Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents

Tschirren, Barbara; Raberg, Lars; Westerdahl, Helena

Abstract: Patterns of selection acting on immune defence genes have recently been the focus of considerable interest. Yet, when it comes to vertebrates, studies have mainly focused on the acquired branch of the immune system. Consequently, the direction and strength of selection acting on genes of the vertebrate innate immune defence remain poorly understood. Here, we present a molecular analysis of selection on an important receptor of the innate immune system of vertebrates, the Toll-like receptor 2 (TLR2), across 17 rodent species. Although purifying selection was the prevalent evolutionary force acting on most parts of the rodent TLR2, we found that codons in close proximity to pathogen-binding and TLR2–TLR1 heterodimerization sites have been subject to positive selection. This indicates that parasite-mediated selection is not restricted to acquired immune system genes like the major histocompatibility complex, but also affects innate defence genes. To obtain a comprehensive understanding of evolutionary processes in host–parasite systems, both innate and acquired immunity thus need to be considered.

DOI: https://doi.org/10.1111/j.1420-9101.2011.02254.x

Posted at the Zurich Open Repository and Archive, University of Zurich
ZORA URL: https://doi.org/10.5167/uzh-73244
Journal Article
Accepted Version

Originally published at:
DOI: https://doi.org/10.1111/j.1420-9101.2011.02254.x
Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (TLR2) during the evolutionary history of rodents

Barbara Tschirren*, Lars Råberg & Helena Westerdahl
Molecular Ecology and Evolution Lab, Department of Biology, Lund University,
Sölvegatan 37, 223 62 Lund, Sweden

*Correspondence: Barbara Tschirren, Institute of Evolutionary Biology and
Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich,
Switzerland; Tel: +41 44 635 47 77, Fax: +41 44 635 68 18, Email:
barbara.tschirren@ieu.uzh.ch

Running title: Positive selection on TLR2 in rodents
Abstract

Patterns of selection acting on immune defence genes have recently been the focus of considerable interest. Yet, when it comes to vertebrates, studies have mainly focused on the acquired branch of the immune system. Consequently, the direction and strength of selection acting on genes of the vertebrate innate immune defence remain poorly understood. Here, we present a molecular analysis of selection on an important receptor of the innate immune system of vertebrates, the Toll-like receptor 2 (TLR2), across 17 rodent species. Although purifying selection was the prevalent evolutionary force acting on most parts of the rodent TLR2, we found that codons in close proximity to pathogen-binding and TLR2-TLR1 heterodimerisation sites have been subject to positive selection in rodents. This indicates that parasite-mediated selection is not restricted to acquired immune system genes like the Major histocompatibility complex (MHC), but also affects innate defence genes. To obtain a comprehensive understanding of evolutionary processes in host-parasite systems, both innate and acquired immunity thus need to be considered.

Key words: arms race, coevolution, disease ecology, host-parasite interactions, innate immunity, parasite-mediated selection, rodents, Toll-like receptor 2
Introduction

The strength and direction of parasite-mediated selection acting on hosts can vary in time and space as a result of host-parasite coevolution and/or variation in the composition of the parasite assemblage. As a result, parasites are often assumed to be important drivers of genetic diversification within and among host species (Hedrick, 2002; Woolhouse et al., 2002). In line with this, a number of studies have demonstrated directional selection (i.e. selective sweeps that favour advantageous mutations and thereby results in divergence between species (Hughes, 1999)) or balancing selection (which maintains genetic diversity within species) on immune defence genes in both plants and animals (Tanaka & Nei, 1989; Tiffin & Moeller, 2006; Obbard et al., 2009). In the case of plants and invertebrates, general patterns of selection on immune defence genes are beginning to emerge (Tiffin & Moeller, 2006). For example, a recent, comprehensive analysis of immune defence genes across Drosophila species revealed that pathogen recognition genes, rather than immune signalling or effector genes, are the primary targets of parasite-mediated selection (Sackton et al., 2007). However, when it comes to vertebrates, it is still largely unclear which types of host defence genes are subject to positive selection from parasites (Tiffin & Moeller, 2006).

The vertebrate immune system consists of two parts, the innate and acquired immune defence. The innate branch of the immune system is ancient and has homologous components in both invertebrates and plants (Beutler, 2004). It provides a fast but generally non-specific defence against parasites. The acquired immune system, on the other hand, is characterized by high specificity and memory, and occurs exclusively in vertebrates (Cooper & Alder, 2006). Some acquired immune defence genes have been shown to be subject to strong selective sweeps and/or
balancing selection in a wide range of taxa, with the Major histocompatibility complex (MHC) genes being exceptionally well studied in this context (Apanius et al., 1997; Hughes & Yeager, 1998; Piertney & Oliver, 2006; Wegner, 2008).

In contrast, patterns of selection acting on vertebrate innate immune genes are less clear. As reviewed above, studies of selection on defence genes in plants and invertebrates have shown that innate immunity genes, in particular genes involved in pathogen recognition, can be subject to positive selection (Jiggins & Kim, 2006; Tiffin & Moeller, 2006; Sackton et al., 2007; Obbard et al., 2009). Yet, in vertebrates the presence of a highly specific defence, mediated by the acquired immune system, might affect the role, and consequently the patterns of selection acting on their innate immune defence. Indeed, it is often argued that because innate defence receptors recognize conserved pathogen structures, they should be subject to purifying (i.e. negative) rather than positive selection (Medzhitov & Janeway, 1997; Mukherjee et al., 2009).

One of the best-studied types of innate defence receptors are Toll-like receptors (TLRs) (Vasselon & Detmers, 2002; Takeda et al., 2003). TLRs form an extended gene family, which has evolved by gene duplication (Zhou et al., 2007). The general structure of TLRs is characterized by an extracellular domain, which consists of a number of leucine-rich repeats (LRRs) that are involved in pathogen recognition (Jin et al., 2007), a transmembrane region, and a highly conserved intracellular Toll/Interleukin-1 receptor (TIR) domain (Kobe & Kajava, 2001; West et al., 2006).

After binding to a pathogen, TLRs form homo- or heterodimers and initiate an intracellular signalling cascade, which leads to the activation of transcription factors, the production of pro-inflammatory cytokines and, ultimately, inflammation in the infected host tissue (Medzhitov, 2001; Akira & Takeda, 2004). Most mammals have
10 – 12 different TLRs, each recognizing different ligands (Roach et al., 2005).

TLR2, for example, recognizes lipoproteins from cell walls of bacteria, whereas TLR3 targets double-stranded RNA of viruses (Garantziotis et al., 2008).

Because Toll-like receptors, like MHC molecules, are involved in pathogen recognition, it has been suggested that they might be involved in coevolutionary processes with pathogens (Hughes & Friedman, 2008). Indeed, recent studies on primates have found evidence for positive selection on TLRs (Ferrer-Admetlla et al., 2008; Hughes & Piontkivska, 2008; Wlasiuk et al., 2009; Enard et al., 2010). Yet, in agreement with the hypothesis that there is strong selection against functional change in innate immune genes in vertebrates, several studies in humans and livestock found that TLRs have been subject to purifying selection (Yilmaz et al., 2005; Barreiro et al., 2009; Mukherjee et al., 2009; Seabury et al., 2010). Hence, analyses of patterns of selection on TLRs have yielded surprisingly mixed results.

To shed further light on the strength and direction of selection acting on the vertebrate innate immune defence in general, and TLRs in particular, and to identify putatively positively selected TLR sites, we studied signatures of selection in the Toll-like receptor 2 (TLR2) gene across 17 rodent species. Genetic polymorphisms in TLR2 are associated with resistance to a variety of disease causing agents, including *Staphylococcus aureus*, *Mycobacterium tuberculosis*, *Mycobacterium leprae*, *Treponema pallidum*, and *Borrelia burgdorferi* in humans and lab mice (Cook et al., 2004; Schröder et al., 2005; Texereau et al., 2005; Garantziotis et al., 2008; Schröder et al., 2008), which makes TLR2 a particularly good candidate for the study of selection acting on innate immune defence in wild populations.
Material and Methods

Samples

Tissue samples from wild populations of 15 rodent species (Table 1) were obtained using live-traps (Ugglan Special No1, Grahnab, Gnosjö, Sweden) or from colleagues working with the species. In addition, we used TLR2 sequences of *Mus musculus* (Rodentia, Murinae, NM_011905.3), *Rattus norvegicus* (Rodentia, Murinae, NM_198769.2) and *Cavia porcellus* (Rodentia, Caviinae, ENSCPOG00000025374) publicly available from GenBank and Ensembl.

Primer design

Murinae primers were designed using the *Mus musculus* (NM_011905.3) TLR2 sequence and the program Primer3 (Rozen & Skaletsky, 2000). For the Arvicolinae primers, we first aligned the *Mus musculus* and *Rattus norvegicus* TLR2 sequences and identified highly conserved regions. Primers designed to anneal to this conserved part of the TLR2 sequence were then used to amplify the homologous region in the bank vole (*Myodes glareolus*). Based on this partial bank vole TLR2 sequence, we designed new bank vole specific primers and performed 5’- and 3’-RACE (Rapid Amplification of C-DNA Ends) with total RNA from bank vole spleen using the SMART RACE cDNA Amplification Kit (Clontech, Mountain View, CA, USA). This technique allowed us to obtain the full-length cDNA sequence of the bank vole TLR2. Specific Arvicolinae primers were then designed based on this bank vole sequence using the program Primer3 (Rozen & Skaletsky, 2000).
DNA amplification and sequencing

Total DNA was extracted from the biopsies following the protocol of Laird et al. (1991). For each rodent species, we amplified the entire coding region of TLR2 in two overlapping amplicons. PCR reactions were performed in a total volume of 25 µl including 25 ng of total genomic DNA, 0.125 mM of each dNTP, 2.0 mM MgCl₂, 1x PCR Buffer (Applied Biosystems, Foster City, CA, USA), 1 mM of each primer, and 2.5 U AmpliTaq DNA polymerase (Applied Biosystems, Foster City, CA, USA) on a GeneAmp PCR Systems 9700 thermocycler (Applied Biosystems, Foster City, CA, USA). For the first amplicon we used the primer pairs Mur1TLR2F 5’- MRSGTCAATCTCAGAGGATG-3’ and Mur1TLR2R 5’- GAGTYACACMKRTAGCTGTCTG-3’ for the Murinae, and Arv1TLR2F 5’- CGTGTTCGAGCCTTGTG-3’ and Arv1TLR2R 5’- CTAACATCCAGCACTCCAG-3’ for the Arvicolinae. For the second amplicon, we used the primer pairs Mur2TLR2F 5’- CAAACTGRAGACTYTGGAAGC-3’ and Rod2TLR2R 5’- GAACTTAGGACTTTATTGCAGTTCTC-3’ for the Murinae, and Arv2TLR2F 5’- CTTGACATCAGCCGGAACAG-3’ and Rod2TLR2R for the Arvicolinae. Additional internal primers were used for sequencing.

The PCR protocol included an initial denaturation step at 94 °C for 5 minutes, followed by 35 cycles of denaturation at 94 °C for 30 seconds, annealing at 56 °C for 30 seconds and extension at 72 °C for 150 seconds. The program ended with a final extension step at 72 °C for 10 minutes. The PCR products were sequenced in both directions on an ABI Prism 3730 capillary sequencer (Applied Biosystems, Foster City, CA, USA) using Big Dye terminator v3.1 chemistry (Applied Biosystems, Foster City, CA, USA).
Sequence analysis

Sequences were processed, assembled and aligned using ClustalW in the program Geneious 5.0. (Drummond et al., 2009). Individual alignments were checked and improved by eye. Consensus sequences were created and all inconsistencies and polymorphisms were examined by eye. The intracellular region of TLR2 is very highly conserved across taxa, we never observed a mismatch between overlapping amplicons and there was no indication that we amplified more than one locus. Together, this strongly suggests that sequences obtained from the different rodent species indeed represent true homologues. The full alignments of the rodent TLR2 sequences are presented in Supplementary material 1. The length of the predicted TLR2 protein was 782 – 784 amino acids in all sequenced rodent species. Sequences were submitted to NCBI GenBank (see Table 1 for accession numbers).

Tests of selection

To test for signatures of past selection in the TLR2 DNA sequence of rodents, we performed a number of standard tests (see e.g. Yang & Bielawski, 2000; Nielsen, 2005). The established structure and inferred function of the TLR2 protein (Xu et al., 2000; Gautam et al., 2006; Jin et al., 2007) allowed us to make a priori predictions about which sites are expected to evolve under positive selection (i.e. sites involved in pathogen recognition), and thus to put our results into a functional perspective. First, to identify regions that may have evolved under positive selection, we performed a sliding window analysis of the ratio ($K_A / K_S$ or $\omega$) of non-synonymous substitutions per non-synonymous site ($K_A$) to synonymous substitutions per
synonymous site (Ks) along the TLR2 gene of Murinae, Myodini and Arvicolini, compared to the outgroup Cavia porcellus using DnaSP 5 (Librado & Rozas, 2009). The Jukes-Cantor correction (JC) was applied when calculating KA and KS. The ratio was calculated for the entire coding region of TLR2 and was averaged for a window of 30 bp, with a step size of 10 bp. A ratio ($\omega$) > 1 is indicative of positive selection (i.e. promoting change at the amino acid level), whereas $\omega$ < 1 is indicative of purifying selection (Nielsen, 2005).

Second, we used the phylogeny-based maximum likelihood analysis of $\omega$ as implemented in the program CODEML of the package PAML 4.3 (Yang, 1997) to statistically test for positive selection acting on TLR2 codons. We generated log likelihood values for models where $\omega$ is allowed to vary among sites within the interval 0 – 1 (neutral models), and for models, which allow $\omega$ to be > 1 for some sites (selection models) following Yang et al. (2000). First, we tested if $\omega$ differs among sites by comparing model M0, which assumes a constant $\omega$ across all sites to model M3, which allows $\omega$ to vary among sites. To formally test for the presence of sites evolving under positive selection, we then compared a nearly neutral model of $\omega$ variation (M1a) to a model that allows for positive selection (M2a) (Wong et al., 2004; Yang et al., 2005), and a neutral model M7, which estimates $\omega$ with a beta distribution over the interval 0 – 1 to a selection model M8, which additionally allows for positively selected sites ($\omega$ > 1) (Yang et al., 2000; Yang & Nielsen, 2002). A beta distribution of $\omega$ has been suggested to more accurately reflect the distribution of $\omega$ among sites in biological data (Friedman & Hughes, 2007). Furthermore, we compared beta model M8a (a special case of M8, which fixes $\omega$ of the highest site class to 1) to M8 (Swanson et al., 2003). We compared the models using likelihood
ratio tests. Twice the log-likelihood difference ($2\Delta \ell$) between models follows a $\chi^2$ distribution with the degrees of freedom equal to the difference in the number of parameters between the models (Yang & Nielsen, 2002). All models were run multiple times with different starting values for $\omega$ to ensure the correct estimation of the model parameters. The unrooted tree input file (Supplementary material 2) for these analyses was created by calculating a distance matrix in the program DNAdist and a tree file in the program NEIGHBOR implemented in the package PHYLIP (Felsenstein, 2005).

Finally, we used Empirical Bayes approaches implemented in CODEML (Yang, 1997) to infer which sites of the TLR2 sequence may have evolved under positive selection (Yang et al., 2005). This algorithm estimates for each site the posterior probability of belonging to one of three site classes: low $\omega$, intermediate $\omega$ and high $\omega$. We considered two different approaches to determine sites under selection, the naive-empirical Bayes and the Bayes-empirical Bayes method (Yang et al., 2005). Positive selection was interfered if the posterior probability of $\omega > 1$ for a site was $\geq 0.95$.

**Results**

The sliding window analysis of the ratio $K_A / K_S$ along the entire TLR2 coding region produced a strong peak in LRRs 10 – 11. This signal of positive selection was repeatable across Murinae, Myodini and Arvicolini (Fig. 1). A second prominent $K_A / K_S$ peak was observed in LRR 14 in Murinae and Arvicolini, but not in Myodini (Fig. 1). A high $K_A / K_S$ can be the result of high $K_A$, low $K_S$, or a combination of both. To investigate what caused the peaks we observed, we therefore calculated $K_A$ and $K_S$.
separately for each of the peak regions and compared these values to the average $K_A$ and $K_S$ values over different functional regions of TLR2 (Table 2). It turned out that both peaks were caused by a combination of high $K_A$ and low $K_S$. To evaluate how extreme the $K_A$ and $K_S$ at the observed peaks were, we generated 100 random windows of 30 bp (i.e. the window size of the sliding window analysis) along TLR2, and calculated $K_S$ and $K_A$ for each window (Supplementary material 3). In this generated dataset we did neither observe $K_A$ values that were as high, nor $K_S$ values that were as low as at the two observed peaks obtained in the sliding window analysis (Table 2, Supplementary material 3), indicating that the probability of observing such extreme values by chance is $< 1\%$.

In addition to the sliding window analysis, phylogeny-based maximum likelihood approaches (Yang, 1997) provided evidence that positive selection has acted on codons of the rodent TLR2 sequence. We detected significant $\omega$ heterogeneity along the TLR2 sequence (Model 0 vs. 3: $\chi^2_4 = 200.32, P < 0.001$) with 68% of the sites evolving under strong purifying selection ($\omega = 0.06$), 32% of the sites evolving neutrally ($\omega = 0.88$), and 0.5% of the sites evolving under strong positive selection ($\omega = 4.8$) (Table 3). A first selection test, in which we compared a nearly neutral model (M1a, two site classes) with a selection model (M2a, three site classes), did not provide statistical support for positive selection ($\chi^2_2 = 0, P = 1$, Table 3). However, when comparing models that estimate $\omega$ with a beta distribution the selection model M8 (eleven site classes), which allows $\omega > 1$, performed significantly better than the neutral model M7 (ten site classes), which restricts $\omega$ to the interval 0 – 1 ($\chi^2_2 = 6.66, P = 0.036$, Table 3). Furthermore, the comparison of model M8a vs. M8 (both eleven site classes; $\chi^2_1 = 4.34, P = 0.037$) provided evidence that $\omega$ of the
highest site class is significantly $> 1$, and thus that parts of TLR2 are evolving under positive selection. Selection models did not perform significantly better than neutral models when analysing Murinae, Arvicolini, and Myodini separately (results not shown).

To identify specific sites of TLR2 that have likely evolved under positive selection, we used Empirical Bayes approaches (Yang et al., 2005). Even though both model M8 and model M3 indicated that $0.5 – 0.6\%$ of all TLR2 sites evolve under positive selection ($\omega = 4.6$, Table 3), which corresponds to 3 – 5 codons, we had the statistical power to identify only one site, amino acid 354, as being positively selected with a posterior probability $\geq 0.95$ (Table 3, Fig. 2).

Discussion

Unlike in plants and invertebrates, where it is well documented that positive selection has shaped innate immunity genes (Jiggins & Kim, 2006; Tiffin & Moeller, 2006; Sackton et al., 2007; Obbard et al., 2009), it is often assumed that in vertebrates purifying selection is the predominant selective force acting on the innate branch of the immune system (Yilmaz et al., 2005; Barreiro et al., 2009; Mukherjee et al., 2009; Seabury et al., 2010). Nonetheless, a few recent studies have demonstrated that also in vertebrates positive selection has acted on functionally relevant sites of innate defence genes (Ferrer-Admetlla et al., 2008; Hughes & Plontkivska, 2008; Wlasiuk et al., 2009; Enard et al., 2010). In agreement with these latter studies, we found indication that positive selection has shaped parts of the rodent TLR2 gene.
Positive selection is predicted to act on regions of immune genes that are involved in pathogen recognition (e.g. Hughes et al., 1990; Hedrick, 2002; Hughes & Friedman, 2008). In agreement with this prediction, the sliding window analysis of $K_A / K_S$ along the TLR2 coding region produced a pronounced signal of positive selection in LRR 10 – 11. This signal of selection was repeatable across all three analysed taxonomic groups. A second peak was observed in LRR 14 in Murinae and Arvicolini. The LRRs 10 – 11 regions contain 10 to 13 putative pathogen-binding sites in humans and lab mice (Jin et al., 2007), and variation in these binding sites might directly influence the pathogen-binding capacity or specificity of the receptor. LRR14, on the other hand, contains sites that are involved in the heterodimerisation of TLR2 with TLR1 (Jin et al., 2007). Interestingly, the common human TLR2 variant Arg753Gln, which is associated with resistance to late stage Lyme disease (Schröder et al., 2005), also affects TLR2 heterodimerisation (Gautam et al., 2006). This indicates that even though heterodimerisation sites are not directly involved in pathogen-binding, they might indirectly influence the pathogen-binding ability of the receptor by changing the conformation of the TLR2 – TLR1 heterodimer structure (Gautam et al., 2006).

Sliding window analyses of $K_A / K_S$ have been criticised (see e.g. Hughes & Friedman, 2008; Nozawa et al., 2009; Wolf et al., 2009) because $K_A / K_S$ peaks could reflect regions with particularly low $K_S$ rather than high $K_A$, something which can occur by chance given the high variance of $K_S$ (Parmley & Hurst, 2007). In such cases, $K_A / K_S$ peaks will not be indicative of positive selection. A detailed examination of our data, however, showed that stochastic variation of $K_S$ alone cannot explain the observed $K_A / K_S$ peaks. Rather the $K_A / K_S$ peaks are a result of both exceptionally high $K_A$ as well as exceptionally low $K_S$. An outlier analysis
demonstrated that such extreme $K_A$ and $K_S$ values would be expected by chance in < 1%. The reason for the low $K_S$ is not entirely clear, but one potential explanation is that selective sweeps (indicated by the high $K_A$), in combination with a high recombination rate, have depleted these particular regions of neutral variation. In any case, the fact that the $K_A / K_S$ peaks are at least partly a result of high $K_A$, in combination with the repeatability of the peaks across the analysed clades, and the location of the peaks in the predicted regions of TLR2 (Jin et al., 2007) provides strong evidence that the observed $K_A / K_S$ peaks are indicative of positive selection rather than chance events.

The notion that positive selection shapes parts of the rodent TLR2 was further corroborated by phylogeny-based maximum likelihood approaches. Although 68% of the TLR2 sequence was estimated to evolve under strong purifying selection and another 32% evolving neutrally, 0.5 – 0.6% of all sites (which corresponds to 3 – 5 amino acids) were found to be shaped by positive selection. Using empirical Bayes approaches we were able to identify one of these positively selected sites, amino acid 354. This site is located in LRR 12 in close proximity to both pathogen-binding and heterodimerisation sites (Jin et al., 2007). The nonsynonymous mutations observed at this site have led to pronounced changes in the amino acids’ physico-chemical properties (i.e. polarity, charge, volume; Supplementary material 1), which are thought to be important in the determination of protein structure (Grantham, 1974; Miyata et al., 1979). Such radical changes are expected if natural selection promotes functional change of proteins, but they are incompatible with the substitution pattern predicted under relaxed purifying selection (see e.g. Nielsen, 2005). Remarkably, LRR12 showed a very low $K_A / K_S$ ratio in the sliding window analysis, indicating that codons in close proximity to amino acid 354 are subject to
strong purifying selection, and thus that patterns of selection can differ markedly on a
very small scale.

To conclude, our results indicate that purifying selection is not the only evolutionary
force that has shaped the rodent TLR2 sequence, as a small number of TLR2
codons have evolved under positive selection. There are numerous examples of
single nucleotide polymorphisms having pronounced evolutionary consequences, for
example in genes involved in pathogen virulence (Brault et al., 2007) or pigmentation
(Hoekstra et al., 2004; Rosenblum et al., 2004; Linnen et al., 2009). Since the
positively selected sites in the TLR2 sequence were in close proximity to pathogen-
binding and TLR heterodimerisation sites (Jin et al., 2007), it is thus plausible that
functional changes in these regions may have direct consequences for host-
pathogen interactions. So far, ecological and evolutionary research into the
interactions between vertebrate hosts and their parasites has almost exclusively
focused on the acquired branch of the immune system (i.e. associations between
host MHC polymorphisms and parasite resistance (Paterson et al., 1998; Wegner et
al., 2003; Meyer-Lucht & Sommer, 2005; Oliver et al., 2009)), whereas the host’s
innate immune defence has received considerably less attention. Associating
putatively positively selected amino acid differences in innate immunity genes with
parasite resistance and tolerance in wild vertebrate populations might thus be a
fruitful endeavour to obtain a more comprehensive understanding of evolutionary
processes in natural host-parasite systems.

Furthermore, future studies will need to examine whether patterns of selection on
TLR2 are exceptional, or if also other members of the Toll-like receptor gene family,
and other innate immunity genes in general, have been subject to positive selection
during the evolutionary past, and if differences in selection patterns can be
associated with functional difference between immune genes. New technologies
(such as next generation sequencing) will greatly facilitate such comparative studies.

Acknowledgements

We are grateful to Hitoshi Suzuki, Zbyszek Boratynski, Gerald Heckel, Anna
Lindholm, Melanie Monroe, Mohammad Nafi S. Al-Sabi, Alice Remy and Peter
Wandeler for providing samples from their study populations, Kristin Scherman for
providing bank vole RNA samples and Martin Stervander for help with MrBayes. We
thank Staffan Bensch and Bengt Hansson for discussion, and Phil Hedrick, Erik
Postma, Pedro Vale and two anonymous reviewers for comments on earlier versions
of the manuscript. The project was funded by the Swedish Research Council (grants
621-2206-2876 and 621-2006-4551 to HW and LR). BT was supported by a Swiss
National Science Foundation Postdoctoral Fellowship (PA0033-121466).
Table legends

Table 1 Origin of rodent samples, number of sequenced individuals (N) and NCBI GenBank accession numbers. *Sequences obtained from NCBI GenBank.

Table 2 Non-synonymous substitutions per non-synonymous site (K_\text{\textalpha}) and synonymous substitutions per synonymous site (K_\text{S}) averaged over the whole TLR2 coding region, the leucine-rich repeat 9 – 14 region (LRR 9 – 14), where most of the pathogen-binding and dimerisation sites are situated (Jin et al. 2007), the intracellular TIR domain and peak 1 and peak 2 identified in the sliding-window analysis (Fig. 1) for all species combined and the Murinae, Myodini and Arvicolini separately. Peak 1 and 2 are characterised by both high K_\text{\textalpha} and low K_\text{S}.

Table 3 Results of PAML site-models and positively selected sites identified by empirical Bayes approaches. Parameters are p_0 = proportion of sites where \omega < 1, p_1 = proportion of sites where \omega = 1 and p_2 = proportion of sites where \omega > 1 (selection models only). For models M7, M8 and M8a, p and q represent parameters of the beta distribution. Positive selection was inferred if the posterior probability of \omega > 1 for a site was 0.95 or higher (bold). Sites with a posterior probability of \omega > 1 between 0.50 – 0.949 are also shown (italic). We considered two different approaches to determine sites under selection, the naive-empirical Bayes and the Bayes-empirical Bayes method (Yang et al., 2005). A site was included if one or both of the approaches gave statistical support for the site. Amino acids correspond to the \textit{Mus musculus} sequence.
**Figure legends**

**Fig. 1** Signatures of positive selection on TLR2. Sliding window analysis of the ratio of non-synonymous substitutions per non-synonymous site \((K_A)\) to synonymous substitutions per synonymous site \((K_S)\) along the entire coding region of the TLR2 gene of A) all species combined, and B) Murinae, C) Myodini and D) Arvicolini separately compared to *Cavia porcellus* (window size 30 bp, step size 10 bp). A ratio > 1 is indicative of positive selection. Note the different scales on the y-axes. The schema of the functional domains and the putative positions of the pathogen-binding and TLR2 – TLR1 heterodimerisation sites are based on Gautam *et al.* (2006), Jin *et al.* (2007), and Jann *et al.* (2008). The positively selected site (PSS) identified by empirical Bayes approaches is shown in red. LRR: leucine-rich repeat, TM: transmembrane domain, TIR: Toll/ Interleukin-1 receptor domain.

**Fig. 2** Expected \(\omega\) of codons along the TLR2 sequence. Estimates are based on Naive empirical Bayes probabilities for three classes.
References


doi:10.1371/journal.pgen.1000840.

Distributed by the author. Department of Genome Sciences, University of
Washington Seattle.

Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marques-Bonet, T., Ramirez-Soriano, A.,
Muntasell, A., Navarro, A., Lazarus, R., Calafell, F., Bertranpetit, J. & Casals,
F. 2008. Balancing selection is the main force shaping the evolution of innate

Friedman, R. & Hughes, A.L. 2007. Likelihood-ratio tests for positive selection of
human and mouse duplicate genes reveal nonconservative and anomalous


and functional evidence for the role of the TLR2 DD loop in TLR1/TLR2

*Science* 185: 862-864.

*Evolution* 56: 1902-1908.


