Adaptation and Constraint at Toll-Like Receptors in Primates

Gabriela Wlasiuk*†,[1] and Michael W. Nachman[1]

[1]Department of Ecology and Evolutionary Biology, University of Arizona

†Present address: Arizona Center for the Biology of Complex Diseases, Arizona Research Labs, University of Arizona.

***Corresponding author:** E-mail: wlasiuk@email.arizona.edu.

Associate editor: Takashi Gojobori

Abstract

Frequent positive selection is a hallmark of genes involved in the adaptive immune system of vertebrates, but the incidence of positive selection for genes underlying innate immunity in vertebrates has not been well studied. The toll-like receptors (TLRs) of the innate immune system represent the first line of defense against pathogens. TLRs lie directly at the host–environment interface, and they target microbial molecules. Because of this, they might be subject to frequent positive selection due to coevolutionary dynamics with their microbial counterparts. However, they also recognize conserved molecular motifs, and this might constrain their evolution. Here, we investigate the evolution of the ten human TLRs in the framework of these competing ideas. We studied rates of protein evolution among primate species and we analyzed patterns of polymorphism in humans and chimpanzees. This provides a window into TLR evolution at both long and short timescales. We found a clear signature of positive selection in the rates of substitution across primates in most TLRs. Some of the implicated sites fall in structurally important protein domains, involve radical amino acid changes, or overlap with polymorphisms with known clinical associations in humans. However, within species, patterns of nucleotide variation were generally compatible with purifying selection, and these patterns differed between humans and chimpanzees and between viral and nonviral TLRs. Thus, adaptive evolution at TLRs does not appear to reflect a constant turnover of alleles and instead might be more episodic in nature. This pattern is consistent with more ephemeral pathogen–host associations rather than with long-term coevolution.

Key words: toll-like receptors, molecular evolution, adaptive evolution, functional constraint, primates.

Research article

Introduction

Toll-like receptors (TLRs) recognize and bind conserved molecular motifs in pathogens to initiate an innate immune response and to prime the adaptive immune system (Akira and Takeda 2004). TLRs have received considerable attention recently because of the discovery of many polymorphisms in humans associated with susceptibility or resistance to both infectious and complex diseases, including autoimmune disorders (Lorenz et al. 2000; Hawn et al. 2003, 2005; Lazarus et al. 2004; Schroder and Schumann 2005; Johnson et al. 2007). TLRs are also interesting from an evolutionary point of view because they lie directly at the host–pathogen interface. Thus, they have the potential to be subject to coevolutionary dynamics. However, they have also been cited as an example of evolutionary conservation and strong functional constraint (Roach et al. 2005).

Although there is some overlap in the classes of ligands they recognize, TLRs expressed within endosomal compartments (TLR3, TLR7, TLR8, and TLR9) target predominantly viral components such as single- and double-stranded RNA and CpG DNA, whereas TLRs expressed in the cell membrane (TLR1, TLR2, TLR4, TLR5, TLR6, and TLR10) target predominantly bacterial (but also fungal and parasite) components such as lipopolysaccharide (LPS), peptidoglycan, and flagellin (Akira et al. 2006; Carpenter and O'Neill 2007). We will refer to these two subclasses as viral and nonviral TLRs. Viral and nonviral TLRs might be subject to different evolutionary pressures. Although vertebrate nucleic acids usually have chemical modifications that reduce the likelihood of activating TLRs (Kariko et al. 2005), they can sometimes induce an autoimmune response. Viral TLRs face the challenge of remaining functional while avoiding autoimmunity, and thus, we hypothesize that they are under stronger functional constraint than nonviral TLRs.

Despite several studies on the evolution of TLRs in humans and nonhuman primates, a clear picture of the evolution of this gene family has not emerged, in part because previous studies have generally focused on a subset of the TLRs or have been sampled within species or between species, but not both. For example, Ferrer-Admetlla et al. (2008) concluded that balancing selection is the best explanation for the pattern of sequence variation in a series of human innate immunity genes that included five TLRs. In contrast, Mukherjee et al. (2009) studied six TLRs and found no evidence of selection in a human population from India. They argued that purifying selection is the predominant force in TLR evolution in agreement with an earlier study of TLR4 (Smirnova et al. 2001). Barreiro et al. (2009) studied patterns of variation at all ten TLRs in three human populations and found no evidence of positive selection acting at most TLRs. At the interspecific level, Ortiz et al. (2008) failed to find evidence of positive selection at TLRs among five primate species (except for TLR1), but Nakajima et al. (2008), using a broader taxonomic sampling, reported that TLR4 has been under selection in Old World primates.

Population samples and interspecific comparisons provide information about evolutionary processes acting over

© The Author 2010. Published by Oxford University Press on behalf of the Society for Molecular Biology and Evolution. All rights reserved. For permissions, please e-mail: journals.permissions@oxfordjournals.org

different timescales. Population samples may provide evidence of very recent or population-specific selection. However, the history of pathogenic diseases during primate evolution undoubtedly played a role in shaping the present-day immune system, and the forces acting on immune genes over this deeper timescale can only be studied from interspecific comparisons.

Our goal was to provide a comprehensive picture of TLR evolution in primates over both short and long timescales. We gathered coding sequences for 8–11 primate species per gene from public databases to evaluate positive and negative selection across the primate phylogeny. We also sequenced both coding and noncoding regions of all ten TLRs in a population sample of western chimpanzees and analyzed these data in conjunction with published sequence data for the same genes in humans. In particular, we sought to 1) look for evidence of positive selection both within and between species, 2) compare the behavior of mildly deleterious polymorphisms in two closely related species (humans and chimpanzees) that differ in a number of population characteristics, and 3) investigate the idea that the "viral" and "nonviral" TLRs might display different patterns of molecular evolution.

We found compelling evidence of recurrent positive selection across primates, but very little evidence of positive selection within humans or chimpanzees. Humans had relatively more polymorphisms predicted to negatively affect protein function than did chimpanzees, consistent with a recent relaxation of constraint or smaller long-term effective population size in humans compared with chimpanzees. Viral TLRs were generally more constrained than nonviral TLRs as predicted by their more complex functional trade-offs.

Materials and Methods

Samples

DNA samples from 19 *Pan troglodytes verus* from the Y-Chromosome Consortium DNA collection were provided by Dr Michael Hammer at the University of Arizona. Human sequence data (24 African Americans and 23 European Americans) for the same genes sequenced in chimpanzees were gathered from the Innate Immunity Database (https://regepi.bwh.harvard.edu/IIPGA2/).

The sequences of the primate TLRs used in the phylogenetic analyses were taken from Genbank and Ensembl. For each TLR, a subset of 8–11 of the following species was used: *Homo sapiens*, *P. troglodytes*, *Gorilla gorilla*, *Pongo pygmaeus*, *Hylobates lar*, *H. pileatus*, *Cercocebus torquatus*, *Macaca mulatta*, *Saguinus oedipus*, *S. fuscicollis*, *Callithrix jacchus*, *Aotus nancymaae*, *Tarsius syrichta*, *Microcebus murinus*, and *Otolemur garnetti*. The species used for each gene and the accession numbers are presented in supplementary table S1, Supplementary Material online.

DNA Sequencing

For TLR1–TLR4 and TLR6–TLR10, the coding region and a noncoding fragment of comparable length (total ~4–5 kb) were polymerase chain reaction (PCR) amplified and sequenced in 19 *P. t. verus*. For TLR5, sequence data from Wlasiuk et al. (2009) were used. PCR was performed in 25–50 μl reactions using Platinum *Taq* High Fidelity DNA Polymerase (Invitrogen, San Diego, CA). PCR products were purified using the Qiagen PCR purification kit (Qiagen, Valencia, CA) and sequenced using an ABI 3700 automated sequencer (Applied Biosystems, Foster City, CA). Amplification and sequencing primers are provided in supplementary table S2, Supplementary Material online. Sequences have been deposited in GenBank under the following accession numbers: TLR1 (GQ343345–GQ343363), TLR2 (GQ343364–GQ343382), TLR3 (GQ343383–GQ343401), TLR4 (GQ343402–GQ343420), TLR6 (GQ343421–GQ343439), TLR7 (GQ343440–GQ343458), TLR8 (GQ343459–GQ343477), TLR9 (GQ343478–GQ343494), and TLR10 (GQ343495–GQ343512).

The human data from the Innate Immunity Database consists of the complete resequencing of all exons (including 5′ and 3′ UTRs), some intronic sequence, and flanking regions.

Sequence editing and assembly were performed using SEQUENCHER (Gene Codes, Ann Arbor, MI). DNA sequences were aligned using CLUSTAL X (Thompson et al. 1997). Primate DNA sequence alignments were adjusted based on the protein sequence using the RevTrans Web server (http://www.cbs.dtu.dk/services/RevTrans/).

Codon-Based Analyses of Positive Selection

To evaluate positive and negative selections at all the TLRs during primate evolution, we compared the rate per site of nonsysnonymous substitution (*dN*) to the rate per site of synonymous substitutions (*dS*) in a maximum likelihood (ML) framework. A ratio of $dN/dS > 1$ is interpreted as strong evidence of positive selection, whereas a $dN/dS < 1$ is evidence of purifying selection.

We tested for positive selection at individual codons in primate samples that include 8–11 species per gene including human, apes, Old World primates, New World primates, and prosimians. For each gene, a neighbor joining or ML tree was used as the working topology. With the exception of a couple of misplaced or unresolved branches, these trees were the same as the accepted phylogeny for these species (Bininda-Emonds et al. 2007).

We implemented two alternative models in CODEML (PAML version 4) (Yang 1997, 2007), one of which (M7) only allows codons to evolve neutrally or under purifying selection (dN/dS values ≤ 1) and one which (M8) adds a class of sites under positive selection with $dN/dS > 1$. The two previous nested models were compared using a likelihood ratio test (LRT) with 2 degrees of freedom (Nielsen and Yang 1998; Yang et al. 2000). To ensure convergence, all analyses were run twice, with starting values of dN/dS of 0.5 and 1.5. For all the analyses, we assumed the F3x4 model of codon frequencies. Amino acids under selection for model M8 were identified using a Bayes empirical Bayes approach (BEB) (Yang et al. 2005). Because it is inherently more difficult to identify specific sites under selection than to show that a proportion of sites is under

selection, we considered as candidates sites with a posterior probability >90%.

Next, a series of ML methods proposed by Kosakovsky Pond and Frost (2005) were implemented in the DATAMONKEY Web server (Pond and Frost 2005). The single likelihood ancestor counting (SLAC) model is based on the reconstruction of the ancestral sequences and the counts of synonymous and nonsynonymous changes at each codon position in a phylogeny. The fixed-effect likelihood (FEL) model estimates the ratio of nonsynonymous to synonymous substitution on a site-by-site basis, without assuming an a priori distribution of rates across sites. The random effect likelihood (REL) model first fits a distribution of rates across sites and then infers the substitution rate for individual sites. FEL and REL have the advantage that they can improve the estimation of the dN/dS ratio by incorporating variation in the rate of synonymous substitution (Pond and Muse 2005). Because a reduced number of sequences typically tends to result in a high false-positive rate, we used more stringent significance thresholds than the ones suggested by simulation to correspond to true type I errors of ~0.5% (Kosakovsky Pond and Frost 2005). We accepted sites with P values <0.1 for SLAC and FEL, and Bayes Factor >50 for REL as candidates for selection.

For the sites identified as under selection by more than one ML method, the amino acid changes were mapped onto the phylogeny by parsimony using MacClade (Sinauer Associates, Sunderland, MA). Crystal structures or theoretical models were used, when available, to map these residues onto the protein three-dimensional structures using the NCBI application Cn3D.

To explore possible heterogeneity in dN/dS among lineages, we ran "free-ratio" models in CODEML (PAML version 4) that allow each branch to have a separate dN/dS value while keeping variation among sites constant (Yang 1998).

Population Genetic Analyses

For both the human and the chimpanzee data sets, we estimated nucleotide heterozygosity, π (Nei and Li 1979), and the proportion of segregating sites, θ_w (Waterson 1975). The script "compute" from the libsequence library (Thornton 2003) was used to calculate Tajima's D (Tajima 1989), Fu and Li's D^* and F^* (Fu and Li 1993), and Fay and Wu's H (Fay and Wu 2000). These statistics evaluate deviations of the allele frequency spectrum from those expected under neutrality. Coalescent simulations, conditioned on the observed number of segregating sites, were used to generate the null distributions of these test statistics in DnaSP version 5 (Librado and Rozas 2009).

We quantified the amount of differentiation between human populations (African Americans, European Americans) using F_{ST} calculated for each gene as $(\pi_T - \pi_W)/\pi_T$, where π_T is the nucleotide diversity for both populations combined and π_W is the average nucleotide diversity within populations. To obtain significance values, we generated an empirical distribution using 23 genes in the Seattle SNPs database sampled in the same individuals.

The extent of linkage disequilibrium (LD) associated with particular variants in the human HapMap data was evaluated with the integrated haplotype score (iHS) statistic (Voight et al. 2006). Using Haplotter (http://haplotter.uchicago.edu/selection/), we screened windows of 50 single nucleotide polymorphisms (SNPs) centered on each gene, looking for an accumulation of SNPs with $|iHS| > 2$, as in Voight et al. (2006).

Levels of polymorphism and divergence were contrasted in two ways. First, the ratio of nonsynonymous to synonymous polymorphisms within humans and within chimpanzees was compared with the ratio of nonsynonymous to synonymous fixed differences between each of the species and macaque (McDonald and Kreitman 1991). Second, the ratio of polymorphism to divergence was compared between each TLR and a control set of genes using human–chimpanzee divergence (Hudson et al. 1987) with the software HKA (http://lifesci.rutgers.edu/~heylab/). In the case of humans, this control set consisted of 50 concatenated noncoding segments from Yu et al. (2002). In the case of chimpanzees, the control set consisted of 26 concatenated noncoding segments from Fischer et al. (2006). For both species, the concatenated sets of viral and nonviral TLRs were compared against each other. The use of the macaque sequence as the interspecific comparison in the McDonald and Kreitman (M-K) test provided more power due to increased divergence. In the Hudson, Kreitman, and Aguade (HKA) test, however, we used human–chimpanzee divergence because the lower divergence resulted in more reliable alignments over long regions on noncoding DNA. In the HKA comparisons, the lower divergence was offset by the use of longer sequences.

Prediction of Deleterious Polymorphisms in Humans and Chimpanzees

To identify sites under purifying selection within species, we predicted the functional consequences of human and chimpanzee polymorphisms using a method described by Sunyaev et al. (2001) and implemented in the Polyphen Web server (http://genetics.bwh.harvard.edu/pph/). Polyphen uses a combination of structural information, sequence annotation, and patterns of sequence conservation among species to classify polymorphisms as "benign" (no predicted effect on protein function), "possibly damaging" (weak evidence of an effect on function), or "probably damaging" (strong evidence of a functional effect). We recognize, however, that some (presumably a small fraction) of the amino acid changes predicted by Polyphen as "damaging" could actually improve protein function.

Relative Levels of Purifying Selection among Genes and Protein Domains

To assess the relative levels of functional constraint among the genes and the different protein domains (signal peptide, leucine-rich repeat domain [endosomal or extracellular],

transmembrane domain, and cytoplasmic domain), we estimated the global *dN*/*dS* for each gene and domain separately using the M_0 site model (no variation among branches or sites) in CODEML. We used the domains inferred by Matsushima et al. (2007). We also estimated the *dN*/*dS* ratio of the human and chimpanzee lineages separately using the macaque sequence as an outgroup.

Results

Inference of Positive Selection from Substitution Patterns

Using ML approaches, we addressed whether recurrent positive selection has been common in the TLR gene family. First, we compared nested models with and without positive selection using LRTs and found that for six of the ten genes (TLR1, TLR4, TLR6, TLR7, TLR8, and TLR9), a model that includes sites with $dN/dS > 1$ fits the data significantly better than a neutral model (table 1). This group of six genes contains an equal number of viral and nonviral receptors. For each of these six genes, the proportion of sites under selection according to the M8 model was relatively low. The specific codons identified by the BEB approach with a posterior probability of 90% constitute an even smaller fraction of that proportion (table 1).

The other ML methods also detected sites under selection for the six genes, some of which coincide with the codons previously identified by M8. To identify robust candidates for sites under selection, we considered sites with evidence of selection in at least two of the ML methods. Each of the six genes presents at least one site that was concordant among methods (table 1).

TLR4 stands out because the proportion of selected sites under M8 (15% with a *dN*/*dS* ratio of ~2.4) is the highest among the six positively selected genes. Using the data set from Nakajima et al. (2008), which consists of a smaller fragment (~600 bp) of the extracellular domain in 20 primate species, we repeated the analyses above and also rejected a neutral model in favor of a model with selection. Several of the putative sites under selection are shared between the two data sets (table 1).

To gain insight into the functional significance of the putatively selected sites, we looked at the location of all the sites identified by ML methods in three-dimensional structures (crystal structures or theoretical models) when available. For most TLRs, we found several sites that fall in or immediately adjacent to regions or residues postulated to affect function (table 2). Figure 1 shows the location of the selected residues in the crystal structures of the extracellular domains of TLR4 (which forms a homodimer) and of TLR1 (which forms a heterodimer with TLR2). The evidence for positive selection is particularly strong for these two genes. For both, numerous sites are identified as positively selected by different methods (table 1). Moreover, some of these sites are known to participate in dimerization or ligand binding.

To examine the phylogenetic distribution of the inferred positively selected changes among the main primate clades (lemurs, New World primates, Old World primates, and apes), we mapped the unambiguous amino acid substitutions onto the phylogeny only including sites that were implicated in positive selection in two or more methods (table 1). We compared the observed and expected counts for each clade, where the expected values were obtained by multiplying the number of unambiguous changes in a clade by its relative divergence time (sum of all branches in a clade divided by the sum of all branches in the entire phylogeny). TLR8 was not included in this analysis because of the low number of unambiguous amino acid changes and the lack of a New World primate sequence for that gene. For four of the five remaining genes, the phylogenetic distribution of positively selected substitutions did not differ significantly from the null model. At TLR4, however, we found an excess of positively selected changes in Old World primates ($P = 0.004$, data not shown) and more specifically in the *C. torquatus* branch, where 5 of the 31 nonambiguous changes fall (fig. 2). For TLR4 therefore, we also investigated models that allow the *dN*/*dS* ratio to vary among lineages. We found that the best-fit model that accommodates heterogeneity in the rates of protein evolution had five different rates (data not shown). Although not significantly better than the five-rate model, the most complex model that assigns a different rate to every branch in the phylogeny helps to evaluate rate changes in specific lineages. Figure 2 shows the lineage-specific *dN*/*dS* values on the TLR4 phylogeny. In line with the observed accumulation of positively selected sites in *Catarrhini* (the clade that groups Old World primates, apes, and humans), four branches within that clade had *dN*/*dS* values above 1.

Levels and Patterns of Variation within Species

In chimpanzees, the nucleotide heterozygosity per site (π) for the coding and noncoding regions together ranged between 0.03% and 0.07% (supplementary table S3, Supplementary Material online), with individual values similar to reported genome-wide averages (Yu et al. 2003; Fischer et al. 2006). For the coding sequences, the levels of polymorphism were generally lower, with π values between 0.01% and 0.06% (supplementary table S3, Supplementary Material online). In humans, the polymorphism levels in the combined coding and noncoding regions (Africans 0.03–0.23%, Europeans 0.03–0.12%; supplementary table S3, Supplementary Material online) were unremarkable and similar to genome-wide patterns (Akey et al. 2004).

We used several statistics to assess departures from a neutral model of evolution in the distribution of allele frequencies. Tajima's *D* compares the number of polymorphisms with the mean pairwise difference between sequences (Tajima 1989). Fu and Li's *D** and *F** compare the number of derived singletons with two different estimators of the overall derived polymorphism (Fu and Li 1993). Fay and Wu's *H* compares the number of low- and high-frequency polymorphisms with the number of intermediate frequency polymorphisms (Fay and Wu 2000).

In chimpanzees, Tajima's *D* did not differ significantly from neutral expectations for any of the genes, and only

Table 1. Phylogenetic Tests of Recurrent Positive Selection.

Gene	No. of Species	Test of Selection[a]					Sites under Selection Identified by Different Methods[b]			
		lnL M7 (neutral)	lnL M8 (selection)	$-2\ln\Delta L$[c]	Significance	p_s, ω_s[d]	PAML M8[e]	SLAC[f]	FEL[g]	REL[h]
TLR1	11	−7,830.4	−7,820.62	19.56	**	0.05, 3.08	61, 106, 174, 321, 392, 396, 466	66, 174	49, 174, 236, 308, 313, 321, 351, 584, 621, 626, 649	34, 49, 66, 73, 174, 177, 203, 236, 248, 289, 293, 308, 313, 321, 345, 346, 370, 401, 417, 458, 466, 540, 574, 584, 621, 626, 649, 653, 663
TLR2	11	−6,674.56	−6,672.78	3.56	NS				24, 220, 354, 475	24, 32, 37, 52, 63, 127, 177, 185, 220, 221, 235, 267, 270, 275, 276, 321, 324, 326, 331, 354, 376, 390, 403, 424, 453, 475, 490, 500, 636, 770, 771
TLR3	8	−6,108.17	−6,108.18	−0.02	NS				79, 715	7, 79, 86, 356, 715
TLR4	11	−8,156.32	−8,142.58	27.48	**	0.15, 2.39	139, 204, 297, 298, 299, 319, 321, 322, 327, 351, 354, 437, 471, 496, 514, 520, 537, 542, 544, 611		204, 300, 319, 323, 327, 368, 437, 471, 475, 514, 542, 606, 639	75, 96, 139, 184, 186, 201, 204, 216, 229, 269, 271, 274, 292, 295, 296, 297, 298, 299, 300, 308, 319, 321, 322, 324, 327, 331, 349, 351, 365, 368, 371, 394, 402, 410, 415, 423, 437, 450, 460, 468, 471, 474, 475, 487, 494, 496, 505, 514, 517, 520, 521, 533, 537, 542, 544, 561, 566, 606, 611, 616, 626, 639, 673, 833
TLR4 b[i]	20	−2,629.41	−2,616.18	26.46	**	0.09, 3.74	229, 295, 319, 321, 322, 349, 360	360	204, 295, 308, 319, 323, 360, 368	229, 319, 322, 360
TLR5	11	−7,560.73	−7,559.07	3.32	NS				400, 407, 567	
TLR6	11	−6,652.14	−6,647.37	9.54	**	0.03, 3.40	293, 470, 471		2, 589	2, 72, 118, 134, 186, 293, 308, 315, 350, 406, 421, 439, 470, 471, 570, 579, 589, 626
TLR7	9	−6,827.06	−6,823.54	7.04	**	0.05, 9.35	486, 542, 693		528, 542	2, 37, 39, 42, 43, 44, 111, 113, 218, 233, 239, 283, 307, 341, 364, 421, 455, 456, 457, 462, 486, 487, 490, 496, 514, 517, 520, 528, 542, 566, 597, 637, 684, 696, 697, 700, 737, 826, 856, 944
TLR8	9	−8,401.8	−8,394.55	14.5	**	0.03, 4.20	159, 225, 237, 469, 738, 765		267, 522, 738, 783	

Table 1. Continued

Gene	No. of Species	lnL M7 (neutral)	Test of Selection[a] lnL M8 (selection)	$-2\ln\Delta L$[c]	Significance	p_s, ω_s[d]	Sites under Selection Identified by Different Methods[b] PAML M8[e]	SLAC[f]	FEL[g]	REL[h]
TLR9	9	−7,787.99	−7,776.32	23.34	**	0.03, 3.57	522, 646, 649, 674, 864	649	278, 311, 522, 649, 863, 864	58, 91, 186, 278, 322, 443, 467, 522, 649, 674, 675, 753, 863, 864
TLR10	8	−4,746.47	−4,746.46	0.02	NS					261

NS, not significant.

[a] Two alternative nested models, one "neutral" and the other including one class of sites with $dN/dS > 1$, were compared in a LRT.

[b] Codons identified by more than one ML method are underlined. See Materials and Methods for details on each of the methods. Sites with clinical associations in humans are shown in shadow.

[c] $-2\ln\Delta L$ is distributed approximately as χ^2 with 2 degrees of freedom.

[d] p_s, proportion of the sites under selection and ω_s, estimated dN/dS of the sites under selection in M8.

[e] Codons with posterior probabilities >90% in the BEB analyses.

[f] Codons with P values <0.1.

[g] Codons with P values <0.1.

[h] Codons with Bayes factors >50.

[i] Data set from Nakajima et al. (2008), fragment of ~600 nt.

**p < 0.01.

three genes (TLR1, TLR2, and TLR6) showed significant deviations in any of these four statistics (supplementary table S3, Supplementary Material online). TLR6 was the most striking case, with a significant excess of low- and high-frequency derived variants in coding and noncoding regions, a pattern expected during or after a selective sweep. However, examination of the table of polymorphism (supplementary table S4, Supplementary Material online) reveals that the excess of rare variants is due to the presence of a unique divergent haplotype that carries three of the six singletons. Although Fay and Wu's *H* is relatively insensitive to demography, specific demographic scenarios can result in an excess of high-frequency variants (i.e., when only a few individuals migrate between two divergent populations) (Fay and Wu 2000). Gene flow between chimpanzee subspecies is probably rare, but introgression has been described previously (e.g., Won and Hey 2005), and it seems a plausible explanation for the observed pattern.

In humans, the distributions of allele frequencies (supplementary table S3, Supplementary Material online) were in overall agreement with the accepted demographic history of Africans and Europeans (i.e., a population expansion in Africans that resulted in greater numbers of rare polymorphisms, and a bottleneck in Europeans that resulted in greater number of intermediate frequency polymorphisms). To take population-level effects into account, we compared Tajima's *D* at each TLR to the empirical distribution of Tajima's *D* in 132 genes sampled in the same individuals (Akey et al. 2004). The same was done for Fu and Li's *D** and *F** and Fay and Wu's *H*. In spite of the observation of several significant values (supplementary table S3, Supplementary Material online), with the exception of TLR10, most of the genes do not seem remarkable in the context of the genome-wide distributions of these statistics (fig. 3). TLR10 shows a departure from neutrality in Europeans for all four statistics, and the deviation is in the opposite direction of that expected from the demographic history of Europeans. Similar to TLR6 in chimpanzees, the excess of low- and high-frequency derived mutations at TLR10 in humans is caused by a divergent haplotype that is present in only one copy in Europeans but is relatively frequent in Africans (supplementary table S5, Supplementary Material online), suggesting that population structure and migration may explain the observed patterns.

Selection can act to maintain the same alleles in different populations or to fix different alleles in a population-specific manner, leading to very low or very high population differentiation, respectively. We estimated F_{ST} between human populations for each TLR and compared this with the empirical distribution of F_{ST} for all the genes in the Seattle SNPs database (supplementary table S6, Supplementary Material online). None of the TLRs fall in the lower or upper 5% of the distribution. TLR9 showed the lowest level of differentiation among TLRs ($F_{ST} = 0.014$), but approximately 9% of the genes have lower F_{ST} values than TLR9. TLR1 had the highest F_{ST} value (0.085), but this was close to the genome-wide average (0.07).

Table 2. Sites Predicted to Affect Function Based on Their Location in the Three-Dimensional Structure.

Gene	Position[a]	Functional Information[b]	Site Identified by[c]	Reference Three-Dimensional model
TLR1	284	Adjacent to site involved in ligand binding	M8 (<0.9)	Jin et al. (2007)
	303	Adjacent to site involved in ligand binding	M8 (<0.9)	
	308	Adjacent to site involved in ligand binding	REL, FEL, M8 (<0.9)	
	313[d]	Ligand binding	REL, FEL, M8 (<0.9)	
	321[d]	Dimerization surface (ionic interaction)	REL, FEL, M8	
	337[d]	Dimerization surface (ionic interaction and hydrogen bond) and adjacent to site involved in ligand binding	M8 (<0.9)	
TLR2	267	Adjacent to site involved in ligand binding	REL, M8 (<0.9)	Jin et al. (2007)
	294[d]	Ligand binding	M8 (<0.9)	
	296	Adjacent to site involved in ligand binding	M8 (<0.9)	
	270[d]	Ligand binding	M8 (<0.9)	
	318[d]	Dimerization surface (ionic interaction) and adjacent to site involved in ligand binding	M8 (<0.9)	
	321[d]	Dimerization surface (ionic interaction)	REL, M8	
	324[d]	Dimerization surface (hydrophobic interaction)	REL, M8 (<0.9)	
	326[d]	Dimerization surface (hydrophobic interaction)	REL, M8 (<0.9)	
	329	Adjacent to site involved in ligand binding	M8 (<0.9)	
	338[d]	Ligand binding	M8 (<0.9)	
	354	Adjacent to site involved in ligand binding	FEL, REL, M8 (<0.9)	
	373[d]	Dimerization surface (hydrophobic interaction)	M8 (<0.9)	
	376	Adjacent to site involved in dimerization surface (hydrogen bond)	REL, M8	
TLR3	86[d]	Ligand binding	REL, M8 (<0.9)	Liu et al. (2008)
TLR4	295	Adjacent to site involved in ligand binding	SLAC, FEL, REL, M8 (<0.9)	Park et al. (2009)
	296[d]	Ligand binding	REL, M8 (<0.9)	
	297	Adjacent to site involved in ligand binding	REL, M8	
	339	Adjacent to site involved in interaction with MD2 (hydrogen bond)	REL, M8 (<0.9)	
	341[d]	Ligand binding and interaction with MD2 (ionic interaction)	REL, M8 (<0.9)	
	415	Adjacent to site involved in interaction with MD2 (hydrogen bond)	REL, M8 (<0.9)	
	437	Adjacent to site involved in ligand binding and interaction with MD2	FEL, REL, M8	
TLR6	315[d]	Dimerization surface	REL, M8 (<0.9)	Jin et al. (2007)
TLR8	491	Adjacent to predicted site involved in ligand binding	M8 (<0.9)	Wei et al. (2009)
TLR9	484	Adjacent to predicted site involved in ligand binding	M8 (<0.9)	Wei et al. (2009)

[a] Relative to human protein sequence.
[b] Based on structural information from crystallographic studies or homology modeling.
[c] Sites identified by model M8 in CODEML but with posterior probabilities <0.90 (not reported in table 1) were included here.
[d] Sites more likely to affect function based on location in protein structure.
Sites identified by more than one method from table 1 are underlined.

Another signature of recent positive selection is the presence of an extended haplotype at relatively high frequency, associated with a selected allele. We tested for long-range LD in the phase II HapMap data using the iHs (Voight et al. 2006). iHs is based on the ratio of the integrated haplotype homozygosities (the area under the curve of a extended haplotype homozygosity by distance plot) of the ancestral and derived alleles at a specific SNP. None of the TLRs display an unusual accumulation of SNPs with high iHs, as would be expected under ongoing or recent selection.

Analyses of Polymorphism and Divergence

The ratio of replacement to silent polymorphism within humans or chimpanzees was compared with the ratio of replacement to silent fixed differences with macaque. None of the genes, individually, nor combined, deviated significantly from the neutral expectation of equal replacement to silent ratios within and between species (table 3). Several genes, however, show a slight excess of replacement polymorphisms with respect to fixations. This can be summarized using the neutrality index (NI), a ratio of the replacement to silent ratios within and between species (Rand and Kann 1996). NI values above 1 indicate an excess of replacement changes within species, whereas values between 0 and 1 indicate an excess of replacement fixations between species. The average NI for nonviral TLRs (1.41) is higher than the average value for viral TLRs (1.15; table 3). This is consistent with the pattern reported by Barreiro et al. (2009) for different populations using a modification of the M-K framework. Most of these replacement polymorphisms are at low frequency, resulting in average values of Tajima's *D* that are slightly but not significantly more negative for replacement sites than for silent sites in humans and chimpanzees (data not shown), as previously reported for other genes (Hughes et al. 2003).

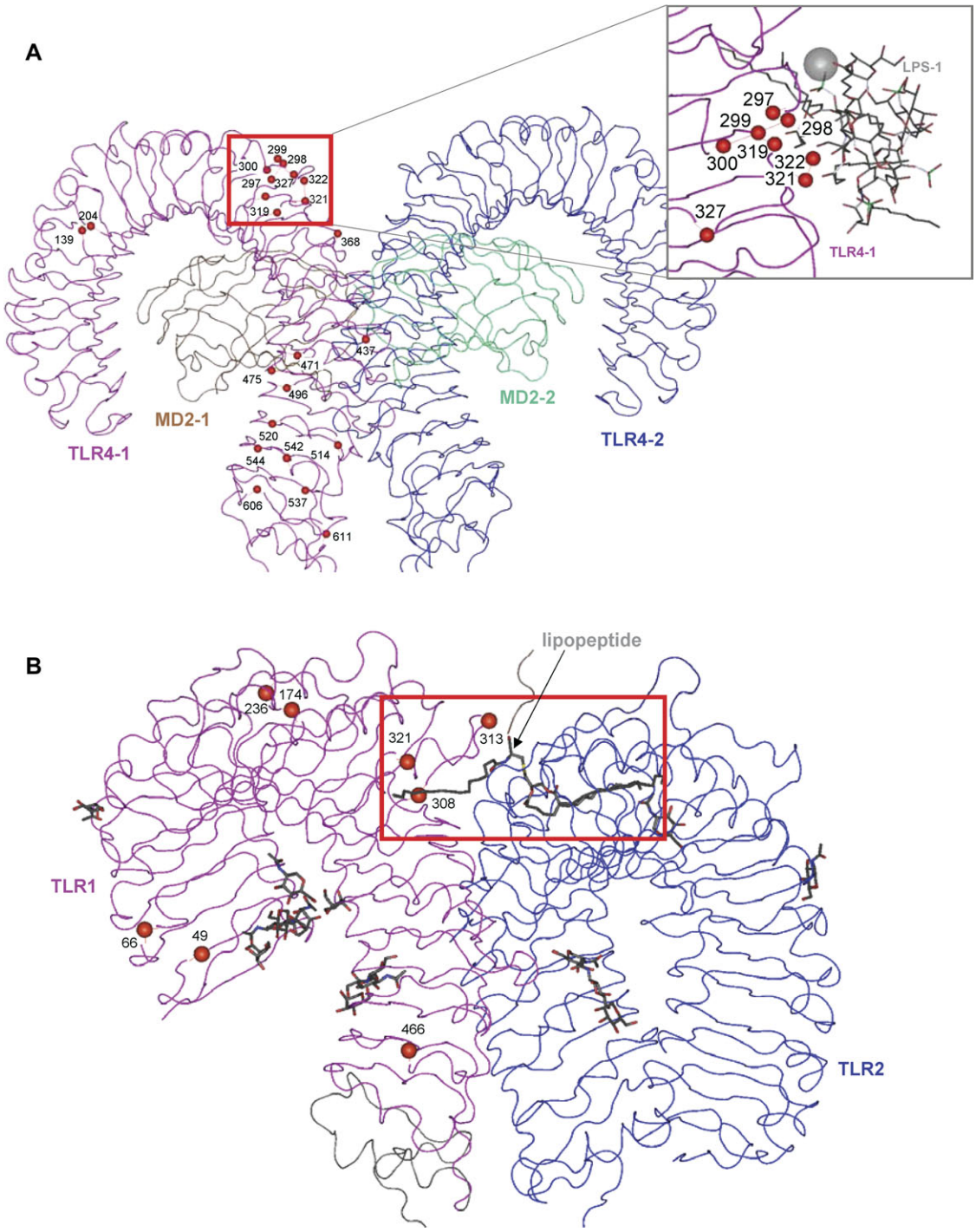

FIG. 1. Positively selected sites in the three-dimensional structures of TLRs. In each case, areas important for ligand binding that contain a concentration of sites under selection are squared. Amino acid positions of positively selected sites are labeled in red. (*A*) TLR4–TLR4 dimer; (*B*) TLR1–TLR2 dimer.

We also used the HKA test to assess whether individual TLRs have been subject to positive selection. Only TLR1 showed a weak but nonsignificant excess of polymorphism relative to divergence (HKA $\chi^2 = 3.5$, $P = 0.06$). No significant deviations were observed for the chimpanzee TLRs (supplementary table S7, Supplementary Material online). Viral and nonviral subsets were not significantly different from each other.

Functional Consequences of Replacement Polymorphisms

Several methods have been developed to computationally predict the functional consequences of replacement polymorphisms (reviewed in Ng and Henikoff 2006). In humans, we found a total of 11 probably damaging, 20 possibly damaging, and 31 benign polymorphisms, whereas in chimpanzees, we found 3 probably damaging, 3 possibly damaging, and 19 benign polymorphisms (table 4). The proportion of

A

Position	aa 1	Properties	aa 2	Properties
139	Ala	SM, NP, NEU	Leu	NP, NEU
	Ala	SM, NP, NEU	Ile	NP, NEU
204	Leu	NP, NEU	Arg	P, POS
	Leu	NP, NEU	Ile	NP, NEU
297	Pro	SM, NP, NEU	Leu	NP, NEU
	Pro	SM, NP, NEU	Ser	SM, P, NEU
	Pro	SM, NP, NEU	Phe	NP, NEU
298	Asp	SM, P, NEG	Asn	SM, P, NEU
299	Asp	SM, P, NEG	Asn	SM, P, NEU
300	Asp	SM, P, NEG	Gly	SM, NP, NEU
321	Asp	SM, P, NEG	Lys	P, POS
	Lys	P, POS	Glu	P, NEG
322	Arg	P, POS	Gly	SM, NP, NEU
327	Pro	SM, NP, NEU	Ser	SM, P, NEU
351	Lys	P, POS	Glu	P, NEG
368	Ala	SM, NP, NEU	Ser	SM, P, NEU
437	Met	NP, NEU	Thr	SM, P, NEU
	Met	NP, NEU	Val	NP, NEU
	Ala	SM, NP, NEU	Ile	NP, NEU
471	Ser	SM, P, NEU	Leu	NP, NEU
	Ser	SM, P, NEU	Phe	NP, NEU
475	Val	NP, NEU	Thr	SM, P, NEU
	Val	NP, NEU	Ile	NP, NEU
514	Glu	P, NEG	Lys	P, POS
	Glu	P, NEG	Thr	SM, P, NEU
	Thr	SM, P, NEU	Lys	P, POS
537	Thr	SM, P, NEU	Met	NP, NEU
542	His	P, POS	Cys	SM, NP, NEU
	Cys	SM, NP, NEU	Ser	SM, P, NEU
606	Gln	P, NEU	Asn	SM, P, NEU
	Gln	P, NEU	Arg	P, POS
639	Leu	NP, NEU	Val	NP, NEU
	Leu	NP, NEU	Phe	NP, NEU

B

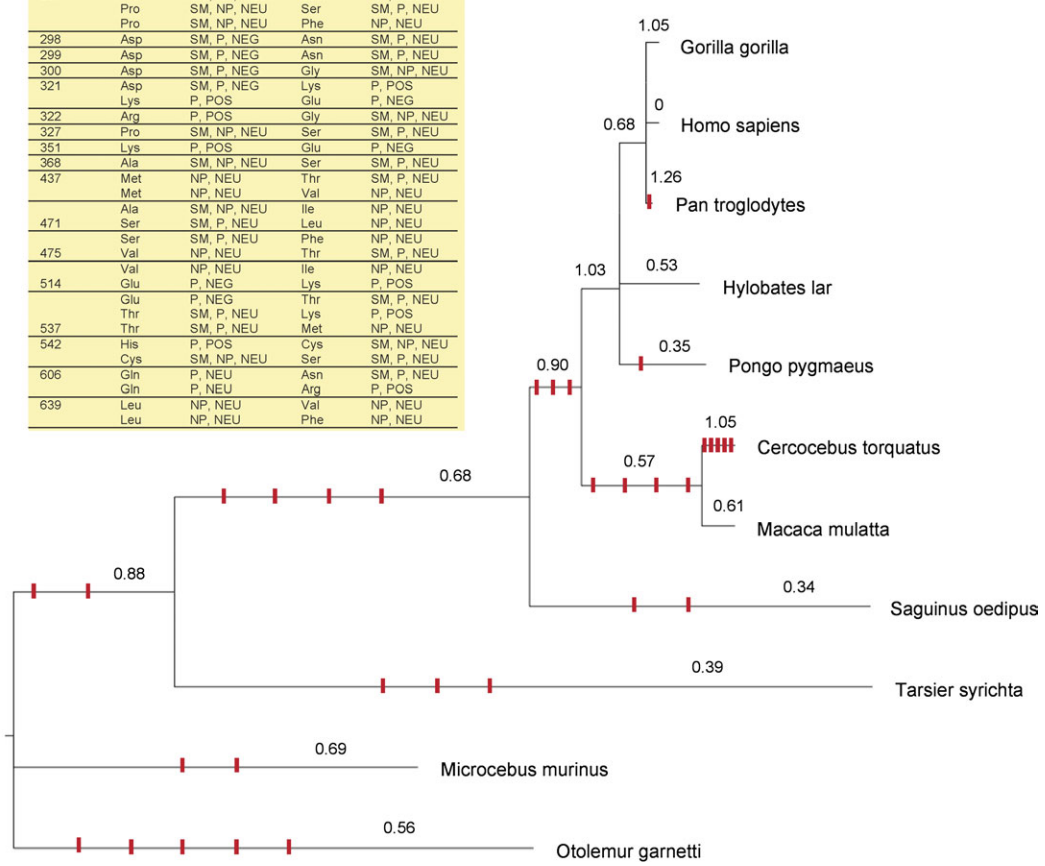

Fig. 2. Positive selection at TLR4. (*A*) Physicochemical properties of the amino acids at the positively selected sites at TLR4. SM, small; NP, nonpolar; P, polar; NEU, neutral; POS, positively charged; NEG, negatively charged. (*B*) Estimated lineage-specific *dN*/*dS* ratios from the branch-based analysis are shown above the branches of the TLR4 phylogeny. For the sites under selection from the codon-based analysis, the amino acid changes reconstructed by parsimony are mapped. Each red mark represents one amino acid substitution. Branch lengths are proportional to *dS*.

probably damaging/possibly damaging/benign SNPs was different between human and chimpanzee TLRs, with a marginal excess of probably and possibly damaging polymorphisms in humans ($P = 0.07$). Interestingly, these ratios are also significantly different between human TLRs and the human genome ($P = 0.02$). The excess of damaging SNPs in human TLRs is more evident for the nonviral TLRs, although the viral TLRs show a weaker but similar trend (table 5). When the probably and possibly damaging polymorphisms were collapsed into one class, the difference between human and chimpanzee TLRs or between human TLRs and the human genome became more significant (table 5), and for the human TLRs, this difference was driven by nonviral TLRs.

Negative Selection at Viral and Nonviral TLRs

We evaluated the levels of functional constraint among TLRs by estimating the global *dN*/*dS* ratio for each gene over the primate phylogeny as well as for the human lineage and the chimpanzee lineage (table 6). In each case, nonviral TLRs displayed a faster average rate of evolution than viral TLRs. Because of the low divergence between humans and chimpanzees, there is little statistical power in comparisons involving these lineages, but the average *dN*/*dS* for viral TLRs was significantly lower than the average *dN*/*dS* for nonviral TLRs across the primate phylogeny (*t* test $P = 0.007$). This indicates that viral TLRs are under stronger purifying selection than nonviral TLRs. The domain-specific *dN*/*dS* values show that on average, the leucine-rich repeat domain evolves faster than the signaling domain. This pattern is shared between viral and nonviral TLRs. On the other hand, the signal peptide and transmembrane domains show a higher *dN*/*dS* than the other domains.

Discussion

We analyzed the patterns of divergence among primates and of polymorphism in humans and chimpanzees for the entire TLR family with the goal of providing a general picture of the evolution of TLRs over different timescales. A summary of these results is provided in supplementary table S8, Supplementary Material online. We found a clear signature of positive selection in the rates of substitution across primates in most TLRs. However, within species, the

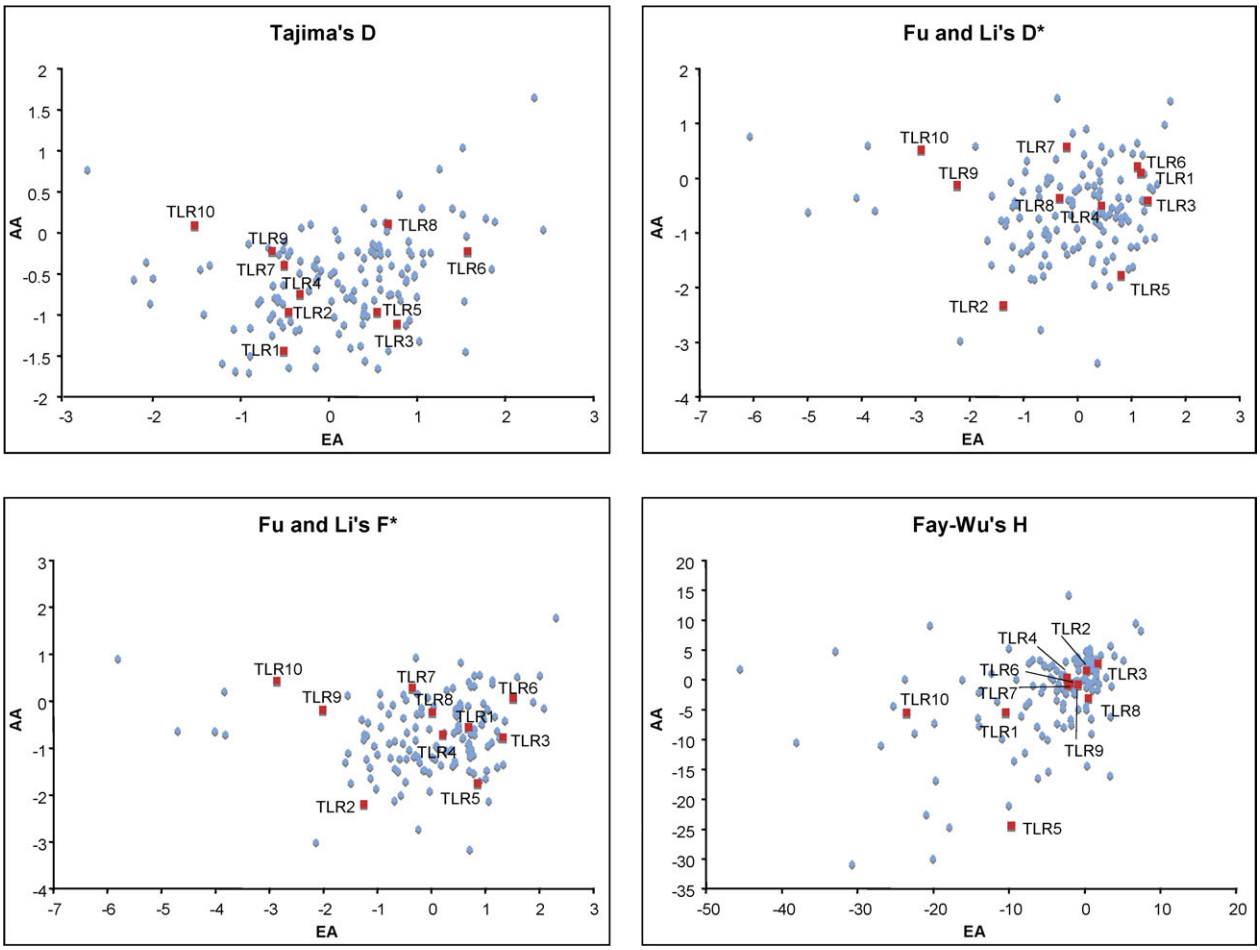

FIG. 3. Summary statistics of the allele frequency spectrum in TLRs compared with the empirical distribution of Akey et al. (2004) for the same populations. AA, African Americans; EA, European Americans.

patterns of nucleotide variation were generally compatible with purifying selection. Thus, adaptive evolution at TLRs is not necessarily characterized by a constant turnover of alleles, as predicted by an arms race model of coevolution but might be more episodic in nature. We found that humans had a higher proportion of deleterious mutations than chimpanzees. We also found that viral TLRs were under stronger purifying selection than nonviral TLRs.

Recurrent Positive Selection is Common in Primate TLRs

Our analyses provide strong evidence that several TLRs have been subject to positive selection during primate

Table 3. Polymorphism and Divergence for Silent and Replacement Sites in Humans and Chimpanzees.

	Human Polymorphism		Human–Macaque Divergence				Chimpanzee Polymorphism		Chimpanzee–Macaque Divergence			
Gene	*R*	*S*	*R*	*S*	NI[a]	*P*	*R*	*S*	*R*	*S*	NI[a]	*P*
TLR1	9	5	35	25	1.29	0.77	6	1	35	26	4.46	0.23
TLR2	5	5	33	30	0.91	1	3	5	35	35	0.60	0.7
TLR3	3	2	42	47	1.68	0.67	2	1	48	48	2.00	1
TLR4	5	3	51	30	0.98	1	1	1	53	30	0.57	1
TLR5	10	5	40	32	1.60	0.56	3	1	45	33	2.20	0.64
TLR6	7	6	31	42	1.58	0.54	3	2	36	43	1.79	0.65
TLR7	2	4	23	34	0.74	1	1	0	23	33	n/a	0.43
TLR8	2	9	33	37	0.25	0.1	0	5	32	40	0.00	0.074
TLR9	5	1	47	76	8.09	0.08	4	2	51	82	3.22	0.43
TLR10	13	9	27	31	1.66	0.45	2	1	24	29	2.42	0.59
Viral	12	16	145	194	1.00	1	7	9	154	203	1.15	0.79
Nonviral	49	33	217	190	1.30	0.33	18	11	228	196	1.41	0.43
ALL	61	49	362	384	1.32	0.18	25	19	382	399	1.37	0.36

[a] $NI = (R/S)_{pol}/(R/S)_{div}$.

Table 4. Number of "Damaging" and "Benign" Polymorphisms within Humans and Chimpanzees, Predicted by Polyphen.

		Humans			Chimpanzees		
Gene	Class	Probably Damaging	Possibly Damaging	Benign	Probably Damaging	Possibly Damaging	Benign
TLR1	Nonviral	2	2	5	0	1	5
TLR2	Nonviral	3	1	1	0	0	3
TLR3	Viral	0	2	1	0	0	2
TLR4	Nonviral	0	4	1	0	1	0
TLR5	Nonviral	4	3	3	1	0	2
TLR6	Nonviral	1	0	6	1	0	2
TLR7	Viral	0	0	2	0	0	1
TLR8	Viral	0	1	2	0	0	0
TLR9	Viral	0	2	3	0	1	3
TLR10	Nonviral	1	5	7	1	0	1
All	All	11	20	31	3	3	19

evolution. Neutral models of evolution were rejected for six of the ten genes, and several ML methods identified specific codons with a high probability of being under selection. In comparison, Dean et al. (2008) found that only 3.4% of 6,110 reproductive genes showed evidence of recurrent positive selection in five mammalian species using a similar approach. Positive selection at TLRs may also account for the relatively high *dN*/*dS* values averaged over the entire tree (table 6) in relation to the mean *dN*/*dS* of 0.25 for the human–chimpanzee–macaque trio (Gibbs et al. 2007). Finally, several of the putatively selected sites fall in regions important for function based on structural information. The fact that the codon-based approaches only detect selection acting on the same sites repeatedly makes these conclusions conservative.

An open question is whether the innate immune systems of vertebrates and invertebrates are under similar selective pressures. In vertebrates, the acquisition of adaptive immunity might have altered the evolutionary pressures on innate immunity genes. In *Drosophila*, pattern recognition receptors display more evidence of positive selection across species than other innate immunity genes (Sackton et al. 2007). Here, we showed that six of ten TLRs have been subject to positive selection in vertebrates in contrast to the prevailing view that TLRs are largely conserved (Roach et al. 2005). Nonetheless, it is important to bear in mind that even for these six positively selected genes, only a small proportion of sites show evidence of selection, whereas much of each gene shows strong functional constraint. This is essentially the same pattern we reported earlier for TLR5 (Wlasiuk et al. 2009). In the present study, we failed to find evidence of selection on TLR5 using fewer species. This is not particularly surprising given that the power to detect recurrent selection with codon-based approaches depends on the number of taxa, and the present study focused on more genes but fewer species.

The strongest evidence for positive selection was seen at TLR4 and TLR1. At TLR4, 24 codons (table 1) were concordant between at least two ML methods and thus constitute robust candidates for positive selection. Some of the non-ambiguous amino acid changes at these sites are radical in terms of their physicochemical properties (size, polarity, and charge) (fig. 2) strengthening the case for positive selection. In association with the myeloid differentiation factor 2 (MD2), TLR4 not only responds to LPS from Gram-negative bacteria but also targets components of yeast, *Trypanosoma*, and even viruses (Kumar et al. 2009). The crystal structure of the extracellular portion of the TLR4–MD2 complex has been resolved (Park et al. 2009), and several of the putative sites under selection (295, 297, 298, 299, 300, and 360) are in a region that participates in interactions between TLR4, MD2, and LPS (fig. 1). Moreover, many of the observed amino acid changes affect polarity or charge (fig. 2). Site 296, identified by REL but not by other methods, directly participates in the binding of LPS to TLR4 by forming a hydrogen bond with the inner core of LPS (Park et al. 2009).

TLR1 interacts TLR2 for the recognition of triacyl lipopeptides from Gram-negative bacteria (Takeuchi et al. 2002) and also showed extensive evidence of recurrent positive selection. In this case, 12 sites appear robust among analyses. Of these, Site 313 falls directly in the ligand-

Table 5. Ratios of Damaging to Benign Polymorphisms in Humans and Chimpanzees for Different Subclasses of TLRs.

	Probably Damaging	Possibly Damaging	Benign	P^a	P^b
Human genome[c]	11,174	14,187	49,795	—	
All human TLRs	11	20	31	0.02	<0.001
Human viral TLRs	0	5	8	0.08	0.8
Human nonviral TLRs	11	15	23	0.02	0.005
All chimpanzee TLRs	3	3	19	0.58 (0.07)[d]	0.4 (0.03)[d]
Chimpanzee viral TLRs	0	1	6	0.47	0.4
Chimpanzee nonviral TLRs	3	2	13	0.74	0.6

[a] Compared with the human genome.
[b] Compared with the human genome, collapsing the probably and possibly damaging mutations.
[c] From Sunyaev et al. (2001).
[d] Comparison with all human TLRs in parentheses.

Table 6. *dN*/*dS* Values for the Human Lineage, the Chimpanzee Lineage, and across Primates.

	Human Branch[a]	Chimpanzee Branch[a]	Primates[b]				
			Global	SP	EXT	TM	CYT
TLR1	1.04	1.35	0.43	0.59	0.45	0.33	0.21
TLR2	n/a[c]	0.31	0.44	0.40	0.52	0.21	0.13
TLR3	0.20	0.64	0.32	0.87	0.30	n/a[c]	0.23
TLR4	0.20	1.39	0.57	0.48	0.71	2.35	0.17
TLR5	0.80	0.54	0.42	1.04	0.43	0.46	0.32
TLR6	0.33	0.59	0.40	0.62	0.38	0.57	0.28
TLR7	0.29	0.60	0.34	1.33	0.36	1.37	0.10
TLR8	0.12	0.06	0.39	0.78	0.44	0.45	0.10
TLR9	0.22	0.15	0.17	0.23	0.17	0.40	0.11
TLR10	0.18	0.26	0.48	0.36	0.46	0.46	0.52
Viral average	0.21	0.36	0.31	0.80	0.32	0.74	0.13
Nonviral average	0.51	0.74	0.46	0.58	0.49	0.73	0.27
P[d]	0.09	0.10	0.007	0.17	0.02	0.49	0.05

n/a, not applicable.

SP, signal peptide; EXT, extracellular domain; TM, transmembrane domain; CYT, cytoplasmic domain.

[a] Macaque sequence used to polarize changes along the human or chimpanzee branches.

[b] Average *dN*/*dS* across primates.

[c] $dS = 0$.

[d] Viral and nonviral averages were compared with a *t* test.

binding site of the extracellular domain, although Sites 308 and 321 are also in close physical proximity to the ligand-binding site in the three-dimensional structure of the dimer (Jin et al. 2007) (fig. 1).

We observed a concentration of positively selected sites at TLR4 in *Catarrhini*, both when we looked at variation of *dN*/*dS* along lineages and among sites, in agreement with previous reports (Nakajima et al. 2008). Similarly, the branches with $dN/dS > 1$ in TLR1 and TLR8 are found among Old World primates and apes (data not shown). Stronger signals of selection in these groups have also been observed in antiviral genes such as APOBEC3G, TRIM5, and PKR (Sawyer et al. 2004, 2005; Elde et al. 2009), suggesting that these radiations might have been associated with major changes in pathogen abundance, diversity, or both.

Interestingly, some of the sites with evidence of recurrent positive selection are polymorphic in humans and have been reported to be associated with different phenotypes. For example, TLR4 D299G has been linked to differences in responsiveness to LPS (Arbour et al. 2000), susceptibility to bacterial infections (Kiechl et al. 2002), and higher prevalence of asthma (Bottcher et al. 2004), whereas TLR1 S248N shows a weak impairment in response to bacteria in vitro (Omueti et al. 2007, but see Hawn et al. 2007; Barreiro et al. 2009) and has been reported to confer increased risk of leprosy (Schuring et al. 2009) and atopic asthma (Kormann et al. 2008).

In spite of the evidence for selection documented here, the selective agents that have shaped TLR evolution are not easy to pinpoint. Because TLRs recognize molecular patterns shared by general classes of microorganisms, the variety of microbes that TLRs can target is large.

No Clear Evidence of Selection within Species

Very little information is available about polymorphism in wild populations of apes, and most efforts have been directed toward sequencing putatively neutral regions of the genome to infer historical demography (e.g., Yu et al. 2003, 2004; Fischer et al. 2006). However, from both an evolutionary and a medical perspective, it is important to understand how two closely related species with very different ecologies differ in a set of genes that constitute the first defense against pathogens. Despite their different habitats, life history, and population attributes likely to affect exposure to pathogens, overall patterns of nucleotide variation at TLRs were fairly similar in humans and chimpanzees.

We found no strong evidence of positive selection at TLRs within species from consideration of the distribution of allele frequencies, patterns of LD, F_{ST}, or relative levels of polymorphism and divergence. Although some genes showed weak departures from neutral expectations in the distribution of allele frequencies in humans (supplementary table S3, Supplementary Material online), these deviations generally disappeared when the effects of demography were taken into account (fig. 3). The evidence of positive selection in interspecific comparisons but not within humans or chimpanzees suggests that selection might be episodic. Positive selection might be more episodic if most infections are sporadic rather than caused by pathogens that establish more permanent or stable associations with their hosts.

Nonetheless, several previous studies suggested that segregating variants at the TLR6–TLR1–TLR10 cluster could be under selection. Burton et al. (2007) reported high differentiation among British people for a SNP in the TLR6–TLR1–TLR10 cluster, and TLR1 N248S presents striking north-south clinal variation (Todd et al. 2007). Barreiro et al. (2009) concluded that TLR1 I602S has been the target of selection in non-Africans based on extreme differentiation, reduced polymorphism, long-range haplotype homozygosity, and functional assays. Finally, SNPs in this region have been linked to disease phenotypes (Zhou et al. 2006;

Hawn et al. 2007; Johnson et al. 2007; Ma et al. 2007; Stevens et al. 2008). Given these observations, we cannot discard the possibility that geographically restricted selection, perhaps in conjunction with a more complicated demographic history has shaped the observed patterns of variation at these loci.

Relaxed Selection at Human TLRs

We predicted computationally the degree of functional disruption caused by each SNP in the human and chimpanzee TLRs and found a striking difference in the relative proportion of damaging to benign SNPs between species. Human TLRs showed a higher ratio of damaging/benign changes (1:1) compared with chimpanzees (1:3) or with the human genome as a whole (1:2). We note, however, that the smaller sample size in chimpanzees could bias against sampling low-frequency polymorphisms. The excess of damaging changes in human TLRs compared with the human genome and chimpanzee TLRs suggests a very recent relaxation of selective constraint on TLRs in the human lineage. If purifying selection had been relaxed a long time ago, we would expect a consistent increase in dN/dS in the human lineage, but this is not seen.

For neutral mutations, an estimate of their age based on their frequency is given by $E(t) = (-2q)(\ln q)/(1 - q)$, where age is measured in units of $2N$ generations (Kimura and Ohta 1973). Assuming that $N = 10{,}000$ and the generation time is 25 years, we estimated that the youngest replacement mutations in our sample are ~50,000 years old. This time frame is inconsistent with very recent changes in sanitary conditions or changes associated with the advent of agriculture. However, the variance associated with this estimate is large. The migration of modern humans out of Africa around 50,000 years is roughly coincident with the estimated age of the low-frequency replacement changes and suggests that the extreme reduction in population size associated with this migration (Garrigan et al. 2007) might have resulted in a relaxation of purifying selection. Lohmueller et al. (2008) showed that Europeans carry a significantly higher proportion of deleterious polymorphisms than Africans, supporting this idea.

Viral Receptors Are under Stronger Purifying Selection

We uncovered consistent differences between viral and nonviral TLRs that imply that viral TLRs are under stronger evolutionary constraint. Viral TLRs showed lower levels of polymorphism and lower rates of protein evolution than nonviral TLRs. Viral TLRs also have a smaller proportion of damaging polymorphisms in both human and chimpanzees. This suggests that viral TLR polymorphisms are mostly neutral, whereas nonviral TLRs also segregate some slightly deleterious polymorphisms. This observation is in line with the NI, which revealed a weak excess of replacement polymorphisms in nonviral TLRs. Most of these polymorphisms are rare, suggesting that they have mildly deleterious effects. Similar results were obtained by Barreiro et al. (2009), who also reported consistent differences between the two classes of receptors based on the pattern of nonsynonymous polymorphisms and their predicted functional effects in African, European, and Asian populations.

Despite these differences, the domain-specific patterns of negative selection revealed important similarities between viral and nonviral receptors. For both, the cytoplasmic region that contains the signaling domain is the most constrained portion of the protein, followed by the leucine-rich repeat domain containing the pathogen recognition site. All TLRs except for TLR3 signal through the MyD88 pathway (reviewed in Kumar et al. 2009). Moreover, MyD88 has relative low rates of protein evolution between species, and one possibility is that sharing this interacting partner results in a lower degree of flexibility.

Viral and nonviral TLRs have important biological differences in terms of their ligands, localization, and potential for self-reactivity that might help to explain their differences in patterns of molecular evolution. Nonviral TLRs localize in the cell membrane to recognize lipids, flagellin, and other molecules (mostly of bacterial origin) that are absent in the host. Viral TLRs, on the other hand, locate intracellularly to recognize nucleic acids mostly from viruses. Unlike TLRs expressed in the cell membrane, viral TLRs remain in the endoplasmic reticulum in a resting state and traffic to endosomal vesicles after ligand-induced stimulation (Latz et al. 2004), where they might undergo further processing to produce a functional receptor (Ewald et al. 2008). Restricted activation of viral TLRs to endosomal compartments has been proposed as an evolutionary strategy to minimize the dangerous encounter with host nucleic acids. Viral recognition by TLRs is based on a nongeneric type of response that needs to be reliable enough to ensure that a correct response is developed, but safe enough to avoid reaction against self-derived nucleic acids, as inappropriate activation can lead to autoimmune disorders (Krieg and Vollmer 2007). This delicate trade-off might constrain the evolution of viral TLRs and help to explain the observed patterns.

Supplementary Material

Supplementary tables S1–S8 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

We thank Dr M. Hammer for providing the chimpanzee DNA samples and Dr W. Klimecki for providing access to the raw human sequence data deposited in the Innate Immunity Database; O. Savina and A. Woerner for providing scripts to handle/analyze polymorphism data; A. Geraldes, M. Carneiro, M. Dean, T. Salcedo, and other members of the Nachman Lab for fruitful discussions. This work was supported by the National Institute of Health grant to M.W.N.

References

Akey JM, Eberle MA, Rieder MJ, Carlson CS, Shriver MD, Nickerson DA, Kruglyak L. 2004. Population history and natural selection shape patterns of genetic variation in 132 genes. *Plos Biol.* 2:1591–1599.

Akira S, Takeda K. 2004. Toll-like receptor signalling. *Nat Rev Immunol.* 4:499–511.

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

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

Barreiro LB, Ben-Ali M, Quach H, et al. (18 co-authors). 2009. Evolutionary dynamics of human Toll-like receptors and their different contributions to host defense. *PLoS Genet.* 5:e1000562.

Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A. 2007. The delayed rise of present-day mammals. *Nature* 446:507–512.

Bottcher MF, Hmani-Aifa M, Lindstrom A, Jenmalm MC, Mai XM, Nilsson L, Zdolsek HA, Bjorksten B, Soderkvist P, Vaarala O. 2004. A TLR4 polymorphism is associated with asthma and reduced lipopolysaccharide-induced interleukin-12(p70) responses in Swedish children. *J Allergy Clin Immunol.* 114:561–567.

Burton PR, Clayton DG, Cardon LR, et al. (192 co-authors). 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678.

Carpenter S, O'Neill LAJ. 2007. How important are Toll-like receptors for antimicrobial response? *Cell Microbiol.* 9:1891–1901.

Dean MD, Good JM, Nachman MW. 2008. Adaptive evolution of proteins secreted during sperm maturation: an analysis of the mouse epididymal transcriptome. *Mol Biol Evol.* 25: 383–392.

Elde NC, Child SJ, Geballe AP, Malik HS. 2009. Protein kinase R reveals an evolutionary model for defeating mimicry. *Nature* 457:485–489.

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

Fay JC, Wu CI. 2000. Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413.

Ferrer-Admetlla A, Bosch E, Sikora M, et al. (11 co-authors). 2008. Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol.* 181:1315–1322.

Fischer A, Pollack J, Thalmann O, Nickel B, Paabo S. 2006. Demographic history and genetic differentiation in apes. *Curr Biol.* 16:1133–1138.

Fu YX, Li WH. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.

Garrigan D, Kingan SB, Pilkington MM, et al. (12 co-authors). 2007. Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics* 177:2195–2207.

Gibbs RA, Rogers J, Katze MG, et al. (176 co-authors). 2007. Evolutionary and biomedical insights from the rhesus macaque genome. *Science* 316:222–234.

Hawn TR, Misch EA, Dunstan SJ, et al. (13 co-authors). 2007. A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. *Eur J Immunol.* 37:2280–2289.

Hawn TR, Verbon A, Lettinga KD, et al. (13 co-authors). 2003. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. *J Exp Med.* 198:1563–1572.

Hawn TR, Wu H, Grossman JM, Hahn BH, Tsao BP, Aderem A. 2005. A stop codon polymorphism of Toll-like receptor 5 is associated with resistance to systemic lupus erythematosus. *Proc Natl Acad Sci U S A.* 102:10593–10597.

Hudson RR, Kreitman M, Aguade M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116: 153–159.

Hughes AL, Packer B, Welch R, Bergen AW, Chanock SJ, Yeager M. 2003. Widespread purifying selection at polymorphic sites in human protein-coding loci. *Proc Natl Acad Sci U S A.* 100: 15754–15757.

Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, Lee HY, Lee JO. 2007. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* 130:1071–1082.

Johnson CM, Lyle EA, Omueti KO, Stepensky VA, Yegin C, Alpsoy E, Hamann L, Schumann RR, Tapping RI. 2007. Cutting edge: a common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. *J Immunol.* 178:7520–7524.

Kariko K, Buckstein M, Ni HP, Weissman D. 2005. Suppression of RNA recognition by Toll-like receptors: the impact of nucleoside modification and the evolutionary origin of RNA. *Immunity* 23:165–175.

Kiechl S, Lorenz E, Reindl M, Wiedermann CJ, Oberhollenzer F, Bonora E, Willeit J, Schwartz DA. 2002. Toll-like receptor 4 polymorphisms and atherogenesis. *New Engl J Med.* 347:185–192.

Kimura M, Ohta T. 1973. Age of a neutral mutant persisting in a finite population. *Genetics* 75:199–212.

Kormann MSD, Depner M, Harti D, Klopp N, Illig T, Adamski J, Vogelberg C, Weiland SK, von Mutius E, Kabesch M. 2008. Toll-like receptor heterodimer variants protect from childhood asthma. *J Allergy Clin Immunol.* 122:86–92.

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

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

Kumar H, Kawai T, Akira S. 2009. Pathogen recognition in the innate immune response. *Biochem J.* 420:1–16.

Latz E, Schoenemeyer A, Visintin A, Fitzgerald KA, Monks BG, Knetter CF, Lien E, Nilsen NJ, Espevik T, Golenbock DT. 2004. TLR9 signals after translocating from the ER to CpG DNA in the lysosome. *Nat Immunol.* 5:190–198.

Lazarus R, Raby BA, Lange C, Silverman EK, Kwiatkowski DJ, Vercelli D, Klimecki WJ, Martinez FD, Weiss ST. 2004. TOLL-like receptor 10 genetic variation is associated with asthma in two independent samples. *Am J Resp Crit Care Med.* 170:594–600.

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

Liu L, Botos I, Wang Y, Leonard JN, Shiloach J, Segal DM, Davies DR. 2008. Structural basis of toll-like receptor 3 signaling with double-stranded RNA. *Science* 320:379–381.

Lohmueller KE, Indap AR, Schmidt S, et al. (12 co-authors). 2008. Proportionally more deleterious genetic variation in European than in African populations. *Nature* 451:994–997.

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

Ma X, Liu Y, Gowen BB, Graviss EA, Clark AG, Musser JM. 2007. Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. *PLoS One.* 2:e1318.

Matsushima N, Tanaka T, Enkhbayar P, Mikami T, Taga M, Yamada K, Kuroki Y. 2007. Comparative sequence analysis of

leucine-rich repeats (LRRs) within vertebrate toll-like receptors. *BMC Genomics* 8:124.
McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the Adh locus in Drosophila. *Nature* 351:652–654.
Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP. 2009. Signatures of natural selection are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc Natl Acad Sci U S A.* 106:7073–7078.
Nakajima T, Ohtani H, Satta Y, Uno Y, Akari H, Ishida T, Kimura A. 2008. Natural selection in the TLR-related genes in the course of primate evolution. *Immunogenetics* 60:727–735.
Nei M, Li WH. 1979. Mathematical-model for studying genetic-variation in terms of restriction endonucleases. *Proc Natl Acad Sci U S A.* 76:5269–5273.
Ng PC, Henikoff S. 2006. Predicting the effects of amino acid substitutions on protein function. *Annu Rev Genomics Hum Genet.* 7:61–80.
Nielsen R, Yang ZH. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. *Genetics* 148:929–936.
Omueti KO, Mazur DJ, Thompson KS, Lyle EA, Tapping RI. 2007. The polymorphism P315L of human toll-like receptor 1 impairs innate immune sensing of microbial cell wall components. *J Immunol.* 178:6387–6394.
Ortiz M, Kaessmann H, Zhang K, Bashirova A, Carrington M, Quintana-Murci L, Telenti A. 2008. The evolutionary history of the CD209 (DC-SIGN) family in humans and non-human primates. *Genes Immun.* 9:483–492.
Park BS, Song DH, Kim HM, Choi BS, Lee H, Lee JO. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. *Nature* 458:1191–1195.
Pond SK, Muse SV. 2005. Site-to-site variation of synonymous substitution rates. *Mol Biol Evol.* 22:2375–2385.
Pond SLK, Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533.
Rand DM, Kann LM. 1996. Excess amino acid polymorphism in mitochondrial DNA: contrasts among genes from Drosophila, mice, and humans. *Mol Biol Evol.* 13:735–748.
Roach JC, Glusman G, Rowen L, Kaur A, Purcell MK, Smith KD, Hood LE, Aderem A. 2005. The evolution of vertebrate Toll-like receptors. *Proc Natl Acad Sci U S A.* 102:9577–9582.
Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007. Dynamic evolution of the innate immune system in Drosophila. *Nat Genet.* 39:1461–1468.
Sawyer SL, Emerman M, Malik HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. *PLoS Biol.* 2:1278–1285.
Sawyer SL, Wu LI, Emerman M, Malik HS. 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. *Proc Natl Acad Sci U S A.* 102:2832–2837.
Schroder NW, Schumann RR. 2005. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. *Lancet Infect Dis.* 5:156–164.
Schuring RP, Hamann L, Faber WR, Pahan D, Richardus JH, Schumann RR, Oskam L. 2009. Polymorphism N248S in the human Toll-like Receptor 1 gene is related to leprosy and leprosy reactions. *J Infect Dis.* 199:1816–1819.
Smirnova I, Hamblin MT, McBride C, Beutler B, Di Rienzo A. 2001. Excess of rare amino acid polymorphisms in the toll-like receptor 4 in humans. *Genetics* 158:1657–1664.
Stevens VL, Hsing AW, Talbot JT, Zheng SL, Sun JL, Chen JB, Thun MJ, Xu JF, Calle EE, Rodriguez C. 2008. Genetic variation in the toll-like receptor gene cluster (TLR10-TLR1-TLR6) and prostate cancer risk. *Int J Cancer.* 123:2644–2650.
Sunyaev S, Ramensky V, Koch I, Lathe W, Kondrashov AS, Bork P. 2001. Prediction of deleterious human alleles. *Hum Mol Genet.* 10:591–597.
Tajima F. 1989. Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
Takeuchi O, Sato S, Horiuchi T, Hoshino K, Takeda K, Dong ZY, Modlin RL, Akira S. 2002. Cutting edge: role of Tll-like receptor 1 in mediating immune response to microbial lipoproteins. *J Immunol.* 169:10–14.
Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG. 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25:4876–4882.
Thornton K. 2003. Libsequence: a C++ class library for evolutionary genetic analysis. *Bioinformatics* 19:2325–2327.
Todd JA, Walker NM, Cooper JD, et al. (40 co-authors). 2007. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet.* 39:857–864.
Voight BF, Kudaravalli S, Wen XQ, Pritchard JK. 2006. A map of recent positive selection in the human genome. *Plos Biol.* 4:446–458.
Waterson GA. 1975. On the number of segregating sites in genetical models without recombination. *Theor Popul Biol.* 7:256–276.
Wei T, Gong J, Jamitzky F, Heckl WM, Stark RW, Rössle SC. 2009. Homology modeling of human Toll-like receptors TLR7, 8, and 9 ligand-binding domains. *Protein Sci.* 18:1684–1691.
Wlasiuk G, Khan S, Switzer WM, Nachman MW. 2009. A history of recurrent positive selection at the toll-like Receptor 5 in primates. *Mol Biol Evol.* 26:937–949.
Won YJ, Hey J. 2005. Divergence population genetics of chimpanzees. *Mol Biol Evol.* 22:297–307.
Yang ZH. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci.* 13:555–556.
Yang ZH. 1998. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol.* 15:568–573.
Yang ZH. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* 24:1586–1591.
Yang ZH, Nielsen R, Goldman N, Pedersen AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155:431–449.
Yang ZH, Wong WSW, Nielsen R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. *Mol Biol Evol.* 22:1107–1118.
Yu N, Chen FC, Ota S, Jorde LB, Pamilo P, Patthy L, Ramsay M, Jenkins T, Shyue SK, Li WH. 2002. Larger genetic differences within Africans than between Africans and Eurasians. *Genetics* 161:269–274.
Yu N, Jensen-Seaman MI, Chemnick L, Kidd JR, Deinard AS, Ryder O, Kidd KK, Li WH. 2003. Low nucleotide diversity in chimpanzees and bonobos. *Genetics* 164:1511–1518.
Yu N, Jensen-Seaman MI, Chemnick L, Ryder O, Li WH. 2004. Nucleotide diversity in gorillas. *Genetics* 166:1375–1383.
Zhou XX, Jia WH, Shen GP, Qin HD, Yu XJ, Chen LZ, Feng QS, Shugart YY, Zeng YX. 2006. Sequence variants in toll-like receptor 10 are associated with nasopharyngeal carcinoma risk. *Cancer Epidemiol Biomarkers Prev.* 15:862–866.