The evolution of vertebrate Toll-like receptors

Jared C. Roach*†, Gustavo Glusman*, Lee Rowen*, Amardeep Kaur*, Maureen K. Purcell‡§, Kelly D. Smith*¶, Leroy E. Hood*, and Alan Aderem*

*Institute for Systems Biology, Seattle, WA 98103; ‡School of Aquatic and Fishery Sciences, and ¶Department of Pathology, University of Washington, Seattle, WA 98195; and §Western Fisheries Research Center/U.S. Geological Survey, Seattle, WA 98115

Edited by Emil R. Unanue, Washington University School of Medicine, St. Louis, MO, and approved May 12, 2005 (received for review March 18, 2005)

The complete sequences of *Takifugu* Toll-like receptor (TLR) loci and gene predictions from many draft genomes enable comprehensive molecular phylogenetic analysis. Strong selective pressure for recognition of and response to pathogen-associated molecular patterns has maintained a largely unchanging TLR recognition in all vertebrates. There are six major families of vertebrate TLRs. This repertoire is distinct from that of invertebrates. TLRs within a family recognize a general class of pathogen-associated molecular patterns. Most vertebrates have exactly one gene ortholog for each TLR family. The family including *TLR1* has more species-specific adaptations than other families. A major family including *TLR11* is represented in humans only by a pseudogene. Coincidental evolution plays a minor role in TLR evolution. The sequencing phase of this study produced finished genomic sequences for the 12 *Takifugu rubripes* TLRs. In addition, we have produced >70 gene models, including sequences from the opossum, chicken, frog, dog, sea urchin, and sea squirt.

coincidental evolution | multigene family | concerted evolution

The Toll-like receptor (TLR) multigene family encodes important recognition receptors of the innate immune system that have been conserved in both the invertebrate and vertebrate lineages (1, 2). TLRs recognize a variety of endogenous and exogenous ligands; many of the latter are conserved molecules essential for pathogen survival. TLR genes have been recognized in a number of vertebrate genomes, and many partial and full-length sequences are available. Recent additions include draft predictions from the Japanese pufferfish *Takifugu rubripes* (3), the zebrafish *Danio rerio* (4–6), and the chicken *Gallus gallus* (7), and partially or fully sequenced mRNAs, including one from the goldfish *Carassius auratus* (8), several from the Japanese flounder *Paralichthys olivaceus* (9), and several from the rainbow trout *Oncorhynchus mykiss* (10). These papers provide incremental molecular phylogenetic analyses, and several reviews are available (11–13). Additionally, the draft genomes of the frog *Xenopus tropicalis*, chicken *G. gallus*, and opossum *Monodelphis domesticus* are now available. We present a complete molecular phylogenetic analysis of the known vertebrate TLR genes in the context of the complete genomic sequences of the *T. rubripes* TLRs.

Methods

Sequencing and Assembly. A draft genome sequence of *T. rubripes* was obtained by pairwise shotgun sequencing (14) through the efforts of an international collaboration (15). Sequence finishing was performed in part as described (16), with additional details provided in *Supporting Text*, which is published as supporting information on the PNAS web site.

Bioinformatics. TLRs were identified as genes coding for both an N-terminal leucine-rich repeat (LRR) domain and a C-terminal Toll-IL-resistance (TIR) domain. To form the basis of our study, vertebrate sequences from the nonredundant DDBJ/EMBL/NCBI database (GenBank) were identified by similarity to known TLRs (Data Set 1, which is published as supporting information on the PNAS web site). Amino acid sequence alignments were generated with CLUSTALX. Molecular distances and trees were computed by using PROTDIST from the PHYLIP package. Multidimensional scaling was performed as previously described (17). HMMER 2.3.2 was used to search for PFAM domains (hmmer.wustl.edu) (18). Synonymous/nonsynonymous substitution ratios were computed with PAML (19). Additional details, and information on draft genome predictions, are provided in *Supporting Text*.

Results

Molecular Tree. We constructed a molecular tree from all complete vertebrate TLRs in GenBank, including our recently added complete *Takifugu* sequences, and high-confidence gene models from the draft genome of *X. tropicalis* and *Monodelphis domestica* (Fig. 1). The multiple alignment supporting this tree (Fig. 4, which is published as supporting information on the PNAS web site) demonstrates that the major TLR families each have distinctive sequence characteristics. In particular, the TLR families vary considerably in the length of their leucine-rich extracellular domains. The extracellular domain of TLR1-family members is <600 amino acid residues, whereas TLR7-family members have an extracellular domain of >800 residues (see *Supporting Text*).

The molecular tree demonstrates six major families containing nearly all vertebrate TLRs, each drawn with a unique color in Fig. 1. TLRs within a family recognize a general class of pathogen-associated molecular pattern (PAMP) associated with that family. For convenience in this paper, we will refer to families by the lowest ordinal TLR contained in that family (e.g., we refer to the family containing TLR7–9 as the "TLR7 family").

An overview of the tree indicates that all of the families, and all of the genes within each family, are about equally distant from the center of the tree, where the progenitor vertebrate TLR gene or set of genes is inferred. This "star phylogeny" implies that all TLRs are evolving at about the same rate. This observation is somewhat unusual for multigene families, where often some members take on new functions; vertebrate TLRs are not fast-evolving genes. Furthermore, the discrepancies in molecular distances between species with shorter and longer generation times are relatively muted. This muting implies that selection is dominant over mutation in governing the rate of evolution of the TLRs, and thus that TLRs are under strong selection for maintenance of function. Even so, mutation is not completely eclipsed by selection, because the two TLRs most distant from the inferred ancestor are from the fast-generation murine lineage (mouse *TLR11* and *TLR12*).

Selective pressure presumably for maintenance of specific PAMP recognition has dominated the TLR2 subfamily (for lipopeptide), the TLR3 family (for dsRNA), the TLR4 family (for LPS) and the TLR5 family (for flagellin), and the TLR7–9 subfamilies (for nucleic acid and heme motifs). The evolution of genes in each of these clades recapitulates the phylogeny of species (Fig. 5, which is

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TLR, Toll-like receptor; PAMP, pathogen-associated molecular pattern; LRR, leucine-rich repeat; TIR, Toll-IL-resistance.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AC156430–AC156440).

†To whom correspondence should be addressed. E-mail: jroach@systemsbiology.org.

IMMUNOLOGY

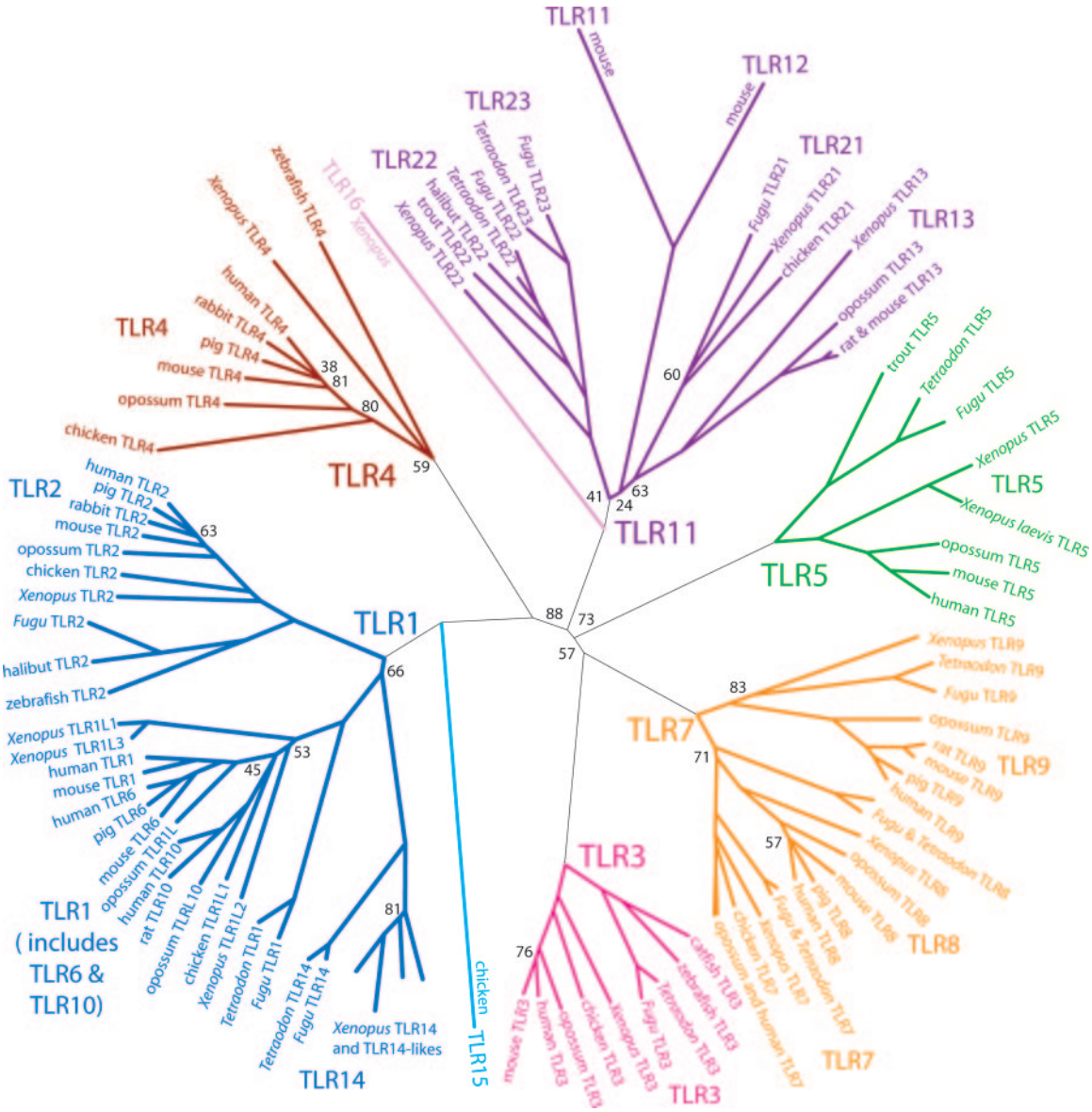

Fig. 1. Molecular tree of the vertebrate TLR. Branches of each major family are shown in a unique color. TLR16 may belong in the TLR11 family; TLR15 may belong in the TLR1 family. ''*Xenopus*'' without a species name indicates *X. tropicalis*. The species is listed along the branch for subfamilies with only one member. To avoid crowding, some known TLRs are not displayed. Bootstrap values <90 are shown as percentages; the preponderance of bootstrap values are 100%. The few low bootstrap values present tend to be associated with uncertainties in placement of very divergent TLRs (e.g., TLR16) or with very short branches.

published as supporting information on the PNAS web site). Most teleost vertebrates, including humans, have exactly one gene orthologous to each of these TLRs. There are occasional exceptions. *Takifugu* lacks TLR4. Chicken lacks TLR9 and possibly TLR5.

Amphibians and fish have a putatively soluble short form of the TLR5 gene (*TLRS5*) that arose by duplication from the LRR domains of TLR5. Although technically not a TLR, as it lacks a TIR domain, it is closely related to TLR5 and is often considered together with the TLRs in phylogenetic analysis. Because TLRS5 cannot be aligned across the full length of the TLR gene, we did not include it in our molecular distance calculations for Fig. 1. When included, TLRS5 is about as distant from TLR5 as TLR9 is from TLR8 and can thus be considered a subfamily of the TLR5 family.

The TLR7 family can be split into three subfamilies: TLR7–9. These TLR subfamilies recognize nucleic acid PAMPs. In addition, TLR9 may recognize heme derivatives (20). The divergence of the TLR7 family into three subfamilies occurred before the divergence of teleosts, because all teleosts appear to have a single ortholog for each of these subfamilies.

The TLR family specific for lipopeptide PAMPs includes *TLR1*, *TLR2*, *TLR6*, *TLR10*, and *TLR14*. The division of the TLR1 family into subfamilies occurred before the divergence of teleosts (Fig. 6, which is published as supporting information on the PNAS web site). Like the other TLR families, this family has also evolved under strong selection but has more species-specific adaptations than other families. The TLRs of the TLR1 family function as a heterodimeric receptor, with TLR2 paired with a member of one of the other TLR1 subfamilies. The TLR2 subfamily appears to operate under more selection constraints, because it has evolved following the phylogeny of species, with apparently no gene loss in any species. Intriguingly, the heterodimer mates of TLR2 appear to evolve under a freedom that many non-TLR multigene families enjoy, with expansions and contractions in gene number obscuring one-to-one orthologies. The TLR14 subfamily, present in fish,

appears to have been lost in amniotes but expanded in amphibians. Because of its relatedness to the TLR1 subfamily, we hypothesize that TLR14 also partners with TLR2. The chicken *TLR15* is molecularly distant from all other TLRs. It may be derived from the TLR1 family.

The remaining major family, including the TLR11–13, TLR21–23 subfamilies, is represented in humans only by a pseudogene. The major divisions of the TLR11 family are clearly very ancient, because most TLR11 subclades have representatives from fish and frogs. Enough sequences from mammals and birds are known to suggest that they, too, may be represented in many or all of these subclades. Little is known about the PAMPs for this family, but TLR11 apparently recognizes uropathogenic bacteria (21). The TLR16 subfamily, molecularly distant from all other TLRs and found only in *Xenopus*, may belong to the TLR11 family. The TLR11 family has more subfamilies than any other family, with diversity comparable to the TLR1 family. It also contains mouse *TLR11* and *TLR12*, the most divergent of all vertebrate TLRs. Thus the TLR11 family is perhaps under less purifying selective pressure than the other TLR families. The high divergence of *TLR11*, *TLR12*, and *TLR16* could conceivably obscure orthology to *TLR21*, *TLR22*, or *TLR23*. The similar number and diversity of subfamilies in the TLR11 family to that of the TLR1 family may indicate that the TLR11 family members function, analogously to the TLR1 family, as heterodimeric partners with each other.

It appears that, with few exceptions, vertebrates have at least one member gene representative from each of the six major TLR families. Where these families have major subfamilies, in many cases most, if not all, vertebrates have at least one representative.

Coincidental Evolution. Multigene families often evolve in ways that violate assumptions necessary for simple and objective gene phylogeny estimation. Their molecular clock may not be regular. In particular, some members of the family may evolve at much faster rates and as such are dubbed "fast-evolving genes." This happens when one member gene takes on a significantly novel function and thus encounters significantly different selective pressures from the other multigene family members. Vertebrate lactate dehydrogenase C is a classic example of a fast-evolving gene. Another usual assumption of molecular tree construction is that each branch of the tree evolves independently from the other branches. "Coincidental evolution" is a term describing phylogenies with branches that do not evolve independently. Multigene families often show coincidental evolution, either indirectly through biased mutational and selective forces or directly by mechanisms such as gene conversion (17). By comparing the molecular distance of pairs of paralogs present in different species with pairs of paralogs present in the same species, we can gain a sense of the amount of within-species coincidental evolution. Our analysis, detailed in *Supporting Text*, suggests that little if any coincidental evolution has occurred during the evolution of vertebrate TLRs, except perhaps between *TLR5* and *TLRS5*. This lack of coincidental evolution makes TLRs a textbook example of multigene evolution and an exception to the extensive coincidental evolution seen in most other multigene families of the immune system.

If there is not coincidental evolution, and TLRs evolve at a conservative and constant rate, then we can use a molecular clock to infer certain aspects of the timing of TLR evolution. We can infer that the divergence of the major families was more than twice as long ago as the divergence of fish and tetrapods. The major TLR families probably diverged during or before the Cambrian Period.

Evaluation of synonymous/nonsynonymous substitution ratios yielded no support for positive selection in the vertebrate TLR phylogeny (see *Supporting Text*).

Metazoan and Early Chordate TLR Evolution. For the most part, we have focused our attention on vertebrate TLRs. However, TLRs also exist in invertebrates (22). *Caenorhabditis elegans* and *Caenorhabditis briggsae* possess a TLR (23). Inamori *et al.* (24) sequenced a TLR cDNA from the horseshoe crab *Tachypleus tridentatus*. Azumi *et al.* (25) recognized TLRs in the urochordate sea squirt *Ciona intestinalis*. Also, for this paper, we have identified TLRs in the draft genomes of *Ciona savignyi* and the echinoderm sea urchin *Strongylocentrotus purpuratus* (Table 1). We were not able to identify TLRs in GenBank for nonteleost vertebrates such as sharks and lamprey. However, because TLRs are found in other vertebrates as well as other chordates, we expect that TLRs will be found in nonteleost vertebrates once a completely sequenced genome is available for rigorous study.

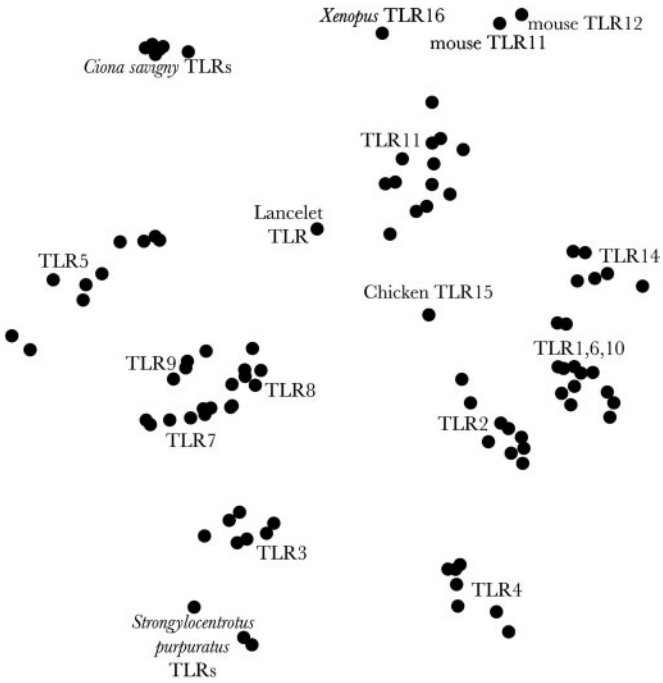

Fig. 2. Multidimensional scaling (MDS) of the molecular distances between TLRs. The distance between the gene families compared with the distances within the gene families is so great that portraying this information as a molecular tree could be misleading. Note that, like geographical maps of intercity distances, MDS representations have no axes. Not all TLRs are shown (e.g., *Strongylocentrotus* has hundreds of TLRs that cluster together).

Ecdysomes, such as nematodes and flies, appear to have at most a dozen or so TLRs. Likewise, a dozen is a typical complement of TLRs for a vertebrate. *C. intestinalis* appears to have only three, whereas *C. savignyi* has between 8 and 20. Strikingly, *Strongylocentrotus* has several hundred.

Construction of molecular phylogenies that include both nonvertebrate and vertebrate sequences is seldom possible and is fraught with peril (17). Great changes in selection pressure over time and between subphyla tend to invalidate most models of protein evolution that are used to compute molecular distances. Sequences diverge to an extent that reliable alignment is not possible. In cases where selection pressure is strong, molecular distances may saturate. These difficulties make it difficult to reliably assign orthology for members of multigene families between species of different subphyla.

However, molecular distance between such sequences may provide hints to relationships. We illustrate the relationships between the known urochordate, cephalochordates, and vertebrate TLRs in Fig. 2. We use multidimensional scaling to portray the relative molecular distances between genes. The distance between gene families is so great compared with the distance within each of them that portraying this information as a molecular tree could possibly be misleading. The large inter-

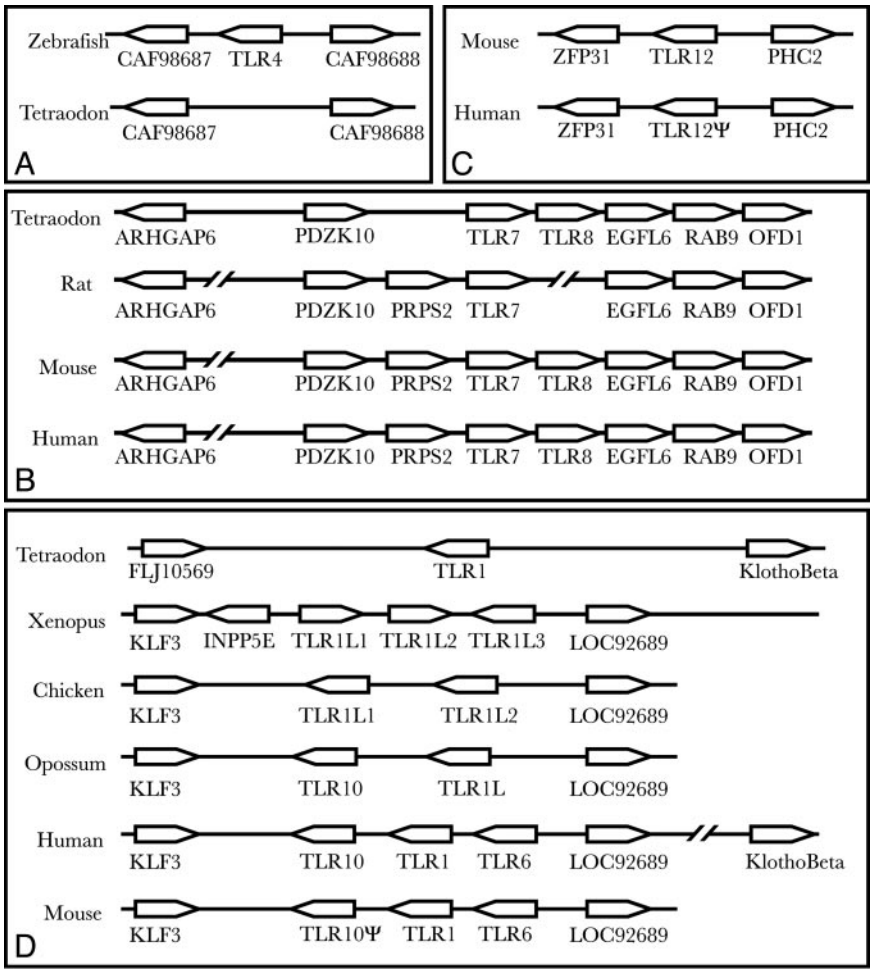

Fig. 3. Order and orientation of genes syntenic to (*A*) *TLR4*, (*B*) *TLR7* and *TLR8*, (*C*) *TLR12*, and (*D*) *TLR1*. For unfinished genomes, a small possibility exists that any gene portrayed as absent is actually present. Orthologs are identified by the Human Genome Organisation (HUGO) symbol of the human ortholog. Klotho Beta in humans is GeneID no. 152831. Genes are intentionally aligned in columns to facilitate visualization of synteny. Such alignments help confirm orthology. Select genomes are chosen to illustrate the dynamics of each locus.

family distances are inclusively either due to (*i*) extremely ancient divergence of the families, (*ii*) significant selection pressure that has pulled the families apart, or (*iii*) coincidental evolution tightening the clusters. The TLRs from *Ciona*, *Strongylocentrotus*, and lancelet all form tight clusters distinct from any of the vertebrate TLR clades. It is unlikely that one-to-one orthologies can ever be convincingly drawn between vertebrate and invertebrate TLRs.

The LRR domains of nonchordate TLRs are not reliably alignable with those of chordate TLRs, so phylogenetics must be based on alignments of the TIR domains. Nonchordate TLRs form multidimensionally scaled clusters distinct from the vertebrate TLRs (Fig. 7, which is published as supporting information on the PNAS web site). As expected, insect TLRs and the *C. elegans* TLR are more distant from the vertebrate than nonvertebrate chordate TLRs. Thus, even if there once was a one-to-one correspondence between a subset of contemporary vertebrate TLRs and contemporary invertebrate TLRs, the primary sequence divergence is now so great that there would no longer be any reason to suspect commonality of function even if orthology could be demonstrated with a technique such as syntenic analysis.

Conservation of Synteny. Many of the orthologous relationships of the TLRs can be confirmed by observations of conserved syntenies. Preservation between species of the order and orientation of orthologous genes also adds confidence to selections of noncoding sequence in searches for regulatory and other conserved elements.

TLR7 and *TLR8* are present as a tandem duplication in all genomes studied to date (Fig. 3*B*). The local gene order is preserved in humans and mice, but the rat genome has an assembly gap where *TLR8* would be anticipated. The gene order in *Tetraodon* is similar but lacks some of the genes in the mammalian locus. *TLR8* lies in tandem with *TLR7* in the chicken genome, but because the draft chicken *TLR8* locus has a gap where the TIR domain should lie, it may be a pseudogene.

Mouse *TLR12* is sandwiched between *ZFP31* and *PHC2* (Fig. 3*C*). Humans have a pseudogene in the orthologous position. Synteny is useful in demonstrating that this gene was once the ortholog to mouse *TLR12*. Dogs also have a pseudogene for *TLR12*, and one for *TLR11* as well. *TLR11* and *TLR12* are comparatively distant from all of the other TLRs, suggesting that they may be fast-evolving. If so, they may represent orthologs to *TLR21* and *TLR22*, for which no mammalian orthologs are known despite extensive searches, as part of our study, through all publicly available sequences.

The TLR1 subfamily of the TLR1 family also maintains syntenic relationships (Fig. 3*D*). Members of the TLR1 subfamily all lie adjacent to each other in every genome for which

sufficient data are available. Fish possess only one TLR1 gene; tetrapods appear to possess two to three TLR1 paralogs in tandem on their genomes. In humans and rodents, these are *TLR1*, *TLR6*, and *TLR10*. The opossum and chicken have only two such paralogs. The three *Xenopus* TLR1 subfamily paralogs are also in tandem.

The gene order in humans is *KLF3* followed by *TLR10*, *TLR1*, and *TLR6*, then *LOC92689*, a nameless but highly conserved gene. The gene order in *Xenopus* is *KLF3*, then *INPP5E* followed by *TLR10*, *TLR1*, and *TLR6*, then *LOC92689*. *INPP5E* in *Xenopus* thus appears to have recently moved into the locus. In mammals, *INPP5E* is found in a locus bounded by *CARD9* and *NOTCH1*. Significantly, the gene orientations in *Xenopus* do not correspond to those in mammals. Together, the lack of correspondence between gene orientations and the lack of correspondence between molecular distances (Fig. 1) suggests that the *Xenopus* TLR1-like genes are not in a one-to-one orthology with mammalian TLR1 genes and that local genome rearrangements have operated in this locus during tetrapod evolution. The TLR1-locus gene order in *Tetraodon* is not identical to that of tetrapods, but like the human locus retains the gene "Klotho Beta" on one of its flanks, indicating some maintenance of synteny.

Mouse *TLR10* is disrupted by two retroelements, a B1 SINE and an ERV LTR with identical intact LTRs. The B1 is unlikely to be older than 5–10 million years but probably not younger than 2–5 million years. The LTR is probably younger than 100,000 years and likely much younger (Arian Smit, personal communication).

Syntenic relationships can also add confidence to predictions of the absence of a gene ortholog from a genome. The conclusion of absence of *TLR4* from pufferfish (*T. rubripes* and *Tetraodon nigroviridis*) is supported by syntenic considerations (Fig. 3*A*). In zebrafish, *TLR4* is flanked on scaffold NA54426.1 by apparent orthologs to *Tetraodon CAF98687* and *CAF98688*. In *Tetraodon*, these two genes are separated by <3 kb on chromosome 17 with no evidence of a TLR between (see Fig. 3*A*). In an unlikely alternative, *TLR4* could conceivably be found in gaps in both pufferfish genome assemblies.

Yilmaz *et al.* (7) predict the absence of a chicken ortholog to *TLR9*. In *Xenopus*, *TLR9* lies between a cadherin and *TRIP*. In chicken, there is no gap between the orthologs of these genes. In humans, *TLR9* lies between *ALAS1* and *PTK9L*. In chicken, there is no gap between the orthologs of these genes. Therefore, *TLR9* is not found in the two most likely locations for a chicken *TLR9* gene. The best TBLASTN match in the chicken genome to the LRR of *Xenopus TLR9* or to its TIR domain is to chicken *TLR7*, again supporting the argument for the absence of *TLR9* from chicken.

All vertebrate species appear to have exactly one functional copy of *TLR2*. However, a TLR2 pseudogene appears upstream in tandem with the functional *TLR2* in opossum, dog, and human. Assuming the pseudogene is from a duplication event before the divergence of marsupials, its signature has been obliterated in mouse. We speculate that a constraint on genome rearrangement has allowed the pseudogene to recognizably persist. For example, regulatory regions 5′ to the pseudoTLR2 may be required by the functional *TLR2*.

Discussion

The TLR family members are capable of recognizing several classes of pathogens and orchestrating appropriate innate and adaptive immune responses. Lipoproteins in which the N-terminal cysteine is triacylated are recognized by TLR2 in combination with TLR1. Diacylated lipoproteins are recognized by TLR2 in combination with TLR6 (13, 26). Double-stranded RNA is recognized by TLR3. LPS is recognized by TLR4. Flagellin is recognized by TLR5. Cyclic compounds such as nucleic acids and heme are recognized by the family consisting of TLR7–9 (20, 27–30). It seems likely that each major family of TLRs represented in contemporary vertebrates is descended from a TLR for that class in the prototypical vertebrate and perhaps in more primitive metazoans.

We demonstrate six major families to which most vertebrate TLRs belong. TLRs within a family recognize a general class of PAMP associated with that family. Selective pressure for maintenance of specific PAMP recognition has clearly dominated the evolution of most of these families and often within subfamilies as well. Supporting this is the observation that most, but not all, vertebrates have exactly one gene orthologous to each of the TLR2–5 and TLR7–9 subfamilies. This assumes that sequenced genomes are representative of unsequenced genomes. The key prediction from our phylogenetic analyses is that strong selective pressure for recognition of specific classes of PAMPs has, and will maintain, a largely unchanging repertoire of TLR recognition in all vertebrates. The high conservation of TLRs is, in most cases, almost certainly because microbes cannot easily mutate their PAMPs (31). TLR gene dosage may also be under selection (*Supporting Text*). Some TLRs may recognize endogenous patterns, either absolutely or facultatively (32). For example, this is the case for *Drosophila* Tl (Toll). In these cases, high conservation may not be driven primarily by negative selection due to pathogens but could rather be driven by pressure to maintain an endogenous signaling network. Such pressure could include a need to conserve binding with accessory receptor molecules such as CD36, LY86 (RP105), and LY96 (MD2), or to conserve recognition of diffusible signaling intermediates such as cytokines.

The family specific for lipopeptide PAMPs consists of *TLR1*, *TLR2*, *TLR6*, *TLR10*, and probably *TLR14*. These PAMPs include zymosan, derived from yeast, and peptidoglycan, a surface component of Gram-positive bacteria. This family has also evolved under strong purifying selection but has more species-specific adaptations than other families. The remaining family, TLR11, has evolved faster and is represented in humans only by a pseudogene; its characterizing PAMP is unknown. The invertebrate repertoire of TLRs appears to have evolved under very different constraints.

Species undoubtedly adapt their immune defenses in a Red Queen's race with their pathogens. The evolutionary changes we see in the TLR repertoire may reflect changes in the spectrum of species-specific pathogens and their respective structural adaptations in PAMPs. To better answer the extent to which host TLRs coevolve with pathogens, a more comprehensive list of all pathogens and their molecular PAMPs for many species would be needed, ideally including pathogens encountered at various evolutionary epochs. It is unclear whether enough such data can ever be accumulated.

Tsujita *et al.* (10) demonstrated that rainbow trout *TLR5* recognizes bacterial flagellin. Rainbow trout possess two forms of *TLR5*, a membrane form and a putatively soluble form. The membrane-bound form was constitutively expressed, whereas the soluble form was induced after stimulation with flagellin. Tsujita *et al.* (10) propose that the soluble form may function analogously to LPS-binding protein and *CD14*, which are required for *TLR4* recognition of LPS in mammals.

Soluble LRR domains may function more widely as immune receptors. LeBouder *et al.* (33) report a soluble form of mammalian TLR2. A polymorphic form of *TLR5* could conceivably be secreted (34). Yilmaz *et al.* (7) report expressed alternatively spliced forms of chicken *TLR3* and *TLR5* without membrane spanning or TIR domains. Lamprey have a humoral immune response possibly based on variable lymphocyte receptors (VLRs). VLRs contain arrays of LRRs and may function in soluble as well as membrane-anchored forms (35). These VLRs might be derived from TLRs.

It is computationally difficult to detect genes that consist solely of LRRs, particularly if they are fast-evolving. Thus there may be

many more immunologically functional soluble LRR proteins to be found in vertebrates, including humans. Such proteins may be posttranslationally processed from membrane-bound forms or directly encoded in the genome. They might exist as polymorphic variants of membrane-bound genes. They may be evolutionarily derived from TLRs or have independent origins. We are not yet in a position to estimate how many such genes there might be.

Conclusion

The coding sequences and function of the vertebrate TLRs are highly conserved. Likewise, the signaling pathways initiated by TLRs are highly conserved (36, 37). Thus, TLRs are an example of evolutionary conservation of a biological system at multiple levels: gene, protein, and network. Comparative genomic analyses, such as those presented here, can play an important role in the identification of parts lists for systems biology (38).

Sydney Brenner inspired and led the *Fugu* Finishing Consortium. J.C.R. is supported by a grant from the National Institute of Allergy and Infectious Diseases (5K08AI056092). Rich Bonneau contributed to the discussion of LRR evolution. Brad Davidson contributed to the analysis of *Ciona* TLRs.

1. Aderem, A. & Ulevitch, R. J. (2000) *Nature* **406,** 782–787.
2. Medzhitov, R. & Janeway, C. A. (2000) *Immunol. Rev.* **173,** 89–97.
3. Oshiumi, H., Tsujita, T., Shida, K., Matsumoto, M., Ikeo, K. & Seya, T. (2003) *Immunogenetics* **54,** 791–800.
4. Jault, C., Pichon, L. & Chluba, J. (2004) *Mol. Immunol.* **40,** 759–771.
5. Meijer, A. H., Krens, S. F. G., Rodriguez, I. A. M., He, S. N., Bitter, W., Snaar-Jagalska, B. E. & Spaink, H. P. (2004) *Mol. Immunol.* **40,** 773–783.
6. Trede, N. S., Langenau, D. M., Traver, D., Look, A. T. & Zon, L. I. (2004) *Immunity* **20,** 367–379.
7. Yilmaz, A., Shen, S., Adelson, D. L., Xavier, S. & Zhu, J. J. (2005) *Immunogenetics* **56,** 743–753.
8. Stafford, J. L., Ellestad, K. K., Magor, K. E., Belosevic, M. & Magor, B. G. (2003) *Dev. Comp. Immunol.* **27,** 685–698.
9. Hirono, I., Takami, M., Miyata, M., Miyazaki, T., Han, H.J., Takano, T., Endo, M. & Aoki, T. (2004) *Immunogenetics* **56,** 38–46.
10. Tsujita, T., Tsukada, H., Nakao, M., Oshiumi, H., Matsumoto, M. & Seya, T. (2004) *J. Biol. Chem.* **279,** 48588–48597.
11. Kimbrell, D. A. & Beutler, B. (2001) *Nat. Rev. Genet.* **2,** 256–267.
12. Kanzok, S. M., Hoa, N. T., Bonizzoni, M., Luna, C., Huang, Y., Malacrida, A. R. & Zheng, L. (2004) *J. Mol. Evol.* **58,** 442–448.
13. Takeda, Y., Kaisho, T. & Akira, S. (2003) *Annu. Rev. Immunol.* **21,** 335–376.
14. Roach, J. C., Boysen, C., Wang, K. & Hood, L. (1995) *Genomics* **26,** 345–353.
15. Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J. M., Dehal, P., Christoffels, A., Rash, S., Hoon, S., Smit, A., *et al.* (2002) *Science* **297,** 1301–1310.
16. Glusman, G., Kaur, A., Hood, L. & Rowen, L. (2004) *BMC Evol. Biol.* **4,** 43.
17. Roach, J. C., Wang, K., Gan, L. & Hood, L. (1997) *J. Mol. Evol.* **45,** 640–652.
18. Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E. L., *et al.* (2004) *Nucleic Acids Res.* **32,** D138–D141.
19. Yang, Z. (1997) *CABIOS* **13,** 555–556.
20. Coban, C., Ishii, K. J., Kawai, T., Hemmi, H., Sato, S., Uematsu, S., Yamamoto, M., Takeuchi, O., Itagaki, S., Kumar, N., *et al.* (2005) *J. Exp. Med.* **201,** 19–25.
21. Zhang, D., Zhang, G., Hayden, M. S., Greenblatt, M. B., Bussey, C., Flavell, R. A. & Ghosh, S. (2004) *Science* **303,** 1522–1526.
22. Imler, J. L. & Zheng, L. (2004) *J. Leukocyte Biol.* **75,** 18–26.
23. Pujol, N., Link, E. M., Liu, L. X., Kurz, C. L., Alloing, G., Tan, M. W., Ray, K. P., Solari, R., Johnson, C. D. & Ewbank, J. J. (2001) *Curr. Biol.* **11,** 809–821.
24. Inamori, K., Koori, K., Mishima, C., Muta, T. & Kawabata, S. (2000) *J. Endotoxin Res.* **6,** 397–399.
25. Azumi, K., De Santis, R., De Tomaso, A., Rigoutsos, I., Yoshizaki, F., Pinto, M. R., Marino, R., Shida, K., Ikeda, M., Ikeda, M., *et al.* (2003) *Immunogenetics* **55,** 570–581.
26. Ozinsky, A., Underhill, D. M., Fontenot, J. D., Hajjar, A. M., Smith, K. D., Wilson, C. B., Schroeder, L. & Aderem, A. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 13766–13771.
27. Aderem, A. & Hume, D. A. (2000) *Cell* **103,** 993–996.
28. Diebold, S., Kaisho, T., Hemmi, H., Akira, S. & Sousa, C. R. (2004) *Science* **303,** 1529–1531.
29. Heil, F., Hemmi, H., Hochrein, H., Ampenberger, F., Kirschning, C., Akira, S., Lipford, G., Wagner, H. & Bauer, S. (2004) *Science* **303,** 1526–1528.
30. Wagner, H. (2004) *Trends Immunol.* **25,** 381–386.
31. Smith, K. D., Andersen-Nissen, E., Hayashi, F., Strobe, K., Bergman, M. A., Barrett, S. L., Cookson, B. T. & Aderem, A. (2004) *Nat. Immunol.* **4,** 1247–1253.
32. Rifkin, I. R., Leadbetter, E. A., Busconi, L., Viglianti, G. & Marshak-Rothstein, A. (2005) *Immunol. Rev.* **204,** 27–42.
33. LeBouder, E., Rey-Nores, J. E., Rushmere, N. K., Grigorov, M., Lawn, S. D., Affolter, M., Griffin, G. E., Ferrara, P., Schiffrin, E. J., Morgan, B. P., *et al.* (2003) *J. Immunol.* **171,** 6680–6689.
34. Hawn, T. R., Verbon, A., Lettinga, K. D., Zhao, L. P., Li, S. S., Laws, R. J., Skerrett, S. J., Beutler, B., Schroeder, L., Nachman, A., *et al.* (2003) *J. Exp. Med.* **198,** 1563–1572.
35. Pancer, Z., Amemiya, C. T., Ehrhardt, G. R., Ceitlin, J., Gartland, G. L. & Cooper, M. D. (2004) *Nature* **430,** 174–180.
36. Kim, D. H. & Ausubel, F. M. (2005) *Curr. Opin. Immunol.* **17,** 4–10.
37. Phelan, P. E., Mellon M. T. & Kim C. H. (2005) *Mol. Immunol.* **42,** 1057–1071.
38. Aderem, A. & Smith, K. D. (2004) *Semin. Immunol.* **16,** 55–67.