



Published in final edited form as:

Immunogenetics. 2008 May ; 60(5): 249–256.

Functional Diversification of the Toll-like Receptor Gene Family

Austin L. Hughes^{1,*} and Helen Piontkivska²

¹*Department of Biological Sciences, University of South Carolina, Columbia SC 29208*

²*Department of Biological Sciences, Kent State University, Kent OH 44242*

Abstract

Phylogenetic analyses supported the hypothesis that the vertebrate toll-like receptors (TLRs) include two very ancient groups that arose by gene duplication prior to the divergence of protostomes and deuterostomes: (1) the TLR1 family (including mammalian TLR1, -2, -6, and -10); and (2) a clade including the remainder of mammalian TLRs. Correlating data on ligand type, subcellular localization, and gene expression in leukocytes and other tissues with the phylogeny provided evidence that certain major functional specializations within the TLRs occurred after ancient gene duplication events and that these traits have been retained through further events of gene duplication. For example, the recognition of bacterial lipoproteins appears to have arisen in the ancestor of the TLR1 family, and continues to characterize members of that family whose ligands are known. Likewise, expression on the endosomal membrane and the recognition of nucleic acids appears to have been arisen in the ancestor of the TLR7 family and some related TLRs. On the other hand, gene expression patterns across tissues appear to have been much more volatile over the evolution of the vertebrate TLRs, since genes may show expression profiles similar to those of distantly related genes but dissimilar to those of closely related genes. Thus, the vertebrate TLRs provide an example of a multi-gene family in which gene duplication has been followed by extensive changes in certain aspects of gene function while others have been conserved throughout vertebrate history.

The innate immune system of vertebrates includes a number of receptors that recognize distinctive pathogen-associated molecular patterns (PAMPs). The toll-like receptors (TLRs) constitute a multi-gene family in vertebrates, whose different members recognize distinctive PAMPs and play a key role in defense against both viral and bacterial infections (Aderem and Ulevitch 2000; Albiger et al. 2007; Carpenter and O'Neill 2007; Dunne and O'Neill 2005; Ishii et al. 2005; Krishnan et al. 2007; Pichlmair and Reis e Sousa 2007). TLRs are named for homology with the Toll protein of *Drosophila* and other insects, a receptor involved in developmental and immune signaling (Tanji and Ip 2005). The PAMPs recognized by TLRs include lipopolysaccharides, nucleic acids, and the bacterial flagelin protein (Dunne and O'Neill 2005). Like Toll and related insect molecules, the TLRs are characterized by leucine-rich repeats (LRR) involved in PAMP recognition and by the intracellular Toll/interleukin-1 receptor domain (TIR domain) involved in signal transduction (Matsushima et al. 2007; Watters et al. 2007).

In humans, TLR 4 recognizes lipopolysaccharide, while TLR2 can form a heterodimer with TLR6 that recognizes diacyl lipopeptides or a heterodimer with TLR1 that recognizes triacyl lipopeptides (Dunne and O'Neill 2005; Krishnan et al. 2007). TLR5, on the other hand, recognizes bacterial flagellins. TLR1, TLR2, TLR4, TLR5, and TLR6 are all expressed on the plasma membrane; by contrast, there are other toll-like receptors expressed on the endosomal membrane (TLR3, TLR7, TLR8, and TLR9), which recognize nucleic acids. TLR3 recognizes

*Address correspondence to: Austin L. Hughes, Ph.D. Department of Biological Sciences, University of South Carolina, Coker Life Sciences Bldg. 700 Sumter St. Columbia SC 29208 USA, Tel: 1-803-777-9186, Fax: 1-803-777-4002 Email: austin@biol.sc.edu

double-stranded RNA, while TLR7 and TLR8 recognize single-stranded RNA, and TLR9 recognizes the unmethylated CpG dinucleotide in bacterial DNA.

TLRs use a variety of intracellular adaptors to initiate a response to PAMP recognition (Krishnan et al. 2007; Watters et al. 2007). MyD88 is used as an adaptor protein by every human TLR whose adaptor is known, with the exception of TLR3. By contrast, TLR3 uses the TRIF (or TICAM-1) protein (Krishnan et al. 2007). TLR4 also can use TRIF, as well as the related adaptor TRAM (or TICAM-2; Krishnan et al. 2007). TLR1, TLR2, TLR4, and TLR6, all of which are expressed on the plasma membrane, are known to use an additional adaptor protein MAL (or TIRAP) related to MyD88 (Kagan and Medzhitov 2006; Krishnan et al. 2007).

Thus, the TLRs provide an example of a multi-gene family whose members have diversified functionally (Hughes 1994). Evolutionary biologists have long theorized that gene duplication is a major source of evolutionary novelty (Nei 1969; Ohno 1970), but how new functions arise after gene duplication remains poorly understood (Hughes 1994, 1999; Lynch 2007; Piatigorsky 2007). It is increasingly clear that gene function is a multi-dimensional variable, involving the patterns of expression in different tissues and under different environmental conditions, as well as a multitude of interactions among proteins and between proteins and non-protein ligands. Gene duplication permits duplicate genes to explore different aspects of the multi-dimensional functional space (Hughes 2005; Piatigorsky 2007). The TLRs represent an excellent model system for studying these processes because information is available regarding several aspects of their function, including PAMP recognized, subcellular localization, and pattern of gene expression across tissues.

Although the vertebrate TLRs have been the subject of a number of phylogenetic studies (e.g., Friedman and Hughes 2002; Ishii et al 2007; Roach et al. 2005; Sanghavi et al. 2004; Wiens et al. 2006), these analyses have not attempted to use phylogenetic methods to reconstruct major events of functional differentiation within this gene family. Here we use phylogenetic methods to reconstruct the patterns of gene duplication and functional differentiation of major clades of mammalian TLRs and the relationship of vertebrate genes with insect homologues. By combining data on PAMP recognition, subcellular localization, adaptor use, and gene expression pattern across different tissues with the results of our phylogenetic analyses, we reconstruct the major functional shifts that have occurred over the evolution of this gene family.

Methods

Phylogenetic Analyses

A total of 128 representative vertebrate Toll-like receptor (TLR) sequences and eight insect Toll receptor sequences were used in analyses (Supplementary Table S1). Vertebrate TLR sequences were chosen so as to emphasize as complete a selection as possible from vertebrate species with completely sequenced genomes or ongoing genome projects; and we used insect sequences to provide information for rooting the vertebrate TLR families. Amphibian sequences were not included because available amphibian sequences all fall into just one of several major TLR families found in other vertebrates (Roach et al. 2005). Non-insect invertebrate sequences (e.g., Wiens et al. 2006) were not included because including additional highly diverged sequences decreased the number of aligned amino acid sites available for the analysis. Chromosomal locations, where known, were obtained from NCBI Map Viewer (Wheeler et al. 2006).

The phylogenetic analysis was based on an alignment of the Toll/interleukin-1 receptor domain (TIR domain) because other domains, particularly the leucine-rich repeats (LRR) cannot be reliably aligned between distantly related members of the family (Friedman and Hughes

2002; Roach et al. 2005). The amino acid sequences of the TIR domain were aligned by the Clustal X program (Thompson et al. 1997; Supplementary Figure S1). In all analyses, any site at which the alignment postulated a gap in any sequence was excluded from the analysis; so that a comparable set of sites was used.

Phylogenetic analyses were conducted using two methods: (1) A minimum evolution (ME) tree (Rzhetsky and Nei 1992) was constructed on the basis of the JTT amino acid distance (Jones et al. 1992), with the rates assumed to vary among sites following a gamma (Γ) distribution, using the MEGA 4 program (Tamura et al. 2007). The shape parameter (α) of the distribution was estimated from the data using the TREEPUZZLE program (Schmidt et al. 2002). The standard error test of branch lengths was conducted by the bootstrap method (Dopazo 1994). (2) A Bayesian tree (Huelsenbeck and Ronquist 2001) was reconstructed using the JTT + model (Rodriguez et al. 1990). All parameters were estimated from the data; the gamma distribution was approximated using 4 categories. Four chains were run for 1,000,000 generations, and trees were sampled every 100 generations. Bayesian posterior probabilities were inferred from the last 5000 sampled trees.

Nucleotide Substitution

In order to examine the pattern of nucleotide substitution in different domains of primate genes encoding TLRs, we separately aligned the amino acid sequences of orthologous gene pairs from human and rhesus monkey (*Macaca mulatta*). We then imposed the alignments on the DNA sequences and estimated the number of synonymous nucleotide substitutions per synonymous site (d_S) and the number of nonsynonymous nucleotide substitutions per nonsynonymous site (d_N) in different gene regions by Nei and Gojobori's (1986) method.

The LRR and TIR domains were defined following annotations in the Swissprot accessions Q15399, O60603, O15455, O00206, Q9JLF7, Q9Y2C9, Q9NYK1, Q9NR97, Q9NR96, and Q9BXR5.

Gene Expression Data

In order to correlate phylogenetic relationships with gene expression patterns, we examined two data sets on expression of human *TLR* genes: (1) a human leukocyte data set, providing data on gene expression in major leukocyte types in human (Jeffrey et al. 2006); and (2) a human tissue data set, providing data on TLR expression in various human tissues (Nishimura and Naito 2005). Gene expression is a multi-dimensional phenomenon with large dimensionality; as with any gene expression data set, the data available for our analyses provided information on only a subset of the cell types and conditions under which gene expression might be measured.

In the human leukocyte data set, there were 17 individual combinations of cell-type and treatment: basophils; B cells; cord blood-derived mast cells; cord blood-derived mast cells (IGE); central memory T cells; immature dendritic cells; dendritic cells (LPS); effector memory T cells; eosinophils; eosinophils (PMA); macrophages; macrophages (LPS); neutrophils; neutrophils (LPS); NK cells; Th1 cells; and Th2 cells. In the above list, treatments are indicated in parentheses; cell-types with no treatment indicated are unstimulated controls. In cases where the original data included two replicates for a given cell-type or cell-type and treatment combination, we averaged the two values. The available data (Gene Expression Omnibus series GES3982), using the Affymetrix Genome U133 Array Set HG-U133A, provided gene expression data for eight human *TLR* genes (excluding *TLR9* and *TLR10*). We computed correlations among genes across the 17 cell-type/treatment combinations and clustered genes by a hierarchical clustering algorithm on the basis of the correlation matrix. The pairwise distance between two genes was defined as $1-r$, where r is the correlation coefficient. The

clustering procedure was performed using the Minitab statistical software package, release 13 (<http://www.minitab.com/>).

The human tissue data set included data on the following 18 adult human tissues: adrenal gland; brain; heart; kidney; liver; lung; placenta; prostate; salivary gland; skeletal muscle; small intestine; spinal cord; spleen; testis; thymus; thyroid gland; trachea; and uterus. The data were obtained using high-sensitivity real-time reverse transcription PCR (RT-PCR) in total RNA from pooled specimens (Nishimura and Naito 2005). We computed correlations among genes across the 18 tissues and clustered genes as described above.

Results

Phylogenetic Analyses

In the minimum evolution (ME) tree of the TIR domain of vertebrate TLRs, there were two major clusters, each of which included both insect and vertebrate TLRs: (1) a cluster including insect TOLL9 and mammalian TLR1, -2, -6, and -10; and (2) a cluster including all other mammalian TLRs (TLR3, -4, -5, -7, -8, and -9), along with the remaining insect sequences (Figure 1). The two clusters were separated by a highly significant internal branch (Figure 1). The same clusters were observed in the Bayesian tree, and the posterior probability of the internal branch separating them was 100% (Figure 2).

The cluster including mammalian TLR1, -2, -6, and -10 included all mammalian TLRs known to be involved in the recognition of bacterial lipopeptides; namely TLR1, -2, and -6 (Figures 1 and 2). We designate this cluster as the TLR1 family, following Roach et al. (2005). The TLR sequences from lamprey clustered with the TLR1 family, consistent with the phylogenetic analysis of Ishii et al. (2006). The relationship of insect TOLL9 to the vertebrate sequences in the TLR1 family was not well resolved in the ME tree (Figure 1); but in the Bayesian tree TOLL9 formed an outgroup to all vertebrate TLR1 family members, and the posterior probability of the branch supporting this pattern was 97% (Figure 2). In the ME tree, TLR2 fell outside the cluster containing TLR1, -6, and -10; and this pattern was strongly supported in both analyses (Figures 1 and 2).

The relationships among TLR1, -6, and -10 were not well resolved in the ME tree (Figure 1). The Bayesian tree was consistent with the hypothesis that the *TLR10* gene arose before duplication of the *TLR1* and *TLR6* genes (Beutler and Rehli 2002). TLR10 clustered outside TLR1 and TLR6 in the Bayesian tree, although the posterior probability of the branch supporting this pattern was only 83% (Figure 2). The phylogenetic analyses suggested that no orthologs to the *TLR1*, -6, and -10 genes are found outside of mammals, consistent with the hypothesis that these genes duplicated within the mammalian lineage.

TLR3, -7, -8, and -9 formed a monophyletic group in the ME tree, and TLR5 fell outside this group; but the group was not supported by a significant internal branch (Figure 1). In the Bayesian tree, by contrast, TLR5 clustered with TLR3, although the posterior probability of the branch supporting this pattern was only 72% (Figure 2). In the Bayesian tree, TLR5 formed part of a larger cluster with TLR3, -7, -8, and -9; and the posterior probability of the branch supporting this pattern was 93% (Figure 2). We designate TLR7, -8, and -9 as the TLR7 family, following Roach et al. (2005). The members of this family formed a monophyletic group in both the ME and Bayesian trees (Figures 1 and 2). In the Bayesian tree, the posterior probability of the branch supporting this pattern was 100% (Figure 2).

Within the TLR7 family, both trees were consistent with the hypothesis that the *TLR7* and *TLR8* genes duplicated prior to the origin of mammals because chicken TLR7 grouped with apparent mammalian orthologs (Figures 1 and 2). In the Bayesian tree, Fugu TLR7 also

grouped with bird and mammal TLR7 (Figure 2). However, neither of the trees showed strong support for the clustering of non-mammalian sequences with mammalian TLR7 (Figures 1 and 2); and thus the age of the duplication of the *TLR7* and *TLR8* genes was not resolved.

Nucleotide Substitution in Primate Genes

We compared numbers of synonymous nucleotide substitutions per synonymous site (d_S) and of nonsynonymous nucleotide substitutions per nonsynonymous site (d_N) in the LRR and TIR between orthologous human and rhesus monkey *TLR* gene pairs (Table 1). None of the 10 genes showed a significant difference between LRR and TIR with respect to d_S (Table 1), supporting the hypothesis that there was no difference in mutation rate between LRR and TIR. By contrast, d_N in the TIR was significantly lower than that in the LRR in the case of five of the genes (Table 1), indicating stronger purifying selection on the TIR than the LRR in these cases and thus greater functional constraint on the former than on the latter. The genes of the TLR1 family did not show a difference in d_N between the TIR and the LRR (Table 1), indicating a relatively relaxed constraint on the TIR in the primate members of the TLR1 family. On the other hand, d_N in the TIR was significantly lower than that in the LRR in all members of the TLR7 family (Table 1), indicating strong constraint on the TIR in that group.

Expression Data

Gene expression in the human leukocyte data set showed highly similar expression patterns in two pairs of *TLR* genes: (1) *TLR2* and *TLR4*; and (2) *TLR1* and *TLR8* (Figure 3A). *TLR1* and *TLR8* also showed a pattern very similar to that seen in *TLR6* and *TLR5* (Figure 3A). By contrast, *TLR3* and *TLR7* did not show patterns similar to those of other *TLR* genes in the human leukocyte data set (Figure 3A). In the human tissue data set, *TLR1* and *TLR8* showed very similar expression patterns (Figure 3B). *TLR6* and *TLR10*, both members of the *TLR1* family, also showed very similar expression patterns in the human tissue data set (Figure 3B).

Discussion

Phylogenetic analyses of vertebrate toll-like receptors (TLRs) by both minimum evolution (ME) and Bayesian methods showed two strongly supported clusters including both vertebrate and insect sequences. One cluster corresponded to the TLR1 family as defined by Roach et al. (2005), which includes mammalian TLR1, -2, -6, and -10, while the other included all other mammalian TLRs. This topology supports the hypothesis that the vertebrate TLRs include two very ancient groups that arose by gene duplication prior to the divergence of protostomes (including insects) and deuterostomes (including vertebrates). The phylogenetic analysis of Sanghavi et al. (2004) also supported the hypothesis that the TLR1 family diverged from other vertebrate TLRs prior to the divergence of protostomes and deuterostomes.

The absence of orthologs to the *TLR1*, -6, and -10 genes outside of mammals supports the hypothesis that these genes duplicated within the mammalian lineage. In human, all members of the TLR1 family map to chromosome 4; and *TLR6*, *TLR1*, and *TLR10* are located in consecutive order (Kruithof et al. 2007). Close linkage of *TLR6*, *TLR1*, and *TLR10* is also found in dog and bovine (Opsal et al. 2006; Supplementary Table S1). Kruithof et al. (2007) present evidence of gene conversion events between *TLR1* and *TLR6*. Gene conversion is more likely in the case of recent duplicates; thus, the results of Kruithof et al. (2007) are consistent with the hypothesis that these genes duplicated within the mammalian lineage. By contrast, although *TLR7* and *TLR8* are closely linked on the X chromosome of human and several other placental mammals, the phylogenetic analyses suggested that these genes may have duplicated prior to the origin of mammals.

Our phylogenetic analysis supported the hypothesis that cellular localization on the plasma membrane and recognition of bacterial lipopeptides are ancestral characteristics of the TLR1 family. In humans, TLR1, TLR2, and TLR6 are known to interact with both MyD88 and MAL adapter proteins; and the phylogeny supports the hypothesis that this trait also was ancestral to the TLR1 family. The phylogenetic analysis thus supported the hypothesis that the ancestral traits of the TLR1 family were present in the ancestry of human TLR10, the function of which remains poorly known (Hasan et al. 2005; Bell et al. 2007).

TLR3, -7, -8, and -9 formed a monophyletic group in the ME tree, although the group was not supported by a significant internal branch (Figure 1). Given this topology, it is parsimonious to assume that common ancestor of TLR3, -7, -8, and -9 possessed the distinctive functional traits of these TLRs; namely, the recognition of nucleic acids and expression on the endosomal membrane (Dunne and O'Neill 2005). In the Bayesian tree, by contrast, TLR5 formed part of a larger cluster with TLR3, -7, -8, and -9, supported by a posterior probability of 93% (Figure 2). Thus, the Bayesian tree was consistent with the hypothesis that the recognition of nucleic acids and expression on the endosomal membrane were present in the common ancestor of TLR3, -5, -7, -8, and -9 but were lost in TLR5. If so, expression on the plasma membrane would be a secondarily acquired trait in the case of TLR5.

Two gene expression data sets, from human leukocytes and from a variety of other human tissues, showed that patterns of expression were similar between closely related *TLR* genes in only a few cases, all involving the TLR1 family. Two members of the TLR1 family, *TLR1* and *TLR6*, had similar expression patterns in both data sets (Figure 3). The human leukocyte data set lacked information on the expression of *TLR10*; but in the human tissue data set, *TLR10* showed an expression pattern very similar to that of *TLR6* and fairly similar to that of *TLR1* (Figure 3B).

On the other hand, very distantly related genes in some cases had similar expression profiles in certain cases. The very similar expression patterns of *TLR1* (TLR1 family) and *TLR8* (TLR7 family) in both data sets provided the most striking example of phylogenetically distant genes (Figures 1-2) with similar expression patterns. Note also that human *TLR1* (chromosome 4) and *TLR8* (X chromosome) are not physically linked; nor are the *TLR1* and *TLR8* genes linked in other mammals (Supplementary Table S1). Likewise, although the phylogenetic trees (Figures 1-2) supported a very distant relationship between *TLR2* and *TLR4* genes, these two genes showed very similar expression patterns in the human leukocyte data set (Figure 3A). *TLR2* (chromosome 2) and *TLR4* (chromosome 4) are not linked in humans; nor are the *TLR2* and *TLR4* genes linked in other mammals (Supplementary Table S1).

Analysis of the patterns of nucleotide substitution in LRR and TIR regions of primate TLR genes showed relatively relaxed functional constraint on the TIR of members of the TLR1 family in comparison to most other genes (Table 1). Because the TLR1 family members interact with both MyD88 and MAL, these results suggest that interaction with both of these adapters leads to a relaxation of constraint on the TIR. The members of the TLR7 family were consistent in showing evidence of strong functional constraint on the TIR. As far as is known, each mammalian member of this family has a single adapter protein with which it interacts; namely, MyD88 in the case of TLR7, -8, and -9 and TRIF in the case of TLR3 (Krishnan et al. 2007). Thus, interaction with a single intracellular adapter protein was associated with strong constraint on the amino acid sequence of the TIR.

Our analyses supported the hypothesis that certain major functional specializations within the TLRs occurred after phylogenetically ancient gene duplication events and that these traits have been retained through further events of gene duplication. For example, the recognition of bacterial lipoproteins appears to have arisen in the ancestor of the TLR1 family, and continues

to characterize members of that family whose ligands are known. Likewise, expression on the endosomal membrane and the recognition of nucleic acids appears to have been arisen in the ancestor of the TLR7 family and some related TLRs. On the other hand, gene expression patterns appear to have been much more volatile over the evolution of the vertebrate TLRs, since genes may show expression profiles similar to those of distantly related genes but dissimilar to those of closely related genes. Thus, the vertebrate TLRs provide an example of a multi-gene family in which gene duplication has been followed by extensive changes in certain aspects of gene function while others have been conserved throughout vertebrate history.

Acknowledgments

This research was supported by grant GM43940 from the National Institutes of Health to A.L.H. and by the Research Council of Kent State University.

References

- Aderem A, Ulevitch RJ. Toll-like receptors in the induction of the innate immune response. *Nature* 2000;406:782–787. [PubMed: 10963608]
- Albiger B, Dahlberg S, Henriques-Normark B, Normark S. Role of the innate immune system in host defence against bacterial infections: focus on the Toll-like receptors. *J Int Med* 2007;261:511–528.
- Bell MP, Svingen PA, Rahman MK, Xiong Y, Faubion WA Jr. FOXP3 regulates TLR10 expression in human T regulatory cells. *J Immunol* 2007;179:1893–1900. [PubMed: 17641056]
- Beutler B, Rehli M. Evolution of the TIR, tolls, and TLRs: functional inferences from computational biology. *Curr Top Microbiol Immunol* 2002;270:1–21. [PubMed: 12467241]
- Carpenter S, O’Neill LA. How important are Toll-like receptors for antimicrobial responses? *Cell Microbiol* 2007;9:1891–1901. [PubMed: 17521328]
- Dopazo J. Estimating errors and confidence intervals for branch lengths in phylogenetic trees by a bootstrap approach. *J Mol Evol* 1994;38:300–304. [PubMed: 8006997]
- Dunne A, O’Neill LA. Adaptor usage and toll-like receptor signaling specificity. *FEBS Lett* 2005;579:3330–3335. [PubMed: 15876435]
- Friedman R, Hughes AL. Molecular evolution of the NF- κ B signaling system. *Immunogenetics* 2002;53:964–974. [PubMed: 11862396]
- Hasan U, Chaffois C, Gaillard C, Saulnier V, Merck E, Tancredi S, Guiet C, Brière F, Vlach J, Lebecqu S, Trinchieri G, Bates EEM. Human TLR10 is a functional receptor, expressed by B cells and plasmacytoid dendritic cells, which activates gene transcription through MyD88. *J Immunol* 2005;174:2942–2950. [PubMed: 15728506]
- Huelsenbeck JP, Ronquist F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* 2001;17:754–755. [PubMed: 11524383]
- Hughes AL. The evolution of functionally novel proteins after gene duplication. *Proc. R. Soc. Lond* 1994;B 256:119–124.
- Hughes, AL. Adaptive evolution of genes and genomes. Oxford University Press; New York: 1999.
- Hughes AL. Gene duplication and the origin of novel proteins. *Proc Natl Acad Sci USA* 2005;102:8791–8792. [PubMed: 15956198]
- Ishii KJ, Coban C, Akira S. Manifold mechanisms of toll-like receptor-ligand recognition. *J Clin Immunol* 2005;25:511–521. [PubMed: 16380815]
- Ishii A, Matsuo A, Sawa H, Tsujita T, Shida K, Matsumoto M, Seya T. Lamprey TLRs with properties distinct from those of the variable lymphocyte receptors. *J Immunol* 2007;178:397–406. [PubMed: 17182578]
- Jeffrey KL, Brummer T, Rolph MS, Liu SM, Callejas NA, Grumont RJ, Gillieron C, Mackay F, Grey S, Camps M, Rommel C, Gerondakis SD, Mackay CR. Positive regulation of immune cell function and inflammatory responses by phosphate PAC-1. *Nature Immunol* 2006;7:274–283. [PubMed: 16474395]

- Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci* 1992;8:275–282. [PubMed: 1633570]
- Kagan JC, Medzhitov R. Phosphoinositide-mediated adaptor recruitment controls toll-like receptor signaling. *Cell* 2006;125:943–955. [PubMed: 16751103]
- Krishnan J, Selvarajoo K, Tsuchiya M, Lee G, Choi S. Toll-like receptor signal transduction. *Exp Mol Med* 2007;39:421–438. [PubMed: 17934330]
- Kruithof EK, Satta N, Liu JW, Dunoyer-Geindre S, Fish RJ. Gene conversion limits the divergence of mammalian TLR1 and TLR6. *BMC Evol Biol* 2007;2007;7:148. [PubMed: 17727694]
- Lynch, M. *The origins of genome architecture*. Sinauer; Sunderland MA: 2007.
- Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, Kuroki Y. Comparative sequence analysis of leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics* 2007;2007;8:14. [PubMed: 17214878]
- Nei M. Gene duplication and nucleotide substitution in evolution. *Nature* 1969;221:40–42. [PubMed: 5782607]
- Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* 1986;3:418–426. [PubMed: 3444411]
- Nishimura M, Naito S. Tissue-specific mRNA expression profiles of human toll-like receptors and related genes. *Biol Pharm Bull* 2005;28:886–892. [PubMed: 15863899]
- Ohno, S. *Evolution by gene duplication*. Springer-Verlag; Berlin: 1970.
- Opsal MA, Våge DI, Hayes B, Berget I, Lien S. Genomic organization and transcript profiling of the bovine toll-like receptor gene cluster *TLR6-TLR1-TLR10*. *Gene* 2006;384:45–50. [PubMed: 16950576]
- Piatigorsky, J. *Gene sharing an evolution: the diversity of protein functions*. Harvard University Press; Cambridge MA: 2007.
- Pichlmair A, Reis e Sousa C. Innate recognition of viruses. *Immunity* 2007;27:370–383. [PubMed: 17892846]
- Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Adair A. The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci USA* 2005;102:9577–9582. [PubMed: 15976025]
- Rodriguez F, Oliver JF, Marí A, Medina JR. The general stochastic model of nucleotide substitution. *J Theor Biol* 1990;142:485–501. [PubMed: 2338834]
- Rzhetsky A, Nei M. A simple method for estimating and testing minimum-evolution trees. *J Mol Evol* 1992;35:367–375. [PubMed: 1404422]
- Sanghavi SK, Shankarappa R, Reinhart TA. Genetic analysis of Toll/interleukin-1 receptor (TIR) domain sequences from rhesus macaque Toll-like receptors (TLRs) 1-10 reveals high homology to human TLR/TIR sequences. *Immunogenetics* 2004;56:667–674. [PubMed: 15578265]
- Schmidt HA, Strimmer K, Vingron M, von Haeseler A. TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics* 2002;18:502–504. [PubMed: 11934758]
- Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 2007;24:1596–1599. [PubMed: 17488738]
- Tanji T, Ip YT. Regulators of the toll and imd pathways in the *Drosophila* innate immune response. *Trends Immunol* 2005;26:193–198. [PubMed: 15797509]
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Diggins DG. The CLUSTAL X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 1997;25:4876–4882. [PubMed: 9396791]
- Watters TM, Kenny EF, O'Neill LAJ. Structure, function and regulation of the Toll/IL-1 receptor adaptor proteins. *Immunol Cell Biol* 2007;85:411–419. [PubMed: 17667936]
- Wheeler DL, Barrett T, Benson DA, Bryant SH, Canese K, Chetvernin V, Church DM, DiCuccio M, Edgar R, Federhen S, Geer LY, Helmberg W, Kapustin Y, Kenton DL, Khovayko O, Lipman DJ, Madden TL, Maglott DR, Ostell J, Pruitt KD, Schuler GD, Schrimi LM, Sequeira E, Sherry ST, Sirotkin K, Souvorov A, Starchenko G, Suzek TO, Tatusov R, Tatusova TA, Wagner L, Yaschenko

E. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Research* 2006;34:D173–D180. [PubMed: 16381840]

Wiens M, Korzhev M, Perovi-Ottstadt S, Lthringer B, Brandt D, Klein S, Müller WE. Toll-like receptors are part of the innate immune defense system of sponges (Desmospongiae: Porifera). *Mol Biol Evol* 2007;24:792–804. [PubMed: 17190971]

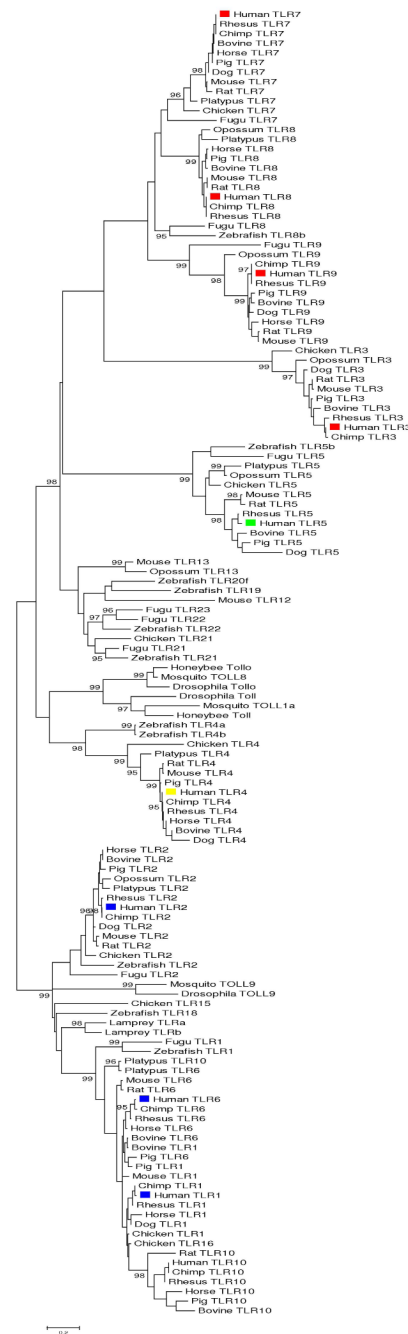


Figure 1. Minimum evolution (ME) tree of TIR domains (119 aligned amino acid sites) based on JTT distance, with Γ correction and shape parameter $a = 2.27$. Numbers on branches are confidence levels of the standard error test of the branch length; only values $\geq 95\%$ are shown. Color coding indicates PAMPs recognized by human TLRs: blue, lipopeptides; yellow, lipopolysaccharides; green, flagellin; red, nucleic acids.

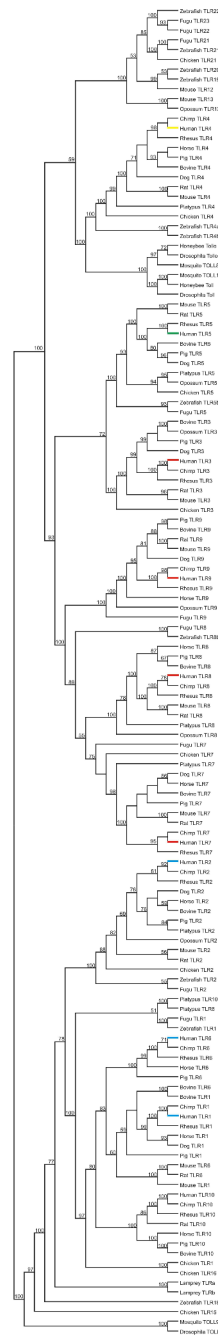
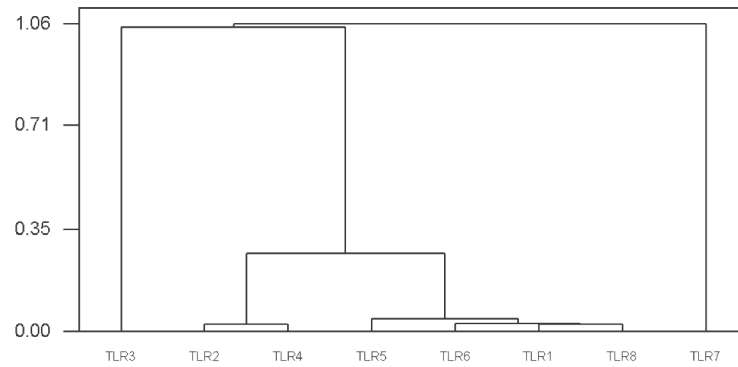


Figure 2. 50% majority-rule consensus Bayesian tree based on the JTT + Γ model. Only the topology, not branch lengths, is shown. Numbers on the branches are Bayesian posterior probabilities. Color coding is as in Figure 1.

A)



B)

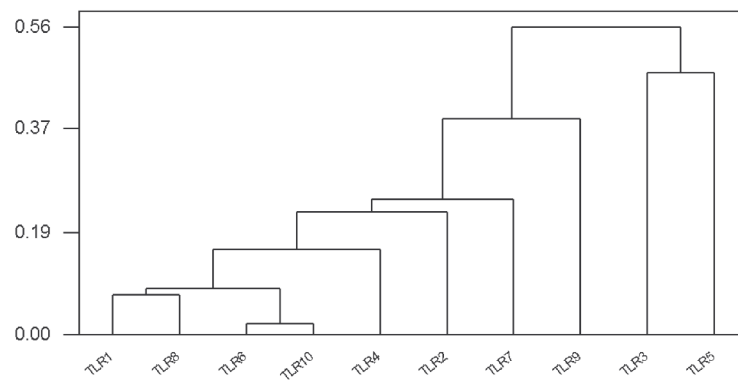


Figure 3. Cluster diagrams, based on the correlation matrix, of gene expression data for (A) the human leukocyte data set; and (B) the human tissue data set. In each case, the Y-axis is a distance based on 1 minus the correlation coefficient.

Table 1

Numbers of synonymous substitutions per synonymous site (d_S) and number of nonsynonymous substitutions per nonsynonymous site (d_N), with their standard errors, in LRR and TIR domains in comparisons of orthologous *TLR* gene pairs between human and rhesus monkey

Gene	LRR		TIR	
	$d_S \pm \text{S.E.}$	$d_N \pm \text{S.E.}$	$d_S \pm \text{S.E.}$	$d_N \pm \text{S.E.}$
<i>TLR1</i>	0.0584 \pm 0.0210	0.0088 \pm 0.0041	0.0712 \pm 0.0292	0.0116 \pm 0.0054
<i>TLR2</i>	0.0723 \pm 0.0163	0.0171 \pm 0.0045	0.0568 \pm 0.0252	0.0087 \pm 0.0050
<i>TLR3</i>	0.0917 \pm 0.0170	0.0232 \pm 0.0044	0.0592 \pm 0.0268	0.0118 \pm 0.0059
<i>TLR4</i>	0.0468 \pm 0.0117	0.0361 \pm 0.0063	0.0852 \pm 0.0315	0.0029 \pm 0.0027***
<i>TLR5</i>	0.0464 \pm 0.0134	0.0289 \pm 0.0059	0.0941 \pm 0.0340	0.0086 \pm 0.0050**
<i>TLR6</i>	0.0937 \pm 0.0244	0.0185 \pm 0.0055	0.1771 \pm 0.0478	0.0205 \pm 0.0078***
<i>TLR7</i>	0.0488 \pm 0.0109	0.0112 \pm 0.0028	0.0543 \pm 0.0245	0.0000 \pm 0.0000***
<i>TLR8</i>	0.0709 \pm 0.0139	0.0176 \pm 0.0036	0.0543 \pm 0.0245	0.0029 \pm 0.0029**
<i>TLR9</i>	0.1055 \pm 0.0159	0.0261 \pm 0.0044	0.1649 \pm 0.0413	0.0030 \pm 0.0030***
<i>TLR10</i>	0.0743 \pm 0.0188	0.0100 \pm 0.0036	0.0554 \pm 0.0250	0.0145 \pm 0.0065

Tests of the hypothesis that d_S or d_N in TIR equals the corresponding value in LRR:

**
P < 0.01

P < 0.001.