lindhorst) carried the Mmd-specific restriction site, suggesting introgression of Mmd mtDNA across the hybrid zone into Mmm range (Table S1). These three specimens (SLINT-WDS, SK843 and SK837) were analyzed as Mmm in the *Tlr4* dataset and as Mmd in the mtDNA dataset (see below for details).

Genetic variation within subspecies

In total 101 specimens (free-living mice together with WDSs and CLSs) were successfully sequenced for both *Tlr4* and *mt-Cytb* genes. We sequenced exon 3 of *Tlr4* (2,244 bp), encompassing 90% (748 of 835 amino acid residues) of the gene coding sequence, following the protocol described in Fornuskova et al. (2013). Almost complete *mt-Cytb* (1,123 bp) was sequenced after amplification by universal primers L14724 and H15915 (Lecompte et al. 2002). Sequences were manually checked and aligned using SEQSCAPE v.2.5 (Applied Biosystems) and BIOEDIT v.7.1.3 (Hall 1999).

Individual *Tlr4* alleles (thereafter called haplotypes for simplicity of comparison with *mt-Cytb*) were reconstructed from the complete alignment using the Bayesian PHASE routine implemented in DnaSP v.5.10 (Stephens and Donnelly 2003; Librado and Rozas 2009). This analysis was carried out using 1,000 iterations, 10 thinning intervals and 1,000 burn-in iterations. Four heterozygous *Tlr4* sequences resolved with low support were checked by cloning using the protocol of pGEM®-T Easy Vector System II (Promega). Initially, two clones from each individual were sequenced and this number was later increased until we obtained both sequences of each heterozygote (identification of the four cloned cases can be found in Table S1). Positions of TLR4 domains (ECD, TMD, ICD/TIR) were determined using the on-line program SMART according to Fornuskova et al. (2013). Amino acids were numbered according to a GenBank *M. musculus* TLR4 protein sequence (GenBank Number: AGA16686.1).

The numbers of nucleotide haplotypes (N) and amino acid variants (A) for both *Tlr4* and *mt-Cytb* genes were estimated using Fabox DNA collapser (Villesen 2007). Nucleotide diversity (π), average number of nucleotide differences (k), number of polymorphic sites (S) and haplotype diversity (Hd) were computed in DnaSP v.5.10. Haplotypes were assigned to haplogroups (HG) based on their phylogenetic interrelationships inferred with MrBayes v 3.1 (Huelsenbeck and Ronquist 2001) and according to a median joining network constructed

with Network v. 4.6.1.1. (Bandelt et al. 1999). The HKY+Γ (Hasegawa et al. 1985) and GTR+Γ (Tavaré 1986) models, determined using jModelTest 0.1.1. (Posada 2008), were applied to *Tlr4* and *mt-Cytb* data, respectively. For both genes we ran 10,000,000 MCMC generations of which 2,500,000 generations were discarded as burn-in. Geographical distribution of the haplogroups was projected onto a map using the PanMap software (http://www.pangaea.de/software/PanMap). All these computations were based on a subset of wild and WDS mice (i.e. we excluded all sequences from CLS).

Analysis of molecular evolution of Tlr4

For detection of recombination breakpoints in the *Tlr4* gene we used two algorithms, the single breakpoint recombination (SBP) and genetic algorithm recombination detection (GARD), provided on the DataMonkey web server (Pond and Frost 2005; Pond et al. 2006a; b; c; Delport et al. 2010). The *Tlr4* dataset was partitioned according to the breakpoints detected with the SBP and GARD methods. Because it is now widely recognized that the evolutionary process is not homogeneous across sites, we performed also an analysis partitioned by three codon positions.

Selection on *Tlr4* was analysed at both the inter- and intrasubspecific level. First, we calculated the number of synonymous substitutions per synonymous site (*dS*) and the number of non-synonymous substitutions per non-synonymous site (*dN*) in SNAP (Tamura et al. 2011) following the methods of Nei and Gojobori (1986) and Ota and Nei (1994). Moreover, we aimed to identify codons subject to positive or negative selection using three tests implemented in the DataMonkey program (Pond and Frost 2005a; Pond et al. 2006a): iFEL (internal fixed effects likelihood, convenient for testing selection within species), SLAC (single likelihood ancestor counting), and REL (random effects likelihood). For REL, the Bayes factor was set up to 50. However, since the first two approaches are recommended for larger data sets (> 50 unique sequences) our results based on 18 and 15 unique sequences (Mmm and Mmd, respectively; see Results) should be taken with caution. As pointed out by Pond and Frost (2005b), for data sets of moderate size (20-50 sequences) iFEL is less conservative than SLAC. Likewise, the REL test tends to be somewhat susceptible to Type 1 errors, especially for small datasets, where parameter estimates are likely to have large associated errors (Pond and Frost 2005a). Finally, we employed the McDonald-Kreitman test (MKT) which compares variation within species to the amount of divergence between species at putatively neutral (synonymous) and non-synonymous sites (McDonald and Kreitman

1991). Four types of comparisons were used in the MKT: Mmm/Mmd vs. rats of the tribe Rattini; Mmm/Mmd vs. *R. norvegicus*; Mmm/Mmd vs. southeastern-Asian mouse species *M. caroli*, *M. cooki*, *M. cervicolor*; and Mmm vs. Mmd (results available upon request). All selection tests were applied to a set of wild and WDS mice (i.e. excluding CLSs). Sequences of Asiatic species of *Mus* and Rattini were taken from Fornuskova et al. (2013).

The crystal structure of mouse TLR4 ECD (PDB 2z64) was adopted and modified from the RCSB PDB Protein Data Bank (http://www.rcsb.org/pdb/explore.do?structureId =2z64) (Kim et al. 2007). Subsequently, non-synonymous substitutions, sites under positive and negative selection detected by REL, and previously described binding sites for LPS and MD-2 (lymphocyte antigen 96) (Kim et al. 2007; Park et al. 2009; Ohto et al. 2012) were visualized using The PyMOL Molecular Graphics System, Version 1.6 (Schrödinger).

3.2.3 Results

Genetic diversity of Tlr4

We successfully amplified *Tlr4* sequences of 44 wild Mmd (27 homozygotes and 17 heterozygotes) and 30 wild Mmm (17 homozygotes and 13 heterozygotes; see Table S1 for the number of heterozygous sites for each individual). We found neither heterozygotes between Mmd and Mmm subspecific variants nor trans-subspecific polymorphism. Phylogenetic analysis of amplified sequences of both genes (*Tlr4* and *mt-Cytb*) showed divergence of genetic diversity into two clades corresponding to the Mmm and Mmd subspecies (Table S1, Fig. S1 and Fig. S2). In total, we found 18 and 15 *Tlr4* haplotypes for Mmm and Mmd, respectively (including WDSs, CLSs and wild mice). Similarly, we identified 23 and 37 haplotypes of *mt-Cytb*, for Mmm and Mmd respectively. All Mmd with the present *Bam*HI restriction site harboured an Mmd-related *mt-Cytb* haplotype, and the same holds for Mmm mice (Table S1, Figs. S1b and S2b).

Genetic variation in the *Tlr4* locus was considerably higher in Mmm (N_{Mmm} = 18, A_{Mmm} = 15, π_{Mmm} = 0.0025±0.00016 S.D.) than in Mmd (N_{Mmd} = 15, A_{Mmd} = 7, π_{Mmd} = 0. 0009±0.00007 S.D.). This is even more noticeable for the extracellular domain with four-fold nucleotide diversity and two-fold number of segregating sites in Mmm relative to Mmd (Table 1). Contrary to *Tlr4*, genetic variation in *mt-Cytb* was comparable for both subspecies

(N_{Mmm} = 23, N_{Mmd} = 37; π_{Mmm} = 0.0047±0.00046 S.D., π_{Mmd} = 0.0046±0.00022 S.D.; Table 1).

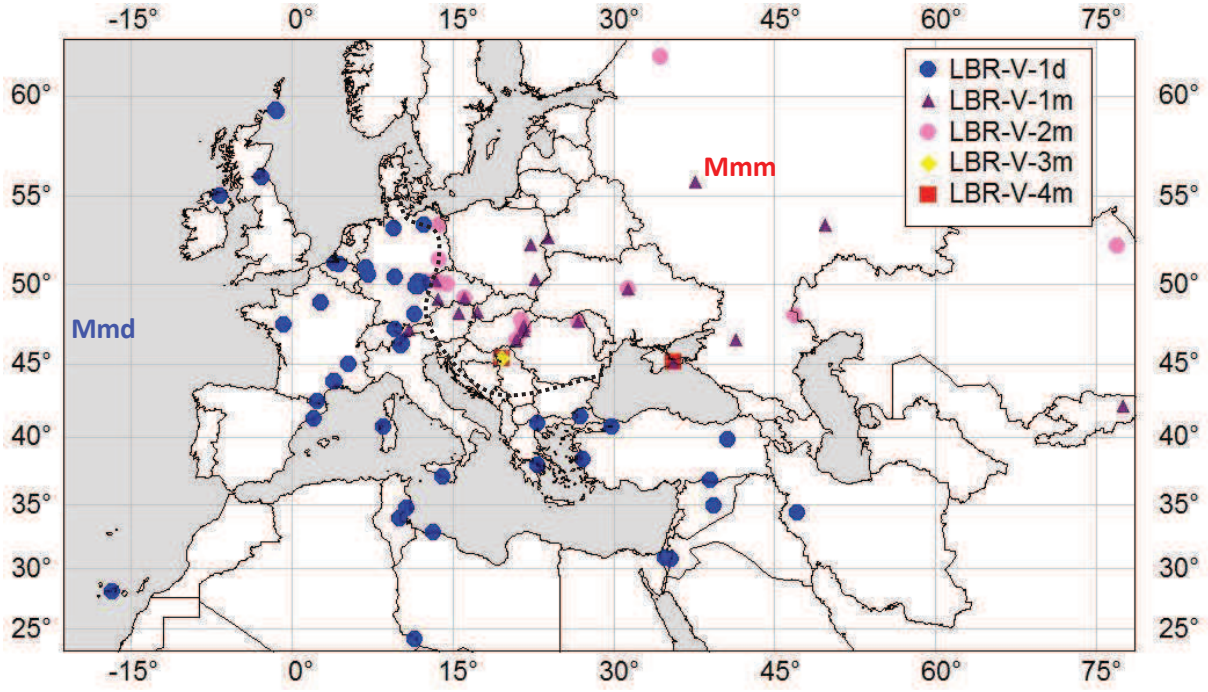


Fig. 1. Distribution of samples analyzed in this study. Blue circles: *Mus musculus domesticus* (Mmd); yellow diamond, red squares, violet triangles and pink circles: *M. m. musculus* (Mmm). Different symbols represent distinct protein variants of the ligand binding region (LBR). Individuals were assigned to the subspecies using HI based on five X-linked loci. Dashed line represents the hybrid zone between two subspecies. Besides sampling localities of free-living populations, the localities of WDSs origin are shown.

Moreover, in all but one Mmd samples we identified a single protein variant of the LBR. The only exception was the A/J laboratory strain which possessed the conservative substitution V254I. This lack of polymorphism is in stark contrast to variation in Mmm where four different variants of LBR were found, with two of them being equally frequent in the Mmm distribution area (Fig. 1). These variants differed at three codons (F350C, D462N and I464V) (Table 2). Nevertheless, all substitutions in the LBR brought about exchanges between biochemically similar amino acids. An overview of all amino acid substitutions, their physicochemical properties and distribution are presented in Fig. 2 and Table 3.

		*N/N**	*A/A**	*π*±S.D.*	*k**	*S**	*Hd*±S.D.*
***Tlr*4-exon 3 2244bp**	Mmd	15/14	7/6	0.0009 ±0.00007	1.929	10	0.736 ±0.052
	Mmm	18/16	15/13	0.0025 ±0.00016	5.595	18	0.882 ±0.028
***Tlr*4-ECD 1644bp**	Mmd	9/8	5/4	0.0005 ±0.00007	0.845	6	0.554 ±0.066
	Mmm	12	7/7	0.0020 ±0.00015	3.267	12	0.800 ±0.043
***Tlr*4-LBR 666bp**	Mmd	2/1	2/1	0.0000	0.000	0	0.000
	Mmm	7/7	4/4	0.0022 ±0.00022	1.473	6	0.627 ±0.063

***Tlr4*-ICD 531bp**	Mmd	5/5	2/2	0.0020 ±0.00014	1.085	4	0.568 ±0.039
	Mmm	8/7	8/7	0.0026 ±0.00016	1.398	4	0.784 ±0.028
***Tlr4*-TIR 435bp**	Mmd	4/4	2/2	0.0024 ±0.00015	1.052	3	0.551 ±0.038
	Mmm	3/2	3/2	0.0001 ±0.00010	0.047	1	0.047 ±0.044
***Cyt* b 1123bp**	Mmd	37/36	15/15	0.0046 ±0.00022	5.105	49	0.983 ±0.009
	Mmm	23/20	9/9	0.0047 ±0.00046	5.254	36	0.974 ±0.016

Table 1, Genetic diversity of *Tlr4* and *mt-Cytb* in two house mouse subspecies; Mmd, *Mus musculus domesticus*; Mmm, *Mus musculus musculus*; 62 and 39 specimens were analysed for Mmd and Mmm respectively; *N*, number of nucleotide haplotypes; *A*, number of amino acid variants; π, nucleotide diversity; *k*, average number of nucleotide differences; *S*, number of polymorphic sites; *Hd*, haplotype diversity; S.D., standard deviation; * indicate analysis without WDS and CLS.

LBR variants	I254V	F350C	D462N	I464V
LBR-V-1d ●	**V**	F	D	I
LBR-V-2d A/J	**I**	F	D	I
LBR-V-1m ▲	V	**F**	**D**	**I**
LBR-V-2m ●	V	**F**	**N**	**I**
LBR-V-3m ♦	V	**C**	**D**	**I**
LBR-V-4m ■	V	**F**	**D**	**V**

Table 2, Description of LBR variants. Coloured symbols correspond to Fig. 1. The distribution of particular variants among sampled specimens is shown in Table S1.

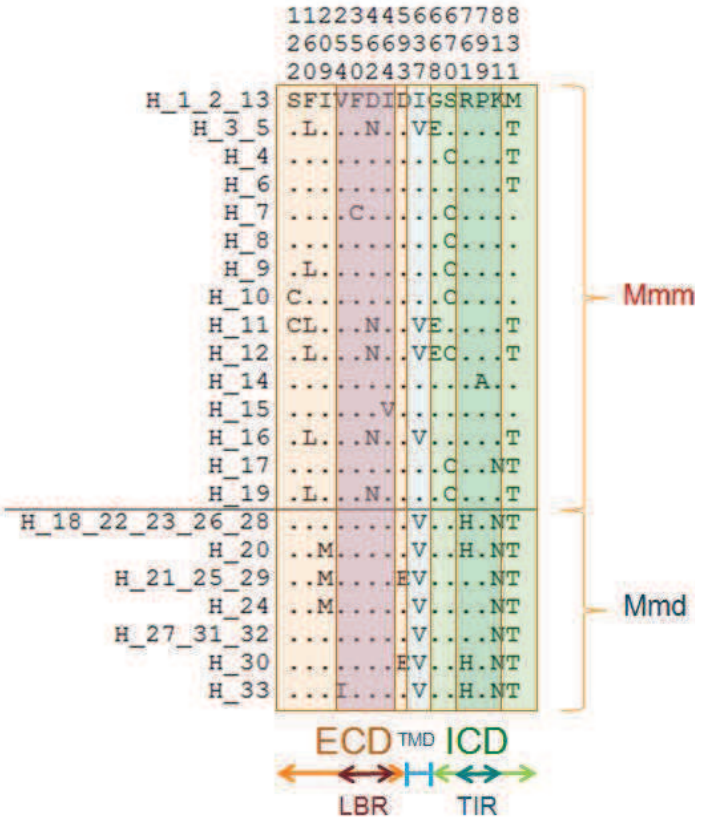


Fig. 2. Overview of *Tlr4* non-synonymous substitutions in Mmd and Mmm. Numbers above alignment indicate amino acid position. ECD, extracellular domain; TMD, transmembrane domain; ICD, intracellular domain; LBR, ligand binding region; TIR, TIR domain. Distribution of individual haplotypes (= alleles) among sampled specimens is presented in Table S1.

Position	aa1	Properties	aa2	Properties	
122	S	SM, P, NEU	C	SM, NP, NEU	
160	F	NP, NEU	L	NP, NEU	
209	I	NP, NEU	M	NP, NEU	
254	V	NP, NEU	I	NP, NEU	LBR
350	F	NP, NEU	C	SM, NP, NEU	
462	D	SM, P, NEG	N	SM, P, NEU	
464	I	NP, NEU	V	NP, NEU	
593	D	SM, P, NEG	E	P, NEG	
637	I	NP, NEU	V	NP, NEU	
668	G	SM, NP, NEU	E	P, NEG	
670	S	SM, P, NEU	C	SM, NP, NEU	
761	R	P, POS	H	P, POS	TIR
799	P	SM, NP, NEU	A	SM, NP, NEU	
811	K	P, POS	N	SM, P, NEU	
831	M	NP, NEU	T	SM, P, NEU	

Table 3. Physicochemical properties of the amino acids involved in non-synonymous substitutions of *Tlr4*; SM, small; NP, non-polar; P, polar; NEU, neutral; POS, positively charged; NEG, negatively charged.

Haplotype network analysis and distribution of genetic groups

The haplotype networks based on nucleotide sequences of exon 3 of *Tlr4* were strikingly different in the two mouse subspecies. In Mmd, there was a single most frequent haplotype (H_18; Fig. 4a). It was present in 71% of all individuals (including CLSs and WDSs) and in 66% of wild mice only (in wild mice it was present in 18 specimens in the homozygote state and in 11 specimens as heterozygotes). Conversely in Mmm individual haplotypes were more evenly represented, none of them occurring in more than 39% of all specimens. The most common Mmm haplotype (H_5) was represented in 33% of wild mice only. Based on the phylogenetic analysis and topology of the haplotype network (Fig. S1 and Fig. S2) we defined three and two haplogroups (HG) for each subspecies, respectively (HG-Id, HG-IId, HG-IIId for Mmd and HG-Im and HG-IIm for Mmm). HG-Id, distributed in the eastern Mediterranean area and northern Africa, seems to be basal for HG-IId and HG-IIId (Fg. 3a). The star-like distribution of HG-IIId, centred on H_18, suggests a recent spatial/demographic expansion of the group. In Mmm, there were two distinct haplotype clouds separated at least by eight substitutions (HG-Im and HG-IIm). Both groups were interconnected by H_2 (CZ, Buškovice) and H_19 (WDS, DE, Lindhorst) which were not included in any HG (see below). The geographical distribution of HG-Im and HG-IIm is very wide, from central Asia to central Europe and they are largely overlapping in most of the Mmm distribution area. Interestingly, the distance between HG-Im and HG-IIm is higher (minimum 8 substitutions) than the distance between HG-Im and HG-Id (minimum 4 substitutions). In contrast to *Tlr4*, the pattern of the *mt-Cytb* haplotype network was very similar for both subspecies with several star-like branching patterns suggesting local

spatial/demographic expansions (Fig. 3b). The geographic distribution of both Mmm and Mmd haplogroups seems to be more intermingled than that of *Tlr4* HGs (see the inset in Fig. 3b). Identification of haplotypes in particular specimens is detailed in Table S1.

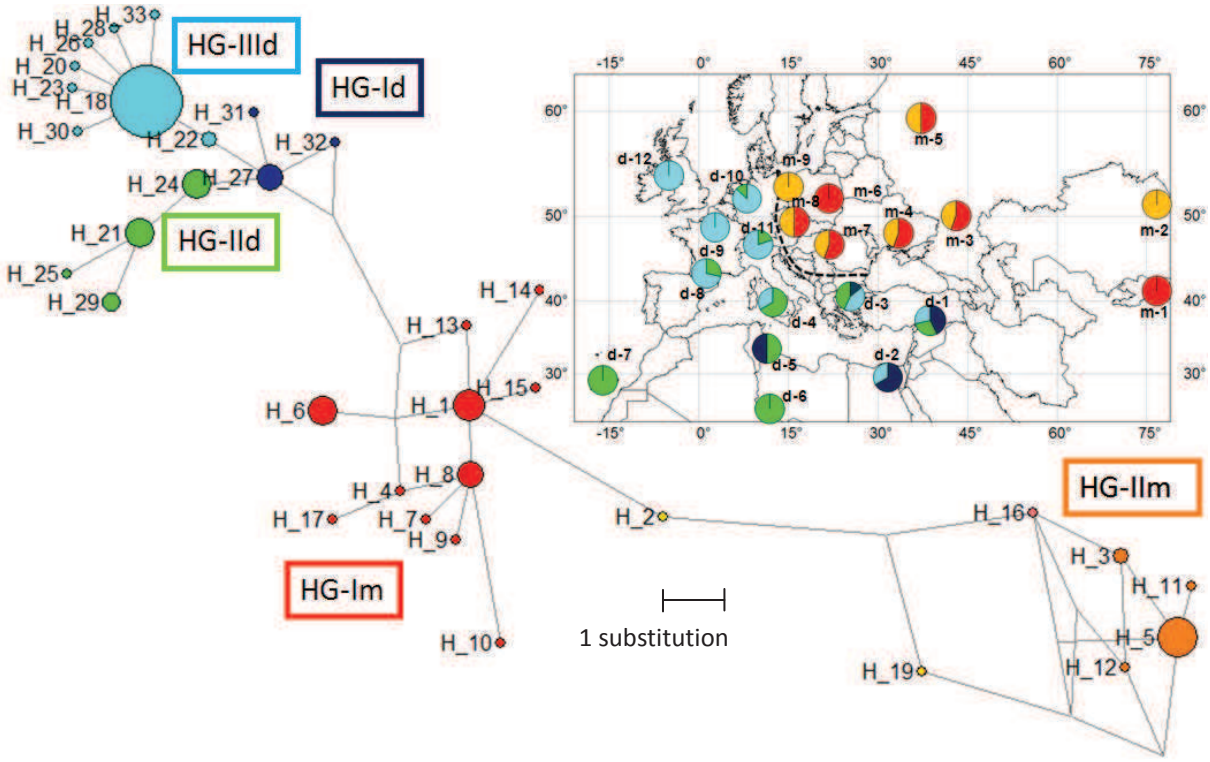


Fig. 4a. Haplotype network and haplogroup distribution of *Tlr4*, **H_** identified haplotypes, **HG-**, identified haplogroup. The size of circles corresponds to the frequency of haplotypes; length of lines is related to the number of substitutions. More detailed information can be found in Table S1. The inset figure represents the geographical distribution of HGs. Colour circles on the map represent the proportion of particular HG (colours correspond to the haplotype network), labels indicate geographic assignment to population groups detailed in Table S1; dashed line shows the position of the house mouse hybrid zone. H_2 and H_19 were excluded from HG due to recombination (see the text for more details).

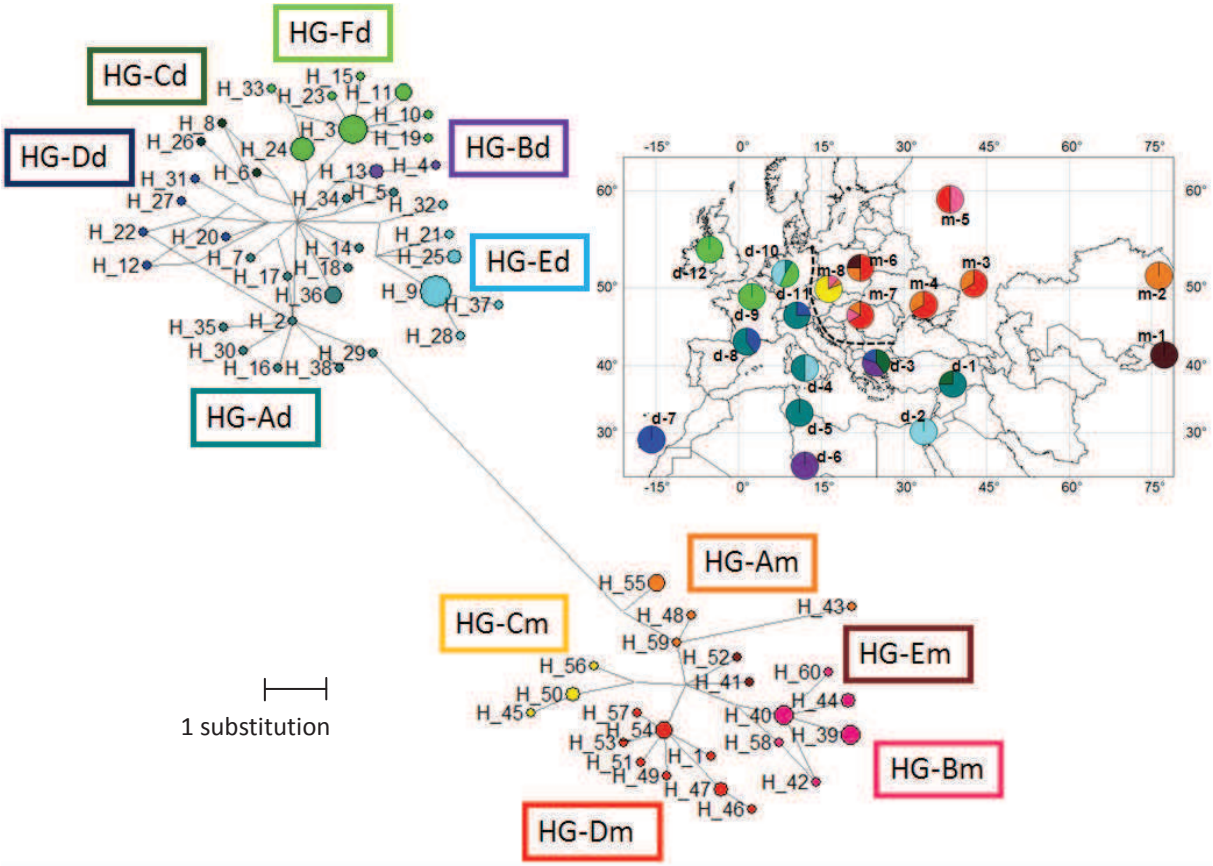


Fig. 4b. Haplotype network and haplogroup distribution of *mt-Cytb*, **H_** identified haplotypes, **HG-**, identified haplogroup. The size of circles corresponds to the frequency of haplotypes; length of lines is related to the number of substitutions. More detailed information can be found in Supplementary materials Tab. S1. The inset figure represents the geographical distribution of HGs. Colour circles on the map represent the proportion of particular HG (colours correspond to the haplotype network); labels indicate geographic assignment to population groups detailed in Table S1; dashed line shows the position of the house mouse hybrid zone.

Recombination and selection in the Tlr4 *gene*

A recombination breakpoint between Mmd and Mmm at position 1779 bp was detected by both tests implemented in DataMonkey. This breakpoint was recognized in one Mmm individual (ST8335, PL, H_13). However, it is based only on a single synonymous substitution at position 849 and more plausible explanation seems convergence. At the intrasubspecific level we detected recombination in two individuals of Mmm. This breakpoint was identified in a conserved region between the LBR and ICD (the SBP algorithm located the recombination breakpoint to nucleotide position 1587 = AA 529, while GARD placed it to position 1611 = AA 537). Haplotypes H_2 and H_19 likely represent recombinant haplotypes between two main Mmm haplogroups (Fig. S3).

The overall *dN/dS* ratios based on Nei and Gojobori's (1986) method using the SNAP program indicated negative selection both for Mmm (*dN* = 0.00173±0.00066; *dS* = 0.00578±0.00236; *dN/dS* = 0.29981) and Mmd (*dN* = 0.00060±0.00036; *dS* =

0.00136±0.00099; *dN/dS* = 0.44436). Similarly, the likelihood-based SLAC approach (Pond and Frost 2005) yielded *dN/dS* = 0.57988 (95% CI = 0.35153-0.89060) for Mmm and *dN/dS* = 0.18097 (95% CI = 0.06489-0.38898) for Mmd. More specifically, the REL test detected eight positively and 14 negatively selected sites (Table 4). Four of the positively selected sites were placed in the ECD; however, none of them was in the LBR (Fig. 4, Table 4). Ten of the 14 negatively selected sites were located in the ECD, three of these codons being in LBR (Fig. 4, Table 4). However, when each subspecies was tested separately, no sites under positive selection were identified in either subspecies by any method used. The iFEL test detected four AA sites under negative selection in Mmm: F416, N529, H537, F647. The first three sites were in the ECD (F416 in LBR) (Fig. 4). Site F647, positioned in the TMD domain, was also confirmed by the SLAC method. In Mmd, only one negatively selected site (L811, TIR region) was suggested by SLAC. None of positively selected site was detected within subspecies by iFEL. The MK test revealed mostly signs of negative selection (not shown).

REL **(Mmm+Mmd+WDSs)**	**ECD (88-635)**	**TMD (636-658)**	**ICD (659-835)**
Positively selected sites	122, 160, 209, 593	637	670, 811, 831
Negatively selected sites	104, 132, 139, 192, 370, 416, 463, 529, 537, 575	647	690, 719, 833

Table 4. Selection tested by REL in both subspecies together, including wild derived strains (WDSs); classic laboratory strains (CLSs) were excluded for this analysis, ECD, extracellular domain, TMD, transmembrane domain, ICD, intracellular domain. Underlined sites in ECD are placed in LBR (248-469). Underlined sites in ICD are placed in TIR domain (671-816). Numbers in brackets indicate position of domains in protein (ECD start with codon 88, first 87 codons are in exon 1 and 2). All sites detected by REL had pp = 0.99.

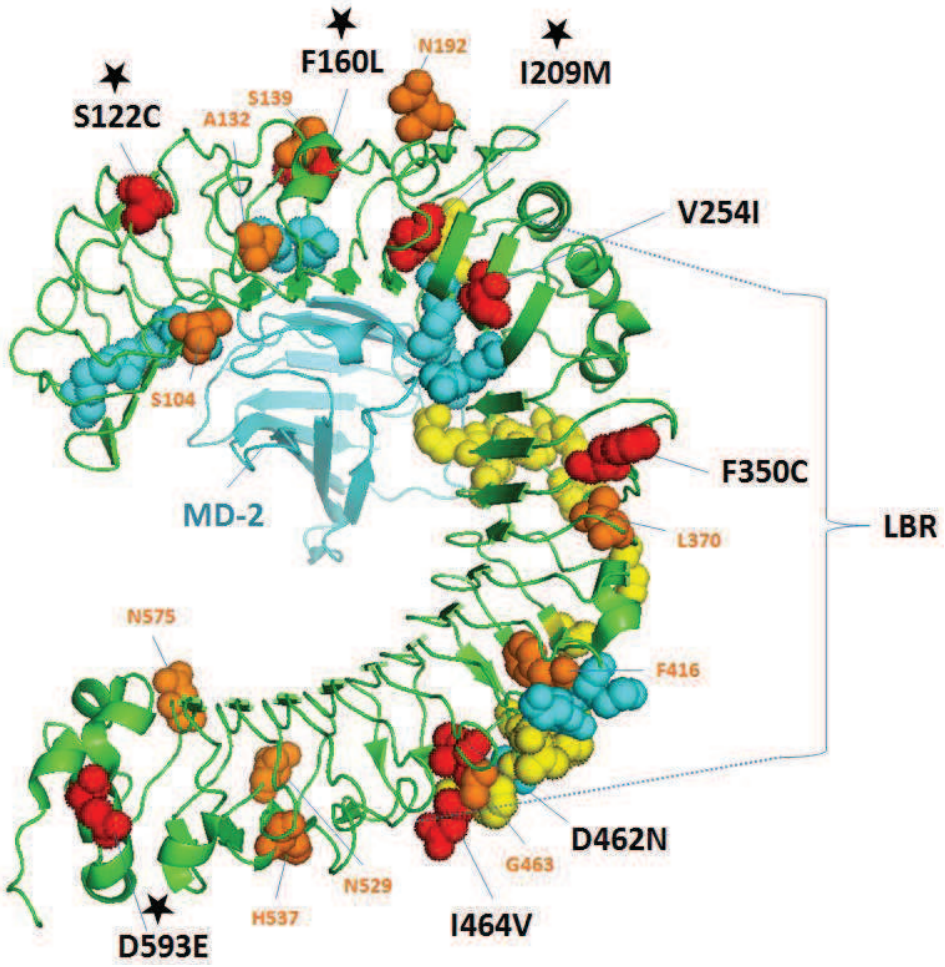


Fig. 3, Ribbon diagram of the TLR4 ECD 3D structure (PDB 2z64 from RCSB PDB Protein Data Bank, http://www.rcsb.org/pdb/explore.do?structureId=2z64, functional sites were described according to Kim et al. (2007); Park et al. (2009); Ohto et al. (2012) and important substitutions were visualised as amino acid space-fill models: cyan corresponds to binding positions for MD-2, yellow represents binding sites for LPS, red represents non-synonymous amino acid changes, orange represents sites under negative selection detected between subspecies by REL; black stars represent detected sites under positive selection in both subspecies by REL; the TLR4 ECD is represented by green colour, MD-2 is represented by cyan colour, LBR, Ligand Binding Region is marked by dashed lines. Description of sites responsible for LPS binding and MD-2 binding is in Table S2, modelling of different sites and design correction made in PyMOL Version 1.5.

3.2.4 Discussion

In the present study we analyzed genetic variation, molecular evolution, and selection patterns in the *Tlr4* gene at the inter- and intrasubspecific level in two subspecies of the house mouse (*Mus musculus musculus*, Mmm, and *M. m. domesticus,* Mmd). We show that although in both subspecies genetic variability is present in the *Tlr4* gene, Mmd exhibits much lower levels of polymorphism than Mmm. This is in contrast with the pattern of genetic variability in *mt-Cytb*. Since we identified several sites under positive selection that differentiate Mmm

from Mmd and some of those lie in functionally important regions of this receptor, we hypothesise that differentiating selection may be responsible for the observed levels of *Tlr4* polymorphism in the two subspecies. Moreover, in Mmm recombination between alleles seems to increase the TLR4 coding sequence variability.

*Tlr*s are generally believed to evolve mainly under purifying selection and, thus, it has been predicted that these genes are relatively uniform within species (e.g. Mukherjee et al. 2009). Contrary to this expectation, we found a moderate intrasubspecific level of *Tlr4* polymorphism. With 15 protein variants in Mmm and 7 protein variants in Mmd, this finding holds true more for Mmm than for Mmd. Indeed, we revealed a striking difference between the two house mouse subspecies in the level of genetic variation. Decreased variability is apparent in Mmd *Tlr4* both at the nucleotide and amino acid levels (Table 1), especially in the ligand-binding region where we found only a single variant in this subspecies, whereas much higher polymorphism level (four variants of LBR) is maintained in Mmm populations. Given the crucial function of TLR4 in mammalian innate immune defence, we may assume that the single LBR variant of TLR4 was advantageous in the past, before or during expansion of Mmd into Western Mediterranean and western/north Europe. On the other hand, we observed similar levels and geographic patterns of genetic variation of *mt-Cytb* in both subspecies. This indicates that the observed pattern does not result from a generally decreased level of genetic polymorphism in Mmd. Moreover, our preliminary investigation of other bacterial-sensing *Tlr* genes in WDSs did not show any marked difference in the level of genetic variation between the two subspecies (data not shown in the present study). Taken altogether, our results may imply the action of contrasting types of selection posing specifically on *Tlr4* in the two house mouse subspecies. A similar contrast in selection on TLR4 across geographically distinct populations is known also in other species. For instance, in humans it has been shown that different haplotypes are positively selected in Sub-Saharan Africa and Eurasia (Ferwerda et al. 2007). Identifying selective forces differentiating subspecies and populations thus appears an intriguing question of current evolutionary biology.

TLR4 is probably the most extensively studied TLR and, more broadly, PRR. The role of TLR4 in lipopolysaccharide (LPS) signalling is indisputable and molecular mechanisms of LPS binding were very well described in human and/or mouse (Kim et al. 2007; Park et al. 2009; Resman et al. 2009; Ohto et al. 2012). LPSs are present in the outer membrane of Gram-negative bacteria and immunologically act as endotoxins, i.e. substances eliciting a strong immune response in animals. Variability of LPS may affect adhesive properties of a microorganism to the cells of its host but also the induced release of inflammatory mediators.

Modifications of LPSs (mainly acylation in the lipid A region) play an important role in the infection process, evasion of the host immune response, and serotypification of Gram–negative bacteria (Robinson et al. 2008). Polymorphism of LPSs has been already shown to be associated with differences in virulence of bacterial strains, for example, *Francisella tularensis, Pseudomonas aeruginosa* or *Yersinia pestis* (Day and Marrceau-Day 1982; Ray et al. 1991; Hajjar et al. 2006; Knirel et al. 2006; Montminy et al. 2006), and as such may be responsible also for evolution and maintenance of recognition mechanisms. This applies especially to *Tlr4* variation. Since the genetic variation of human and livestock TLR4 is associated with susceptibility to various infectious and inflammatory diseases (e.g. Leveque et al. 2003; Hawn et al. 2005; Achyut et al. 2007; Sentitutla et al. 2012; Zaki et al. 2012) and several non-synonymous single nucleotide substitutions (nsSNP) has been identified as immunologically relevant (Ferwerda et al. 2007), we focused on physical properties of the nsSNPs we detected in the house mouse *Tlr4*. In total, we detected 15 nsSNP positions which were distributed evenly across the whole sequenced region including the ECD, TMD and ICD. Of these 15 nsSNPs we found four (V254I, F350C, D462N and I464V) that were located in the LBR close to the ligand binding site of LPSs (Fig. 4). Out of these, the substitution V254I has been identified only in the LBR of the A/J laboratory strain and not in any WDS and/or free-living mice (see also Smirnova et al. 2000). We, therefore, suggest that this substitution does not represent a naturally occurring polymorphism and may have originated in laboratory breeds. On the other hand, particularly functionally important might be the residues 462 and 464 that lie in immediate topological proximity to site F461 which has been previously identified as a residuum essential for LPS binding through hydrophobic interactions in mammals (Park et al. 2009; Resman et al. 2009). We, therefore, hypothesise that these nsSNPs can influence the protein function. Our tests of selection, however, did not support this view as no positively selected sites were identified in the LBR. This suggests that D462N and I464V substitutions either have no functional impact or, at least, that there is no selection differentiating these sites in Mmm and Mmd. Nonetheless, our selection analysis showed that three out of eight sites positively selected on the intersubspecific level were present in the MD-2-binding region, indicating selection differentiating Mmm and Mmd in the TLR4-MD-2 co-evolution. Recent data showed that mouse subspecies harbour genetically different parasites (e.g. *Cryptosporidium tyzzeri*; Kváč et al. 2013). Both subspecies may therefore differ in immune response to specific pathogens. Preliminary laboratory experiments have already shown differences in immunological response between two WDSs derived from

both subspecies (Mmm BULS and Mmd STRA) by stimulating in vitro by ConA and LPS (Piálek et al. 2008).

Although most substitutions identified in the present study involve physically very similar amino acids, it has been shown that even subtle changes in the topological proximity of the binding interface may have substantial impact on the protein function and binding affinity (Zhang et al. 2012). Further studies are, however, needed to test the functional significance of the nsSNPs for recognition of LPS variants.

Previous studies showed that genes encoding TLRs exhibit moderate levels of polymorphism even at intraspecific level (Smirnova et al. 2000; Tschirren et al. 2011; Bergman et al. 2012; Grueber et al. 2012) and that this can have important fitness consequences. In free-living populations it was documented that selection caused by pathogens can vary across different geographic regions and over time (Tschirren et al. 2012). Polymorphism in immune receptors is thought to be primarily maintained by pathogen-evoked balancing selection. This may be viewed as an evolutionary key-and-lock process described by the Matching alleles model (Frank 1993). Applied to receptor-ligand co-evolution, this model proposes that polymorphism in ligands protecting parasites from recognition is mirrored by adaptive host polymorphism allowing detection of ligand-variants by specifically matching receptor alleles (Agrawal and Lively 2002, 2003). Nonetheless, the present analysis performed at the intrasubspecific level in Mmm and Mmd did not reveal any codon under positive selection within the subspecies. This result suggests no clear diversifying selection acting in alleles present within the Mmm and Mmd populations.

In addition to nucleotide substitutions, also intragenic recombination can very quickly create new allele variants. In house mouse, the effect of recombination in the evolution of immune genes is well documented, for example, in the MHC genes (Cizkova et al. 2011). However, in most recent studies on intraspecific TLR polymorphism the relevant tests of recombination have not been performed. Using two alternative approaches our study detected recombination events in the ECD located close to the boundary with the TMD in Mmm. This finding adds another piece of information to the puzzle of PRR polymorphism in free-living rodents showing that recombination might be an important factor increasing TLR variability. Our results are consistent with studies of several other mammals reporting signals of recombination in the ECD in human TLR4 (Zaki et al. 2012) or bovine TLR3, TLR4 and TLR10 (Seabury et al. 2010). Detailed analysis of our sequences suggests that haplotypes H_2 and H_19 are recombinants composed of the ECD from haplogroup HG-IIm and ICD of HG-Im. These two Mmm haplotypes are genetically dissimilar and were found in two specimens

separated by 500 km (see Table S1). Assuming that they represent two independent recombination events, we suggest that recombination in this genic region may be relatively frequent in nature. Because the recombination breakpoints combine different ECDs and ICDs, the WDS SLINT bearing H_19 (in combination with other WDSs from *Tlr4* haplogroups HG-Im and HG-IIm) provides a unique opportunity to discriminate the role of LBR-mediated LPS recognition from the transduction of the signal by the TIR domain.

Although pathogens likely play an important role in evolution of *Tlr4* variability, it may be admitted that the observed difference between the subspecies in TLR4 polymorphism might have arisen as a result of non-adaptive evolutionary processes during mouse colonization of the Western Palearctic. For example, in some avian populations affected by bottlenecks the dominant force influencing evolution of TLRs seems to be genetic drift, outweighing the effect of selection (Grueber et al. 2012, 2013). Similarly, genetic drift also shaped the genetic history of human TLR4 during population expansion out of Africa (Netea et al. 2012). Thus, the pattern observed in mice might result, e. g., from differences between subspecies in historical demographic processes (quick expansion of Mmd and two founder populations or refuges for Mmm). In such a case we would, however, expect similar contrasting patterns in *mt-Cytb*. Since this was not the case, we may consider the explanation of the observed pattern by genetic drift as unlikely. Finally, we must also bear in mind that Mmd *Tlr4* may not be the positively selected gene itself but only a gene involved in gene hitchhiking. Nevertheless, this hypothesis is in contradiction with results of selection analysis which have detected eight positively selected sites in ECD of free living *Mus musculus*.

Acknowledgement

We thank the following colleagues for donating mouse samples: Stuart J. E. Baird, Joëlle Goüy de Bellocq, Nina Bulatova, Jörg Burgstaller, Dagmar Čížková, Barbara Dod, Ľudovít Ďureje, Jiří Forejt, Sophie Gryseels, Heidi C. Hauffe, Sonja Ihle, Adam Konečný, Stanislav Martínek, Natália Martínková, Ferhat Matur, Peter Mikulíček, George Mitsainas, Alina Mishta, Martina Mrkvicová Vyskočilová, Pavel Munclinger, Jana Piálková, Alexis Ribas, Jana Svobodová, Vitaly Volobouev, Mladen Vujošević, Jan M. Wójcik and Jan Zima. Lucie Vlčková genotyped X-linked loci and the *Nd1* gene. This research was supported by the Czech Science Foundation (project 206/08/0640) to JP. We thank Stuart J. E. Baird for language corrections and significant improvement of the previous versions of the manuscript and Jean-François Cosson for general support and commentaries. Phylogenetic

reconstructions were computed at the CBGP HPC computational platform located at Centre de Biologie et Gestion des Populations, Montpellier.

Authors' Contributions

Conceived and designed the experiments: AF JP JB. Performed the sequencing: AF. Analysed the data: AF. Contributed samples: JP AF. Wrote the paper: AF JB JP MM MV (all authors contributed equally, sorted by alphabetic order). All authors read and approved the final manuscript.

SUPPLEMENTARY MATERIALS

Table S1. Summary of sampled specimens, identification of haplotypes and NCBI GenBank accession numbers.

ID	Genome HI	*Bam*HI = *mt-Cytb*	Locality	Country	Sex	G	Latitude	Longitude	*Colour pies, Fig.2 Tlr4/ mt-Cytb*	*Tlr4*_h1/ h2	*Tlr4*_ HG1/HG2	*Tlr4*_LBR-V	*mt-Cytb* _h	*mt-Cytb* _HG	N°of het positions	GenBank Acc. *Tlr4*	GenBank Acc. *mt-Cytb*
MU103	1.00	-	Tosor	KG	M	0	42° 09' 56"	77° 26' 40"	m-1	14	Im	1m	41	Em	0	KF696929	KF697033
SU4736	1.00	-	Petrozavodsk	RU	M	0	61° 47' 00"	34° 21' 00"	m-5	5	Im	2m	54	Dm	0	KF696930	KF697034
SU4391	1.00	-	Bakhilova Polyana	RU	M	0	53° 26' 00"	49° 40' 00"	m-3	1	Im	1m	57	Dm	0	KF696931	KF697035
SU5214	1.00	-	Sal'sk	RU	M	0	46° 35' 00"	41° 21' 00"	m-3	1	Im	1m	55	Am	0	KF696932	KF697036
SU5220	1.00	-	Moscow	RU	M	0	55° 44' 14"	37° 38' 02"	m-5	1	Im	1m	56	Cm	0	KF696933	KF697037
SU5218	1.00	-	Shkili Sands, Middle Volga	RU	F	0	48° 08' 00"	46° 54' 00"	m-3	5	IIm	2m	54	Dm	0	KF696934	KF697038
SK1347	1.00	-	Pavlodar	KZ	F	0	52° 16' 48"	76° 56' 25"	m-2	5	IIm	2m	43	Am	0	KF696935	KF697039
SU1235	1.00	-	Bialowieza	PL	F	0	52° 42' 00"	23° 51' 00"	m-2	8/1	Im	1m	55	Am	1	KF696936	KF697040
ST8337	1.00	-	Majdan Sieniawski	PL	F	0	50° 17' 37"	22° 43' 10"	m-6	8/1	Im	1m	52	Em	1	KF696937	KF697041
ST8303	1.00	-	Bielany	PL	F	0	52° 19' 47"	22° 14' 34"	m-6	8/6	Im	1m	46	Dm	3	KF696938	KF697042
ST8335	1.00	-	Majdan Sieniawski	PL	F	0	50° 17' 26"	22° 43' 06"	m-6	13	Im	1m	47	Dm	0	KF696939	KF697043
SU5158	1.00	-	Bački Petrovac	RS	M	0	45° 20' 29"	19° 34' 58"	m-7	7/8	Im	3m/1m	54	Dm	1	KF696940	KF697044
JPC2803	1.00	-	Studenec	CZ	F	0	49° 12' 00"	16° 04' 00"	m-8	5	IIm	2m	39	Bm	0	KF696941	KF697045
JPC2821	1.00	-	Studenec	CZ	F	0	49° 12' 00"	16° 04' 00"	m-8	6	Im	1m	39	Bm	0	KF696942	KF697046
SK1469	1.00	-	Milhostov	CZ	F	0	50° 09' 32"	12° 27' 30"	m-8	1/5	Im/IIm	1m/2m	44	Bm	10	KF696943	KF697047
MM1178	1.00	-	Kokrdy	CZ	F	0	50° 06' 41"	13° 42' 41"	m-8	3	IIm	2m	40	Bm	0	KF696944	KF697048
MM21	1.00	-	Zdíkovec	CZ	F	0	49° 05' 47"	13° 41' 02"	m-8	4	Im	1m	40	Bm	0	KF696945	KF697049
SK1004	1.00	-	Buškovice	CZ	M	0	50° 13' 18"	13° 22' 27"	m-8	1/2	Im/NCS	1m	42	Bm	3	KF696946	KF697050
SU5607	1.00	-	Stepantsi	UA	M	0	49° 42' 14"	31° 18' 11"	m-4	1/16	Im/IIm	1m/2m	47	Dm	8	KF696947	KF697051
SU5209	1.00	-	Primorskoe Pheodosiya	UA	M	0	45° 07' 31"	35° 30' 26"	m-4	1/15	Im	1m/4m	55	Am	1	KF696948	KF697052
360	1.00	-	Botosani	RO	M	0	47° 45' 00"	26° 40' 00"	m-4	6/5	Im/IIm	1m/2m	1	Dm	10	KF696949	KF697053
ST8358	1.00	-	Hajdúnánás	HU	F	0	47° 50' 40"	21° 25' 45"	m-7	3/5	IIm	2m	53	Dm	1	KF696950	KF697054
ST8360	1.00	-	Szepes	HU	F	0	47° 28' 39"	21° 35' 24"	m-7	8/12	Im/IIm	1m/2m	48	Am	9	KF696951	KF697055
ST8387	1.00	-	Gabortelep	HU	M	0	46° 31' 39"	20° 54' 59"	m-7	8	Im	1m	51	Dm	0	KF696952	KF697056
ST8389	1.00	-	Gáborteleр	HU	F	0	46° 32' 00"	20° 56' 00"	m-7	9/5	Im/IIm	1m/2m	50	Cm	cloned 10	KF696953	KF697057
ST8381	1.00	-	Szomolyon	HU	F	0	47° 07' 43"	21° 34' 08"	m-7	10/11	Im/IIm	1m/2m	49	Dm	10	KF696954	KF697058
SK18	0.80	-	Staudach	AT	F	0	47° 16' 17"	10° 57' 23"	m-8	6	Im	1m	45	Cm	0	KF696955	KF697059
SK1787	1.00	-	Thallern	AT	F	0	48° 17' 08"	15° 33' 11"	m-8	6	Im	1m	50	Cm	0	KF696956	KF697060
SK843	1.00	+	Lindhorst	DE	M	0	53° 26' 34"	13° 46' 06"	m-9/d-10	5	IIm	2m	9	Ed	0	KF696957	KF697031
SK837	0.70	+	Lauchhammer	DE	F	0	51° 28' 06"	13° 44' 24"	m-9/d-10	5	IIm	2m	9	Ed	0	KF696958	KF697032
STUF-WDS	1.00	-	Studenec	CZ	F,M	31	49° 12' 00"	16° 04' 00"	m-8	6	Im	1m	60	Bm	0	KF696959	KF697061
STUP-WDS	1.00	-	Studenec	CZ	F,M	30	49° 12' 00"	16° 04' 00"	m-8	5	IIm	2m	39	Bm	0	KF696960	KF697062
STUS-WDS	1.00	-	Studenec	CZ	F,M	22	49° 12' 00"	16° 04' 00"	m-8	6	Im	1m	39	Bm	0	KF696961	KF697063
BULS-WDS	1.00	-	Buškovice	CZ	F,M	25	50° 13' 18"	13° 22' 27"	m-8	5	IIm	2m	58	Bm	0	KF696962	KF697064
BUSNA-WDS	1.00	-	Buškovice	CZ	F,M	30	50° 13' 18"	13° 22' 27"	m-8	5	IIm	2m	40	Bm	0	KF696963	KF697065
PWD-WDS	1.00	-	Kunratice	CZ	F,M	x+12	50° 00' 47"	14° 29' 07"	m-8	5	IIm	2m	40	Bm	0	KF696964	KF697066
SENK-WDS	1.00	-	Šenkvice	SK	F,M	1	48° 17' 55"	17° 21' 05"	m-8	17	Im	1m	59	Am	0	KF696965	KF697067
SMIL-WDS	1.00	-	Milhostov	CZ	F,M	3	50° 09' 33"	12° 27' 30"	m-8	5	IIm	2m	44	Bm	0	KF696966	KF697068
SLINT-WDS	1.00	+	Lindhorst	DE	F,M	1	53° 26' 34"	13° 46' 06"	m-9/d-10	19	NCS	2m	9	Ed	0	KF696967	KF697030
SHY-WDS	0.00	+	Hohenberg	DE	F,M	4	50° 06' 07"	12° 12' 46"	d-10	18	IIId	1d	9	Ed	0	KF696867	KF696968
STRA-WDS	0.00	+	Straas	DE	F,M	29	50° 10' 53"	11° 45' 44"	d-10	18	IIId	1d	3	Fd	0	KF696868	KF696969
STRB-WDS	0.00	+	Straas	DE	F,M	26	50° 10' 53"	11° 45' 44"	d-10	18	IIId	1d	3	Fd	0	KF696869	KF696970
STLT-WDS	0.00	+	Straas	DE	F,M	19	50° 10' 53"	11° 45' 44"	d-10	18	IIId	1d	3	Fd	0	KF696870	KF696971

SCHUNT-WDS	0.00	+	Schweben	DE	F,M	7	50° 26' 00"	09° 35' 00"	d-10	18	IIId	1d	9	Ed	0	KF696871	KF696972
SCHEST-WDS	0.00	+	Schweben	DE	F,M	7	50° 26' 10"	09° 35' 10"	d-10	18	IIId	1d	9	Ed	0	KF696872	KF696973
SIN-WDS	0.00	+	Scar, Sanday Isl., Orkneys	UK	F,M	5	59° 18' 00"	-02° 33' 00''	d-12	18	IIId	1d	24	Fd	0	KF696873	KF696974
SIT-WDS	0.00	+	Scar, Sanday Isl., Orkneys	UK	F,M	6	59° 18' 00"	-02° 33' 00''	d-12	18	IIId	1d	24	Fd	0	KF696874	KF696975
SFEL-WDS	0.00	+	Feldkirch	AT	F,M	2	47° 15' 42"	09° 35' 10"	d-10	18	IIId	1d	38	Ad	0	KF696875	KF696976
SPOS-WDS	0.00	+	Migiondo	IT	F,M	3	46° 19' 23"	10° 18' 17"	d-11	18	IIId	1d	36	Ad	0	KF696876	KF696977
SUV-WDS	0.00	+	Sernio	IT	F,M	3	46° 13' 26"	10° 12' 20"	d-11	18	IIId	1d	36	Ad	0	KF696877	KF696978
SCHEFE-WDS	0.00	+	Schweben	DE	F,M	7	50° 26' 10"	09° 35' 10"	d-10	18	IIId	1d	9	Ed	0	KF696878	KF696979
STAIL-WDS	0.00	+	Schweben	DE	F,M	9	50° 26' 00"	09° 35' 00"	d-10	18	IIId	1d	9	Ed	0	KF696879	KF696980
SOTT-WDS	0.00	+	Ottmannsreuth	DE	F,M	1	49° 53' 27"	11° 37' 04"	d-10	18	IIId	1d	11	Fd	0	KF696880	KF696981
SPLY-WDS	0.00	+	Plössen	DE	F,M	2	49° 51' 18"	11° 47' 10"	d-10	18	IIId	1d	11	Fd	0	KF696881	KF696982
C57BL/6J-CLS	0.20	+	classical laboratory strain	lab	F,M	-	-	-	-	29	IId	1d	3	Fd	0	KF696882	KF696983
A/J-CLS	0.00	+	classical laboratory strain	lab	F,M	-	-	-	-	33	IIId	2d	3	Fd	0	KF696883	KF696984
C3Ha-CLS	0.00	+	classical laboratory strain	lab	F,M	-	-	-	-	29	IId	1d	3	Fd	0	KF696884	KF696985
SK1482	0.00	+	Plössen	DE	F	0	49° 51' 18"	11° 47' 10"	d-10	18	IIId	1d	11	Fd	0	KF696885	KF696986
JPC2705	0.00	+	Straas	DE	F	0	50° 10' 53"	11° 45' 44"	d-10	18	IIId	1d	3	Fd	0	KF696886	KF696987
ST5068	0.00	+	Arzdorf	DE	F	0	50° 36' 00"	07° 05' 00"	d-10	18	IIId	1d	21	Ed	0	KF696888	KF696989
ST6688	0.00	+	Arzdorf	DE	F	0	50° 36' 00"	07° 05' 00"	d-10	18/23	IIId	1d	3	Fd	1	KF696891	KF696992
SU607	0.00	+	Schweben	DE	F	0	50° 26' 00"	09° 35' 00"	d-10	18	IIId	1d	9	Ed	0	KF696887	KF696988
SU627	0.00	+	Schweben	DE	F	0	50° 26' 10"	09° 35' 10"	d-10	18	IIId	1d	37	Ed	0	KF696893	KF696994
SK6	0.10	+	München	DE	F	0	48° 09' 00"	11° 26' 00"	d-10	24	IId	1d	16	Ad	0	KF696889	KF696990
SK899	0.00	+	Hamersen	DE	M	0	53° 15' 15"	09° 28' 50"	d-10	18	IIId	1d	19	Fd	0	KF696890	KF696991
ST7519	0.00	+	Köln	DE	F	0	50° 58' 37"	06° 57' 18"	d-10	18	IIId	1d	23	Fd	0	KF696892	KF696993
SK957	0.00	+	Suckow	DE	F	0	53° 24' 44"	12° 19' 40"	d-10	18	IIId	1d	9	Ed	0	KF696894	KF696995
ST9613	0.00	+	Scar, Whitemill Bay, Sanday Isl.	UK	F	0	59° 18' 00"	-02° 33' 00"	d-12	18	IIId	1d	24	Fd	0	KF696895	KF696996
ST9597	0.00	+	Bay of Brough, Sanday Isl.	UK	F	0	59° 16' 00"	-02° 36' 00"	d-12	18	IIId	1d	24	Fd	0	KF696896	KF696997
ST9600	0.00	+	Little Sea, Sanday Isl.	UK	F	0	59° 15' 00"	-02° 35' 00"	d-12	18/20	IIId	1d	24	Fd	1	KF696897	KF696998
SK2046	0.00	+	Edinburgh	UK	M	0	55° 57' 03"	-03° 10' 53"	d-12	18	IIId	1d	15	Fd	0	KF696898	KF696999
SU5122	0.00	+	Londonderry	IE(UK)	F	0	55° 00' 25"	-07° 17' 20"	d-12	18	IIId	1d	24	Fd	0	KF696899	KF697000
SU4648	0.00	+	Antwerp	BE	M	0	51° 11' 50"	04° 24' 35"	d-10	29/30	IId/IIId	1d	28	Ed	4	KF696900	KF697001
SK1349	0.00	+	Stekene	BE	M	0	51° 13' 53"	04° 00' 33"	d-10	21/18	IId/IIId	1d	10	Fd	4	KF696901	KF697002
SK1515	0.00	+	Saint Jean-et-Royans	FR	F	0	45° 01' 10"	05° 16' 20"	d-8	18	IIId	1d	12	Dd	0	KF696902	KF697003
ST7605	0.00	+	Brouzet-les-Quissac	FR	F	0	43° 50' 00"	03° 58' 00"	d-8	21/22	IId/IIId	1d	35	Ad	3	KF696903	KF697004
SU5046	0.00	+	Lagny	FR	F	0	48° 52' 00"	02° 43' 00"	d-9	18	IIId	1d	33	Fd	0	KF696904	KF697005
SU4886	0.00	+	Pi de Conflent	FR	F	0	42° 30' 00"	02° 21' 00"	d-8	18/26	IIId	1d	29	Ad	1	KF696905	KF697006
JPC2912	0.00	+	Valflaunes	FR	F	0	43° 48' 00"	03° 52' 00"	d-8	21/18	IId/IIId	1d	2	Ad	4	KF696906	KF697007
SU5313	0.00	+	Les Brosses	FR	F	0	47° 30' 26"	-00° 46' 03"	d-9	18	IIId	1d	3	Fd	0	KF696907	KF697008
SER1047	0.00	+	Sernio	IT	F	0	46° 13' 26"	10° 12' 20"	d-11	18	IIId	1d	36	Ad	0	KF696908	KF697009
SU5019	0.10	+	Falconara, Sicily	IT	F	0	37° 06' 32"	14° 02' 16"	d-4	24/18	IId/IIId	1d	32	Ed	3	KF696909	KF697010
SU4894	0.00	+	Sardinia	IT	M	0	40° 43' 00"	08° 35' 00"	d-4	21	IId	1d	30	Ad	0	KF696910	KF697011
ST8495	0.10	+	Tovo S Agata	IT	F	0	46° 14' 32"	10° 14' 33"	d-11	21/22	IId/IIId	1d	22	Dd	3	KF696911	KF697012
SU4991	0.00	+	Alojera, Canary Archipel.	ES	M	0	28° 09' 00"	-17° 19' 00"	d-7	25	IId	1d	31	Dd	0	KF696912	KF697013
SK916	0.00	+	El Prat del Llobregat	ES	M	0	41° 18' 00"	02° 04' 00"	d-8	18	IIId	1d	20	Dd	0	KF696913	KF697014
MM760	0.00	+	Kilkis	GR	M	0	40° 59' 42"	22° 52' 14"	d-3	21/18	IId/IIId	1d	6	Cd	cloned 4	KF696914	KF697015
R307	0.10	+	Korinthos	GR	F	0	37° 56' 19"	22° 55' 43"	d-3	27	Id	1d	13	Bd	0	KF696915	KF697016
TU21	0.00	+	Harran	TR	F	0	36° 51' 00"	39° 01' 00"	d-1	27/24	Id/IId	1d	34	Ad	1	KF696916	KF697017
SK1763	0.20	+	Kuleli	TR	M	0	41° 29' 39"	26° 56' 31"	d-3	18	IIId	1d	13	Bd	0	KF696917	KF697018
SU3770	0.00	+	Izmir	TR	F	0	38° 24' 07"	27° 07' 12"	d-3	24/18	IId/IIId	1d	27	Dd	3	KF696918	KF697019
SK1768	0.00	+	Gölcük	TR	M	0	40° 42' 55"	29° 49' 37"	d-3	24/18	IId/IIId	1d	8	Cd	3	KF696919	KF697020
SU6834	0.00	+	Ciftlik, Askale	TR	M	0	39° 47' 11"	40° 38' 31"	d-1	24/18	IId/IIId	1d	26	Cd	3	KF696920	KF697021
SU5229	0.20	+	Hazeva	IL	F	0	30° 46' 02"	35° 16' 39"	d-2	31	Id	1d	25	Ed	0	KF696921	KF697022
SU5224	0.20	+	Nachal Nizzana	IL	F	0	30° 51' 30"	34° 45' 05"	d-2	28/27	IIId/Id	1d	25	Ed	cloned 3	KF696922	KF697023

SK1773	0.00	+	Kermanshah	IR	M	0	34° 23' 00"	47° 06' 00"	d-1	27	Id	1d	14	Ad	0	KF696923	KF697024
M273	0.00	+	Palmyra	SY	M	0	34° 56' 00"	39° 16' 00"	d-1	32/18	Id/IIId	1d	5	Ad	cloned 3	KF696924	KF697025
SK811	0.20	+	Gabes	TN	M	0	33° 53' 35"	10° 06' 06"	d-5	27	Id	1d	17	Ad	0	KF696925	KF697026
SK813	0.20	+	Sfax	TN	M	0	34° 45' 52"	10° 45' 11"	d-5	21/27	IId/Id	1d	18	Ad	2	KF696926	KF697027
PB1599	0.20	+	Al Qusbat	LY	M	0	32° 53' 00"	13° 09' 00"	d-5	21	IId	1d	7	Ad	0	KF696927	KF697028
LIB11	0.20	+	Al Awayna	LY	M	0	24° 10' 00"	11° 30' 00"	d-6	24	IId	1d	4	Bd	0	KF696928	KF697029

NOTE. ID - identification of specimens, **WDS** - wild-derived strain; C**LS** - laboratory strain; **Genome** – classification of mice based on hybrid index, **HI**, estimated from 5 X-linked loci (for details see Methods) to either subspecies (ranging from 0.00 for *Mus musculus domesticus* to 1.00 for *M. m. musculus*); ***Bam*HI** - presence (+) or absence (-) of restriction site; **Country codes**: **AT** - Austria, **CZ** - Czech Republic, **BE** - Belgium, **DE** - Germany, **ES** - Spain, **FR** - France, **GR** - Greece, **HU** - Hungary, **IT** - Italy, **IE** - Northern Ireland, **IL** - Israel, **IR** - Iran, **KG** - Kyrgyzstan, **KZ** - Kazakhstan, **LY** - Libyan Arab Jamahiriya, **PL** - Poland, **RO** - Romania, **RS** - Serbia, **RU** - Russian Federation, **SK** - Slovakia, **SY** - Syrian Arab Republic, **TN** - Tunisia, **TR** - Turkey, **UA** - Ukraine, **UK** - United Kingdom; **NCS** - non-classified sample; **G** – number of generation under brother x sister matings, G0 stands for wild mice, x+ indicates that the strain was kept for some generations in different laboratory; **Colour pies, Fig.2 *Tlr4*/*mt-Cytb*,** labels m-1-9 and d-1-12 correspond to pies at **Fig. 2a** and **2b**, discrepancies between *Tlr4* and *mt-Cytb* are represented by /; ***Tlr4*_h1/h2** - haplotype identification, doubled labels indicate heterozygotes; ***Tlr4*_HG1/HG2** - identification to haplogroups, doubled labels indicate heterozygotes; ***Tlr4*_LBR-V** – indicates type of LBR variants, two variants are marked by /, labels correspond to **Fig. 1** and **Table 2**; ***mt-Cytb*_h** - cytochrome b haplotype identification; ***mt-Cytb*_HG** - cytochrome b haplogroups identification; **N°of het positions** - number of heterozygous sites for *Tlr4*; **GenBank Acc.** - GenBank Accession numbers.

Fig. S1-a, *Tlr4* phylogeny based on Bayesian inference (MrBayes v.3.1), 10,000,000 iterations and 2,500,000 burn-in iterations, distribution of haplotypes among individuals can be found in Table S1. Numbers above branches indicate posterior support of particular branch.

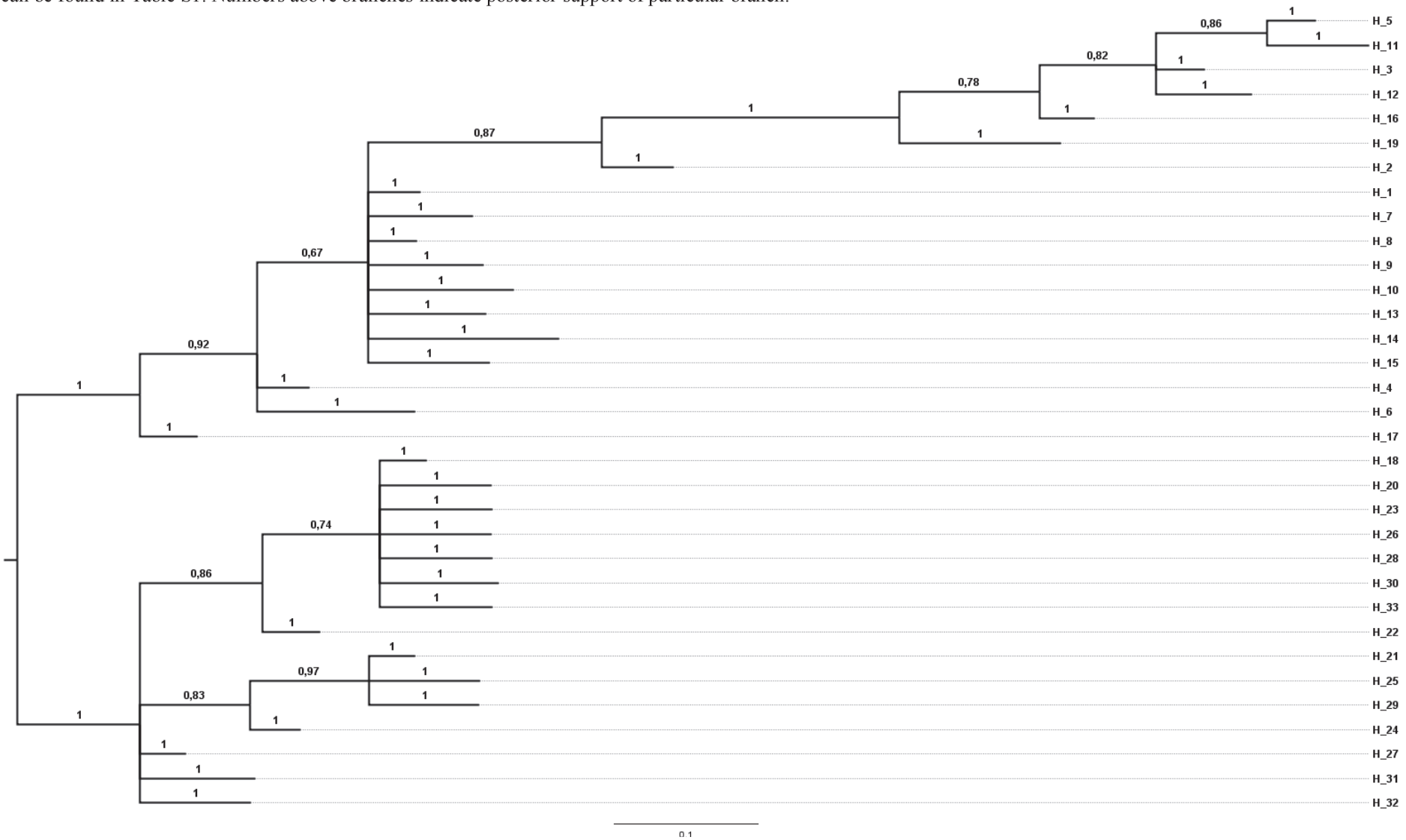


Fig. S1-b, *mt-Cytb* phylogeny was based on Bayesian inference (MrBayes v.3.1), 10,000,000 iterations and 2,500,000 burn-in iterations; distribution of haplotypes among individuals can be found in Table S1. Numbers above branches indicate posterior support of particular branch.

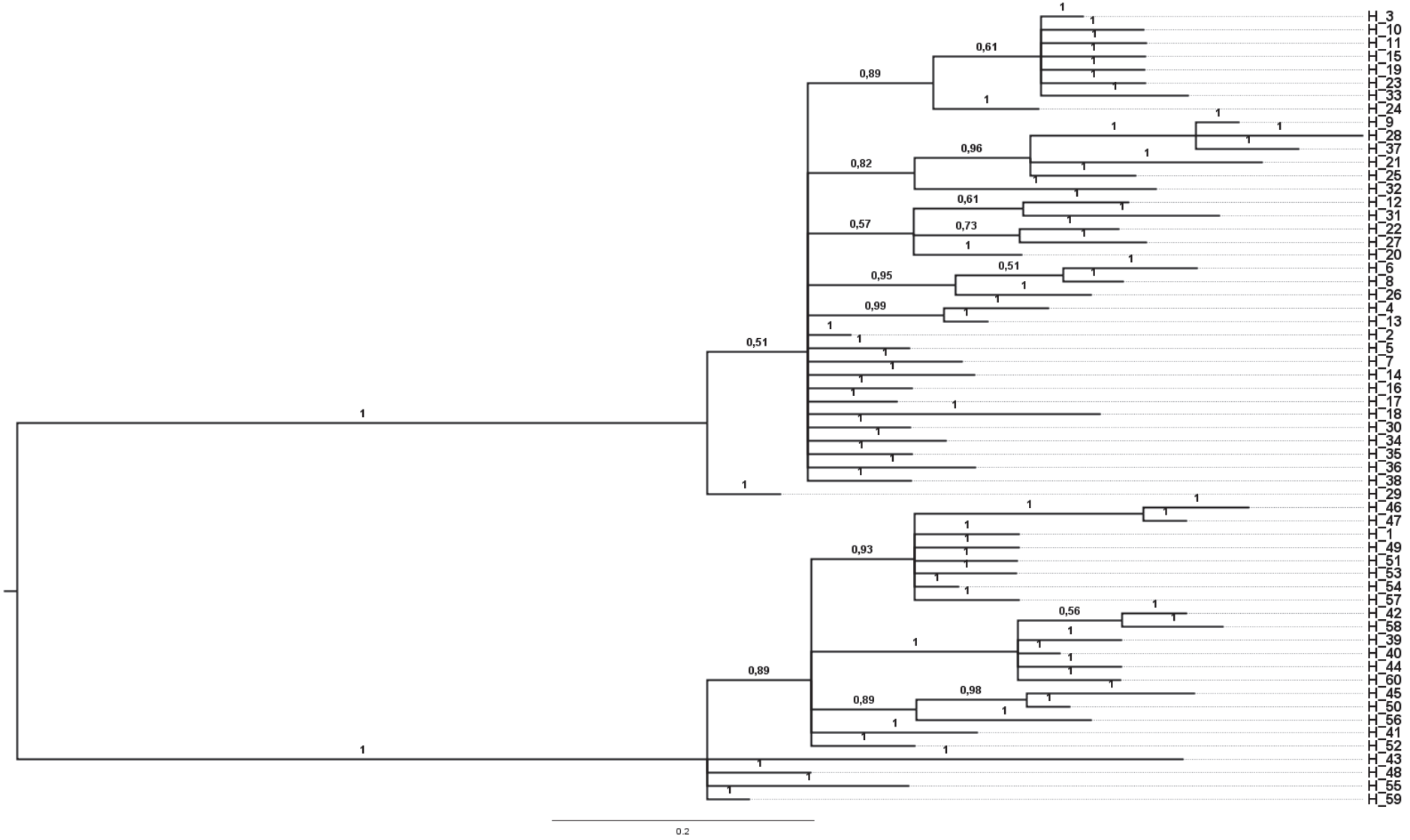


Fig. S2-a, *Tlr4* haplogroup definition; Haplotype network reconstructed in Network v. 4.6.1.1., phylogeny taken from Fig. S 1-a, distribution of haplotypes among sampled individuals can be found in Table S1.

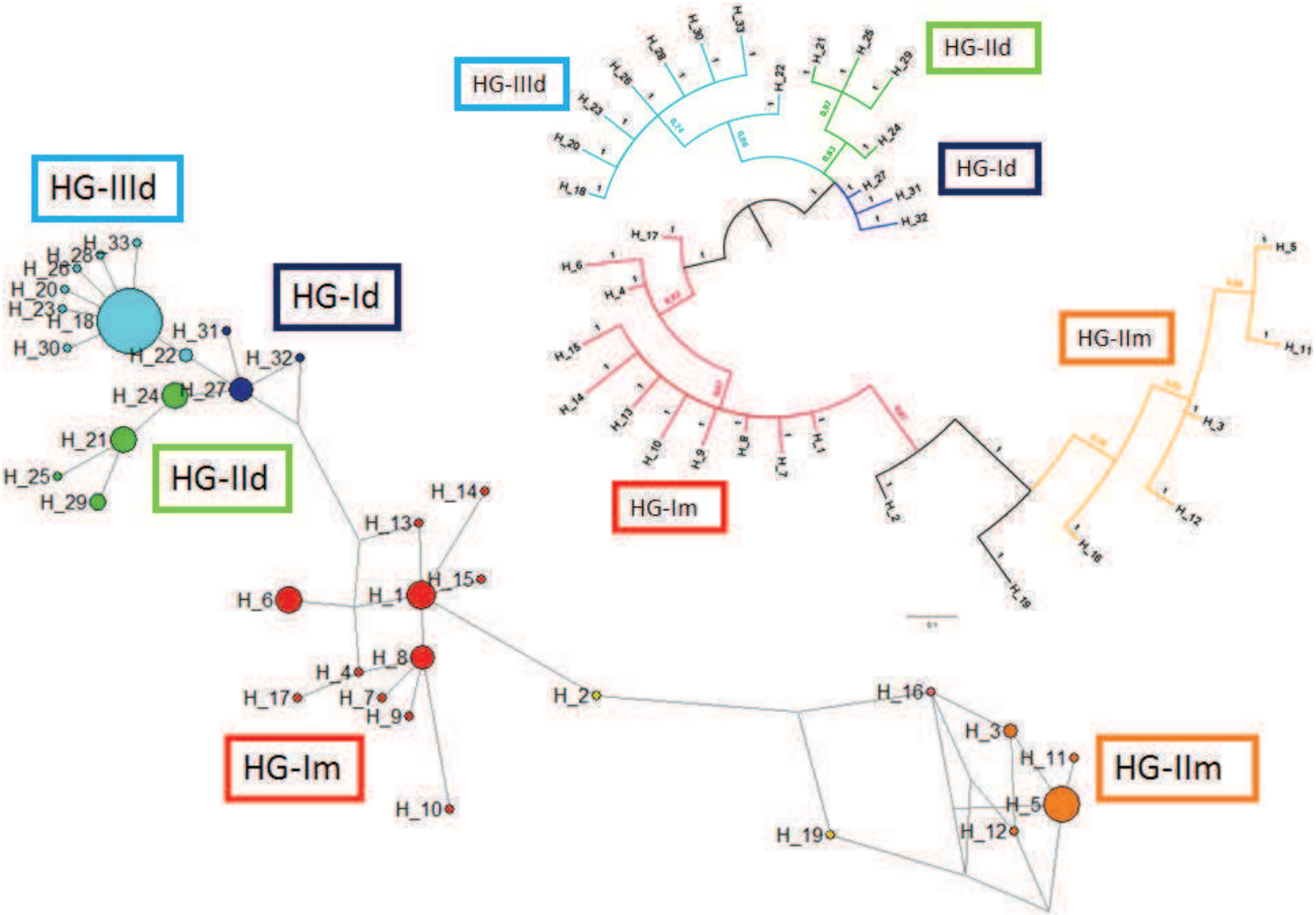


Fig. S2-b, *mt-Cytb* haplogroup definition; haplotype network reconstructed in Network v. 4.6.1.1., phylogeny taken from Fig. S 1-b, , distribution of haplotypes among sampled individuals can be found in Table S1.

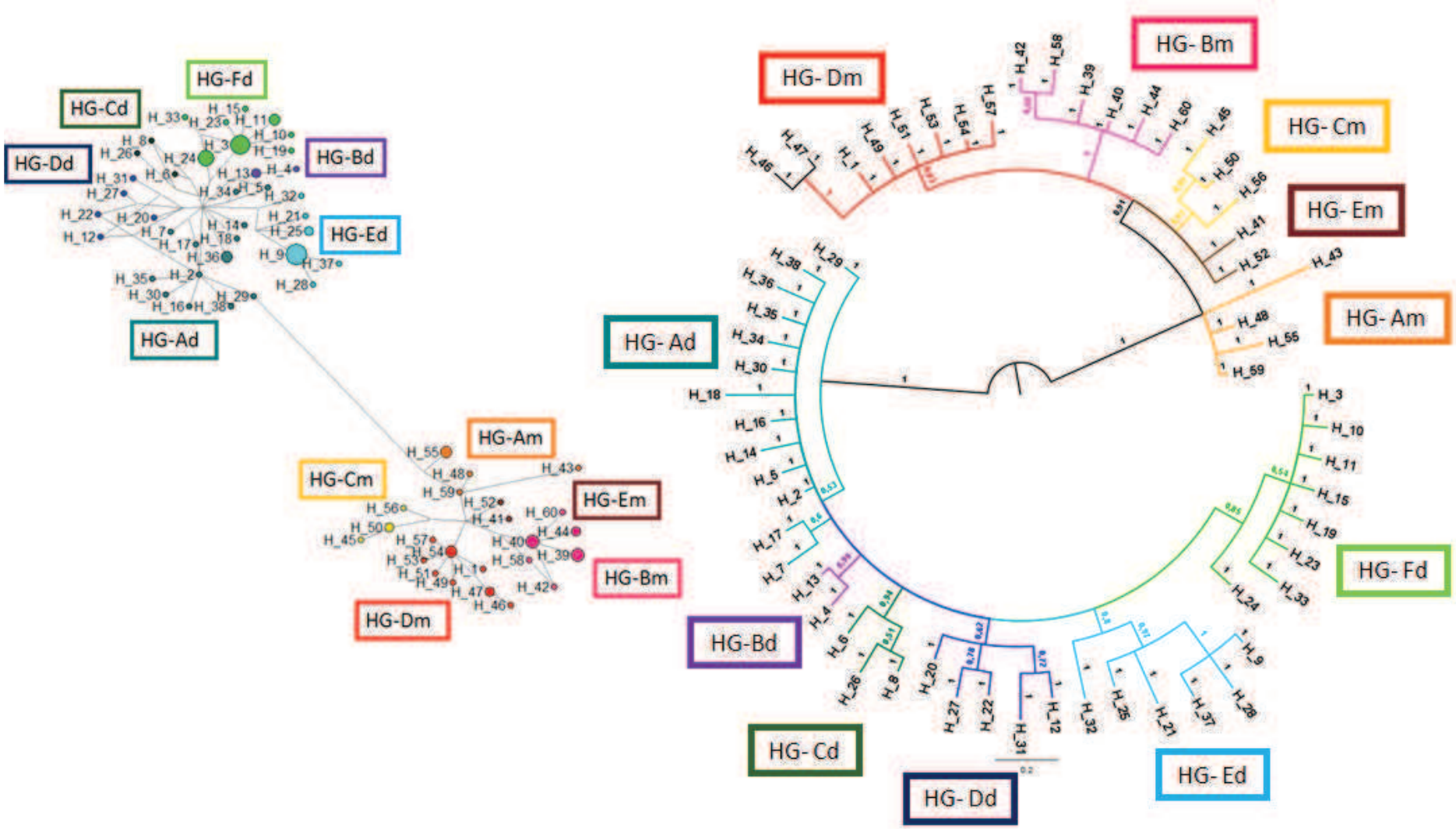


Fig. S3, Evidence of recombination between HG-Im and HG-IIm of Mmm, Breakpoints in Mmm were detected in our alignment by SBP at position 1326bp (529aa) and by GARD at position 1350 (537aa), these breakpoints are represented by the vertical black dashed line between positions 1326 and 1350. RH, presumably recombinant haplotypes (H_2 and H_19) are separated by the horizontal black dashed line from the rest of haplotypes. Varible sites were extracted by Fabox DNA collapser (Villesen 2007).

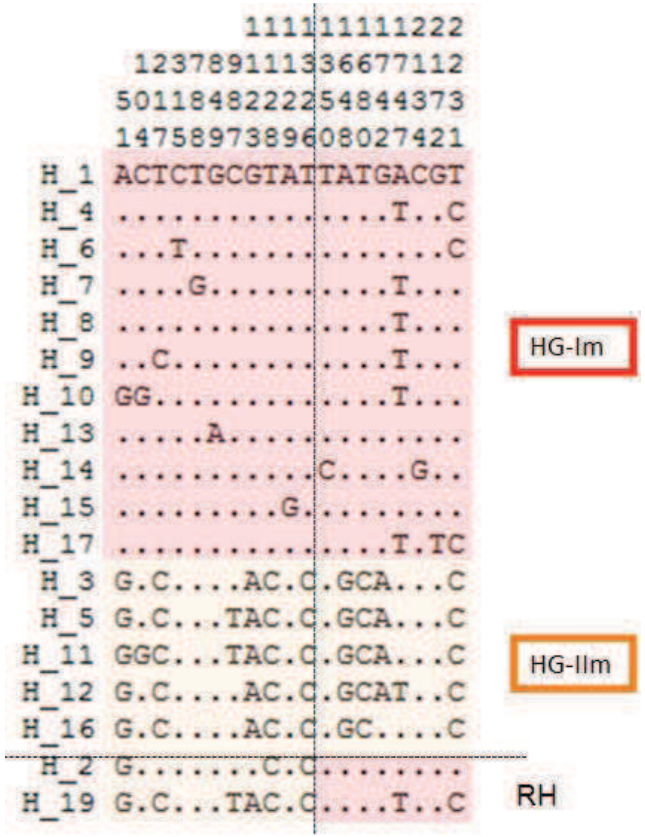


Table S2, Binding sites between TLR4/LPS/MD-2, based on knowledge of 3D-crystalography in human (predicted by Kim et al. 2007; Park et al. 2009; Resman et al. 2009; Ohto et al. 2012); substitutions with suggested key role in ligand binding interactions are in bold, substitutions with questioned role are in grey. Exon 2 was not analyzed in our study.

Position in mus	Localization	Numbering in human sequence	Function of AA	Publication
C28	exon2	hTLR4_C29	key sites for MD-2 dimerization, documented in human	Nishitani et al. 2006
C39	exon2	hTLR4_C40	key sites for MD-2 dimerization, documented in human	Nishitani et al. 2006
D41	exon2	hTLR4_E42	key sites for MD-2 dimerization	Kim et al. 2007
D83	exon2	hTLR4_D84	key sites for MD-2 dimerization	Kim et al. 2007
E134	exon3	hTLR4_E135	key sites for MD-2 dimerization	Kim et al. 2007
H158	exon3	hTLR4_H159	key sites for MD-2 dimerization	Kim et al. 2007
R233	exon3	hTLR4_R234	key sites for MD-2 dimerization	Kim et al. 2007
K263	exon3	hTLR4_R264	LPS (Charge interaction with phosphates)	Park et al. 2009
R266	exon3	hTLR4_G267	Residues Involved in the Species Specificity for Lipid IVa	Ohto et al. 2012
R288	exon3	hTLR4_R300	key sites for MD-2 dimerization	Kim et al. 2007
K319	exon3	hTLR4_E321	Residues Involved in the Species Specificity for Lipid IVa	Ohto et al. 2012
R337	exon3	hTLR4_N339	Residues Involved in the Species Specificity for Lipid IVa	Ohto et al. 2012
Q339	exon3	hTLR4_K341	LPS (Charge interaction with phosphates)	Park et al. 2009
K341	exon3	hTLR4_G343	Residues Involved in the Species Specificity for Lipid IVa	Ohto et al. 2012
K360	exon3	hTLR4_K362	LPS (Charge interaction with phosphates)	Park et al. 2009
K367	exon3	hTLR4_E369	Residues Involved in the Species Specificity for Lipid IVa	Ohto et al. 2012
S386	exon3	hTLR4_K388	LPS (Charge interaction with phosphates)	Park et al. 2009, Resman et al. 2009
S413	exon3	hTLR4_ S415	LPS, hydrogen bond with 1-PO4	Ohto et al. 2012
A414	exon3	hTLR4_ S416	MD-2 (Hydrogen bond)	Park et al. 2009, Ohto et al. 2012
N415	exon3	hTLR4_ N417	MD-2 (Hydrogen bond)	Park et al. 2009, Ohto et al. 2012
M417	exon3	hTLR4_ L419	MD-2 (Hydrophobic interaction)	Ohto et al. 2012
R434	exon3	hTLR4_Q436	LPS , MD-2 (Hydrogen bond)	Park et al. 2009
E437	exon3	hTLR4_E439	MD-2 (Hydrogen bond)	Park et al. 2009
F438	exon3	hTLR4_F440	LPS, MD-2 (Hydrophobic interaction)	Park et al. 2009, Resman et al. 2009
L442	exon3	hTLR4_L444	LPS, MD-2 (Hydrophobic interaction)	Park et al. 2009, Resman et al. 2009
S443	exon3	hTLR4_ S445	MD-2 (Hydrophobic interaction)	Ohto et al. 2012
F461	exon3	hTLR4_F463	LPS, MD-2 (Hydrophobic interaction)	Park et al. 2009, Resman et al. 2009

3.3 Analysis of variability at interspecific level of wild rodents

Annotation

Mechanisms of selection acting of TLRs are different at intraspecific and interspecific levels. In this chapter, I screened the diversity of two TLRs (bacterial sensing TLR4 and viral sensing TLR7) in 23 species from two murine lineages, with the aim to understand evolution of TLRs at interspecific level. Previous studies showed that bacterial-sensing TLRs might be more polymorphic with more signals of positive selection. This was explained by more relaxed constraints imposed by pathogens and redundant function with other PRRs and/or by different selection pressure imposed by bacterial pathogens and variability of their structures from those imposed by viruses. Comparisons of phylogenetic trees for *Tlr* genes and for nearly neutral part of mitochondrial DNA performed in this chapter were very similar, suggesting predominat action of neutral evolutionary processes. However, we also found significant signatures of positive selection in some lineages (more intensive in *Tlr4* than *Tlr7*) and these signals were concentrated particularly in protein parts responsible for antigen recognition.

CONTRASTED EVOLUTIONARY HISTORIES OF TWO TOLL-LIKE RECEPTORS (TLR4 AND TLR7) IN WILD RODENTS (MURINAE)

Alena Fornůsková [1,2,3], Michal Vinkler[4], Marie Pagès [3,5], Maxime Galan [3], Emmanuelle Jousselin [3], Frederique Cerqueira [6], Serge Morand [3,7,8], Nathalie Charbonnel [3], Josef Bryja [1,2], Jean-François Cosson [3]

[1] Institute of Vertebrate Biology, Research Facility Studenec, Academy of Sciences of the Czech Republic; [2] Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic; [3] INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus International de Baillarguet, CS 30016, 34988 Montferrier-sur-Lez Cedex, France; [4] Department of Zoology, Faculty of Science, Charles University in Prague, Czech Republic; [5] Laboratoire de génétique des microorganismes, Université de Liège, 4000 Liège, Belgique [6] Labex CeMEB, Plateforme Génotypage-Séquençage, Université Montpellier2, France, [7] ISEM, Montpellier, France, [8] CIRAD, UR AGIRs, Montpellier, France

E-mail addresses: AF, Alena.Fornuskova@supagro.inra.fr; MV, michal.vinkler@natur.cuni.cz; MP, Marie.Pages@supagro.inra.fr; MG, galan@supagro.inra.fr; EJ, jousseli@supagro.inra.fr; FC, Frederique.Cerqueira@univ-montp2.fr; SM, serge.morand@univ-montp2.fr; NCH, Nathalie.Charbonnel@supagro.inra.fr; JB, bryja@brno.cas.cz; J-FC, cosson@supagro.inra.fr

Published in BMC Evolutionary Biology 2013, 13 :194

Abstract

Background: In vertebrates, it has been repeatedly demonstrated that genes encoding proteins involved in pathogen-recognition by adaptive immunity (*e.g.* MHC) are subject to intensive diversifying selection. On the other hand, the role and the type of selection processes shaping the evolution of innate-immunity genes are currently far less clear. In this study we analysed the natural variation and the evolutionary processes acting on two genes involved in the innate-immunity recognition of pathogen-associated molecular patterns (PAMPs).

Results: We sequenced genes encoding Toll-like receptor 4 (*Tlr4*) and 7 (*Tlr7*), two of the key bacterial- and viral-sensing receptors of innate immunity, across 23 species within the subfamily Murinae. Although we have shown that the phylogeny of both *Tlr* genes is largely congruent with the phylogeny of rodents based on a comparably sized non-immune sequence dataset, we also identified several potentially important discrepancies. The sequence analyses revealed that major parts of both *Tlr*s are evolving under strong purifying selection, likely due to functional constraints. Yet, also several signatures of positive selection have been found in both genes, with more intense signal in the bacterial-sensing *Tlr4* than in the viral-sensing *Tlr7*. 92% and 100% of sites evolving under positive selection in *Tlr4* and *Tlr7*, respectively,

were located in the extracellular domain. Directly in the Ligand-Binding Region (LBR) of TLR4 we identified two rapidly evolving amino acid residues and one site under positive selection, all three likely involved in species-specific recognition of lipopolysaccharide of gram-negative bacteria. In contrast, all putative sites of LBR_{TLR7} involved in the detection of viral nucleic acids were highly conserved across rodents. Interspecific differences in the predicted 3D-structure of the LBR of both *Tlr*s were not related to phylogenetic history, while analyses of protein charges clearly discriminated Rattini and Murini clades.

Conclusions: In consequence of the constraints given by the receptor protein function purifying selection has been a dominant force in evolution of *Tlr*s. Nevertheless, our results show that episodic diversifying parasite-mediated selection has shaped the present species-specific variability in rodent *Tlr*s. The intensity of diversifying selection was higher in *Tlr4* than in *Tlr7*, presumably due to structural properties of their ligands.

Keywords: Arms race, host-pathogen interaction, pattern recognition receptors, adaptive evolution, pathogen-associated molecular pattern (PAMP)

3.3.1 Introduction

An effective immune defence is dependent on well-timed activation of an appropriate immune response. Pathogen recognition by innate immunity Pattern Recognition Receptors (PRRs) is crucial in this process (Zak and Aderem 2009; Barreiro and Quintana-Murci 2010). The PRRs detect molecular structures named pathogen-associated molecular patterns (PAMPs) that are conservatively present among individual microorganism taxa, because they are essential for their survival (such as, *e.g.*, bacterial lipopolysaccharides, muramyl dipeptide, peptidoglycan, flagellin, mannose, bacterial, fungal, parasitic and viral nucleic acids) (Akira et al. 2006). Recent studies have associated polymorphism in genes encoding PRRs with variability in resistance or susceptibility to several infectious diseases in humans, laboratory mice and poultry (e.g. Schröder and Schumann 2005; Pandey and Agrawal 2006; Bochud et al. 2007; Loo and Gale Jr. 2011; Netea et al. 2012). However, in wildlife, molecular variation in PRR genes is still poorly documented (Wlasiuk and Nachman 2010; Alcaide and Edwards 2011; Tschirren et al. 2011; Grueber et al. 2012; Tschirren et al. 2012; Tschirren et al. 2013).

Understanding the evolution of the immune system in general has been a challenge for evolutionary biologists and ecologists since JBS Haldane associated natural selection with

infectious diseases (Haldane). In vertebrates, the study of selection patterns was mostly oriented towards genes of acquired immunity which are now intensively studied even in wild populations. Among them, genes of the major histocompatibility complex (MHC) are the most explored and the role of balancing selection in their evolution is generally accepted and well understood (Apanius et al. 1997; Bernatchez and Landry 2003; Aguilar et al. 2004; Bryja et al. 2006; Piertney and Oliver 2006; Spurgin and Richardson 2010; Čížková et al. 2011; Smith et al. 2011). The quite late discovery of genes involved in the second branch of vertebrate immunity, *i.e.* innate immunity, among which the most important PRRs are Toll-like receptors (hereafter abbreviated according to the mouse gene and protein nomenclature as *Tlr*s and TLRs, respectively) (Medzhitov et al. 1997; Hedrick 2004; O'Neill 2004; Bassett and Rich), has resulted in modest research of their evolution in wildlife populations (Acevedo-Whitehouse and Cunningham 2006).

Generally, two subclasses of TLRs are distinguished in vertebrates according to the ligands they target (Akira et al. 2006; Barreiro et al. 2009; Vinkler and Albrecht 2009; Wlasiuk and Nachman 2010). The first subclass includes TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10. These TLRs predominantly detect bacterial components (but also fungal and to lesser extent viral components) and are expressed on the outer cell membrane. Throughout this paper we term them "bacterial-sensing" TLRs. The second subclass includes TLR3, TLR7, TLR8 and TLR9 and targets mainly viral components (*e.g.* ssRNA, dsRNA, DNA containing unmethylated CpG), hereafter termed "viral-sensing" TLRs. These TLRs are expressed mostly within cells into the membranes of endosomal compartments. This current spectrum of genes for TLRs arose by multiple gene duplication and during the last 700 Mya diversified to recognize distinct PAMPs (Janssens and Beyaert 2003; Roach et al. 2005; Hughes and Piontkivska 2008; Leulier and Lemaitre 2008; Temperley et al. 2008; Barreiro et al. 2009; Huang et al. 2011).

TLRs of both subclasses are transmembrane proteins composed of three domains (Leulier and Lemaitre 2008; Werling et al. 2009). The extra-cellular domain (ECD) consists of a varying number of leucin-rich repeat motifs (LRRs) that form a horseshoe-shaped tertiary structure of the ECD. This domain contains the ligand binding region (LBR) which is directly responsible for physical interactions with the pathogen-derived structures and as such it is likely subject to intensive selection. The ECD is followed by a short transmembrane domain (TMD), and an intracellular domain (ICD) containing the Toll/Interleukin-1 Receptor (TIR) domain responsible for TLR signaling (Akira et al. 2006). As previously shown (Burke et al. 2007), non-synonymous SNPs located in LBR may affect the 3D structure of the protein and

its surface charge. This may have important functional consequences, influencing receptor ability to bind pathogens (Park et al. 2009; Huang et al. 2011; Tschirren et al. 2013), and may even lead to the evolution of species-specific ligand recognition (Keestra and van Putten 2008; Walsh et al. 2008). Appropriate binding of PAMPs by LBR is connected with changes in receptor dimerization (Zhu et al. 2009; Botos et al. 2011; Kang and Lee 2011) that induce signaling and release of cytokines triggering mainly Th1 and Th17 inflammation, fever and phagocytosis (Pasare and Medzhitov 2003; Parker et al. 2007; Kawai and Akira 2010). The TLR signaling ensures an immediate response to invading microorganisms that, in a second step, further directs the following adaptive immune response (Pasare and Medzhitov 2004b; Netea et al. 2005).

Previous studies, mostly based on investigation in humans, primates and domestic or laboratory animals, provided information regarding some general patterns of TLR evolution and maintenance of their genetic polymorphism (Smirnova et al. 2000; Ferwerda et al. 2007; Vinkler et al. 2009; Barreiro and Quintana-Murci 2010; Wlasiuk and Nachman 2010). These studies revealed that the ECD is more frequently a target of positive selection than the TIR domain. Moreover, in general the viral-sensing TLRs seem to evolve under stronger purifying selection than the bacterial-sensing ones (Krieg and Vollmer 2007; Barrat and Coffman 2008; Waldner 2009; Mikami et al. 2012). However, up to now, the evidence of TLR polymorphism and the type of selection that shapes this polymorphism in natural populations remain rare (Alcaide and Edwards 2011; Tschirren et al. 2011; Grueber et al. 2012; Tschirren et al. 2012; Tschirren et al. 2013). Besides, to our knowledge the precise investigation of the LBR variability and evolution is missing. Such information could nevertheless be important to better understand species-specific differences in the susceptibility to various pathogens (Worobey et al. 2007).

In the present study we focused on the molecular variation of the genes encoding the bacterial-sensing TLR4 (binding mainly bacterial lipopolysaccharides, LPS, as a ligand) (Poltorak et al. 1998) and the viral-sensing TLR7 (binding viral ssRNA) (Diebold et al. 2004; Heil et al. 2004) in 23 species of the subfamily Murinae. Murine rodents are largely distributed over the world and several species (such as rats and mice) live in close proximity to humans. A recent review showed that 60% of the agents of emerging diseases in humans circulate in animals (Jones et al. 2008) and most of the natural reservoirs of a number of serious viral and bacterial emerging agents of zoonoses are rodents (Mills 2006; Luis et al. 2013). Species-specific molecular variability in immune-related genes may be responsible for differences in the ability of rodent species to transmit these pathogens. Herein we aimed to

document evolutionary histories of these two *Tlr*s during murine diversification. We implemented statistical approaches to infer *Tlr* phylogeny and to detect selection acting on DNA and amino acid (AA) sequences. We searched for deviations from "species" phylogeny based on a comparably sized non-immune sequence dataset by contrasting phylogenetic trees reconstructed from *Tlr* sequences with those reconstructed from "neutral" genes (both mitochondrial and nuclear). Deviations would indicate the occurrence of non-neutral patterns during the *Tlr* evolutionary history, *e.g.* adaptive selection (Wlasiuk and Nachman 2010; Fornarino et al. 2011; Govindaraj et al. 2011). Next we estimated putative functional changes in the LBR by examining variability in predicted tertiary 3D-structures of the proteins, and in biophysical properties of proteins (charge and structural characteristics) at polymorphic binding sites. Finally, we compared the evolutionary histories of the two TLRs to reveal potentially distinct evolutionary pressures shaping these proteins.

3.3.2 Materials and methods

Sampling

Murine rodents from 23 species belonging to the Rattini and Murini (sensu Lecompte et al. (2008)) tribes were sampled mainly in South-East Asia, and three synanthropic species (*i.e. Rattus rattus*, *Mus m. muscululus* and *Mus m. domesticus*) were also sampled in Europe and Africa. In our sampling area, *Rattus tanezumi* specimens corresponded to two divergent mitochondrial lineages although they could not be distinguished according to their nuclear pool (Pagès et al. 2013). These samples were further referred to clades *R. tanezumi R2* and *R3* according to their mitotype. *Rattus sakeratensis* corresponds to the lineage previously referred to as *R. losea* and found in central, northern Thailand and Vientiane Plain of Lao PDR (*Rattus losea*-like by Pagès et al. (2010)). This lineage was recently distinguished from the true *R. losea*, which is restricted to Cambodia, Vietnam, China and Taiwan (Aplin et al. 2011).

Species identification was initially based on morphological criteria and thereafter confirmed using molecular barcoding for problematic lineages (Pagès et al. 2010; Galan et al. 2012). We sequenced two to 10 individuals per species. In total 103 specimens were analysed (Table S1 in SUPPLEMENTARY MATERIALS).

Toll-like receptor sequencing and sequence alignments

We sequenced the complete exon 3 of *Tlr4* (2.250 bp) and *Tlr7* (3.150 bp) as it encompasses the LBR in both genes. Exon 3 corresponds to 89.7% and 99.0% of the total coding sequence for *Tlr4* and *Tlr7*, respectively. Short exons 1 and 2 (241 bp encoding 5′-untranslated (UT) region and first 257 bp of ECD in $Tlr4_{exon2}$ and 154 bp of 5′-UT regions and 3bp of ECD in $Tlr7_{exon2}$) were not analysed in present study, because we were preferentially interested by functional regions (*e.g.* LBR and TIR). For all analyses and discussion the codon numbering follows the sequences of *Rattus norvegicus* available in GenBank [GenBank Acc. NP_062051.1 for *Tlr4*, and NP_001091051.1, for *Tlr7*].

Primers for Polymerase Chain Reaction (PCR) and sequencing were designed according to the sequences available in the Ensembl database for *Mus musculus* [*Tlr4* ENSMUSE00000354724/MGI:96824, *Tlr7* ENSMUSE00000405820/ MGI:2176882] and *Rattus norvegicus* [*Tlr4* ENSRNOE00000099045/NP_062051, *Tlr7* ENSRNOE00000039897/NP_001091051]. We used the software PRIMER3 (Rozen and Skaletsky 2000) to design primers (see their sequences in Table S2 and positions in Figure S1 in Additional files 1). Total DNA was extracted from rodent tissue (biopsy from ear or necropsy from liver) using the DNeasy Blood & Tissue Kit (Qiagen AB, Hilden, Germany). Amplifications were carried out in a final volume of 25 µl containing 12.5 µl of Multiplex Kit PCR master mix (Qiagen), 9.3 µl of H_2O, 0.5 µM of each of primer pairs and 2 µl of DNA. Cycling conditions included an initial denaturation at 95°C for 15 min, followed by 10 cycles of denaturation at 95°C for 40 s, annealing with touchdown at 65°C to 55°C (-1°C/cycle) for 45 s and extension at 72°C for 90 s, followed by 30 cycles of denaturation at 95°C for 40 s, annealing at 55°C for 45 s and extension at 72°C for 90 s, with a final extension phase at 72°C for 10 min. The final extension was performed for 10 min at 72°C. The lengths of amplicons were checked on 1.5 % agarose gels. Sequencing was carried out using an ABI3130 automated DNA sequencer (Applied Biosystems). DNA sequences were aligned and edited using SEQSCAPE v.2.5 (Applied Biosystems) and BIOEDIT v.7.1.3 (Hall 1999). All sequences have been submitted to NCBI GenBank (Accession numbers are presented in Table S1 in SUPPLEMENTARY MATERIALS).

Sequence analysis

Diploid genotypes were resolved using the Bayesian PHASE platform (Stephens and Donnelly 2003) implemented in DnaSP ver. 5.10 (Librado and Rozas 2009). Calculations were carried out using 1000 iterations, 10 thinning intervals, and 1000 burn-in iterations. Sequences were collapsed into individual alleles by Fabox DNA collapser, an online FASTA sequence toolbox (Villesen 2007). The identification and visualization of main domains (ECD, TM and ICD with TIR domain and ICD-DP) was performed in SMART (Schultz et al. 1998) based on *Rattus norvegicus* sequences provided in GenBank [NP_062051.1 for *Tlr4* and NP_001091051.1 for *Tlr7*]. 3D structure was predicted in PHYRE2 (Kelley and Sternberg 2009) and then visualized using FirstGlance in Jmol v.1.9. Finally, we estimated nucleotide diversity (π), number of polymorphic sites (S) and total number of mutations (ε) with DnaSP, and the number of nucleotide alleles (hN) and amino acid variants (hA) using Fabox DNA collapser.

Phylogenetic reconstructions and congruence between the tree based on a comparably sized non-immune sequence dataset and Tlr *trees*

We first tested *Tlr* sequences for recombination using SBP, to avoid further false positive events of selection. This method (implemented in DATAMONKEY, Pond and Frost 2005; Delport et al. 2010) allowed the screening of *Tlr* sequences for recombination breakpoints. SBP identify non-recombinant regions and allowed each region to have its own phylogenetic reconstruction (Pond et al. 2006b; Pond et al. 2006c).

Phylogenies were reconstructed independently for each gene using the alignment of complete exon 3 sequences. A phylogeny inferred from the combination of one nuclear (the first exon of the gene encoding the interphotoreceptor retinoid binding protein, *Irbp*) and two mitochondrial genes (the cytochrome *b* gene, *Cytb*, and the cytochrome *c* oxidase I gene, *Co*I), taken from Pagès et al. (Pagès et al. 2010), was used for comparison of "neutral" evolution of the studied rodents with trees obtained from the immune gene alignments. Both Maximum likelihood (ML) and Bayesian (BA) methods were applied to infer phylogenetic relationships from each *Tlr* alignments. The best evolutionary model of nucleotide substitution was determined using jModelTest 0.1.1 (Posada 2008). Phylogenies based on ML analyses were reconstructed using RAxML 7.2.6 (Stamatakis et al. 2008). Analyses were run

as the rapid bootstrap procedure (option –f a) with bootstraps defined by option –NautoMR. For both *Tlr*s we used nucleotide substitution model GTR + Γ (option –m GTRGAMMA) selected by jModelTest 0.1.1 as the most appropriate to our data. Bayesian analyses were performed using a parallel version of MrBayes v3.1 (Huelsenbeck and Ronquist 2001) at the University of Oslo Bioportal (Kumar et al. 2009) and CBGP HPC computational platform located at Centre de Biologie et Gestion des Populations, Montpellier. Two runs of 50,000,000 generations in each were adopted, applying the best fitted model of substitution (GTR+ Γ). A burn-in period of 10,000,000 generations was determined using Tracer v1.4 (Rambaut and Drummond 2007). Convergence was also evaluated using Tracer v1.4. After discarding samples from the burnin period, results were based on the pooled samples from the stationary phases of the two independent runs. Trees were edited using FigTree v1.3.1. (Rambaut 2009).

We tested the congruence between the rodent phylogeny and the *Tlr*s phylogeny based on the MrBayes approach using reconciliation analyses. Reconciliation analyses explore all possible mappings of one tree onto another, assigning different costs to evolutionary events and find optimal (*i.e.* yielding minimal costs) solutions. These analyses were conducted using JANE 4 (Conow et al. 2010). This software was initially built to reconcile parasite and host trees, yet it can also be used for comparative analysis of species and gene trees. In the context of host-parasite relationships, five evolutionary events between parasites and host can be taken into account in JANE 4: co-speciation, host switches, duplication, failure to diverge and parasite loss. These events are analogous to co-divergence, convergence, duplication, purifying selection and gene loss (respectively) when considered in the context of species and gene tree reconciliation. For each of these events the specific costs can be set. The lowest cost is attributed to the event considered as most likely. In order to obtain reconciliations that maximize the number of co-divergences we set the cost of a co-divergence event to 0 while other costs were set to 1 (see Cruaud et al. 2012 for similar approach). The cost of the best solution is then compared with costs found in reconciliations in which tip mappings are permuted at random. This generates a null distribution of the costs of reconciliation. If the cost of the best solution is lower than that expected by the chance it means that the two phylogenies are significantly congruent. The following parameters were used: the number of generations (iterations of the algorithm) was set to 100 and the "population" (number of samples per generation) was set to 100. Input phylogenies were those obtained by the Bayesian inference. The cost of the best solution was compared to distribution of the costs of 1000 randomizations.

Moreover, we tested the congruence between genes and tree based on a comparably sized non-immune sequence dataset using SH test (Shimodaira and Hasegawa 1999) as implemented in PAUP. Alternative topologies required for ML SH test were reconstructed by ML approach in the software GARLI v. 2.0 (Zwickl 2006). Two different ML trees were estimated for each *Tlr*; a first one inferred under non-constrained conditions with default options and a second one constrained by the tree topology based on a comparably sized non-immune sequence dataset. Mouse species (genus *Mus*) were excluded from the analysis of co-divergence in order to match data with the study of Pagès et al. (Pagès et al. 2010) where the mice are missing.

Search for signatures of selection on Tlr *sequences*

We estimated separately the number of synonymous (*dS*) and non-synonymous (*dN*) substitutions per site for the whole exon 3, ECD, LBR and the TIR domains, and for both *Tlr*s. Computations were made with 1,000 bootstraps and Nei-Gojobori method (with Jukes-Cantor correction) in MEGA 5 (Tamura et al. 2011). We then estimated the overall ratio *dN/dS* for each domain and for the whole exon 3 of both *Tlr*s by Single Likelihood Ancestor Counting (SLAC) implemented in DATAMONKEY. The p-value was 0.05. As the SLAC method tends to be a very conservative test, the actual rate of false positives (*i.e.* neutrally evolving sites incorrectly classified as selected) can be much lower than the significance level (Murrell et al. 2012). In the next step we estimated selection at each codon by SLAC to find which codons of the exons 3 have been subject to positive and negative selection. As a default tree we used a NJ tree and appropriate substitution model proposed by automatic model selection tool in DATAMONKEY.

Finally, we used the Mixed Effects Model of Evolution (MEME) algorithm in the HYPHY package accessed on the website of DATAMONKEY interface (Delport et al. 2010) to detect codons evolved under positive selection along the branches of the phylogenies. This method is recently recommended as a replacement for the Fixed Effects Likelihood (FEL) and SLAC models (Murrell et al. 2012). It allows the detection of signatures of episodic selection, even when the majority of lineages are subject to purifying selection. This test permits ω to vary from site to site and also from branch to branch in phylogeny (Murrell et al. 2012). Tests of episodic diversifying selection were performed at significance level $p < 0.05$ and MrBayes trees were used as working topologies. Only events of positive selection with Empirical Bayes Factor (EBF) estimated by MEME near to 100 were mapped on to the phylogeny.

Functional analysis of ligand binding region

Positions of LBR in both TLRs have been previously described in humans (Park et al. 2009; Wei et al. 2009). The corresponding LBR position in rodents was predicted based on the human-rodent alignment. The LBR was located between codons AA248 and AA469 in TLR4 and between codons AA495 and AA597 in TLR7.

We first explored the evolutionary conservation of each amino acid position in LBR using the CONSURF algorithm (Ashkenazy et al. 2010). CONSURF estimates the evolutionary rate of amino acid positions in a protein molecule, based on the phylogenetic relationships between homologous sequences. Conservation scale is defined from the most variable amino acid positions (grade 1, color represented by turquoise) which are considered as rapidly evolving to conservative positions (grade 9, color represented by maroon) which are considered as slowly evolving. We used the proposed substitution matrix and computation was based on the empirical Bayesian paradigm. MrBayes trees were used as the working topology. Protein tertiary structure was adopted from *R. norvegicus* [Gene Bank Acc. TLR4/KC811688 and TLR7/KC811786].

Because protein tertiary structure is essential for its biological function we finally explored the variability in the 3D structures of LBRs in the different AA variants. The prediction of 3D structures of the variants was performed by homology modeling using PHYRE2 (Kelley and Sternberg 2009). Differences in 3D protein structure among variants were then evaluated using the root mean square deviations (RMSD) calculated by the DALI pairwise comparison tool (Holm et al. 2008). The RMSD-based distance matrices were analysed in STATISTICA v. 8.0 (StatSoft, Inc., Tulsa) by joining tree clustering using Unweighted Pair Group Method with Arithmetic Mean (UPGMA, Kalinowski 2009). We then analysed the variability of the charge of each LBR variant, which could be another key indicator of functional changes, because differences in protein charge could influence the ability to bind ligands (Walsh et al. 2008; Govindaraj et al. 2011). LBR charge of each variant was estimated at predefined neutral pH = 7 using LRRFINDER (Offord et al. 2010).

3.3.3 Results

Sequence analyses

Amplification and sequencing were successful in 96 samples representing 23 rodent species for *Tlr4* and in 96 samples representing 22 species for *Tlr7* (Table S1 in Additional files 1). Only samples from one species - *Maxomys surifer* could not be completely sequenced for *Tlr7* - the first 180 bp were missing and we excluded this species from the *Tlr7* analyses. No stop codons, indels nor recombination were detected in these data using SBP (DATAMONEKY).

For the whole *Tlr4* coding sequence (CDS), the three different domains were predicted by SMART as follows: ECD from AA position 1 to 635, TM from position 636 to 658 and ICD from position 659 to 835 in which the TIR domain (from position 671 to 816) and ICD distal part (ICD-DP; from 817 to 835) may be identified (Fig. S1 in SUPPLEMENTARY MATERIALS). For *Tlr7*, the predicted location of the three domains was the following: ECD from position 1 to 850, TM from position 851 to 873 and ICD from position 874 to 1050 (TIR from 894 to 1033 and ICD-DP from 1034 to 1050; Fig. S1). In general, *Tlr4* was more diverse than *Tlr7*, and within each *Tlr*, the ECD domain was more variable than the TIR domain in both molecules (Table 1). Surprisingly, ICD-DP located on the C-terminal end of *Tlr4* represented the most variable region of exon 3 ($\pi_{\text{ICD-DP-}Tlr4}$ = 0.102±0.015).

***Tlr* domains**	***n***	***L***	***π*±S.E.**	***hN***	***hA***	***S***	***Eta***	***dN*±S.E.**	***dS*±S.E.**	***dN/dS***
Tlr4										
Exon 3	96	2247	0.049±0.003	122	90	545	625	0.038±0.003	0.102 ±0.008	**0.481**
ECD	96	1647	0.053±0.003	112	83	441	504	0.045±0.004	0.098 ±0.009	**0.597**
LBR	96	666	0.072±0.006	67	50	203	242	0.070±0.008	0.108±0.015	**0.787**
TIR	96	435	0.031±0.002	54	11	68	79	0.004±0.002	0.143 ±0.024	**0.067**
Tlr7										
Exon 3	96	3147	0.034±0.003	79	49	466	518	0.021±0.002	0.088 ±0.007	**0.398**
ECD	96	2547	0.037±0.003	75	48	407	455	0.025±0.002	0.089±0.007	**0.468**
LBR	96	311	0.035±0.003	19	8	37	38	0.018±0.006	0.107±0.024	**0.196**
TIR	96	420	0.026±0.003	26	6	43	47	0.007±0.003	0.105±0.021	**0.070**

Table 1. Estimates of sequence diversity and average codon-based evolutionary divergence over all sequence pairs for the exon 3 and particular domains of *Tlr4* and *Tlr7* genes, **ECD** - extracellular domain; **LBR** - ligand biding region; **TIR** - Toll/interleukin-1 receptor domain; ***n*** - the number of sequenced individuals; ***L*** - length of analysed sequences in base pairs; ***π*** - average number of nucleotide differences per site between two sequences; ***S.E.*** - Standard error; ***hN*** - number of nucleotide alleles; ***hA*** - number of amino acid variants; ***S*** - number of polymorphic sites; ***Eta*** - total number of mutations; ***dS*** - number of synonymous substitutions per synonymous site (estimated by MEGA); ***dN*** - number of non-synonymous substitutions per non-synonymous site (estimated by MEGA). Analyses were conducted using the Nei-Gojobori model; ***S.E.*** of *dN* and *dS* - were obtained by a bootstrap procedure (1000 replicates); ***dN/dS*** were computed by SLAC (DATAMONKEY).

Phylogeny and co-divergence between the tree based on a comparably sized non-immune sequence dataset and TLR trees

Both phylogenetic approaches (MrBayes and RAxML) displayed similar trees for both *Tlr*s (Figs. S2 and S3 in SUPPLEMENTARY MATERIALS). Minor differences between ML and Bayesian trees were found only at the intraspecific level. *Tlr4* topology was well-supported with posterior probabilities (pp) $\geq$ 0.95 despite a lack of resolution within the black rat species complex (including *Rattus rattus*, *R. tanezumi*, *R. sakeratensis*, *R. tiomanicus*, *R. argentiventer*, *R. andamanensis*), between two *Bandicota* species (*Bandicota savilei* and *B. indica* did not form reciprocal monophyletic clades) and between two subspecies of the house mouse (Figs. S2a and S3a in Additional files 2). Sequences of *Tlr7* were also predominantly clustered according to species with strong supports (pp $\geq$ 0.95). Relationships between Asiatic mouse species were not fully resolved (monophyly of *Mus caroli*, *M. cooki* and *M. cervicolor* supported with a moderate pp value of 0.86 and Bootstrap values, Bp = 81) as well as those between *Leopoldamys* species (*L. edwardsi* appeared more closely related to *L. neilli*, rather than to *L. sabanus* but with a low pp of 0.6, Bp = 48). Similarly to *Tlr4*, branching orders within the genus *Rattus* were not resolved: *Rattus exulans* (clade I) was retrieved monophyletic without ambiguity (pp = 1, Bp = 100), *R. norvegicus* and *R. nitidus* were grouped together with the highest support (clade II, pp = 1, Bp = 100) and the remaining *Rattus* species formed a moderately supported group (clade III, pp = 0.7, Bp = 98, for more details see Figs. S2b and S3b in SUPPLEMENTARY MATERIALS).

At the first glance, *Tlr* phylogenies (based on MrBayes approach) of the black rat complex was congruent to the the tree based on a comparably sized non-immune sequence dataset (Fig. 1). The number of co-divergence events inferred using JANE 4 was significantly higher than expected by chance, meaning that the two phylogenies were similar (Fig. S4 in Additional files 1). However, the Shimodaire-Hasegawa test showed significant disagreement between the species tree and both *Tlr*s phylogenies (Δln L = 257, ddl = 1, $p < 0.001$ for *Tlr4*; Δln L = 76, ddl = 0.008, $p < 0.05$ for *Tlr7*), indicating that neither of the *Tlr* trees coincided precisely with the tree based on a comparably sized non-immune sequence dataset. The incongruence was mainly caused by recently diverged species of *Rattus*. However, we revealed several other differences, such as the misplacement of the genus *Bandicota* (occurring within *Rattus* in the *Tlr4* tree) and the different positions of *R. sakeratensis* and *R. exulans* in species and *Tlr7* trees (Fig. 1).

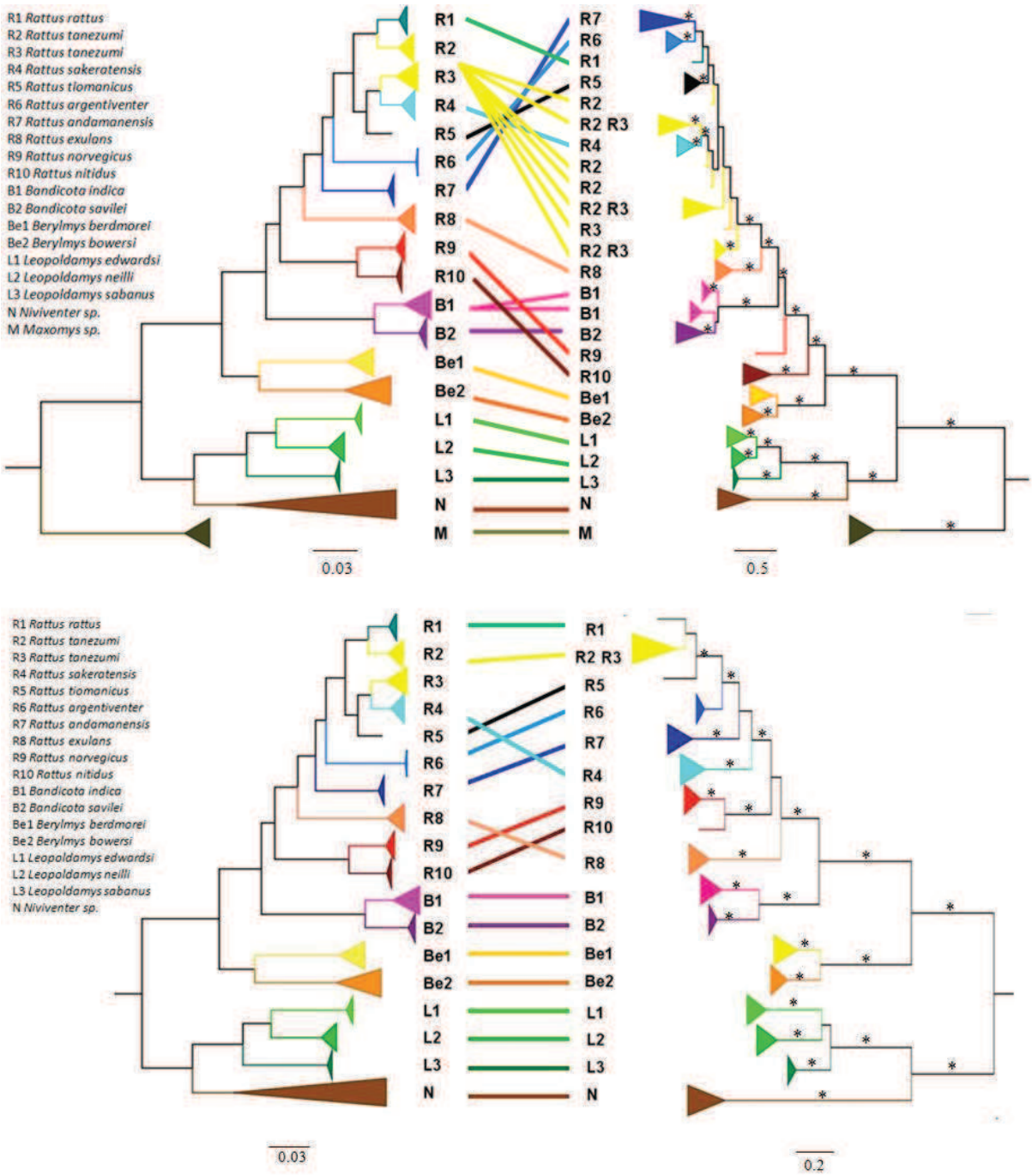


Fig. 1, Comparison of phylogenetic trees based on *Tlr*s and neutral markers,comparison of the Bayesian phylogenetic trees of *Tlr4* (a) and *Tlr7* (b) on the right with phylogenetic trees based on presumably neutral markers (*Cytb*, *Co*I, *Irbp*; for more details see Pagès et al. 2010) on the left. Abbreviations (R1, R2….M) indicate species assignment used in Pagès et al. 2010; corresponding legend is on the left. Color lines link the supported clades represented by the same species; * indicates posterior probabilities (pp) > 0.95.

Evidence of signatures of selection

The comparison of ω (*dN/dS*) revealed substantial differences between the two *Tlr*s, as well as between gene parts encoding different domains (for details see Table 1). The difference between gene parts was mainly due to variations in the number of non-synonymous substitutions (which was higher in ECDs than in the TIR), while they both had similar numbers of synonymous substitutions.

The highly conservative SLAC (Single Likelihood Ancestor Counting) analysis (DATAMONKEY) (Pond and Frost 2005) revealed two codon positions evolving under positive selection in *Tlr4* and only one in *Tlr7*, all of them being located within the ECD domain ($p < 0.05$, Table 2, Fig. 2). We found 26 and 10 negatively selected sites for *Tlr4* and *Tlr7* respectively ($p < 0.05$, Table 2, Fig. 2), distributed evenly over the whole sequences.

The imprint of natural selection on protein coding gene is often difficult to reveal because selection is frequently episodic (*i.e.* it affects only a subset of lineages) (Murrell et al. 2012). We therefore looked for evidence of episodic diversifying selection at individual sites along the evolutionary branches of the trees using the MEME algorithm. Thirteen codon positions were found to be affected by episodic selection for *Tlr4* (1.7% of all analysed codons) while only 4 codon positions showed this signature for *Tlr7* (0.38% of all analysed codons). In *Tlr4*, 12 of these sites were located directly in LBR, while in *Tlr7* none of the sites evolving under positive selection were in LBR. Whatever the *Tlr* gene considered, all sites found to evolve under positive selection using the SLAC were identified also by the MEME algorithm.

The signs of positive selection were scattered over whole *Tlr* trees, affecting nearly all branches of the *Tlr4* phylogeny, both basal and terminal, while they mostly concerned the terminal branches for the *Tlr7* phylogeny (Fig. 3). Interestingly, one site evolving under positive selection ($p < 0.05$) was located in the ICD-DP of *Tlr4* gene (Table 2, Fig 2a). We found that this part (*i.e.* the last 57bp of C-terminal end of the protein following the TIR domain) was highly variable (19 nucleic acid alleles and 16 AA variants) with a mean ω = 1.11.

	ECD (and LBR)	TD	TIR	ICD-DP
Tlr4	**1-635 (248-469)**	**636-658**	**671-816**	**817-835**
MEME	273, 335, 345, 347, 361, 363, 366, 368, 394, 398, 442, 469	-	-	818
SLAC-PS	347, 469			
SLAC-NS	99, 105, 149, 240, 253, 364, 457, 461, 463, 518, 522, 529, 549, 616, 635	-	679, 688, 691, 721, 740, 772, 782, 785, 793, 811	822
Tlr7	**1-850 (495-597)**	**851-873**	**894-1033**	**1034-1050**
MEME	128, 308, 461, 772	-	-	-
SLAC-PS	308			
SLAC-NS	156, 272, 455, 528, 541, 671, 676, 709,	-	963, 971	-

Table 2. Positively (MEME and SLAC-PS) and negatively (SLAC-NS) selected sites detected for the exon 3 of *Tlr4* and *Tlr7* at $p < 0.05$, **ECD** - extracellular domain; **LBR** - ligand binding domain; **TD** - transmembrane domain; **TIR** -TIR domain; **ICD-DP** - distal part of intracellular domain. Prediction of domains and numbering of sites are according to the reference protein sequence of *Rattus norvegicus* taken from GenBank [NP_062051.1 for *Tlr4* and NP_001091051.1 for *Tlr7*]. Sites located in LBR are underlined.

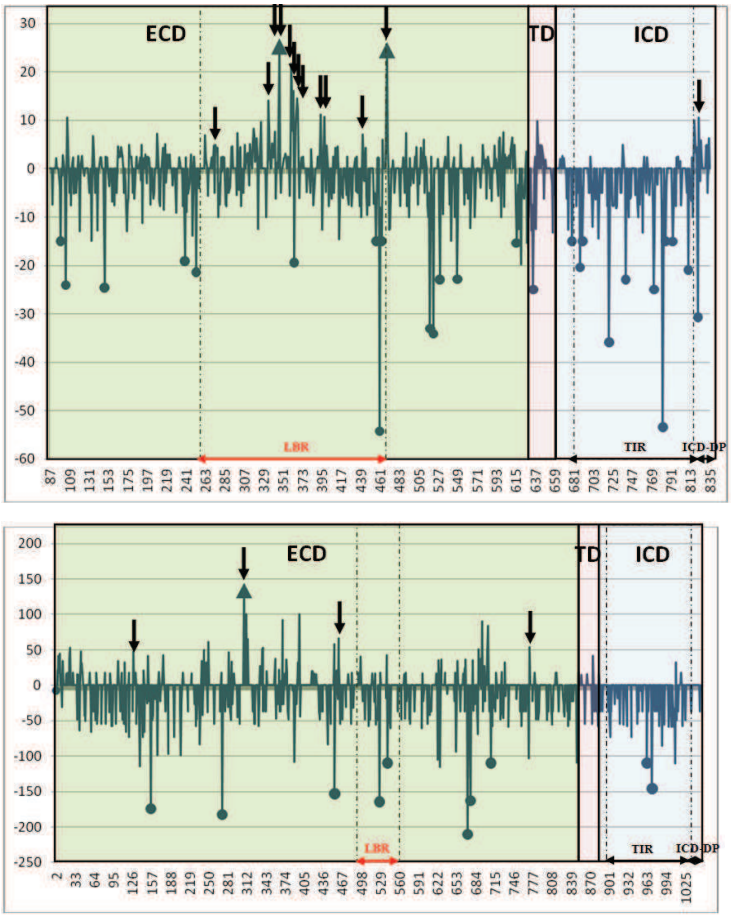


Fig. 2, Distribution of sites under selection identified by SLAC and MEME, Intensity of selection acting on *Tlr4* (a) and *Tlr7* (b) exon3 with $p < 0.05$; the blue line is normalized *dN-dS* calculated in SLAC (DATAMONKEY); blue arrows-up - sites under positive selection detected by SLAC; black arrows-down - sites under positive selection detected by MEME (DATAMONKEY); blue full circles - sites under negative selection detected by SLAC. **ECD** - extracellular domain; **LBR** - ligand binding region; **TD** - transmembrane domain; **TIR** - TIR domain; **ICD** - intracellular domain; **ICD-DP** - distal part of intracellular domain.

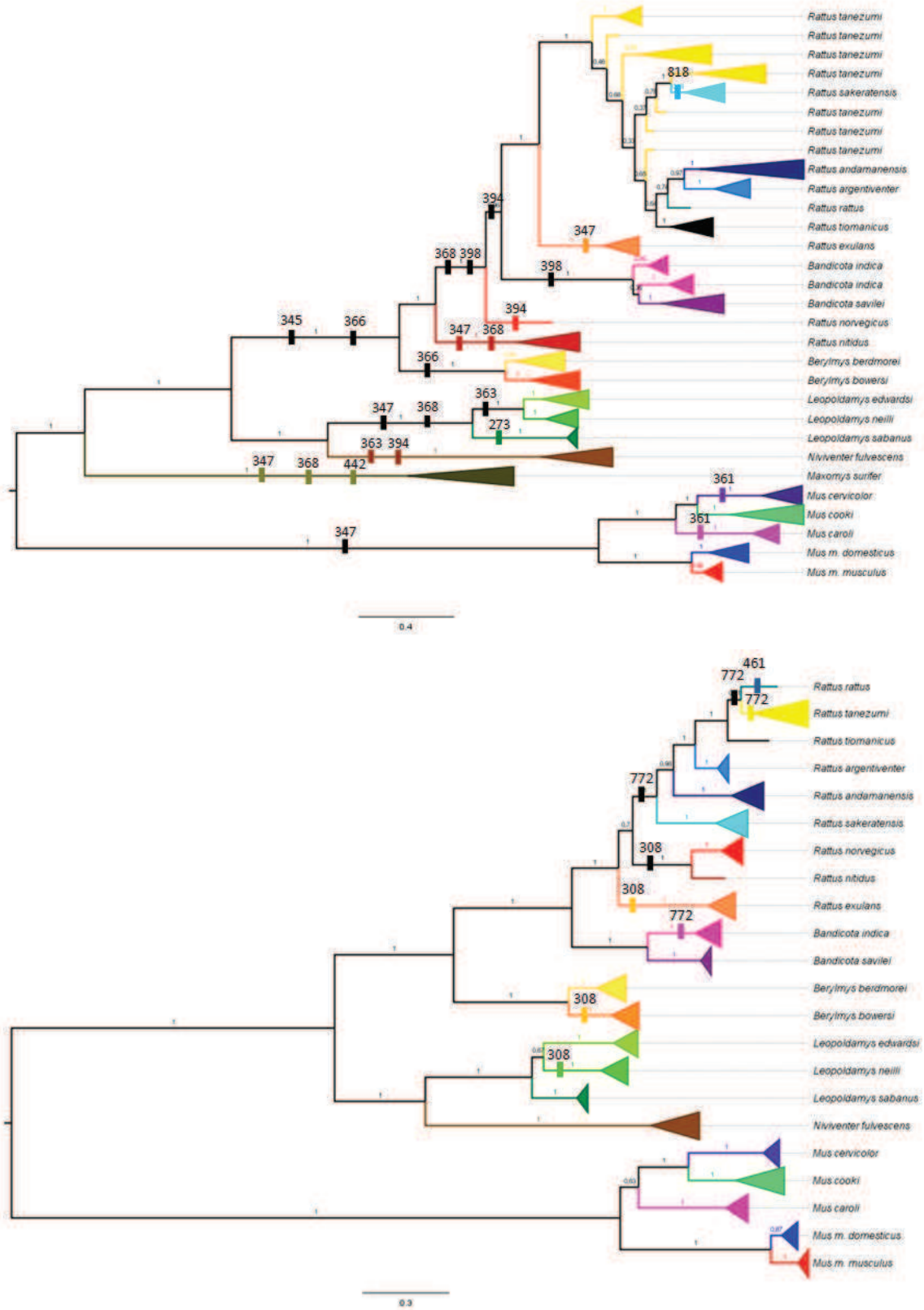


Fig. 3, Sites under positive selection identified in evolutionary lineages by MEME, *Tlr4* (a), *Tlr7* (b) (significance level at $p < 0.05$), positively selected sites are marked and numbered above branches at simplify phylogeny based on MrBayes.

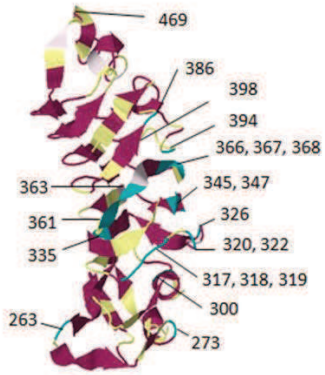
Analysis of the ligand binding regions

In general, the ligand binding region (LBR) was much more variable in *Tlr4* than in *Tlr7* genes. We detected 50 different AA variants of the LBR in the TLR4 dataset, while only eight different AA variants were detected in TLR7. Out of the 222 AA sites of $LBR_{TLR4,}$ 43% were polymorphic, while among the 103 AA sites of $LBR_{TLR7,}$ only 10% exhibited genetic variations. The CONSURF analysis performed to estimate the degree of evolutionary conservation of each amino acid position in LBR revealed 10% of phylogenetically variable positions (*i.e.* 22 positions assigned to grade 1 and corresponding to the most variable and rapidly evolving amino acid positions out of 222 positions in total) in TLR4, but only 2% (2 positions with grade 1 out of 103) in TLR7 (Fig. 4). Other positions were assigned as conservative (57% and 79% in TLR4 and TLR7, respectively) or had insufficient support (33% and 19%, respectively; Fig. 4).

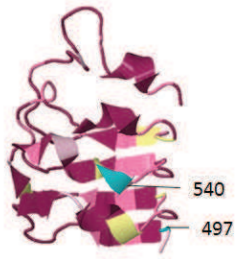
Ligand-binding positions in rodents were predicted by comparison with those identified in humans by Park et al. (2009). In TLR4, two out of eight LPS-binding amino acid positions were identical to humans and strictly conserved among rodents (F438 and F461). Three other were conserved in terms of amino acid features (*i.e.* polarity, hydrophobicity) but distinct from human residue and variable among rodents (R263K, K360R and K434R). Interestingly, one LPS binding site that was uniform in human was found to be evolving under positive selection using the MEME algorithm. We found hydrophobic and hydrophilic residues, although this position, L442Y, is known to be involved in hydrophobic interactions. Finally, two remaining positions were found to be highly variable in rodents (339 and 386) (Table S3 in SUPPLEMENTARY MATERIALS). In TLR7, the nine ligand binding residues predicted following Wei et al. (2009) were strictly conserved within rodents and seven of them were common to both rodents and human TLR7 (Table S4 in SUPPLEMENTARY MATERIALS).

The pairwise RMSD that allowed estimating the differences in 3D protein structure among variants varied from 0 to 1.5Å in TLR4 variants, and from 0.6 to 1.7Å in TLR7 variants (Fig. S5 in SUPPLEMENTARY MATERIALS). Yet, in the phenetic diagram of TLR4, 3D-structures of *Rattus sakeratensis* and *Rattus nitidus* were distinct from each other and also from all other species. Similarly for TLR7, the 3D-structure of the protein of *Rattus exulans* was separated from other species (Fig. S5 in Additional files 1). To provide wider context we performed additional comparison between PDB structures (obtained from The RCSB Protein Data Bank http://www.rcsb.org/pdb/home/home.do) of human (HoSaTLR4-

3fxi_A) and mouse (MuMuTLR4-3vq2_A) ECD_{TLR4} and between ECD of mouse TLR4 and TLR3 (MuMuTLR3-3ciy_A). The comparison between species of the same TLR was 1.7Å (HoSaTLR4-MuMuTLR4). Comparison between two TLRs from most distant TLR families of the same species was 4.6Å (MuMuTLR4-MuMuTLR3). The analysis of electric charge of LBR revealed higher variation in TLR4 (from -7.7 to 1.5) when compared with TLR7 (from -1.6 to 0.6). Detailed analyses of LBR_{TLR4} revealed that *Mus* and *Rattus* species were well differentiated from each other (*Mus*: from -7.7 to -3.7; *Rattus* and related genera: from -3 to 1.5, Fig. S6a in Additional files 1). Similar pattern was found for LBR_{TLR7} (*Mus*: -1.6, *Rattus* and related genera: from -1.4 to 0.6, Fig. S6b in SUPPLEMENTARY MATERIALS).



Codon	Residue Variety
263	R, K
273*	H, R, P
300	D, N, Y, V
317	S, H, D, P, Y
318	I, L, V
319	Q, R, K, E
320	H, Q, Y
322	A, K, G, E
326	S, N, K, R
335*	I, V
345*	A, T, P, K
347*	H, S, T, D, N, R, K
361*	A, Q, K, R, G, L, E
362	S, H, T, D, P
363*	T, M, I, L
366*	D, R, K, G
367	A, Q, T, D, R, K
368*	M, L, V
386	S, T, I, R
394*	F, M, L, V
398*	S, D, N, I, G
469*	A, T, I, G, V



Codon	Residue Variety
497	Q, K
540	Q, W, R

Fig. 4, Mapping of evolutionary conservation of amino acid positions in a protein molecule based on the phylogenetic relations between homologous sequences, conserved amino acid positions in LBR of TLR4 (a) and TLR7 (b). Structure of LBR was analysed in CONSURF; computations were based on MrBayes phylogenetic trees and tertiary protein structures of *R. norvegicus* [Gen Bank Acc. KC811688/ KC811786]; most variable positions are highlighted in turquoise and numbered (grade 1); most conserved sites are in violet; yellow sites mark insufficient data; white sites have average conservation score; tables show residue variants at the phylogenetically variable positions with grade 1; codons with asterisk have been identified as those under positive selection by MEME.

3.3.4 Discussion

In this study we analysed the variability of two important vertebrate immune genes involved in innate immunity across wild murine rodents and we looked for evidence of selection. Overall, we found that *Tlr4* was much more variable than *Tlr7* and that the evolution of both genes had been influenced mostly by purifying selection. However, comparison of both *Tlrs* revealed contrasting evolutionary patterns. *Tlr7*, which is involved in the recognition of viral nucleic acids, was highly conserved across rodents and its evolution seemed to be strongly shaped by purifying selection. Predicted ligand binding sites in LBR_{TLR7} were identical across all species and only few sites were detected to evolve under

positive selection within the whole molecule. By contrast, *Tlr4*, which detects several different pathogen ligands, was more variable and was affected by numerous events of episodic selection. Positively selected sites mostly occurred in LBR, probably as a result of co-evolution with pathogens. Analyses of the LBR variability in surface charge revealed a potential for interspecific differences in ligand binding capacities of both *Tlr*s.

Differences in TLRs evolution - phylogenetic approach

We found that both *Tlr*s were conserved genes as their phylogeny almost correctly recapitulated species phylogeny. In spite of this conservatism we revealed some incongruence between gene and species topologies, especially in branches represented by the shallow genealogy of the black rat complex and *Bandicota* spp. (Fig. 1a). These species have experienced recent and rapid radiation during the Early Pleistocene about 1 Mya (Pagès et al. 2010; Aplin et al. 2011). Discrepancies between a gene genealogy and the species phylogeny in recently diverged species often results from incomplete lineage sorting (ILS) of ancestral polymorphism and/or episodic gene flow and hybridization (Moore 1995; Hobolth et al. 2011). Indeed, *R. tanezumi R2* and *R. tanezumi R3* were recently proposed as conspecifics or were suspected to hybridize in Southeast Asia (Pagès et al. 2013). In addition, hybridization with introgression occurred between the invasive populations of *R. tanezumi* and *R. rattus* in the United States (Lack et al. 2012). These phenomena could explain incongruence between *Tlr*s and species trees. However, directional selection could also be involved. Discrepancies in *Tlr7* phylogeny represented by *R. exulans* and *R. sakeratensis* seem more likely to be caused by pathogen selective pressure (Fig. 1b). ILS and hybridization are unlikely to result in such deeper changes, whereas the influence of directional selection (positive or negative) on non-neutrally evolving genes could be at more likely explanation (Nichols 2001). The rejection of co-divergence (concerning basal nodes) between *Tlr*s and species phylogenies could reflect the occurrence of pathogen-driven selection on *Tlr*s during the evolutionary history of the murine rodents (Roach et al. 2005; Edwards 2009). The former hypothesis should now be tested by a detailed analysis of spectrum of pathogens from rodents to determine if the species producing the incongruent topology displayed specific pathogens that could mediate this selection.

Tlr variability and signatures of selection

We found that 92% and 100% sites (respectively for *Tlr4* and *Tlr7*) evolving under positive selection were located in the ECD, which is responsible for pathogen recognition. For *Tlr*4 92% of these positively selected sites found by MEME algorithm were located in the LBR. This is in concordance with several recent studies conducted on primates, birds and rodents, that have suggested a high accumulation of positively selected sites at LBR (Wlasiuk and Nachman 2010; Alcaide and Edwards 2011; Areal et al. 2011; Tschirren et al. 2011; Smith et al. 2012). Surprisingly, none of the sites evolving under positive selection was identified directly in the LBR of *Tlr7*.

The TIR domain of both *Tlr*s was evolving under much stronger functional constraint than the ECD in both genes. We found only 11 amino acid variants of TIR_{TLR4} in 23 species and six different variants of TIR_{TLR7} in 22 species. Altogether our results support the observation that *Tlr* exodomains evolve more rapidly than the intracellular TIR domain (Wlasiuk and Nachman 2010; Areal et al. 2011; Mikami et al. 2012; Smith et al. 2012). The requirement of sites within ECD, which would be involved in ligand recognition and able to recognize permanently fast-evolving pathogens, could explain this pattern. Besides, the high conservation of the TIR domain could be adapted to maintain a functional response of signal transduction (see, e.g. Poltorak et al. 1998; Smirnova et al. 2000; Hughes and Piontkivska 2008; Downing et al. 2010; Wlasiuk and Nachman 2010; Mikami et al. 2012).

Both genes showed non-significant differences between ECD and TIR with respect to *dS,* supporting the hypothesis that there was no difference in mutation rate between ECD and TIR. The same result has been found in comparative studies of 10 vertebrate TLRs (Hughes and Piontkivska 2008). The distal part of ICD in *Tlr4* was surprisingly highly variable among rodent species. The reason for such a high level of variability is still unknown; however some authors suggest that this region at the carboxy-terminal end of *Tlr4* could be responsible for interspecific differences in LPS sensitivity (Smirnova et al. 2000).

Positive selection we also detected using the MEME approach that individually considers each codon along the *Tlr*s phylogeny (Murrell et al. 2012). We found that episodic positive selection affected most lineages in the phylogenetic tree of *Tlr4*, while the situation was quite different in *Tlr7*, where the sites evolving under positive selection were mostly distributed only along the terminal branches. Episodic diversifying selection could have affected *Tlr4* throughout its evolution and this process could still be in operating, while in

Tlr7 diversifying selection seemed to have appeared more recently and the gene history was mostly maintained by the stronger purifying selection (Fig. 3).

Analysis of the Ligand binding region

In TLR4 variants we found 22 rapidly evolving positions distributed all over the LBR. While TLR4 is able to detect several ligands, the most studied one is LPS of Gram negative bacteria. TLR4 does not interact with LPS alone directly but forms stable heterodimers with MD-2 (Kim et al. 2007). Analysis of the crystallographic structure of mouse TLR4-MD-2-ligand complex has shown that the interactions between TLR4, -LPS and MD-2 take place on the concave surface of TLR4 (Kim et al. 2007). We predicted that sites involved in the TLR4-MD-2 interaction should be highly conserved to maintain the receptor function in LPS binding and these sites were thus not identified in the present study. Among the eight known LPS-binding sites, identified by Park et al. (2009) in humans, two residues (F438 and F461) were conserved between humans and rodents as well as among rodents. These key residues are jointly involved also in hydrophobic interactions between TLR4 and MD-2 (Park et al. 2009; Resman et al. 2009). It is possible that negative selection might maintain an invariable combination at these sites to preserve MD-2 binding, which supports our hypothesis mentioned above. One exception was the controversial site L442Y which was suggested by Park et al. (2009) to be also involved in hydrophobic interactions between TLR4 and MD-2, but Resman et al. (2009) challenged the importance of its function. Among the studied rodents this codon was found to be polymorphic and has been shown to be affected by episodic positive selection during rodent evolution. A hydrophobic nonpolar residue (Leucine, L) was commonly shared between rodent species except for *Maxomys surifer* that harbored a hydrophobic and polar Tyrosine (Y). For three LPS-binding sites, R263K, K360R and K434R, the biochemical features of the residue were maintained between rodents (all were positively charged residues) but distinct amino acids were detected. The important role of these residues was supported also by Ohto et al. (2012) and the potential functional importance of substitution R263K was beside confirmed by conservation analysis. Finally, we have identified in TLR4 two ligand binding positions, 339 and 386, with important amino acid substitutions that might be responsible for variability in LPS binding. No signature of positive selection was detected for these sites; however functional importance of position 386 was supported by the CONSURF analysis. Intriguingly, both residues form charge interactions with the same lipid A phosphate of the LPS, which might indicate that the evolution of this

position is associated with phosphate binding. However, this interpretation must be taken cautiously since Resman et al. (2009) have questioned the role of the site 386 (in human K388) in LPS binding.

LBR_{TLR7} sequence was much shorter than LBR_{TLR4} one (103 vs. 222 codons, respectively), which could be explained by the smaller size of LBR_{TLR7} ligand, the viral ssRNA (Wei et al. 2009). LBR_{TLR7} was highly conserved at the interspecific level. Only two rapidly evolving positions (out of 103 analysed sites) were detected and neither of them corresponded to the predicted ligand binding residues (Wei et al. 2009). Generally the conserved sites (sites evolving under negative selection), have important evolutionary roles for example in protein-protein interactions (TIR domain) or in the preservation of protein structure (*e.g.* LRR forming horseshoe structure).

We found that structural variation between rodent LBR of both TLRs (TLR4 - 1.5Å and TLR7 - 1.7Å) was comparable with the variation observed between ECD_{TLR4} of human and mouse (1.7Å). The 3D-protein structure modeling revealed that LBR_{TLR4} differed between *Rattus sakeratensis, R. nitidus* and all other rodent species. The analysis of LBR_{TLR4} sequences did not reveal any specific or unique substitution that could be responsible for this clustering. The same analysis performed on LBR_{TLR7} revealed that *Rattus exulans* substantially differed from other species. This difference could be explained by substitutions found at position H516Y, one being specific of *R. exulans* (Y at position 516) while other *Rattus* and *Mus* species harbored an H amino acid at this position. These inter-specific differences in LBR 3D structure were not related to the phylogenetic distance between species. They could be better explained by similar pathogen exposition and thus similar pathogen-mediated selection.

The results of charge analyses might be more important as they revealed interspecific variation in LBRs of both receptors. *Mus* species had generally a more negative overall charge at LBR than *Rattus* species (Fig. S6 in SUPPLEMENTARY MATERIALS). Differences in protein charges were previously shown to be associated with differences in protein-ligand interactions (Walsh et al. 2008; Govindaraj et al. 2011). Likewise, differences between these two groups were also found in LBR_{TLR4} at positions that directly bind to LPS. However, some caution is needed, since variation of TLR4 and TLR7 in sensitivity to LPS or ssRNA, respectively, between rats and mice has not been investigated.

Differences in evolution of bacterial-sensing and viral-sensing Tlrs

Our results showed that the bacterial-sensing *Tlr4* was more variable than the viral-sensing *Tlr7*, and that *Tlr4* evolution was more intensively shaped by positive selection than in *Tlr7*. *Tlr4* had 1.7% of codons under positive selection, while in *Tlr7* it was only 0.38%. These differences are likely to be explained by *Tlr*s' specificity to different groups of PAMPs with which they co-evolved (Mikami et al. 2012). *Tlr4* detects more types of ligands (*e.g.* bacterial LPS, envelope viral components, fungal cell wall components – Mannan) (Vinkler and Albrecht 2009) and it seems that these pathogen structures have exerted more diversifying selective pressures on *Tlr4* than the viral ssRNA affecting *Tlr7*. Recent studies of parasites show that there is an important structural variability in PAMPs between bacterial species (*e.g.* flagellin and LPS) (Raetz and Whitfield 2002; Woude and Bäumler 2004; Andersen-Nissen et al. 2005; Sun et al. 2006; Resman et al. 2009; Kang and Lee 2011; Maeshima and Fernandez 2013). We propose that the ligand binding region of *Tlr4* detecting these PAMPs should reflect higher ligand variability observed in our data.

Reduced genetic variability in important genes generally results from strong purifying selection acting against deleterious mutations in these genes. It can result in a smaller effective population size and a lower amount of incomplete lineage sorting (Charlesworth et al. 1993; Hobolth et al. 2011). These two phenomena were found to be more pronounced when analysing *Tlr7* phylogeny. Moreover the *Tlr7* gene is located on the X chromosome in mammals, which can be advantageous during evolution (*e.g.* lower polymorphism is maintained by quicker fixation of beneficial mutations and elimination of deleterious ones by stronger selection and more intense genetic drift) (Salcedo et al. 2007). We suggest that the tension between diversifying and purifying selection, caused by adaptation to the variability of viral motifs detected by viral-sensing *Tlr7* and maintenance of function together played an important role in the distribution of *Tlr7* polymorphisms.

Conclusion

This study brings a unique insight into the natural variability and molecular history of two Toll-like receptors in free-living populations of 23 murine species. Purifying selection seems to be the dominant evolutionary force shaping *Tlr4* and *Tlr7* polymorphism. However, specific sites putatively evolving under diversifying selection were detected in both *Tlr*s. These sites accumulated within *Tlr4* LBR, and detailed analyses revealed that several

important amino-acid substitutions might alter LPS binding. These substitutions were often species-specific and differentiated between the Rattini and Murini tribes. Interspecific charge variability of LBR and to lesser extent the variability in 3D structure indicated the potential differences in protein-ligand interaction. By contrast, the evolution of *Tlr7* was strongly shaped by purifying selection. All predicted ligand binding residues in this receptor were uniform across all studied mammals to date. The contrasting evolutionary histories of these two *Tlr*s are likely to result from different structural variability of ligands they target. Since the crystallography of certain ligands (*e.g.* biglycans, hyaluronans and heparin sulphates, ssRNA) (Wei et al. 2009; Kang and Lee 2011) remains unknown and the precise positions of corresponding binding sites are still missing, our data provide important avenues towards understanding which codons might be candidates for ligand binding residues.

Availability of supporting data section

All sequences have been submitted to NCBI GenBank under Accession numbers from KC811609 to KC811800 (Individual accession numbers are presented in Table S1 in SUPPLEMENTARY MATERIALS). *Tlr* phylogenies based on MrBayes (*Tlr4*_MrBayes_final.nex, *Tlr7*_MrBayes_final.nex) and RAxML (*Tlr4*_RAxML_final.nex, *Tlr7*_RAxML_final.nex) approach were added to the TreeBase database (http://treebase.org/treebase-web/home.html). Trees are available at URL: http://purl.org/phylo/treebase/phylows/study/TB2:S14659

Acknowledgments

This work was supported by the French National Agency for Research projects CERoPath (grant number 00121 0505, 07 BDIV 012) http://www.ceropath.org/ and BioDivHealthSEA (grant number ANR 11 CPEL 002), and the Czech Science Foundation (grant number 206/08/0640). Cooperation on this project was also partly supported by bilateral project BARRANDE (grant number MEB021130/24504WM). The thesis of A. Fornůsková was partly funded by a three year French government fellowship and the fellowship from Masaryk University. MP is currently funded by an FRS - FNRS fellowship (Belgian Fund for Scientific Research).We are grateful to Anna Bryjová, Yannick Chaval, Gael Kergoat, Marian Novotný, Sylvain Piry, Lucie Vlčková for their help during various stages of the manuscript preparation

and to Jamie Caroline Winternitz for language corrections. We also thank to the CBGP HPC computational platform and to the Centre Méditerranéen Environnement Biodiversité.

Authors' Contributions

Conceived and designed the experiments: AF JFC JB NCH MV. Performed the sequencing: AF MG FC. Analysed the data: AF MV MP EJ. Contributed samples: SM JFC AF. Wrote the paper: AF MV JFC JB NCH MP EJ (sorted by the significance of contributions). All authors read and approved the final manuscript.

Table S1. Summary of sampled specimens and identification of haplotypes.

ID	Species	Country	Province	Sex	LBR_{TLR4} variants	LBR_{TLR7} variants	Hap Exon 3 *Tlr4*	Hap Exon 3 *Tlr7*	GenBank Acc. *Tlr4/Tlr7*
C0443	*Bandicota indica*	Cambodia	Mondolkiri	F	Bain1_Bain2	Rattus sp.	**bainh5_bainh6**	**bainh1**	**KC811609/KC811705**
L0276	*Bandicota indica*	Lao PDR	Champasak	M	Bain3	Rattus sp.	**bainh1_bainh2**	**bainh2**	**KC811610/KC811706**
R4000	*Bandicota indica*	Thailand	Kalasin	M	Bain1	Rattus sp.	**bainh4**	**bainh3**	**KC811611/KC811707**
R5313	*Bandicota indica*	Thailand	Nan	M	Bain3	Rattus sp.	**bainh3**	**bainh4**	**KC811612/KC811708**
C0709	*Bandicota savilei*	Cambodia	Mondolkiri	F	Basa2	Basa	**basah1**	basah2	**KC811613/KC811709**
L0331	*Bandicota savilei*	Lao PDR	Champasak	F	Basa1_Basa2	Basa	**basah2_basah3**	**basah1**_basah2	**KC811614/KC811710**
R4141	*Bandicota savilei*	Thailand	Phrae	M	Basa3	Basa	**basah4**	**basah2**	**KC811615/ KC811711**
R5475	*Bandicota savilei*	Thailand	Buriram	F	Basa1	Basa	**basah5_basah6**	basah2	**KC811616/ KC811712**
C0333	*Berylmys berdmorei*	Cambodia	Sihanouk	M	Bebe	Be sp.	**bebeh2**	**bebeh1**	**KC811617/ KC811713**
C0481	*Berylmys berdmorei*	Cambodia	Mondolkiri	M	Bebe	Be sp.	**bebeh1**	**bebeh2**	**KC811618/ KC811714**
L0006	*Berylmys berdmorei*	Lao PDR	Luang Prabang	M	Bebe	Be sp.	bebeh3_**bebeh4**	**bebeh3**	**KC811619/ KC811715**
R3441	*Berylmys berdmorei*	Thailand	Loei	M	Bebe	Be sp.	**bebeh3**	**bebeh4**	**KC811620/ KC811716**
L0151	*Berylmys bowersi*	Lao PDR	Luang Prabang	F	Bebo	Be sp.	**beboh4**_beboh6	**beboh1**_beboh4	**KC811621/ KC811717**
R4400	*Berylmys bowersi*	Thailand	Loei	M	Bebo	Be sp.	**beboh1**	beboh4	**KC811622/ KC811718**
R4804	*Berylmys bowersi*	Thailand	Loei	M	Bebo	Be sp.	**beboh2_beboh3**	**beboh4**	**KC811623/ KC811719**
R5410	*Berylmys bowersi*	Thailand	Nan	F	Bebo	Be sp.	**beboh5_beboh6**	**beboh2_beboh3**	**KC811624/ KC811720**
C0421	*Leopoldamys edwardsi*	Cambodia	Mondolkiri	M	Leed	Le sp.	**leedh4_leedh5**	**leedh3**	**KC811625/ KC811721**
R4070	*Leopoldamys edwardsi*	Thailand	Loei	M	Leed	Le sp.	**leedh6_leedh7**	**leedh1**	**KC811626/ KC811722**
R4276	*Leopoldamys edwardsi*	Thailand	Phrae	F	Leed	Le sp.	**leedh1_leedh3**	leedh2	**KC811627/ KC811723**
R4296	*Leopoldamys edwardsi*	Thailand	Phrae	F	Leed	Le sp.	leedh1_**leedh2**	leedh2	**KC811628/ KC811724**
R4350	*Leopoldamys edwardsi*	Thailand	Phrae	?	x	Le sp.	x	leedh2	x/ **KC811725**
R5057	*Leopoldamys edwardsi*	Thailand	Loei	M	x	Le sp.	x	**leedh2**	x/ **KC811726**
R4477	*Leopoldamys neilli*	Thailand	Phrae	F	Lene	Le sp.	**leneh2**_leneh4	**leneh1**	**KC811629/ KC811727**
R4486	*Leopoldamys neilli*	Thailand	Phrae	F	Lene	Le sp.	**leneh1**_leneh4	leneh1	**KC811630/ KC811728**
R4527	*Leopoldamys neilli*	Thailand	Loei	F	Lene	Le sp.	**leneh4**	leneh2	**KC811631/ KC811729**
R4530	*Leopoldamys neilli*	Thailand	Loei	M	Lene	Le sp.	**leneh3**_leneh4	**leneh2**	**KC811632/ KC811730**
R3033	*Leopoldamys sabanus*	Thailand	Kanchanaburi	F	Lesa	Le sp.	**lesah1**_lesah2	**lesah1**_lesah2	**KC811633/ KC811731**
R3111	*Leopoldamys sabanus*	Thailand	Kanchanaburi	M	Lesa	Le sp.	lesah1_**lesah2**	**lesah2**	**KC811634/ KC811732**
C0118	*Maxomys surifer*	Cambodia	Sihanouk	M	Masu1	x	**masuh5_masuh7**	x	**KC811635/** x
C0478	*Maxomys surifer*	Cambodia	Mondolkiri	M	Masu2	x	**masuh1_masuh2**	x	**KC811636/** x
L0274	*Maxomys surifer*	Lao PDR	Champasak	F	Masu1	x	**masuh6**	x	**KC811637/** x
R4099	*Maxomys surifer*	Thailand	Loei	M	Masu3_Masu4	x	**masuh3_masuh4**	x	**KC811638/**x
C0423	*Mus caroli*	Cambodia	Mondolkiri	M	Muca2	Mus sp.	**mucah3**	**mucah4**	**KC811639/ KC811733**
L0014	*Mus caroli*	Lao PDR	Luang Prabang	F	Muca1	Mus sp.	**mucah2**	**mucah1**	**KC811640/ KC811734**
L0211	*Mus caroli*	Lao PDR	Luang Prabang	M	Muca1	Mus sp.	mucah3_**mucah4**	mucah1	**KC811641/ KC811735**
L0275	*Mus caroli*	Lao PDR	Champasak	M	Muca3	Mus sp.	**mucah1**_mucah3	**mucah2**	**KC811642/ KC811736**
R5642	*Mus caroli*	Thailand	Buriram	M	x	Mus sp.	x	**mucah3**	x/ **KC811737**
R4864	*Mus cervicolor*	Thailand	Loei	F	Muce1_Muce2	Mus sp.	**muceh4_muceh5**	**muceh1**	**KC811643/ KC811738**

R5644	*Mus cervicolor*	Thailand	Buriram	M	Muce1_Muce2	Mus sp.	**muceh1_muceh2**	**muceh2**	**KC811644/ KC811739**
R5666	*Mus cervicolor*	Thailand	Buriram	F	x	Mus sp.	x	muceh2	x/ **KC811740**
R5671	*Mus cervicolor*	Thailand	Buriram	F	Muce2	Mus sp.	**muceh3**_muceh2	**muceh3**	**KC811645/ KC811741**
L0103	*Mus cookii*	Lao PDR	Luang Prabang	F	Muco1	Mus sp.	mucoh3	**mucoh3**	**KC811646/ KC811742**
L0178	*Mus cookii*	Lao PDR	Luang Prabang	M	Muco1	Mus sp.	**mucoh3**	mucoh3	**KC811647/ KC811743**
R4106	*Mus cookii*	Thailand	Loei	F	Muco1_Muco2	Mus sp.	**mucoh1_mucoh2**	**mucoh1_mucoh2**	**KC811648/ KC811744**
SK1515	*Mus domesticus*	France	Saint Jean-et-Royans	F	Mus sp.	Mus sp.	**mudoh5**	**mudoh3**	**KC811649/ KC811745**
SU3770	*Mus domesticus*	Turkey	Izmir	F	Mus sp.	Mus sp.	**mudoh2**_mudoh5	mudoh1_mudoh3	**KC811650/ KC811746**
M273	*Mus domesticus*	Syria	Palmyra	M	Mus sp.	Mus sp.	**mudoh4**_mudoh5	**mudoh1**	**KC811651/ KC811747**
MISC352	*Mus domesticus*	Egypt	Sabha Oasis	M	Mus sp.	Mus sp.	**mudoh1_mudoh3**	**mudoh2**	**KC811652/ KC811748**
JPC2821	*Mus musculus*	Czech rep.	Buskovice	F	Mus sp.	Mus sp.	**mumuh3**	mumuh2	**KC811653/ KC811749**
SK843	*Mus musculus*	Germany	Lindhorst	M	Mumu2	Mus sp.	mumuh4	mumuh2	**KC811654/ KC811750**
SU5218	*Mus musculus*	Russia	Shkili	F	Mumu2	Mus sp.	mumuh4	mumuh1_mumuh2	**KC811655/ KC811751**
360	*Mus musculus*	Romania	Botosani	M	Mumu2	Mus sp.	mumuh3_**mumuh4**	**mumuh2**	**KC811656/ KC811752**
SU5209	*Mus musculus*	Ukraine	Primorskoe	M	Mus sp._Mumu1	Mus sp.	**mumuh1_mumuh2**	**mumuh1**	**KC811657/ KC811753**
C0322	*Niviventer fulvescens*	Cambodia	Sihanouk	F	Nifu1	Nifu	**nifuh1_nifuh2**	**nifuh2**	**KC811658/ KC811754**
C0430	*Niviventer fulvescens*	Cambodia	Mondolkiri	F	x	Nifu	x	**nifuh5_nifuh8**	x/ **KC811755**
L0273	*Niviventer fulvescens*	Lao PDR	Champasak	F	Nifu1	Nifu	**nifuh3_nifuh4**	**nifuh6_nifuh7**	**KC811659/ KC811756**
R4071	*Niviventer fulvescens*	Thailand	Loei	M	Nifu1_Nufu2	Nifu	**nifuh5_nifuh6**	**nifuh1**	**KC811660/ KC811757**
R4497	*Niviventer fulvescens*	Thailand	Phrae	F	Nifu1	Nifu	nifuh3_**nifuh7**	**nifuh3_nifuh4**	**KC811661/ KC811758**
L0052	*Rattus andamanensis*	Lao PDR	Luang Prabang	F	Raan1	Rattus sp.	**raanh1**	raanh1_raanh2	**KC811662/ KC811759**
L0149	*Rattus andamanensis*	Lao PDR	Luang Prabang	F	Raan2_Raan3	Rattus sp.	**raanh2_raanh3**	**raanh1_raanh2**	**KC811663/ KC811760**
R2953	*Rattus andamanensis*	Thailand	Kanchanaburi	F	Raan3	Rattus sp.	raanh4_**raanh5**	raanh3_**raanh4**	**KC811664/ KC811761**
R3087	*Rattus andamanensis*	Thailand	Kanchanaburi	M	Raan3	Rattus sp.	**raanh4**_raanh5	**raanh3**	**KC811665/ KC811762**
C0014	*Rattus argentiventer*	Cambodia	Sihanouk	M	Raan1	Rattus sp.	**raarh5**	**raarh1**	**KC811666/ KC811763**
C0048	*Rattus argentiventer*	Cambodia	Sihanouk	F	Raan1	Rattus sp.	**raarh3**_raanr5	raarh1_**raarh2**	**KC811667/ KC811764**
C0104	*Rattus argentiventer*	Cambodia	Sihanouk	M	Raan1	Rattus sp.	**raarh2**	raarh1	**KC811668/ KC811765**
R5674	*Rattus argentiventer*	Thailand	Buriram	F	Raan1	Rattus sp.	**raarh1**	raarh1	**KC811669/ KC811766**
R5679	*Rattus argentiventer*	Thailand	Buriram	M	Raar	Rattus sp.	**raarh4**	raarh1	**KC811670/ KC811767**
C0278	*Rattus exulans*	Cambodia	Sihanouk	M	Raex1	Raex	**raexh1_raexh3**	**raexh5**	**KC811671/ KC811768**
C0353	*Rattus exulans*	Cambodia	Mondolkiri	F	Raex3	Raex	raexh1_**raexh2**	**raexh2**_raexh5	**KC811672/ KC811769**
L0217	*Rattus exulans*	Lao PDR	Champasak	F	Raex1	Raex	**raexh7**	**raexh4**_raexh5	**KC811673/ KC811770**
R1805	*Rattus exulans*	Thailand	Bangkok	M	Raex1	Raex	**raexh5_raexh6**	**raexh3**	**KC811674/ KC811771**
R4103	*Rattus exulans*	Thailand	Loei	M	Raex2	Raex	**raexh4**_raexh7	**raexh1**	**KC811675/ KC811772**
L0277	*Rattus sakeratensis*	Lao PDR	Champasak	F	Rasa	Rattus sp.	rasah5_**rasah6**	**rasah2**	**KC811676/ KC811773**
R0237	*Rattus sakeratensis*	Thailand	Ratchaburi	F	Rasa	Rattus sp.	**rasah1**_rasah6	**rasah1**_rasah5	**KC811677/ KC811774**
R1015	*Rattus sakeratensis*	Thailand	Nakhon Ratchasima	M	Rasa	Rattus sp.	**rasah3_rasah5**	**rasah3**	**KC811678/ KC811775**
R4402	*Rattus sakeratensis*	Thailand	Loei	F	Rasa	Rattus sp.	**rasah4**_rasah5	**rasah5**	**KC811679/ KC811776**
R4568	*Rattus sakeratensis*	Thailand	Phrae	M	Rasa	Rattus sp.	**rasah2**	**rasah4**	**KC811680/ KC811777**
L0180	*Rattus nitidus*	Lao PDR	Luang Prabang	F	Rani	Rano-Rani	**ranih5**	**ranih1**	**KC811681/ KC811778**
L0191	*Rattus nitidus*	Lao PDR	Luang Prabang	M	Rani	Rano-Rani	**ranih3**_ranih5	ranih1	**KC811682/ KC811779**
L0192	*Rattus nitidus*	Lao PDR	Luang Prabang	F	Rani	Rano-Rani	ranih1_**ranih4**	ranih1	**KC811683/ KC811780**
L0196	*Rattus nitidus*	Lao PDR	Luang Prabang	M	x	Rano-Rani	x	ranih1	x/ **KC811781**
R4846	*Rattus nitidus*	Thailand	Loei	M	Rani	Rano-Rani	**ranih1_ranih2**	ranih1	**KC811684/ KC811782**

C0141	*Rattus norvegicus*	Cambodia	Sihanouk	F	Rano	Rano-Rani	ranoh1	**ranoh1**_ranoh3	**KC811685/ KC811783**
C0210	*Rattus norvegicus*	Cambodia	Sihanouk	F	Rano	Rano-Rani	ranoh1	**ranoh2**_ranoh3	**KC811686/ KC811784**
C0211	*Rattus norvegicus*	Cambodia	Sihanouk	F	Rano	Rano-Rani	ranoh1	**ranoh3**	**KC811687/ KC811785**
C0224	*Rattus norvegicus*	Cambodia	Sihanouk	F	Rano	Rano-Rani	**ranoh1**	ranoh3	**KC811688/ KC811786**
C0028	*Rattus tanezumi R3*	Cambodia	Sihanouk	F	Rata7_Rata8	Rattus sp.	**ratah1_ratah7**	**ratah7_ratah8**	**KC811689/ KC811787**
C0250	*Rattus tanezumi R3*	Cambodia	Sihanouk	F	Rati_Rata9	Rattus sp.	**ratah4_ratah5**	**ratah1**	**KC811690/ KC811788**
C0477	*Rattus tanezumi R3*	Cambodia	Mondolkiri	M	Rata9	Rattus sp.	**ratah6**	**ratah4**	**KC811691/ KC811789**
L0313	*Rattus tanezumi R3*	Lao PDR	Champasak	F	x	Rattus sp.	x	**ratah11_ratah12**	x/ **KC811790**
L0242	*Rattus tanezumi R3*	Lao PDR	Champasak	F	Rata9	Rattus sp.	ratah6_**ratah8**	**ratah5_ratah6**	**KC811692/ KC811791**
R5051	*Rattus tanezumi R3*	Thailand	Loei	F	Rata10_Rata11	Rattus sp.	**ratah2_ratah3**	**ratah2_ratah3**	**KC811693/ KC811792**
NK37	*Rattus rattus*	Senegal	?	F	Rara	Rattus sp.	rarah1	rarah1	**KC811694/ KC811793**
R197	*Rattus rattus*	Guadaloupe	?	?	Rara	Rattus sp.	**rarah1**	**rarah1**	**KC811695/ KC811794**
R2	*Rattus rattus*	Riov	?	?	Rara	Rattus sp.	rarah1	rarah1	**KC811696/ KC811795**
L0100	*Rattus tanezumi R2*	Lao PDR	Luang Prabang	F	Rata1_Rata2	Rattus sp.	**ratah10_ratah11**	ratah9_**ratah10**	**KC811697/ KC811796**
R1831	*Rattus tanezumi R2*	Thailand	Nakhon Sri Thammarat	F	Rata3_Rata4	Rattus sp.	**ratah15_ratah16**	**ratah9**	**KC811698/ KC811797**
R3560	*Rattus tanezumi R2*	Thailand	Samui	F	Rati_Rata5	Rattus sp.	**ratah13_ratah14**	**ratah13**	**KC811699/ KC811798**
R4377	*Rattus tanezumi R2*	Thailand	Loei	M	Rata1_Rata6	Rattus sp.	**ratah9_ratah12**	ratah9	**KC811700/ KC811799**
Li0249	*Rattus tiomanicus*	Sumatra	Minas	M	Rati	x	**ratih1_ratih2**	x	**KC811701/** x
Li0258	*Rattus tiomanicus*	Sumatra	Minas	M	Rati	x	**ratih3**_ratih6	x	**KC811702/** x
Li0259	*Rattus tiomanicus*	Sumatra	Minas	M	Rati	x	**ratih4**_ratih6	x	**KC811703/** x
Li0315	*Rattus tiomanicus*	Sumatra	Minas	?	Rati	Rattus sp.	**ratih5_ratih6**	**ratih1**	**KC811704/ KC811800**

NOTE. - ID - identification of specimens; **LBR_{TLR4} and LBR_{TLR7} variants** - variants of ligand-binding region, double indication mean heterozygote specimen; **Hap_Exon3_*Tlr4*** and ***Tlr7***- alleles of exon 3 for each species, double indication means heterozygote specimen; haplotypes used for all analysis are in bold; **x** - sequence not complete;**? -** no exact information; **GenBank Acc.** - Gen Bank Accession numbers.

Table S2. Primer description.

Primer ID	Sequence of primers 5' – 3'	Function
rTLR4-F	AGT TTA TCA TCA CTG YA GCA AG	amplification
rTLR4-XF2	CCC AAT TGA CTC CAT TCA AGC CC	amplification
rTLR4-XF3	CCC TCA GGA CTC TTG ATT GCA G	amplification
rTLR4-R-1	ATT CTC CCA AGA TCA ACC GAT G	amplification
rTLR4-R-3	CTG KTC CTT GAC CCA CTG C	amplification
rTLR4-R	AGA RMC CCA GRT GAR CTG TAG CAT T	amplification
rTLR7-F	AAG ACC YRT GTT GYT TAG TTT TAA TAA TG	amplification
rTLR7-1F	CAG ATT AGA CCT GGA AGC TTT AGT G	amplification
rTLR7-4F	TCT TGA CCT TGG CAC TAA CTT CAT A	amplification
rTLR7-5F	CCA TTG GCC AAA CTC TTA ATG G	sequencing
rTLR7-6F	GGT GAT AAC AGA TAC TTG GAC TTC T	sequencing
rTLR7-7F	CTG GCC ACT GAT GTG ACT TGT	sequencing
rTLR7-2R	GTT AGC CTC AAG GCT CAG AAG	amplification
rTLR7-9R	TAT CGG AAA TAG TGT AAG GCC TCA AG	amplification
rTLR7-R	AGA AAG AAR TTA TCK TCT ATC AGT CTC	amplification

Table S3. Residues binding to LPS in TLR4 based on knowledge of 3D-crystalography in human predicted by Park et al. 2009.

Position in rodent alignment	Numbering in human sequence	Function of AA	Residue variety in rodents
263	hTLR4_R264	LPS (Charge interaction with phosphates)	positively charged R, K
339	hTLR4_K341	LPS (Charge interaction with phosphates)	positively charged R, H, uncharged NH2 residue Q
360	hTLR4_K362	LPS (Charge interaction with phosphates)	positively charged K, R
386	hTLR4_K388	LPS (Charge interaction with phosphates)	positively charged R, uncharged hydrophilic T, small uncharged hydrophilic S, uncharged hydrophobic I
434	hTLR4_Q436	LPS (Hydrogen bond)	uncharged R, K
438	hTLR4_F440	LPS, MD-2 (Hydrophobic interaction)	uniformly aromatic hydrophobic F
442*	hTLR4_L444	LPS, MD-2 (Hydrophobic interaction)	aliphatic hydrophobic L, aromatic polar hydrophobic Y
461	hTLR4_F463	LPS, MD-2 (Hydrophobic interaction)	uniformly aromatic hydrophobic F

NOTE. - Variable sites detected by CONSURF are underlined; * indicates sites identified by MEME.

Table S4. Potential residues binding ssRNA predicted by Wei et al. 2009.

Position in rodent alignment	Numbering in human sequence	Function of AA	Residue variety in rodents
503	hTLR7_K502	Potential ligand binding residue	uniformly positively charged R
505	hTLR7_S504	Potential ligand binding residue	uniformly uncharged NH_2 N
527	hTLR7_G526	Potential ligand binding residue	uniformly tiny G
532	hTLR7_Q531	Potential ligand binding residue	uniformly uncharged NH_2 Q
552	hTLR7_N551	Potential ligand binding residue	uniformly uncharged NH_2 N
554	hTLR7_R553	Potential ligand binding residue	uniformly positively charged R
557	hTLR7_L556	Potential ligand binding residue	uniformly aliphatic hydrophobic L
576	hTLR7_S575	Potential ligand binding residue	uniformly small uncharged hydrophilic S
579	hTLR7_H578	Potential ligand binding residue	uniformly positively charged aromatic H

FIGURES

Fig. S1, Protein structure of TLR4 (a, c) and TLR7 (b, d) identified by SMART (http://smart.embl-heidelberg.de/) (a, b) and CONSURF (c, d). SMART (a, b) identified following types of domains: **LRR** - Leucine rich repeat; **LRRCT** - Leucine rich repeat C-terminal domain; **TIR** - TIR domain, **Fulfilled blue box (TD)** - transmembrane domain; **LRRNT** - Leucine rich repeat N-terminal domain. **Red box** - LBR (from AA248 to AA469 for TLR4 and from AA495 to AA597 for TLR7). **ECD** - extracellular domain is represented by solid black double arrow; **ICD** - intracellular domain is represented by dashed double arrow. Distal part of ICD (**ICD-DP**) is indicated by a simple solid arrow. Positions of forward and reverse primers used for amplification are shown by arrows. Arrows of same color indicates primer pairs. Description of crystallographic structure (c, d) **LBR** is represented by red polygon; **TD** is present between two dashed lines. To the right from TD is **ICD**, to the left is **ECD**.

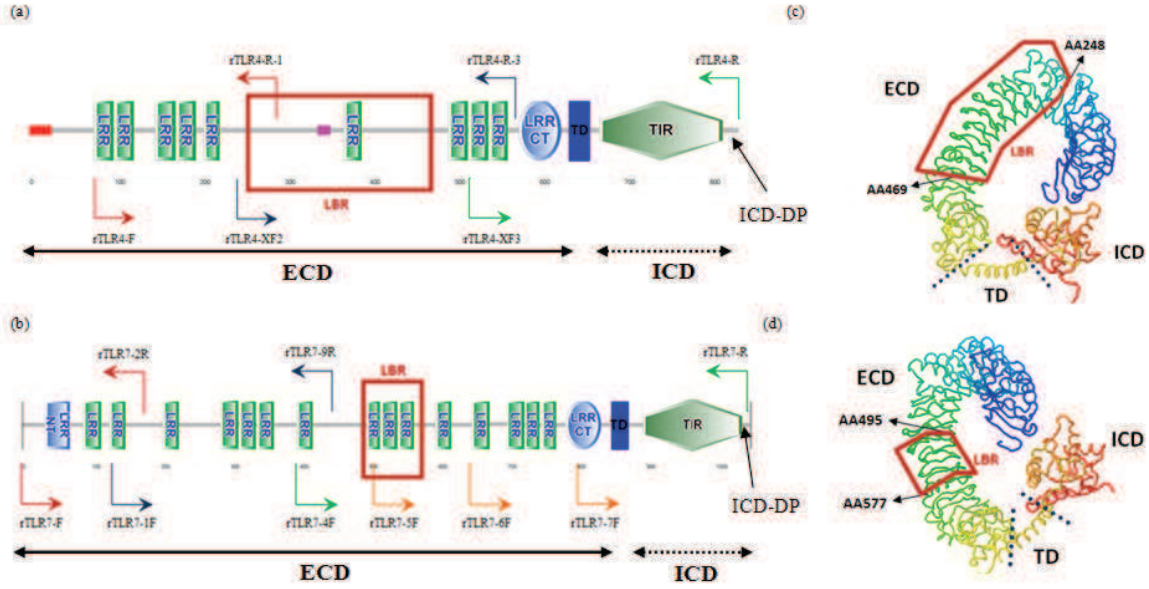


Fig. S2a, Phylogenetic trees based on the exon 3 of *Tlr4* gene reconstructed by Bayesian inference method in MrBayes (using GTR+G model of evolution). Numbers above the branches indicates Bayesian posterior probabilities supporting the branches. Names of haplotypes are explained in the Table S1.

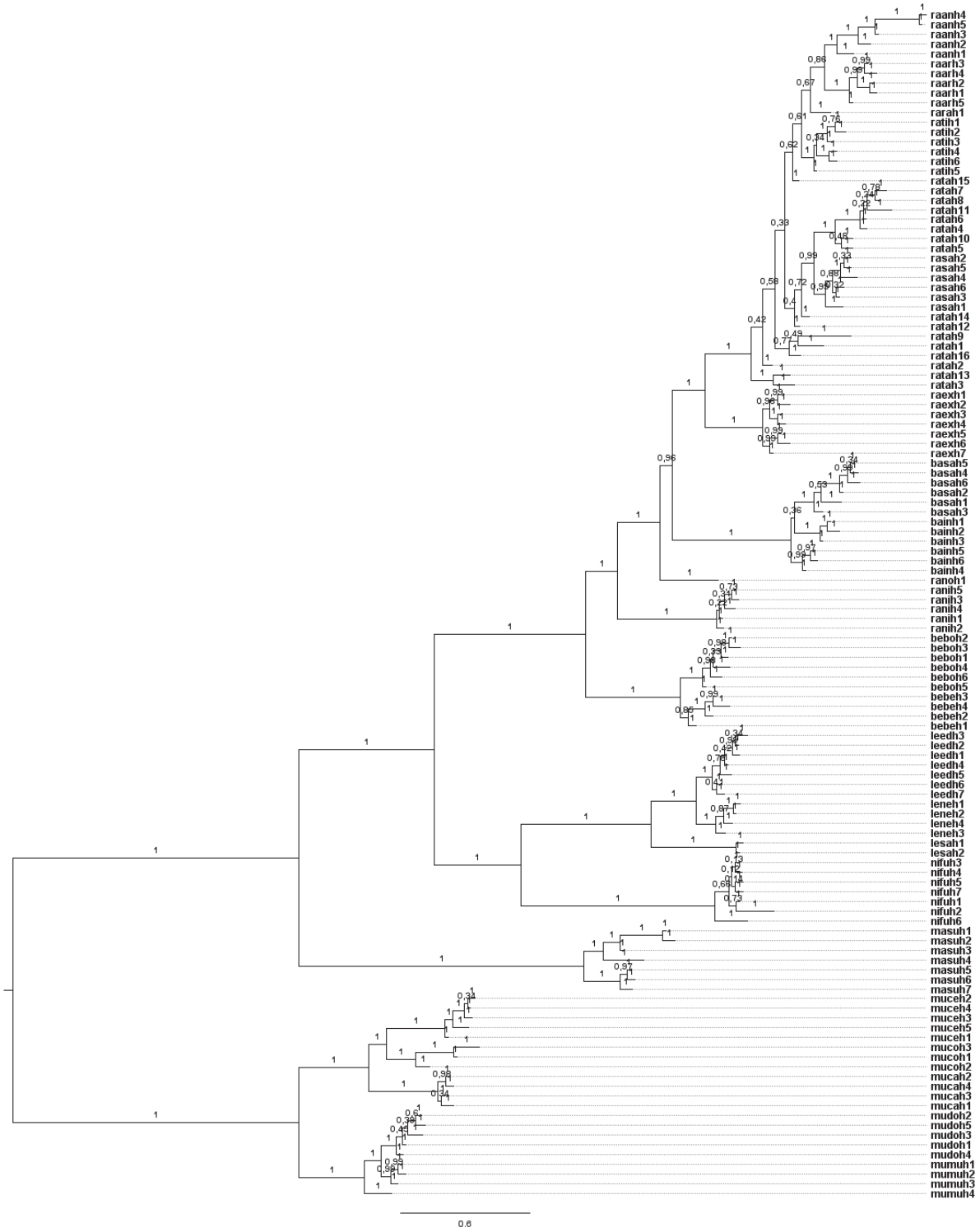


Fig. S2b, Phylogenetic trees based on the exon 3 of *Tlr7* gene reconstructed by Bayesian inference method in MrBayes (using GTR+G model of evolution). Numbers above the branches indicates Bayesian posterior probabilities supporting the branches. Names of haplotypes are exlained in the Table S1.

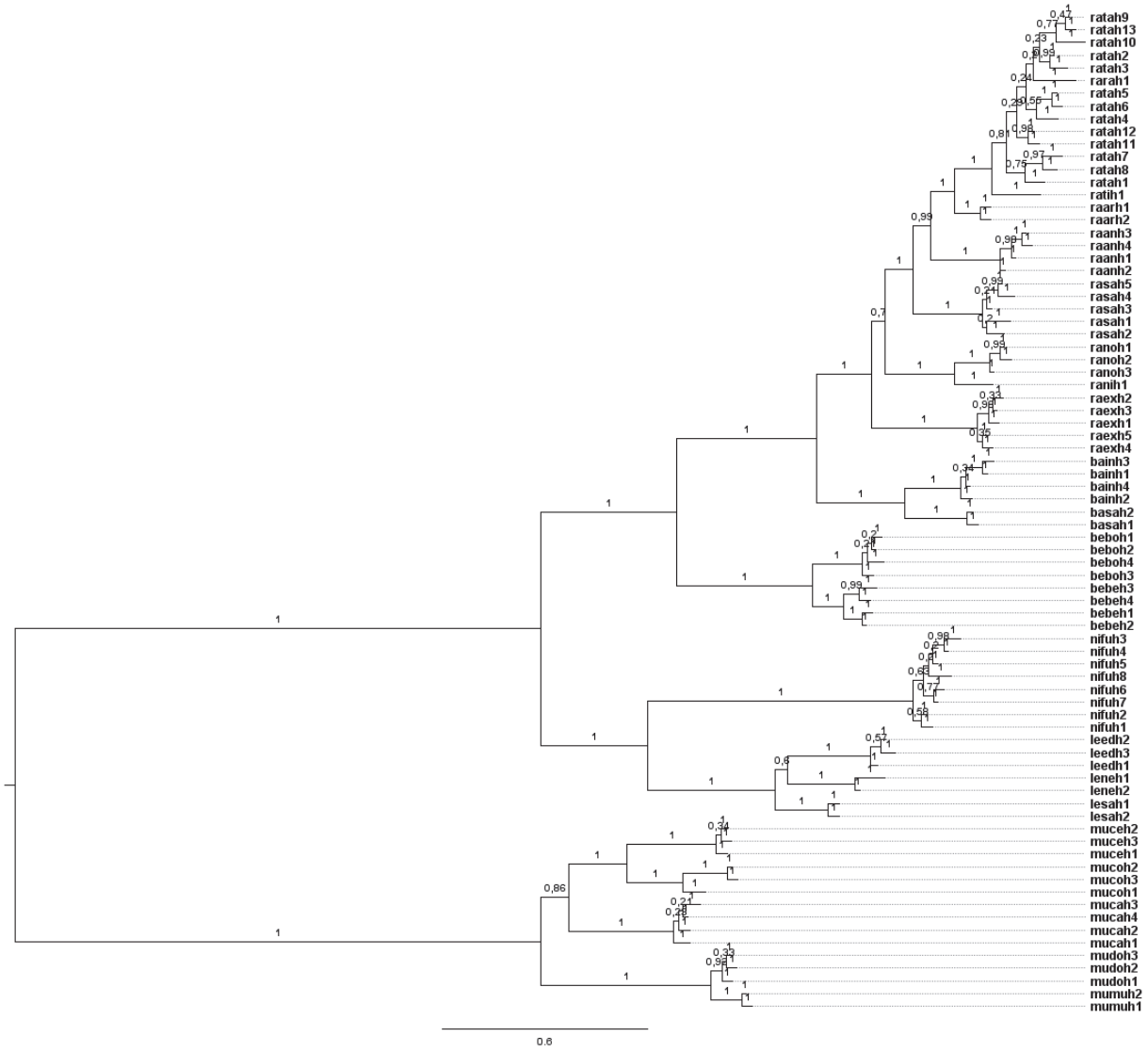


Fig. S3a, Phylogenetic trees based on the exon 3 of *Tlr4* gene reconstructed by maximum likelihood method in RAxML, Numbers above the branches indicates bootstrap support values, Names of haplotypes are explained in the Table S1.

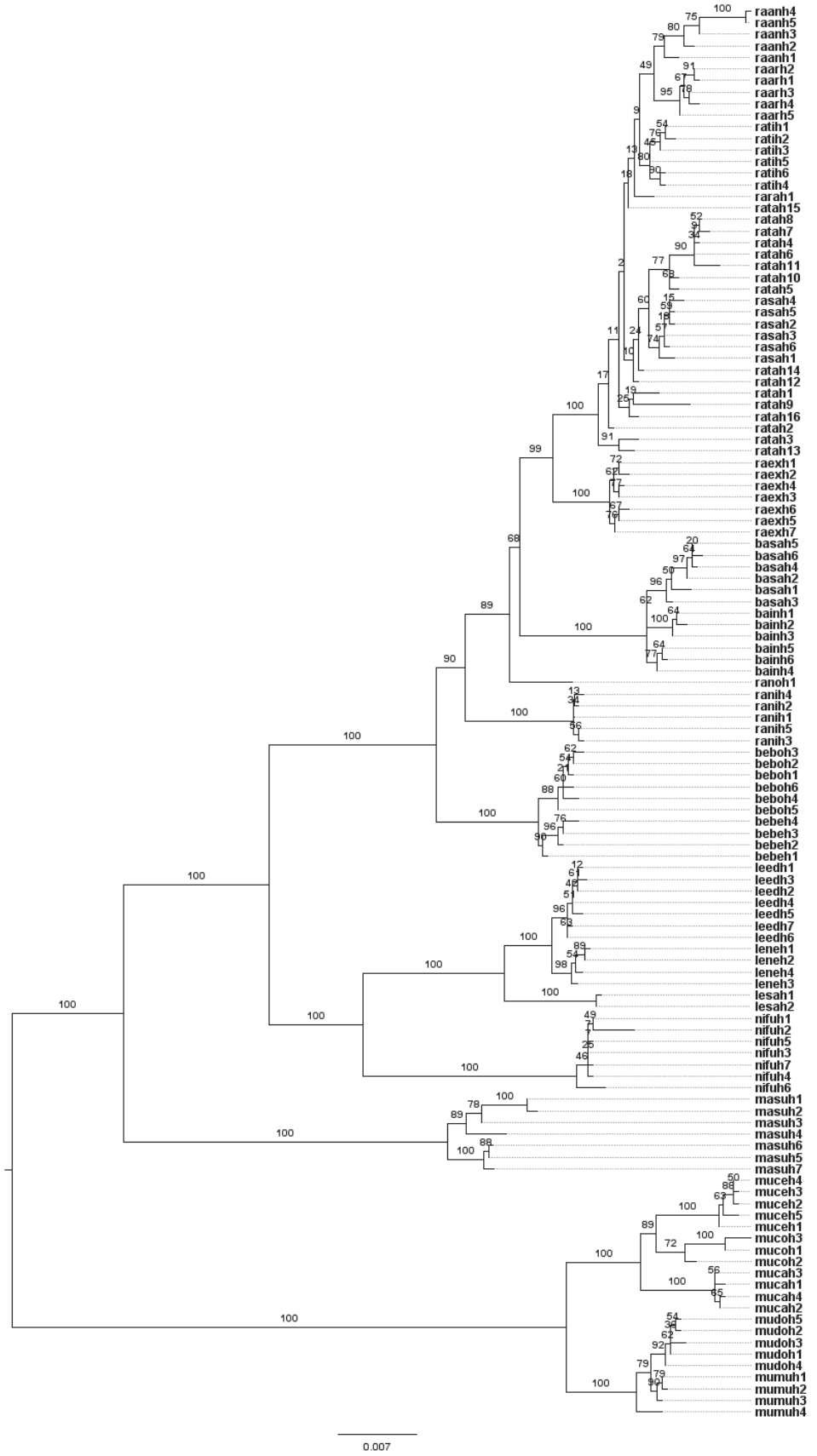


Fig. S3B, Phylogenetic trees based on the exon 3 of *Tlr7* gene reconstructed by maximum likelihood method in RAxML, Numbers above the branches indicates bootstrap support values, Names of haplotypes are explained in the Table S1.

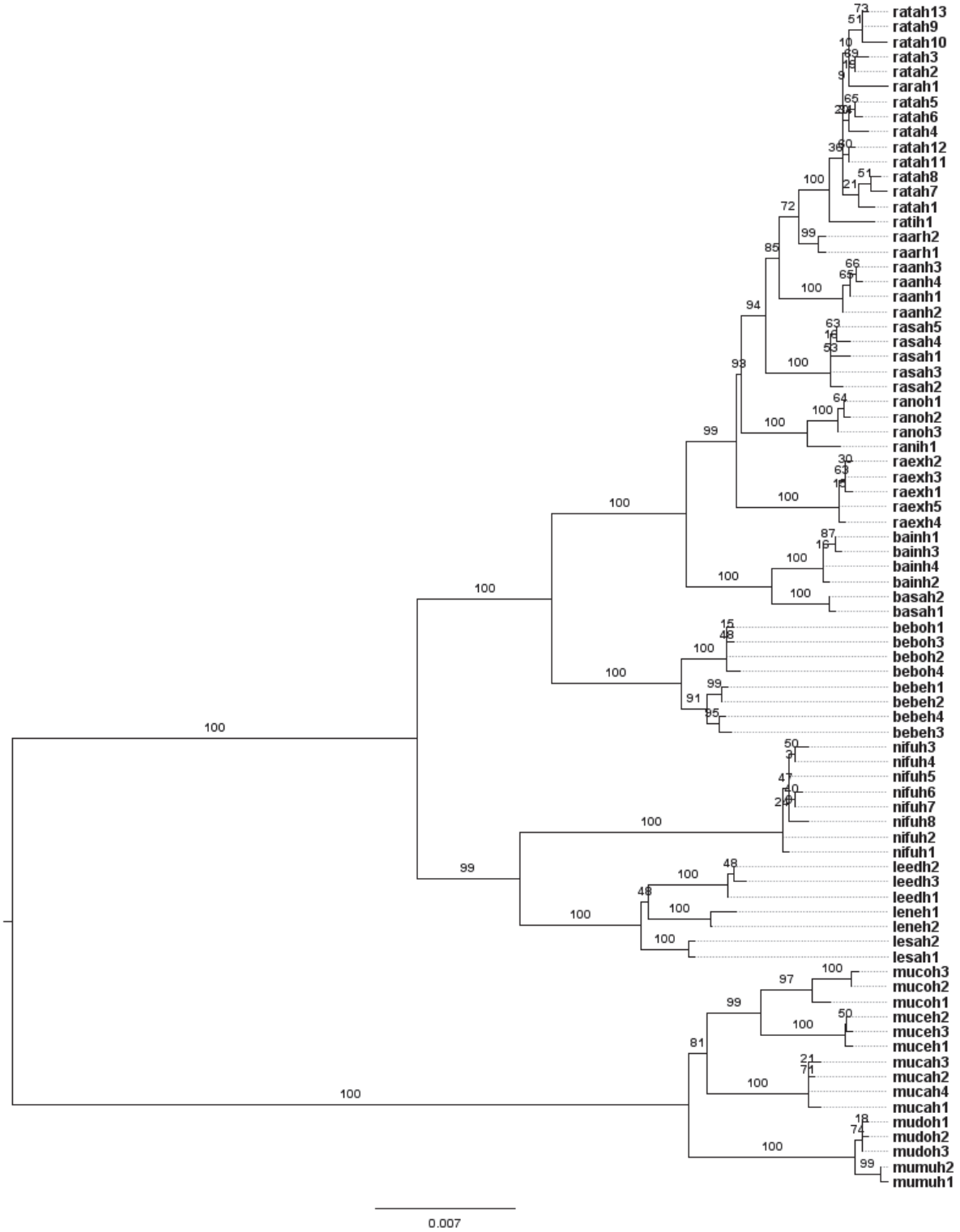
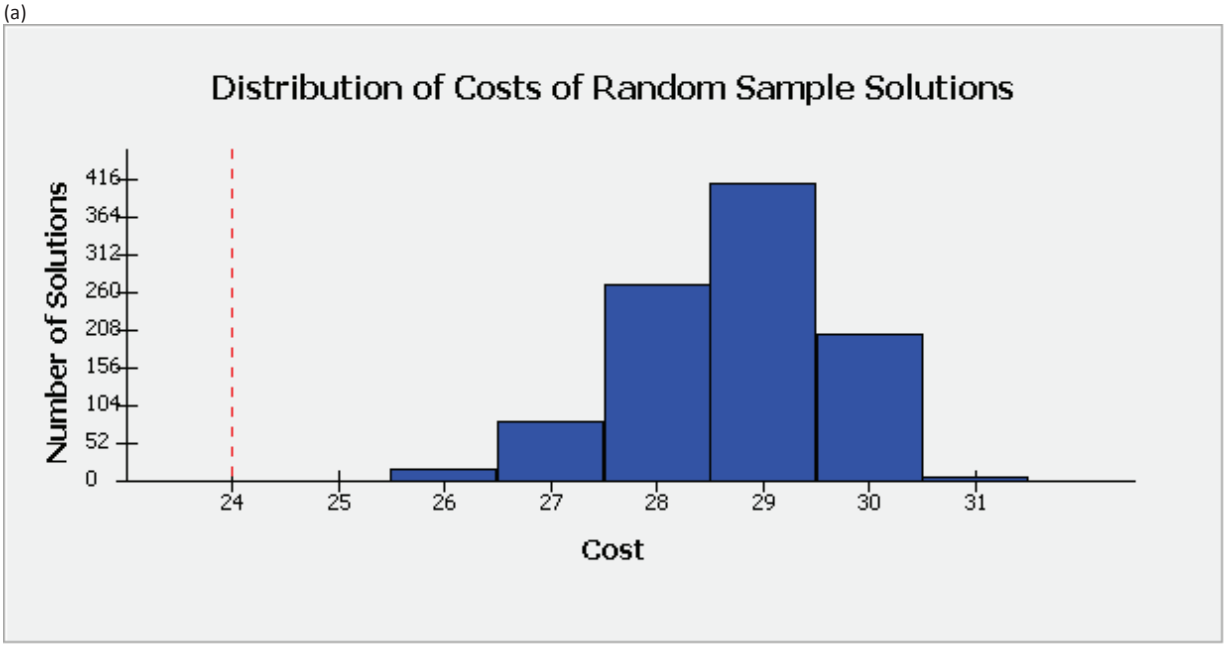


Fig. S4, Test of congruence between the presumably neutral and *Tlr* phylogenies (*Tlr4* (a), *Tlr7* (b) following JANE 4). Number at X axis represents costs of co-divergence. The red dashed line represents the cost observed in our data. The blue columns represent the random distributions of costs. Lower cost than random observed in our data signified higher congruence between species and gene topologies.

(a)



(b)

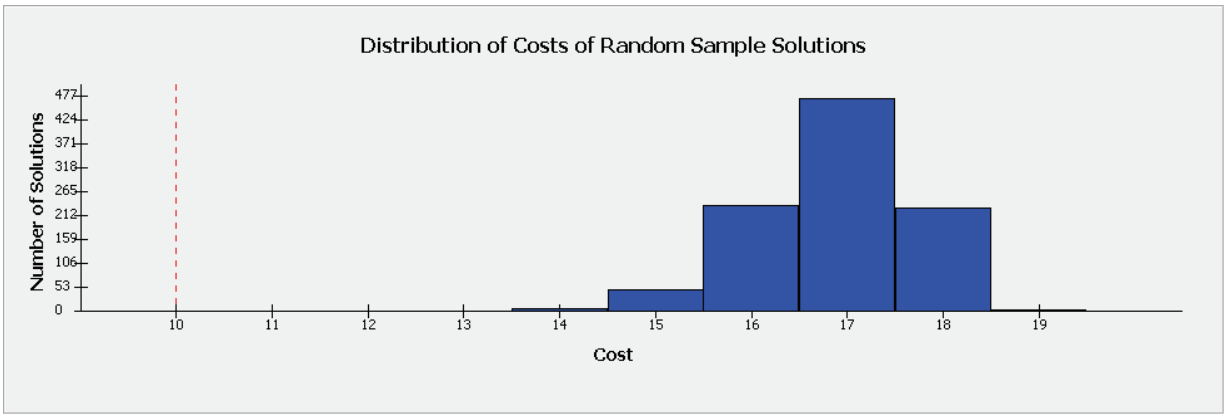
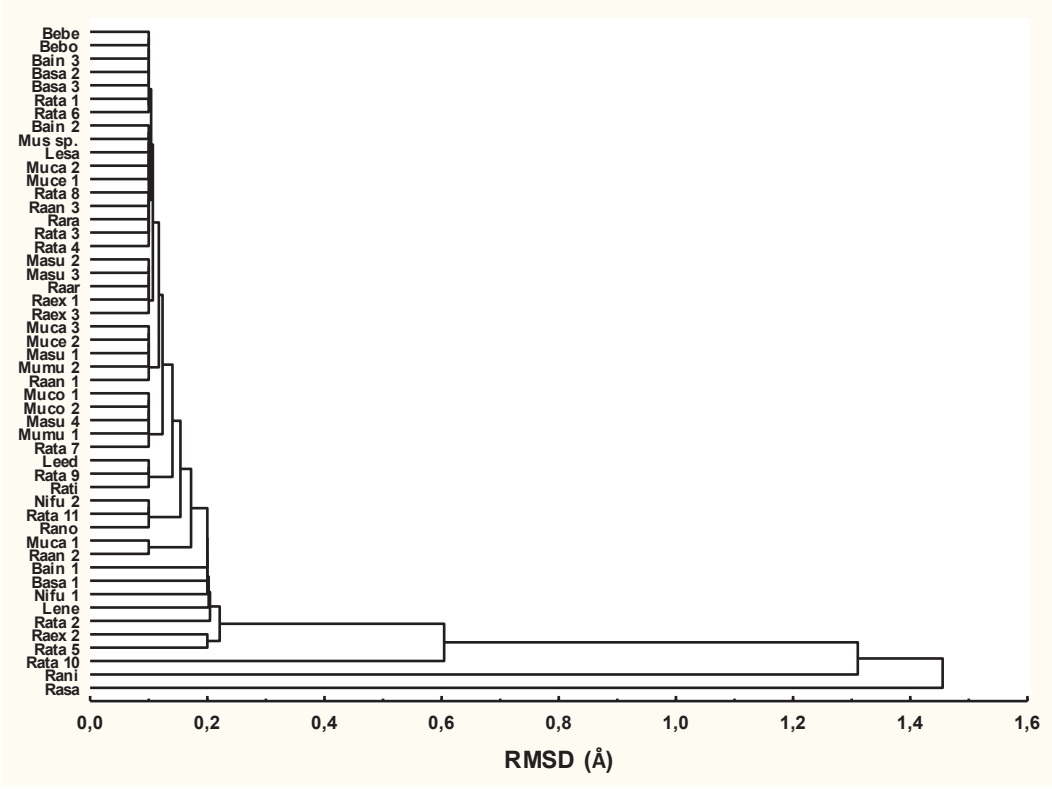


Fig. S5, Superimposition of structures, tree clustering diagrams based on linkage distance, (a) LBR_{TLR4} and (b) LBR_{TLR7}; individual LBR-variants often unify more species; description of LBR-variants labels is in the Table S1 under Hap_LBR_{TLR4} and Hap_LBR_{TLR7}.

(a)



(b)

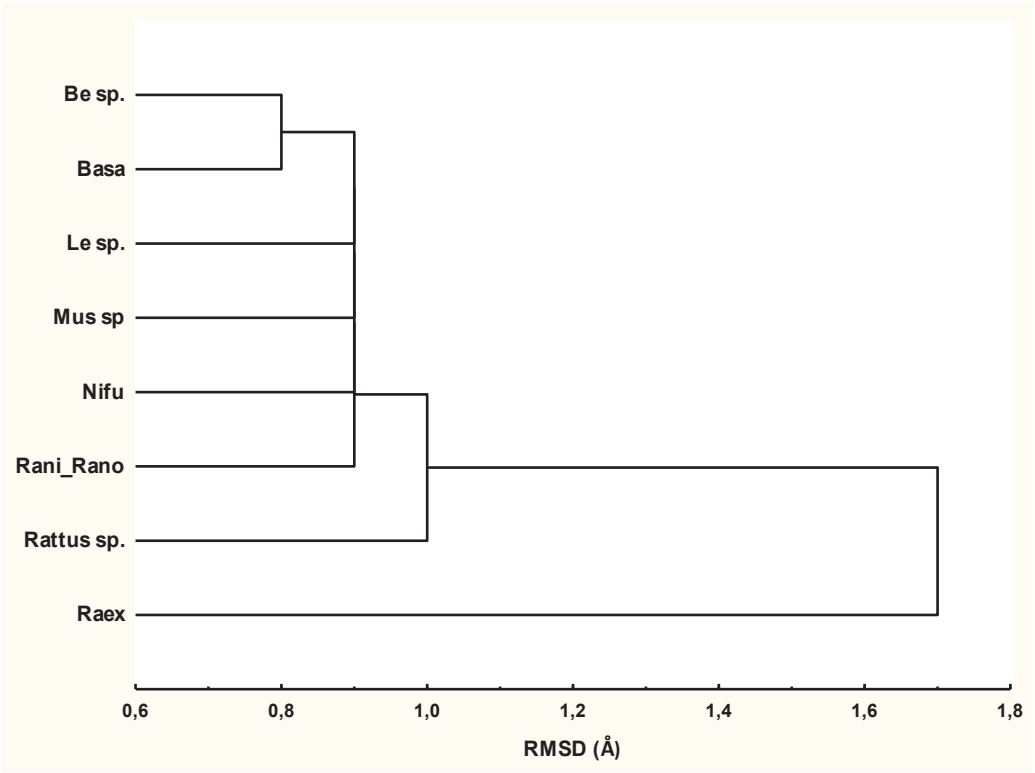
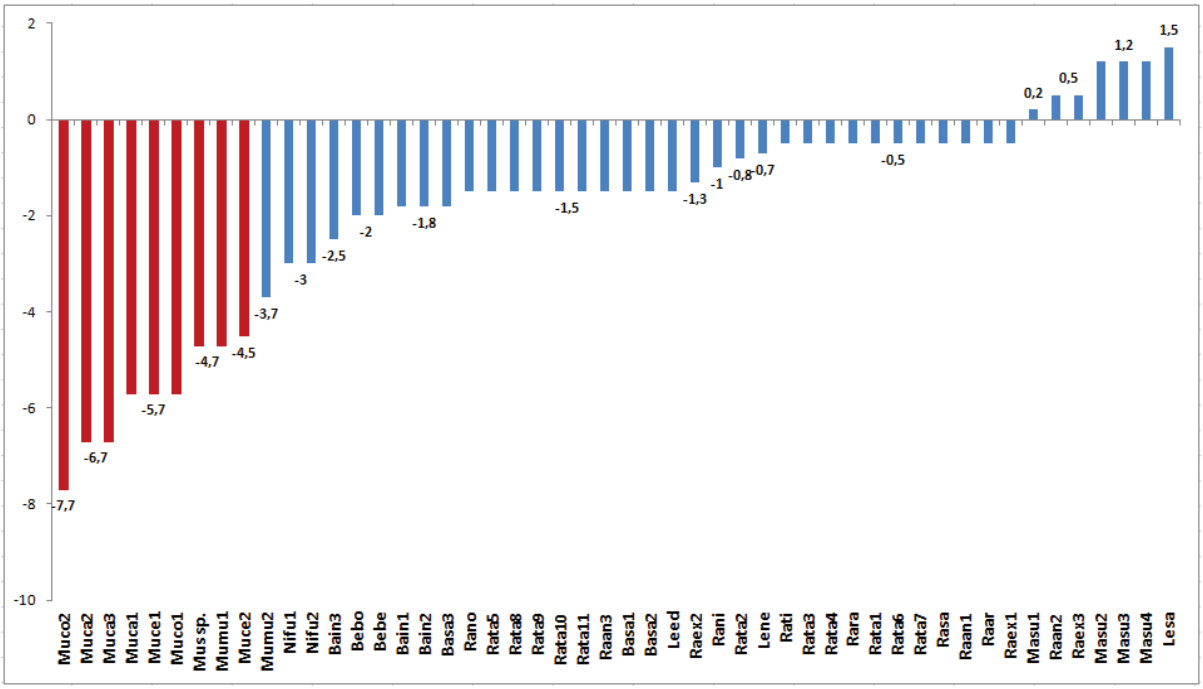
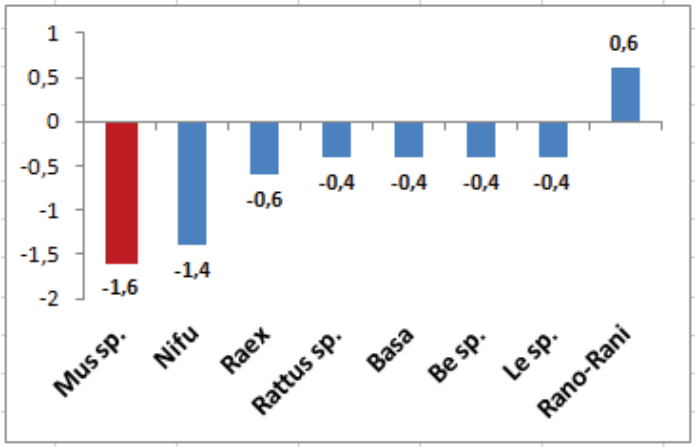


Fig. S6, Analysis of LBR amino acid sequence charge at pH 7 (LRRFinder) for (a) LBR_{TLR4} and (b) LBR_{TLR7}, individual LBR-variants often unify more species; description of LBR-variants labels is in the Table S1 under Hap_LBR_{TLR4} and Hap_LBR_{TLR7}. Mouse species are in red, *Rattus* spp. and related genera are in blue.

(a)



(b)



4 GENERAL DISCUSSION

Recapitulation of main objectives

The aim of this thesis was to describe Toll-like receptor variability at intra- and inter-specific levels in murine rodents and to search for imprints of selection at these two levels. Firstly I have described the polymorphism of five bacterial sensing TLRs of wild derived strains of Mus musculus. This first insight into the TLR variability was followed by detailed analysis of TLR4 in allopatric population of M. m. musculus and M. m. domesticus. Finally two receptors from different TLR families (bacterial sensing TLR4 and viral sensing TLR7) were studied at macroevolutionary level represented by 23 wild murine rodents from Southeast Asia. Therefore obtained results allow to analyse the evolutionary processes at two time scales: first is provided by diversification of Mmm and Mmd (0.5mya) and second by separation of Mus and Rattus lineages (12.5mya).

At the beginning of my PhD study (in 2010-2011) information concerning evolutionary processes acting on TLRs was rather scarce. In general TLR research was based mostly on laboratory or domestic animals or human and many questions concerning TLR evolution in natural conditions remained without answer. In first reviews TLRs were presented as ancient and conservative immune receptors common for all eumetazoans, providing bridge between both arms of immune system (Akira et al. 2001; Roach et al. 2005; Leulier and Lemaitre 2008). The role of two main subfamilies (bacterial and viral sensing) was well documented as well as the TLR structure and domain organization (Akira et al. 2006; Werling et al. 2009). Strong negative selection was detected in intracellular TIR domain, which is responsible for signal transduction, in contrast to extracellular domain which is directly responsible for ligand binding (Smirnova et al. 2000; Matsushima et al. 2007; Zhou et al. 2007; Wlasiuk and Nachman 2010). First studies showing important *Tlr* polymorphism focused on *Tlr4* in 35 laboratory mice strains and six domestic chicken inbred lines (Smirnova et al. 2000; Leveque et al. 2003). In both studies non-synonymous substitutions were detected representing potential for diversity in pathogen detection and immune response. Non-synonymous polymorphism was described also in TLRs of other domestic animals (Jann et al. 2008; Palermo et al. 2009; Seabury et al. 2010; Jungi et al. 2011; Raja et al. 2011; Malik 2011) and human (Ferwerda et al. 2007). Differences in signalling with functional significance were confirmed for some non-synonymous SNPs for example in

TLR3 in wild derived mouse strains (Stephan et al. 2007) and many others non-synonymous SNPs were associated with variable response to diverse diseases in human (Arbour et al. 2000; Lorenz et al. 2000; Kang and Chae 2001; Ogus et al. 2004; Hawn et al. 2005; Schröder and Schumann 2005; Rezazadeh et al. 2006; Ferwerda et al. 2007). However studies of free-living animals were missing in spite of the known reality that they differ in immune response from domestic or laboratory animals (Vinkler and Albrecht 2009; Abolins et al. 2011; Pedersen and Babayan 2011). This was the main reason, why I decided to study TLR polymorphism in free living murine rodents.

My main aim was to track signals, direction and strength of different evolutionary processes shaping TLRs in free living animals. At the beginning of my thesis I asked myself many questions: What is the variability in mouse WDS? Is this variability represented also in free living rodents? Could we observe some pattern of selection acting on TLRs in recently derived mouse subspecies? What about interspecific level? Are TLR variants species-specific or can we see trans-species polymorphism, which is commonly observed in MHC genes? Will we be able to detect signals of recombination, a mechanism successfully applied in adaptive immunity (e.g. Takahata and Satta 1998)? Are there differences between extracellular and intracellular domains? Are sites under selection (if there are any) accumulated to the specific region? Are there differences between different TLRs? etc..

At the end of my thesis I can tell that I am able to answer many of these questions; however some still remain with question mark and the new ones came into my mind so further research is surely required to understand evolutionary mechanisms acting on these important parts of innate immunity.

4.1 Selection forces acting on TLRs in free living populations: intra- vs. inter-specific level

Intraspecific or population level can provide evidence about ongoing selection processes and local adaptations, while interspecific comparison takes into consideration deeper timescale. Studies based at interspecific level detected signals of positive selection acting on distinct TLRs in mammals, i.e. rodents, primates or for example cetaceans (Nakajima et al. 2008; Ortiz et al. 2008; Wlasiuk and Nachman 2010; Tschirren et al. 2011; Shen et al. 2012) as well as in birds where few sites under positive selection were also detected in LBR (Alcaide and Edwards 2011), and in fish in TLR9 and TLR22 (Chen et al. 2008; Sundaram et al. 2012). It was also found that selection shaping TLRs do not act in the same way and direction and we can observe this pattern by signals of positive selection at distinct lineages. Great example was shown in cetaceans, where strong positive selection ($dN/dS > 1$) was found at TLR4 in the branch leading to hippos and whales. Authors suggest drastic change in pathogene environment due to habitat transition of the terrestrial ancestor of cetaceans from terrestrial to semi-aquatic habitat (Shen et al. 2012). Another strong signature of positive selection was detected in branch leading to oceanic dolphins which experienced rapid adaptive diversification. Therefore Shen et al. (2012) suggest an important role of adaptive evolution of TLR4 in cetaceans. Similar pattern was observed also in extracellular domain of TLR4 in primates during evolution of *Catarrhini* lineage (Nakajima et al. 2008; Wlasiuk and Nachman 2010). Wlasiuk and Nachman (2010) conclude that adaptive evolution play an important role in *Tlr* evolution, but is rather episodic in the nature.

Our results suggest that evolution of murine TLR4 and TLR7 is shaped predominatly by purifying selection, which eliminates majority of non-synonymous substitutions to maintain correct function of proteins. However signals of positive selection were found in both receptors. Overall, the phylogeny of vertebrate TLRs correctly recapitulates the phylogeny of vertebrate species and trans-species polymorphism seems to be very rare pattern. Therefore discrepancies between species phylogeny and genealogy can result from strong pathogen selective pressures acting on several lineages. Departure from neutrality represented in our dataset by phylogeny based on three non-immune genes was found in both TLRs. In TLR4 discrepancie between species phylogeny and genealogy concerned *Bandicota* spp. Relationships between *Bandicota* spp., which diverged quite recently, and the rest of *Rattus* genus are still not well resolved, therefore observed pattern could be the result of ILS (Pagès et al. 2010). In TLR7 discrepancies were found between *Rattus exulans* and *Rattus*

sakeratensis. Both species belong to well supported lineages and therefore more probable seem the explanation of coevolution with some pathogens.

At the intraspecific population level, the evidence for positive selection is much more complicated and results of previous studies are not congruent. In human it was found that some populations have evolved under strong purifying selection (TLR2, TLR4 and TLR9) while the TLR evolution of the other populations show patterns of balancing selection (TLR1, TLR6, TLR9 and TLR10) (Ferrer-Admetlla et al. 2008; Mukherjee et al. 2009). Differences between populations were explained by local adaptation to specific pathogens and to regions with a high microbial load in the study of Mukherjee et al. (2009). In contrast Ferrer-Admettla et al. (2008) rejected the hypothesis of adaptation to the pathogenic environment as they did not find any differences between studied populations. Signals of positive selection were detected also within species of primates or rodents (Wlasiuk and Nachman 2010; Tschirren et al. 2012; Tschirren et al. 2013), however authors hypothesized potential role of population demographic changes or selective sweep with some linked locus rather than selection on *Tlr* itself. Tschirren et al. (2013) revealed causal relationship between TLR2 polymorphism and bank vole resistance to *Borrelia afzelii*. The associations with parasite prevalence showed that the most frequent allele is also the most advantageous. Nevertheless they admitted also the possibility of hitchhiking by linked locus, which could have reduced TLR2 polymorphism in yellow-necked mouse (Tschirren et al. 2012).

In our data we revealed moderate variability at intraspecific level in two sub-species of *Mus musculus*, but no signature of selection appeared at molecular level when we analysed both subspecies separately despite the extensive geographic sampling. By comparison of subspecies we found that *Tlr4* variability was lower in Mmd than in Mmm. Our results thus confirmed the pattern observed previously in WDS. Mmd harboured one dominant haplotype (prevalent in 71% of Mmd) widespread across all Western Palaearctic region and only single variant of the ligand binding region. In contrast, *Tlr4* of Mmm was much more polymorphic with four haplotypes at intermediate frequencies (around 20-30%) and we also found clear signal of recombination between two principal *Tlr4* haplogroups. Majority of amino acid substitutions between both subspecies were rather functionally neutral with respect to their biochemical proparties, but it was shown that even such substitutions can change significantly protein function (Nei and Nozawa 2011).

Analysis of mitochondrial cytochrome *b* gene sequence showed comparable variability and pattern of haplotype network between two subspecies. Even if we can not reject completely the effect of demographic processes (and some indication of western expansion

during colonization are visible in *Tlr4* of Mmd, but not in Mmm), we conclude that observed differences in *Tlr4* diversity could be caused by contrasting parasite-mediated selection acting on the two subspecies. Several hypothesis could be formulated. For example, we can hypothesize that observed pattern was caused by single parasite specific to Mmd, i.e. not occuring in Mmm, which could have resulted in expansion of one advantageous dominant *Tlr4* allele in Mmd. Western Europe was colonized by Mmd across southern regions with higher biodiversity, including presumably higher prevalence of certain pathogens (see for example distribution area of visceral leishmaniasis Fig. 21) (Lysenko 1971) that could have formed strong selective pressure to Mmd immune system. Nevertheless it is only a speculation, even if the role of TLR4 during recognition and elimination of *Leishmania* was already described in laboratory mice (Tuon et al. 2008). Alternative scenario suggests cyclic long-term co-evolution with pathogens, which is characteristic by occasional selective sweeps in some lineages or populations. According to this hypothesis both subspecies can be now situated in different stages of *Tlr4* evolution manifested in Mmd by overdominance of one allele while in Mmm higher variability can be observed. Hypothetically, if this sweep would occur during the stay of Mmd in Levant before expansion to Western Europe, then we could observe reduction in Mmd *Tlr4*, but not in *mt-Cytb*. This hypothesis could be confirmed by extensive sampling in Mediterranean basin and in North India to compare *Tlr4* variability in both regions. However with present data we are not able to explain our results definitively.

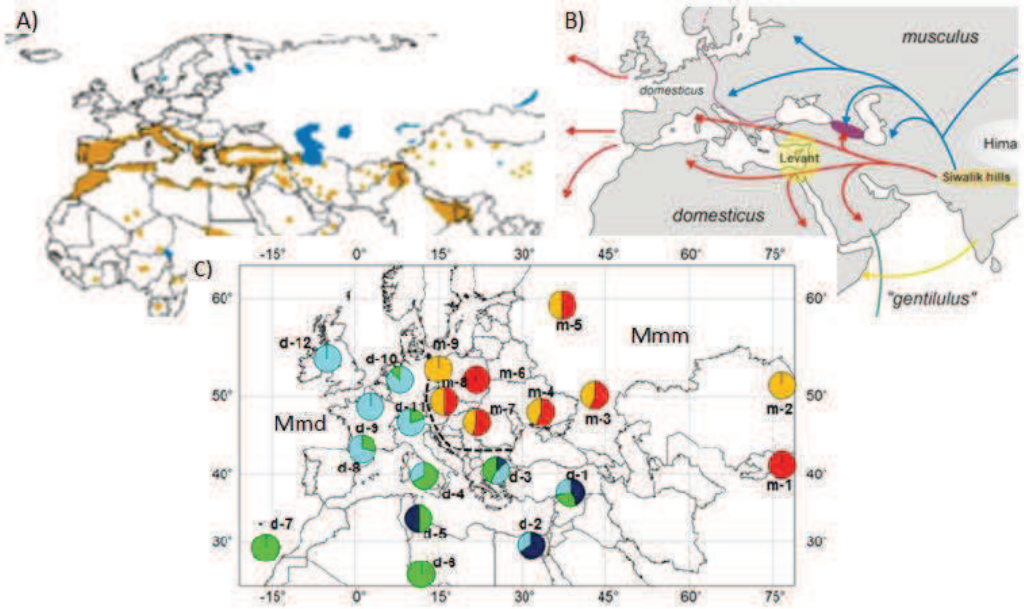


Fig. 21, Hypothesis about the origin of reduced *Tlr4* variability in Mmd, A) Geographical distribution of visceral leishmaniasis in the Old World. http://www.who.int/leishmaniasis/leishmaniasis_maps/en/, Data source & Map production: WHO/NTD/IDM HIV/AIDS, Tuberculosis and Malaria (HTM) World Health Organization, October 2010. B) Colonization of western Palearctic by the house mouse; Mmd is in red. C) Distribution of main *Tlr4* haplogroups of Mmd and Mmm.

4.2 The role of recombination: instrument of stochastic processes or selection

In the arm race between parasites and hosts, maintaining of the polymorphism of genes involved in pathogen recognition is crucial for survival of hosts. Therefore all mechanisms favouring arising of new advantageous alleles and their combination would be favoured. Recombination can be such mechanism and can play an important role in producing genetic diversity at intraspecific level. Recombination can very quickly create new variants and its important role in the evolution of immune genes was shown for example in the adaptive immune branch, especially in MHC genes (Takahata and Satta 1998). Recombination is the source of new combinations and therefore might has an appreciable role also in creating polymorphism of TLRs.

In our data we revealed the signal of the recombination in Mmm where recombination takes place between two main haplogroups (HG-Im and HG-IIm). Detailed analysis of our sequences suggests that two haplotypes H_2 and H_19 are recombinants comprised of ECD from haplogroup HG-IIm and ICD of HG-Im. Both haplotypes arised independently in specimanes separated by distance of at least 500km and therefore we suggest that recombination may be frequent in this region of *Tlr4*. Combination of variable ECD and conservative ICD with TIR domain can be in fact very succesfull strategy how to maintain the variability of ECD and at the same time correct function during signalling. Even if we detected signals of recombination only in Mmm, recombination could occur in both subspecies of *Mus*. The signal was visible only in Mmm probably due to low variability in Mmd. In spite of the fact that signals of recombination were already detected in the ECD of human TLR4 (Zaki et al. 2012) and bovine TLR3, TLR4 and TLR10 or in birds (Seabury et al. 2010; Alcaide and Edwards 2011), the relevant tests of recombination have not been performed in most studies of TLR evolution. In contrast to MHC genes where trans-species polymorphism if very often, we did not find any pattern in our data.

4.3 Selection forces acting on TLRs: bacterial sensing vs. viral sensing

Strength of selection vary also between individual TLRs. For example positive selection was detected in primate TLR4 and TLR1, while no signal of positive selection was found in TLR5 (Wlasiuk and Nachman 2010). Diversity in selection between distinct TLRs is generally explained by spectrum of targeted PAMPs.

One of the classifications of TLRs is based on the TLR localization, which is related to ligand types. Intracellular or viral-sensing TLRs are considered as essential receptors with non-redundant biological role in host survival. Viral proteins are in general poor targets for innate recognition, because they are evolving very rapidly. Intracellular TLRs thus trigger specific properties of viral nucleic acids which are difficult to change for viruses. Because they are involved in recognition of nucleic acids they have to prevent self nucleic acid recognition leading to autoimmune responses. Self RNAs (which could be detected by TLR7) are subject to degradation by extracellular RNases and only rarely reach the endocytic compartment where TLR resides (Diebold et al. 2004). Another viral sensing TLR, TLR9, recognize non-methylated DNA which is specific to pathogens, because mammals have their DNA methylated (Jong et al. 2010). Therefore it is not surprising that they are under strong selective pressure to maintain correct function (Krieg and Vollmer 2007; Barrat and Coffman 2008; Waldner 2009; Wlasiuk and Nachman 2010; Casanova et al. 2011). Moreover it is possible that viruses are putting stronger selective pressure on immunity sensors than other microbes.

In their recent study, Areal et al. (2011) contradicts the general observations by the finding that viral sensing TLR8 in mammals had the same number of positively selected codons as TLR4, which was considered as one of the most variable TLRs showing strongest pattern of positive selection (Wlasiuk and Nachman 2010; Areal et al. 2011). TLR7 and TLR8 arose in mammals by duplication and are located in tandem at chromosom X, therefore their role is analogous and TLR8 detects similar ligands as TLR7 (ssRNA). Therefore one would expect that evolutionary pattern will be also similar between them. Surprisingly, Areal et al. (2011) showed that TLR7 belongs among the most conservative TLRs. Their finding is explained by wider sampling across different mammalian groups such as ungulates, carnivores, primates, lagomorphs or rodents and potential broader spectrum of pathogens involved in TLR8 signalling, in comparison to previous studies where TLR8 was considered to evolve under negative selection (e.g primates, Wlasiuk and Nachman 2010 and birds,

Alcaide and Edwards 2011). Other viral sensing TLRs had low proportion of sites under positive selection (e.g. TLR7 detecting the same ligand as TLR8 has only 0.67%, TLR3 = 0.99%, TLR9 = 0.39%) compare to 2.5% sites under positive selection found in TLR8 (Areal et al. 2011). It seems that TLR8 is rather exception among viral-sensing TLRs. However we admit that description of ligands for TLR8 is not definitive and that spectrum of viral motifs detected by viral-sensing TLRs could play an important role in TLR variability and further research should focus at this problematic.

In mammals two viral sensing receptors, TLR7 and TLR8, are placed at chromosome X. Because X chromosome evolve under specific selective constraints, their position could be thus considered as possible evolutionary advantage in the arm-race, because deleterious recessive mutations are removed more efficiently (Vicoso and Charlesworth 2006). Nevertheless position on X chromosome can cause also problems. Several studies documented differences between females and males in immune response to viral and autoimmune diseases, probably due to TLR localization on chromosome X (Shen et al. 2010; Wang et al. 2011). On the other hand two viral sensing TLRs in birds (TLR7 and TLR21), which also evolved under strong negative selection, are placed at autosomes, therefore effect of X chromosome on the TLR evolution is questioned (Alcaide and Edwards 2011).

Constrained evolution was detected also in other PRRs involved in viral detection. For example human RIG-like receptors (RLRs) were shaped by natural selection in different direction and intensity (Vasseur et al. 2011; Vasseur et al. 2012). RIG-I which is involved in detection of viruses is the most constrained receptor in this group. Low level of nucleotide diversity, low tolerance to amino acid-alternation, lower frequencies of non-synonymous mutations were explained by pressure of viruses. Viral dsRNA represent only short ligand and therefore its detection requires more strict binding sites. Vasseur et al. (2011) suggested also alternative role of RLR beside PAMP recognition for example in embryonal development, what has been already shown for NOD receptor. Other two receptors of RLRs were found to evolve under more relaxed constraints (IFIH1 and LGP2) (Vasseur et al. 2011). In the study of Vasseur et al. (2012) endosomal TLRs and NALPs were found under stronger selective constraint than cell-surface TLRs and for example NOD/IPAF subfamily. For detailed overview of PRRs and their evolutionary dynamics see Fig. 22.

In my thesis I studied only one receptor of this interesting subfamily, TLR7. We confirmed the general observation among bacterial and viral sensing TLRs. In comparison to TLR4, which belong among bacterial sensing TLRs, TLR7 was under much stronger negative selection across murine rodents. Overall *dN/dS* ratio was in TLR7 = 0.398 (similar to ratio

found in birds, where *dN/dS* was 0.386, Alcaide and Edwards 2011) while in TLR4 at the same dataset *dN/dS* was 0.481 (in birds the mean ratio *dN/dS* of TLR4 was 0.517). Both receptors differed especially in region responsible for ligand binding where TLR4 had $dN/dS_{LBR} = 0.787$, while TLR7 only $dN/dS_{LBR} = 0.196$.

Bacterial-sensing (or cell-surface) TLRs, on the contrary, have been shown to evolve under much more relaxed selective constraints, what is also the case of TLR4 in our study. Among these TLRs we can found pseudogenes (caused mostly by substitutions causing stop codons, for example TLR5) (Roach et al. 2005, this thesis - Chapter 3.1) and higher number of non-synonymous substitutions, sometime tolerated and positively selected and sometime with slightly deleterious effects (Wlasiuk and Nachman 2010). This is often explained by greater redundancy and overlapping functions with other PRRs. Partial redundancy might be therefore the important strategy, which permits to part of PRRs to evolve and run with pathogens while others receptors are in the guard. Relaxed selective constraints may also enable fine-tuning of host defenses.

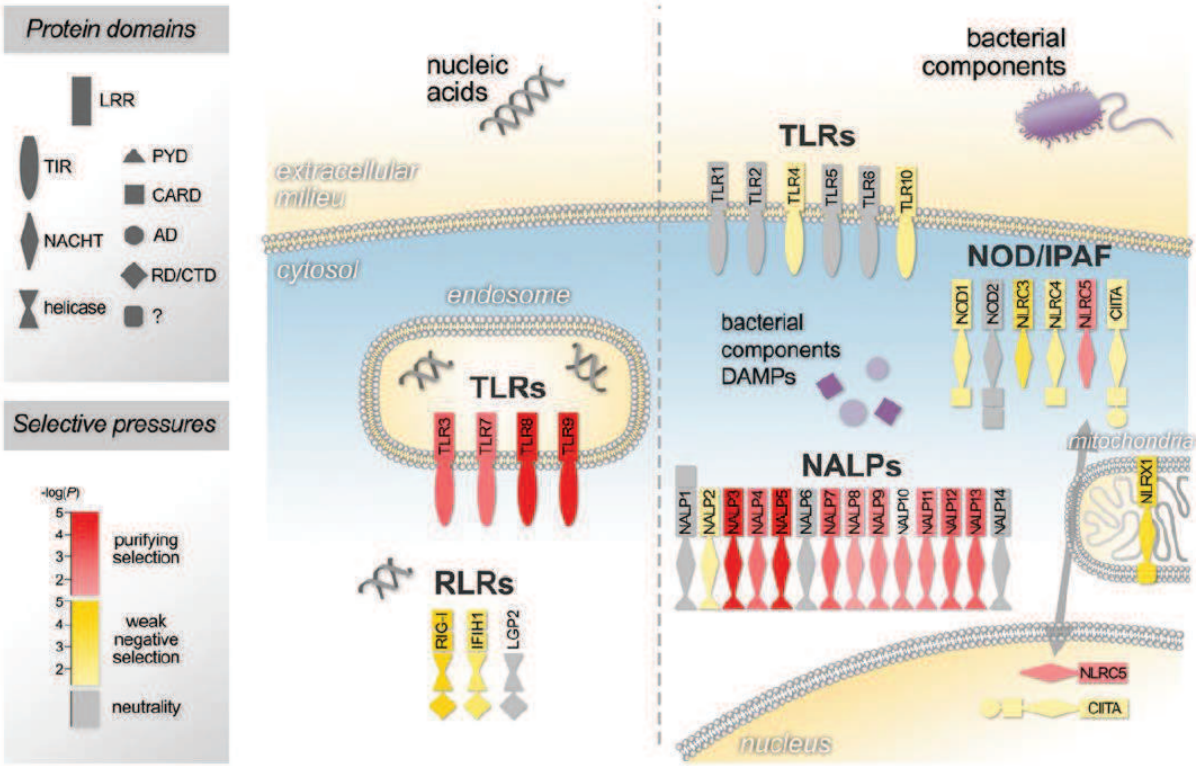


Fig. 22, Hierarchical model outlining the evolutionary dynamics and biological relevance of the various families of PRRs. This representation is based on the intensity of the selective constraints detected for the 34 PRRs. These analyses allowed us to distinguish three groups of genes: genes under purifying selection ($u < 1$, in red), genes under weaker selective constraints ($g < 0$, in yellow), and genes for which no deviation from neutrality was detected (in gray). Color intensity is proportional to the –log(p-value) of u or g tests. Cellular sublocalization, protein domains, and ligands are given as an indication but are not exhaustive. Adopted from Vasseur et al. 2012.

4.4 Selection forces acting on TLRs: ECD vs. ICD

Previous interspecific comparisons of TLRs revealed different selective pressure acting on different parts of TLR proteins (Mikami et al. 2012). Comparison of *dN* and *dS* in ten primate TLRs revealed that with respect to *dS* there is no difference between TIR and LRR in ECD. This indicates that both domains are evolving at similar mutation rate. However when both domains were compared with respect to *dN*, TIR contains lower number of *dN* in comparison to LRR (Hughes and Piontkivska 2008). The observation that cytoplasmic TIR domain evolve under stronger purifying selection than ECD (Matsushima et al. 2007; Wlasiuk and Nachman 2010; Alcaide and Edwards 2011; Smith et al. 2012) was explained by the essential role of TIR domain in signalling pathway and interaction with other binding partners such as MyD88, TRIF, MyL and TRAM. Moreover TIR domain does not interact with PAMPs and therefore there is no reason to keep variability in this part of TLR protein. By contrast faster evolution of LBR was confirmed by many studies (Zhou et al. 2007; Jann et al. 2008; Wlasiuk and Nachman 2010; Alcaide and Edwards 2011; Huang et al. 2011; Mikami et al. 2012; Smith et al. 2012). Mikami et al. (2012) also found that evolutionary rate of LRR domains differ between diverse vertebrate TLRs, probably as the result of co-evolution with distinct PAMPs. Lowest evolutionary rate of LRR domains was found in two viral sensing TLRs TLR7 and TLR3 (Mikami et al. 2012). Moreover 81% changes in ECD of bird TLRs were non-synonymous (Downing et al. 2010).

In our data of murine rodents we revealed similar differences between ECD and TIR domain with respect to the number of *dN*, but not to the number of *dS*. The ratio *dN/dS* was significantly higher in ECD and especially in specific LBR than in TIR domain (for example in murine TLR4 $dN/dS_{LBR} = 0.787$, while $dN/dS_{TIR} = 0.067$). Comparison of TLR4 and TLR7 revealed similar pattern in both receptros ($dN/dS_{LBR} > dN/dS_{TIR}$), however evolutionary rate was higher in LBR of TLR4 than TLR7 (TLR7 $dN/dS_{LBR} = 0.196$). In TLR4 92% of positively selected sites were placed directly in ECD and overwhelming majority of these sites were directly in LBR. In TLR7 all of those sites were found in ECD, but non in LBR. All these results together suggest much faster evolution of ECD resulting from arm race with pathogens and important role of TIR domain involved in signalling. Strong purifying selection acting at TIR domain was confirmed also in murine TLRs, where we found in 22 species only 6 amino acid variants of TLR7 TIR domain and in 23 species only 11 variants of TLR4 TIR domain. Rather than trans-species polymorphism, we suggest that strong purifying selection on the correct signalling is responsible for observed pattern.

4.5 TLRS IN SPECIATION RESEARCH - FUTURE PROSPECTS

Generally, immunity genes, or their part interacting with pathogens, are assumed to evolve rapidly due to selective pressure posed by pathogens with faster life cycles than hosts. Present debate concerning the role of immune-related genes in evolution of reproductive isolation in diverging taxa has generated two alternative viewpoints. The first view considers immune genes as elements accelerating speciation. Due to breaking-down the co-adapted immune gene complexes and/or excess of allelic diversity inter(sub)specific hybrids are expected to exhibit higher parasite loads and selection then acts against these hybrid/recombinant hosts strengthening the isolation (Sage et al. 1986; Eizaguirre et al. 2009). The second hypothesis, on the contrary, views the immune genes as elements likely to flow across (sub) specific barriers. This hypothesis considers breaking-down of the co-adapted immune gene complex as less probable due to the fact that in source population polymorphism is maintained. Therefore bigger is the polymorphism of studied genes less likely would be hybrids penalized by presence of gene copies from source populations. Rather than reinforcing the isolation, the higher numbers of alleles brought by diverged taxa into the contact zones, together with novel variants produced by recombination, ensure a vigorous immune function and lower parasite loads in hybrids (Baird et al. 2012). The European house mouse hybrid zone seems an ideal study system in which to evaluate these alternatives. The first alternative predicts that immune genes will behave in similar way as genes contributing to the species barrier (e.g. those in the X chromosome, Dufková et al. 2011; Macholán et al. 2011; Janoušek et al. 2012). Under the second alternative immune genes will resemble more closely neutral genes, introgressing freely across the hybrid zone. The clear phylogenetic structure of *Tlr4* across the subspecies should give us the analytic power to test these two alternatives in the hybrid zone.

4.6 CONCLUSION

We found that evolutionary dynamics of TLRs genes in murine rodents are driven by various evolutionary forces (genetic drift, directional selection, negative selection etc.) which differ in strength and direction between species.

1) In general negative or purifying selection seems to be the dominant type of selection, which is common evolutionary mechanism responsible for maintaining correct function of important genes. Signals of positive selection were detected only at interspecific or "intersubspecific" level and mainly in extracellular domain. Contrary analysis of data based on intraspecific level did not revealed any site under positive selection and only few sites under purifying selection. Nevertheless pattern of genetic variation suggests strong effect of directional selection in Mmd and possibly balancing selection in Mmm. Inspite of the weak signal of selection we revealed moderate variability of TLR4 in both subspecies.
2) Evidence of recombination between ECD and TIR domain was detected in Mmm and we suggest that this mechanism contributed to the TLR polymorphism in the house mouse.
3) Negative selection was more pronounced in TLR7 representing viral-sensing family in this study. The most parsimonious explanation is the necessity to maintain their correct function and at the same time to prevent self detection causing disorders in immunity response. Differences between TLRs can be also caused by variance in immunological redundancy with other PRRs due to distinct contributions to host defense.
4) TIR domain was under stronger selective constraint than the rest of TLRs in all studied genes. The explenation is probably consist in its important function during signalling and close cooperation with other partners. Signalls of positive selection were concentrated to the ECD which is responsible for ligand recognition and therefore involved in the arm race with pathogens. The majority of sites under positive selection in TLR4 was in LBR.

5 REFFERENCES

Abolins SR, Pocock MJO, Hafalla JCR, Riley EM, Viney ME. 2011. Measures of immune function of wild mice, *Mus musculus*. Mol. Ecol. 20:881–892.

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

Achyut BR, Ghoshal UC, Moorchung N, Mittal B. 2007. Association of Toll-like receptor-4 (Asp299Gly and Thr399Ileu) gene polymorphisms with gastritis and precancerous lesions. Hum. Immunol. 68: 901–907.

Agrawal AF, Lively CM. 2002. Infection genetics: gene-for-gene versus matching-alleles models and all points in between. Evol. Ecol. Research, 4, 79–90.

Agrawal AF, Lively CM. 2003. Modelling infection as a two-step process combining gene-for-gene and matching-allele genetics. Proc. R. Soc. B Biol. Sci. 270:323–334.

Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK. 2004. High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal. Proc. Natl. Acad. Sci. U. S. A. 101:3490–3494.

Akira S, Takeda K, Kaisho T. 2001. Toll-like receptors: critical proteins linking innate and acquired immunity. Nat. Immunol. 2:675–680.

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

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

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

Andersen-Nissen E, Hawn TR, Smith KD, Nachman A, Lampano AE, Uematsu S, Akira S, Aderem A. 2007. Cutting Edge: Tlr5−/− mice are more susceptible to escherichia coli urinary tract infection. J. Immunol. 178:4717–4720.

Anderson KV, Nüsslein-Volhard C. 1984. Information for the dorsal–ventral pattern of the Drosophila embryo is stored as maternal mRNA. Nature 311:223–227.

Andrews CA. 2010. Natural selection, genetic drift, and gene flow do not act in isolation in natural populations. Nature Education Knowledge 3(10):5.

Apanius V, Penn D, Slev PR, Ruff LR, Potts WK. 1997. The nature of selection on the major histocompatibility complex. Crit. Rev. Immunol. 17:179–224.

Aplin KP, Suzuki H, Chinen AA, et al. 2011. Multiple geographic origins of commensalism and complex dispersal history of Black Rats. PloS One 6:e26357.

Arbour NC, Lorenz E, Schutte BC, Zabner J, Kline JN, Jones M, Frees K, Watt JL, Schwartz DA. 2000. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat. Genet. 25:187–191.

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

Areschoug T, Gordon S. 2008. Pattern recognition receptors and their role in innate immunity: focus on microbial protein ligands. Contrib. Microbiol. 15:45–60.

Areschoug T, Gordon S. 2009. Scavenger receptors: role in innate immunity and microbial pathogenesis. Cell. Microbiol. 11:1160–1169.

Armelagos GJ, Goodman AH, Jacobs KH. 1991. The origins of agriculture: Population growth during a period of declining health. Popul. Environ. 13:9–22.

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:W529–533.

Auffray JC, Britton-Davidian J. 2012. The house mouse and its relatives: systematics and taxonomy. In *Evolution of the House Mouse* (ed. M. Macholan, S.J.E. Baird, P. Munclinger & J. Piálek). Cambridge Series in Morphology and Molecules. Cambridge University Press, pp 1–35.

Austin CM, Ma X, Graviss EA. 2008. Common nonsynonymous polymorphisms in the NOD2 gene are associated with resistance or susceptibility to tuberculosis disease in African Americans. J. Infect. Dis. 197:1713–1716.

Babayan SA, Allen JE, Bradley JE, et al. 2011. Wild immunology: converging on the real world. Ann. N. Y. Acad. Sci. 1236:17–29.

Badenhorst D, Tatard C, Suputtamongkol Y, Robinson TJ, Dobigny G. 2012. Host cell/Orientia tsutsugamushi interactions: evolution and expression of syndecan-4 in Asian rodents (Rodentia, Muridae). Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 12:1136–1146.

Baird SJE, Ribas A, Macholán M, Albrecht T, Piálek J, Goüy de Bellocq J. 2012. Where are the wormy mice? A reexamination of hybrid parasitism in the European house mouse hybrid zone. Evol. Int. J. Org. Evol. 66:2757–2772.

Baird SJE, Macholán M. 2012. What can the *Mus musculus musculus/M. m. domesticus* hybrid zone tell us about speciation? In: Macholán M, Baird SJE, Munclinger P, Piálek J, eds. Evolution of the House Mouse. Cambridge, Cambridge University Press: 334–372.

Balamayooran T, Balamayooran G, Jeyaseelan S. 2010. Toll-like receptors and NOD-like receptors in pulmonary antibacterial immunity. Innate Immun. 16:201–210.

Bandelt HJ, Forster P, Röhl A. 1999. Median-joining networks for inferring intraspecific phylogenies. Mol. Biol. Evol. 16:37–48.

Barrat FJ, Coffman RL. 2008. Development of TLR inhibitors for the treatment of autoimmune diseases. Immunol. Rev. 223:271–283.

Barreiro LB, Ben-Ali M, Quach H, et al. 2009. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. PLoS Genet. 5:e1000562.

Barreiro LB, Quintana-Murci L. 2010. From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat. Rev. Genet. 11:17–30.

Barton NH, Hewitt GM. 1985. Analysis of Hybrid Zones. Annu. Rev. Ecol. Syst. 16:113–148.

Bassett EH, Rich T. Introduction. In: Toll and Toll-like receptors: an immunologic perspective. Boston, MA: Springer US. p. 1–17. Available from: http://link.springer.com/content/pdf/10.1007%2F0-387-27445-6_1.pdf#page-1

Bazykin AD. 1969. Hypothetical mechanism of speciaton. Evolution 23:685–687.

Beck JA, Lloyd S, Hafezparast M, Lennon-Pierce M, Eppig JT, Festing MF, Fisher EM. 2000. Genealogies of mouse inbred strains. Nat. Genet. 24:23–25.

Bell JK, Botos I, Hall PR, Askins J, Shiloach J, Segal DM, Davies DR. 2005. The molecular structure of the Toll-like receptor 3 ligand-binding domain. Proc. Natl. Acad. Sci. U. S. A. 102:10976–10980.

Benton MJ, Donoghue PCJ. 2007. Paleontological evidence to date the tree of life. Mol. Biol. Evol. 24:26–53.

Van den Berg TK, Yoder JA, Litman GW. 2004. On the origins of adaptive immunity: innate immune receptors join the tale. Trends Immunol. 25:11–16.

Bergman I-M, Edman K, Ekdahl KN, Rosengren KJ, Edfors I. 2012. Extensive polymorphism in the porcine Toll-like receptor 10 gene. Int. J. Immunogenet. 39:68–76.

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.

Beutler BA. 2009. TLRs and innate immunity. Blood 113:1399–1407.

Blasdell K, Cosson JF, Chaval Y, et al. 2011. Rodent-borne hantaviruses in Cambodia, Lao PDR, and Thailand. Ecohealth 8:432–443.

Bochud P-Y, Bochud M, Telenti A, Calandra T. 2007. Innate immunogenetics: a tool for exploring new frontiers of host defence. Lancet Infect. Dis. 7:531–542.

Boldt AB, Messias-Reason IJ, Lell B, Issifou S, Pedroso ML, Kremsner PG, Kun JF. 2009. Haplotype specific-sequencing reveals MBL2 association with asymptomatic Plasmodium falciparum infection. Malar. J. 8:97.

Bonhomme F, Orth A, Cucchi T, Rajabi-Maham H, Catalan J, Boursot P, Auffray J-C, Britton-Davidian J. 2011. Genetic differentiation of the house mouse around the Mediterranean basin: matrilineal footprints of early and late colonization. Proc. R. Soc. B Biol. Sci. 278:1034–1043.

Boraschi D, Meltzer MS. 1979a. Macrophage activation for tumor cytotoxicity: genetic variation in macrophage tumoricidal capacity among mouse strains. Cell. Immunol. 45:188–194.

Boraschi D, Meltzer MS. 1979b. Defective tumoricidal capacity of macrophages from A/J mice. II. Comparison of the macrophage cytotoxic defect of A/J mice with that of lipid A-unresponsive C3H/HeJ mice. J. Immunol. Baltim. Md 1950 122:1592–1597.

Boraschi D, Meltzer MS. 1979c. Defective tumoricidal capacity of macrophages from A/J mice. I. Characterization of the macrophage cytotoxic defect after in vivo and in vitro activation stimuli. J. Immunol. Baltim. Md 1950 122:1587–1591.

Borghesi L, Milcarek C. 2007. Innate versus Adaptive Immunity: A paradigm past its prime? Cancer Res. 67:3989–3993.

Bosch TCG. 2013. Cnidarian-Microbe interactions and the origin of innate immunity in Metazoans. Annu. Rev. Microbiol. 67:499–518.

Botos I, Segal DM, Davies DR. 2011. The structural biology of Toll-like receptors. Struct. Lond. Engl. 1993 19:447–459.

Boursot P, Auffray JC, Britton-Davidian J, Bonhomme F. 1993. The Evolution of House Mice. Annu. Rev. Ecol. Syst. 24:119–152.

Boursot P, Din, W, Anand R, Darviche D, Dod B, Von Deimling F *et al.* 1996. Origin and radiation of the house mouse: mitochondrial DNA phylogeny. J. Evol. Biol. 9: 391–415.

Bowie AG, Haga IR. 2005. The role of Toll-like receptors in the host response to viruses. Mol. Immunol. 42:859–867.

Bowie AG, Unterholzner L. 2008. Viral evasion and subversion of pattern-recognition receptor signalling. Nat. Rev. Immunol. 8:911–922.

Božíková E, Munclinger P, Teeter KC, Tucker PK, MacholáN M, Piálek J. 2005. Mitochondrial DNA in the hybrid zone between *Mus musculus musculus* and *Mus musculus domesticus*: a comparison of two transects: mtDNA in the house mouse hybrid zone. Biol. J. Linn. Soc. 84:363–378.

Brand S, Staudinger T, Schnitzler F, et al. 2005. The role of Toll-like receptor 4 Asp299Gly and Thr399Ile polymorphisms and CARD15/NOD2 mutations in the susceptibility and phenotype of Crohn's disease. Inflamm. Bowel Dis. 11:645–652.

Bryja J, Galan M, Charbonnel N, Cosson JF. 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.

Burke DF, Worth CL, Priego E-M, Cheng T, Smink LJ, Todd JA, Blundell TL. 2007. Genome bioinformatic analysis of nonsynonymous SNPs. BMC Bioinformatics 8:301.

Carius HJ, Little TJ, Ebert D. 2001. Genetic variation in a host-parasite association: potential for coevolution and frequency-dependent selection. Evol. Int. J. Org. Evol. 55:1136–1145.

Carlton JM. 2003. Genome sequencing and comparative genomics of tropical disease pathogens. Cell. Microbiol. 5:861–873.

Carneiro LAM, Magalhaes JG, Tattoli I, Philpott DJ, Travassos LH. 2008. Nod-like proteins in inflammation and disease. J. Pathol. 214:136–148.

Casanova J-L, Abel L, Quintana-Murci L. 2011. Human TLRs and IL-1Rs in host defense: natural insights from evolutionary, epidemiological, and clinical genetics. Annu. Rev. Immunol. 29:447–491.

Charlesworth B, Morgan MT, Charlesworth D. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics 134:1289–1303.

Chen J-M, Cooper DN, Chuzhanova N, Férec C, Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. Nat. Rev. Genet. 8:762–775.

Chen JS-C, Wang T-Y, Tzeng T-D, Wang C-Y, Wang D. 2008. Evidence for positive selection in the TLR9 gene of teleosts. Fish Shellfish Immunol. 24:234–242.

Chevret P, Veyrunes F, Britton-Davidian J. 2005. Molecular phylogeny of the genus Mus (Rodentia: Murinae) based on mitochondrial and nuclear data. Biol. J. Linn. Soc. 84:417–427.

Choudhuri K, Kearney A, Bakker TR, van der Merwe PA. 2005. Immunology: How Do T Cells Recognize Antigen? Curr. Biol. 15:R382–R385.

Cizkova D, de Bellocq JG, Baird SJE, Pialek J, Bryja J. 2011. Genetic structure and contrasting selection pattern at two major histocompatibility complex genes in wild house mouse populations. Heredity 106:727–740.

Cížková D, Gouy de Bellocq J, Baird SJE, Piálek J, Bryja J. 2011. Genetic structure and contrasting selection pattern at two major histocompatibility complex genes in wild house mouse populations. Heredity (Edinb) 106:727–740.

Clayton DH, Moore J. 1997. Host-parasite evolution: General principles and avian models. Oxford Univ. Press, England. 473 pp. 20 chapters by 27 authors.

Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R. 2010. Jane: a new tool for the cophylogeny reconstruction problem. Algorithms Mol. Biol. AMB 5:16.

Criscitiello MF, de Figueiredo P. 2013. Fifty shades of immune defense. PLoS Pathog 9:e1003110.

Cruaud A, Rønsted N, Chantarasuwan B, et al. 2012. An extreme case of plant-insect codiversification: figs and fig-pollinating wasps. Syst. Biol. 61:1029–1047.

Cucchi T, Vigne J-D, Auffray J-C. 2005. First occurrence of the house mouse (*Mus musculus domesticus* SCHWARTZ & SCHWARTZ, 1943) in Western Mediterranean: a revision of sub-fossil house mice occurrences using a zooarchaeological critical grid. Biol. J. Linn. Soc. 84:429–445.

Cucchi T, Auffray JC, Vigne V. 2012. History of house mouse synanthropyand dispersal in the Near East and Europe: a zooarchaeological insight. Available from: http://www.academia.edu/1949058/Neighbourhood_The_House_Mouse_as_a_Model_in_Evolutionary

Cucchi T, Auffray JC, Vigne JD. 2012. On the origin of the house mouse synanthropy and dispersal in the Near East and Europe: zooarchaeological review and perspectives. In: Macholán M, Baird SJE, Munclinger P and Piálek J, eds. Evolution of the House Mouse. Cambridge: Cambridge University Press: 65–93.

Cucchi T, Kovács ZE, Berthon R, et al. 2013. On the trail of Neolithic mice and men towards Transcaucasia: zooarchaeological clues from Nakhchivan (Azerbaijan). Biol. J. Linn. Soc. 108:917–928.

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

Day DF, Marrceau-Day ML. 1982. Lipopolysaccharide Variability in *Pseudomonas aeruginosa*. Curr. Microbiol. 7:93–98.

Delport W, Poon AFY, Frost SDW, Pond SLk. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. Bioinforma. Oxf. Engl. 26:2455–2457.

Dembic Z. 2005. The Function of Toll-like receptors. In: Toll and Toll-like receptors: An immunologic perspective. Molecular Biology Intelligence Unit. Springer US. p. 18–55. Available from: http://link.springer.com/chapter/10.1007/0-387-27445-6_2

Diebold SS, Kaisho T, Hemmi H, Akira S, Reis e Sousa C. 2004. Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA. Science 303:1529–1531.

Doherty PC, Zinkernagel RM. 1975. Enhanced immunological surveillance in mice heterozygous at the H-2 gene complex. Nature 256:50–52.

Downing T, Lloyd Andrew T, O'Farrelly C, Bradley DG. 2010. The differential evolutionary dynamics of avian cytokine and TLR gene classes. J. Immunol. Baltim. Md 1950 184:6993–7000.

Dufková P, Macholán M, Piálek J. 2011. Inference of selection and stochastic effects in the house mouse hybrid zone. Evolution 65:993–1010.

Ďureje L, Macholán M, Baird SJE, Piálek J. 2012. The mouse hybrid zone in Central Europe: from morphology to molecules. Folia Zool. 61:308–318.

Duret L. 2008. Neutral theory: The null hypothesis of molecular evolution. Nature Education 1(1)

Duvaux L, Belkhir K, Boulesteix M, Boursot P. 2011. Isolation and gene flow: inferring the speciation history of European house mice. Mol. Ecol. 20:5248–5264.

Edwards SV. 2009. Natural selection and phylogenetic analysis. Proc. Natl. Acad. Sci. 106:8799–8800.

Eizaguirre C, Lenz TL, Traulsen A, Milinski M. 2009. Speciation accelerated and stabilized by pleiotropic major histocompatibility complex immunogenes. Ecol. Lett. 12:5–12.

Ewald PW. 1994. Evolution of infectious disease. Oxford [England]; New York: Oxford University Press

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

Eyre-Walker A, Keightley PD. 2007. The distribution of fitness effects of new mutations. Nat. Rev. Genet. 8:610–618.

Ferrer-Admetlla A, Bosch E, Sikora M, et al. 2008. balancing selection is the main force shaping the evolution of innate immunity genes. J. Immunol. 181:1315–1322.

Ferwerda B, McCall MBB, Alonso S, et al. 2007. TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans. Proc. Natl. Acad. Sci. U. S. A. 104:16645–16650.

Fisher CA, Bhattarai EK, Osterstock JB, et al. 2011. Evolution of the bovine tlr gene family and member associations with mycobacterium avium subspecies paratuberculosis infection. PLoS ONE 6:e27744.

Flajnik MF, Kasahara M. 2010. Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat. Rev. Genet. 11:47–59.

Fornarino S, Laval G, Barreiro LB, Manry J, Vasseur E, Quintana-Murci L. 2011. Evolution of the TIR domain-containing adaptors in humans: swinging between constraint and adaptation. Mol. Biol. Evol. 28:3087–3097.

Fornuskova A, Vinkler M, Pagès M, et al. 2013. Contrasted evolutionary histories of two Toll-like receptors (Tlr4 and Tlr7) in wild rodents (Murinae). BMC Evol. Biol.:13:194.

Fortune SM, Solache A, Jaeger A, Hill PJ, Belisle JT, Bloom BR, Rubin EJ, Ernst JD. 2004. Mycobacterium tuberculosis inhibits macrophage responses to IFN-gamma through myeloid differentiation factor 88-dependent and -independent mechanisms. J. Immunol. Baltim. Md 1950 172:6272–6280.

Fox JG, Beck P, Dangler CA, Whary MT, Wang TC, Shi HN, Nagler-Anderson C. 2000. Concurrent enteric helminth infection modulates inflammation and gastric immune responses and reduces helicobacter-induced gastric atrophy. Nat. Med. 6:536–542.

Franchi L, Amer A, Body-Malapel M, et al. 2006. Cytosolic flagellin requires Ipaf for activation of caspase-1 and interleukin 1beta in salmonella-infected macrophages. Nat. Immunol. 7:576–582.

Frank SA. 1993. Specificity versus detectable polymorphism in host-parasite genetics. Proc. Biol. Sci. 254:191–197.

Fraser IP, Koziel H, Ezekowitz RA. 1998. The serum mannose-binding protein and the macrophage mannose receptor are pattern recognition molecules that link innate and adaptive immunity. Semin. Immunol. 10:363–372.

Freeman S, Herron JC. 2007. Evolutionary Analysis. Pearson Prentice Hall

Fumagalli M, Sironi M, Pozzoli U, Ferrer-Admettla A, Pattini L, Nielsen R. 2011. Signatures of Environmental Genetic Adaptation Pinpoint Pathogens as the Main Selective Pressure through Human Evolution. PLoS Genet. [Internet] 7. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3207877/

Futuyma DJ. 2005. Evolution. Sinauer Associates Incorporated

Galan M, Pagès M, Cosson J-F. 2012. Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples. PloS One 7:e48374.

Garrigan D, Hedrick PW. 2003. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evol. Int. J. Org. Evol. 57:1707–1722.

Gay NJ, Gangloff M, Weber ANR. 2006. Toll-like receptors as molecular switches. Nat. Rev. Immunol. 6:693–698.

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

Georgel P, Jiang Z, Kunz S, Janssen E, Mols J, Hoebe K, Bahram S, Oldstone MBA, Beutler B. 2007. Vesicular stomatitis virus glycoprotein G activates a specific antiviral Toll-like receptor 4-dependent pathway. Virology 362:304–313.

Geraldes A, Basset P, Gibson B, Smith KL, Harr B, Yu H-T, Bulatova N, Ziv Y, Nachman MW. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. Mol. Ecol. 17:5349–5363.

Getz GS. 2005. Bridging the innate and adaptive immune systems. J. Lipid Res. 46:619–622.

Gordon S. 2004. Pathogen recognition or homeostasis? APC receptor functions in innate immunity. C. R. Biol. 327:603–607.

Gough PJ, Gordon S. 2000. The role of scavenger receptors in the innate immune system. Microbes Infect. Inst. Pasteur 2:305–311.

Govindaraj RG, Manavalan B, Basith S, Choi S. 2011. Comparative analysis of species-specific ligand recognition in Toll-like receptor 8 signaling: a hypothesis. PloS One 6:e25118.

Govindaraj RG, Manavalan B, Lee G, Choi S. 2010. Molecular modeling-based evaluation of hTLR10 and identification of potential ligands in Toll-like receptor signaling. PLoS ONE 5:e12713.

Grueber CE, Wallis GP, King TM, Jamieson IG. 2012. Variation at Innate Immunity Toll-Like Receptor Genes in a Bottlenecked Population of a New Zealand Robin. PLoS ONE 7:e45011.

Grueber CE, Wallis GP, Jamieson IG. 2013. Genetic drift outweighs natural selection at Toll-like receptor (TLR) immunity loci in a re-introduced population of a threatened species. Mol. Ecol. 22:4470–4482.

Guénet JL. 1998. Wild mice as a source of genetic polymorphism. Pathol. Biol. (Paris) 46:685–688.

Guénet JL, Bonhomme F. 2003. Wild mice: an ever-increasing contribution to a popular mammalian model. Trends Genet. TIG 19:24–31.

Häcker H, Mischak H, Miethke T, Liptay S, Schmid R, Sparwasser T, Heeg K, Lipford GB, Wagner H. 1998. CpG-DNA-specific activation of antigen-presenting cells requires stress kinase activity and is preceded by non-specific endocytosis and endosomal maturation. EMBO J. 17:6230–6240.

Hajjar AM, Harvey MD, Shaffer SA, et al. 2006. Lack of in vitro and in vivo recognition of *Francisella tularensis* subspecies lipopolysaccharide by Toll-like receptors. Infect. Immun. 74:6730–6738.

Haldane JBS. 2006. Disease and Evolution. In: Malaria: Genetic and Evolutionary Aspects. Boston: Kluwer Academic Publishers. p. 175–187. Available from: http://link.springer.com/content/pdf/10.1007/0-387-28295-5_9.pdf

Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95–98.

Hartl DL, Clark AG. 1997. Principles of population genetics. Sinauer Associates, Incorporated

Hasan U, Chaffois C, Gaillard C, et al. 2005. Human TLR10 is a functional receptor, expressed by B cells and plasmacytoid dendritic cells, which activates gene transcription through MyD88. J. Immunol. Baltim. Md 1950 174:2942–2950.

Hasegawa M, Kishino H, Yano T. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Hawn TR, Verbon A, Janer M, Zhao LP, Beutler B, Aderem A. 2005. Toll-like receptor 4 polymorphisms are associated with resistance to Legionnaires' disease. Proc. Natl. Acad. Sci. U. S. A. 102:2487–2489.

Hawn TR, Verbon A, Lettinga KD, et al. 2003. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. J. Exp. Med. 198:1563–1572.

Hayashi F, Smith KD, Ozinsky A, et al. 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature 410:1099–1103.

He Y, Li J, Jiang S. 2006. A single amino acid substitution (R441A) in the receptor-binding domain of SARS coronavirus spike protein disrupts the antigenic structure and binding activity. Biochem. Biophys. Res. Commun. 344:106–113.

Hedrick PW. 2002. Pathogen resistance and genetic variation at MHC loci. Evol. Int. J. Org. Evol. 56:1902–1908.

Hedrick SM. 2004. The acquired immune system: a vantage from beneath. Immunity 21:607–615.

Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C, Akira S, Lipford G, Wagner H, Bauer S. 2004. Species-specific recognition of single-stranded RNA via Toll-like receptor 7 and 8. Science 303:1526–1529.

Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T. 2011. Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection. Genome Res. 21:349–356.

Holm L, Kääriäinen S, Rosenström P, Schenkel A. 2008. Searching protein structure databases with DaliLite v.3. Bioinforma. Oxf. Engl. 24:2780–2781.

Honnay O. 2013. Genetic drift. In: Maloy S, Hughes K, editors. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press. p. 251–253. Available from: http://www.sciencedirect.com/science/article/pii/B9780123749840006161

Huang S, Yuan S, Guo L, et al. 2008. Genomic analysis of the immune gene repertoire of amphioxus reveals extraordinary innate complexity and diversity. Genome Res. 18:1112–1126.

Huang Y, Temperley ND, Ren L, Smith J, Li N, Burt DW. 2011. Molecular evolution of the vertebrate TLR1 gene family-a complex history of gene duplication, gene conversion, positive selection and co-evolution. BMC Evol. Biol. 11:149.

Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinforma. Oxf. Engl. 17:754–755.

Hughes AL, Piontkivska H. 2008. Functional diversification of the toll-like receptor gene family. Immunogenetics 60:249–256.

Hugot JP, Plyusnina A, Herbreteau V, Nemirov K, Laakkonen J, Lundkvist Å, Supputamongkol Y, Henttonen H, Plyusnin A. 2006. Genetic analysis of Thailand hantavirus in *Bandicota indica* trapped in Thailand. Virol. J. 3:72.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11:97–108.

Jameson SC, Bevan MJ. 1998. T-cell selection. Curr. Opin. Immunol. 10:214–219.

Janeway CA Jr, Medzhitov R. 2002. Innate immune recognition. Annu. Rev. Immunol. 20:197–216.

Janeway CA Jr. 1989. Approaching the asymptote? Evolution and revolution in immunology. Cold Spring Harb. Symp. Quant. Biol. 54 Pt 1:1–13.

Jann OC, Werling D, Chang J-S, Haig D, Glass EJ. 2008. Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. BMC Evol. Biol. 8:288.

Janoušek V, Wang L, Luzynski K, et al. 2012. Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between Mus musculus musculus and M. m. domesticus. Mol. Ecol. 21:3032–3047.

Jansa SA, Weksler M. 2004. Phylogeny of muroid rodents: relationships within and among major lineages as determined by IRBP gene sequences. Mol. Phylogenet. Evol. 31:256–276.

Janssens S, Beyaert R. 2003. Role of Toll-like receptors in pathogen recognition. Clin. Microbiol. Rev. 16:637–646.

Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik S-G, Lee H, Lee J-O. 2007. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. Cell 130:1071–1082.

Jones EP, Eager HM, Gabriel SI, Jóhannesdóttir F, Searle JB. 2013. Genetic tracking of mice and other bioproxies to infer human history. Trends Genet. 29:298–308.

Jones EP, Jóhannesdóttir F, Gündüz İ, Richards MB, Searle JB. 2011. The expansion of the house mouse into north-western Europe. J. Zool. 283:257–268.

Jones EP, Skirnisson K, McGovern TH, Gilbert MTP, Willerslev E, Searle JB. 2012. Fellow travellers: a concordance of colonization patterns between mice and men in the North Atlantic region. BMC Evol. Biol. 12:35.

Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008. Global trends in emerging infectious diseases. Nature 451:990–993.

De Jong SD, Basha G, Wilson KD, Kazem M, Cullis P, Jefferies W, Tam Y. 2010. The immunostimulatory activity of unmethylated and methylated CpG oligodeoxynucleotide is dependent on their ability to colocalize with TLR9 in late endosomes. J. Immunol. Baltim. Md 1950 184:6092–6102.

Jungi TW, Farhat K, Burgener IA, Werling D. 2011. Toll-like receptors in domestic animals. Cell Tissue Res. 343:107–120.

Kalinowski ST. 2009. How well do evolutionary trees describe genetic relationships among populations? Heredity (Edinb) 102:506–513.

Kanaan Z, Ahmad S, Roberts H, Thé T, Girdler S, Pan J, Rai SN, Weller EB Jr, Galandiuk S. 2012. Crohn's disease in Caucasians and African Americans, as defined by clinical predictors and single nucleotide polymorphisms. J. Natl. Med. Assoc. 104:420–427.

Kang JY, Lee JO. 2011. Structural biology of the Toll-like receptor family. Annu. Rev. Biochem. 80:917–941.

Kang JY, Nan X, Jin MS, et al. 2009. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity 31:873–884.

Kang TJ, Chae GT. 2001. Detection of Toll-like receptor 2 (TLR2) mutation in the lepromatous leprosy patients. FEMS Immunol. Med. Microbiol. 31:53–58.

Karvonen A, Seehausen O. 2012. The role of parasitism in adaptive radiations; When might parasites promote and when might they constrain ecological speciation? Int. J. Ecol. 2012. Available from: http://www.hindawi.com/journals/ijecol/2012/280169/abs/

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

Keestra AM, van Putten JPM. 2008. Unique properties of the chicken TLR4/MD-2 complex: selective lipopolysaccharide activation of the MyD88-dependent pathway. J. Immunol. Baltim. Md 1950 181:4354–4362.

Kelley LA, Sternberg MJE. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. 4:363–371.

Key KHL. 1968. The concept of stasipatric speciation. Syst. Zool. 17:14–22.

Kim HM, Park BS, Kim JI, et al. 2007. Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran. Cell 130:906–917.

Kimbrell DA, Beutler B. 2001. The evolution and genetics of innate immunity. Nat. Rev. Genet. 2:256–267.

Kimura M. 1985. The neutral theory of molecular evolution, New Scientist, pp41-46.

Kimura M. 1991. The neutral theory of molecular evolution: a review of recent evidence. Idengaku Zasshi 66:367–386.

Knirel YA, Dentovskaya SV, Senchenkova SN, Shaikhutdinova RZ, Kocharova NA, Anisimov AP. 2006. Structural features and structural variability of the lipopolysaccharide of *Yersinia pestis*, the cause of plague. J. Endotoxin Res. 12:3–9.

Konečný A, Estoup A, Duplantier JM, Bryja J, Bâ K, Galan M, Tatard C, Cosson JF. 2013. Invasion genetics of the introduced black rat (*Rattus rattus*) in Senegal, West Africa. Mol. Ecol. 22:286–300.

Kopp E, Medzhitov R. 2003. Recognition of microbial infection by Toll-like receptors. Curr. Opin. Immunol. 15:396–401.

Korber B. 2000. HIV signature and sequence variation analysis. Computational analysis of HIV molecular sequences, Chapter 4, pages 55-72. Allen G. Rodrigo and Gerald H. Learn, eds. Dordrecht, Netherlands: Kluwer Academic Publishers.

Kositanont U, Naigowit P, Imvithaya A, Singchai C, Puthavathana P. 2003. Prevalence of antibodies to Leptospira serovars in rodents and shrews trapped in low and high endemic areas in Thailand. J. Med. Assoc. Thail. Chotmaihet Thangphaet 86:136–142.

Krieg AM. 2000. The role of CpG motifs in innate immunity. Curr. Opin. Immunol. 12:35–43.

Krieg AM, Vollmer J. 2007. Toll-like receptors 7, 8, and 9: linking innate immunity to autoimmunity. Immunol. Rev. 220:251–269.

Kruithof EK, Satta N, Liu JW, Dunoyer-Geindre S, Fish RJ. 2007. Gene conversion limits divergence of mammalian TLR1 and TLR6. BMC Evol. Biol. 7:148.

Kumar S, Skjaeveland A, Orr RJS, Enger P, Ruden T, Mevik B-H, Burki F, Botnen A, Shalchian-Tabrizi K. 2009. AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses. BMC Bioinformatics 10:357.

Kurt-Jones EA, Popova L, Kwinn L, et al. 2000. Pattern recognition receptors TLR4 and CD14 mediate response to respiratory syncytial virus. Nat. Immunol. 1:398–401.

Kváč M, McEvoy J, Loudová M, et al. 2013. Coevolution of *Cryptosporidium tyzzeri* and the house mouse (*Mus musculus*). Int. J. Parasitol. 43:805–817.

Lack JB, Greene DU, Conroy CJ, Hamilton MJ, Braun JK, Mares MA, Van Den Bussche RA. 2012. Invasion facilitates hybridization with introgression in the *Rattus rattus* species complex. Mol. Ecol. 21:3545–3561.

Lai J, Bernhard OK, Turville SG, Harman AN, Wilkinson J, Cunningham AL. 2009. Oligomerization of the Macrophage Mannose Receptor Enhances gp120-mediated Binding of HIV-1. J. Biol. Chem. 284:11027–11038.

Lecompte E, Granjon L, Peterhans JK, Denys C. 2002. Cytochrome b-based phylogeny of the Praomys group (Rodentia, Murinae): a new African radiation? C. R. Biol. 325:827–840.

Lecompte E, Aplin K, Denys C, Catzeflis F, Chades M, Chevret P. 2008. Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily. BMC Evol. Biol. 8:199.

Lederberg J. 1999. J. B. S. Haldane (1949) on infectious disease and evolution. Genetics 153:1–3.

Lemaitre B, Nicolas E, Michaut L, Reichhart J-M, Hoffmann JA. 1996. The dorsoventral regulatory gene cassette spätzle/Toll/cactus controls the potent antifungal response in Drosophila adults. Cell 86:973–983.

Leonard JN, Ghirlando R, Askins J, Bell JK, Margulies DH, Davies DR, Segal DM. 2008. The TLR3 signaling complex forms by cooperative receptor dimerization. Proc. Natl. Acad. Sci. 105:258–263.

Letunic I, Doerks T, Bork P. 2011. SMART 7: recent updates to the protein domain annotation resource. Nucleic Acids Res. 40:D302–D305.

Leulier F, Lemaitre B. 2008. Toll-like receptors-taking an evolutionary approach. Nat. Rev. Genet. 9:165–178.

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

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

Linnenbrink M, Wang J, Hardouin EA, Künzel S, Metzler D, Baines JF. 2013. The role of biogeography in shaping diversity of the intestinal microbiota in house mice. Mol. Ecol. 22:1904–1916.

Loo YM, Gale Jr. M. 2011. Immune signaling by RIG-I-like receptors. Immunity 34:680–692.

Lorenz E, Mira JP, Cornish KL, Arbour NC, Schwartz DA. 2000. A Novel polymorphism in the Toll-like receptor 2 gene and its potential association with staphylococcal infection. Infect. Immun. 68:6398–6401.

Luis AD, Hayman DTS, O'Shea TJ, et al. 2013. A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special? Proc. Biol. Sci. 280:20122753.

Lundrigan BL, Jansa SA, Tucker PK. 2002. Phylogenetic relationships in the genus mus, based on paternally, maternally, and biparentally inherited characters. Syst. Biol. 51:410–431.

Lynch M. 2006. The origins of eukaryotic gene structure. Mol. Biol. Evol. 23:450–468.

Lynch M. 2007. The frailty of adaptive hypotheses for the origins of organismal complexity, PNAS May 15, 2007 vol. 104 free full text: "Most biologists are so convinced that all aspects of biodiversity arise from adaptive processes that virtually no attention is given to the null hypothesis of neutral evolution".

Lysenko AJ. 1971. Distribution of leishmaniasis in the Old World. Bull. World Health Organ. 44:515–520.

MacDonald AS, Araujo MI, Pearce EJ. 2002. Immunology of parasitic helminth infections. Infect. Immun. 70:427–433.

Maeshima N, Fernandez RC. 2013. Recognition of lipid A variants by the TLR4-MD-2 receptor complex. Front. Cell. Infect. Microbiol. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3569842/

Macholán M, Vyskocilová M, Bonhomme F, Krystufek B, Orth A, Vohralík V. 2007. Genetic variation and phylogeography of free-living mouse species (genus Mus) in the Balkans and the Middle East. Mol. Ecol. 16:4774–4788.

Macholán M. 2008. The mouse skull as a source of morphometric data for phylogeny inference. Zool. Anz. - J. Comp. Zool. 247:315–327.

Macholán M, Baird SJE, Munclinger P, Dufková P, Bímová B, Piálek J. 2008. Genetic conflict outweighs heterogametic incompatibility in the mouse hybrid zone? BMC Evol. Biol. 8:271.

Macholán M, Baird SJE, Dufková P, Munclinger P, Bímová BV, Piálek J. 2011. Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe. Evol. Int. J. Org. Evol. 65:1428–1446.

Macholán M, Pialek J, Baird SJE, Munclinger P. 2012. Evolution of the House Mouse. 1st ed. Cambridge University Press Available from: http://www.uread.com/book/evolution-house-mouse-milos-macholan/9780521760669

Macholán M, Mrkvicová Vyskočilová M, Bejček V, Šťastný K. 2012b: Mitochondrial DNA sequence variation and evolution of Old World house mice (*Mus musculus*). Folia Zool., 61(3-4), 284-307.

Macholán M. 2013. Hybrid Zone, Mouse. In: Maloy S, Hughes K, editors. Brenner's Encyclopedia of Genetics (Second Edition). San Diego: Academic Press. p. 588–591. Available from: http://www.sciencedirect.com/science/article/pii/B9780123749840007555

Maizels RM, Balic A, Gomez-Escobar N, Nair M, Taylor MD, Allen JE. 2004. Helminth parasites--masters of regulation. Immunol. Rev. 201:89–116.

Malik SC. 2011. Genomic Analyses of Toll-like Receptor 4 and 7 Exons of Bos indicus from Temperate Sub-himalayan Region of India. Asian Australas. J. Anim. Sci. 24:1019 – 1025.

Manavalan B, Basith S, Choi S. 2011. Similar structures but different roles - an updated perspective on TLR structures. Front. Physiol. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3146039/

Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, Kuroki Y. 2007. Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate Toll-like receptors. BMC Genomics 8:124.

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. Nature 351:652–654.

McTaggart SJ, Obbard DJ, Conlon C, Little TJ. 2012. Immune genes undergo more adaptive evolution than non-immune system genes in Daphnia pulex. BMC Evol. Biol. 12:63.

Medzhitov R, Janeway CA. 1999. Innate Immune Induction of the Adaptive Immune Response. Cold Spring Harb. Symp. Quant. Biol. 64:429–436.

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

Medzhitov R, Janeway CA Jr. 2002. Decoding the patterns of self and nonself by the innate immune system. Science 296:298–300.

Medzhitov R. 2007. Recognition of microorganisms and activation of the immune response. Nature 449:819–826.

Meerburg BG, Singleton GR, Kijlstra A. 2009. Rodent-borne diseases and their risks for public health. Crit. Rev. Microbiol. 35:221–270.

Mempel M, Kalali BN, Ollert M, Ring J. 2007. Toll-like receptors in dermatology. Dermatol. Clin. 25:531–540.

Meyer TF. 1991. Evasion mechanisms of pathogenic Neisseriae. Behring Inst. Mitt.:194–199.

Michallet MC, Rota G, Maslowski K, Guarda G. 2013. Innate receptors for adaptive immunity. Curr. Opin. Microbiol. 16:296–302.

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

Milinski M. 2006. The major histocompatibility complex, sexual selection, and mate choice. Annu. Rev. Ecol. Evol. Syst. 37:159–186.

Mills JN. 2006. Biodiversity loss and emerging infectious disease: An example from the rodent-borne hemorrhagic fevers. Biodiversity 7:9–17.

Mizel SB, West AP, Hantgan RR. 2003. Identification of a sequence in human Toll-like receptor 5 required for the binding of Gram-negative flagellin. J. Biol. Chem. 278:23624–23629.

Mockenhaupt FP, Hamann L, von Gaertner C, Bedu-Addo G, von Kleinsorgen C, Schumann RR, Bienzle U. 2006. Common polymorphisms of Toll-like receptors 4 and 9 are associated with the clinical manifestation of malaria during pregnancy. J. Infect. Dis. 194:184–188.

Mogensen TH. 2009. Pathogen recognition and inflammatory signaling in innate immune defenses. Clin. Microbiol. Rev. 22:240–273.

Montminy SW, Khan N, McGrath S, et al. 2006. Virulence factors of *Yersinia pestis* are overcome by a strong lipopolysaccharide response. Nat. Immunol. 7:1066–1073.

Moore WS. 1995. Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726.

Moulia C, Aussel JP, Bonhomme F, Boursot P, Nielsen JT, Renaud F. 1991. Wormy mice in a hybrid zone: A genetic control of susceptibility to parasite infection. J. Evol. Biol. 4:679–687.

Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP. 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. Proc. Natl. Acad. Sci. 106:7073–7078.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Pond SLK. 2012. Detecting Individual Sites Subject to Episodic Diversifying Selection. PLoS Genet 8:e1002764.

Musser GG, Carleton MD. 2005. Superfamily Muroidea, p. 894-1531. In Wilson, D.E. and Reeder, D. (eds.) *Mammal Species of the World A Taxonomic and Geographic Reference, 3rd edition*, 2 vol. Johns Hopkins University Press, Baltimore, Maryland.

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

Nei M, Tajima F. 1981. Genetic drift and estimation of effective population size. Genetics 98:625–640.

Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418–426.

Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. Annu. Rev. Genet. 39:121–152.

Nei M, Nozawa M. 2011. Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era. Genome Biol. Evol. 3:812–829.

Netea MG, Ferwerda G, de Jong DJ, et al. 2005. Nucleotide-binding oligomerization domain-2 modulates specific TLR pathways for the induction of cytokine release. J. Immunol. Baltim. Md 1950 174:6518–6523.

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

Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. Annu. Rev. Genomics Hum. Genet. 7:61–80.

Nielsen R, Bustamante C, Clark AG, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3:e170.

Nichols R. 2001. Gene trees and species trees are not the same. Trends Ecol. Evol. 16:358–364.

Nunome M, Ishimori C, Aplin KP, Tsuchiya K, Yonekawa H, Moriwaki K, Suzuki H. 2010. Detection of recombinant haplotypes in wild mice (*Mus musculus*) provides new insights into the origin of Japanese mice. Mol. Ecol. 19:2474–2489.

O'Neill LAJ. 2004. TLRs: Professor Mechnikov, sit on your hat. Trends Immunol. 25:687–693.

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

Ogus AC, Yoldas B, Ozdemir T, Uguz A, Olcen S, Keser I, Coskun M, Cilli A, Yegin O. 2004. The Arg753GLn polymorphism of the human Toll-like receptor 2 gene in tuberculosis disease. Eur. Respir. J. 23:219–223.

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

Okamura Y, Watari M, Jerud ES, Young DW, Ishizaka ST, Rose J, Chow JC, Strauss JF 3rd. 2001. The extra domain A of fibronectin activates Toll-like receptor 4. J. Biol. Chem. 276:10229–10233.

Opitz B, Eitel J, Meixenberger K, Suttorp N. 2009. Role of Toll-like receptors, NOD-like receptors and RIG-I-like receptors in endothelial cells and systemic infections. Thromb. Haemost. 102:1103–1109.

Ortiz M, Kaessmann H, Zhang K, Bashirova A, Carrington M, Quintana-Murci L, Telenti A. 2008. The evolutionary history of the CD209 (DC-SIGN) family in humans and non-human primates. Genes Immun. 9:483–492.

Ota T, Nei M. 1994. Variance and covariances of the numbers of synonymous and nonsynonymous substitutions per site. Mol. Biol. Evol. 11:613–619.

Ozinsky A, Underhill DM, Fontenot JD, Hajjar AM, Smith KD, Wilson CB, Schroeder L, Aderem A. 2000. The repertoire for pattern recognition of pathogens by the innate immune system is defined by cooperation between Toll-like receptors. Proc. Natl. Acad. Sci. 97:13766–13771.

Pagès M, Bazin E, Galan M, et al. 2013. Cytonuclear discordance among Southeast Asian black rats (*Rattus rattus* complex). Mol. Ecol. 22:1019–1034.

Pagès M, Chaval Y, Herbreteau V, Waengsothorn S, Cosson JF, Hugot JP, Morand S, Michaux J. 2010. Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries. BMC Evol. Biol. 10:184.

Palermo S, Capra E, Torremorell M, Dolzan M, Davoli R, Haley C, Giuffra E. 2009. Toll-like receptor 4 genetic diversity among pig populations. Anim. Genet. 40:289–299.

Pandey S, Agrawal DK. 2006. Immunobiology of Toll-like receptors: emerging trends. Immunol. Cell Biol. 84:333–341.

Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191–1195.

Park S, Park D, Jung Y, Chung E, Choi S. 2010. Positive selection signatures in the TLR7 family. GENES GENOMICS 32:143–150.

Parker LC, Prince LR, Sabroe I. 2007. Translational mini-review series on Toll-like receptors: networks regulated by Toll-like receptors mediate innate and adaptive immunity. Clin. Exp. Immunol. 147:199–207.

Pasare C, Medzhitov R. 2003. Toll pathway-dependent blockade of CD4+CD25+ T cell-mediated suppression by dendritic cells. Science 299:1033–1036.

Pasare C, Medzhitov R. 2004a. Toll-like receptors: linking innate and adaptive immunity. Microbes Infect. Inst. Pasteur 6:1382–1387.

Pasare C, Medzhitov R. 2004b. Toll-like receptors and acquired immunity. Semin. Immunol. 16:23–26.

Patnaik R. 2011. Fossil murine rodents as ancient monsoon indicators of the Indian subcontinent. Quat. Int. 229:94–104.

Payseur BA, Krenz JG, Nachman MW. 2004. Differential patterns of introgression across the X chromosome in a hybrid zone between two species of house mice. Evol. Int. J. Org. Evol. 58:2064–2078.

Pearce EJ, Sher A. 1987. Mechanisms of immune evasion in schistosomiasis. Contrib. Microbiol. Immunol. 8:219–232.

Pedersen AB, Babayan SA. 2011. Wild immunology. Mol. Ecol. 20:872–880.

Peiser L, Mukhopadhyay S, Gordon S. 2002. Scavenger receptors in innate immunity. Curr. Opin. Immunol. 14:123–128.

Phifer-Rixey M, Bonhomme F, Boursot P, Churchill GA, Pialek J, Tucker PK, Nachman MW. 2012. Adaptive Evolution and Effective Population Size in Wild House Mice. Mol. Biol. Evol. 29:2949–2955.

Piálek J, Vyskočilová M, Bímová B, et al. 2008. Development of unique house mouse resources suitable for evolutionary studies of speciation. J. Hered. 99:34–44.

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

Plyusnina A, Ibrahim I-N, Plyusnin A. 2009. A newly recognized hantavirus in the Asian house rat (*Rattus tanezumi*) in Indonesia. J. Gen. Virol. 90:205–209.

Polley L, Thompson RCA. 2009. Parasite zoonoses and climate change: molecular tools for tracking shifting boundaries. Trends Parasitol. 25:285–291.

Poltorak AHX, Smirnova I, et al. 1998. Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene. Science 282:2085–2088.

Pond SLK, Frost SDW. 2005a. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinforma. Oxf. Engl. 21: 2531–2533.

Pond SLK, Frost SDW. 2005b. Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. 22:1208–1222.

Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006a. GARD: a genetic algorithm for recombination detection. Bioinformatics 22:3096–3098.

Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006b. Automated phylogenetic detection of recombination using a genetic algorithm. Mol. Biol. Evol. 23:1891–1901.

Pond SLK, Frost SDW, Grossman Z, Gravenor MB, Richman DD, Brown AJL. 2006c. Adaptation to different human populations by HIV-1 revealed by codon-based analyses. PLoS Comput Biol 2:e62.

Posada D. 2008. jModelTest: Phylogenetic model averaging. Mol. Biol. Evol. 25:1253–1256.

Quach H, Wilson D, Laval G, et al. 2013. Different selective pressures shape the evolution of Toll-like receptors in human and African great ape populations. Hum. Mol. Genet.

Quintana-Murci L, Clark AG. 2013. Population genetic tools for dissecting innate immunity in humans. Nat. Rev. Immunol. 13:280–293.

Raetz CRH, Whitfield C. 2002. Lipopolysaccharide endotoxins. Annu. Rev. Biochem. 71:635–700.

Raja A, Vignesh AR, Mary BA, Tirumurugaan KG, Raj GD, Kataria R, Mishra BP, Kumanan K. 2011. Sequence analysis of Toll-like receptor genes 1-10 of goat (*Capra hircus*). Vet. Immunol. Immunopathol. 140:252–258.

Rajabi-Maham H, Orth A, Siahsarvie R, Boursot P, Darvish J, Bonhomme F. 2012. The south-eastern house mouse *Mus musculus castaneus* (Rodentia: Muridae) is a polytypic subspecies. Biol. J. Linn. Soc. 107:295–306.

Rambaut A, Drummond A. 2007. Tracer v1.4, Available from http://beast.bio.ed.ac.uk/Tracer.

Rambaut A. 2009. FigTree v1.3.1 2009, Available with the program package at http://tree.bio.ed.ac.uk/software/figtree/.

Ray A, Redhead K, Selkirk S, Poole S. 1991. Variability in LPS composition, antigenicity and reactogenicity of phase variants of *Bordetella pertussis*. FEMS Microbiol. Lett. 79:211–218.

Rebl A, Goldammer T, Seyfert HM. 2010. Toll-like receptor signaling in bony fish. Vet. Immunol. Immunopathol. 134:139–150.

Reeder JC, Brown GV. 1996. Antigenic variation and immune evasion in Plasmodium falciparum malaria. Immunol. Cell Biol. 74:546–554.

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

Reyburn HT, Mandelboim O, Valés-Gómez M, Davis DM, Pazmany L, Strominger JL. 1997. The class I MHC homologue of human cytomegalovirus inhibits attack by natural killer cells. Nature 386:514–517.

Rezazadeh M, Hajilooi M, Rafiei A, Haidari M, Nikoopour E, Kerammat F, Mamani M, Ranjbar M, Hashemi H. 2006. TLR4 polymorphism in Iranian patients with brucellosis. J. Infect. 53:206–210.

Riley EM, Viney ME. 2011. Wild mice provide insights into natural killer cell maturation and memory. Mol. Ecol. 20:4827–4829.

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

Robinson RT, Khader SA, Locksley RM, Lien E, Smiley ST, Cooper AM. 2008. Yersinia pestis evades TLR4-dependent induction of IL-12(p40)2 by dendritic cells and subsequent cell migration. J. Immunol. Baltim. Md 1950 181:5560–5567.

Rowe KC, Reno ML, Richmond DM, Adkins RM, Steppan SJ. 2008. Pliocene colonization and adaptive radiations in Australia and New Guinea (Sahul): Multilocus systematics of the old endemic rodents (Muroidea: Murinae). Mol. Phylogenet. Evol. 47:84–101.

Roy BA, Widmer A. 1999. Floral mimicry: a fascinating yet poorly understood phenomenon. Trends Plant Sci. 4:325–330.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for biologist programmers. Methods Mol. Biol. Clifton NJ 132:365–386.

Sage RD, Heyneman D, Lim KC, Wilson AC. 1986. Wormy mice in a hybrid zone. Nature 324:60–63.

Sage RD, Prager EM, Tichy H, Wilson AC. 1990. Mitochondrial DNA variation in house mice, Mus domesticus (Rutty). Biol. J. Linn. Soc. 41:105–123.

Salcedo T, Geraldes A, Nachman MW. 2007. Nucleotide variation in wild and inbred mice. Genetics 177:2277–2291.

Schmid-Hempel P. 2011. Evolutionary parasitology: the integrated study of infections, immunology, ecology, and genetics. Oxford [England]; New York: Oxford University Press.

Seabury CM, Seabury PM, Decker JE, Schnabel RD, Taylor JF, Womack JE. 2010. Diversity and evolution of 11 innate immune genes in *Bos taurus taurus* and *Bos taurus indicus* cattle. Proc. Natl. Acad. Sci. U. S. A. 107:151–156.

Sentitula, Kumar R, Yadav BR. 2012. Molecular analysis of TLR4 gene and its association with intra-mammary infections in Sahiwal cattle and Murrah buffaloes. Indian J Biotechnol 11(3), 267-273.

Sheldon B. 1998. Host-parasite evolution: general principles and avian models. Parasitol. Today Pers. Ed 14:84.

Shen N, Fu Q, Deng Y, et al. 2010. Sex-specific association of X-linked Toll-like receptor 7 (TLR7) with male systemic lupus erythematosus. Proc. Natl. Acad. Sci. U. S. A. 107:15838–15843.

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

Shimodaira H, Hasegawa M. 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16:1114.

Shinkai H, Arakawa A, Tanaka-Matsuda M, Ide-Okumura H, Terada K, Chikyu M, Kawarasaki T, Ando A, Uenishi H. 2012. Genetic variability in swine leukocyte antigen class II and Toll-like receptors affects immune responses to vaccination for bacterial infections in pigs. Comp. Immunol. Microbiol. Infect. Dis. 35:523–532.

Schenten D, Medzhitov R. 2011. The control of adaptive immune responses by the innate immune system. In: Advances in Immunology. Vol. 109. Elsevier. p. 87–124. Available from: http://linkinghub.elsevier.com/retrieve/pii/B9780123876645000030

Schönrich G, Rang A, Lütteke N, Raftery MJ, Charbonnel N, Ulrich RG. 2008. Hantavirus-induced immunity in rodent reservoirs and humans. Immunol. Rev. 225:163–189.

Schröder NWJ, Schumann RR. 2005. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. Lancet Infect. Dis. 5:156–164.

Schultz J, Milpetz F, Bork P, Ponting CP. 1998. SMART, a simple modular architecture research tool: identification of signaling domains. Proc. Natl. Acad. Sci. U. S. A. 95:5857–5864.

Schwensow N, Axtner J, Sommer S. 2011. Are associations of immune gene expression, body condition and parasite burden detectable in nature? A case study in an endemic rodent from the Brazilian Atlantic Forest. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 11:23–30.

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

Smith C, Ondračková M, Spence R, Adams S, Betts DS, Mallon E. 2011. Pathogen-mediated selection for MHC variability in wild zebrafish. Evol. Ecol. Res. Vol.67:217–218.

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

Spurgin LG, Richardson DS. 2010. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc. Biol. Sci. 277:979–988.

Stamatakis A, Hoover P, Rougemont J. 2008. A rapid bootstrap algorithm for the RAxML web servers. Syst. Biol. 57:758–771.

Stephens M, Donnelly P. 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73:1162–1169.

Stephan K, Smirnova I, Jacque B, Poltorak A. 2007. Genetic analysis of the innate immune responses in wild-derived inbred strains of mice. Eur. J. Immunol. 37:212–223.

Steppan SJ, Adkins RM, Anderson J. 2004. Phylogeny and divergence-date estimates of rapid radiations in muroid rodents based on multiple nuclear genes. Syst. Biol. 53:533–553.

Steppan SJ, Adkins RM, Spinks PQ, Hale C. 2005. Multigene phylogeny of the Old World mice, Murinae, reveals distinct geographic lineages and the declining utility of mitochondrial genes compared to nuclear genes. Mol. Phylogenet. Evol. 37:370–388.

Stevens L. 2001. Selection: Frequency-dependent. In: eLS. John Wiley & Sons, Ltd. Available from: http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0001763.pub2/abstract

Sun JC, Lanier LL. 2009. Natural killer cells remember: An evolutionary bridge between innate and adaptive immunity? Eur. J. Immunol. 39:2059.

Sun W, Dunning FM, Pfund C, Weingarten R, Bent AF. 2006. Within-species flagellin polymorphism in Xanthomonas campestris pv campestris and its impact on elicitation of Arabidopsis FLAGELLIN SENSING2-dependent defenses. Plant Cell 18:764–779.

Sundaram AY, Kiron V, Dopazo J, Fernandes JM. 2012. Diversification of the expanded teleost-specific toll-like receptor family in Atlantic cod, Gadus morhua. BMC Evol. Biol. 12:256.

Suzuki H, Shimada T, Terashima M, Tsuchiya K, Aplin K. 2004. Temporal, spatial, and ecological modes of evolution of Eurasian Mus based on mitochondrial and nuclear gene sequences. Mol. Phylogenet. Evol. 33:626–646.

Tada H, Nemoto E, Shimauchi H, et al. 2002. Saccharomyces cerevisiae- and Candida albicans-derived mannan induced production of tumor necrosis factor alpha by human monocytes in a CD14- and Toll-like receptor 4-dependent manner. Microbiol. Immunol. 46:503–512.

Tajima F. 1989. The effect of change in population size on DNA polymorphism. Genetics 123:597–601.

Takada T, Ebata T, Noguchi H, et al. 2013. The ancestor of extant Japanese fancy mice contributed to the mosaic genomes of classical inbred strains. Genome Res. 23:1329–1338.

Takahata N, Satta Y. 1998. Selection, convergence, and intragenic recombination in HLA diversity. Genetica 102-103:157–169.

Takahata N, Nei M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histocompatibility complex loci. Genetics 124:967–978.

Takeda K, Akira S. 2005. Toll-like receptors in innate immunity. Int. Immunol. 17:1–14.

Tal G, Mandelberg A, Dalal I, et al. 2004. Association between common Toll-like receptor 4 mutations and severe respiratory syncytial virus disease. J. Infect. Dis. 189:2057–2063.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Tavaré, S. 1986. Some probabilistic and statisical problems on the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences 17: 57-86.

Teeter KC, Payseur BA, Harris LW, et al. 2008. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 18:67–76.

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

Thompson RCA. 2013. Parasite zoonoses and wildlife: One health, spillover and human activity. Int. J. Parasitol. 43:1079–1088.

Tipping PG. 2006. Toll-like receptors: The interface between innate and adaptive immunity. J. Am. Soc. Nephrol. 17:1769–1771.

Tollenaere C, Brouat C, Duplantier JM, et al. 2010. Phylogeography of the introduced species *Rattus rattus* in the western Indian Ocean, with special emphasis on the colonization history of Madagascar. J. Biogeogr. 37:398–410.

Travis J. 2009. Origins. On the origin of the immune system. Science 324:580–582.

Tschirren B, Råberg L, Westerdahl H. 2011. Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents. J. Evol. Biol. 24:1232–1240.

Tschirren B, Andersson M, Scherman K, Westerdahl H, Råberg L. 2012. Contrasting patterns of diversity and population differentiation at the innate immunity gene Toll-like receptor 2 (TLR2) in two sympatric rodent species. Evolution 66:720–731.

Tschirren B, Andersson M, Scherman K, Westerdahl H, Mittl Peer RE, Råberg L. 2013. Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population. Proc. Biol. Sci. 280:20130364.

Tucker PK, Sage RD, Warner J, Wilson AC, Eicher EM. 1992. Abrupt cline for sex chromosomes in a hybrid zone between two species of mice. Evolution 46:1146–1163.

Tuon FF, Amato VS, Bacha HA, AlMusawi T, Duarte MI, Neto VA. 2008. Toll-like receptors and leishmaniasis. Infect. Immun. 76:866–872.

Turner AK, Begon M, Jackson JA, Bradley JE, Paterson S. 2011. Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population. PLoS Genet 7:e1002343.

Van Valen L. 1973. A new evolutionary law. Evol. Theory 1, 1-30.

Vasseur E, Patin E, Laval G, Pajon S, Fornarino S, Crouau-Roy B, Quintana-Murci L. 2011. The selective footprints of viral pressures at the human RIG-I-like receptor family. Hum. Mol. Genet. 20:4462–4474.

Vasseur E, Boniotto M, Patin E, Laval G, Quach H, Manry J, Crouau-Roy B, Quintana-Murci L. 2012. The evolutionary landscape of cytosolic microbial sensors in humans. Am. J. Hum. Genet. 91:27–37.

Vazquez-Mendoza A, Carrero JC, Rodriguez-Sosa M. 2013. Parasitic Infections: A role for C-Type lectins receptors. BioMed Res. Int. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3581113/

Verneau O, Catzeflis F, Furano AV. 1998. Determining and dating recent rodent speciation events by using L1 (LINE-1) retrotransposons. Proc. Natl. Acad. Sci. 95:11284–11289.

Vicoso B, Charlesworth B. 2006. Evolution on the X chromosome: unusual patterns and processes. Nat. Rev. Genet. 7:645–653.

Villaseñor-Cardoso MI, Ortega E. 2011. Polymorphisms of innate immunity receptors in infection by parasites. Parasite Immunol. 33:643–653.

Villesen P. 2007. FaBox: an online toolbox for fasta sequences. Mol. Ecol. Notes 7:965–968.

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

Vyskocilová M, Prazanová G, Piálek J. 2009. Polymorphism in hybrid male sterility in wild-derived *Mus musculus musculus* strains on proximal chromosome 17. Mamm. Genome Off. J. Int. Mamm. Genome Soc. 20:83–91.

Wakelin D. 1996. Helminths: Pathogenesis and defenses. In: Baron S, editor. Medical Microbiology. 4th ed. Galveston (TX): University of Texas Medical Branch at Galveston. Available from: http://www.ncbi.nlm.nih.gov/books/NBK8191/

Waldner H. 2009. The role of innate immune responses in autoimmune disease development. Autoimmun. Rev. 8:400–404.

Walsh C, Gangloff M, Monie T, Smyth T, Wei B, McKinley TJ, Maskell D, Gay N, Bryant C. 2008. Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa. J. Immunol. Baltim. Md 1950 181:1245–1254.

Wang CH, Eng HL, Lin KH, Chang CH, Hsieh CA, Lin YL, Lin TM. 2011. TLR7 and TLR8 Gene variations and susceptibility to hepatitis C virus infection. PLoS ONE 6:e26235.

Wang JR, Villena FP de, McMillan L. 2012. Comparative analysis and visualization of multiple collinear genomes. BMC Bioinformatics 13:S13.

Watts CHS, Baverstock PB. 1995. Evolution in the Murinae (Rodentia) assessed by microcomplement fixation of albumin. Aust. J. Zool. 43.

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

Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ. 2009. Variation matters: TLR structure and species-specific pathogen recognition. Trends Immunol. 30:124–130.

Williams NM, Timoney PJ. 1994. Variation in susceptibility of ten mouse strains to infection with a strain of *Ehrlichia risticii*. J. Comp. Pathol. 110:137–143.

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

Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. Nat. Genet. 32:569–577.

Worobey M, Bjork A, Wertheim JO. 2007. Point, counterpoint: The evolution of pathogenic viruses and their human hosts. Annu. Rev. Ecol. Evol. Syst. 38:515–540.

Woude MW van der, Bäumler AJ. 2004. Phase and antigenic variation in bacteria. Clin. Microbiol. Rev. 17:581–611.

Yamada E, Montoya M, Schuettler CG, et al. 2005. Analysis of the binding of hepatitis C virus genotype 1a and 1b E2 glycoproteins to peripheral blood mononuclear cell subsets. J. Gen. Virol. 86:2507–2512.

Yang H, Wang JR, Didion JP, et al. 2011. Subspecific origin and haplotype diversity in the laboratory mouse. Nat. Genet. 43:648–655.

Yonekawa H, Moriwaki K, Gotoh O, Miyashita N, Matsushima Y, Shi LM, Cho WS, Zhen XL, Tagashira Y. 1988. Hybrid origin of Japanese mice "*Mus musculus molossinus*": evidence from restriction analysis of mitochondrial DNA. Mol. Biol. Evol. 5:63–78.

Zaccone P, Fehervari Z, Phillips Jm, Dunne Dw, Cooke A. 2006. Parasitic worms and inflammatory diseases. Parasite Immunol. 28:515–523.

Zak DE, Aderem A. 2009. Systems biology of innate immunity. Immunol. Rev. 227:264–282.

Zaki HY, Leung KH, Yiu WC, Gasmelseed N, Elwali NEM, Yip SP. 2012. Common polymorphisms in TLR4 gene associated with susceptibility to pulmonary tuberculosis in the Sudanese. Int. J. Tuberc. Lung Dis. Off. J. Int. Union Tuberc. Lung Dis. 16:934–940.

Zhang D, Zhang G, Hayden MS, Greenblatt MB, Bussey C, Flavell RA, Ghosh S. 2004. A Toll-like receptor that prevents infection by uropathogenic bacteria. Science 303:1522–1526.

Zhang Z, Miteva MA, Wang L, Alexov E. 2012. Analyzing Effects of Naturally Occurring Missense Mutations. Comput. Math. Methods Med. [Internet] 2012. Available from: http://www.hindawi.com/journals/cmmm/2012/805827/abs/

Zhou H, Gu J, Lamont SJ, Gu X. 2007. Evolutionary analysis for functional divergence of the Toll-like receptor gene family and altered functional constraints. J. Mol. Evol. 65:119–123.

Zhu J, Brownlie R, Liu Q, Babiuk LA, Potter A, Mutwiri GK. 2009. Characterization of bovine Toll-like receptor 8: Ligand specificity, signaling essential sites and dimerization. Mol. Immunol. 46:978–990.

Zimmer C. 2001. Parasite rex: inside the bizarre world of nature's most dangerous creatures. New York: Simon & Schuster

Zwickl D. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion.

6 ANNEX

6.1 PRIMERS

Gene	Primer ID	Sequence of primers 5' – 3'	Function
Tlr1	TLR1_F1	CCT ATA CCC ATG TGG CAA TGC TC	amplification
Tlr1	TLR1_R1	GCA GCA ACA TCA TTG AGG TGG ATA TTC	amplification
Tlr1	MuMuTLR4_Ex1_F1_-327i	TTG CTC CAC GCC AGT TCC CT	amplification
Tlr1	MuMuTLR4_Ex1_R1_263i	TGC CTG TTC CTG CCC TGT GT	amplification
Tlr1	TLR1_S1_F	CCA ACT ACA GTT CCT GGG GTT GAG C	sequencing
Tlr1	TLR1_S2_F	AAG GCT TTG TCG ATA CAT CAA GTT GTC A	sequencing
Tlr1	TLR1_S3_F	CAG GCT TTG CAG GAA CTC AAT GTA GC	sequencing
Tlr1	TLR1_S4_F	TCC AGA GAA ACC TCC AGT TCC ATG C	sequencing
Tlr2	TLR2_F1	ATA AGC GTG ATA ATA ATG ATA TGT CC	amplification
Tlr2	TLR2_R1	AAC AGT ATT CAA GAC AAA ACC CAT AGT T	amplification
Tlr2	TLR2_F2	TGC CTG TAA CTT ATT CCT TGC ATG AGG	amplification
Tlr2	TLR2_R2	AGG AAG TCA GGA ACT GGG TGG AGA	amplification
Tlr2	TLR2-S1_F	TGG GAA ATC CTT ACC AGA CAC TGG GG	sequencing
Tlr2	TLR2-S2_F	ACG TAG TGA GCG AGC TGG GTA AAG T	sequencing
Tlr2	TLR2-S3_F	AGA CGC TGG AGG TGT TGG ATG T	sequencing
Tlr2	TLR2-S4_F	TCC AGG CCA AGA GGA AGC CCA A	sequencing
Tlr2	TLR2_Rs1_R	GAG TCA GGT GAT GGA TGT CG	sequencing
Tlr4	rTLR4-F	AGT TTA TCA TCA CTG YA GCA AG	amplification
Tlr4	rTLR4-XF2	CCC AAT TGA CTC CAT TCA AGC CC	amplification
Tlr4	rTLR4-XF3	CCC TCA GGA CTC TTG ATT GCA G	amplification
Tlr4	rTLR4-R-1	ATT CTC CCA AGA TCA ACC GAT G	amplification
Tlr4	rTLR4-R-3	CTG KTC CTT GAC CCA CTG C	amplification
Tlr4	rTLR4-R	AGA RMC CCA GRT GAR CTG TAG CAT T	amplification
Tlr4	MuMuTLR4_Ex1_F1_-327i	TTG CTC CAC GCC AGT TCC CT	amplification
Tlr4	MuMuTLR4_Ex1_R1_263i	TGC CTG TTC CTG CCC TGT GT	amplification
Tlr4	MuMuTLR4_Ex2_F1_-248i	TGG AGA AGG ATG GGT GTG ATA C	amplification
Tlr4	MuMuTLR4_Ex2_R1_632i	ACA GTG TGT TGC CCT TAT TTT CAC	amplification
Tlr5	MuMuTLR5_Ex3_F1_-113i	GCC ATT CTT CCT TGA ACC ACC ACA GA	amplification
Tlr5	MuMuTLR5_Ex3_R1_84i	AGC CTG CTC CAT GCC TGA CA	amplification
Tlr5	mTLR5_F	ACT GGG CAT TTC TGT TCC AC	amplification
Tlr5	mTLR5_R	TGG TGT TGT TTT CAT TGT CAG A	amplification
Tlr5	mTLR5_F1	GTG CTG TGT TAA GTG ACG GTT A	amplification
Tlr5	mTLR5_F2	TCG CAC GGC TTT ATC TTC TC	amplification
Tlr5	mTLR5_F3	GCC TCT GTT GGG ATG TTT TT	amplification
Tlr5	mTLR5_F4	GCT GGT GTT CAA GGA CAA GG	amplification
Tlr5	mTLR5_R1	AGG CTC GAG TTC ATC TTC ACA	amplification
Tlr5	mTLR5_R2	CAC GGT CAG CTT GTT AGC AC	amplification
Tlr6	TLR6_F1	TTA AAG TCT GTC TTA TTT CTA GAG CTT GGA	amplification
Tlr6	TLR6_R1	GCC CAG GTT GAC AGT TTA TTA AGT TA	amplification
Tlr6	TLR6-S1_F	GCC TGT GTG TAA GGA ATT TGG CAA CC	sequencing
Tlr6	TLR6-S2_F	AGT CAC TGA TGA TAG AGC ACG TCA A	sequencing
Tlr6	TLR6-S3_F	TGT CAC CCA CCT GCA GGC TTT	sequencing
Tlr6	TLR6-S4_F	GGC ACA TCC CCT TAG AGG AAC TCC AG	sequencing
Tlr7	rTLR7-F	AAG ACC YRT GTT GYT TAG TTT TAA TAA TG	amplification
Tlr7	rTLR7-1F	CAG ATT AGA CCT GGA AGC TTT AGT G	amplification
Tlr7	rTLR7-4F	TCT TGA CCT TGG CAC TAA CTT CAT A	amplification
Tlr7	rTLR7-5F	CCA TTG GCC AAA CTC TTA ATG G	sequencing
Tlr7	rTLR7-6F	GGT GAT AAC AGA TAC TTG GAC TTC T	sequencing
Tlr7	rTLR7-7F	CTG GCC ACT GAT GTG ACT TGT	sequencing
Tlr7	rTLR7-2R	GTT AGC CTC AAG GCT CAG AAG	amplification
Tlr7	rTLR7-9R	TAT CGG AAA TAG TGT AAG GCC TCA AG	amplification
Tlr7	rTLR7-R	AGA AAG AAR TTA TCK TCT ATC AGT CTC	amplification

Table 1. Primers for *Tlr4* and *Tlr7* exon 3, primers are common for whole Murine rodents; the rest of primers is tested only for *Mus musculus*.

6.2 PCR PROTOCOLS

	Reaction (ml)	Reaction (ml)				
	1	1	Stock		Final conc.	Dilution (x)
dd H2O	7,68	7,68				
buffer	1,30	1,30	10	x	1	10,00
MgCl2	0,78	0,78	25	mM	1,5	16,67
dNTPs	0,26	0,26	10	mM	0,2	50,00
TLR-R1	0,39	0,39	10	uM	0,3	33,33
TLR-F1	0,39	0,39	10	uM	0,3	33,33
Taq polymerase	0,20	0,20	5	U/ul		
DNA	2,00					
Volume	13	11				
Program						
1) 94°C	2 min					
2) 92°C	30 s					
3) 65°C	1 min	Number of cycles		35x		
4) 72°C	1min 30 s					
5) 72°C	10 min					

Table 2. Common PCR protocol for *Tlr1*, *Tlr2*, *Tlr4* exon 1 and 2, *Tlr5* exon 3, *Tlr6*, sequenced by Mgr. Zuzana Bainová

	Reaction (ml)	Reaction (ml)				
	1	1	Stock		Final conc.	Dilution (x)
dd H2O	9.3	9.3				
PCR Mix	12.5	12.5	10	x	1	10,00
TLR-R1	0,6	0,6	10	uM	0,24	20,00
TLR-F1	0,6	0,6	10	uM	0,24	20,00
DNA	2,00					
DNA	2,00					
Volume	25	23				
Program						
1) 95°C	15 min					
2) 95°C	40 s					
3) 68-58°C	45 s	Number of cycles		30x		
4) 72°C	45-1min 30 s					
5) 72°C	10 min					

Table 3. Common PCR protocol for *Tlr4* exon 3, *Tlr7* exon 3 and *Tlr5* exon 2.

6.3 CURRICULUM VITAE

Name: ALENA FORNŮSKOVÁ
Address: Litava 4, Olší, 59261
Date of birth: 4. 7. 1984
Nationality: Czech
Telephone: +420 605 464 704
Mail: fornuskova@ivb.cz, AlenaFornuskova@supagro.inra.fr

EDUCATION AND TRAINING

2009-2013 (dissertation defense-December 2013)

Ph.D. study – dissertation in double supervision: variability of innate immunity genes in wild rodent populations
Title of thesis: Genes of innate immunity and their significance in evolutionary ecology of free living rodents.
Institutions:

- Université Montpellier 2, Sciences et Techniques, Place Eugène Bataillon, 34095 Montpellier, France, successful application for grant of French government, 2010-2013 (supervisors: Dr. Jean-François Cosson and Dr. Nathalie Charbonnel)
- CBGP, Campus International de Baillarguet, 34988 Montferrier-sur-Lez, France
- Masaryk University, Faculty of Science, Department of Botany and Zoology, Kotlářská 2, 611 37 Brno, CZ (supervisor: Dr. Josef Bryja)
- Institute of Vertebrate Biology (IVB), Research Facility Studenec; AS CR, Studenec 122, 675 02 Koněšín (supervisor: Dr. Josef Bryja)

2007-2009

Master study: Population genetic structure of two cryptic species of the genus *Pipistrellus* in Central Europe.
Institutions:

- Université de Rennes 1, EFCE, 2, rue du Thabor, 35065 Rennes, France, successful application for grant of French government, 2008-2009 (supervisor: Dr. Eric Petit)
- Masaryk University, Faculty of Science, Department of Botany and Zoology, Kotlářská 2, 611 37 Brno,CZ (supervisor: Dr. Josef Bryja)

PROFESSIONAL EXPERIENCE

2009-2012
Position: Research assistant, project number GA206/08/0640
Content: Research in immunogenetics and population genetics,
Institution: IVB, Research Facility Studenec; AS CR, Studenec 122, 675 02 Koněšín,

2012→
Position: Project manager, PROVAZ, OP VK: CZ.1.07/2.4.00/17.0138
Content: Organization of workshops, seminars, student conferences, administration of project
Institution: IVB, Research Facility Studenec; AS CR, Studenec 122, 675 02 Koněšín,

FIELD EXPERIENCE

2009-2013: several field trips within the Czech Republic; sampling of small terrestrial mammals and birds

PERSONAL SKILLS AND COMPETENCES

Languages: **French** - Effective Operational Proficiency (D.E.L.F.), **English** - Effective Operational Proficiency, **Spanish** - Waystage, **Russian** – Waystage

Technical skills and competences: Application of molecular techniques to genetic of populations (DNA, RNA extraction, PCR, sequencing, cloning)

Bioanalysis: genotyping and sequence analysis-GeneMapper, SeqScape, BioEdit, population genetic analysis-Genepop, Genetix, FSTAT, DnaSP, Network, phylogenetic and selection analysis - MrBayes, RAxML, PAML, DataMonkey, FigTree, Mega, Jane 4, ParaFit,…

2013-2020: Certificate of professional competence in designing experiments and experimental project (Ministry of Agriculture, CR)

EXTRA ACTIVITIES AND AWARDS

2013 - Eight popularization lectures for the public
2013-2009: Teaching assistance: class - Mammals of CR, field excursions
2012 - Leading of high school student in secondary vocational activity (2nd place at regional level)
2010 - Successful application for Barrande project N°24504WM
2010 - Rector Price of Masaryk University
2010 - Successful application for Grant of French government (doctorate in double supervision between CBGP, Montpellier, FR and UBZ, Masaryk University, supervisors Dr. Nathalie Charbonnel, Dr. Jean-Francois Cosson, Dr. Josef Bryja)
2009 - 2nd. Price for poster at 1st International Symposium on Bat Migration, Berlin, 16-18 January: FORNŮSKOVÁ A., PETIT E., KAŇUCH P., BARTONIČKA T., ŘEHÁK Z. & BRYJA J.: Population genetic structure of two cryptic species (*Pipistrellus pipistrellus* and *P. pygmaeus*) in continental Europe suggests their long distance gene flow.
2008 - Successful application for Grant of French government (master studies, 11 months at University of Rennes I, supervisors Dr. Alain Butet and Dr. Eric Petit)

PUBLICATIONS

2013

- **FORNŮSKOVÁ A.**, VINKLER M., PAGES M., GALAN M., JOUSSELIN E., CERQUEIRA F., MORAND S., BRYJA J., CHARBONNEL N., COSSON J-F.: Contrasted evolutionary histories of two Toll-like receptors (*Tlr4* and *Tlr7*) in wild rodents (Murinae). *BMC Evol. Biol.* **13** (2013) 194. IF. 3.52
- **FORNŮSKOVÁ A.**, BRYJA J., VINKLER M., MACHOLÁN, M., PIALEK J., Contrasting patterns of polymorphism and selection in bacterial-sensing Toll-like receptor 4 in two house mouse subspecies Submitted in Ecology and Evolution.
- **FORNŮSKOVÁ A.**, PETIT E., BARTONIČKA T., KAŇUCH P., BUTET A., ŘEHÁK Z., BRYJA J.: Strong matrilineal structure in common pipistrelle bats is associated with variability in echolocation calls. In prep. for resubmission.

2010-2007

- HULVA P., **FORNŮSKOVÁ A.**, CHUDÁRKOVÁ A., EVIN A., ALLEGRINI B., BENDA P., BRYJA J.: Mechanisms of radiation in a bat group from the genus Pipistrellus inferred by phylogeography, demography and population genetics, *Molecular Ecology* **19** (2010) 5417-5431. IF. 5.52
- KAŇUCH P., **FORNŮSKOVÁ A.**, BARTONIČKA T., BRYJA J., ŘEHÁK Z.: Do two cryptic pipistrelle bat species differ in their autumn and winter roosting strategies within the range of sympatry?, *Folia zoologica* **59** (2010) 102-107. IF.0.548
- BRYJA J., KAŇUCH P., **FORNŮSKOVÁ A.**, BARTONIČKA T., ŘEHÁK Z.: Low population genetic structuring of two cryptic bat species suggests their migratory behaviour in continental Europe, *Biological Journal of the Linnean Society* **96** (2009) 103-114. IF. 2.19
- KAŇUCH P., **FORNŮSKOVÁ A**., BARTONIČKA, T., BRYJA J.: Multiplex panels of polymorphic microsatellite loci for two cryptic bat species of the genus *Pipistrellus*, developed by cross-species amplification within the family Vespertilionidae. *Molecular Ecology Notes*, **7**(2007): 871-873. IF. 2.38

INTERNATIONAL CONFERENCES – LECTURES

2013

- Fornůsková A., Vinkler M., Bryja J., Pialek J. Contrasting patterns in variability of Toll-like receptor 4, in the two subspecies of the wild house mouse, *Mus musculus*, Lecture in Eng., The 11th International Mammalogical Congress at Queens University in Belfast. 11. -16. august 2013 (lecture ENG)

2012

- Fornůsková A., Vinkler M., Galan M., Charbonnel N., Bryja J., Cosson J.F. Differences in evolution of Toll-like receptors 4 and 7 genes in wild rodents (murinae), Rodens and Spatium, Finland, 16-20. July 2012, (lecture ENG)

- Fornůsková A., Vinkler M., Bryja J., Pialek J. Intraspecific variability of the innate immunity gene *Tlr4* in the wild house mouse, *Mus musculus*, CBGP Montpellier, Réunion du groupe rongeurs du CBGP, 24. - 25. zari 2012, (lecture FR)

- Fornůsková A., Vinkler M., Bryja J., Pialek J. Intraspecific variability of the innate immunity gene *Tlr4* in the wild house mouse, *Mus musculus*, Mus Studenticus, Český Šternberk, 7.-8. června 2012, (lecture ENG)

2011

- Fornůsková A., Galan M., Cerqueira F., Bryja J. Charbonnel N., Cosson J.F. Evolution of Toll-like receptors 4 and 7 genes in wild rodents (Murinae). VIth European Congress of Mammalogy, 19-23 July 2011, Paris, France: 12. (lecture in ENG)

- Fornůsková A., Galan M., Cerqueira F., Bryja J. Charbonnel N., Cosson J.F., Evolution of Toll-like receptors 4 and 7 genes in wild rodents (Murinae). Nové Hrady 2011, 2. -4. listopadu, Mouse Genes, Gene expression, and behaviour, (lecture ENG)

2010

- Fornůsková A.,Vinkler M., Bryja J., Pialek J. Distribution de *Tlr4* dans la zone d'hybridation de *Mus musculus musculus* et *Mus musculus domesticus* en République Tchèque, CBGP Montpellier, Réunion du groupe rongeurs du CBGP 9 et 10 décembre 2010 (lecture FR)

INTERNATIONAL CONFERENCES – POSTERS

2012

- Fornůsková A., Vinkler M., Bryja, J., Pialek, J. Intraspecific variability of innate immunity gene, *Tlr4*, in the wild house mouse, *Mus musculus*, Rodens and Spatium, Finland, 16-20. July 2012, (poster in ENG)

2011

- Bainová Z., Fornůsková A., Čížková D., Bryja J., Piálek J., Vinkler M. Allelic variability in Toll-like receptor 1, 2, 4 and 6 in wild-derived mice (*Mus musculus*). 13th Congress of the European Society for Evolutionary Biology, 20-24 August 2011, Tübingen, Germany, (poster in ENG)

2009-2007

- Fornůsková A., Petit E., Kaňuch P., Bartonička T., Řehák Z., Bryja J. Population genetic structure of two cryptic species (*Pipistrellus pipistrellus* and *P. pygmaeus*) in continental Europe suggests their long distance gene flow. In: 1st International Symposium on Bat Migration, Berlin, 16-18 January 2009: 87. (poster in ENG)

- Kaňuch P., Fornůsková A., Bartonička T., Bryja J., Řehák Z. Different patterns of roosting behaviour in two cryptic pipistrelle bats (*Pipistrellus pipistrellus* and *P. pygmaeus*). XIth European Bat Research Symposium, 18-22 August, 2008, Cluj-Napoca, Romania: 79. (poster in ENG)

- Bryja J., Fornůsková A., Kaňuch P., Bartonička T., Patzenhauerová H., Řehák Z. No evidence of genetic structuring and isolation by distance in central European populations of migratory common pipistrelles (*Pipistrellus pipistrellus*). In: Prigioni C. & Sforzi A.. (eds.): Abstracts V European Congress of Mammalogy, Hystrix It. J. Mamm., (n.s.) Vol. I-2, Supp.: 195. (poster in ENG)

- Bryja J., Fornůsková A., Kaňuch P., Bartonička T., Patzenhauerová H., Řehák Z. Nízká genetická strukturovanost a geografická izolace u středoevropských populací netopýra hvízdavého (Pipistrellus pipistrellus). In: 8. celoštátna odborná konferencia s medzinárodnou účasťou. Výskum a ochrana cicavcov na Slovensku, Zvolen, 12. - 13.10.2007. (poster in ENG)

NATIONAL CONFERENCES

- Fornůsková A., Bainová Z., Vinkler M., Čížková D., Piálek J., Bryja J., 2011: Polymorfismus Toll-like receptoru 1, 2, 4 a 6 u dvou poddruhů myši domácí. In: Bryja J., Řehák Z. & Zukal J. (eds.): Zoologické dny Brno 2011. Sborník abstraktů z konference 17.-18. února 2011: 62. (poster in CZ)

- Fornůsková A., 2010: Analýza Toll like receptoru 4 u myších kmenů; genetic variation at the mouse *Tlr4* locus, exon 3, Mus studenticus Hynčice (lecture in CZ)

- Fornůsková A., Petit E., Kaňuch P., Bartonička T., Řehák Z., Bryja J., 2010: Comparison of nuclear and mitochondrial markers reveals strong female philopatry of two cryptic bat species of genus pipistrellus. In: Bryja J. & Zasadil P. (eds.): Zoologické dny Praha 2010. Sborník abstraktů z konference 11.-12. února 2010: 68. (lecture in CZ)

- Chudárková A., Fornůsková A., Benda P., Bryja J., Hulva P., 2010 : Phylogeography and demography of Pipistrellus pipistrellus and P. pygmaeus. In: Bryja J. & Zasadil P. (eds.): Zoologické dny Praha 2010. Sborník abstraktů z konference 11.-12. února 2010: 95-96. (poster in CZ)

- Vinkler M., Bainová H., Bainová Z., Fornůsková A., Tomášek O., Promerová M., Bryjová A. In: Bryja J., Řehák Z. & Zukal J. (eds.). Mohou Toll-like receptory přispět k poodhalení rozdílů v imunoekologii obratlovců?: Zoologické dny Brno 2011. Sborník abstraktů z konference 17.-18. února 2011: 62. (lecture by Dr Vinkler in CZ)

TRAININGS

- Winterschool, Brno 5-9.11 2011, International workshop, Fornůsková: Evolution of Toll-like receptors 4 and 7 genes in wild rodents (Murinae) (lecture ENG during student colloquium)
- Bioinformatics – theory and practical training, IVB, Research Facility Studenec, AS CR, Studenec 122, 22. April 2013, (sequence analysis, alignments, phylogeography, co-evolutionary analysis, molecular dating, Bayesian inference)
- Adaptive evolution in immune defence, IVB, Research Facility Studenec AS CR, Studenec 122, 7.-8. November 2013, (genetic variability of immune genes, natural selection shaping immune genes, maintenance of variability, immunopathology, speciation, sexual selection and immunity)

6.4 ACCEPTED ARTICLE

Fornůsková et al. BMC Evolutionary Biology 2013, 13:194
http://www.biomedcentral.com/1471-2148/13/194

BMC Evolutionary Biology

RESEARCH ARTICLE Open Access

Contrasted evolutionary histories of two Toll-like receptors (*Tlr4* and *Tlr7*) in wild rodents (MURINAE)

Alena Fornůsková[1,2,3*], Michal Vinkler[4], Marie Pagès[3,5], Maxime Galan[3], Emmanuelle Jousselin[3], Frederique Cerqueira[6], Serge Morand[3,7,8], Nathalie Charbonnel[3], Josef Bryja[1,2] and Jean-François Cosson[3]

Abstract

Background: In vertebrates, it has been repeatedly demonstrated that genes encoding proteins involved in pathogen-recognition by adaptive immunity (*e.g.* MHC) are subject to intensive diversifying selection. On the other hand, the role and the type of selection processes shaping the evolution of innate-immunity genes are currently far less clear. In this study we analysed the natural variation and the evolutionary processes acting on two genes involved in the innate-immunity recognition of Microbe-Associated Molecular Patterns (MAMPs).

Results: We sequenced genes encoding Toll-like receptor 4 (*Tlr*4) and 7 (*Tlr*7), two of the key bacterial- and viral-sensing receptors of innate immunity, across 23 species within the subfamily Murinae. Although we have shown that the phylogeny of both *Tlr* genes is largely congruent with the phylogeny of rodents based on a comparably sized non-immune sequence dataset, we also identified several potentially important discrepancies. The sequence analyses revealed that major parts of both *Tlr*s are evolving under strong purifying selection, likely due to functional constraints. Yet, also several signatures of positive selection have been found in both genes, with more intense signal in the bacterial-sensing *Tlr*4 than in the viral-sensing *Tlr*7. 92% and 100% of sites evolving under positive selection in *Tlr*4 and *Tlr*7, respectively, were located in the extracellular domain. Directly in the Ligand-Binding Region (LBR) of TLR4 we identified two rapidly evolving amino acid residues and one site under positive selection, all three likely involved in species-specific recognition of lipopolysaccharide of gram-negative bacteria. In contrast, all putative sites of LBR_{TLR7} involved in the detection of viral nucleic acids were highly conserved across rodents. Interspecific differences in the predicted 3D-structure of the LBR of both *Tlr*s were not related to phylogenetic history, while analyses of protein charges clearly discriminated Rattini and Murini clades.

Conclusions: In consequence of the constraints given by the receptor protein function purifying selection has been a dominant force in evolution of *Tlr*s. Nevertheless, our results show that episodic diversifying parasite-mediated selection has shaped the present species-specific variability in rodent *Tlr*s. The intensity of diversifying selection was higher in *Tlr*4 than in *Tlr*7, presumably due to structural properties of their ligands.

Keywords: Arms race, Host-pathogen interaction, Pattern recognition receptors, Adaptive evolution, Pathogen-Associated Molecular Pattern (PAMP)

* Correspondence: Alena.Fornuskova@supagro.inra.fr
[1]Institute of Vertebrate Biology, Research Facility Studenec, Academy of Sciences, Prague, Czech Republic
[2]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic
Full list of author information is available at the end of the article

BioMed Central

© 2013 Fornůsková et al.; licensee BioMed Central Ltd. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/2.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background

An effective immune defence is dependent on well-timed activation of an appropriate immune response. Pathogen recognition by innate immunity Pattern Recognition Receptors (PRRs) is crucial in this process [1,2]. The PRRs detect molecular structures named Microbe-Associated Molecular Patterns (MAMPs) that are conservatively present among individual microorganism taxa, because they are essential for their survival (such as, *e.g.*, bacterial lipopolysaccharides, muramyl dipeptide, peptidoglycan, flagellin, mannose, bacterial, fungal, parasitic and viral nucleic acids) [3]. Recent studies have associated polymorphism in genes encoding PRRs with variability in resistance or susceptibility to several infectious diseases in humans, laboratory mice and poultry *e.g.* [4-8]. However, in wildlife, molecular variation in PRR genes is still poorly documented [9-14].

Understanding the evolution of the immune system in general has been a challenge for evolutionary biologists and ecologists since JBS Haldane associated natural selection with infectious diseases [15]. In vertebrates, the study of selection patterns was mostly oriented towards genes of acquired immunity which are now intensively studied even in wild populations. Among them, genes of the major histocompatibility complex (MHC) are the most explored and the role of balancing selection in their evolution is generally accepted and well understood [16-23]. The quite late discovery of genes involved in the second branch of vertebrate immunity, *i.e.* innate immunity, among which the most important PRRs are Toll-like receptors (hereafter abbreviated according to the mouse gene and protein nomenclature as *Tlr*s and TLRs, respectively) [24-27], has resulted in modest research of their evolution in wildlife populations [28].

Generally, two subclasses of TLRs are distinguished in vertebrates according to the ligands they target [3,9,29,30]. The first subclass includes TLR1, TLR2, TLR4, TLR5, TLR6 and TLR10. These TLRs predominantly detect bacterial components (but also fungal and to lesser extent viral components) and are expressed on the outer cell membrane. Throughout this paper we term them "bacterial-sensing" TLRs. The second subclass includes TLR3, TLR7, TLR8 and TLR9 and targets mainly viral components (*e.g.* ssRNA, dsRNA, DNA containing unmethylated CpG), hereafter termed "viral-sensing" TLRs. These TLRs are expressed mostly within cells into the membranes of endosomal compartments. This current spectrum of genes for TLRs arose by multiple gene duplication and during the last 700 Mya diversified to recognize distinct MAMPs [29,31-36].

TLRs of both subclasses are transmembrane proteins composed of three domains [34,37]. The Extra-Cellular Domain (ECD) consists of a varying number of Leucin-Rich Repeat motifs (LRRs) that form a horseshoe-shaped tertiary structure of the ECD. This domain contains the Ligand Binding Region (LBR) which is directly responsible for physical interactions with the pathogen-derived structures and as such it is likely subject to intensive selection. The ECD is followed by a short Transmembrane Domain (TM), and an Intracellular domain (ICD) containing the Toll/Interleukin-1 Receptor (TIR) domain responsible for TLR signaling [3]. As previously shown [38], non-synonymous SNPs located in LBR may affect the 3D structure of the protein and its surface charge. This may have important functional consequences, influencing receptor ability to bind pathogens [14,36,39], and may even lead to the evolution of species-specific ligand recognition [40,41]. Appropriate binding of MAMPs by LBR is connected with changes in receptor dimerization [42-44] that induce signaling and release of cytokines triggering mainly Th1 and Th17 inflammation, fever and phagocytosis [45-47]. The TLR signaling ensures an immediate response to invading microorganisms that, in a second step, further directs the following adaptive immune response [48,49].

Previous studies, mostly based on investigation in humans, primates and domestic or laboratory animals, provided information regarding some general patterns of TLR evolution and maintenance of their genetic polymorphism [2,9,50-52]. These studies revealed that the ECD is more frequently a target of positive selection than the TIR domain. Moreover, in general the viral-sensing TLRs seem to evolve under stronger purifying selection than the bacterial-sensing ones [53-56]. However, up to now, the evidence of TLR polymorphism and the type of selection that shapes this polymorphism in natural populations remain rare [10-14]. Besides, to our knowledge the precise investigation of the LBR variability and evolution is missing. Such information could nevertheless be important to better understand species-specific differences in the susceptibility to various pathogens [57].

In the present study we focused on the molecular variation of the genes encoding the bacterial-sensing TLR4 (binding mainly bacterial lipopolysaccharides, LPS, as a ligand) [58] and the viral-sensing TLR7 (binding viral ssRNA) [59,60] in 23 species of the subfamily Murinae. Murine rodents are largely distributed over the world and several species (such as rats and mice) live in close proximity to humans. A recent review showed that 60% of the agents of emerging diseases in humans circulate in animals [61] and most of the natural reservoirs of a number of serious viral and bacterial emerging agents of zoonoses are rodents [62,63]. Species-specific molecular variability in immune-related genes may be responsible for differences in the ability of rodent species to transmit these pathogens. Herein we aimed to document evolutionary histories of these two *Tlr*s during murine

diversification. We implemented statistical approaches to infer *Tlr* phylogeny and to detect selection acting on DNA and amino acid (AA) sequences. We searched for deviations from "species" phylogeny based on a comparably sized non-immune sequence dataset by contrasting phylogenetic trees reconstructed from *Tlr* sequences with those reconstructed from "neutral" genes (both mitochondrial and nuclear). Deviations would indicate the occurrence of non-neutral patterns during the *Tlr* evolutionary history, *e.g.* adaptive selection [9,64,65]. Next we estimated putative functional changes in the LBR by examining variability in predicted tertiary 3D-structures of the proteins, and in biophysical properties of proteins (charge and structural characteristics) at polymorphic binding sites. Finally, we compared the evolutionary histories of the two TLRs to reveal potentially distinct evolutionary pressures shaping these proteins.

Results

Sequence analyses

Amplification and sequencing were successful in 96 samples representing 23 rodent species for *Tlr*4 and in 96 samples representing 22 species for *Tlr*7 (Additional file 1: Table S1). Only samples from one species - *Maxomys surifer* could not be completely sequenced for *Tlr7* - the first 180 bp were missing and we excluded this species from the *Tlr*7 analyses. No stop codons, indels nor recombination were detected in these data using SBP (DATAMONKEY).

For the whole *Tlr*4 coding sequence (CDS), the three different domains were predicted by SMART as follows: ECD from AA position 1 to 635, TM from position 636 to 658 and ICD from position 659 to 835 in which the TIR domain (from position 671 to 816) and ICD distal part (ICD-DP; from 817 to 835) may be identified (Additional file 1: Figure S1). For *Tlr*7, the predicted location of the three domains was the following: ECD from position 1 to 850, TM from position 851 to 873 and ICD from position 874 to 1050 (TIR from 894 to 1033 and ICD-DP from 1034 to 1050; Additional file 1: Figure S1). In general, *Tlr*4 was more diverse than *Tlr*7, and within each *Tlr*, the ECD domain was more variable than the TIR domain in both molecules (Table 1). Surprisingly, ICD-DP located on the C-terminal end of *Tlr*4 represented the most variable region of exon 3 ($\pi_{\text{ICD-DP-}Tlr4}$ = 0.102±0.015).

Phylogeny and co-divergence between the tree based on a comparably sized non-immune sequence dataset and TLR trees

Both phylogenetic approaches (MrBayes and RAxML) displayed similar trees for both *Tlr*s (Additional file 2: Figures S2 and S3). Minor differences between ML and Bayesian trees were found only at the intraspecific level. *Tlr*4 topology was well-supported with posterior probabilities (pp) ≥ 0.95 despite a lack of resolution within the black rat species complex (including *Rattus rattus*, *R. tanezumi*, *R. sakeratensis*, *R. tiomanicus*, *R. argentiventer*, *R. andamanensis*), between two *Bandicota* species (*Bandicota savilei* and *B. indica* did not form reciprocal monophyletic clades) and between two subspecies of the house mouse (Additional file 2: Figures S2a and S3a). Sequences of *Tlr*7 were also predominantly clustered according to species with strong supports (pp ≥ 0.95). Relationships between Asiatic mouse species were not fully resolved (monophyly of *Mus caroli*, *M. cooki* and *M. cervicolor* supported with a moderate pp

Table 1 Estimates of sequence diversity and average codon-based evolutionary divergence over all sequence pairs for the exon 3 and particular domains of *Tlr*4 and *Tlr*7 genes

Tlr domains	*n*	*L*	π±S.E.	*hN*	*hA*	*S*	*Eta*	*dN*±S.E.	*dS*±S.E.	*dN/dS*
Tlr4										
Exon 3	96	2247	0.049±0.003	122	90	545	625	0.038±0.003	0.102 ±0.008	**0.481**
ECD	96	1647	0.053±0.003	112	83	441	504	0.045±0.004	0.098 ±0.009	**0.597**
LBR	96	666	0.072±0.006	67	50	203	242	0.070±0.008	0.108±0.015	**0.787**
TIR	96	435	0.031±0.002	54	11	68	79	0.004±0.002	0.143 ±0.024	**0.067**
Tlr7										
Exon 3	96	3147	0.034±0.003	79	49	466	518	0.021±0.002	0.088 ±0.007	**0.398**
ECD	96	2547	0.037±0.003	75	48	407	455	0.025±0.002	0.089±0.007	**0.468**
LBR	96	311	0.035±0.003	19	8	37	38	0.018±0.006	0.107±0.024	**0.196**
TIR	96	420	0.026±0.003	26	6	43	47	0.007±0.003	0.105±0.021	**0.070**

NOTE. - ***ECD*** extracellular domain, ***LBR*** - ligand biding region, ***TIR*** Toll/interleukin-1 receptor domain, ***n*** the number of sequenced individuals, ***L*** length of analysed sequences in base pairs, π average number of nucleotide differences per site between two sequences, ***S.E.*** Standard error, ***hN*** number of nucleotide alleles, ***hA*** number of amino acid variants, ***S*** number of polymorphic sites, ***Eta*** total number of mutations, ***dS*** number of synonymous substitutions per synonymous site (estimated by MEGA), ***dN*** number of non-synonymous substitutions per non-synonymous site (estimated by MEGA). Analyses were conducted using the Nei-Gojobori model; **S.E**. of *dN* and *dS* - were obtained by a bootstrap procedure (1000 replicates); ***dN*/*dS*** were computed by SLAC (DATAMONKEY).

value of 0.86 and Bootstrap values, Bp = 81) as well as those between *Leopoldamys* species (*L. edwardsi* appeared more closely related to *L. neilli*, rather than to *L. sabanus* but with a low pp of 0.6, Bp = 48). Similarly to *Tlr4*, branching orders within the genus *Rattus* were not resolved: *Rattus exulans* (clade I) was retrieved monophyletic without ambiguity (pp = 1, Bp = 100), *R. norvegicus* and *R. nitidus* were grouped together with the highest support (clade II, pp = 1, Bp = 100) and the remaining *Rattus* species formed a moderately supported group (clade III, pp = 0.7, Bp = 98, for more details see Additional file 2: Figures S2b and S3b).

At the first glance, *Tlr* phylogenies (based on MrBayes approach) of the black rat complex was congruent to the tree based on a comparably sized non-immune sequence dataset (Figure 1). The number of co-divergence events inferred using JANE 4 was significantly higher than expected by chance, meaning that the two phylogenies were similar (Additional file 1: Figure S4). However, the Shimodaire-Hasegawa test showed significant disagreement between the species tree and both *Tlr*s phylogenies (Δln L = 257, ddl = 1, $p < 0.001$ for *Tlr4*; Δln L = 76, ddl = 0.008, $p < 0.05$ for *Tlr7*), indicating that neither of the *Tlr* trees coincided precisely with the tree based on a comparably sized non-immune sequence dataset. The incongruence was mainly caused by recently diverged species of *Rattus*. However, we revealed several other differences, such as the misplacement of the genus *Bandicota* (occurring within *Rattus* in the *Tlr4* tree) and the different positions of *R. sakeratensis* and *R. exulans* in species and *Tlr7* trees (Figure 1).

Evidence of signatures of selection

The comparison of ω (dN/dS) revealed substantial differences between the two *Tlr*s, as well as between gene parts encoding different domains (for details see Table 1). The difference between gene parts was mainly due to variations in the number of non-synonymous substitutions (which was higher in ECDs than in the TIR), while they both had similar numbers of synonymous substitutions.

The highly conservative SLAC (Single Likelihood Ancestor Counting) analysis (DATAMONKEY) [66] revealed two codon positions evolving under positive selection in *Tlr4* and only one in *Tlr7*, all of them being located within the ECD domain ($p < 0.05$, Table 2, Figure 2). We found 26 and 10 negatively selected sites for *Tlr4* and *Tlr7* respectively ($p < 0.05$, Table 2, Figure 2), distributed evenly over the whole sequences.

The imprint of natural selection on protein coding gene is often difficult to reveal because selection is frequently episodic (*i.e.* it affects only a subset of lineages) [67]. We therefore looked for evidence of episodic diversifying selection at individual sites along the evolutionary branches of the trees using the MEME algorithm. Thirteen codon positions were found to be affected by episodic selection for *Tlr4* (1.7% of all analysed codons) while only 4 codon positions showed this signature for *Tlr7* (0.38% of all analysed codons). In *Tlr4*, 12 of these sites were located directly in LBR, while in *Tlr7* none of the sites evolving under positive selection were in LBR. Whatever the *Tlr* gene considered, all sites found to evolve under positive selection using the SLAC were identified also by the MEME algorithm.

The signs of positive selection were scattered over whole *Tlr* trees, affecting nearly all branches of the *Tlr4* phylogeny, both basal and terminal, while they mostly concerned the terminal branches for the *Tlr7* phylogeny (Figure 3). Interestingly, one site evolving under positive selection ($p < 0.05$) was located in the ICD-DP of *Tlr4* gene (Table 2, Figure 2a). We found that this part (*i.e.* the last 57 bp of C-terminal end of the protein following the TIR domain) was highly variable (19 nucleic acid alleles and 16 AA variants) with a mean ω = 1.11.

Analysis of the ligand binding regions

In general, the Ligand Binding Region (LBR) was much more variable in *Tlr4* than in *Tlr7* genes. We detected 50 different AA variants of the LBR in the TLR4 dataset, while only eight different AA variants were detected in TLR7. Out of the 222 AA sites of $LBR_{TLR4,}$ 43% were polymorphic, while among the 103 AA sites of $LBR_{TLR7,}$ only 10% exhibited genetic variations. The CONSURF analysis performed to estimate the degree of evolutionary conservation of each amino acid position in LBR revealed 10% of phylogenetically variable positions (*i.e.* 22 positions assigned to grade 1 and corresponding to the most variable and rapidly evolving amino acid positions out of 222 positions in total) in TLR4, but only 2% (2 positions with grade 1 out of 103) in TLR7 (Figure 4). Other positions were assigned as conservative (57% and 79% in TLR4 and TLR7, respectively) or had insufficient support (33% and 19%, respectively; Figure 4).

Ligand-binding positions in rodents were predicted by comparison with those identified in humans by Park et al. [39]. In TLR4, two out of eight LPS-binding amino acid positions were identical to humans and strictly conserved among rodents (F438 and F461). Three other were conserved in terms of amino acid features (*i.e.* polarity, hydrophobicity) but distinct from human residue and variable among rodents (R263K, K360R and K434R). Interestingly, one LPS binding site that was uniform in human was found to be evolving under positive selection using the MEME algorithm. We found hydrophobic and hydrophilic residues, although this position, L442Y, is known to be involved in hydrophobic interactions. Finally, two remaining positions were found to be highly variable in rodents (339 and 386) (Additional file 1: Table S3). In TLR7, the

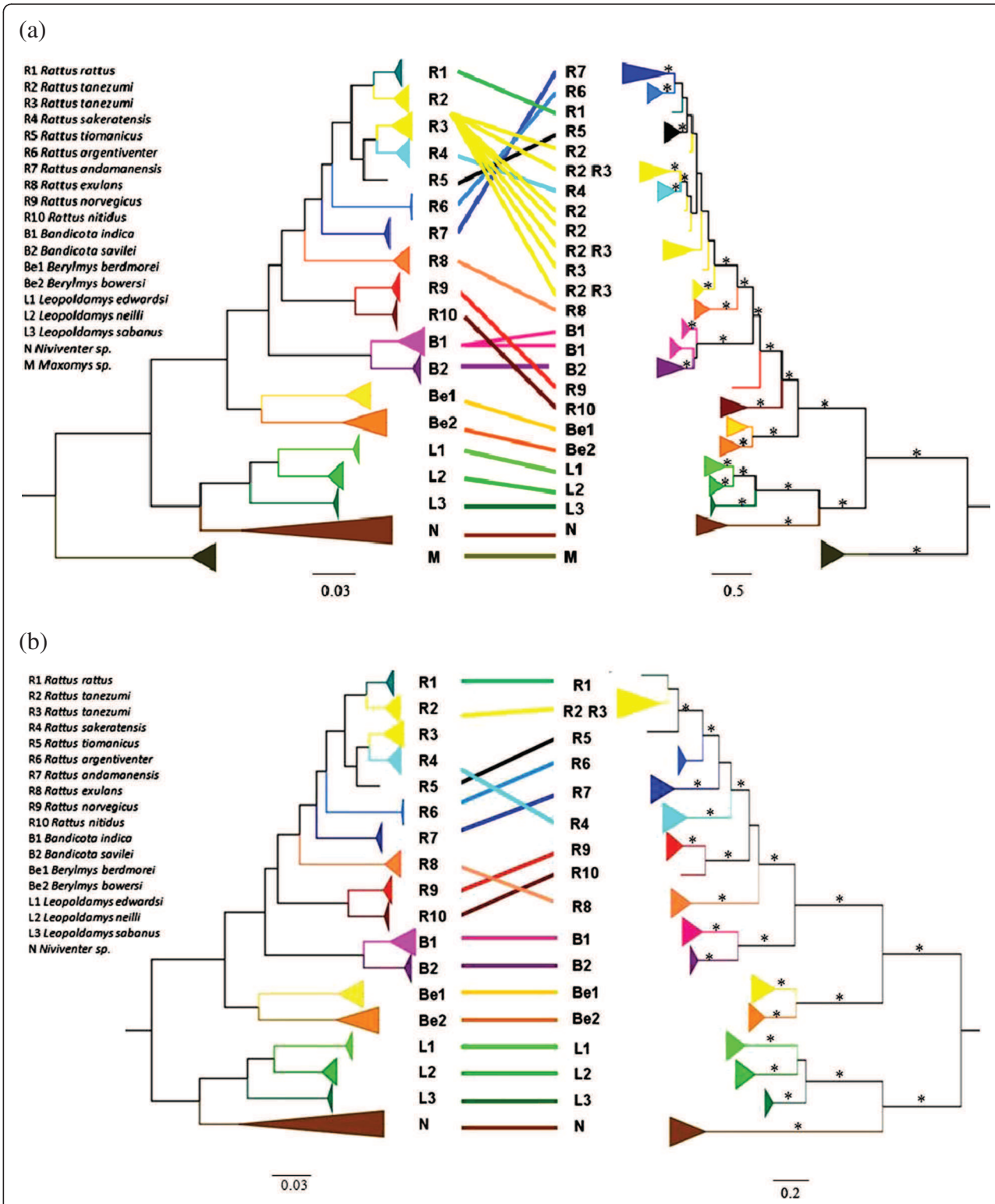


Figure 1 Comparison of phylogenetic trees based on *Tlr*s and neutral markers. Comparison of the Bayesian phylogenetic trees of *Tlr*4 **(a)** and *Tlr*7 **(b)** on the right with phylogenetic trees based on presumably neutral markers (*Cytb*, *Col*, *Irbp*; for more details see Pagès et al. 2010) on the left. Abbreviations (R1, R2….M) indicate species assignment used in Pagès et al. 2010; corresponding legend is on the left. Color lines link the supported clades represented by the same species; * indicates posterior probabilities (pp) > 0.95.

Table 2 Positively (MEME and SLAC-PS) and negatively (SLAC-NS) selected sites detected for the exon 3 of *Tlr4* and *Tlr7* at p < 0.05

	ECD (and LBR)	TD	TIR	ICD-DP
Tlr4	**1-635 (248–469)**	**636-658**	**671-816**	**817-835**
MEME	273, 335, 345, 347, 361, 363, 366, 368, 394, 398, 442, 469	-	-	818
SLAC-PS	347, 469			
SLAC-NS	99, 105, 149, 240, 253, 364, 457, 461, 463, 518, 522, 529, 549, 616, 635	-	679, 688, 691, 721, 740, 772, 782, 785, 793, 811	822
Tlr7	**1-850 (495–597)**	**851-873**	**894-1033**	**1034-1050**
MEME	128, 308, 461, 772	-	-	-
SLAC-PS	308			
SLAC-NS	156, 272, 455, 528, 541, 671, 676, 709,	-	963, 971	-

NOTE. - ***ECD*** extracellular domain, ***LBR*** ligand binding domain, ***TD*** transmembrane domain, ***TIR*** TIR domain, ***ICD-DP*** distal part of intracellular domain. Prediction of domains and numbering of sites are according to the reference protein sequence of *Rattus norvegicus* taken from GenBank [NP_062051.1 for *Tlr*4 and NP_001091051.1 for *Tlr7*]. Sites located in LBR are underlined.

nine ligand binding residues predicted following Wei et al. [68] were strictly conserved within rodents and seven of them were common to both rodents and human TLR7 (Additional file 1: Table S4).

The pairwise RMSD that allowed estimating the differences in 3D protein structure among variants varied from 0 to 1.5Å in TLR4 variants, and from 0.6 to 1.7Å in TLR7 variants (Additional file 1: Figure S5). Yet, in the phenetic diagram of TLR4, 3D-structures of *Rattus sakeratensis* and *Rattus nitidus* were distinct from each other and also from all other species. Similarly for TLR7, the 3D-structure of the protein of *Rattus exulans* was separated from other species (Additional file 1: Figure S5). To provide wider context we performed additional comparison between PDB structures (obtained from The RCSB Protein Data Bank http://www.rcsb.org/pdb/home/home.do) of human (HoSaTLR4-3fxi_A) and mouse (MuMuTLR4-3vq2_A) ECD_{TLR4} and between ECD of mouse TLR4 and TLR3 (MuMuTLR3-3ciy_A). The comparison between species of the same TLR was 1.7Å (HoSaTLR4-MuMuTLR4). Comparison between two TLRs from most distant TLR families of the same species was 4.6Å (MuMuTLR4-MuMuTLR3). The analysis of electric charge of LBR revealed higher variation in TLR4 (from −7.7 to 1.5) when compared with TLR7 (from −1.6 to 0.6). Detailed analyses of LBR_{TLR4} revealed that *Mus* and *Rattus* species were well differentiated from each other (*Mus*: from −7.7 to −3.7; *Rattus* and related genera: from −3 to 1.5, Additional file 1: Figure S6a). Similar pattern was found for LBR_{TLR7} (*Mus*: -1.6, *Rattus* and related genera: from −1.4 to 0.6, Additional file 1: Figure S6b).

Discussion

In this study we analysed the variability of two important vertebrate immune genes involved in innate immunity across wild murine rodents and we looked for evidence of selection. Overall, we found that *Tlr4* was much more variable than *Tlr7* and that the evolution of both genes had been influenced mostly by purifying selection. However, comparison of both *Tlrs* revealed contrasting evolutionary patterns. *Tlr7*, which is involved in the recognition of viral nucleic acids, was highly conserved across rodents and its evolution seemed to be strongly shaped by purifying selection. Predicted ligand binding sites in LBR_{TLR7} were identical across all species and only few sites were detected to evolve under positive selection within the whole molecule. By contrast, *Tlr4*, which detects several different pathogen ligands, was more variable and was affected by numerous events of episodic selection. Positively selected sites mostly occurred in LBR, probably as a result of co-evolution with pathogens. Analyses of the LBR variability in surface charge revealed a potential for interspecific differences in ligand binding capacities of both *Tlr*s.

Differences in TLRs evolution - phylogenetic approach

We found that both *Tlr*s were conserved genes as their phylogeny almost correctly recapitulated species phylogeny. In spite of this conservatism we revealed some incongruence between gene and species topologies, especially in branches represented by the shallow genealogy of the black rat complex and *Bandicota* spp. (Figure 1a). These species have experienced recent and rapid radiation during the Early Pleistocene about 1 Mya [69,70]. Discrepancies between a gene genealogy and the species phylogeny in recently diverged species often results from Incomplete Lineage Sorting (ILS) of ancestral polymorphism and/or episodic gene flow and hybridization [71,72]. Indeed, *R. tanezumi R2* and *R. tanezumi R3* were recently proposed as conspecifics or were suspected to hybridize in Southeast Asia [73]. In addition, hybridization with introgression occurred between the invasive populations of *R. tanezumi* and *R. rattus* in the United States [74]. These phenomena could

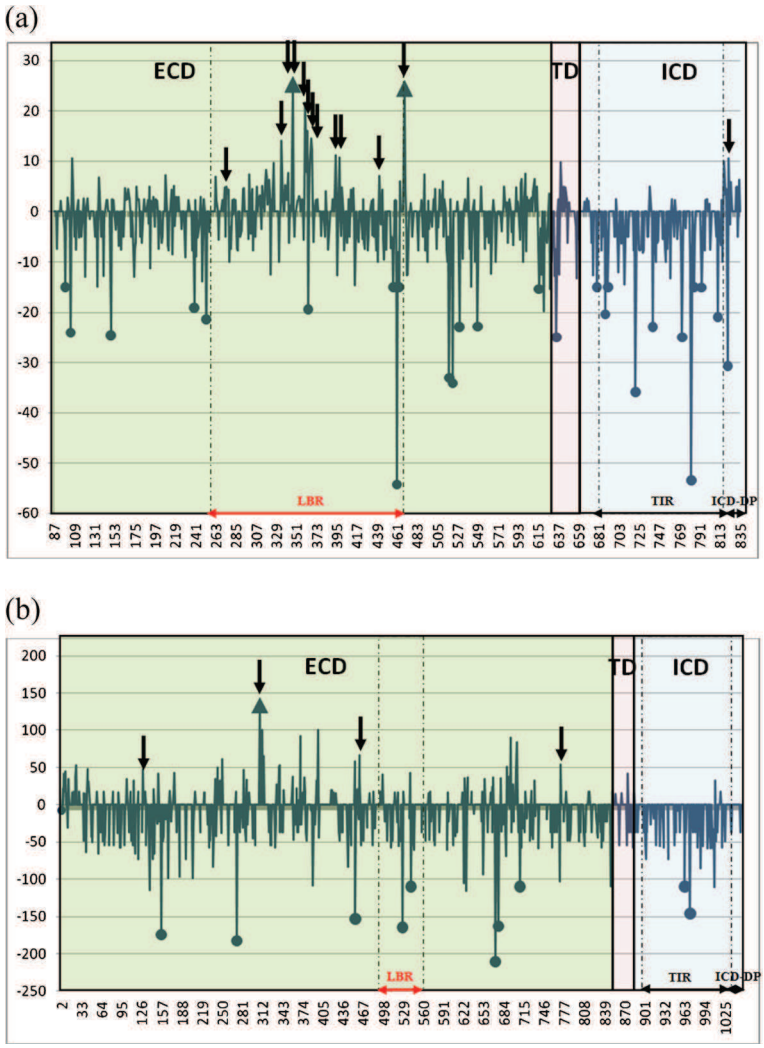


Figure 2 Distribution of sites under selection identified by SLAC and MEME. Intensity of selection acting on *Tlr4* **(a)** and *Tlr7* **(b)** exon3 with $p < 0.05$; the blue line is normalized dN-dS calculated in SLAC (DATAMONKEY); blue arrows-up - sites under positive selection detected by SLAC; black arrows-down - sites under positive selection detected by MEME (DATAMONKEY); blue full circles - sites under negative selection detected by SLAC. ECD - extracellular domain; LBR - ligand binding region; TD - transmembrane domain; TIR - TIR domain; ICD - intracellular domain; ICD-DP - distal part of intracellular domain.

explain incongruence between *Tlr*s and species trees. However, directional selection could also be involved. Discrepancies in *Tlr7* phylogeny represented by *R. exulans* and *R. sakeratensis* seem more likely to be caused by pathogen selective pressure (Figure 1b). ILS and hybridization are unlikely to result in such deeper changes, whereas the influence of directional selection (positive or negative) on non-neutrally evolving genes could be at more likely explanation [75]. The rejection of co-divergence (concerning basal nodes) between *Tlr*s and species phylogenies could reflect the occurrence of pathogen-driven selection on *Tlr*s during the evolutionary history of the murine rodents [32,76]. The former hypothesis should now be tested by a detailed analysis of spectrum of pathogens from rodents to determine if the species producing the incongruent topology displayed specific pathogens that could mediate this selection.

Tlr variability and signatures of selection

We found that 92% and 100% sites (respectively for *Tlr4* and *Tlr7*) evolving under positive selection were located in the ECD, which is responsible for pathogen recognition. For *Tlr4* 92% of these positively selected sites found by MEME algorithm were located in the LBR. This is in concordance with several recent studies conducted on primates, birds and rodents, that have suggested a high accumulation of positively selected sites at LBR [9-11,77,78]. Surprisingly, none of the sites evolving under positive selection was identified directly in the LBR of *Tlr7*.

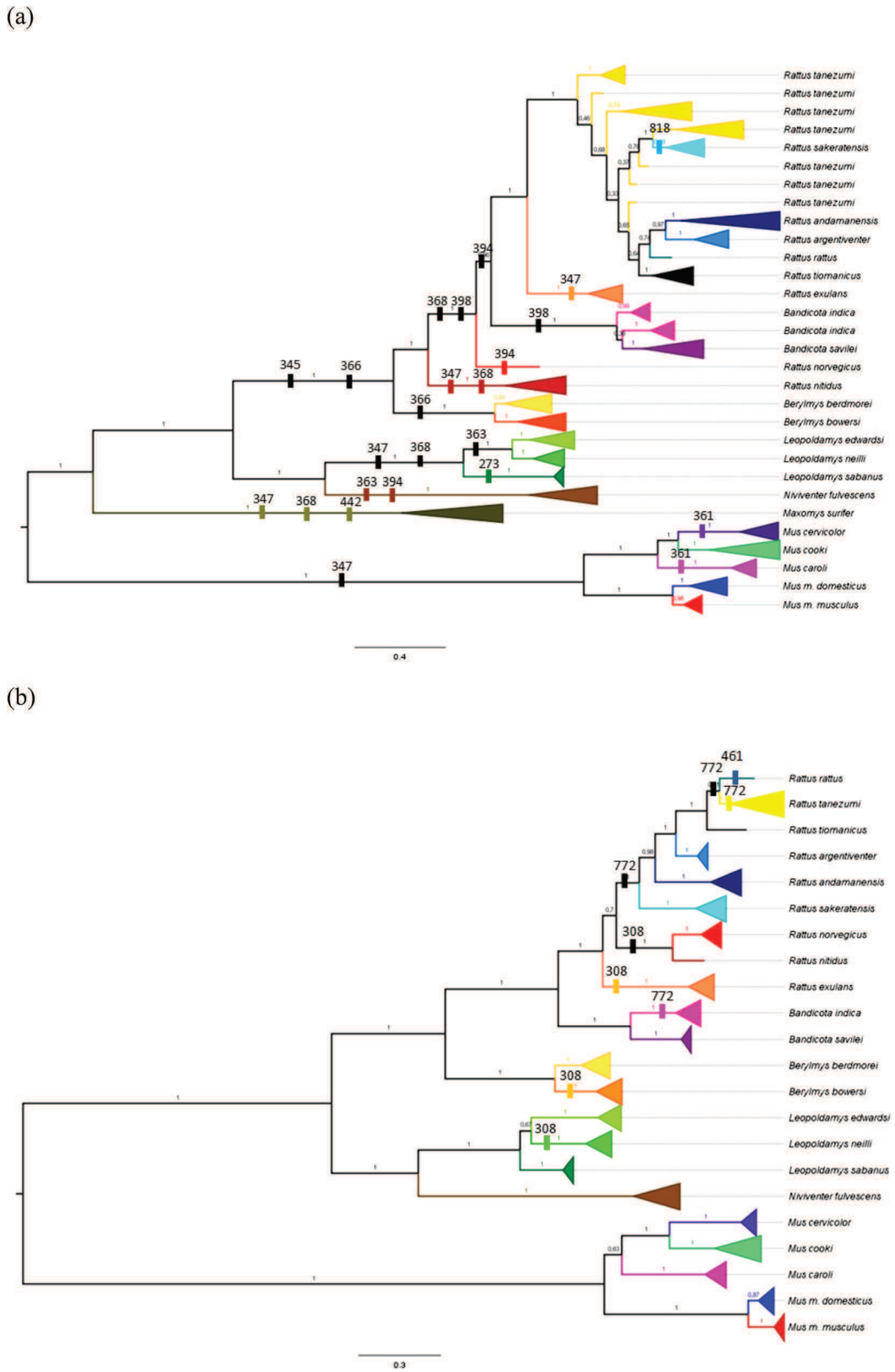
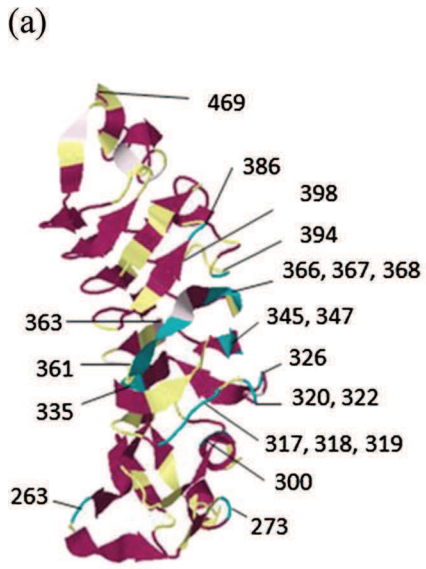


Figure 3 Sites under positive selection identified in evolutionary lineages by MEME. *Tlr4* **(a)**, *Tlr7* **(b)** (significance level at $p < 0.05$), positively selected sites are marked and numbered above branches at simplify phylogeny based on MrBayes.



Codon	Residue Variety
263	R, K
273*	H, R, P
300	D, N, Y, V
317	S, H, D, P, Y
318	I, L, V
319	Q, R, K, E
320	H, Q, Y
322	A, K, G, E
326	S, N, K, R
335*	I, V
345*	A, T, P, K
347*	H, S, T, D, N, R, K
361*	A, Q, K, R, G, L, E
362	S, H, T, D, P
363*	T, M, I, L
366*	D, R, K, G
367	A, Q, T, D, R, K
368*	M, L, V
386	S, T, I, R
394*	F, M, L, V
398*	S, D, N, I, G
469*	A, T, I, G, V

Codon	Residue Variety
497	Q, K
540	Q, W, R

Figure 4 Mapping of evolutionary conservation of amino acid positions in a protein molecule based on the phylogenetic relations between homologous sequences. Conserved amino acid positions in LBR of TLR4 **(a)** and TLR7 **(b)**. Structure of LBR was analysed in CONSURF; computations were based on MrBayes phylogenetic trees and tertiary protein structures of *R. norvegicus* [Gen Bank Acc. KC811688/ KC811786]; most variable positions are highlighted in turquoise and numbered (grade 1); most conserved sites are in violet; yellow sites mark insufficient data; white sites have average conservation score; tables show residue variants at the phylogenetically variable positions with grade 1; codons with asterisk have been identified as those under positive selection by MEME.

The TIR domain of both *Tlr*s was evolving under much stronger functional constraint than the ECD in both genes. We found only 11 amino acid variants of TIR_{TLR4} in 23 species and six different variants of TIR_{TLR7} in 22 species. Altogether our results support the observation that *Tlr* exodomains evolve more rapidly than the intracellular TIR domain [9,56,77,78]. The requirement of sites within ECD, which would be involved in ligand recognition and able to recognize permanently fast-evolving pathogens, could explain this pattern. Besides, the high conservation of the TIR domain could be adapted to maintain a functional response of signal transduction see, e.g. [9,33,50,56,58,79].

Both genes showed non-significant differences between ECD and TIR with respect to *dS*, supporting the hypothesis that there was no difference in mutation rate between ECD and TIR. The same result has been found in comparative studies of 10 vertebrate TLRs [33]. The distal part of ICD in *Tlr*4 was surprisingly highly variable among rodent species. The reason for such a high level of variability is still unknown; however some authors suggest that this region at the carboxy-terminal end of *Tlr*4 could be responsible for interspecific differences in LPS sensitivity [50].

Positive selection we also detected using the MEME approach that individually considers each codon along the *Tlr*s phylogeny [67]. We found that episodic positive selection affected most lineages in the phylogenetic tree of *Tlr*4, while the situation was quite different in *Tlr*7, where the sites evolving under positive selection were mostly distributed only along the terminal branches. Episodic diversifying selection could have affected *Tlr*4

throughout its evolution and this process could still be in operating, while in *Tlr*7 diversifying selection seemed to have appeared more recently and the gene history was mostly maintained by the stronger purifying selection (Figure 3).

Analysis of the Ligand binding region

In TLR4 variants we found 22 rapidly evolving positions distributed all over the LBR. While TLR4 is able to detect several ligands, the most studied one is LPS of Gram negative bacteria. TLR4 does not interact with LPS alone directly but forms stable heterodimers with MD-2 [80]. Analysis of the crystallographic structure of mouse TLR4-MD-2-ligand complex has shown that the interactions between TLR4, -LPS and MD-2 take place on the concave surface of TLR4 [80]. We predicted that sites involved in the TLR4-MD-2 interaction should be highly conserved to maintain the receptor function in LPS binding and these sites were thus not identified in the present study. Among the eight known LPS-binding sites, identified by Park et al. [39] in humans, two residues (F438 and F461) were conserved between humans and rodents as well as among rodents. These key residues are jointly involved also in hydrophobic interactions between TLR4 and MD-2 [39,81]. It is possible that negative selection might maintain an invariable combination at these sites to preserve MD-2 binding, which supports our hypothesis mentioned above. One exception was the controversial site L442Y which was suggested by Park et al. [39] to be also involved in hydrophobic interactions between TLR4 and MD-2, but Resman et al. [81] challenged the importance of its function. Among the studied rodents this codon was found to be polymorphic and has been shown to be affected by episodic positive selection during rodent evolution. A hydrophobic nonpolar residue (Leucine, L) was commonly shared between rodent species except for *Maxomys surifer* that harbored a hydrophobic and polar Tyrosine (Y). For three LPS-binding sites, R263K, K360R and K434R, the biochemical features of the residue were maintained between rodents (all were positively charged residues) but distinct amino acids were detected. The important role of these residues was supported also by Ohto et al. [82] and the potential functional importance of substitution R263K was beside confirmed by conservation analysis. Finally, we have identified in TLR4 two ligand binding positions, 339 and 386, with important amino acid substitutions that might be responsible for variability in LPS binding. No signature of positive selection was detected for these sites; however functional importance of position 386 was supported by the CONSURF analysis. Intriguingly, both residues form charge interactions with the same lipid A phosphate of the LPS, which might indicate that the evolution of this position is associated with phosphate binding. However, this interpretation must be taken cautiously since Resman et al. [81] have questioned the role of the site 386 (in human K388) in LPS binding.

LBR_{TLR7} sequence was much shorter than LBR_{TLR4} one (103 vs. 222 codons, respectively), which could be explained by the smaller size of LBR_{TLR7} ligand, the viral ssRNA [68]. LBR_{TLR7} was highly conserved at the inter-specific level. Only two rapidly evolving positions (out of 103 analysed sites) were detected and neither of them corresponded to the predicted ligand binding residues [68]. Generally the conserved sites (sites evolving under negative selection), have important evolutionary roles for example in protein-protein interactions (TIR domain) or in the preservation of protein structure (*e.g.* LRR forming horseshoe structure).

We found that structural variation between rodent LBR of both TLRs (TLR4 - 1.5Å and TLR7 - 1.7Å) was comparable with the variation observed between ECD_{TLR4} of human and mouse (1.7Å). The 3D-protein structure modeling revealed that LBR_{TLR4} differed between *Rattus sakeratensis*, *R. nitidus* and all other rodent species. The analysis of LBR_{TLR4} sequences did not reveal any specific or unique substitution that could be responsible for this clustering. The same analysis performed on LBR_{TLR7} revealed that *Rattus exulans* substantially differed from other species. This difference could be explained by substitutions found at position H516Y, one being specific of *R. exulans* (Y at position 516) while other *Rattus* and *Mus* species harbored an H amino acid at this position. These inter-specific differences in LBR 3D structure were not related to the phylogenetic distance between species. They could be better explained by similar pathogen exposition and thus similar pathogen-mediated selection.

The results of charge analyses might be more important as they revealed interspecific variation in LBRs of both receptors. *Mus* species had generally a more negative overall charge at LBR than *Rattus* species (Additional file 1: Figure S6). Differences in protein charges were previously shown to be associated with differences in protein-ligand interactions [41,65]. Likewise, differences between these two groups were also found in LBR_{TLR4} at positions that directly bind to LPS. However, some caution is needed, since variation of TLR4 and TLR7 in sensitivity to LPS or ssRNA, respectively, between rats and mice has not been investigated.

Differences in evolution of bacterial-sensing and viral-sensing Tlrs

Our results showed that the bacterial-sensing *Tlr*4 was more variable than the viral-sensing *Tlr*7, and that *Tlr*4 evolution was more intensively shaped by positive selection than in *Tlr*7. *Tlr*4 had 1.7% of codons under positive selection, while in *Tlr*7 it was only 0.38%. These differences are likely to be explained by *Tlrs*' specificity

to different groups of MAMPs with which they co-evolved [56]. *Tlr*4 detects more types of ligands (*e.g.* bacterial LPS, envelope viral components, fungal cell wall components – Mannan) [30] and it seems that these pathogen structures have exerted more diversifying selective pressures on *Tlr*4 than the viral ssRNA affecting *Tlr*7. Recent studies of parasites show that there is an important structural variability in MAMPs between bacterial species (*e.g.* flagellin and LPS) [44,81,83-87]. We propose that the ligand binding region of *Tlr*4 detecting these MAMPs should reflect higher ligand variability observed in our data.

Reduced genetic variability in important genes generally results from strong purifying selection acting against deleterious mutations in these genes [88]. It can result in a smaller effective population size and a lower amount of incomplete lineage sorting [72,89]. These two phenomena were found to be more pronounced when analysing *Tlr*7 phylogeny. Moreover the *Tlr*7 gene is located on the X chromosome in mammals, which can be advantageous during evolution (*e.g.* lower polymorphism is maintained by quicker fixation of beneficial mutations and elimination of deleterious ones by stronger selection and more intense genetic drift) [90]. We suggest that the tension between diversifying and purifying selection, caused by adaptation to the variability of viral motifs detected by viral-sensing *Tlr*7 and maintenance of function together played an important role in the distribution of *Tlr*7 polymorphisms.

Conclusion

This study brings a unique insight into the natural variability and molecular history of two Toll-like receptors in free-living populations of 23 murine species. Purifying selection seems to be the dominant evolutionary force shaping *Tlr*4 and *Tlr*7 polymorphism. However, specific sites putatively evolving under diversifying selection were detected in both *Tlr*s. These sites accumulated within *Tlr*4 LBR, and detailed analyses revealed that several important amino-acid substitutions might alter LPS binding. These substitutions were often species-specific and differentiated between the Rattini and Murini tribes. Interspecific charge variability of LBR and to lesser extent the variability in 3D structure indicated the potential differences in protein-ligand interaction. By contrast, the evolution of *Tlr*7 was strongly shaped by purifying selection. All predicted ligand binding residues in this receptor were uniform across all studied mammals to date. The contrasting evolutionary histories of these two *Tlr*s are likely to result from different structural variability of ligands they target. Since the crystallography of certain ligands (*e.g.* biglycans, hyaluronans and heparin sulphates, ssRNA) [44,68] remains unknown and the precise positions of corresponding binding sites are still missing, our data provide important avenues towards understanding which codons might be candidates for ligand binding residues.

Methods

Sampling

Murine rodents from 23 species belonging to the Rattini and Murini (sensu Lecompte et al. [91]) tribes were sampled mainly in South-East Asia, and three synanthropic species (*i.e. Rattus rattus*, *Mus m. musculus* and *Mus m. domesticus*) were also sampled in Europe and Africa. In our sampling area, *Rattus tanezumi* specimens corresponded to two divergent mitochondrial lineages although they could not be distinguished according to their nuclear pool [73]. These samples were further referred to clades *R. tanezumi R2* and *R3* according to their mitotype. *Rattus sakeratensis* corresponds to the lineage previously referred to as *R. losea* and found in central, northern Thailand and Vientiane Plain of Lao PDR (*Rattus losea*-like by Pagès et al. [69]). This lineage was recently distinguished from the true *R. losea*, which is restricted to Cambodia, Vietnam, China and Taiwan [70].

Species identification was initially based on morphological criteria and thereafter confirmed using molecular barcoding for problematic lineages [69,92]. We sequenced two to 10 individuals per species. In total 103 specimens were analysed (Additional file 1: Table S1).

Toll-like receptor sequencing and sequence alignments

We sequenced the complete exon 3 of *Tlr*4 (2.250 bp) and *Tlr*7 (3.150 bp) as it encompasses the LBR in both genes. Exon 3 corresponds to 89.7% and 99.0% of the total coding sequence for *Tlr*4 and *Tlr*7, respectively. Short exons 1 and 2 (241 bp encoding 5´- untranslated (UT) region and first 257 bp of ECD in $Tlr4_{exon2}$ and 154 bp of 5´-UT regions and 3bp of ECD in $Tlr7_{exon2}$) were not analysed in present study, because we were preferentially interested by functional regions (*e.g.* LBR and TIR). For all analyses and discussion the codon numbering follows the sequences of *Rattus norvegicus* available in GenBank [GenBank Acc. NP_062051.1 for *Tlr*4, and NP_001091051.1, for *Tlr*7].

Primers for Polymerase Chain Reaction (PCR) and sequencing were designed according to the sequences available in the Ensembl database for *Mus musculus* [*Tlr*4 ENSMUSE00000354724/MGI:96824, *Tlr*7 ENSMUSE00000405820/ MGI:2176882] and *Rattus norvegicus* [*Tlr*4 ENSRNOE00000099045/NP_062051, *Tlr*7 ENSRNOE00000039897/NP_001091051]. We used the software PRIMER3 [93] to design primers (see their sequences in Additional file 1: Table S2 and positions in Additional file 1: Figure S1). Total DNA was extracted from rodent tissue (biopsy from ear or necropsy from liver) using the DNeasy Blood & Tissue Kit (Qiagen AB, Hilden, Germany). Amplifications were carried out in a final volume of 25 μl

containing 12.5 μl of Multiplex Kit PCR master mix (Qiagen), 9.3 μl of H_2O, 0.5 μM of each of primer pairs and 2 μl of DNA. Cycling conditions included an initial denaturation at 95°C for 15 min, followed by 10 cycles of denaturation at 95°C for 40 s, annealing with touchdown at 65°C to 55°C (-1°C/cycle) for 45 s and extension at 72°C for 90 s, followed by 30 cycles of denaturation at 95°C for 40 s, annealing at 55°C for 45 s and extension at 72°C for 90 s, with a final extension phase at 72°C for 10 min. The final extension was performed for 10 min at 72°C. The lengths of amplicons were checked on 1.5% agarose gels. Sequencing was carried out using an ABI3130 automated DNA sequencer (Applied Biosystems). DNA sequences were aligned and edited using SEQSCAPE v.2.5 (Applied Biosystems) and BIOEDIT v.7.1.3 (Hall 1999). All sequences have been submitted to NCBI GenBank (Accession numbers are presented in Additional file 1: Table S1).

Sequence analysis

Diploid genotypes were resolved using the Bayesian PHASE platform [94] implemented in DnaSP ver. 5.10 [95]. Calculations were carried out using 1000 iterations, 10 thinning intervals, and 1000 burn-in iterations. Sequences were collapsed into individual alleles by Fabox DNA collapser, an online FASTA sequence toolbox [96]. The identification and visualization of main domains (ECD, TM and ICD with TIR domain and ICD-DP) was performed in SMART [97] based on *Rattus norvegicus* sequences provided in GenBank [NP_062051.1 for *Tlr*4 and NP_001091051.1 for *Tlr*7]. 3D structure was predicted in PHYRE2 [98] and then visualized using FirstGlance in Jmol v.1.9. Finally, we estimated nucleotide diversity (π), number of polymorphic sites (S) and total number of mutations (ε) with DnaSP, and the number of nucleotide alleles (*hN*) and amino acid variants (*hA*) using Fabox DNA collapser.

Phylogenetic reconstructions and congruence between the tree based on a comparably sized non-immune sequence dataset and Tlr trees

We first tested *Tlr* sequences for recombination using SBP, to avoid further false positive events of selection. This method (implemented in DATAMONKEY, [66,99]) allowed the screening of *Tlr* sequences for recombination breakpoints. SBP identify non-recombinant regions and allowed each region to have its own phylogenetic reconstruction [100,101].

Phylogenies were reconstructed independently for each gene using the alignment of complete exon 3 sequences. A phylogeny inferred from the combination of one nuclear (the first exon of the gene encoding the interphotoreceptor retinoid binding protein, *Irbp*) and two mitochondrial genes (the cytochrome *b* gene, *Cytb*, and the cytochrome *c* oxidase I gene, *Co*I), taken from Pagès et al. [69], was used for comparison of "neutral" evolution of the studied rodents with trees obtained from the immune gene alignments. Both Maximum likelihood (ML) and Bayesian (BA) methods were applied to infer phylogenetic relationships from each *Tlr* alignments. The best evolutionary model of nucleotide substitution was determined using jModelTest 0.1.1 [102]. Phylogenies based on ML analyses were reconstructed using RAxML 7.2.6 [103]. Analyses were run as the rapid bootstrap procedure (option –f a) with bootstraps defined by option –NautoMR. For both *Tlr*s we used nucleotide substitution model GTR + Γ (option –m GTRGAMMA) selected by jModelTest 0.1.1 as the most appropriate to our data. Bayesian analyses were performed using a parallel version of MrBayes v3.1 [104] at the University of Oslo Bioportal [105] and CBGP HPC computational platform located at Centre de Biologie et Gestion des Populations, Montpellier. Two runs of 50,000,000 generations in each were adopted, applying the best fitted model of substitution (GTR+ Γ). A burn-in period of 10,000,000 generations was determined using Tracer 1.4 [106]. Convergence was also evaluated using Tracer v1.4. After discarding samples from the burnin period, results were based on the pooled samples from the stationary phases of the two independent runs. Trees were edited using FigTree v1.3.1. [107].

We tested the congruence between the rodent phylogeny and the *Tlr*s phylogeny based on the MrBayes approach using reconciliation analyses. Reconciliation analyses explore all possible mappings of one tree onto another, assigning different costs to evolutionary events and find optimal (*i.e.* yielding minimal costs) solutions. These analyses were conducted using JANE 4 [108]. This software was initially built to reconcile parasite and host trees, yet it can also be used for comparative analysis of species and gene trees. In the context of host-parasite relationships, five evolutionary events between parasites and host can be taken into account in JANE 4: co-speciation, host switches, duplication, failure to diverge and parasite loss. These events are analogous to co-divergence, convergence, duplication, purifying selection and gene loss (respectively) when considered in the context of species and gene tree reconciliation. For each of these events the specific costs can be set. The lowest cost is attributed to the event considered as most likely. In order to obtain reconciliations that maximize the number of co-divergences we set the cost of a co-divergence event to 0 while other costs were set to 1 (see Cruaud et al. [109] for similar approach). The cost of the best solution is then compared with costs found in reconciliations in which tip mappings are permuted at random. This generates a null distribution of the costs of reconciliation. If the cost of the best solution is lower

than that expected by the chance it means that the two phylogenies are significantly congruent. The following parameters were used: the number of generations (iterations of the algorithm) was set to 100 and the "population" (number of samples per generation) was set to 100. Input phylogenies were those obtained by the Bayesian inference. The cost of the best solution was compared to distribution of the costs of 1000 randomizations.

Moreover, we tested the congruence between genes and tree based on a comparably sized non-immune sequence dataset using SH test [110] as implemented in PAUP. Alternative topologies required for ML SH test were reconstructed by ML approach in the software GARLI v. 2.0 [88]. Two different ML trees were estimated for each *Tlr*; a first one inferred under non-constrained conditions with default options and a second one constrained by the tree topology based on a comparably sized non-immune sequence dataset. Mouse species (genus *Mus*) were excluded from the analysis of co-divergence in order to match data with the study of Pagès et al. [69] where the mice are missing.

Search for signatures of selection on Tlr sequences

We estimated separately the number of synonymous (*dS*) and non-synonymous (*dN*) substitutions per site for the whole exon 3, ECD, LBR and the TIR domains, and for both *Tlr*s. Computations were made with 1000 bootstraps and Nei-Gojobori method (with Jukes-Cantor correction) in MEGA 5 [111]. We then estimated the overall ratio *dN*/*dS* for each domain and for the whole exon 3 of both *Tlr*s by Single Likelihood Ancestor Counting (SLAC) implemented in DATAMONKEY. The p-value was 0.05. As the SLAC method tends to be a very conservative test, the actual rate of false positives (*i.e.* neutrally evolving sites incorrectly classified as selected) can be much lower than the significance level [67]. In the next step we estimated selection at each codon by SLAC to find which codons of the exons 3 have been subject to positive and negative selection. As a default tree we used a NJ tree and appropriate substitution model proposed by automatic model selection tool in DATAMONKEY.

Finally, we used the Mixed Effects Model of Evolution (MEME) algorithm in the HYPHY package accessed on the website of DATAMONKEY interface [99] to detect codons evolved under positive selection along the branches of the phylogenies. This method is recently recommended as a replacement for the Fixed Effects Likelihood (FEL) and SLAC models [67]. It allows the detection of signatures of episodic selection, even when the majority of lineages are subject to purifying selection. This test permits ω to vary from site to site and also from branch to branch in phylogeny [67]. Tests of episodic diversifying selection were performed at significance level $p < 0.05$ and MrBayes trees were used as working topologies. Only events of positive selection with Empirical Bayes Factor (EBF) estimated by MEME near to 100 were mapped on to the phylogeny.

Functional analysis of ligand binding region

Positions of LBR in both TLRs have been previously described in humans [39,68]. The corresponding LBR position in rodents was predicted based on the human-rodent alignment. The LBR was located between codons AA248 and AA469 in TLR4 and between codons AA495 and AA597 in TLR7.

We first explored the evolutionary conservation of each amino acid position in LBR using the CONSURF algorithm [112]. CONSURF estimates the evolutionary rate of amino acid positions in a protein molecule, based on the phylogenetic relationships between homologous sequences. Conservation scale is defined from the most variable amino acid positions (grade 1, color represented by turquoise) which are considered as rapidly evolving to conservative positions (grade 9, color represented by maroon) which are considered as slowly evolving. We used the proposed substitution matrix and computation was based on the empirical Bayesian paradigm. MrBayes trees were used as the working topology. Protein tertiary structure was adopted from *R. norvegicus* [Gene Bank Acc. TLR4/KC811688 and TLR7/KC811786].

Because protein tertiary structure is essential for its biological function we finally explored the variability in the 3D structures of LBRs in the different AA variants. The prediction of 3D structures of the variants was performed by homology modeling using PHYRE2 [98]. Differences in 3D protein structure among variants were then evaluated using the root mean square deviations (RMSD) calculated by the DALI pairwise comparison tool [113]. The RMSD-based distance matrices were analysed in STATISTICA v. 8.0 (StatSoft, Inc., Tulsa) by joining tree clustering using Unweighted Pair Group Method with Arithmetic Mean (UPGMA, [114]). We then analysed the variability of the charge of each LBR variant, which could be another key indicator of functional changes, because differences in protein charge could influence the ability to bind ligands [41,65]. LBR charge of each variant was estimated at predefined neutral pH = 7 using LRRFINDER [115].

Availability of supporting data section

All sequences have been submitted to NCBI GenBank under Accession numbers from KC811609 to KC811800 (Individual accession numbers are presented in Additional file 1: Table S1). *Tlr* phylogenies based on MrBayes (Tlr4_MrBayes_final.nex, Tlr7_MrBayes_final.nex) and RAxML (Tlr4_RAxML_final.nex, Tlr7_RAxML_final.nex) approach were added to the TreeBase database

(http://treebase.org/treebase-web/home.html). Trees are available at URL: http://purl.org/phylo/treebase/phylows/study/TB2:S14659.

Additional files

Additional file 1: Table S1. Summary of sampled specimens and identification of haplotypes. **Table S2.** Primer description. **Table S3.** Residues binding to LPS in TLR4 based on knowledge of 3D-crystalography in human predicted by Park et al. 2009. **Table S4.** Potential residues binding ssRNA predicted by Wei et al. 2009. **Figure S1.** Protein structure of TLR4 (a, c) and TLR7 (b, d) identified by SMART (http://smart.embl-heidelberg.de/) (a, b) and CONSURF (c, d). SMART (a, b) identified following types of domains: **LRR** - Leucine rich repeat; **LRRCT** - Leucine rich repeat C-terminal domain; **TIR** - TIR domain, **Fulfilled blue box (TD)** - transmembrane domain; **LRRNT** - Leucine rich repeat N-terminal domain. **Red box** - LBR (from AA248 to AA469 for TLR4 and from AA495 to AA597 for TLR7). **ECD** - extracellular domain is represented by solid black double arrow; **ICD** - intracellular domain is represented by dashed double arrow. Distal part of ICD **(ICD-DP)** is indicated by a simple solid arrow. Positions of forward and reverse primers used for amplification are shown by arrows. Arrows of same color indicates primer pairs. Description of crystallographic structure (c, d) **LBR** is represented by red polygon; **TD** is present between two dashed lines. To the right from TD is **ICD**, to the left is **ECD**. **Figures S4.** Test of congruence between the presumably neutral and *Tlr* phylogenies (*Tlr4* (a), *Tlr7* (b) following JANE 4). Number at X axis represents costs of co-divergence. The red dashed line represents the cost observed in our data. The blue columns represent the random distributions of costs. Lower cost than random observed in our data signified higher congruence between species and gene topologies. **Figure S5.** Superimposition of structures, tree clustering diagrams based on linkage distance, (a) LBRTLR4 and (b) LBRTLR7; individual LBR-variants often unify more species; description of LBR-variants labels is in the Table S1 under Hap_LBRTLR4 and Hap_LBRTLR7. **Figure S6.** Analysis of LBR amino acid sequence charge at pH 7 (LRRFinder) for (a) LBRTLR4 and (b) LBRTLR7, individual LBR-variants often unify more species; description of LBR-variants labels is in the Table S1 under Hap_LBRTLR4 and Hap_LBRTLR7. Mouse species are in red, *Rattus* spp. and related genera are in blue.

Additional file 2: Figures S2 and S3. (Phylogenetic trees).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

Conceived and designed the experiments: AF JFC JB NCH MV. Performed the sequencing: AF MG FC. Analysed the data: AF MV MP EJ. Contributed samples: SM JFC AF. Wrote the paper: AF MV JFC JB NCH MP EJ (sorted by the significance of contributions). All authors read and approved the final manuscript.

Acknowledgments

This work was supported by the French National Agency for Research projects CERoPath (grant number 00121 0505, 07 BDIV 012) http://www.ceropath.org/ and BioDivHealthSEA (grant number ANR 11 CPEL 002), and the Czech Science Foundation (grant number 206/08/0640). Cooperation on this project was also partly supported by bilateral project BARRANDE (grant number MEB021130/24504WM). The thesis of A. Fornůsková was partly funded by a three year French government fellowship and the fellowship from Masaryk University. MP is currently funded by an FRS - FNRS fellowship (Belgian Fund for Scientific Research).We are grateful to Anna Bryjová, Yannick Chaval, Gael Kergoat, Marian Novotný, Sylvain Piry, Lucie Vlčková for their help during various stages of the manuscript preparation and to Jamie Caroline Winternitz for language corrections. We also thank to the CBGP HPC computational platform and to the Centre Méditerranéen Environnement Biodiversité.

Author details

[1]Institute of Vertebrate Biology, Research Facility Studenec, Academy of Sciences, Prague, Czech Republic. [2]Department of Botany and Zoology, Faculty of Science, Masaryk University, Brno, Czech Republic. [3]INRA, UMR CBGP (INRA/IRD/Cirad/Montpellier SupAgro), Campus International de Baillarguet, CS 30016, 34988 Montferrier-sur-Lez Cedex, France. [4]Department of Zoology, Faculty of Science, Charles University in Prague, Prague, Czech Republic. [5]Laboratoire de génétique des microorganismes, Université de Liège, 4000 Liège, Belgique. [6]Labex CeMEB, Plateforme Génotypage-Séquençage, Université Montpellier2, Montpellier, France. [7]ISEM, Montpellier, France. [8]CIRAD, UR AGIRs, Montpellier, France.

Received: 11 July 2013 Accepted: 6 September 2013
Published: 12 September 2013

References

1. Zak DE, Aderem A: **Systems biology of innate immunity.** *Immunol Rev* 2009, **227:**264–282.
2. Barreiro LB, Quintana-Murci L: **From evolutionary genetics to human immunology: how selection shapes host defence genes.** *Nat Rev Genet* 2010, **11:**17–30.
3. Akira S, Uematsu S, Takeuchi O: **Pathogen recognition and innate immunity.** *Cell* 2006, **124:**783–801.
4. Schröder NWJ, Schumann RR: **Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease.** *Lancet Infect Dis* 2005, **5:**156–164.
5. Pandey S, Agrawal DK: **Immunobiology of Toll-like receptors: emerging trends.** *Immunol Cell Biol* 2006, **84:**333–341.
6. Bochud P-Y, Bochud M, Telenti A, Calandra T: **Innate immunogenetics: a tool for exploring new frontiers of host defence.** *Lancet Infect Dis* 2007, **7:**531–542.
7. Loo Y-M, Gale M Jr: **Immune signaling by RIG-I-like receptors.** *Immunity* 2011, **34:**680–692.
8. Netea MG, Wijmenga C, O'Neill LAJ: **Genetic variation in Toll-like receptors and disease susceptibility.** *Nat Immunol* 2012, **13:**535–542.
9. Wlasiuk G, Nachman MW: **Adaptation and constraint at Toll-like receptors in primates.** *Mol Biol Evol* 2010, **27:**2172–2186.
10. Alcaide M, Edwards SV: **Molecular evolution of the Toll-like receptor multigene family in birds.** *Mol Biol Evol* 2011, **28:**1703–1715.
11. Tschirren B, Råberg L, Westerdahl H: **Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents.** *J Evol Biol* 2011, **24:**1232–1240.
12. Grueber CE, Wallis GP, King TM, Jamieson IG: **Variation at innate immunity Toll-like receptor genes in a bottlenecked population of a New Zealand robin.** *PLoS ONE* 2012, **7:**e45011.
13. Tschirren B, Andersson M, Scherman K, Westerdahl H, Råberg L: **Contrasting patterns of diversity and population differentiation at the innate immunity gene Toll-like receptor 2 (TLR2) in two sympatric rodent species.** *Evolution* 2012, **66:**720–731.
14. Tschirren B, Andersson M, Scherman K, Westerdahl H, Mittl PRE, Råberg L: **Polymorphisms at the innate immune receptor TLR2 are associated with Borrelia infection in a wild rodent population.** *Proc Biol Sci* 2013, **280:**20130364.
15. Haldane JBS: **Malaria: disease and evolution.** In *Genetic and Evolutionary Aspects.* Boston: Kluwer Academic Publishers; 2006:175–187.
16. Apanius V, Penn D, Slev PR, Ruff LR, Potts WK: **The nature of selection on the major histocompatibility complex.** *Crit Rev Immunol* 1997, **17:**179–224.
17. Bernatchez L, Landry C: **MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years?** *J Evol Biol* 2003, **16:**363–377.
18. Aguilar A, Roemer G, Debenham S, Binns M, Garcelon D, Wayne RK: **High MHC diversity maintained by balancing selection in an otherwise genetically monomorphic mammal.** *Proc Natl Acad Sci USA* 2004, **101:**3490–3494.
19. Bryja J, Galan M, Charbonnel N, Cosson JF: **Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae).** *Immunogenetics* 2006, **58:**191–202.
20. Piertney SB, Oliver MK: **The evolutionary ecology of the major histocompatibility complex.** *Heredity (Edinb)* 2006, **96:**7–21.

21. Spurgin LG, Richardson DS: **How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings.** *Proc Biol Sci* 2010, **277:**979–988.
22. Cížková D, Gouy de Bellocq J, Baird SJE, Piálek J, Bryja J: **Genetic structure and contrasting selection pattern at two major histocompatibility complex genes in wild house mouse populations.** *Heredity (Edinb)* 2011, **106:**727–740.
23. Smith C, Ondračková M, Spence R, Adams S, Betts DS, Mallon E: **Pathogen-mediated selection for MHC variability in wild zebrafish.** *Evol Ecol Res* 2011, **67:**217–218.
24. Medzhitov R, Preston-Hurlburt P, Janeway CA Jr: **A human homologue of the Drosophila Toll protein signals activation of adaptive immunity.** *Nature* 1997, **388:**394–397.
25. Hedrick SM: **The acquired immune system: a vantage from beneath.** *Immunity* 2004, **21:**607–615.
26. O'Neill LAJ: **TLRs: Professor Mechnikov, sit on your hat.** *Trends Immunol* 2004, **25:**687–693.
27. Bassett EH, Rich T: **Introduction.** In *Toll and Toll-Like Receptors: An Immunologic Perspective.* Boston, MA: Springer US; 2005:1–17.
28. Acevedo-Whitehouse K, Cunningham AA: **Is MHC enough for understanding wildlife immunogenetics?** *Trends Ecol Evol (Amst)* 2006, **21:**433–438.
29. Barreiro LB, Ben-Ali M, Quach H, Laval G, Patin E, Pickrell JK, Bouchier C, Tichit M, Neyrolles O, Gicquel B, Kidd JR, Kidd KK, Alcaïs A, Ragimbeau J, Pellegrini S, Abel L, Casanova J-L, Quintana-Murci L: **Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense.** *PLoS Genet* 2009, **5:**e1000562.
30. Vinkler M, Albrecht T: **The question waiting to be asked: innate immunity receptors in the perspective of zoological research.** *Folia Zool* 2009, **58:**15–28.
31. Janssens S, Beyaert R: **Role of Toll-like receptors in pathogen recognition.** *Clin Microbiol Rev* 2003, **16:**637–646.
32. Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A: **The evolution of vertebrate Toll-like receptors.** *Proc Natl Acad Sci USA* 2005, **102:**9577–9582.
33. Hughes AL, Piontkivska H: **Functional diversification of the toll-like receptor gene family.** *Immunogenetics* 2008, **60:**249–256.
34. Leulier F, Lemaitre B: **Toll-like receptors–taking an evolutionary approach.** *Nat Rev Genet* 2008, **9:**165–178.
35. Temperley ND, Berlin S, Paton IR, Griffin DK, Burt DW: **Evolution of the chicken Toll-like receptor gene family: a story of gene gain and gene loss.** *BMC Genomics* 2008, **9:**62.
36. Huang Y, Temperley ND, Ren L, Smith J, Li N, Burt DW: **Molecular evolution of the vertebrate TLR1 gene family–a complex history of gene duplication, gene conversion, positive selection and co-evolution.** *BMC Evol Biol* 2011, **11:**149.
37. Werling D, Jann OC, Offord V, Glass EJ, Coffey TJ: **Variation matters: TLR structure and species-specific pathogen recognition.** *Trends Immunol* 2009, **30:**124–130.
38. Burke DF, Worth CL, Priego E-M, Cheng T, Smink LJ, Todd JA, Blundell TL: **Genome bioinformatic analysis of nonsynonymous SNPs.** *BMC Bioinforma* 2007, **8:**301.
39. Park BS, Song DH, Kim HM, Choi B-S, Lee H, Lee J-O: **The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex.** *Nature* 2009, **458:**1191–1195.
40. Keestra AM, van Putten JPM: **Unique properties of the chicken TLR4/MD-2 complex: selective lipopolysaccharide activation of the MyD88-dependent pathway.** *J Immunol* 2008, **181:**4354–4362.
41. Walsh C, Gangloff M, Monie T, Smyth T, Wei B, McKinley TJ, Maskell D, Gay N, Bryant C: **Elucidation of the MD-2/TLR4 interface required for signaling by lipid IVa.** *J Immunol* 2008, **181:**1245–1254.
42. Zhu J, Brownlie R, Liu Q, Babiuk LA, Potter A, Mutwiri GK: **Characterization of bovine Toll-like receptor 8: ligand specificity, signaling essential sites and dimerization.** *Mol Immunol* 2009, **46:**978–990.
43. Botos I, Segal DM, Davies DR: **The structural biology of Toll-like receptors.** *Structure* 2011, **19:**447–459.
44. Kang JY, Lee J-O: **Structural biology of the Toll-like receptor family.** *Annu Rev Biochem* 2011, **80:**917–941.
45. Pasare C, Medzhitov R: **Toll pathway-dependent blockade of CD4+CD25+ T cell-mediated suppression by dendritic cells.** *Science* 2003, **299:**1033–1036.
46. Parker LC, Prince LR, Sabroe I: **Translational mini-review series on Toll-like receptors: networks regulated by Toll-like receptors mediate innate and adaptive immunity.** *Clin Exp Immunol* 2007, **147:**199–207.
47. Kawai T, Akira S: **The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors.** *Nat Immunol* 2010, **11:**373–384.
48. Pasare C, Medzhitov R: **Toll-like receptors and acquired immunity.** *Semin Immunol* 2004, **16:**23–26.
49. Netea MG, Ferwerda G, de Jong DJ, Jansen T, Jacobs L, Kramer M, Naber THJ, Drenth JPH, Girardin SE, Kullberg BJ, Adema GJ, Van der Meer JWM: **Nucleotide-binding oligomerization domain-2 modulates specific TLR pathways for the induction of cytokine release.** *J Immunol* 2005, **174:**6518–6523.
50. Smirnova I, Poltorak A, Chan EK, McBride C, Beutler B: **Phylogenetic variation and polymorphism at the Toll-like receptor 4 locus (TLR4).** *Genome Biol* 2000, **1:**002.1–002.10.
51. Ferwerda B, McCall MBB, Alonso S, Giamarellos-Bourboulis EJ, Mouktaroudi M, Izagirre N, Syafruddin D, Kibiki G, Cristea T, Hijmans A, Hamann L, Israel S, ElGhazali G, Troye-Blomberg M, Kumpf O, Maiga B, Dolo A, Doumbo O, Hermsen CC, Stalenhoef AFH, van Crevel R, Brunner HG, Oh D-Y, Schumann RR, de la Rúa C, Sauerwein R, Kullberg B-J, van der Ven AJAM, van der Meer JWM, Netea MG: **TLR4 polymorphisms, infectious diseases, and evolutionary pressure during migration of modern humans.** *Proc Natl Acad Sci USA* 2007, **104:**16645–16650.
52. Vinkler M, Bryjová A, Albrecht T, Bryja J: **Identification of the first Toll-like receptor gene in passerine birds: TLR4 orthologue in zebra finch (Taeniopygia guttata).** *Tissue Antigens* 2009, **74:**32–41.
53. Krieg AM, Vollmer J: **Toll-like receptors 7, 8, and 9: linking innate immunity to autoimmunity.** *Immunol Rev* 2007, **220:**251–269.
54. Barrat FJ, Coffman RL: **Development of TLR inhibitors for the treatment of autoimmune diseases.** *Immunol Rev* 2008, **223:**271–283.
55. Waldner H: **The role of innate immune responses in autoimmune disease development.** *Autoimmun Rev* 2009, **8:**400–404.
56. Mikami T, Miyashita H, Takatsuka S, Kuroki Y, Matsushima N: **Molecular evolution of vertebrate Toll-like receptors: evolutionary rate difference between their leucine-rich repeats and their TIR domains.** *Gene* 2012, **503:**235–243.
57. Worobey M, Bjork A, Wertheim JO: **Point, counterpoint: the evolution of pathogenic viruses and their human hosts.** *Annu Rev Ecol Evol Syst* 2007, **38:**515–540.
58. Poltorak A, He X, Smirnova I, Liu MY, Van Huffel C, Du X, Birdwell D, Alejos E, Silva M, Galanos C, Freudenberg M, Ricciardi-Castagnoli P, Layton B, Beutler B: **Defective LPS signaling in C3H/HeJ and C57BL/10ScCr mice: mutations in Tlr4 gene.** *Science* 1998, **282:**2085–2088.
59. Diebold SS, Kaisho T, Hemmi H, Akira S: **Reis e Sousa C: Innate antiviral responses by means of TLR7-mediated recognition of single-stranded RNA.** *Science* 2004, **303:**1529–1531.
60. Heil F, Hemmi H, Hochrein H, Ampenberger F, Kirschning C, Akira S, Lipford G, Wagner H, Bauer S: **Species-specific recognition of single-stranded RNA via Toll-like receptor 7 and 8.** *Science* 2004, **303:**1526–1529.
61. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P: **Global trends in emerging infectious diseases.** *Nature* 2008, **451:**990–993.
62. Mills JN: **Biodiversity loss and emerging infectious disease: an example from the rodent-borne hemorrhagic fevers.** *Biodiversity* 2006, **7:**9–17.
63. Luis AD, Hayman DTS, O'Shea TJ, Cryan PM, Gilbert AT, Pulliam JRC, Mills JN, Timonin ME, Willis CKR, Cunningham AA, Fooks AR, Rupprecht CE, Wood JLN, Webb CT: **A comparison of bats and rodents as reservoirs of zoonotic viruses: are bats special?** *Proc Biol Sci* 2013, **280:**20122753.
64. Fornarino S, Laval G, Barreiro LB, Manry J, Vasseur E, Quintana-Murci L: **Evolution of the TIR domain-containing adaptors in humans: swinging between constraint and adaptation.** *Mol Biol Evol* 2011, **28:**3087–3097.
65. Govindaraj RG, Manavalan B, Basith S, Choi S: **Comparative analysis of species-specific ligand recognition in Toll-like receptor 8 signaling: a hypothesis.** *PLoS ONE* 2011, **6:**e25118.
66. Pond SLK, Frost SDW: **Datamonkey: rapid detection of selective pressure on individual sites of codon alignments.** *Bioinformatics* 2005, **21:**2531–2533.

67. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K: **Kosakovsky Pond SL: Detecting individual sites subject to episodic diversifying selection.** *PLoS Genet* 2012, **8:**1002764.
68. Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rössle SC: **Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains.** *Protein Sci* 2009, **18:**1684–1691.
69. Pagès M, Chaval Y, Herbreteau V, Waengsothorn S, Cosson J-F, Hugot J-P, Morand S, Michaux J: **Revisiting the taxonomy of the Rattini tribe: a phylogeny-based delimitation of species boundaries.** *BMC Evol Biol* 2010, **10:**184.
70. Aplin KP, Suzuki H, Chinen AA, Chesser RT, Ten Have J, Donnellan SC, Austin J, Frost A, Gonzalez JP, Herbreteau V, Catzeflis F, Soubrier J, Fang Y-P, Robins J, Matisoo-Smith E, Bastos ADS, Maryanto I, Sinaga MH, Denys C, Van Den Bussche RA, Conroy C, Rowe K, Cooper A: **Multiple geographic origins of commensalism and complex dispersal history of Black Rats.** *PLoS ONE* 2011, **6:**e26357.
71. Moore WS: **Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees.** *Evolution* 1995, **49:**718–726.
72. Hobolth A, Dutheil JY, Hawks J, Schierup MH, Mailund T: **Incomplete lineage sorting patterns among human, chimpanzee, and orangutan suggest recent orangutan speciation and widespread selection.** *Genome Res* 2011, **21:**349–356.
73. Pagès M, Bazin E, Galan M, Chaval Y, Claude J, Herbreteau V, Michaux J, Piry S, Morand S, Cosson J-F: **Cytonuclear discordance among Southeast Asian black rats (*Rattus rattus* complex).** *Mol Ecol* 2013, **22:**1019–1034.
74. Lack JB, Greene DU, Conroy CJ, Hamilton MJ, Braun JK, Mares MA, Van Den Bussche RA: **Invasion facilitates hybridization with introgression in the *Rattus rattus* species complex.** *Mol Ecol* 2012, **21:**3545–3561.
75. Nichols R: **Gene trees and species trees are not the same.** *Trends Ecol Evol* 2001, **16:**358–364.
76. Edwards SV: **Natural selection and phylogenetic analysis.** *PNAS* 2009, **106:**8799–8800.
77. Areal H, Abrantes J, Esteves PJ: **Signatures of positive selection in Toll-like receptor (TLR) genes in mammals.** *BMC Evol Biol* 2011, **11:**368.
78. Smith SA, Jann OC, Haig D, Russell GC, Werling D, Glass EJ, Emes RD: **Adaptive evolution of Toll-like receptor 5 in domesticated mammals.** *BMC Evol Biol* 2012, **12:**122.
79. Downing T, Lloyd AT, O'Farrelly C, Bradley DG: **The differential evolutionary dynamics of avian cytokine and TLR gene classes.** *J Immunol* 2010, **184:**6993–7000.
80. Kim HM, Park BS, Kim J-I, Kim SE, Lee J, Oh SC, Enkhbayar P, Matsushima N, Lee H, Yoo OJ, Lee J-O: **Crystal structure of the TLR4-MD-2 complex with bound endotoxin antagonist Eritoran.** *Cell* 2007, **130:**906–917.
81. Resman N, Vasl J, Oblak A, Pristovsek P, Gioannini TL, Weiss JP, Jerala R: **Essential roles of hydrophobic residues in both MD-2 and Toll-like receptor 4 in activation by endotoxin.** *J Biol Chem* 2009, **284:**15052–15060.
82. Ohto U, Fukase K, Miyake K, Shimizu T: **Structural basis of species-specific endotoxin sensing by innate immune receptor TLR4/MD-2.** *Proc Natl Acad Sci USA* 2012, **109:**7421–7426.
83. Raetz CRH, Whitfield C: **Lipopolysaccharide endotoxins.** *Annu Rev Biochem* 2002, **71:**635–700.
84. van der Woude MW, Bäumler AJ: **Phase and antigenic variation in bacteria.** *Clin Microbiol Rev* 2004, **17:**581–611.
85. Andersen-Nissen E, Smith KD, Strobe KL, Barrett SLR, Cookson BT, Logan SM, Aderem A: **Evasion of Toll-like receptor 5 by flagellated bacteria.** *Proc Natl Acad Sci USA* 2005, **102:**9247–9252.
86. Sun W, Dunning FM, Pfund C, Weingarten R, Bent AF: **Within-species flagellin polymorphism in *Xanthomonas campestris pv campestris* and its impact on elicitation of Arabidopsis flagellin sensinG2-dependent defenses.** *Plant Cell* 2006, **18:**764–779.
87. Maeshima N, Fernandez RC: **Recognition of lipid A variants by the TLR4-MD-2 receptor complex.** *Front Cell Infect Microbiol* 2013, **3**. doi:10.3389/fcimb.2013.00003.
88. Zwickl DJ: *Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion*, The University of Texas at Austin; 2006. Ph.D. dissertation.
89. Charlesworth B, Morgan MT, Charlesworth D: **The effect of deleterious mutations on neutral molecular variation.** *Genetics* 1993, **134:**1289–1303.
90. Salcedo T, Geraldes A, Nachman MW: **Nucleotide variation in wild and inbred mice.** *Genetics* 2007, **177:**2277–2291.
91. Lecompte E, Aplin K, Denys C, Catzeflis F, Chades M, Chevret P: **Phylogeny and biogeography of African Murinae based on mitochondrial and nuclear gene sequences, with a new tribal classification of the subfamily.** *BMC Evol Biol* 2008, **8:**199.
92. Galan M, Pagès M, Cosson J-F: **Next-generation sequencing for rodent barcoding: species identification from fresh, degraded and environmental samples.** *PLoS ONE* 2012, **7:**e48374.
93. Rozen S, Skaletsky H: **Primer3 on the WWW for general users and for biologist programmers.** *Methods Mol Biol* 2000, **132:**365–386.
94. Stephens M, Donnelly P: **A Comparison of bayesian methods for haplotype reconstruction from population genotype data.** *Am J Hum Genet* 2003, **73:**1162–1169.
95. Librado P, Rozas J: **DnaSP v5: a software for comprehensive analysis of DNA polymorphism data.** *Bioinformatics* 2009, **25:**1451–1452.
96. Villesen P: **FaBox: an online toolbox for fasta sequences.** *Mol Ecol Notes* 2007, **7:**965–968.
97. Schultz J, Milpetz F, Bork P, Ponting CP: **SMART, a simple modular architecture research tool: identification of signaling domains.** *Proc Natl Acad Sci USA* 1998, **95:**5857–5864.
98. Kelley LA, Sternberg MJE: **Protein structure prediction on the Web: a case study using the Phyre server.** *Nat Protoc* 2009, **4:**363–371.
99. Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL: **Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology.** *Bioinformatics* 2010, **26:**2455–2457.
100. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW: **Automated phylogenetic detection of recombination using a genetic algorithm.** *Mol Biol Evol* 2006, **23:**1891–1901.
101. Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW: **GARD: a genetic algorithm for recombination detection.** *Bioinformatics* 2006, **22:**3096–3098.
102. Posada D: **jModelTest: phylogenetic model averaging.** *Mol Biol Evol* 2008, **25:**1253–1256.
103. Stamatakis A, Hoover P, Rougemont J: **A rapid bootstrap algorithm for the RAxML web servers.** *Syst Biol* 2008, **57:**758–771.
104. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogenetic trees.** *Bioinformatics* 2001, **17:**754–755.
105. Kumar S, Skjaeveland A, Orr RJS, Enger P, Ruden T, Mevik B-H, Burki F, Botnen A, Shalchian-Tabrizi K: **AIR: a batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses.** *BMC Bioinforma* 2009, **10:**357.
106. Rambaut A, Drummond AJ: *Tracer v1.4*; 2007. Available from http://beast.bio.ed.ac.uk/Tracer.
107. Rambaut A: *FigTree v1.3.1 2006–2009*; 2009. Available with the program package at http://tree.bio.ed.ac.uk/software/figtree.
108. Conow C, Fielder D, Ovadia Y, Libeskind-Hadas R: **Jane: a new tool for the cophylogeny reconstruction problem.** *Algorithms Mol Biol* 2010, **5:**16.
109. Cruaud A, Rønsted N, Chantarasuwan B, Chou LS, Clement WL, Couloux A, Cousins B, Genson G, Harrison RD, Hanson PE, Hossaert-McKey M, Jabbour-Zahab R, Jousselin E, Kerdelhué C, Kjellberg F, Lopez-Vaamonde C, Peebles J, Peng Y-Q, Pereira RAS, Schramm T, Ubaidillah R, van Noort S, Weiblen GD, Yang D-R, Yodpinyanee A, Libeskind-Hadas R, Cook JM, Rasplus J-Y, Savolainen V: **An extreme case of plant-insect codiversification: figs and fig-pollinating wasps.** *Syst Biol* 2012, **61:**1029–1047.
110. Shimodaira H, Hasegawa M: **Multiple comparisons of log-likelihoods with applications to phylogenetic inference.** *Mol Biol Evol* 1999, **16:**1114–1116.
111. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Biol Evol* 2011, **28:**2731–2739.
112. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N: **ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids.** *Nucleic Acids Res* 2010, **38:**W529–533.

113. Holm L, Kääriäinen S, Rosenström P, Schenkel A: **Searching protein structure databases with DaliLite v.3.** *Bioinformatics* 2008, **24:**2780–2781.
114. Kalinowski ST: **How well do evolutionary trees describe genetic relationships among populations?** *Heredity (Edinb)* 2009, **102:**506–513.
115. Offord V, Coffey TJ, Werling D: **LRRfinder: a web application for the identification of leucine-rich repeats and an integrative Toll-like receptor database.** *Dev Comp Immunol* 2010, **34:**1035–1041.

doi:10.1186/1471-2148-13-194
Cite this article as: Fornůsková *et al.*: Contrasted evolutionary histories of two Toll-like receptors (*Tlr4* and *Tlr7*) in wild rodents (MURINAE). *BMC Evolutionary Biology* 2013 **13**:194.

Submit your next manuscript to BioMed Central and take full advantage of:

- **Convenient online submission**
- **Thorough peer review**
- **No space constraints or color figure charges**
- **Immediate publication on acceptance**
- **Inclusion in PubMed, CAS, Scopus and Google Scholar**
- **Research which is freely available for redistribution**

Submit your manuscript at
www.biomedcentral.com/submit

BioMed Central