THE MOLECULAR EVOLUTION OF INNATE IMMUNITY GENES

THE MOLECULAR EVOLUTION OF INNATE IMMUNITY GENES

by

Gabriela Wlasiuk Battagliotti

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF ECOLOGY AND EVOLUTIONARY BIOLOGY

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2009

THE UNIVERSITY OF ARIZONA
GRADUATE COLLEGE

As members of the Dissertation Committee, we certify that we have read the dissertation prepared by GABRIELA WLASIUK BATTAGLIOTTI

entitled THE MOLECULAR EVOLUTION OF INNATE IMMUNITY GENES

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

_____ Date: 10/21/2009
Michael W. Nachman

_____ Date: 10/21/2009
Michael F. Hammer

_____ Date: 10/21/2009
Nancy A. Moran

_____ Date: 10/21/2009
Donata Vercelli

_____ Date: 10/21/2009
Michael Worobey

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copies of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

_____ Date: 10/21/2009
Dissertation Director: Michael Nachman

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at the University of Arizona and is deposited at the University Library to be made available to borrowers under the rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgement of source is made. Request for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the head of the major department or the Dean of the Graduate College when in his or her judgment the proposed use of the material is in the interests of scholarship. In all other instances, however, permission must be obtained from the author.

SIGNED: Gabriela Wlasiuk Battagliotti

ACKNOWLEDGEMENTS

This thesis is the result of the combined effort of many people to which I am extremely grateful. First, I would like to thank my advisor Michael Nachman, for being an excellent teacher, instilling in me his critical thinking and being the best example of scientific rigor. His influence truly changed the way I see and do science.

I am also very thankful to my committee members, who helped in different aspects of the difficult process of becoming intellectually independent. Mike Hammer offered his valuable expertise in human and primate evolution. Nancy Moran contributed with her very broad understanding of evolution and an always super-critical point of view that continuously inspires me. Mike Worobey provided his extensive knowledge in phylogenetics and molecular evolution and generously gave me the opportunity to participate in one of his research projects. Donata Vercelli nicely complemented the population genetics and molecular evolution aspects of my work with the different angle of her functional perspective and opened the doors of her lab from an early stage.

Many past and present, permanent and transitory, members of the Nachman lab have contributed to this thesis and my formation by both offering interesting ideas and/or participating in very useful discussions that helped shape my own ideas. For that, and for their friendship, I would like to thank Armando Geraldes, Jeff Good, Matt Saunders, Tovah Salcedo, Mari Sans-Fuentes, Miguel Carneiro, Matt Dean, Ming Beckwith, Adrian Munguia, Patrick Basset, Barbora Bimova, Radka Storkova, Lisa Kent, Kim Smith, Polly Campbell, Megan Phifer-Rixey, Taichi Susuki and Jeremy Jonas. Over these six years, I helped mentor two undergraduate students: Soofia Khan and Alexandra Permar, and the internships of two graduate students, Elena Crestani and Chinedu Nworu. I really enjoyed our interactions and learned extensively in the process of training them, so I also extend my acknowledgements to them. Soofia and Alexandra worked in the lab for long periods of time and I am further indebted to them for helping generating extensive amounts of data.

I am very thankful to August Woerner and Olga Savina for their friendly attitude and unconditional help with bioinformatics issues. It was always refreshing to go downstairs, knowing that there would be a smile and an answer to my questions. Heather Norton, Fernando Mendez and Adam Bjork also helped me at different stages with data analysis or interpretation. Other fellow graduate students have played important roles in my personal and intellectual development. In particular I would like to thank Becca Young, Kevin Oh, Laura Carsten, Erin Kelleher, Tami Haselkorn, Patrick Degnan, Heather Maughan, Liz Wood, Carlos Flores, Katy Prudic and Jeff Oliver. I am also grateful to several people in the EEB administrative staff that made my life easier over the years, especially Lili Schwartz, but also Kate Riley, Beth Sanchez, Sue Withworth, Suzanne LeClair and Carol Burleson.

I want to thank all the professors for whom I worked as a teaching assistant: Bill Birky, Gio Bosco, Jeremiah Hackett, Carlos Machado, Johanna Masel, Nancy Moran, Michael Nachman, Howard Ochman, Dan Papaj, Peter Reinthal, Linda Restifo, Mary Walkingshaw, Bruce Walsh and Ted Weinert. From their different approaches to teaching I benefited and gained a better appreciation of the labor of teaching.

My work critically depended on the samples or data provided by several people and institutions. Michael Hammer provided human and chimpanzee samples. Walt Klimecki gave access to the raw human sequence data deposited at the Innate Immunity Database. William Switzer, the Museum of Vertebrate Zoology (University of California, Berkeley) the Southwest National Primate Research Center, and the Smithsonian Institution, provided DNA or tissue samples from primate species.

I am thankful to The Rotary Foundation for the Ambassadorial Scholarship that funded my first year at the program, as well as the Department of Ecology and Evolutionary Biology, The Galileo Circle Scholarship, Women in Science Engineering, and my advisor, Michael Nachman, for funding different components of my work.

I would like to thank my family and friends for their continual support. I am very grateful to all the friends who in these six years shortened a little bit the distances from my family and my country, in particular, Ron and Sharon Sciaroni, Claudio Volonte, Jordan Samuel, Wilma Zufelt, Lalo and Sara Ameneiros, the Turco family, Carmen Cueva, Tom Dodd, Carlos Machado and Johanna Barrero, Gaspar Soria and Ana Cinti, Bill and Brenda Fee, Al and Sharyn Chesser, and Bob and Char Ernstein. Finally I would like to thank Leo, for his love and attitude towards life, which is my ultimate source of inspiration.

DEDICATION:

A Leo

TABLE OF CONTENTS

ABSTRACT

It is not clear whether genes of the innate immune system of vertebrates are subject to the same selective pressures as genes of the adaptive immune system, despite the fact that innate immunity genes lie directly at the interface between host and pathogens. The lack of consensus about the incidence, type, and strength of selection acting on vertebrate innate immunity genes motivated this study. The goal of this work was to elucidate the general principles of innate immune receptor evolution within and between species. A phylogenetic analysis of the Toll-like receptor 5 (TLR5) in primates showed an excess of nonsynonymous substitutions at certain codons, a pattern that is consistent with recurrent positive selection. The putative sites under selection often displayed radical substitutions, independent parallel changes, and were located in functionally important regions of the protein. In contrast with this interspecific pattern, population genetic analysis of this gene in humans and chimpanzees did not provide conclusive evidence of recent selection. The frequency and distribution of a TLR5 null mutation in human populations further suggested that TLR5 function might be partially redundant in the human immune system (Appendix A). Comparable analyses of the remaining nine human TLRs produced similar results and further pointed to a biologically meaningful difference in the pattern of molecular evolution between TLRs specialized in the recognition of viral nucleic acids and the other TLRs (Appendix B). The general picture that emerges from these studies challenges the conventional idea that pattern recognition

receptors are subject to an extreme degree of functional constraint dictated by the recognition of molecules that are essential for microbial fitness. Instead, TLRs display patterns of substitution between species that reflect an old history of positive selection in primates. A common theme, however, is that only a restricted proportion of sites is under positive selection, indicating an equally important role for purifying selection as a conservative force in the evolution of this gene family. A comparative analysis of evolutionary rates at fifteen loci involved in innate, intrinsic and adaptive immunity, and mating systems revealed that more promiscuous species are on average under stronger selection at defense genes (Appendix C). Although the effect is weak, this suggests that sexual promiscuity plays some role in the evolution of immune loci by affecting the risk of contracting infectious diseases.

CHAPTER 1: INTRODUCTION

Explanation of the problem and its context

A key goal of evolutionary biology is to understand variation in traits associated with fitness. Immune function is an essential component of fitness. This is supported by several independent lines of evidence, such as the deleterious consequences of immunodeficiencies in humans (Janeway 2001), the large fraction of the vertebrate genome devoted to immune-related functions (~4% of all genes in mouse and human), the recognized tradeoffs between immunity and reproduction (Sheldon and Verhulst 1996), and the rapid evolution of many immunity genes (Gibbs et al. 2004; 2005; Lindblad-Toh et al. 2005; Gibbs et al. 2007). In fact, host-pathogen interactions provide arguably one of the best arenas in which to study some of the central pillars of Darwin's theory of evolution by natural selection.

At the same time, a large amount of polymorphism is frequently observed at immunological traits in natural populations (Hughes and Nei 1988; Hughes and Nei 1989; Lazarus et al. 2002; Lazzaro, Sceurman, and Clark 2004; Hughes et al. 2005; Moeller and Tiffin 2005). In humans for example, the most polymorphic loci in the genome are found among immunity genes (Hughes 2002; Moeller and Tiffin 2005). If indeed the immune system is so closely related to fitness, it seems paradoxical to find this extreme level of polymorphism for immune traits. Unless variation itself is adaptive, we expect that natural selection will purge deleterious mutations and quickly fix beneficial mutations (Fisher 1930), and potentially remove linked neutral variation at the same time

(Maynard Smith and Haigh 1974; Charlesworth, Morgan, and Charlesworth 1993). However, this simplistic view ignores that infectious diseases are extremely dynamic and affected by multiple factors. Lazzaro and Little (2009) suggest that the interplay of several factors (including host and pathogen genotypes and their interactions, gene-environment interactions, fluctuating abiotic environments and pleiotropy) can generate complex selective regimes that potentially result in the maintenance of genetic variation at immunity genes.

Two decades ago, Charles Janeway (1989) envisioned a general theory of innate immune recognition that revolutionized our understanding of the vertebrate immune system. He predicted the existence and properties of pattern recognition receptors, the pathogen sensors of the innate immune system. Twenty years later, several families of pattern recognition receptors have been identified, among which the Toll-like receptors (TLRs) are the best characterized. In spite of an enormous basic and biomedical interest in elucidating the general principles of innate immune recognition and understanding the evolution of pattern recognition receptors, many basic questions remain unanswered. While there is general consensus about the importance of balancing selection in maintaining variation at adaptive immune genes (Hughes 2002), the overall pattern of evolution at innate immune genes is unclear (Holmes 2004). Evolutionary studies can provide perspective into the historical factors that shaped the present day patterns of variation, thereby providing clues about function.

This dissertation describes the molecular evolution of one family of innate immunity genes (the Toll-like receptors), and then addresses the effect of sexual

promiscuity on the rate of protein evolution in a functionally broader set of immune defense genes in primates. Patterns of nucleotide variability at TLRs in two closely related species, humans and chimpanzees, are contrasted with patterns of substitution across the primate radiation to assess the relative importance of negative and positive selection at pattern recognition receptors. By studying variation within and between species it is possible to make inferences about the timescale over which natural selection has acted. The results challenge the predominant view that pattern recognition receptors are subject to strong evolutionary constraint. To start disentangling the contribution of different factors to the evolution of immune loci, I investigated the link between mating system and immunity at the molecular level. Using comparative data in primates I discovered a positive relationship between sexual promiscuity and the rate of evolution in immunity genes.

A review of the literature

The vertebrate immune system: Although all living organisms have evolved effective mechanisms of defense against parasites, the complexity and specificity of the vertebrate immune system has no parallels. Vertebrate immunity consists of two intricately related branches: innate and adaptive immunity. The innate immune system is ancient, with shared pathways between vertebrates and invertebrates (Hoffmann et al. 1999) and some elements even shared between animals and plants. It is based on relatively conserved receptors and leads to an immediate response. Adaptive immunity, on the other hand, is restricted to jawed vertebrates. It is based on hypervariable receptors

whose variability is generated by somatic recombination, and results in a slower response and immunological memory (Janeway et al. 2005). These two arms of vertebrate immunity have fundamentally different strategies of pathogen recognition and elimination. The innate immune system is highly efficient at distinguishing self from non-self because it is based on receptors that mostly recognize microbial components, but it is relatively non-specific. Conversely, the adaptive immune system is less efficient at discriminating self from non-self components because it is essentially self-referential (although in normal conditions is not activated by self-ligands), but it has an extremely specific response. Through the coordinated action of the two systems, an acceptable efficacy of self/non-self discrimination and high degree of specificity are achieved (Janeway 2001; Palm and Medzhitov 2009). The recent recognition that innate and adaptive immunity act in such a tightly coordinated manner has blurred the traditional distinction between the two arms of the vertebrate immune system (Flajnik and Du Pasquier 2004).

Toll-like receptors of the innate immune system: The main targets of innate immune recognition are pathogen-associated molecular patterns (PAMPs), conserved molecular structures produced only by microbial pathogens but not by the host, and shared by general 'classes' of microorganisms (Medzhitov and Janeway 1997). PAMPs are recognized by a limited set of host receptors referred to as pattern recognition receptors (PRRs). A number of PRRs have been described in mammals, among which the TLRs are the most extensively studied. Ten TLR members are known in humans: TLR4 (Medzhitov, Preston-Hurlburt, and Janeway 1997); TLR1, TLR3 (Rock et al. 1998);

TLR2, TLR5 (Chaudhary et al. 1998; Rock et al. 1998), TLR6 (Takeuchi et al. 1999); TLR7, TLR8, TLR9 (Du et al. 2000); TLR10 (Chuang and Ulevitch 2001).

TLRs are transmembrane type I glycoproteins with a leucine-rich repeat ectodomain (LRR) and a Toll-IL-1 receptor cytoplasmic domain (TIR) connected by a single transmembrane domain (Bell et al. 2003). Some TLRs are located on the cell surface whereas others are found in intracellular compartments (Chaturvedi and Pierce 2009). Structurally, their ectodomains share a basic horseshoe shape typical of leucine-rich repeat proteins (Jin et al. 2007; Liu et al. 2008; Park et al. 2009). The active forms of TLRs are homo or heterodimers, and this heterodimerization broadens the repertoire of molecular patterns they can recognize (Ozinsky et al. 2000). The general structure of TLRs, as well as their cellular localization and main ligands are shown in Figure 1. Upon pathogen binding, TLRs induce the expression of several costimulatory cytokines and antimicrobial peptides through the NF-κB pathway (Akira and Takeda 2004). TLRs constitute a good set of genes for studying the molecular evolution of innate immunity because the clinical, structural, and mechanistic information available makes it possible to link evolutionary patterns with functional details.

Molecular evolution of immunity genes: Immunity-related genes usually show pervasive evidence of adaptive evolution. They evolve rapidly between species (Tanaka and Nei 1989; Jansa, Lundrigan, and Tucker 2003; Schlenke and Begun 2003; Gibbs et al. 2004; Sawyer, Emerman, and Malik 2004; Sawyer et al. 2005; Gibbs et al. 2007; Sackton et al. 2007; Elde et al. 2009) and show evidence of positive selection within

species (Ballingall et al. 2001; Lazzaro and Clark 2003; Schlenke and Begun 2003; Hughes et al. 2005), possibly due to host-pathogen coevolution.

Two basic models have been proposed to explain host-pathogen coevolution. In the 'arms race' model, the host evolves resistance and pathogens rapidly counter-evolve mechanisms to avoid that resistance (Van Valen 1973). This results in a continual selective turnover of alleles. This model predicts that the host population will be generally monomorphic at disease resistance loci, resulting in an excess of divergence with respect to polymorphism. Another model is based on the cost of resistance, and this model postulates that in the absence of pathogens, resistant individuals have reduced fitness (Stahl et al. 1999). Under this 'trench warfare' model, in a temporally or spatially varying selective regime, alleles for susceptibility and resistance can coexist, fluctuating in frequency for long periods of time. This model predicts that some alleles will be old and will be maintained as balanced polymorphisms of intermediate frequency, resulting in an excess of polymorphism with respect to divergence. More complex dynamics involving transient polymorphisms might also exist, depending on environmental heterogeneity.

In vertebrates, most molecular evolution studies have focused on effector genes of adaptive immunity such as immunoglobulins, T-cell receptors and MHC loci. These genes often show clear evidence of balancing selection (Hughes and Nei 1988; Hughes and Nei 1989; Hughes and Nei 1990). In contrast, vertebrate innate immunity genes have been less studied. In primates, several antiretroviral genes usually ascribed to the category of ‘intrinsic immunity’ (constitutively expressed cellular proteins that

specifically inhibit or block retroviruses) have been shown to evolve adaptively at extremely fast rates (Hughes and Nei 1988; Hughes and Nei 1989; Hughes and Nei 1990; Sawyer, Emerman, and Malik 2004; Sawyer et al. 2005; Sawyer, Emerman, and Malik 2007). Some antimicrobial peptides show signals of adaptive diversification after gene duplication (Hughes 1999; Tennessen 2005). However, because studies on immunity before the 90's concentrated mostly on adaptive immunity (Hoffmann et al. 1999), we lack a more systematic characterization of the patterns of evolution of innate immune genes.

Drosophila innate immunity genes, particularly PRRs, usually evolve by positive directional selection (Hughes 1999; Schlenke and Begun 2003; Sackton et al. 2007). It remains unclear whether innate immunity genes in vertebrates exhibit the same strength and type of selection compared to their invertebrate counterparts. This question is relevant because most organisms do not posses adaptive immunity, and the acquisition of the adaptive immune system along the vertebrate lineage might have radically changed the selective pressures acting on the innate response.

Mating and immunity: Reproduction and immunity are two functions closely related to fitness but also intimately linked to each other. Numerous connections have been proposed between mating and immunity (Lawniczak et al. 2007). At the precopulatory level, mate choice could be based on secondary sexual traits that indirectly reflect heritable variation in immune condition (Hamilton and Zuk 1982). Post-mating processes could also interact with immunity. For example, the male ejaculate might interfere with female immunity leading to sexual conflict (Fedorka and Zuk 2005).

Another possibility is cryptic female choice mediated by immune response in the female reproductive tract (Reddy, Yedery, and Aranha 2004). Finally, species with more promiscuous mating systems might be under increased risk of acquiring sexually transmitted diseases (Nunn, Gittleman, and Antonovics 2000).

Testing hypotheses about the relationship between mating and immunity is complicated in large vertebrates, in which most species are not amenable to experimental manipulation. In this situation, the comparative approach is a powerful alternative. Although correlational evidence emerging from this type of studies needs to be interpreted with caution, comparative studies can shed light on the complex relationship between immunity, reproduction, and other important life-history traits.

Explanation of dissertation format

Understanding the evolution of genes that underlie vertebrate innate immunity is of fundamental importance but we still do not have a clear picture of the general patterns of evolution of these genes. In Appendix A, I present a detailed study of the evolution of TLR5 among primates and within humans and chimpanzees. I show that the evolutionary history of TLR5 has been driven by recurrent positive selection on a small proportion of codons, against a background of strong purifying selection. The examination of patterns of variation within species shows, in contrast, no clear evidence of positive selection. Appendix B expands upon these results to examine patterns of molecular evolution within and between primates for the entire family of TLRs. In agreement with the evolutionary history of TLR5, most of the genes in the family show evidence of adaptive

evolution across primates, but no evidence of selection within humans or chimpanzees. By comparing the patterns of polymorphism and divergence between humans and chimpanzees and between viral and non-viral receptors, I provide a general picture of the evolution of this important family of innate immune receptors. Finally, Appendix C focuses on lineage specific patterns of evolution, to evaluate the effect of sexual promiscuity on the evolutionary rates of a diverse set of immunity genes (including, but not limited to TLRs). By looking at variation in the rates of protein evolution in the context of mating system and other variable expected to influence disease risk, I find evidence of a weak but significant effect of mating system on the evolution of immune defense genes.

CHAPTER 2: PRESENT STUDY

The methods, results, and conclusions of this study are presented in the papers appended to this dissertation. The following is a summary of the most important findings of these papers.

To characterize the main patterns of evolution of innate immunity genes in primates and to understand the major factors affecting their rate of protein evolution, I have: 1) conducted comprehensive population genetic and molecular evolution analyses of TLRs, and 2) extended ideas about mating system and the evolution of immunity to the molecular level, by testing the hypothesis that levels of female promiscuity influence rates of evolution of immunity genes. Appendix A provides a detailed study of the evolution of TLR5, a receptor that recognizes flagellated bacteria. Using maximum likelihood methods I uncovered clear signatures of positive selection driving the evolution of this gene in primates. Within humans or chimpanzees, however, a multiplicity of approaches and tests of selection did not find deviations from the neutral model of evolution. Moreover, genetic drift seems to be responsible for the relatively high frequency of a loss of function mutation in humans, suggesting some degree of functional redundancy at this gene.

Appendix B is a natural extension of the previous study and examines the patterns of molecular evolution of all the genes in the TLR family, also at deep and recent timescales. Compelling evidence of positive selection among species was generalized to most of the other TLR familiy members, challenging the current paradigm of TLR

evolution. No obvious evidence for positive selection was found at the population level. An increase in the proportion of deleterious polymorphisms was found in humans with respect to chimpanzees, which can be explained by relaxed selection in the former. Viral TLRs were under stronger purifying selection than non-viral TLRs. By dissecting the patterns of positive and negative selection at different timescales I provide a more complete picture of the evolutionary history of this important family of innate immune receptors.

Appendix C tests the hypothesis that mating system is a major determinant of the rate of protein evolution in a set of 15 immune defense genes with a known history of rapid evolution in primates. Primates constitute an excellent system with which to test this idea, because extensive information is available about social and mating systems and other ecological and life history variables that can affect disease risk. The degree of female promiscuity, as determined by the mating system, showed a weak but significant effect on the rate of protein evolution. As predicted by the disease-risk/promiscuity hypothesis, species with higher levels of female promiscuity had on average more evidence of positive selection than less promiscuous species.

REFERENCES

Akira, S, and Takeda, K. 2004. Toll-like receptor signalling. Nature Reviews Immunology 4:499-511.

Ballingall, KT, Waibochi, L, Holmes, EC, Woelk, CH, MacHugh, ND, Lutje, V, and McKeever, DJ. 2001. The CD45 locus in cattle: allelic polymorphism and evidence for exceptional positive natural selection. Immunogenetics 52:276-283.

Bell, JK, Mullen, GED, Leifer, CA, Mazzoni, A, Davies, DR, and Segal, DM. 2003. Leucine-rich repeats and pathogen recognition in Toll-like receptors. Trends in Immunology 24:528-533.

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

Chaturvedi, A, and Pierce, SK. 2009. How Location Governs Toll-Like Receptor Signaling. Traffic 10:621-628.

Chaudhary, PM, Ferguson, C, Nguyen, V, Nguyen, O, Massa, HF, Eby, M, Jasmin, A, Trask, BJ, Hood, L, and Nelson, PS. 1998. Cloning and characterization of two Toll/interleukin-1 receptor-like genes TIL3 and TIL4: Evidence for a multi-gene receptor family in humans. Blood 91:4020-4027.

Chuang, TH, and Ulevitch, RJ. 2001. Identification of hTLR10: a novel human Toll-like receptor preferentially expressed in immune cells. Biochimica Et Biophysica Acta-Gene Structure and Expression 1518:157-161.

Du, X, Poltorak, A, Wei, YG, and Beutler, B. 2000. Three novel mammalian Toll-like receptors: gene structure, expression, and evolution. European Cytokine Network 11:362-371.

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

Fedorka, KM, and Zuk, M. 2005. Sexual conflict and female immune suppression in the cricket, Allonemobious socius. Journal of Evolutionary Biology 18:1515-1522.

Fisher, RA. 1930. The genetical theory of natural selection. Oxford University Press, Oxford, United Kingdom.

Flajnik, MF, and Du Pasquier, L. 2004. Evolution of innate and adaptive immunity: can we draw a line? Trends in Immunology 25:640-644.

Gibbs, RARogers, JKatze, MG et al. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222-234.

Gibbs, RAWeinstock, GMMetzker, ML et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493-521.

Hamilton, WD, and Zuk, M. 1982. Heritable true fitness and bright birds - a role for parasites. Science 218:384-387.

Hoffmann, JA, Kafatos, FC, Janeway, CA, and Ezekowitz, RAB. 1999. Phylogenetic perspectives in innate immunity. Science 284:1313-1318.

Holmes, EC. 2004. Adaptation and immunity. Plos Biology 2:1267-1269.

Hughes, AL. 1999. Evolutionary diversification of the mammalian defensins. Cellular and Molecular Life Sciences 56:94-103.

Hughes, AL. 2002. Natural selection and the diversification of vertebrate immune effectors. Immunological Reviews 190:161-168.

Hughes, AL, and Nei, M. 1988. Pattern of nucleotide substitution at Major Histocompatibility Complex Class-I loci reveals overdominant selection. Nature 335:167-170.

Hughes, AL, and Nei, M. 1989. Nucleotide substitution at Major Histocompatibility Complex Class-II loci - Evidence for overdominant selection. Proc. Natl. Acad. Sci. USA 86:958-962.

Hughes, AL, and Nei, M. 1990. Evolutionary Relationships of Class-Ii Major-Histocompatibility-Complex Genes in Mammals. Mol. Biol. Evol. 7:491-514.

Hughes, AL, Packer, B, Welch, R, Chanock, SJ, and Yeager, M. 2005. High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. Immunogenetics 57:821-827.

Janeway, CA. 1989. Approaching the asymptote - Evolution and revolution in immunology. Cold Spring Harbor Symposia on Quantitative Biology 54:1-13.

Janeway, CA. 2001. How the immune system works to protect the host from infection: A personal view. Proc. Natl. Acad. Sci. USA 98:7461-7468.

Janeway, CA, Travers, P, Walport, M, and Shlomchik, M. 2005. Immunobiology. Garland Science, Oxford, UK.

Jansa, SA, Lundrigan, BL, and Tucker, PK. 2003. Tests for positive selection on immune and reproductive genes in closely related species of the murine genus Mus. J Mol Evol 56:294-307.

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

Lawniczak, MKN, Barnes, AI, Linklater, JR, Boone, JM, Wigby, S, and Chapman, T. 2007. Mating and immunity in invertebrates. Trends in Ecology & Evolution 22:48-55.

Lazarus, R, Vercelli, D, Palmer, LJ et al. 2002. Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. Immunological Reviews 190:9-25.

Lazzaro, BP, and Clark, AG. 2003. Molecular population genetics of inducible antibacterial peptide genes in Drosophila melanogaster. Mol. Biol. Evol. 20:914-923.

Lazzaro, BP, and Little, TJ. 2009. Immunity in a variable world. Philosophical Transactions of the Royal Society B-Biological Sciences 364:15-26.

Lazzaro, BP, Sceurman, BK, and Clark, AG. 2004. Genetic basis of natural variation in D-melanogaster antibacterial immunity. Science 303:1873-1876.

Lindblad-Toh, K, Wade, CM, Mikkelsen, TS et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. Nature 438:803-819.

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

Maynard Smith, J, and Haigh, J. 1974. The hitch-hiking effect of a favorable gene. Genetical Reseach:23-35.

Medzhitov, R, and Janeway, CA. 1997. Innate immunity: The virtues of a nonclonal system of recognition. Cell 91:295-298.

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

Mikkelsen, TS, Hillier, LW, Eichler, EE et al. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87.

Moeller, DA, and Tiffin, P. 2005. Genetic diversity and the evolutionary history of plant immunity genes in two species of Zea. Mol. Biol. Evol. 22:2480-2490.

Nunn, CL, Gittleman, JL, and Antonovics, J. 2000. Promiscuity and the primate immune system. Science 290:1168-1170.

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

Palm, NW, and Medzhitov, R. 2009. Pattern recognition receptors and control of adaptive immunity. Immunological Reviews 227:221-233.

Park, BS, Song, DH, Kim, HM, Choi, BS, Lee, H, and Lee, JO. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191-U1130.

Reddy, KVR, Yedery, RD, and Aranha, C. 2004. Antimicrobial peptides: premises and promises. International Journal of Antimicrobial Agents 24:536-547.

Rock, FL, Hardiman, G, Timans, JC, Kastelein, RA, and Bazan, JF. 1998. A family of human receptors structurally related to Drosophila Toll. Proc. Natl. Acad. Sci. USA 95:588-593.

Sackton, TB, Lazzaro, BP, Schlenke, TA, Evans, JD, Hultmark, D, and Clark, AG. 2007. Dynamic evolution of the innate immune system in Drosophila. Nat. Genet. 39:1461-1468.

Sawyer, SL, Emerman, M, and Malik, HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. Plos Biology 2:1278-1285.

Sawyer, SL, Emerman, M, and Malik, HS. 2007. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. Plos Pathogens 3:1918-1929.

Sawyer, SL, Wu, LI, Emerman, M, and Malik, HS. 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. USA 102:2832-2837.

Schlenke, TA, and Begun, DJ. 2003. Natural selection drives drosophila immune system evolution. Genetics 164:1471-1480.

Sheldon, BC, and Verhulst, S. 1996. Ecological immunology: Costly parasite defenses and trade-offs in evolutionary ecology. Trends in Ecology & Evolution 11:317-321.

Stahl, EA, Dwyer, G, Mauricio, R, and Kreitman, MJ. 1999. Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature 96:302-306.

Takeuchi, O, Kawai, T, Sanjo, H, Copeland, NG, Gilbert, DJ, Jenkins, NA, Takeda, K, and Akira, S. 1999. TLR6: A novel member of an expanding Toll-like receptor family. Gene 231:59-65.

Tanaka, T, and Nei, M. 1989. Positive Darwinian Selection Observed at the Variable-Region Genes of Immunoglobulins. Mol. Biol. Evol. 6:447-459.

Tennessen, JA. 2005. Molecular evolution of animal antimicrobial peptides: widespread moderate positive selection. Journal of Evolutionary Biology 18:1387-1394.

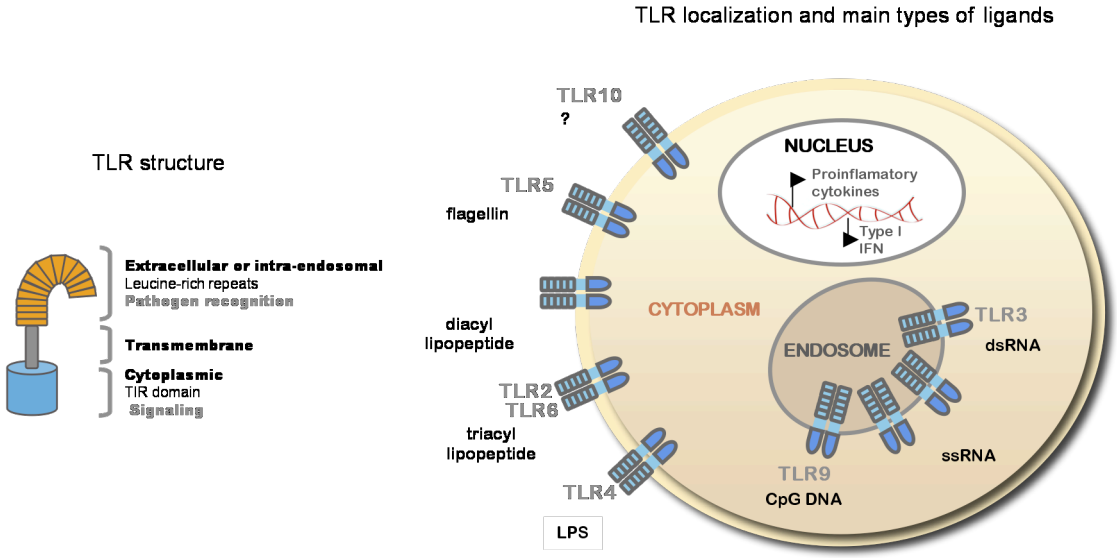
Van Valen, L. 1973. A new evolutionary law. Evolutionary Theory 1:1-30.

APPENDIX A: FIGURE

FIGURE LEGEND

Figure 1. Schematic representation of the typical TLR structure showing the extracellular or intra-endosomal region containing leucine-rich repeats responsible for pathogen recognition, the transmembrane portion and the cytoplasmic region containing a TIR domain responsible for signaling. The cellular localization and main types of ligands recognized by TLRs are shown. Adapted from Carpenter and O'Neill 2007.

Figure 1.


TLR structure
Extracellular or intra-endosomal
Leucine-rich repeats
Pathogen recognition
Transmembrane
Cytoplasmic
TIR domain
Signaling
TLR localization and main types of ligands
TLR10
?
TLR5
flagellin
diacyl
lipopeptide
TLR2
TLR6
triacyl
lipopeptide
TLR4
LPS
NUCLEUS
Proinflamatory
cytokines
Type I
IFN
CYTOPLASM
ENDOSOME
TLR3
dsRNA
ssRNA
TLR9
CpG DNA

APPENDIX B: A HISTORY OF RECURRENT POSITIVE SELECTION AT THE TOLL-LIKE RECEPTOR 5 IN PRIMATES.

Published: Molecular Biology and Evolution (2009) 26: 937-949

OXFORD UNIVERSITY PRESS LICENSE
TERMS AND CONDITIONS

Oct 13, 2009

This is a License Agreement between Gabriela Wlasiuk ("You") and Oxford University Press ("Oxford University Press") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Oxford University Press, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number	2287130236545
License date	Oct 13, 2009
Licensed content publisher	Oxford University Press
Licensed content publication	Molecular Biology and Evolution
Licensed content title	A History of Recurrent Positive Selection at the Toll-Like Receptor 5 in Primates
Licensed content author	Gabriela Wlasiuk, et. al.
Licensed content date	April 2009
Type of Use	Thesis / Dissertation
Institution name	University of Arizona
Title of your work	THE MOLECULAR EVOLUTION OF INNATE IMMUNITY GENES
Publisher of your work	University Microfilms Incorporated
Expected publication date	Dec 2009
Permissions cost	0.00 USD
Value added tax	0.00 USD
Total	0.00 USD
Terms and Conditions	

ABSTRACT

Many genes involved in immunity evolve rapidly. It remains unclear, however, to what extent pattern-recognition receptors (PRRs) of the innate immune system in vertebrates are subject to recurrent positive selection imposed by pathogens, as suggested by studies in *Drosophila*, or whether they are evolutionarily constrained. Here we show that TLR5, a member of the Toll-like receptor family of innate immunity genes that responds to bacterial flagellin, has undergone a history of adaptive evolution in primates. We have identified specific residues that have changed multiple times, sometimes in parallel in primates, and are thus likely candidates for selection. Most of these changes map to the extracellular leucine-rich repeats involved in pathogen recognition and some are likely to have an effect on protein function due to the radical nature of the amino acid substitutions that are involved. These findings suggest that vertebrate PRRs might show similar patterns of evolution to *Drosophila* PRRs, in spite of the acquisition of the more complex and specific vertebrate adaptive immune system. At shorter time scales, however, we found no evidence of adaptive evolution in either humans or chimpanzees. In fact, we found that one mutation that abolishes TLR5 function is present at high frequencies in many human populations. Patterns of variation indicate that this mutation is not young, and its high frequency suggests some functional redundancy for this PRR in humans.

INTRODUCTION

Vertebrate immune systems include acquired and innate components. Pattern-recognition receptors (PRRs) are an essential component of the innate immune system. PRRs recognize and bind pathogen-associated molecular patterns (PAMPs), conserved molecular motifs that are shared by infectious agents but which are absent in the host. The interaction between PRRs and PAMPs illustrates two fundamental aspects of the innate immune system: i) the ability to discriminate between self and non-self and ii) the targeting of components essential for microbial fitness, which are therefore functionally constrained (Medzhitov and Janeway 1997). Toll-like receptors (TLRs) constitute the best-characterized PRRs of the innate immune system of vertebrates, and so far, ten have been described in humans (Akira and Takeda 2004). After stimulation with their ligands, TLRs form homo- or hetero-dimers and trigger intracellular signaling cascades that induce the expression of a variety of genes. This in turn leads to the activation of innate immunity effector mechanisms as well as the development of adaptive immunity (Akira and Takeda 2004).

Because TLRs interact with microbial invaders, theory predicts that over evolutionary time they may be engaged in co-evolutionary arms races with their microbial ligands. Recent results from the comparison of several *Drosophila* genomes support this hypothesis, showing that among innate immunity loci, PRRs constitute a functional class that evolves quickly between species (Sackton et al. 2007). It remains unclear whether vertebrates and invertebrates are similar in this respect. On the other hand, given the extremely conserved nature of the molecular patterns targeted by TLRs,

they might be evolutionarily inflexible. In fact, they are often cited as an example of evolutionary conservation due to the fundamental constraint imposed by the inability of pathogens to tolerate mutations in molecular motifs that are essential to their fitness (Medzhitov and Janeway 1997).

Here, we attempt to distinguish between these two competing hypotheses using the evolution of Toll-like receptor 5 (TLR5) in primates as an example. So far, the limited evidence about the patterns of molecular evolution of TLRs in primates is inconclusive. While Ortiz et al. (2008) claimed that all TLRs, except for TLR1, have evolved under purifying selection in primates, Nakajima et al. (2008) using a more extensive phylogenetic sampling, suggested that TLR4, has been under positive selection in Old World monkeys.

TLR5 targets monomeric flagellin, the main component of the bacterial flagella and a potent virulence factor (Hayashi et al. 2001; Ramos, Rumbo, and Sirard 2004). Recently, Andersen-Nissen et al. (2005) showed that members of the α and ε Proteobacteria that are important human pathogens, such as *Campylobacter jejuni* and *Helicobacter pylori,* are able to evade TLR5 recognition by mutating key residues in the TLR5 recognition site. These mutations abolish flagellar motility, but the pathogens acquire compensatory mutations in other parts of the flagellin molecule that restore motility, which is essential for efficient infectivity. These results demonstrate that pathogens can evolve to evade PRR recognition while remaining fully functional and capable of infection. More importantly, these findings suggest opportunities for co-

evolution between PRRs and their microbial ligands, in spite of some overall functional constraint.

Additional motivation for studying PRRs in primates comes from ideas concerning the relationship between mating systems and disease risk. Based on the finding that the basal number of white blood cells (WBCs) in primates and carnivores is correlated with the degree of sexual promiscuity, but not with other life history traits expected to influence disease risk, Nunn and colleagues proposed the controversial idea that mating system drives the evolution of the immune system (Nunn, Gittleman, and Antonovics 2000; Nunn 2002; Nunn, Gittleman, and Antonovics 2003). The underlying hypothesis is that in promiscuous species the increased risk of acquiring sexually transmitted diseases has resulted in the evolution of stronger immune systems. This hypothesis has not been broadly tested at the molecular level. As a secondary goal, we take advantage of the variation in mating system among primate species to test predictions of this hypothesis.

A final motivation for studying TLR5 comes from association studies in humans which showed that a premature stop codon ($TLR5^{392STOP}$) was linked to susceptibility to Legionnaire's disease, a type of pneumonia produced by the flagellated bacterium *Legionella pneumophila* (Hawn et al. 2003), and resistance to two autoimmune disorders: Crohn's disease (Gewirtz et al. 2006) and Systemic Lupus Erythematosus (SLE) (Hawn et al. 2005). $TLR5^{392STOP}$ results in a loss of function, acts in a negative-dominant fashion (one defective copy is enough to produce a homodimer that is unable to signal), and has been reported to segregate in different populations at frequencies between 5 to 10 %

(Hawn et al. 2003; Hawn et al. 2005; Gewirtz et al. 2006). Hawn et al. (2005) suggested that the high population frequency of TLR5^{392STOP} might be due to an evolutionary advantage associated with defective TLR5-mediated signaling, at least in some situations. The "less is more" hypothesis proposed by Olson (1999) and Olson and Varki (2003) suggests that gene loss might be advantageous and an important engine of evolutionary change. This idea has received considerable attention recently in light of several reports of adaptive pseudogenizations in the human lineage (Tournamille et al. 1995; Ali, Rellos, and Cox 1998; Wang, Grus, and Zhang 2006; Seixas et al. 2007). The idea that TLR5 might be another case of adaptive gene loss in humans is intriguing because of its putative important immunologic function.

Here, we have analyzed the entire TLR5 coding sequence of 22 species of old and new world primates and apes in a phylogenetic framework, and surveyed sequence variation in both coding and non-coding regions in population samples of humans and chimpanzees to answer the following questions: 1) Has TLR5 undergone adaptive evolution in primates? 2) Is there any support for the promiscuity/disease-risk hypothesis in the rates of protein evolution across primates? 3) Are there signatures of positive selection in the patterns of nucleotide variation at TLR5 in humans and chimpanzees? and 4) Has the premature stop codon in humans increased in frequency due to recent positive selection? We found convincing evidence of positive selection at TLR5 throughout the primate phylogeny, involving amino acids that might mediate flagellin recognition, suggesting that innate immunity genes may experience some of the same evolutionary pressures previously described for adaptive immunity genes. Only four out

of six independent transitions to increased sexual promiscuity were associated with increased rates of protein evolution, arguing against the hypothesis that mating system plays a major role in TLR5 evolution. In humans and chimpanzees, patterns of DNA sequence variation are largely consistent with neutral expectations, suggesting that the relatively high frequency and widespread distribution of the human TLR5^{392STOP} mutation might be a consequence of functional redundancy.

MATERIALS AND METHODS

Samples: The species used in the phylogenetic analyses are shown in Figure 1 and the origins of the samples are given in Supplementary Table 1. Samples were collected in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines. Additionally, coding sequences of *Homo sapiens* and *Macaca mulatta* were retrieved from GenBank (accession numbers NM_003268 and XM_001099501 respectively).

DNA Samples from 19 *Pan troglodytes verus,* 3 *P. t. troglodytes* and 18 humans (9 Africans, 9 Europeans) 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 two non-overlapping fragments that together include ~17 kb were gathered from the Innate Immunity Database (www.innateimmunity.net).

Nine hundred and fifty individuals from the Human Genome Diversity Panel (Cann et al. 1999; Cann et al. 2002) were used to estimate the worldwide frequency and geographic distribution of the $TLR5^{392STOP}$ mutation. This HGDP excludes samples previously identified as related individuals or duplicates (Rosenberg 2006).

DNA Amplification and sequencing: The entire coding region of TLR5 (~2.5 kb) was PCR-amplified and sequenced from the 19 primate species listed above, using primers designed in conserved regions of published primate sequences. Together with the *Macaca* sequence from GenBank and the human and chimpanzee sequences (see below), the phylogenetic analyses included 22 species.

Two non-overlapping genomic fragments were PCR-amplified and sequenced from 18 humans (~12 kb and ~5kb) to match similar gene regions available from the Innate Immunity Database (see below). A 5 kb fragment was also PCR-amplified and sequenced in 19 *P. t. verus* and 3 *P. t. troglodytes*. In humans and chimpanzees, the sequenced regions contain the complete coding region as well as adjacent non-coding sequence.

PCR was performed in 25-50 μl reactions using Platinum *Taq* High Fidelity DNA Polymerase (Invitrogen, San Diego, CA). A complete list of amplification and sequencing primers for all fragments and the corresponding annealing temperatures and PCR protocols are provided in Supplementary Tables 2 and 3. 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). Sequences have been deposited in GenBank under the following accession numbers: Primates other than humans and chimpanzees: FJ542200-FJ542219; Chimpanzees: FJ546349-FJ546370; Humans: FJ556974-FJ556991.

Sequence editing and assembly were performed using SEQUENCHER (Gene Codes, Ann Arbor, MI). DNA sequences were aligned using CLUSTAL X (Thompson et al. 1997) with manual alignment of small indels using the amino acid sequence as a reference. Gametic phase was computationally determined using PHASE (Stephens, Smith, and Donnelly 2001).

Phylogeny-based tests of selection: We tested for positive selection in the primate phylogeny by comparing the number of nonsysnonymous substitutions per non-

synonymous site (dN) to the number of synonymous substitutions per synonymous site (dS) in a maximum likelihood (ML) framework. A ratio of dN/dS (ω) grater than one is usually taken as evidence of selection. We used the accepted primate phylogeny (Purvis 1995; Bininda-Emonds et al. 2007) in all analyses. We also used the TLR5 data to estimate phylogenetic relationships using Neighbor Joining. The resulting tree was similar to the well-accepted primate phylogeny (Bininda-Emonds et al. 2007) with only four branches placed in slightly different positions. Analyses of selection using the TLRs tree yielded very similar results to those obtained using the accepted primate phylogeny, so we report only the latter below.

First, we evaluated selection at individual codons, not allowing variation among lineages. We ran a series of nested models implemented in PAML ver 4 (Yang 1997; Yang 2007), in which the 'neutral' models restrict ω to values ≤1, while 'selection' models include a class of sites with dN/dS>1. A likelihood ratio test (LRT) was then used to compare nested models (Table 1). To check for convergence, all analyses were run twice, using initial ω values of 0.5 and 1.5. Amino acids under selection for model M8 were identified using a Bayes Empirical Bayes approach (BEB) (Yang, Wong, and Nielsen 2005). Two models of codon frequencies were used: F3x4 and F61.

A recent improvement in statistical methods to infer selection in a phylogenetic context is the incorporation of variation in the rate of synonymous substitution (Pond and Muse 2005). Kosakovsky Pond and Frost (2005) proposed a series of models to study selection on a codon basis. They classify previous methods as either 'counting methods', 'random effect models' or 'fixed effect models'. Counting methods reconstruct the

ancestral sequences to estimate the number of synonymous and non-synonymous changes at each codon. Random effect models assume a distribution of rates across sites and then infer the rate at which individual sites evolve. Fixed effects models estimate the ratio of non-synonymous to synonymous substitution on a site-by-site basis, without assuming *a priori* a distribution of rates across sites. SLAC (Single Likelihood Ancestor), REL (Random Effects Likelihood) and FEL (Fixed Effects Likelihood) methods, new versions of the 'counting', 'random effect' and 'fixed effect' models, respectively, that allow variation in the synonymous substitution rate (Kosakovsky Pond and Frost 2005), were implemented at the DATAMONKEY web server (Pond and Frost 2005).

Finally, to detect variation in ω among lineages, a model with one ω (M0) was compared with a 'free-ratio' model that allows each branch to have a separate ω value while keeping variation among sites constant (Nielsen and Yang 1998; Yang 1998). Because a parameter-rich model does not necessarily fit the data better than simpler models, a model selection scheme was performed in DATAMONKEY to find the variable-branch model with the best fit to the data.

Parallel amino acid changes were inferred using maximum parsimony in MacClade (Sinauer Associates, Sunderland, MA).

Population genetic analyses and tests of selection: Nucleotide heterozygosity, π (Nei and Li 1979) and the proportion of segregating sites, θ_w (Waterson 1975) were estimated for the entire human and chimpanzee datasets, and also for different functional regions (coding, non-coding), and different human populations separately.

Tajima's D (Tajima 1989) and Fu and Li's D* (Fu and Li 1993) were calculated to assess whether the allele frequency spectrum deviates from neutral expectations. Coalescent simulations, conditioned on the observed number of segregating sites, were used to generate the null distributions of these test statistics. The ratio of non-synonymous to synonymous polymorphisms in humans or chimpanzees was compared to the ratio of non-synonymous to synonymous fixed differences with respect to the orangutan sequence (Mcdonald and Kreitman 1991). These analyses were performed using DnaSP (Rozas et al. 2003) and SITES (Hey and Wakeley 1997). To test for selection at putative regulatory regions as in Andolfatto (2005), we compared the ratio of polymorphism within humans to human-chimpanzee divergence at silent sites in the coding region and at two 1 kb regions directly upstream of two alternative human transcripts.

To study population structure in the chimpanzee data, 50,000 neutral genealogies of 38 chromosomes were simulated under panmixia using the program 'ms' (Hudson 2002) using the observed level of variability and the recombination rate estimated from the data. To test for an excess of linkage disequilibrium (LD) due to admixture/population structure in chimpanzees, we computed the number of congruent sites (*pb*), defined as sites that determine only two haplotypes, and *gd,* defined as the maximum distance between any two congruent sites, using the script *lbcalc* (Garrigan et al. 2005). We then compared these values with the simulated distribution to calculate the probability of obtaining values more extreme than the observed ones.

To further evaluate the likelihood of gene flow between the chimpanzee subspecies, we fitted an isolation with migration model (Nielsen and Wakeley 2001; Hey and Nielsen 2004) using a Markov chain Monte Carlo method implemented in the program IMa (Hey and Nielsen 2007). Under this model, two populations split and diverge in isolation, with some level of gene flow. We used the largest non-recombining region of the combined *verus-troglodytes* sample, which includes 1660 bases of non-coding sequence, to run the program with a burn-in period of 2,000,000 steps using 15 chains with a geometric heating scheme. After the burn-in period, we ran the program for 15,399,385 steps, recording the results every 10 steps. We checked for convergence by comparing multiple runs.

Genotyping assay: The $TLR5^{392STOP}$ mutation was genotyped by restriction analysis with DdeI in the HGDP as in Hawn et al. (2003).

RESULTS

Positive selection on the extracellular domain of primate TLR5

We obtained the coding sequence of TLR5 for a relatively broad array of primates including New World primates, Old World primates and apes. To address whether specific codons in the protein have been subject repeatedly to positive directional selection in different species, we first investigated models in which the dN/dS ratio is allowed to vary among different classes of sites. LRTs showed that models that incorporate selection fit significantly better than neutral models (Table 1). For model 8, the most stringent of the models implemented in PAML, a small proportion of the codons (3.4% or 29 codons) was estimated to be under selection, with a ω value of 4.34, of which 13 were identified by the BEB approach with posterior probabilities above 0.8 (Table 2).

We then compared these results with those from methods that incorporate synonymous rate variation (Table 2). Using significance thresholds of $p<0.2$ for SLAC and FEL [consistent with a true Type I error rate of ~5%, as suggested by Kosakovsky Pond and Frost (2005) and a Bayes factor > 20 for REL (corresponding approximately to a p-value of 0.05)], SLAC and FEL identified 1 and 14 codons, respectively, and REL identified 11 codons as targets of selection. Eleven codons (104, 158, 292, 312, 354, 482, 523, 530, 567, 586 and 847) were picked by at least two methods.

Although not independent from previous results, we also considered parallel amino acid changes (independent changes at the same codon position, from the same initial state to the same final state) as potential candidates for selection. At TLR5, 24

codon sites show parallel evolution in two lineages and eight sites have evolved in parallel in three lineages (Table 2). Most of these do not fall at CpG sites (on either strand) and are thus not likely to be the product of mutational bias and/or increased mutation rate. Ten of the parallel changes (aa 158, 292, 312, 354, 482, 523, 530, 567, 586 and 847) correspond to sites that were identified by more than one ML method as targets of selection. Interestingly, parallel changes have not accumulated on specific branches, but instead are relatively scattered across the primate phylogeny. A possible explanation for the high number of parallel changes is functional constraint due to the presence of many leucine-rich repeats in the extracellular domain. Such motifs typically contain a conserved 11 aa motif (LXXLXLXXNXL, where "L" is Leu, Ile, Val or Phe, "N" is Asn, and X is any amino acid) and a variable region (Matsushima et al. 2007). In this case, all the parallel changes that occurred in the conserved portion of the LRRs, involve "X" residues, suggesting that if functional constraint to maintain this motif exists, it does not seem to be responsible for the high number of parallel changes. We thus infer that selection might have played a role in driving these substitutions.

We investigated the radical or conservative nature of amino acid substitutions using U, an empirically derived universal index based on the genetic code that measures amino acid exchangeability during evolution (Tang et al. 2004) (Table 2). In principle, more radical changes are more likely to affect function. U varies from 0.241 to 2.490 with lower values representing more radical (less common) changes (Tang et al. 2004). U is weakly correlated with other conventional measures, such as Grantham's distance, that determine amino acid exchangeability based on a combination of physicochemical

properties such as volume and polarity (Grantham 1974), but it is a considerably better predictor of the observed pattern of amino acid substitution in a variety of taxa (Tang et al. 2004). Several sites show relatively radical amino acid changes (Table 2).

Of the 11 sites that were identified by more than one ML method, amino acids 292, 312, 530, and 567 show the strongest evidence of selection because they were consistently identified by at least three ML methods, they show parallel changes, and they involve relatively radical amino acid changes. Particularly compelling is the evidence for selection on aa 530. This is the only site identified by all four ML methods, and it displays a radical change occurring in three independent lineages. Of the remaining seven sites, two deserve special attention. Site 354 involves a moderately radical change, and together with site 312, falls within the flagellin recognition domain (Andersen-Nissen et al. 2007). Site 847 also shows the same amino acid transition in 3 independent instances and is located in the very conserved TIR signaling domain.

Disease risk and mating system

Having shown that TLR5 evolution in primates is consistent with recurrent positive selection, we were interested in looking for heterogeneity in rates of protein evolution among different lineages and in investigating whether these differences were correlated with reported levels of sexual promiscuity. The best-fit model that allows variation in dN/dS among lineages grouped branches under four different rates: ω=3.13, ω=0.51, ω=0.25 and ω=0.06. The full model, which assigns a different rate to each branch, had a higher likelihood but not a significantly better fit than a model with a single

rate for all branches. Nonetheless, we compared the dN/dS values obtained in this full model with the variation in mating systems among species (Figure 1). We categorized mating systems as less promiscuous (monogamous + polygynous) or more promiscuous (promiscuous + dispersed), based on information compiled by Dixon (1998) and Lindenfors and Tullberg (1998). To avoid the problem of uncertainty in reconstructing mating system along long branches, we focused only on the terminal branches. We observed an increase in the rate of evolution associated with an increase in promiscuity in four of the six independent transitions from less promiscuous to more promiscuous mating systems (Figure 1). This was true when we included all sites, or when we included only the extracellular domain where most positively selected sites were located. For the extracellular domain, the average ω for more promiscuous branches, ($\bar{\omega}$=0.84; st. dev.=0.79), was higher than the average ω for less promiscuous branches ($\bar{\omega}$=0.46; st. dev.=0.22), but this difference was not significant (t-test, p=0.093). Thus, there is no compelling evidence for a causal link between mating system and molecular evolution at TLR5 in these data.

Human and chimpanzee polymorphism at TLR5

Levels of variation at TLR5 in humans are summarized in Table 3. In general, both coding and non coding regions showed polymorphism levels similar to those seen at other genes (Akey et al. 2004). Overall, humans presented an excess of rare variants with negative values of Tajima's D and Fu and Li's D* for both the coding and non-coding regions (Table 3). The African samples showed strongly negative values while the

European samples showed either less negative values (coding) or slightly positive (non-coding) values. Differences in the level and pattern of variation between the African and European samples in non-coding regions are largely in agreement with well-accepted demographic scenarios for African Americans and European Americans (Stajich and Hahn 2005). For example, our Tajima's D values are not outliers in the distribution of Tajima's D for a large set of genes sequenced in African Americans and European Americans (Stajich and Hahn 2005), suggesting that demographic effects rather than positive selection best explain the deviations from the null model at non-coding sites.

For the coding region, both the African and European samples showed a more pronounced excess of low frequency variants than in the non-coding region (Table 3). Tajima's D for non-synonymous sites was -1.495 and -1.020 for silent sites. This lower value for non-synonymous polymorphisms is consistent with the idea that some of these mutations may be weakly deleterious. This is also supported by a slightly, but not significantly, higher ratio of polymorphism to divergence for non-synonymous mutations than for synonymous mutations (Table 4) using polymorphism data from both humans and chimpanzees.

Of the 13 observed replacement changes observed in humans, three had frequencies above 5% [C1174T ($TLR5^{392STOP}$), freq=0.069; A1775G, freq=0.12; and T1846C, freq=0.29]. The high frequency of these mutations raises the question of whether they represent functional variants maintained at high frequency by selection. Merx et al. (2006) showed that only three of all known non-synonymous single nucleotide polymorphisms (SNPs) at TLR5 had functional effects when tested on a site-

by-site basis in transiently transfected CHO-K1 cells using reporter assays: one was $TLR5^{392STOP}$ and the other two were very rare SNPs not present in our sample. Each of these mutations resulted in a non-responsive receptor (loss of function) after stimulation with flagellin. The T1846C and A1775G mutations, on the other hand, result in an induction of expression comparable to the wildtype TLR5. Although these results have to be interpreted with caution, since they derive from in-vitro assays, they suggest that these mutations do not have a large functional effect. Thus, their high frequency might simply be due to drift.

We used a modified McDonald and Kreitman (MK) test to compare the ratio of polymorphism to divergence for silent versus putative regulatory sites as in Andolfatto (2005) and found no deviation from the neutral expectation (Table 5).

Levels of nucleotide variability in western chimpanzees (*P. t. verus*) are presented in Table 3 and are similar to genome-wide averages (Yu et al. 2003; Fischer et al. 2006). No significant deviations from neutrality were detected using tests of the allele frequency spectrum (Table 3) or the MK test (Table 4).

However, examination of the table of polymorphism revealed the presence of two major haplogroups (Figure 2, Supplementary Table 5). Divergence between these haplogroups was 0.15%, close to the average value between chimpanzee subspecies (Yu et al. 2003; Fischer et al. 2006). To gain more insight into the origin of this variation, we sequenced three individuals of *P. t. troglodytes.* We found that the least frequent haplotype class (8/38) from *P. t. verus* is present in *P. t. troglodytes* in 5 out of 6

chromosomes, while another haplotype present only in a single copy in *P. t. troglodytes* is more closely related to the major haplogoup in *P. t. verus* (Figure 2).

Three possible explanations for divergent haplotypes shared between subspecies are (*i*) unsorted ancestral polymorphism, (*ii*) admixture (i.e. gene flow between groups), or (*iii*) old balancing selection. Distinguishing among these is difficult. We note that the estimated divergence time between *P. t. verus* and *P. t. troglodytes* of 422,000 years (Won and Hey 2005) is less than the average time required to achieve reciprocal monophyly [$E(t) \approx 4N_e$ generations = 530,300 years, using the ancestral population size estimated by Won and Hey (2005) of N_e=5,300 and a generation time of 25 years]. Although the variance associated with this estimation is very large (Tajima 1983), this comparison suggests that ancestral variation could still be segregating between these subspecies. However, variation that is ancestral should have relatively little LD, whereas variation that is due to recent admixture should have higher levels of LD, an idea formalized into a test by Wall (2000) to detect ancient admixture in humans. We applied this test to our data. We computed the number of congruent sites (*pb*) and the maximum distance between any two congruent sites (*gd*), and compared these values with a simulated distribution generated by sampling neutral genealogies conditioned on the observed level of variation. The probability of obtaining both *pb*=6 and *gd*=0.285 under panmixia was 0.039 (using the level of recombination estimated from the data), indicating the existence of population structure or historical gene flow. We also fitted an isolation model with gene flow, as in Won and Hey (2005), and found evidence of gene flow between subspecies, although most of this gene flow was from *P. t. verus* to *P. t.*

troglodytes. In light of the relative excess of LD revealed by the Wall test, some form of admixture or population structure seems to be the most likely explanation for the patterns of variation seen at TLR5 in *P. t. verus*, although we note that more complex scenarios involving retention of ancestral polymorphism and selection could also contribute to the observed patterns.

Distribution, frequency, and haplotype structure of TLR5^{392STOP} in humans

Two lines of evidence suggest that TLR5^{392STOP} has functional consequences. First, *in vitro* assays showed that it encodes a defective receptor (Merx et al. 2006). Second, it is associated with disease phenotypes in human populations (Hawn et al. 2003; Hawn et al. 2005; Gewirtz et al. 2006). Because of these observations we were interested in measuring the frequency of TLR5^{392STOP} in different populations and exploring the idea that this mutation might be under recent strong positive selection in humans. We genotyped the mutation in the Human Genome Diversity Panel (HGDP), and estimated a global frequency of 4.2%. The genotype frequencies were close to Hardy Weinberg expectations (Supplementary Material Table 5). TLR5^{392STOP} is distributed nearly worldwide, with the mutation present in at least one copy in approximately half of the populations sampled (Figure 3). Since the mutation is often relatively rare, it is possible that the mutation is present at low frequencies in more populations than those reported here. Notably, some populations in the Middle East and Southern Asia have considerably higher frequencies, such as Balochi and Baruscho from Pakistan (14.5% and 12.0%

respectively), Miaozu and Naxi from China (10.0% and 11.0% respectively), Cambodia (16.7%), Papua-New Guinea (14.7%) and Melanesia (22.7%).

If $TLR5^{392STOP}$ has increased in frequency due to selection in the recent past, the mutation is expected to be embedded in unusually long haplotypes. For example, selection at G6pd has generated LD over more than 1 Mb (Saunders et al. 2005). However, only two sites (positions 9946 and 11185) show significant LD (measured as D') with $TLR5^{392STOP}$ (position 33309) after Bonferroni correction for multiple testing (Table 6). The distances between $TLR5^{392STOP}$ and these sites are 23,963 and 22,724 nucleotides, respectively. Because $TLR5^{392STOP}$ (or any linked marker) is not present in the Hapmap we were not able to evaluate the extent of LD at longer distances, but the fact that the haplotype containing $TLR5^{392STOP}$ extends less than 25 kb suggests that if selection is responsible for the actual frequencies, it is not recent and strong.

DISCUSSION

Immunity genes are among the fastest evolving classes of genes in mammalian genomes (Gibbs et al. 2004; Mikkelsen et al. 2005; Nielsen et al. 2005), an observation that is usually interpreted as evidence of positive selection due to their potential engagement in host-pathogen co-evolution. Despite this generalization, it has been unclear whether genes of the adaptive and innate branches of immunity show similar patterns of evolution or whether they are characterized by very different levels of functional constraint. By studying both phylogeny-based estimates of evolutionary rates and patterns of nucleotide variation within and between closely related primate species, we sought to provide an integrated understanding of the molecular evolution of an innate immunity receptor at different evolutionary timescales.

Positive selection at the extracellular domain of TLR5 in primates

The results of several ML approaches provide strong evidence that TLR5 has experienced positive selection in primates. Conservatively, we identified 11 sites that show congruence between different ML methods as the strongest candidates of adaptive evolution. Of these, 10 sites are localized in leucine-rich repeats of the extracelullar domain (Table 2), and three are located within a 228 aa region where the putative flagellin recognition site lies (Andersen-Nissen et al. 2007). Although we still do not have a complete picture of the flagellin-TLR5 interaction surface, this observation strengthens the case of adaptive evolution at TLR5. Moreover, based on the modeled three-dimensional structure of the extracellular domain, Andersen-Nissen et al. (2007)

hypothesized that amino acids near a conserved concavity within the 228 aa region could mediate species-specific patterns of TLR5 recognition. It is worth noting that site 268 lies adjacent to a residue (267) that was identified by mutagenesis as responsible for differences in specificity between human and mouse TLR5 (Andersen-Nissen et al. 2007). It is possible that residue 268 or some of the other sites identified as candidates for being under selection are also involved in TLR5 species-specificity, a matter that functional studies will be able to clarify. 'U', the evolutionary index (Tang et al. 2004), provides additional information about the likelihood that specific mutations affect function and thus may be under selection. Six of the 11 sites under selection (aa292, aa312, aa354, aa482, aa530 and aa567) show relatively radical changes, with U ranging between 0.375 and 0.732.

The identification of several sites under selection within the pathogen interaction domain fits the expectation of co-evolutionary models. This is in line with the finding that several flagellated Proteobacteria are able to evade human TLR5 recognition (Andersen-Nissen et al. 2005). However, we note that only a small proportion of sites (11/858=1.3%) show strong evidence of positive selection. Thus, most of the protein, including the TIR (signaling) domain, shows strong functional constraint, in agreement with the most generally accepted paradigm of Toll-like receptor function. This duality of strong positive selection on a few sites against a background of strong purifying selection over most of the TLR5 protein is in sharp contrast with antiretroviral genes such as APOBEC3G, TRIM5α. These genes show a much larger proportion of sites (30% and 18% respectively) under positive selection (Sawyer, Emerman, and Malik 2004; Sawyer

et al. 2005). These differences between TLR5 and APOBEC3G or TRIM5α may reflect general differences between PRRs and genes involved in 'intrinsic' immunity (i.e. genes that typically do not participate in the classic innate immunity pathways but nevertheless can restrict certain retroviruses). These differences might also reflect differences between genes whose products interact with bacteria versus those whose products interact with viruses. It is possible for example, that due to their higher mutation rates and faster turnover, viruses impose stronger selection than do bacteria.

Vertebrate immune systems differ from invertebrate immune systems in many ways, but most notably in the presence of an adaptive immune response. The acquisition of adaptive immunity could have fundamentally changed the evolutionary dynamics of vertebrate PRRs. The recent publication of genome-wide patterns of evolution of innate immunity genes in *Drosophila* by Sackton at al. (2007) allows us to start comparing patterns of evolution of PRRs and other classes of innate immunity genes between *Drosophila* and vertebrates. Using a similar codon-based ML approach as the one used here, Sackton at al. (2007) found that among 245 *Drosophila* immunity genes, PRRs constitute the class with the highest proportion of positively selected sites (followed by signaling peptides and then antimicrobial peptides) in the *D. melanogaster* group. In contrast, Schlenke and Begun (2003) reported that adaptive fixations are also common in signaling molecules in *D. simulans*. In vertebrates, similar genome-wide analyses of innate immunity genes are missing, but some evidence points to the possibility that innate immunity genes are also under strong selection. Recent examples include APOBEC3G and TRIM5α (Sawyer, Emerman, and Malik 2004; Sawyer et al. 2005), TRIM22

(Sawyer, Emerman, and Malik 2007), TLR4 (Nakajima et al. 2008), RNASEL (Summers and Crespi 2008) and PKR (Elde et al. 2008). Our results demonstrate that some PRRs can also evolve rapidly between species.

Mating system and molecular evolution of immunity genes

We tested the mating system/disease risk hypothesis with six phylogenetically independent contrasts between promiscuous and monogamous/polygynous mating systems in the primate phylogeny. We found that dN/dS changed in the predicted direction in four of six cases, and that the average dN/dS was not significantly higher in more promiscuous lineages. Thus, rates of molecular evolution at TLR5 do not seem to support this controversial hypothesis, and suggest that lineage-specific effects are more important than the effect, if any, of mating system. A more complete test will require analysis of similar data from many immunity genes. An interesting observation is that the increase in ω in the more promiscuous group was accompanied by an increased variance. It is possible that promiscuous mating systems are associated with stronger natural selection on immunity genes only some of the time (or only on a subset of these genes) leading to a higher average ω and also to a greater variance in ω in more promiscuous lineages compared to less promiscuous lineages.

Patterns of nucleotide variation in humans and chimpanzees

Patterns of nucleotide variation within humans and chimpanzees were largely consistent with neutral expectations. The deviations from neutral predictions in the

spectrum of allele frequencies were similar to those seen at other genes, suggesting that demographic effects, rather than selection, are responsible for these patterns. Thus, despite the strong evidence for adaptive evolution at TLR5 over deeper evolutionary timescales in primates (see above), we did not find evidence for adaptive evolution within humans or chimpanzees. This suggests that adaptive evolution at TLR5 may be somewhat episodic, or at least not marked by continual turnover of new adaptive alleles as might be expected under an arms race model of host-pathogen co-evolution. A model of episodic selection would be more compatible with systems in which pathogens do not show stable associations with hosts but instead infect hosts sporadically.

We estimated the rate of adaptive fixations from our phylogenetic comparisons to get a sense of the likelihood of detecting selection within species. Using the 11 sites with the strongest evidence of selection (Table 2), we estimated the rate of adaptive fixation by dividing the total number of amino acid substitutions (39) at these sites by the total length of the tree (417.2 MY) using divergence times from Bininda-Emonds at al. (2007). This yielded a value of approximately one adaptive fixation every 10 MY. This is probably an underestimate of the true rate, because the ML methods used here only have power to detect recurrent positive selection on the same sites. However, even if the true rate was an order of magnitude higher than this estimate, it would not be surprising to fail to find evidence of selection within humans or chimpanzees. Polymorphism-based tests of selection typically have power to detect selection over fairly recent time scales, often on the order of less than N_e generations (~250,000 years in humans) (Braverman et al. 1995; Simonsen, Churchill, and Aquadro 1995; Przeworski 2002).

One result worth noting was the observation of low-frequency replacement polymorphisms in humans. These polymorphisms contribute to a ratio of replacement to silent variation that is slightly higher within species than between species (Table 4). Along with negative values of Tajima's D for replacement polymorphisms, this suggests that many of these polymorphisms may be weakly deleterious, consistent with the general pattern of functional constraint revealed by the phylogenetic analysis.

Patterns of nucleotide variation within chimpanzees differed from those seen in humans. Levels of variation were lower in chimpanzees, in spite of similar effective population sizes (or slightly higher in *P. t. verus*) (Yu et al. 2004). The distribution of allele frequencies differed too, with an excess of rare variants in humans and a trend towards an excess of intermediate frequency variants in chimpanzees at non-coding sites. Chimpanzees exhibited two divergent haplogroups in both *P. t. verus* and in *P. t. troglodytes*. The presence of these shared haplotypes is probably best explained by gene flow between subspecies at some point in the recent past or by some more complicated form of population structure.

Is the human TLR5 redundant?

Recently, several cases of adaptive gene loss in humans have been reported (Tournamille et al. 1995; Ali, Rellos, and Cox 1998; Wang, Grus, and Zhang 2006; Seixas et al. 2007). This somewhat counterintuitive idea, positive selection favoring gene loss, has been proposed as a potentially important mechanism in human evolution (Olson 1999; Olson and Varki 2003).

$TLR5^{392STOP}$, a loss-of-function mutation, segregates in humans at a considerable frequency along with the normal variants. This raises the question of whether (*i*) it is being constantly generated by recurrent mutation, (*ii*) it has increased in frequency due to positive selection, in which case there might be a trade-off between the disadvantage of loosing the function and some other benefit, or (*iii*) it has drifted in the population to its present frequency.

The frequency of $TLR5^{392STOP}$ is clearly not compatible with mutation-selection balance. Assuming a mutation rate, μ, of $2x10^{-8}$ (or $\sim10^{-7}$ for a CpG site) (Nachman and Crowell 2000) and an equilibrium frequency, q_e, of 0.042 we can calculate the selection coefficient, s, against a dominant mutation as $s \approx \mu/q_e$ (Haldane 1932). The estimated s (6.0×10^{-7}, or 2.4×10^{-6} for a CpG site) is so small as to be effectively neutral in human populations, where the effective population size is approximately 10,000 (Zhao et al. 2006). If s was 0.01, then the mutation rate would have to be on the order of 10^{-4} to account for the observed frequencies, and this is clearly unrealistic. Moreover, the fact that the $TLR5^{392STOP}$ always appears on the same haplotype argues against recurrent mutation.

We found no evidence of strong, recent selection on $TLR5^{392STOP}$ either in patterns of LD, which were unremarkable, or in levels of variability, which were average. Moreover, the distribution of allele frequencies at TLR5 fits well with generally accepted demographic models. This leaves drift as the most likely explanation for the present frequency of $TLR5^{392STOP}$. Given the difficulties of detecting selection from polymorphism data in humans, we cannot rule out the possibility that $TLR5^{392STOP}$ has

been under weak positive selection, especially in light of the phenotypes associated with this mutation. For instance, SLE has a relatively high prevalence (up to 160/100,000) and mostly affects women in reproductive age (Danchenko, Satia, and Anthony 2006) making the hypothesis of selection for protection against autoimmune diseases at least reasonable. There are marked geographic differences in SLE burden that might reflect underlying genetic variation for resistance/susceptibility or variation in environmental factors. Microbial infections are common triggers of autoimmunity through TLRs (Anders et al. 2005). It would be interesting to correlate worldwide abundance of flagellated pathogens with the prevalence of SLE.

If drift took the mutation to its present frequency, then the mutation must be relatively old. An estimate of the age of an allele based on its frequency, q, is given by $E(t)=(-2q)(\ln q)/(1-q)$, where age is measured in units of 2N generations (Kimura and Ohta 1973). The global frequency of $TLR5^{392STOP}$ is 0.042. Assuming that N=10,000, the estimated age is 5,560 generations, or 139,000 years (assuming a generation of 25 years). Another way to estimate the age of the $TLR5^{392STOP}$ mutation is from the decay of LD as a function of time and recombination rate. The time required to erode linkage to the observed level is given by: $t = \ln (D'_t/D'_0)/ \ln(1-c)$ (Hedrick 2000), where D'_t is the observed LD in the data, D'_0 is the initial LD (assumed to be complete when the $TLR5^{392STOP}$ mutation arose, D'=1), and c is the recombination distance calculated using the average recombination rate for chromosome 1 of 1.2 cM/Mb (Jensen-Seaman et al. 2004). Using five sites that show significant LD (Table 6) t was estimated as 2,096 generations or 52,398 years.

The observation that TLR5^{392STOP}, a null variant, is present at frequencies up to 23% in some populations suggests that TLR5 function might be partially compensated by other genes (i.e. functional redundancy for TLR5). A similar case is provided by Verdu et al. (2006) who, based on the patterns of nucleotide variation and absence of extended LD concluded that MBL2, another innate immune receptor that activates the lectin-complement pathways, is functionally redundant in human innate immune defenses. Redundancy in PAMP recognition might be a common theme in the innate immune response (Miao et al. 2007). The recognition of viral RNA provides a good example in which several TLRs participate in the detection of ssRNA and dsRNA in endosomal compartments, while another suite of genes responds to the same PAMPs in the cytosol. In fact, human carriers of null mutations at TLR3 are susceptible to herpes simplex virus 1 encephalitis but seem to show normal responses against other viruses (Zhang et al. 2007). It is possible that this recognition at multiple levels is an important and previously unappreciated feature of the innate immune system. The recent identification of a second flagellin receptor (cytosolic), Ipaf (Franchi et al. 2006), is consistent with this idea. However, the downstream effects of both genes are quite different, and they also respond to different types of bacteria [reviewed in (Miao et al. 2007)], suggesting that TLR5 and Ipaf might cooperate in recognizing flagellated bacteria rather than being completely functionally redundant.

ACKNOWLEDGEMENTS

We especially thank the following people/institutions for providing DNA or tissue samples: Drs. M. Hammer, O. Ryder, and B. Beer, The Museum of Vertebrate Zoology, Berkeley, California, The Southwest National Primate Research Center, and the Gladys Porter, Toronto, San Diego, and Los Angeles Zoos. We thank Dr. W. Klimecki for providing access to the raw human sequence data deposited in the Innate Immunity Database, Dr. M. Saunders and Dr. H. Norton for help with analysis, A. Woerner for the use of scripts to handle/analyze polymorphism data, and Donata Vercelli, Armando Geraldes, Matt Dean, Jeff Good, Miguel Carneiro, Tovah Salcedo and Mari Sans-Fuentes for very useful discussions of the data. This work was supported by an NIH grant to MWN (GM074245).

REFERENCES

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

Akira, S, and Takeda, K. 2004. Toll-like receptor signalling. Nature Reviews Immunology 4:499-511.

Ali, M, Rellos, P, and Cox, TM. 1998. Hereditary fructose intolerance. Journal of Medical Genetics 35:353-365.

Anders, HJ, Zecher, D, Pawar, RD, and Patole, PS. 2005. Molecular mechanisms of autoimmunity triggered by microbial infection. Arthritis Res Ther 7:215-224.

Andersen-Nissen, E, Smith, KD, Bonneau, R, Strong, RK, and Aderem, A. 2007. A conserved surface on Toll-like receptor 5 recognizes bacterial flagellin. Journal of Experimental Medicine 204:393-403.

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

Andolfatto, P. 2005. Adaptive evolution of non-coding DNA in Drosophila. Nature 437:1149-1152.

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

Braverman, JM, Hudson, RR, Kaplan, NL, Langley, CH, and Stephan, W. 1995. The Hitchhiking effect on the site frequency-spectrum of DNA polymorphisms. Genetics 140:783-796.

Cann, HM, de Toma, C, Cazes, L et al. 2002. A human genome diversity cell line panel. Science 296:261-262.

Cann, HM, De Toma, C, Marcadet-Troton, A, Thomas, G, Dausset, J, and Cavalli-Sforza, LL. 1999. The HGDP-CEPH human genome diversity panel. Am. J. Hum. Genet. 65:A198-A198.

Danchenko, N, Satia, JA, and Anthony, MS. 2006. Epidemiology of systemic lupus erythematosus: a comparison of worldwide disease burden. Lupus 15:308-318.

Dixon, AF. 1998. Primate Sexuality. Oxford University Press, New York.

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

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

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

Fu, YX, and Li, WH. 1993. Statistical Tests of Neutrality of Mutations. Genetics 133:693-709.

Garrigan, D, Mobasher, Z, Kingan, SB, Wilder, JA, and Hammer, MF. 2005. Deep haplotype divergence and long-range linkage disequilibrium at Xp21.1 provide evidence that humans descend from a structured ancestral population. Genetics 170:1849-1856.

Gewirtz, AT, Vijay-Kumar, M, Brant, SR, Duerr, RH, Nicolae, DL, and Cho, JH. 2006. Dominant-negative TLR5 polymorphism reduces adaptive immune response to flagellin and negatively associates with Crohn's disease. American Journal of Physiology-Gastrointestinal and Liver Physiology 290:G1157-G1163.

Gibbs, RA, Weinstock, GM, Metzker, ML et al. 2004. Genome sequence of the Brown Norway rat yields insights into mammalian evolution. Nature 428:493-521.

Grantham, R. 1974. Amino-Acid Difference Formula to Help Explain Protein Evolution. Science 185:862-864.

Haldane, JBS. 1932. The causes of Evolution. Longmans, Green & Co., London.

Hawn, TR, Verbon, A, Lettinga, KD et al. 2003. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. Journal of Experimental Medicine 198:1563-1572.

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

Hayashi, F, Smith, KD, Ozinsky, A, Hawn, TR, Yi, EC, Goodlett, DR, Eng, JK, Akira, S, Underhill, DM, and Aderem, A. 2001. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. Nature 410:1099-1103.

Hedrick, PW. 2000. Genetics of populations. Jones and Bartlett Publishers Inc., Sudbury.

Hey, J, and Nielsen, R. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of Drosophila pseudoobscura and D-persimilis. Genetics 167:747-760.

Hey, J, and Nielsen, R. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. P. Natl. Acad. Sci. USA 104:2785-2790.

Hey, J, and Wakeley, J. 1997. A coalescent estimator of the population recombination rate. Genetics 145:833-846.

Hudson, RR. 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18:337-338.

Jensen-Seaman, MI, Furey, TS, Payseur, BA, Lu, YT, Roskin, KM, Chen, CF, Thomas, MA, Haussler, D, and Jacob, HJ. 2004. Comparative recombination rates in the rat, mouse, and human genomes. Genome Research 14:528-538.

Kimura, M, and Ohta, T. 1973. Age of a Neutral Mutant Persisting in a Finite Population. Genetics 75:199-212.

Kosakovsky Pond, SL, and 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.

Lindenfors, P, and Tullberg, BS. 1998. Phylogenetic analyses of primate size evolution: the consequences of sexual selection. Biol. J Linn. Soc. 64:413-447.

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

Mcdonald, JH, and Kreitman, M. 1991. Adaptive Protein Evolution at the Adh Locus in Drosophila. Nature 351:652-654.

Medzhitov, R, and Janeway, CA. 1997. Innate immunity: The virtues of a nonclonal system of recognition. Cell 91:295-298.

Merx, S, Zimmer, W, Neumaier, M, and Ahmad-Nejad, P. 2006. Characterization and functional investigation of single nucleotide polymorphisms (SNPs) in the human TLR5 gene. Human Mutation 27:293.

Miao, EA, Andersen-Nissen, E, Warren, SE, and Aderem, A. 2007. TLR5 and Ipaf: Dual sensors of bacterial flagellin in the innate immune system. Seminars in Immunopathology 29:275-288.

Mikkelsen, TS, Hillier, LW, Eichler, EE et al. 2005. Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69-87.

Nachman, MW, and Crowell, SL. 2000. Estimate of the mutation rate per nucleotide in humans. Genetics 156:297-304.

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

Nei, M, and Li, WH. 1979. Mathematical-Model for Studying Genetic-Variation in Terms of Restriction Endonucleases. Proc. Natl. Acad. Sci. USA 76:5269-5273.

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

Nielsen, R, and Wakeley, J. 2001. Distinguishing migration from isolation: A Markov chain Monte Carlo approach. Genetics 158:885-896.

Nielsen, R, and Yang, ZH. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936.

Nunn, CL. 2002. A comparative study of leukocyte counts and disease risk in primates. Evolution 56:177-190.

Nunn, CL, Gittleman, JL, and Antonovics, J. 2000. Promiscuity and the primate immune system. Science 290:1168-1170.

Nunn, CL, Gittleman, JL, and Antonovics, J. 2003. A comparative study of white blood cell counts and disease risk in carnivores. P. Roy. Soc. Lond. B Bio. 270:347-356.

Olson, MV. 1999. When less is more: Gene loss as an engine of evolutionary change. Am. J. Hum. Genet. 64:18-23.

Olson, MV, and Varki, A. 2003. Sequencing the chimpanzee genome: Insights into human evolution and disease. Nature Reviews Genetics 4:20-28.

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

Pond, SK, and Muse, SV. 2005. Site-to-Site Variation of Synonymous Substitution Rates. Mol Biol Evol 22:2375-2385.

Pond, SLK, and Frost, SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531-2533.

Przeworski, M. 2002. The signature of positive selection at randomly chosen loci (vol 160, pg 1179). Genetics 162:2053-2053.

Purvis, A. 1995. A Composite Estimate of Primate Phylogeny. Philos T Roy Soc B 348:405-421.

Ramos, HC, Rumbo, M, and Sirard, JC. 2004. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. Trends in Microbiology 12:509-517.

Rosenberg, NA. 2006. Standardized subsets of the HGDP-CEPH human genome diversity cell line panel, accounting for atypical and duplicated samples and pairs of close relatives. Ann Hum Genet 70:841-847.

Rozas, J, Sanchez-DelBarrio, JC, Messeguer, X, and Rozas, R. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496-2497.

Sackton, TB, Lazzaro, BP, Schlenke, TA, Evans, JD, Hultmark, D, and Clark, AG. 2007. Dynamic evolution of the innate immune system in Drosophila. Nat. Genet. 39:1461-1468.

Saunders, MA, Slatkin, M, Garner, C, Hammer, MR, and Nachman, MW. 2005. The extent of linkage disequilibrium caused by selection on G6PD in humans. Genetics 171:1219-1229.

Sawyer, SL, Emerman, M, and Malik, HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. Plos Biology 2:1278-1285.

Sawyer, SL, Emerman, M, and Malik, HS. 2007. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. Plos Pathogens 3:1918-1929.

Sawyer, SL, Wu, LI, Emerman, M, and Malik, HS. 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. USA 102:2832-2837.

Schlenke, TA, and Begun, DJ. 2003. Natural selection drives drosophila immune system evolution. Genetics 164:1471-1480.

Seixas, S, Suriano, G, Carvalho, F, Seruca, R, Rocha, J, and Di Rienzo, A. 2007. Sequence diversity at the proximal 14q32.1 SERPIN subcluster: Evidence for natural selection favoring the pseudogenization of SERPINA2. Mol. Biol. Evol. 24:587-598.

Simonsen, KL, Churchill, GA, and Aquadro, CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141:413-429.

Stajich, JE, and Hahn, MW. 2005. Disentangling the effects of demography and selection in human history. Mol. Biol. Evol. 22:63-73.

Stephens, M, Smith, NJ, and Donnelly, P. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978-989.

Summers, K, and Crespi, B. 2008. Molecular evolution of the prostate cancer susceptibility locus RNASEL: Evidence for positive selection. Infection Genetics and Evolution 8:297-301.

Swanson, WJ, Nielsen, R, and Yang, QF. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. 20:18-20.

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437-460.

Tajima, F. 1989. Statistical-Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. Genetics 123:585-595.

Tang, H, Wyckoff, GJ, Lu, J, and Wu, CI. 2004. A universal evolutionary index for amino acid changes. Mol. Biol. Evol. 21:1548-1556.

Thompson, JD, Gibson, TJ, Plewniak, F, Jeanmougin, F, and 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.

Tournamille, C, Colin, Y, Cartron, JP, and Levankim, C. 1995. Disruption of a Gata Motif in the Duffy Gene Promoter Abolishes Erythroid Gene-Expression in Duffy Negative Individuals. Nat. Genet. 10:224-228.

Verdu, P, Barreiro, LB, Gessain, A et al. 2006. Evolutionary insights into the high worldwide prevalence of MBL2 deficiency alleles. Hum Mol Genet 15:2650-2658.

Wall, JD. 2000. Detecting ancient admixture in humans using sequence polymorphism data. Genetics 154:1271-1279.

Wang, XX, Grus, WE, and Zhang, JZ. 2006. Gene losses during human origins. Plos Biology 4:366-377.

Waterson, GA. 1975. On the number of segregating sites in genetical models without recombination. Theoretical population biology 7.

Won, YJ, and 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, and Pedersen, AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.

Yang, ZH, Wong, WSW, and Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22:1107-1118.

Yu, N, Jensen-Seaman, MI, Chemnick, L, Kidd, JR, Deinard, AS, Ryder, O, Kidd, KK, and Li, WH. 2003. Low nucleotide diversity in chimpanzees and bonobos. Genetics 164:1511-1518.

Yu, N, Jensen-Seaman, MI, Chemnick, L, Ryder, O, and Li, WH. 2004. Nucleotide diversity in gorillas. Genetics 166:1375-1383.

Zhang, SY, Jouanguy, E, Ugolini, S. et al. 2007. TLR3 deficiency in patients with Herpes simplex encephalitis. Science, 317:1522-1527.

Zhao, ZM, Yu, N, Fu, YX, and Li, WH. 2006. Nucleotide variation and haplotype diversity in a 10-kb noncoding region in three continental human populations. Genetics 174:399-409

Table 1. Tests for Positive selection at TLR5 [a, b].

Model Category [b]	Models compared [c]	χ^2 [d]	d.f. [e]	p-value	p_s [f]	ω_{sel} [g]
Site models	M1 vs M2	22.132	2	<0.0001	0.029	4.55
	M7 vs M8	23.002	2	<0.0001	0.034	4.54
	M8a vs M8	22.796	1	<0.0001		
Branch models	M0 vs Full	31.115	40	0.842		

[a] Analysis using the F3x4 or F61 models of codon frequencies yielded virtually identical results; the results presented here refer to the F3x4 model.
[b] LRTs were performed between nested models that allow variation in dN/dS among codons but not branches ("sites" models) or between models that allow variation among branches but not codons ("branch" models).
[c] In the case of "site models" we performed three comparisons, each involving a null model (M1, M7, M8a) and a positive selection model (M2, M8). Specifically, we compared models M1 (two classes of sites with rates, $\omega_0<1$, $\omega_1=1$) vs M2 (three rates $\omega_0<1$, $\omega_1=1$, $\omega_{sel}>1$), and M7 (fit to a beta distribution, 10 rates) vs M8 (fit to a beta distribution with an extra rate that allows $\omega_{sel}>1$) (Nielsen and Yang 1998; Yang et al. 2000). Additionally, the M8a model proposed by Swanson et al. (2003) was compared to M8.
[d] $-2\ln\Delta L$ (where ΔL is the difference in likelihoods between the nested models) is distributed approximately as χ^2
[e] d.f. = Degrees of freedom, equal to the difference in the number of parameters between the models.
[f] Proportion of the sites under selection.
[g] Estimated dN/dS of the sites under selection.

Table 2. TLR5 amino acid sites under positive selection identified using different method.

AA position[a]	AA change (No. of parallel changes)	Protein domain[b]	Maximum Likelihood method				Parallel change	U[f]	Clade[g]
			PAML M8[c]	SLAC	FEL[d]	REL[e]			
14	Val-Met (2)	Signal P.					YES	0.986	A, N
29	Arg-Gln (2)	LRRNT					YES	1.045	A, O
104	Asp-Ser	LRR 3	0.908		0.158	63		*	A, O, amb.
	Ser-Gly							1.360	
	Ser-Asn							2.053	
158	Arg-His (3)	LRR 5			0.173	22	YES	1.317	A, O, N
168	Lys-Glu (2)	LRR 5					YES	0.548	N, amb
[181]	Gln_Lys	LRR6			0.192			1.466	O, N
	Gln-Arg							1.045	
[197]	Thr-Met (2)	LRR 6					YES	1.007	O, N
[207]	Asn-Ser (2)	LRR 7					YES	2.053	A, N
[230]	Thr-Ileu (2)	LRR8					YES	0.750	A, N
[262]	His-Tyr (2)	LRR9					YES	0.665	O, N
[268]	Gly-Ser or Gly-Thr	LRR 9	0.883					1.36 or *	N, amb
	Ser-Thr							2.49	
[280]	Asn-Ser (2)	LRR 9					YES	2.053	A, N
[292]	His-Arg (3)	LRR 10	0.925		0.102	86	YES	1.317	A, N, amb
	His-Leu or Arg-Leu							0.56 or 0.414	
[312]	Gln-Arg	LRR 10	0.990		0.069	113	YES	1.045	A, N, amb
	Arg-Gly (2)							0.534	
	Gln-Lys							1.466	
[354]	Ser-Trp or Ser-Leu	LRR 12	0.972			28	YES	0.375 or 0.725	A, N, amb
	Trp-Leu (2)							0.793	
	Ser-Ala							2.38	
[363]	Ala-Thr (2)	LRR 13					YES	1.587	N, amb
[400]	His-Tyr (2)	LRR 14					YES	0.665	O, amb
407	Asp-Ala (2)	LRR15	0.897				YES	0.657	O, amb
	Asp-Asn (1 or 2)							1.015	
416	Ala-Val (2)	LRR 15					YES	1.017	N, amb

446	Arg-Gln or Arg-Glu Gln-Glu	LRR16			0.187			1.045 or * 1.634	N, amb
460	Leu-Phe	LRR17			0.185			0.732	A
482	Glu-Gly (3)	LRR18			0.185	22	YES	0.553	O
492	Glu-Gln (3) Glu- Ala	LRR18	0.951				YES	1.634 0.906	A, O
496	Asp-Asn (2)	LRR18					YES	1.015	A, O
523	Ser-Lys (2 or 3)	LRR19	0.995		0.110	166	YES	*	O
530	Gly-Arg (3) Gly-Ala	LRR20	0.968	0.181	0.108	73	YES	0.534 1.379	A, O, N
564	Asp-Asn (2)	LRR21					YES	1.015	A
567	Leu-Val (3) Leu-Phe	LRR21	0.971		0.160	36	YES	1.329 0.732	A, O, N
586	Glu-Ala Ala-Thr (2)	LRR22			0.172	22	YES	0.906 1.587	O
592	Asn-His Asn-Ser Asn-Lys His-Arg	LRR22	0.906					1.382 2.053 1.075 1.317	A, O
616	Leu-Phe (1-2)	LRRCT					YES	0.732	A, amb
628	Asp-Gly (2) Asp-Ala or Gly-Ala Ala-Val	LRRCT	0.924				YES	0.548 0.657 or 1.379 1.015	A, O, amb
634	Ileu-Val (2)	LRRCT					YES	2.415	N, amb
644	Ileu-Val (2)	Transm.					YES	2.415	O, N
650	Val-Leu	Transm.			0.153			1.329	O
680	Lys-Arg (2)	Intracel.					YES	1.583	O
690	Thr-Met (2)	Intracel.					YES	1.007	O, amb
847	Ser-Asn (3) Asn-Asp	TIR	0.916		0.130	74	YES	2.053 1.015	N, amb
854	Val-Ileu (2)	TIR					YES	2.415	O, N

For the codons identified as candidates for positive selection, posterior probabilities, p-values or Bayes factors are given (see below). Codons identified by more than one ML method are underlined.

[a] Amino acids between brackets fall within the 228 amino acid region identified by Andersen-Nissen et al. (2007) as important for flagellin recognition.

[b] Signal P= Signal Peptide, LLR = Leucine-rich repeat, NT= N-terminal, CT= C-terminal, Transm.=Transmembrane, Intracel.= Intracellular, TIR= Toll-IL-1 Receptor Domain

[c] Posterior probabilities of the BEB analysis

[d] P-values.

[e] Bayes factor

[f] U is an empirically derived index that measures the likelihood of a nonsynonymous fixation (Tang et al. 2004). * = indicates than more than 1 nucleotide change is needed for the amino acid change

[g] Clade in which amino acid substitution occurred (O=Old World Monkeys, N=New World Monkeys, A= Apes and Humans, amb=more than one equally parsimonious reconstruction.

Table 3. Levels of polymorphism and tests of neutrality in humans and chimpanzees.

		Coding						Non coding					
	n[a]	L[b]	S[c]	π (%)	θ (%)	Tajima's *D*	Fu and Li's *D**	L[b]	S[c]	π (%)	θ (%)	Tajima's *D*	Fu and Li's *D**
Humans													
Africans	66	2573	13	0.048	0.106	-1.560*	-0.124	13725	112	0.123	0.171	-0.984	-1.803*
Europeans	64	2574	12	0.057	0.099	-1.203	-1.387	13572	56	0.092	0.087	0.176	0.698
All	130	2574	19	0.054	0.136	-1.684*	-0.722	13647	122	0.109	0.164	-1.087	-2.110*
Chimpanzees	38	2574	4	0.036	0.037	-0.0588	-0.0235	1861	5	0.082	0.064	0.703	1.1216

[a] No. of chromosomes
[b] No. used sites (including missing data)
[c] No. polymorphic sites
* significant at the 0.05 level

Table 4. Polymorphisms and fixed differences for non-synonymous and synonymous sites.

Comparison	Site	Polymorphisms [a]	Fixed Differences [b]	P-value
Human and	Non-synonymous	15 (16)	23	0.08 (0.08)
chimp together	Synonymous	7	28	
Human alone	Non-synonymous	12 (13)	21	0.17 (0.10).
	Synonymous	6	27	
Chimp alone	Non-synonymous	3	21	0.61
	Synonymous	1	23	

[a] 13 replacement polymorphisms occur in humans. One of them occurs uniquely in the background of the haplotype containing the premature stop. The M-K test was computed including and excluding that replacement change (numbers of polymorphisms and p-values in parenthesis).
[b] All fixed differences are in comparison to the orangutan sequence.

Table 5. Polymorphisms and fixed differences for synonymous and regulatory sites.

Site	Polymorphisms	Fixed Differences [a]	P-value
Regulatory 1 [b]	7	18	0.49
Synonymous (coding)	6	9	
Regulatory 2 [c]	9	13	1.00
Synonymous (coding)	6	9	
Regulatory combined	16	31	0.76
Synonymous (coding)	6	9	

[a] Divergence with respect to the chimpanzee sequence
[b] 1 Kb upstream of the transcription start site of transcript ENST00000342210
[c] 1 Kb upstream of the transcription start site of transcript ENST00000366881

Table 6. Sites that show significant linkage disequilibrium with the premature stop mutation in humans.

Site 1	TLR5^{392STOP}	Distance (bp)	D'	P [a]	Age of TLR5^{392STOP} (years) [b]
9946	33909	23963	0.731	<0.0001(B)	27,237
10021	33909	23888	0.539	<0.001	28,723
11185	33909	22724	0.731	<0.0001(B)	53,893
11970	33909	21939	0.386	<0.0001	90,382
13373	33909	20536	0.544	<0.001	61,754

[a] B=significant after Bonferroni correction
[b] The age reported in the text is the average of the 5 sites.

FIGURE LEGENDS

Figure1. Lineage-specific dN/dS values of TLR5 in primates, A) for the entire gene and B) for the extracellular domain. Estimated dN/dS values from the branch-based model are shown above branches and the estimated number of non-synonymous and synonymous changes are shown below branches. Branches with dN/dS values greater than 1 are shown in red. Mating systems categorized as 'less promiscuous' (polygyny + monogamy) are indicated with a blue circle while 'more promiscuous' (promiscuous + dispersed) mating systems are indicated with a red circle. Arrows show the six unambiguous independent transitions between less and more promiscuous mating systems. For the Old World and New World monkey clades, "circled-pointed arrows" indicate additional transitions between low and high promiscuity according to alternative but equally parsimonious reconstructions.

Figure 2. Haplotype network showing the two divergent haplogroups shared between *P.t.verus* and *P.t.troglodytes*. Each circle represents a different haplotype and its size is proportional to its frequency in the sample. Mutations distinguishing haplotypes are shown as marks along the lines, while missing haplotypes are shown as black dots.

Figure 3. Distribution of $TLR5^{392STOP}$ around the world. The frequency of the allele is shown in red.

Figure 1

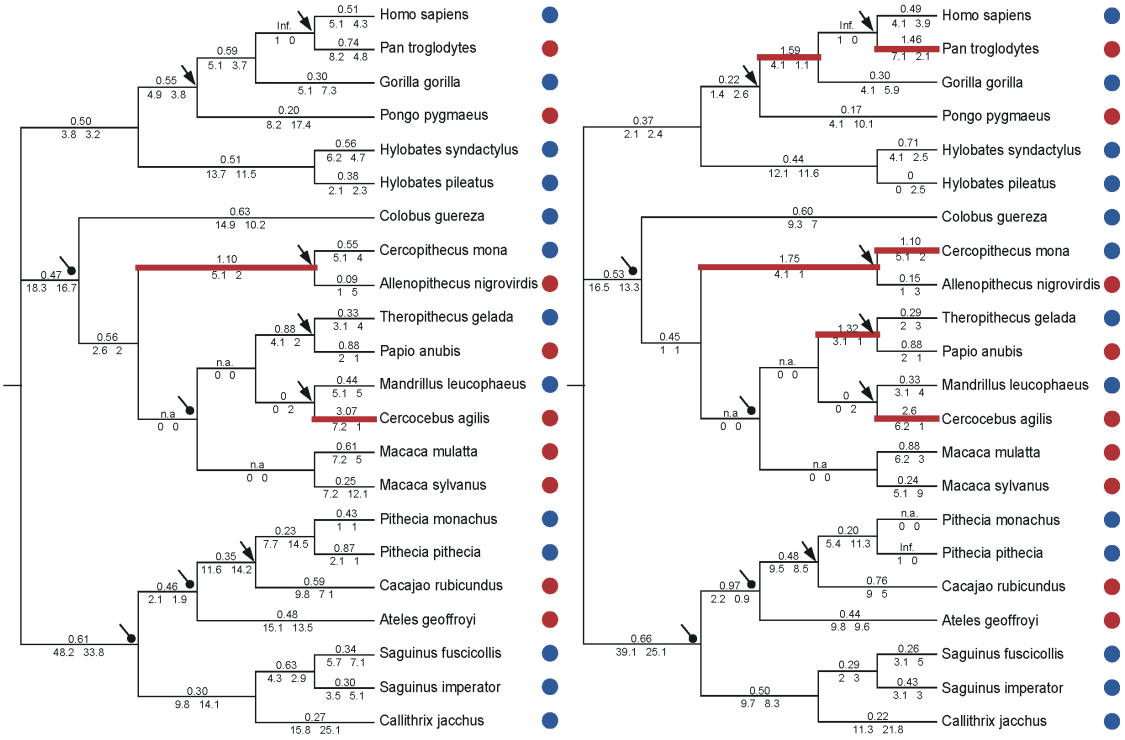

Homo sapiens
Pan troglodytes
Gorilla gorilla
Pongo pygmaeus
Hylobates syndactylus
Hylobates pileatus
Colobus guereza
Cercopithecus mona
Allenopithecus nigrovirdis
Theropithecus gelada
Papio anubis
Mandrillus leucophaeus
Cercocebus agilis
Macaca mulatta
Macaca sylvanus
Pithecia monachus
Pithecia pithecia
Cacajao rubicundus
Ateles geoffroyi
Saguinus fuscicollis
Saguinus imperator
Callithrix jacchus

Figure 2

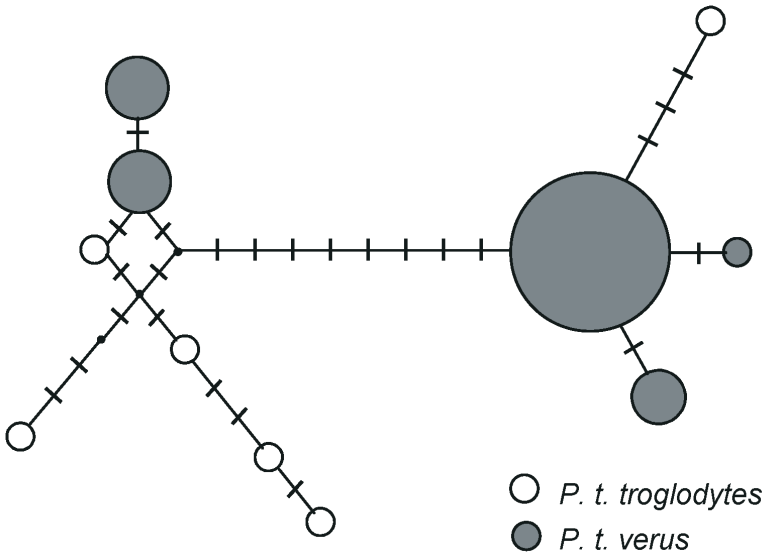


Figure 3

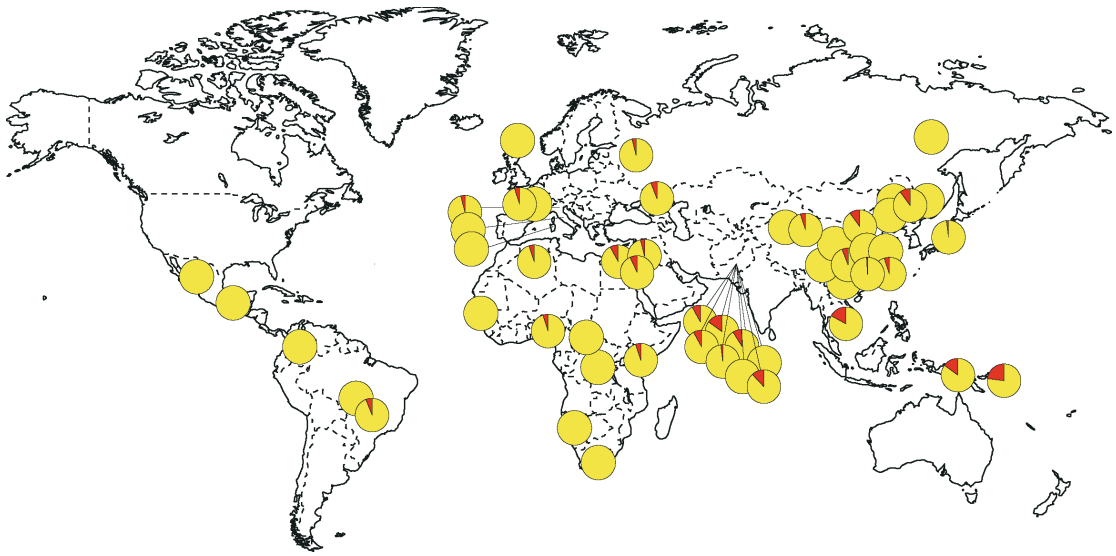


Table S1: Sources of the primate samples for the phylogenetic study.

Species	Origin of tissue or DNA sample
Colobus guereza	Coriell Cell Repositories
Allenopithecus nigrovirdis	Coriell Cell Repositories
Cercocebus agilis	Coriell Cell Repositories
Hylobates syndactylus	Coriell Cell Repositories
Ateles geoffroyi	Coriell Cell Repositories
Pithecia pithecia	Coriell Cell Repositories
Gorilla gorilla	Museum of Vertebrate Zoology, Berkeley
Saguinus imperator	Museum of Vertebrate Zoology, Berkeley
Saguinus fuscicollis	Museum of Vertebrate Zoology, Berkeley
Pithecia monachus	Museum of Vertebrate Zoology, Berkeley
Papio anubis	Southwest National Primate Research Center
Callithrix jacchus	Southwest National Primate Research Center
Pongo pygmaeus	Dr. Oliver Ryder
Hylobates pileatus	Gladys Porter Zoo
Cercopithecus mona	Dr. Brigitte Beer
Mandrillus leucophaeus	San Diego Zoo's Center for Reproduction of Endangered Species
Theropithecus gelada	San Diego Zoo's Center for Reproduction of Endangered Species
Macaca sylvanus	Toronto Zoo
Cacajao rubicundus	Los Angeles Zoo

Table S2: PCR conditions

Humans (Fragment 1) [a]	**Humans (Fragment 2)** [a] **/ Chimps** [b]	**Primates** [c]
15.75 µl water	26.5µl water	26.5µl water
2.5µl PCR Buffer 10X	5µl PCR Buffer 10X	5µl PCR Buffer 10X
2µl dNTPs 10µM	4µl dNTPs 10µM	4µl dNTPs 10µM
0.5µl ea. primer 10 mM	1µl ea. primer 10 mM	1µl ea. primer 10 mM
1µl MgSO4 50 mM	2µl MgSO4 50 mM	2µl MgSO4 50 mM
0.25µl HiFi Platinum Taq	0.5µl HiFi Platinum Taq	0.5µl HiFi Platinum Taq
2.5µl DNA (5ng/µl)	10µl DNA (5ng/µl)	10µl DNA (5ng/µl)

The following PCR cycling conditions were used for all PCRs: 40 cycles of 94°C 30sec, X°C 30sec (annealing temperatures provided in Table 1), and 68°C X min (adjusted according to size of PCR product, about 1min per 1kb).

[a] In humans, two fragments of TLR5 were sequenced. The fragments were PCR-amplified and sequenced using specific primers designed using the human genome sequence and/or published in the IID (www.innateimmunity.net). For the first fragment, we used nine sets of primers for the amplification: TLR05-090 and TLR05-091, TLR05-100 and TLR05-101, TLR05-110 and TLR05-111, TLR05-120 and TLR05-121, TLR05-132 and TLR05-133, TLR05-140 and TLR05-141, TLR05-150 and TLR05-191, TLR05-190 and TLR05-231, and TLR05-230 and TLR05-251. For the second fragment, TLR5F1 and TLR5R1 were used as amplification primers. PCR reactions were run in volumes of 25µl for the first fragment and in volumes of 50µl for the second fragment.

[b] In chimps, one fragment was PCR-amplified and sequenced using specific primers designed using the human genome sequence. TLR5F1 and TLR5R1 were used as amplification primers. PCR reactions were run in volumes of 50µl.

[c] In primates the coding sequence was PCR-amplified and sequenced using specific primers designed in conserved regions of the human-chimp-macaque-orang alignment. TLR5P5F and TLR5P4R were used as amplification primers. PCR reactions were run in volumes of 50µl.

Table S3 - Amplification and Sequencing Primers

Name	Direction	Location	Type	Sequence 5'-3'	Annealing Temp (°C)
Human Frag1[a]					
TLR05-090	F	5' flank	PCR/Seq	CCTTGGTATTTGTATTTCCTTGAA	52
TLR05-100	F	5' flank	PCR/Seq	CTGGCTGTTCAGGTAGAAGTTGT	63
TLR05-110	F	5' flank	PCR/Seq	ATTTCTGTCCTTGAACTGGGTTT	51
TLR05-120	F	5' flank	PCR/Seq	TTTCCAGTCTATGTCCAACTCG	51
TLR05-132	F	5' flank	PCR	TCCGTCCAGCACCGTGAACT	64
TLR05-134	F	5' flank	Seq	AACAAGCAGAGTAACCCTAG	
TLR05-140	F	5' flank	PCR/Seq	GATTACAGGTACCCACCACTACG	63
TLR05-150	F	5' flank	PCR/Seq	AGCACCTGTGAAACAATTAGAGC	54
TLR05-160	F	5' flank	Seq	AGGGATCTTTGACTCACTCACTG	
TLR05-170	F	5' flank	Seq	GGTAGGCTGAACCAGAGTGATAA	
TLR05-180	F	5' flank	Seq	GAAGAAGCCTGAAGCAGAGAAC	
TLR05-190	F	5' flank	PCR/Seq	GCCAAAGTCATGGTATTGCTAAA	53
TLR05-200	F	In 1	Seq	ATCATGTCACTGCACTCTAGCCT	
TLR05-210	F	In 1	Seq	CATGTCTCTCTGCTTTACCCATC	
TLR05-220	F	In 1	Seq	CACCAGTGAAATCATCTTCCTCT	
TLR05-230	F	In 2	PCR/Seq	GCCATTCTGTACCTAAACCATGT	52
TLR05-240	F	In 2	Seq	TCTCTGATGCTTCACTCTTCAGC	
TLR05-250	F	In 2	Seq	CTCATCCTTGTGCTTGAGTCTTT	
TLR05-091	R	5' flank	PCR/Seq	CATGCAACTTTGTGAATATGTGG	52
TLR05-101	R	5' flank	PCR/Seq	TATAAGAACAGTGCTCCTCCTGC	63
TLR05-111	R	5' flank	PCR/Seq	AGGATGTGAGTGACTTCGTCTGT	51
TLR05-121	R	5' flank	PCR/Seq	AGAACAAGGGTGAACTGAGTCAA	51
TLR05-133	R	5' flank	PCR	GATAATTAGGGTCATACGCACAG	64
TLR05-135	R	5' flank	Seq	GTCATACGCACAGGCATG	
TLR05-141	R	5' flank	PCR/Seq	TGCCTTTCATTATTCTGAGCTTC	63
TLR05-151	R	5' flank	PCR/Seq	TGCATTACACTCAGACTCCTCAA	54
TLR05-161	R	5' flank	Seq	ATGTAAGATGTGACTTTGCTCCC	
TLR05-171	R	5' flank	Seq	CATGAACTACTGTACCCAGCCTC	
TLR05-181	R	5' flank	Seq	ATGCATTTGCTATTATTGCTGCT	
TLR05-191	R	In 1	PCR/Seq	CATGGTCTTCTGAGTTCAAATCC	54
TLR05-201	R	In 1	Seq	CCCTCTCCTTGTATTTCTGCTTT	
TLR05-203	R	In 1	Seq	GCATAAGAGATCTGAAATTGTGAC	
TLR05-211	R	In 1	Seq	GAGAGGGATGTGTAGTCTGTGCT	
TLR05-221	R	In 2	Seq	CGCAACTACACCTTACCAGAAAC	
TLR05-231	R	In 2	PCR/Seq	TGTTAGCGGTGAGAACCTAAGAG	53
TLR05-241	R	In 2	Seq	TGGAAGAATTGCAAACTTTCTGT	
TLR05-251	R	In 3	PCR/Seq	CTGGTATTCTGGGTACATTTCCA	52
Humans Frag2 / Chimps					
TLR5F1	F	In 3	PCR/Seq	TCCTAACGATTATTAGATGCCTGAG	52
TLR5F2	F	Ex 4	Seq	TGCTCTCATCATGGTGGTGG	
TLR5F3	F	Ex 4	Seq	TTTCCCTCTTCTCTCTTTCC	
TLR5F4	F	Ex 4	PCR/Seq	CTTCAGAGAATCCCAGCTTA	52
TLR5F5	F	Ex 4	Seq	TGTCTTCTCCCTGAACTCAC	
TLR5F6	F	Ex 4	Seq	TACCTTCATCCTTCATTTGG	
TLR5F7	F	In 3	Seq	TGTGTTTTCATTCTCCCTTC	
TLR5F8	F	In 3	Seq	TCACATCTGTAATCCCAACA	
TLR5F9	F	In 3	Seq	TAAGGTCGGATAAATGGAGA	
TLR5F10	F	In 3	Seq	TCTTCCTTTACCTTCCAACA	
TLR5R1	R	3' flank	PCR/Seq	ACAGAACGGTATTATTGGATCTGAA	52
TLR5R2	R	In 3	Seq	TATCCGACCTTACTCCACAC	
TLR5R3	R	In 3	Seq	TGGCCTATTCTTGCTCTCTA	
TLR5R4	R	In 3/Ex 4	Seq	CCATGATCCTATGGAGAAGA	
TLR5R5	R	Ex 4	PCR/Seq	CCAAATGAAGGATGAAGGTA	52
TLR5R6	R	Ex 4	Seq	GTGAGTTCAGGGAGAAGACA	
TLR5R7	R	Ex 4	Seq	TAAGCTGGGATTCTCTGAAG	
TLR5R8	R	Ex 4	Seq	TCCTCTTCATCACAACCTTC	
TLR5R9	R	Ex 4	Seq	CCTCTGATGGATTGATGTTT	

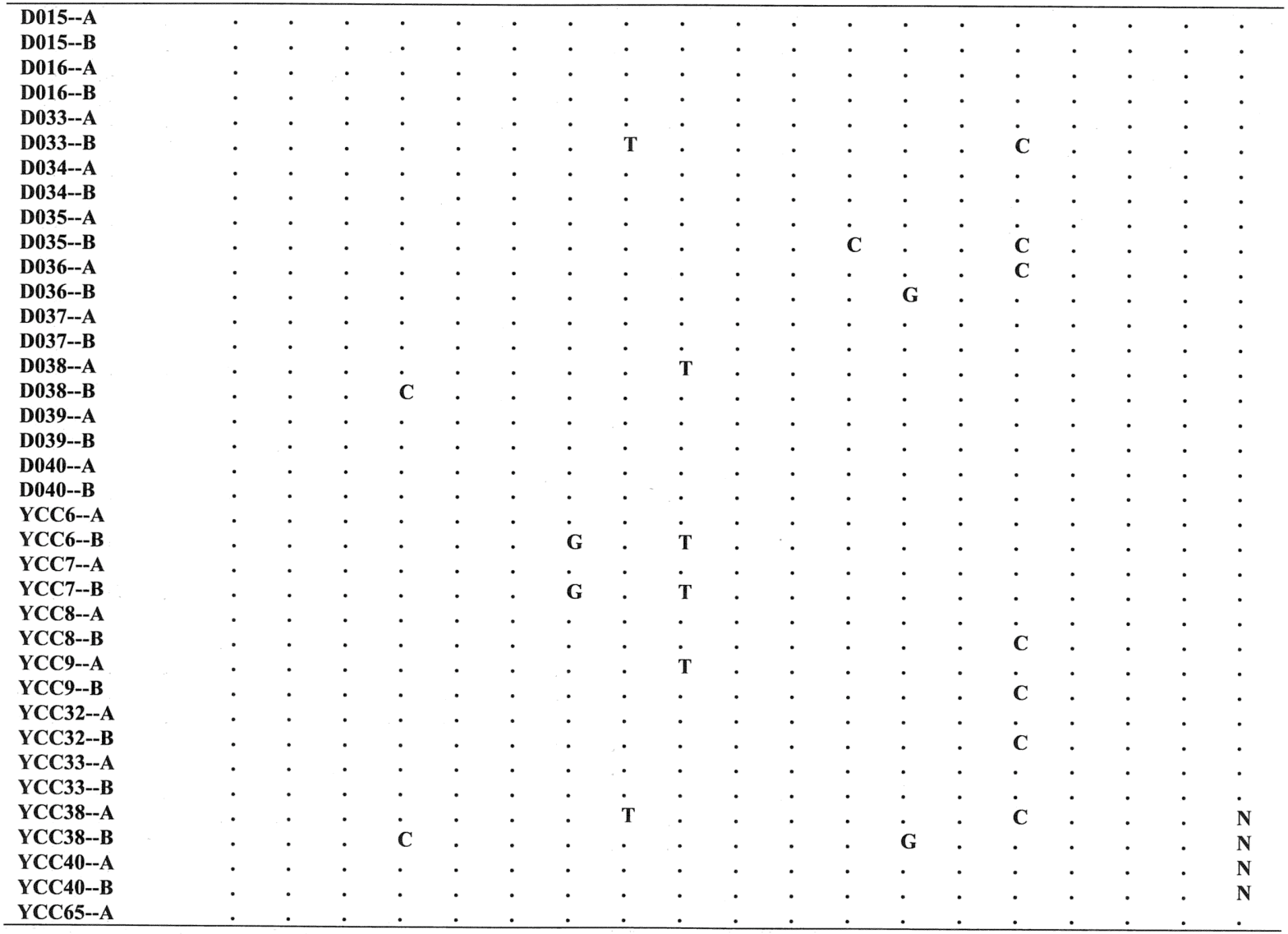
Primates					
TLR5P1F	F	In 3	Seq	TCATTCTCCYTTCTWCTCCATA	
TLR5P3F	F	In 3	Seq	GGAGACCACCTDGACCTTC	
TLR5P5F	F	In 3	PCR	GCCKGKTTYTCATTCTCC	53
TLR5Pi1F	F	Ex 4	Seq	CCTGACCAGARCACATTC	
TLR5Pi2F	F	Ex 4	Seq	GAAACTTYAGCAATGCCATCA	
TLR5Pi3F	F	Ex 4	Seq	GAATGTGMACTTAGCACTTT	
TLR5Pi4F	F	Ex 4	Seq	CTTAATCAYACCAATGTCACTA	
TLR5P2R	R	3' flank	PCR	TGGTGYAAATACAAAGTGAAGA	53
TLR5P4R	R	Ex 4	Seq	GAATGTTAYTGTCTTTCTTCTTTT	
TLR5P6R	R	Ex 4	Seq	TGAGACARAACATKGTGTTGATA	
TLR5Pi1R	R	Ex 4	Seq	TATAGTGACATTGGTRTGATTAAG	
TLR5Pi2R	R	Ex 4	Seq	AAAGTGCTAAGTKCACATTC	
TLR5Pi3R	R	Ex 4	Seq	AATGTGYTCTGGTCAGG	
TLR5Pi4R	R	Ex 4	Seq	TGATGGCATTGCTRAAGTT	

[a] Primer sequences taken from the Innate Immunity Database (www.innateimmunity.net)

F=Forward, R=Reverse, 5' flank=5' flanking region, 3' flank=3' flanking region, In1=Intron 1, In2=Intron 2, In3=Intron 3, Ex4=Exon 4, PCR=Primer used for anplification, Seq=Primer used for sequencing.

Table S4. Table of polymorphism of the coding region in humans

Position	183	245	334	428	541	744	939	1174	1332	1368	1459	1734	1775	1793	1846	1930	2081	2523	2537
Africans + African-Americans																			
consensus	C	C	C	A	C	T	T	C	C	T	C	T	A	A	T	A	A	A	A
D001--A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
D001--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D002--A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
D002--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D003--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D003--B	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
D004--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
D004--B	.	.	.	.	.	.	.	.	.	.	.	.	.	C	C	.	.	.	.
D005--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D005--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
D006--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D006--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D007--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D007--B	.	.	G	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
D008--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D008--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D009--A	.	.	.	.	.	.	.	.	T	C	.	.	.	.	.	.	.	.	.
D009--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D010--A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
D010--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D011--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D011--B	.	.	.	.	.	.	.	.	.	.	A	.	.	.	C	T	.	.	G
D012--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D012--B	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.
D013--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D013--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
D014--A	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.
D014--B	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.


D015--A
D015--B
D016--A
D016--B
D033--A
D033--B
D034--A
D034--B
D035--A
D035--B
D036--A
D036--B
D037--A
D037--B
D038--A
D038--B
D039--A
D039--B
D040--A
D040--B
YCC6--A
YCC6--B
YCC7--A
YCC7--B
YCC8--A
YCC8--B
YCC9--A
YCC9--B
YCC32--A
YCC32--B
YCC33--A
YCC33--B
YCC38--A
YCC38--B
YCC40--A
YCC40--B
YCC65--A

YCC65--B	.	.	.	.	.	.	.	.	.	.	A	.	.	.	C	T	.	.	G
Europeans + European Americans																				
consensus	C	C	C	A	C	T	T	C	C	T	C	T	A	A	T	A	A	A	A
E001--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E001--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
E002--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E002--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E003--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E003--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E004--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E004--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E005--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E005--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E006--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E006--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E007--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E007--B	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.
E008--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
E008--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E009--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E009--B	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E010--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E010--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E011--A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E011--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E012--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
E012--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E013--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E013--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
E014--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E014--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E015--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E015--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E016--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E016--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	G	.	.
E017--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

E017--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E018--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E018--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E019--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E019--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E020--A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
E020--B	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.
E021--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E021--B	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
E022--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
E022--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
E023--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
E023--B	.	.	.	.	.	.	.	T	.	.	.	.	.	.	C	.	.	.	.
YCC26--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YCC26--B	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YCC56--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.
YCC56--B	.	.	.	.	A	C	.	T	.	.	.	.	.	.	C	.	.	.	.
YCC59--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YCC59--B	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YCC61--A	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
YCC61--B	.	.	.	.	A	.	.	T	.	.	.	.	.	.	C	.	.	.	.
YCC62--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
YCC62--B	.	.	.	.	.	.	.	.	.	.	A	.	.	.	C	T	.	.	G
YCC64--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YCC64--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YCC71--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
YCC71--B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
YCC72--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
YCC72--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.
YCC79--A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C	.	.	.	.
YCC79--B	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.	.	.	.	.

Table S5: Table of polymorphism of the non-coding and coding regions in *P.t. verus* and *P.t. troglodytes*

Position	-1458	-1181	-1169	-1102	-1099	-767	-765	-626	-581	-267	65	85	295	888	977	175	1877	2372
	Non-coding												*coding*					
P.t. verus																		
'CH09--B'	C	G	T	T	A	G	A	T	T	C	C	A	A	A	A	G	C	T
'CH09--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.
'CH29--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH87--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH88--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH94--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH94--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH95--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH95--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH96--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH96--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH97--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH98--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH98--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH102--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH103--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH103--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH106--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH106--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH107--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH107--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH108--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH108--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH110--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH112--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH123--A'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH123--B'	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH85--B'	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

'CH109--B'	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH112--B'	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
'CH29--B'	.	A	C	.	G	.	.	C	.	.	T	C	G	.	.	.	.	.
'CH88--B'	.	A	C	.	G	.	.	C	.	.	T	C	G	.	.	.	.	.
'CH102--B'	.	A	C	.	G	.	.	C	.	.	T	C	G	.	.	.	.	.
'CH110--B'	.	A	C	.	G	.	.	C	.	.	T	C	G	.	.	.	.	.
'CH85--A'	.	A	C	.	G	.	.	C	.	.	T	C	.	.	.	.	.	.
'CH87--B'	.	A	C	.	G	.	.	C	.	.	T	C	.	.	.	.	.	.
'CH97--B'	.	A	C	.	G	.	.	C	.	.	T	C	.	.	.	.	.	.
'CH109--A'	.	A	C	.	G	.	.	C	.	.	T	C	.	.	.	.	.	.
P.t. troglodytes																		
'CH79tro--A'	.	.	.	.	.	.	.	.	.	T	.	.	.	.	G	A	.	.
'CH13tro--A'	.	.	C	.	G	.	G	C	C	.	.	C	.	.	.	.	.	.
'CH79tro--B'	.	.	C	.	G	.	G	C	C	.	.	C	.	.	.	.	.	C
'CH99tro--A'	.	.	C	.	G	A	.	C	C	.	T	C	.	C	.	.	.	C
'CH13tro--B'	.	.	C	C	G	.	.	C	.	.	T	C	.	.	.	.	.	C
'CH99tro--B'	.	A	C	.	G	.	.	C	.	.	T	C	.	.	.	.	.	C

Table S6. Estimated allele frequencies TLR5^{392STOP} in the HGDP.

Population	Geographic origin	Sample size (#alleles)	# ind. +/+	# ind. +/S	# ind. S/S	Freq. S	Exp. No. heterozygotes
Biaka Pygmies	Cen. Afr. Republic	46	23	0	0	0.000	0
Mbuti Pygmies	Dem. Rep. Congo	26	13	0	0	0.000	0
Mandenka	Senegal	44	22	0	0	0.000	0
Yoruba	Nigeria	44	20	2	0	0.045	2
Bantu N.E.	Kenya	22	10	1	0	0.045	1
San	Namidia	12	6	0	0	0.000	0
Bantu S.E., S.W.	South Africa	16	8	0	0	0.000	0
Mozabite	Algeria (Mzab)	58	26	3	0	0.052	3
Bedouin	Israel (Negev)	92	40	5	1	0.076	6
Druze	Israel (Carmel)	84	36	6	0	0.071	6
Palestinian	Israel (Central)	92	42	4	0	0.043	4
Brahui	Pakistan	50	21	4	0	0.080	4
Balochi	Pakistan	48	17	7	0	0.146	6
Hazara	Pakistan	44	18	4	0	0.091	4
Makrani	Pakistan	50	25	0	0	0.000	0
Sindhi	Pakistan	48	20	4	0	0.083	4
Pathan	Pakistan	48	23	1	0	0.021	1
Kalash	Pakistan	46	23	0	0	0.000	0
Burusho	Pakistan	50	20	4	1	0.120	5
Han	China	86	42	1	0	0.012	1
Tujia	China	20	9	1	0	0.050	1
Yizu	China	20	10	0	0	0.000	0
Miaozu	China	20	8	2	0	0.100	2
Oroqen	China	18	8	1	0	0.056	1
Daur	China	20	10	0	0	0.000	0
Mongola	China	20	10	0	0	0.000	0
Hezhen	China	18	9	0	0	0.000	0
Xibo	China	18	9	0	0	0.000	0
Uygur	China	20	9	1	0	0.050	1
Dai	China	20	10	0	0	0.000	0
Lahu	China	16	8	0	0	0.000	0
She	China	20	10	0	0	0.000	0
Naxi	China	18	7	2	0	0.111	2
Tu	China	20	10	0	0	0.000	0
Yakut	Siberia	48	24	0	0	0.000	0
Japanese	Japan	58	28	1	0	0.017	1
Cambodian	Cambodia	18	6	3	0	0.167	3
Papuan	New Guinea	34	13	3	1	0.147	4
NAN Melanesian	Bougainville	22	6	5	0	0.227	4
French	France	56	25	3	0	0.054	3
French Basque	France	48	24	0	0	0.000	0
Sardinian	Italy	56	28	0	0	0.000	0
North Italian	Italy (Bergamo)	26	12	1	0	0.038	1
Tuscan	Italy	16	8	0	0	0.000	0
Orcadian	Orkney Islands	30	15	0	0	0.000	0
Adygei	Russia Caucasus	34	15	2	0	0.059	2
Russian	Russia	50	23	2	0	0.040	2

Pima	Mexico	28	14	0	0	0.000	0
Maya	Mexico	42	21	0	0	0.000	0
Colombian	Colombia	14	7	0	0	0.000	0
Karitiana	Brazil	28	14	0	0	0.000	0
Surui	Brazil	18	8	1	0	0.056	1
	TOTAL	1900	873	74	3	0.042	77

APPENDIX C: MOLECULAR EVOLUTION OF INNATE IMMUNITY GENES AT DIFFERENT TIMESCALES: ADAPTATION AND CONSTRAINT AT TOLL-LIKE RECEPTORS IN PRIMATES.

ABSTRACT

Frequent positive selection is a hallmark of genes involved in the adaptive immune system of vertebrates, but this pattern has not been well studied for genes underlying vertebrate innate immunity. 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 co-evolutionary dynamics with their microbial counterparts. However, they recognize conserved molecular motifs, and this might constrain their evolution. Here, we provide a general picture of the evolution of all human TLRs in the framework of these competing ideas. We studied rates of protein evolution among 8-11 primate species. We also analyzed patterns of polymorphism in humans and in chimpanzees. These approaches provide a picture of TLR evolution at different 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 non-viral 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-tern co-evolution.

INTRODUCTION

Toll-like receptors (TLRs) recognize and bind conserved molecular patterns in pathogens both to initiate an innate immune response and to prime the adaptive immune system (Akira and Takeda 2004). The innate immune receptors, as exemplified by the TLR family, have evolved to perform several complex tasks simultaneously. They must discriminate self from foreign (by targeting molecular patterns absent in the host), achieve some degree of specificity (by targeting molecular patterns shared by classes of pathogens) and prevent the evolution of mechanisms of pathogen evasion (by targeting components that are essential for microbial fitness).

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; Lazarus et al. 2004; Hawn et al. 2005; 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).

An interesting aspect of TLRs is their suggested functional redundancy (Ku et al. 2005). TLR deficient mice often display lower cytokine production and reduced survival following microbial challenges, which clearly demonstrate their importance as microbial sensors (reviewed in Carpenter and O'Neill 2007). However, the accumulation of loss of

function mutations in human populations (Barreiro et al. 2009; Wlasiuk et al. 2009) suggests some degree of redundancy. Functional redundancy would exist if the same microorganism activates different TLRs by means of different molecular components (Akira, Uematsu, and Takeuchi 2006) or if different receptors target the same microbial molecule (e.g Miao et al. 2007).

Although there is some overlap in the classes of ligands they recognize, TLRs expressed within endosomal compartments (TLR3, TLR7, TLR8, TLR9) target predominantly viral components such as single and double-stranded RNA and CpG DNA, while TLRs expressed in the cell membrane (TLR1, TLR2, TLR4, TLR5, TLR6, TLR10) target predominantly bacterial (but also fungal and parasite) components such as lipopolysaccharide, peptidoglycan and flagellin (Akira, Uematsu, and Takeuchi 2006; Carpenter and O'Neill 2007). We will refer to these two subclasses as viral and non-viral TLRs. Viral and non-viral 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), self nucleic acids retain some capacity to induce an immune response. Viral TLRs face the challenge of remaining fully functional while avoiding autoimmunity, and thus, we hypothesize that they are under stronger functional constraint than non-viral TLRs.

Despite several studies on the evolution of TLRs in humans and non-human primates, a clear picture of the evolution of this family of innate immune receptors has not emerged. Previous studies have generally focused only 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), using SNP data in multiple human populations and sequence data in Africans and Europeans, concluded that balancing selection is the best explanation for the pattern of sequence variation in a series of human innate immunity genes that include five TLRs. On the other hand Mukherjee et al. (2009) found no evidence of selection in a human population from India, and they argued that purifying selection is the predominant force in TLR evolution, in agreement with an earlier study of TLR4 (Smirnova et al. 2001). The study of Mukherjee et al. (2009) included six TLRs, two of which were also included in Ferrer-Admetlla et al. (2008). Recently, 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) did not 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 at 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 time-scale can only be studied from interspecific comparisons.

Our goal was to provide a comprehensive picture of TLRs 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 non-coding regions of all 10 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 'non-viral' 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 from patterns of nucleotide variation. In spite of similar levels of variation in both species, 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 to chimpanzees. Viral TLRs were generally more constrained than non-viral 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 (www.innateimmunity.net).

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, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus, Hylobates lar, Hylobates pileatus, Cercocebus torquatus, Macaca mulatta, Saguinus oedipus, Saguinus 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 1.

DNA sequencing: For TLR1-TLR4 and TLR6-TLR10, the coding region and a non-coding fragment of comparable length (total ~4-5 kb) were 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 2. 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, some intronic sequence, 5' and 3' UTRs, 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 selection at all the TLRs during primate evolution, we compared the rate of nonsysnonymous substitution (dN) to the rate of synonymous substitution (dS) in a maximum likelihood (ML) framework. A ratio of $dN/dS>1$ is interpreted as strong evidence of positive selection while 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 ver 4) (Yang 1997; Yang 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. 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, Wong, and Nielsen 2005).

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 model (SLAC) is based on the reconstruction of the ancestral sequences and the counts of synonymous and non-synonymous changes at each codon position in a phylogeny. The Fixed Effect Likelihood model (FEL) estimates the ratio of non-synonymous 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 3D structures using PyMOL (Delano Scientific, San Carlos, CA).

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

Population genetic analyses: For both the human and chimpanzee datasets, we estimated the level of polymorphism as measured by the 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 ver 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 the 323 genes in the Seattle SNPs database, sampled in the same individuals.

The extent of linkage disequilibrium (LD) associated with particular variants in the human Hap Map data was evaluated with the iHS statistic (Voight et al. 2006). Using Haplotter (http://pritch.bsd.uchicago.edu/data), we screened windows of 50 SNPs centered on each gene, looking for an accumulation of SNPs with |iHS| >2, as in Voight et al. (2006). To assess LD between genes in the TLR6-TLR1-TLR10 cluster we calculated D' (Lewontin 1964) between all pairs of SNPs using DnaSP v5 (Librado and Rozas 2009).

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 to 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, Kreitman, and Aguade 1987) with the software HKA (http://lifesci.rutgers.edu/~heylab/). In the case of humans, this control set consisted of 10 concatenated putative neutral loci with average levels of polymorphism and divergence (IL20, CSF2, FSBP, IL22, CSF3, MMP9, IFGN, CRP, PLAU, IL6), previously sequenced for the same population samples (Akey et al. 2004). In the case of chimpanzees, the control set consisted of 26 noncoding segments sequenced in a population sample of chimpanzees of roughly the same size as ours

(Fischer et al. 2006). For both species, the concatenated sets of viral and non-viral TLRs were compared against each other. The use of the macaque sequence as the interspecific comparison in the M-K test provided more power due to increased divergence. In the HKA test, however, we used human-chimpanzee divergence because the lower divergence resulted in more reliable alignments over long regions on non-coding 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 determine the effect of 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 Webserver (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 a functional effect) or 'probably damaging' (strong evidence of a functional effect). We subsequently combined polymorphisms predicted by Polyphen as 'possibly' or 'probably' damaging in one class as 'damaging'. We recognize, however, that some (presumably a small fraction) of the amino acid changes predicted by Polyphen as 'damaging' might 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.

We mapped the sites with dN/dS<1 across primates from the previous analysis using SLAC, REL and FEL onto these domains. Then, the observed distribution was compared with the expected distribution obtained by multiplying the total number of sites by the relative length of each domain.

RESULTS

Inference of positive selection from substitution patterns

Using maximum likelihood approaches we addressed whether recurrent positive selection has been common in the TLR family. First, we compared nested models with and without positive selection using likelihood ratio tests, and found that for six out of the ten genes (TLR1, TLR4, TLR6, TLR7, TLR8, TLR9), a model that includes sites with dN/dS>1 fits the data significantly better that a neutral model (Table 1). This group of six genes contains an equal number of viral and non-viral 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 dataset 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 datasets (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 3D 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. SNPs at some of these sites are also associated with various disease phenotypes (Table 1).

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 (Figure 2). This analysis only included sites that were implicated in positive selection in two or more methods (Table 1). We compared the observed and expected counts for each clade. The expected values were generated 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. At 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 (χ^2 p=0.0095), and more specifically in the *C. torquatus* branch, where 5 of the 31 non-ambiguous changes fall (Figure 3). 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 3 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.

Patterns and levels of variation within species

A summary of the polymorphism data for the 10 TLRs in humans and chimpanzees (for coding and non-coding regions) is presented in Table 3. In chimpanzees, the nucleotide heterozygosity per site (π) for the coding and non-coding regions together ranged between 0.03-0.07%, 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%. Tables of polymorphism for the chimpanzee TLRs are presented in Supplementary tables 3-12. In humans, the polymorphism levels in the combined coding

and non-coding regions (Africans 0.03%-0.23%, Europeans 0.03%-0.12) were unremarkable and similar to genome-wide patterns (Akey et al. 2004).

Table 3 shows several summary statistics commonly used to assess departures from a neutral model of evolution. These statistics capture different aspects of the allele frequency spectrum and have different power to detect selection or other alternative hypotheses to the neutral equilibrium model (Simonsen, Churchill, and Aquadro 1995; Fay and Wu 2000). Tajima's D compares the number of polymorphisms with the mean pairwise difference between sequences (Tajima 1989; Fu and Li 1993). 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, three genes show significant deviations in one of these four statistics. TLR6 is the most striking case, with a significant excess of low and high frequency derived variants in coding and non-coding regions, a pattern expected during or after a selective sweep. However, examination of the table of polymorphism (Supplementary Tables 8) reveals that the excess of rare variants is due to the presence of a unique divergent haplotype that carries 3 of the 6 singletons. While 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 invoked previously (e.g. Won and Hey 2005), and it seems a plausible explanation for the observed pattern.

In humans, the population trends are in overall agreement with the accepted demographic history of Africans and Europeans: a population expansion in Africans that resulted in greater numbers of rare polymorphisms and is reflected in more negative values of Tajima's D, and a bottleneck in Europeans that resulted in greater number of intermediate frequency polymorphisms and is reflected in less negative values of Tajima's D. 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 (Table 3) in the human dataset, with the exception of TLR10, most of the genes do not seem remarkable in the context of the genome-wide distributions of these statistics (Figure 4). TLR10 shows a departure from neutrality in Europeans according to all four statistics, and in the opposite direction of the one expected by the known demographic perturbation. Similar to the TLR6 case in chimpanzees, the excess of low and high frequency derived mutations seems to be caused by a divergent haplotype present only in one copy in Europeans but relatively frequent in Africans (Supplementary Table 13), suggesting again that migration might be a more plausible explanation than selection.

Another (not mutually exclusive) possible reason for the excess of low frequency variants that is observed in several of the genes, is the segregation of slightly deleterious polymorphisms, an idea explored in more detail below.

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 to the empirical distribution of F_{ST} for all the genes in the Seattle SNPs database (Table 4). 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).

Another signature of recent positive selection is the presence of an extended haplotype at relatively high frequency, associated with the selected allele. We tested for long-range LD in the Phase II Hapmap data using the integrated haplotype score (iHs) (Voight et al. 2006). iHs is based on the ratio of the integrated haplotype homozygosities (area under the curve of a Extended Haplotype Homozygosity-EEH 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.

TLR1, TLR6 and TLR10 belong to a gene cluster on chromosome 4 that spans ~54 kb. We investigated the patterns of LD between the three genes and found that many sites display significant LD as measured by D'. This degree of LD is not unexpected, since haplotype blocks in humans often extend for tens of kilobases (Gabriel et al. 2002). The LD between replacement polymorphisms is generally due to rare haplotypes. Table 5 shows a few interesting exceptions. Two pairs of SNPs in TLR1 (S248N, I602S) and

TLR10 (H241N, I369L) that show moderate LD are at intermediate frequencies. Notably, site 248 at TLR1 is one of the sites inferred to be under selection in primates by REL (Table 1).

Analyses of polymorphism and divergence

The ratio of replacement to silent polymorphism within humans or chimpanzees was compared to 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 6). Several genes however, show a slight excess of replacement polymorphisms with respect to fixations. This deviation can be described by the Neutrality Index (N.I), a ratio of the replacement to silent ratios within and between species (Rand and Kann 1996). N.I. values above one indicate an excess of replacement changes within species, while values between 0 and 1 indicate an excess of replacement fixations between species. When combined, non-viral TLRs present N.I. values above one, while viral TLRs have N.I. close to 1 (Table 6). 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).

We also used the HKA test to assess whether individual TLRs have been under positive selection. Two human TLRs differed significantly from the control set of genes, with an excess of polymorphism relative to divergence. TLR1was significant (HKA χ^2=4.29 p=0.04) and TLR10 was marginally significant (HKA χ^2= 2.76 p=0.09). No significant deviations were observed for the chimpanzee TLRs (Table 7). Viral and non-viral 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). There is a growing interest in this type of computational approach, because it provides a means to infer function when large-scale biochemical characterization of SNPs is not possible. In human genetics, there is particular interest in predicting deleterious alleles, since rare SNPs might contribute to disease.

In humans, we found a total of 31 damaging and 31 benign polymorphisms, while in chimpanzees we found 6 damaging and 19 benign polymorphisms (Table 8). These ratios are significantly different from each other (Table 9). Interestingly, the ratio of damaging/benign SNPs of human TLRs is significantly different from that in the human genome, in which the number of damaging SNPs is roughly one half of the number of benign SNPs. This excess of damaging SNPs in human TLRs is driven by the non-viral TLRs, since the viral TLRs do not deviate from the human genome trend (Table 9).

Negative selection at viral and non-viral TLRs

We evaluated the levels of functional constraint among TLRs in two ways. First, to get a sense of the overall rates of evolution of the different TLRs, we estimated the global dN/dS ratio for each gene over the primate phylogeny, as well as for the human lineage and the chimpanzee lineage (Table 10). In each case, non-viral 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 non-viral TLRs across the primate phylogeny (t-test $p=0.007$). This indicates that viral TLRs are under stronger purifying selection than non-viral 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 non-viral TLRs. On the other hand, the signal peptide and transmembrane domains show a higher dN/dS than the other domains.

Second, we mapped the codons with dN/dS values <1 onto the predicted protein domains and compared the observed distributions with expectations based on domain length (Table 11). Although not significant, the contrast between the observed and expected numbers revealed the same relative order of functional constraint among domains that was found with the global dN/dS values. For both viral and non-viral TLRs the cytoplasmic domain showed more negatively selected sites than other domains, indicative of stronger purifying selection, followed by the leucine-rich repeat domain, while the signal peptide and transmembrane domains were the least constrained.

DISCUSSION

Here, 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. We found a clear signature of positive selection in the rates of substitution across primates in most TLRs. However, within species, the 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 the 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 non-viral TLRs. These and other results are discussed below.

Recurrent positive selection is common in primate TLRs

Our analyses provide strong evidence that several TLRs have been subject to positive selection during primate 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, Good, and Nachman (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 10), in relation to the mean dN/dS of 0.25 for the human-chimp-macaque trio (Gibbs et

al. 2007). Finally, several of the putatively selected sites fall in regions important for function, based on structural information, and a few of them are linked to clinical phenotypes in humans. The reduced number of taxa, the stringent significance thresholds we used, and the fact that the codon-based approaches only detect selection acting on the same sites repeatedly, make these conclusions conservative.

An open question is whether the innate immune systems of vertebrates and invertebrates are under similar selective regimes. A relatively recent realization is that, far from being independent, the innate and adaptive immune responses of vertebrates act in a highly coordinated manner (Palm and Medzhitov 2009). This led us to speculate that the acquisition of adaptive immunity might have altered the levels or patterns of functional constraint of 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 TLRs have also been subject to positive selection in vertebrates.

The common theme we found among the six positively selected genes is of strong functional constraint along most of the protein, with a small proportion of sites under positive selection. Our results challenge the basic paradigm of TLR conservation and evolution, showing that these genes do have the potential to evolve by positive selection in response to pathogen pressure in spite of their overall 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. TLR4 was the gene with the highest proportion of sites inferred to be under selection. Twenty-four codons (Table 1) were concordant between at least two ML methods and thus constitute robust candidates for positive selection. Moreover, some of the non-ambiguous amino acid changes at these sites are radical in terms of their physicochemical properties (size, polarity, charge) (Figure 3) strengthening the case for positive selection. In association with the myeloid differentiation factor 2 (MD2), TLR4 responds to lipopolysaccharide (LPS) from gram-negative bacteria, but also targets components of yeast, *Trypanosoma* and even viruses (Kumar, Kawai, and Akira 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 reside in a region that participates in hydrophobic and electrostatic interactions and hydrogen bonds between TLR4, MD2 and LPS. Worth noting is the apparent clustering of inferred positively selected sites on two surfaces of the TLR4 ectodomain. Several of the sites (295, 297, 298, 299, 300 360) physically converge in an area important for interaction between TLR4 and LPS (Figure 1) and of these, many involve amino acid changes that affect polarity or charge (Figure 3). Site 296, identified by REL but not 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). Residue 299, identified with high confidence by CODEML and REL, is polymorphic in humans (D299G) and is responsible for

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). It has also been suggested that the otherwise negative effect of the D299G allele might be compensated by the benefit of protection against malaria in Africa (Ferwerda et al. 2007).

Similarly to what was reported by Nakajima et al. (2008), we observed a concentration of positively selected sites at TLR4 in *Catarrhini*, both when we looked at variation of dN/dS along lineages (the only four branches with dN/dS>1 fall in that clade) and among sites (the observed number of non ambiguous amino acid changes at these sites is higher than expected given the available time along these branches). Furthermore, one of the multiple amino acid changes at site 299 is located in the basal Old World primate branch. Similarly, although not as striking as the TLR4 case, the branches with higher dN/dS (above 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, Emerman, and Malik 2004; Sawyer et al. 2005; Elde et al. 2009), suggesting that these radiations might have been associated with major changes in pathogen abundance, diversity, or both.

TLR1, which interacts with TLR2 for the recognition of triacyl lipopeptides from gram-negative bacteria (Takeuchi et al. 2002) 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-binding site of the extracellular domain, although sites 308 and 321 are also in close physical proximity to the ligand binding site in the 3D

structure of the dimer (Jin et al. 2007) (Figure 1). In the same region, two human nonsynonymous polymorphisms (H305L and P315L) exhibit reduced activity in vitro (Omueti et al. 2007), further reinforcing the idea that this region is critical for function. Site 248, although only identified by one ML method, is also polymorphic in humans (S248N) and has been linked to a weak impairment in response to bacteria *in vitro* (Omueti et al. 2007, but see Hawn et al. 2007; Barreiro et al. 2009), 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. This makes any hypothesis about specific selective forces speculative.

Importantly, our results help to reconcile previous seemingly discordant results. In agreement with Ortiz et al. (2008) we found evidence of selection at TLR1 but we also found strong support for selection at TLR4 as reported by Nakajima et al. (2008). The general disagreement between these studies is probably due to the fact that the former used very small number of species, which severely reduces the power to detect selection in codon-based approaches, and the latter examined variation in dN/dS along lineages, which is not powerful when only a few codons are under selection.

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 towards sequencing putatively neutral regions

of the genome to infer historical demography (e.g. Yu et al. 2003; Yu et al. 2004; Fischer et al. 2006). However, from both an evolutionary and 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 that are likely to affect the exposure to pathogens, overall patterns of nucleotide variation at human and chimpanzee TLRs were fairly similar.

A summary of all intraspecific and interspecific tests of selection is given in Table 12. Although at the population level some of the genes showed departures from the neutral expectation, in humans (Table 3 and Table 12) these deviations generally disappeared when the effects of demography were taken into account (Figure 4). 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 lines of evidence suggest that segregating variants at the TLR6-TLR1-TLR10 cluster affect function. (1) TLR10 is robust to the known demographically caused deviations of the allele frequency spectrum, showing extreme patterns of polymorphism in Europeans (Table 3), and it has a weak excess of polymorphism with respect to divergence (Table 7). (2) TLR1 shows a significant excess of polymorphism in an HKA test (Table 7). (3) As part of genome-wide association study, very high differentiation among British people was found for a SNP within the

TLR6-1-10 cluster (Burton et al. 2007). (4) TLR1 N248S, one of the SNPs in LD with TLR10, has been shown to present striking North-South clinal variation (Todd et al. 2007). (5) The four SNPs that we found in moderate LD between TLR1 and TLR10, including TLR1 N248S, have been linked to several disease phenotypes. In addition to the aforementioned clinical associations of TLR1 N248S, the haplotype containing these SNPs has been associated with reduced risk of prostate cancer (Stevens et al. 2008). TLR1 I602S results in severe impairment of NF-κB signaling (Hawn et al. 2007), abnormal trafficking to cell surface and protection against leprosy (Johnson et al. 2007). Both SNPs at TLR1 have been associated, alone and in combination, with tuberculosis (Ma et al. 2007). Finally TLR10 I369L in association with other SNPs has also been implicated with increased risk of nasopharyngeal carcinoma (Zhou et al. 2006).

Many of these observations seem to suggest the action of positive natural selection within Europeans in one or more of the genes within this cluster. However, similar to Todd et al. (2007) we found no evidence of extended haplotype homozygosity associated with these genes. At this moment, we cannot discard the possibility that geographically restricted selection, perhaps in conjunction with a more complicated demographic history has shaped the observed pattern at these loci.

In sum, in contrast with the high incidence of positive selection that we found at deeper divergences, we did not find strong evidence of selection within humans or within chimpanzees (Table 12), although the TLR6-TLR1-TLR10 case deserves further investigation. Barreiro et al. (2009) concluded that TLR1 I602S has been the target of selection in non-Africans. This conclusion is based on extreme differentiation, reduced

polymorphism and long-range haplotype homozygosity associated with the derived allele, and functional assays confirming reduced signaling. Although the evidence is certainly very suggestive, the picture is far from being complete and other scenarios cannot be excluded. First, in their analyses, the gene that globally shows strongest deviations from neutrality is TLR10, not TLR1, but for technical reasons, variants at this gene belonging to the same haplotype were not functionally tested. Second, the true target of selection might be some regulatory element rather than a coding variant, but noncoding variants were not included in the functional assays. While *in vitro* assays are a valuable tool to study function, they do not necessarily reflect the true functional effect *in vivo*. Finally, the link with the putative selective agent is missing. The population patterns and the observations mentioned above suggest that the evolutionary history of the TLR6-1-10 cluster is very complex, and that signatures of selection and demography might be very difficult to disentangle.

Relaxed selection at human non-viral 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. Surprisingly, 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), driven by non-viral TLRs. We note that the chimpanzee smaller sample size could bias against sampling low frequency polymorphisms. However, the trends were very similar when we analyzed the human

sample from Africa separately (results not shown). The human sample from Africa has a comparable size to the chimpanzee sample, so the effect, if any, is probably small.

The excess of damaging changes in human non-viral TLRs compared with the human genome and chimpanzee TLRs suggests a very recent relaxation of selective constraint in the human lineage that mostly or only affected the non-viral TLRs. The increase in the proportion of damaging polymorphisms in humans, without a concomitant change in the dN/dS ratio is compatible with this scenario. 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.

Humans and chimpanzees differ in many aspects of their ecology. The introduction of domestication and agriculture in the last 10,000 years marked a major shift in human lifestyle that was likely linked to changes in selective regimes associated with new diets, social structures and changes in the dynamics of infectious diseases (Larsen 1995). It is difficult to conceive, though, how such early conditions might have resulted in relaxed selection of a subset of the TLRs. On the other hand, it is frequently suggested that deleterious alleles have been accumulating in the human genome at a faster pace recently because of the decreased efficacy of natural selection in removing these mutations in modern populations with greatly reduced mortality (e.g. Crow 1997). However, most of the obvious changes in sanitation and technology that had an impact on mortality (such as the massive administration of antibiotics in the last century) are probably too recent to have left a signal in the pattern of sequence variation.

If indeed these mutations are neutral or slightly deleterious, 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). Using the allele frequencies for the entire human sample, and 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.

It is possible that some of these rare damaging polymorphisms are associated with human diseases (Yue and Moult 2006). The idea that rare variants can contribute significantly to human complex diseases is an appealing hypothesis that is currently being explored (Fearnhead, Winney, and Bodmer 2005).

Viral receptors are under stronger purifying selection

We uncovered consistent differences between viral and non-viral 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 non-viral TLRs.

Although not significant, viral TLRs have a smaller proportion of damaging polymorphisms in both human and chimpanzees. This suggests that viral TLR polymorphisms are mostly neutral while non-viral TLRs also segregate some slightly deleterious polymorphisms. This observation is in line with the N.I., which revealed a weak excess of replacement polymorphisms in non-viral TLRs. Most of these polymorphisms are rare in the populations, suggesting they have mild deleterious effects; while they might remain in the populations at low frequencies, they usually do not become fixed between species. Another observation in support of the stronger constraint of viral TLRs is that premature stop codons (as reported in dbSNP) are more frequent in non-viral TLRs than in viral TLRs (6 vs 1). Very 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 non-viral 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, Kawai, and Akira 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 non-viral 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. Non-viral 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. Usually these nucleotides have modifications that distinguish them from host components, but they can be at least partially self-reactive. Unlike TLRs expressed in the cell membrane, viral TLRs remain in the endoplasmic reticulum in a resting state and traffic to endosomal vesicles upon 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 non-generic 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 have resulted in receptors that are more evolutionarily inflexible than the non-viral counterparts.

ACKNOWLEDGEMENTS

We especially 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 very useful discussions of the data. This work was supported by an NIH grant to M.W. N.

REFERENCES

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

Akira, S, and Takeda, K. 2004. Toll-like receptor signalling. Nature Reviews Immunology 4:499-511.

Akira, S, Uematsu, S, and 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, and Schwartz, DA. 2000. TLR4 mutations are associated with endotoxin hyporesponsiveness in humans. Nat. Genet. 25:187-+.

Barreiro, LB, Ben-Ali, M, Quach, Hln et al. 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, and 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, and Vaarala, O. 2004. A TLR4 polymorphism is associated with asthma and reduced lipopolysaccharide-induced interleukin-12(p70) responses in Swedish children. Journal of Allergy and Clinical Immunology 114:561-567.

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

Carpenter, S, and O'Neill, LAJ. 2007. How important are Toll-like receptors for antimicrobial response? Cellular Microbiology 9:1891-1901.

Crow, JF. 1997. The high spontaneous mutation rate: Is it a health risk? P. Natl. Acad. Sci. USA 94:8380-8386.

Dean, MD, Good, JM, and 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, and 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, and Barton, GM. 2008. The ectodomain of Toll-like receptor 9 is cleaved to generate a functional receptor. Nature 456:658-U688.

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

Fearnhead, NS, Winney, B, and Bodmer, WF. 2005. Rare variant hypothesis for multifactorial inheritance - Susceptibility to colorectal adenomas as a model. Cell Cycle 4:521-525.

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

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

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

Fu, YX, and Li, WH. 1993. Statistical Tests of Neutrality of Mutations. Genetics 133:693-709.

Gabriel, SB, Schaffner, SF, Nguyen, H et al. 2002. The structure of haplotype blocks in the human genome. Science 296:2225-2229.

Garrigan, D, Kingan, SB, Pilkington, MM et al. 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. 2007. Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222-234.

Hawn, TR, Misch, EA, Dunstan, SJ et al. 2007. A common human TLR1 polymorphism regulates the innate immune response to lipopeptides. European Journal of Immunology 37:2280-2289.

Hawn, TR, Verbon, A, Lettinga, KD et al. 2003. A common dominant TLR5 stop codon polymorphism abolishes flagellin signaling and is associated with susceptibility to legionnaires' disease. Journal of Experimental Medicine 198:1563-1572.

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

Hudson, RR, Kreitman, M, and 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, and Yeager, M. 2003. Widespread purifying selection at polymorphic sites in human protein-coding loci. P. Natl. Acad. Sci. USA 100:15754-15757.

Jin, MS, Kim, SE, Heo, JY, Lee, ME, Kim, HM, Paik, SG, Lee, HY, and 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, and Tapping, RI. 2007. Cutting edge: A common polymorphism impairs cell surface trafficking and functional responses of TLR1 but protects against leprosy. Journal of Immunology 178:7520-7524.

Kariko, K, Buckstein, M, Ni, HP, and 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, and Schwartz, DA. 2002. Toll-like receptor 4 polymorphisms and atherogenesis. New England Journal of Medicine 347:185-192.

Kimura, M, and 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, and Kabesch, M. 2008. Toll-like receptor heterodimer variants protect from childhood asthma. Journal of Allergy and Clinical Immunology 122:86-92.

Kosakovsky Pond, SL, and 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, and Vollmer, J. 2007. Toll-like receptors 7, 8, and 9: linking innate immunity to autoimmunity. Immunological Reviews 220:251-269.

Ku, CL, Yang, K, Bustamante, J et al. 2005. Inherited disorders of human Toll-like receptor signaling: immunological implications. Immunological Reviews 203:10-20.

Kumar, H, Kawai, T, and Akira, S. 2009. Pathogen recognition in the innate immune response. Biochemical Journal 420:1-16.

Larsen, CS. 1995. Biological changes in human populations with agriculture. Annual Review of Anthropology 24:185-213.

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

Lazarus, R, Raby, BA, Lange, C, Silverman, EK, Kwiatkowski, DJ, Vercelli, D, Klimecki, WJ, Martinez, FD, and Weiss, ST. 2004. TOLL-like receptor 10 genetic variation is associated with asthma in two independent samples. American Journal of Respiratory and Critical Care Medicine 170:594-600.

Lewontin, R. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49:49–67.

Librado, P, and 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, and 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. 2008. Proportionally more deleterious genetic variation in European than in African populations. Nature 451:994-U995.

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

Ma, X, Liu, Y, Gowen, BB, Graviss, EA, A.G., C, and J.M., M. 2007. Full-exon resequencing reveals toll-like receptor variants contribute to human susceptibility to tuberculosis disease. PLoS One.

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

Mcdonald, JH, and Kreitman, M. 1991. Adaptive Protein Evolution at the Adh Locus in Drosophila. Nature 351:652-654.

Miao, EA, Andersen-Nissen, E, Warren, SE, and Aderem, A. 2007. TLR5 and Ipaf: Dual sensors of bacterial flagellin in the innate immune system. Seminars in Immunopathology 29:275-288.

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

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

Nei, M, and Li, WH. 1979. Mathematical-Model for Studying Genetic-Variation in Terms of Restriction Endonucleases. P. Natl. Acad. Sci. USA 76:5269-5273.

Ng, PC, and Henikoff, S. 2006. Predicting the effects of amino acid substitutions on protein function. Annual Review of Genomics and Human Genetics 7:61-80.

Nielsen, R, and 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, and Tapping, RI. 2007. The polymorphism P315L of human toll-like receptor 1 impairs innate immune sensing of microbial cell wall components. Journal of Immunology 178:6387-6394.

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

Palm, NW, and Medzhitov, R. 2009. Pattern recognition receptors and control of adaptive immunity. Immunological Reviews 227:221-233.

Park, BS, Song, DH, Kim, HM, Choi, BS, Lee, H, and Lee, JO. 2009. The structural basis of lipopolysaccharide recognition by the TLR4-MD-2 complex. Nature 458:1191-U1130.

Pond, SK, and Muse, SV. 2005. Site-to-Site Variation of Synonymous Substitution Rates. Mol Biol Evol 22:2375-2385.

Pond, SLK, and Frost, SDW. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. Bioinformatics 21:2531-2533.

Rand, DM, and 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, and Aderem, A. 2005. The evolution of vertebrate Toll-like receptors. P. Natl. Acad. Sci. USA 102:9577-9582.

Sackton, TB, Lazzaro, BP, Schlenke, TA, Evans, JD, Hultmark, D, and Clark, AG. 2007. Dynamic evolution of the innate immune system in Drosophila. Nat. Genet. 39:1461-1468.

Sawyer, SL, Emerman, M, and Malik, HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. Plos Biology 2:1278-1285.

Sawyer, SL, Wu, LI, Emerman, M, and Malik, HS. 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. P. Natl. Acad. Sci. USA 102:2832-2837.

Schroder, NW, and Schumann, RR. 2005. Single nucleotide polymorphisms of Toll-like receptors and susceptibility to infectious disease. Lancet Infectious Diseases 5:156-164.

Schuring, RP, Hamann, L, Faber, WR, Pahan, D, Richardus, JH, Schumann, RR, and Oskam, L. 2009. Polymorphism N248S in the Human Toll-Like Receptor 1 Gene Is Related to Leprosy and Leprosy Reactions. Journal of Infectious Diseases 199:1816-1819.

Simonsen, KL, Churchill, GA, and Aquadro, CF. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics 141:413-429.

Smirnova, I, Hamblin, MT, McBride, C, Beutler, B, and 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, and Rodriguez, C. 2008. Genetic variation in the toll-like receptor gene cluster (TLR10-TLR1-TLR6) and prostate cancer risk. International Journal of Cancer 123:2644-2650.

Sunyaev, S, Ramensky, V, Koch, I, Lathe, W, Kondrashov, AS, and 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, and Akira, S. 2002. Cutting edge: Role of Toll-like receptor 1 in mediating immune response to microbial lipoproteins. Journal of Immunology 169:10-14.

Thompson, JD, Gibson, TJ, Plewniak, F, Jeanmougin, F, and 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. 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, and Pritchard, JK. 2006. A map of recent positive selection in the human genome. Plos Biology 4:446-458.

Waterson, GA. 1975. On the number of segregating sites in genetical models without recombination. Theoretical population biology 7.

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

Wlasiuk, G, Khan, S, Switzer, WM, and 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, and 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, Wong, WSW, and Nielsen, R. 2005. Bayes empirical Bayes inference of amino acid sites under positive selection. Mol. Biol. Evol. 22:1107-1118.

Yu, N, Jensen-Seaman, MI, Chemnick, L, Kidd, JR, Deinard, AS, Ryder, O, Kidd, KK, and Li, WH. 2003. Low nucleotide diversity in chimpanzees and bonobos. Genetics 164:1511-1518.

Yu, N, Jensen-Seaman, MI, Chemnick, L, Ryder, O, and Li, WH. 2004. Nucleotide diversity in gorillas. Genetics 166:1375-1383.

Yue, P, and Moult, J. 2006. Identification and analysis of deleterious human SNPs. Journal of Molecular Biology 356:1263-1274.

Zhou, XX, Jia, WH, Shen, GP, Qin, HD, Yu, XJ, Chen, LZ, Feng, QS, Shugart, YY, and Zeng, YX. 2006. Sequence variants in toll-like receptor 10 are associated with nasopharyngeal carcinoma risk. Cancer Epidemiology Biomarkers & Prevention 15:862-866

Table 1. Phylogenetic tests of recurrent positive selection

		Test of selection[a]					Sites under selection identified by different methods[b]			
Gene	No. species	lnL M7 (neutral)	lnL M8 (selection)	-2lnΔL[c]	Sig[d]	p_s, ω_s [e]	PAML M8[f]	SLAC[g]	FEL[h]	REL[i]
TLR1	11	-7830.4	-7820.62	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	-6674.56	-6672.78	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	-6108.17	-6108.18	-0.02	ns				79, 715	7, 79, 86, 356, 715
TLR4	11	-8156.32	-8142.58	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[j]	20	-2629.41	-2616.18	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	-7560.73	-7559.07	3.32	ns				400, 407, 567	
TLR6	11	-6652.14	-6647.37	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	-6827.06	-6823.54	7.04	**	0.05, 9.35	486, 542, 693		528, 542	2, 37, 39, 42, 43, 44, 11, 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	-8401.8	-8394.55	14.5	**	0.03, 4.20	159, 225, 237, 469, 738, 765		267, 522, 738, 783	
TLR9	9	-7787.99	-7776.32	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	-4746.47	-4746.46	0.02	ns					261

[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 bold and italics.
[c] $-2\ln\Delta L$ is distributed approximately as χ^2 with 2 degrees of freedom.
[d] **=p<0.01, ns= not significant.
[e] p_s=proportion of the sites under selection, ω_s=estimated dN/dS of the sites under selection in M8.
[f] Codons with posterior probablilities > 90% in the Bayes empirical Bayes analyses.
[g] Codons with p-values < 0.1.
[h] Codons with p-values < 0.1
[i] Codons with Bayes factors > 50.
[j] Dataset from Nakajima et al. (2008), fragment of ~600 nt.

Table 2. Sites predicted to affect function based on their location in the 3D structure.

Gene	Position [a]	Functional information [b]	Site identified by: [c]	Reference 3D 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*	ligand binding	REL, FEL, M8 (<0.9)	
	321*	dimerization surface (ionic interaction)	REL, FEL, M8	
	337*	dimerization surface (ionic interaction and hydrogen bond) and adjacent to site involded in ligand binding	M8 (<0.9)	
TLR2	267	adjacent to site involved in ligand binding	REL, M8 (<0.9)	Jin et al. 2007
	294*	ligand binding	M8 (<0.9)	
	296	adjacent to site involved in ligand binding	M8 (<0.9)	
	270*	ligand binding	M8 (<0.9)	
	318*	dimerization surface (ionic interaction) and adjacent to site involved in ligand binding	M8 (<0.9)	
	321*	dimerization surface (ionic interaction)	REL, M8	
	324*	dimerization surface (hydrophobic interaction)	REL, M8 (<0.9)	
	326*	dimerization surface (hydrophobic interaction)	REL, M8 (<0.9)	
	329	adjacent to site involved in ligand binding	M8 (<0.9)	
	338*	ligand binding	M8 (<0.9)	
	354	adjacent to site involved in ligand binding	FEL, REL, M8 (<0.9)	
	373*	dimerization surface (hydrophobic interaction)	M8 (<0.9)	
	376	adjacent to site involved in dimerization surface (hydrogen	REL, M8	

		bond)		
TLR3	86*	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*	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*	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*	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; *=Sites more likely to affect function based on location in protein structure. Sites identified by more than one method from Table1 are underlined.
[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.

Table 3. Allele frequency spectrum statistics

locus	No. alleles	No. sites	S	Singletons	θw (%)	π (%)	Tajima's D	Fu-Li D*	Fu-Li F*	Fay-Wu H [a]
HUMAN ALL										
TLR1 AA	48	10161	72	16	0.16	0.10	-1.44*	0.10	-0.55	-5.43
TLR1 EA	46	10161	35	3	0.08	0.07	-0.51	1.18	0.69	-10.41*
TLR2 AA	48	5722	18	10	0.07	0.05	-0.96	-2.32*	-2.20*	1.58
TLR2 EAs	46	5722	11	5	0.04	0.04	-0.46	-1.37	-1.25	0.23
TLR3 AA	48	6343	21	6	0.07	0.05	-1.11	-0.41	-0.77	2.69
TLR3 EA	46	6343	8	0	0.03	0.04	0.77	1.30	1.33	1.79
TLR4 AA	48	14523	66	19	0.10	0.08	-0.75	-0.51	-0.71	0.38
TLR4 EA	46	14523	34	6	0.05	0.05	-0.33	0.45	0.21	-2.33
TLR5 AA	48	18640	132	56	0.16	0.12	-0.96	-1.77*	-1.75*	-24.42
TLR5 EA	46	18640	65	9	0.08	0.09	0.55	0.81	0.86	-9.68
TLR6 AA	48	6601	44	9	0.15	0.14	-0.22	0.21	0.07	-0.64
TLR6 EA	46	6601	24	2	0.08	0.12	1.57*	1.12	1.51*	-1.89
TLR7 AA	36	24146	80	14	0.11	0.10	-0.39	0.58	0.29	-0.94
TLR7 EA	34	24145	43	12	0.06	0.05	-0.50	-0.20	-0.35	-2.20
TLR8 AA	36	16622	60	18	0.12	0.12	0.11	-0.36	-0.23	-3.02
TLR8 EA	34	16622	37	11	0.07	0.09	0.67	-0.32	0.02	0.50
TLR9 AA	48	8429	12	3	0.03	0.03	-0.22	-0.13	-0.18	-0.78
TLR9 EA	46	8429	12	7	0.03	0.03	-0.64	-2.23*	-2.01*	-0.87
TLR10 AA	48	6966	70	12	0.23	0.23	0.09	0.52	0.44	-5.46
TLR10 EA	46	6966	54	31	0.18	0.10	-1.52*	-2.89*	-2.86*	-23.50*
HUMAN CODING										
TLR1 AA	48	2358	12	3	0.11	0.09	-0.62	-0.13	-0.34	-0.91
TLR1 EA	46	2358	7	1	0.07	0.06	-0.47	0.48	0.22	-3.23*
TLR2 AA	48	2352	9	5	0.09	0.04	-1.44*	-1.91	-2.06*	0.73
TLR2 EAs	46	2352	6	3	0.06	0.04	-0.82	-1.36*	-1.39	-0.11
TLR3 AA	48	2712	5	3	0.04	0.01	-1.74*	-1.77*	-2.06*	0.31
TLR3 EA	46	2712	1	0	0.01	0.02	1.13	0.55	0.83	0.25
TLR4 AA	48	2470	7	2	0.06	0.04	-0.81	-0.30	-0.54	-0.06
TLR4 EA	46	2470	3	1	0.03	0.01	-1.45	-0.39	-0.82	0.20
TLR5 AA	48	2574	12	7	0.11	0.04	-1.73*	-2.27*	-2.46*	-0.20
TLR5 EA	46	2574	6	1	0.05	0.05	-0.18	0.33	0.20	0.27
TLR6 AA	48	2388	11	2	0.10	0.09	-0.43	0.30	0.07	1.13

TLR6 EA	46	2388	7	2	0.07	0.08	0.58	-0.28	-0.01	0.14
TLR7 AA	36	3144	3	0	0.03	0.04	0.56	0.93	0.95	0.67
TLR7 EA	34	3144	5	3	0.05	0.03	-1.29	-1.56*	-1.72	0.53
TLR8 AA	36	3177	9	2	0.09	0.09	-0.07	0.14	0.09	0.47
TLR8 EA	34	3177	5	1	0.05	0.07	1.16	0.23	0.60	-0.11
TLR9 AA	48	3096	4	2	0.03	0.02	-0.72	-1.18	-1.21	-0.29
TLR9 EA	46	3096	3	2	0.02	0.02	-0.29	-1.68	-1.47	0.20
TLR10 AA	48	2433	19	4	0.18	0.18	0.04	0.14	0.13	-0.74
TLR10 EA	46	2433	17	7	0.16	0.12	-0.87	-1.25	-1.32	-2.38
CHIMP ALL										
TLR1	38	5381	18	3	0.08	0.06	-0.74	0.52	0.12	-0.22
TLR2	38	4461	18	2	0.09	0.07	-0.75	0.88	0.41	-6.48*
TLR3	38	4103	7	2	0.04	0.03	-0.46	-0.21	-0.34	0.08
TLR4	38	5069	11	5	0.05	0.05	-0.20	-1.24	-1.06	0.52
TLR5	38	4427	9	1	0.05	0.06	0.42	0.74	0.75	-1.71
TLR6	38	4312	10	6	0.06	0.04	-0.77	-2.04	-1.92*	-3.49*
TLR7	20	4486	5	1	0.04	0.05	0.35	0.39	0.43	-0.65
TLR8	20	5053	9	3	0.07	0.05	-0.69	-0.17	-0.37	1.63
TLR9	34	4287	11	3	0.06	0.06	-0.19	-0.12	-0.16	0.46
TLR10	36	4397	9	1	0.05	0.04	-0.58	0.75	0.40	-1.93
CHIMP CODING										
TLR1	38	2358	7	0	0.07	0.06	-0.48	1.26*	0.84	-1.84
TLR2	38	2352	8	1	0.08	0.05	-0.96	0.64	0.16	-2.87*
TLR3	38	2712	3	1	0.03	0.02	-0.85	-0.34	-0.56	0.39
TLR4	38	2470	2	1	0.02	0.01	-1.29	-0.81	-1.10	0.15
TLR5	38	2574	4	1	0.04	0.04	-0.06	-0.02	-0.04	-0.47
TLR6	38	2388	5	3	0.05	0.04	-0.58	-1.63*	-1.53	-4.10*
TLR7	20	3147	1	0	0.01	0.01	-0.59	0.65	0.37	-1.52
TLR8	20	3117	5	1	0.06	0.05	-0.56	0.39	0.14	0.94
TLR9	34	3096	6	1	0.05	0.05	0.06	0.41	0.35	1.18
TLR10	36	2433	3	1	0.03	0.03	0.11	-0.32	-0.22	0.49

[a] Chimpanzee sequence used to polarize derived mutations

AA=African-Americans, EA=European-Americans, *= $p<0.05$ (based on coalescent simulations).

TR7 and TLR8 are X-linked, so θw and π were multiplied by 4/3 (the effective populations size of X-linked loci is ¾ that of autosomes) to make the polymorphism levels comparable among genes.

Table 4. Differentiation between human populations

Locus	F_{ST}[a]
TLR1	0.085
TLR2	0.073
TLR3	0.025
TLR4	0.051
TLR5	0.019
TLR6	0.060
TLR7	0.077
TLR8	0.055
TLR9	0.014
TLR10	0.075
Seattle SNPs genes[b]	0.070

[a] Calculated on a per gene basis
[b] Based on 323 genes from Seattle SNPs

Table 5. Linkage disequilibrium between intermediate-frequency [a] nonsynonymous SNPs at TLR1 and TLR10.

TLR1 (SNP1)	Variants [b]	TLR10 (SNP2)	Variants [b]	distance (nt)	D'
rs4833095	A743G (S248N)	rs11096957	A721C (H241N)	23,219	0.642*
rs4833095	A743G (S248N)	rs11096955	A1105C (I369L)	23,603	0.642*
rs5743618	T1805G (I602S)	rs11096957	A721C (H241N)	22,157	-0.647*
rs5743618	T1805G (I602S)	rs11096955	A1105C (I369L)	22,541	-0.647*

[a] Greater than 20% in the combined human populations.
[b] Nucleotide positions with respect to the coding region (amino acid positions).
* significant after Bonferroni correction (including all polymorphic sites).

Table 6. Polymorphism and divergence for silent and replacement sites in humans and chimpanzees.

	Human Polymorphism		Human-Macaque Divergence		N.I. [a]	p-value	Chimp Polymorphism		Chimp-Macaque Divergence		N.I.[a]	p-value
Gene	R	S	R	S			R	S	R	S		
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
Non-viral	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] N.I.= Neutrality index = $(R/S)_{pol} / (R/S)_{div}$

Table 7. HKA tests

		Polymorphisms		Fixed differences [a]		
Species	Gene	Observed	Expected	Observed	Expected	p-value [c]
Humans	TLR1	77	55	91	113	0.04
	TLR2	20	17	35	38	0.53
	TLR3	21	27	67	61	0.36
	TLR4	75	67	139	147	0.54
	TLR5	143	138	294	300	0.81
	TLR6	47	45	99	101	0.85
	TLR7	99	87	199	211	0.42
	TLR8	72	58	145	159	0.23
	TLR9	20	28	71	63	0.23
	TLR10	75	57	100	118	0.09
	viral TLRs	269	279	482	472	0.79
	non-viral TLRs	437	427	758	768	-
Chimpanzees	TLR1	18	16	46	48	0.62
	TLR2	18	12	28	34	0.10
	TLR3	7	9	34	32	0.48
	TLR4	11	9	25	27	0.48
	TLR5	9	11	38	36	0.62
	TLR6	10	16	69	63	0.22
	TLR7	5	6	29	28	0.83
	TLR8	9	8	38	39	0.76
	TLR9	11	11	36	36	0.96
	TLR10	9	12	47	44	0.40
	viral TLRs	37	39	137	135	0.85
	non-viral TLRs	75	73	253	255	-

[a] Divergence was calculated with respect to the chimpanzee sequence
[b] Divergence was calculated with respect to the human sequence
[c] Each TLR was compared against a 'neutral' set of 10 combined human loci in a 2-locus HKA test. Viral and non-viral combined sets were compared against each other.
[d] Each TLR was compared against a 'neutral' set of 26 combined chimpanzee loci in a 2-locus HKA test. Viral and non-viral combined sets were compared against each other.

Table 8. Number of 'damaging' and 'benign' polymorphisms within humans and chimpanzees, predicted by Polyphen.

		Humans		Chimpanzees	
Gene	Class	Damaging	Benign	Damaging	Benign
TLR1	Non-viral	4	5	1	5
TLR2	Non-viral	4	1	0	3
TLR3	Viral	2	1	0	2
TLR4	Non-viral	4	1	1	0
TLR5	Non-viral	7	3	1	2
TLR6	Non-viral	1	6	1	2
TLR7	Viral	0	2	0	1
TLR8	Viral	1	2	0	0
TLR9	Viral	2	3	1	3
TLR10	Non-viral	6	7	1	1
All	All	31	31	6	19

Table 9. Ratios of damaging to benign polymorphisms in humans and chimpanzees for different subclasses of TLRs.

	Damaging	Benign	P-value [b]
Human Genome [a]	25361	49795	-
All human TLRs	31	31	<0.001
Human Viral TLRs	5	8	0.8
Human Non-viral TLRs	26	23	0.005
All Chimp TLRs	6	19	0.4 (0.03) [c]
Chimp Viral TLRs	1	6	0.4
Chimp Non-viral TLRs	5	13	0.6

[a] From Sunyaev et al. 2001
[b] Compared to the human genome
[c] Comparison to all human TLRs in parenthesis

Table 10. dN/dS values for the human lineage, the chimpanzee lineage and across primates.

	Human branch [a]	Chimp branch [a]	Primates [b]				
			global	SP [c]	EXT [c]	TM [c]	CYT [c]
TLR1	1.04	1.35	0.43	0.59	0.45	0.33	0.21
TLR2	Inf. [d]	0.31	0.44	0.40	0.52	0.21	0.13
TLR3	0.20	0.64	0.32	0.87	0.30	Inf. [d]	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
Non-viral average	0.51	0.74	0.46	0.58	0.49	0.73	0.27
p-value [e]	0.09	0.1	0.007	0.17	0.02	0.49	0.05

[a] Macaque sequence used to polarize changes along the human or chimp branches
[b] Average dN/dS across primates
[c] SP=Signal Peptide, EXT=Extracellular domain, TM=Transmembrane domain, CYT=Cytoplasmic domain.
[d] dS=0
[e] Viral and non-viral averages were compared in a t-test.

Table 11. Distribution of sites with dN/dS<1 across protein domains.

		Protein Domains				
		Signal Peptide	**Extracellular**	**Transmembrane**	**Cytoplasmic**	**p-value**
Viral TLRs	Observed	10	395	3	115	0.14
	Expected	13	401	10	99	
Non-viral TLRs	Observed	21	415	12	183	0.21
	Expected	27	439	14	151	

Table 12. Summary of tests of selection at the phylogenetic and population levels.

TYPE	TEST	HUMANS																			
		TLR1		TLR2		TLR3		TLR4		TLR5		TLR6		TLR7		TLR8		TLR9		TLR10	
		AA	EA	AA	EA	AA	EA	AA	EA	AA	EA	AA	EA	AA	EA	AA	EA	AA	EA	AA	EA
Allele frequency spectrum [a]		TD	FWH	FLD FLF						FLD FLF			TD FLF						FLD FLF		TD FLD FLF FWH
Polymorphism-divergence	MK [b]																				
	HKA [c]	*																			
LD	iHS [d]									N/A											
FST [e]																					
		CHIMPANZEES																			
		TLR1		TLR2		TLR3		TLR4		TLR5		TLR6		TLR7		TLR8		TLR9		TLR10	
Allele frequency spectrum				FWH								FLF FWH									
Polymorphism-divergence	MK [b]																				
	HKA [c]																				
		PRIMATES																			
		TLR1		TLR2		TLR3		TLR4		TLR5		TLR6		TLR7		TLR8		TLR9		TLR10	
Phylogeny-based	Paml [f]	*						*				*		*		*		*			
	SLAC [g]	X																X			
	FEL [h]	X		X		X		X		X		X		X		X		X			
	REL [i]	X		X		X		X				X		X		X		X		X	

[a] Significant ($p<0.05$) results for TD=Tajima's D, FLD=Fu-Li's D*, FLF=Fu-Li's F*, FWH=Fay-Wu's H.

[b] For the coding region. Macaque sequence used as outgroup

[c] Chimpanzee sequence used as outgroup. 2 locus HKA against 10 neutral loci from Akey et al. (2004) (humans) or 26 neutral loci from Fischer et al. (2006) (chimpanzees).

[d] Gene based iHS. Significance derived from the empirical distribution of the proportion of SNPs per gene with $|iHS|>2$

[e] On a per gene basis. P-value based on the empirical distribution of F_{ST} for the 323 genes in Seattle SNPs

[f] LTR of M7 (neutral) *vs* M8 (selection)

[g] $dN>dS$ for individual sites with $p<0.1$

[h] $dN>dS$ for individual sites with $p<0.1$

[i] $dN>dS$ for individual sites with Bayes Factor>50

*=$P<0.05$, X= Evidence of selection according to criteria 7, 8 or 9

FIGURE LEGENDS

Figure 1. Positively selected sites in the 3D-structures of TLRs. In each case, areas important for ligand binding that contain a concentration of sites under selection are squared in red. Amino acid positions of positively selected sites are indicated.

A. TLR4-TLR4 dimer. Positively selected sites identified using the smaller phylogenetic sample that includes the entire coding region are labeled as red circles and the ones identified using the Najakima et al. (2004) dataset are labeled in green.

B. TLR1-TLR2 dimer. Positively selected sites are labeled in red.

Figure 2. Positive selection at primate TLRs. Distribution of positively selected changes among the main primate groups. For each major group the number of observed and expected amino acid changes was compared. The expected numbers were obtained by multiplying the total number of non-ambiguous changes by the total time of the clade in millions of years (by summing up times for all the branches) and dividing that by the total amount of time in the entire phylogeny. Divergence times were taken from Bininda-Emonds et al. (2007). Only unambiguous changes at the inferred positively selected sites (concordant between at least two ML methods) were used.

Apes=Apes+human, OWP=Old World primates, NWP=New World Primates, *=$p<0.05$

Figure 3. Positive selection at TLR4.

A. Physicochemical properties of the amino acids at the positively selected sites at TLR4. SM=small, NP=non-polar, 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.

Figure 4. 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.

Figure 1

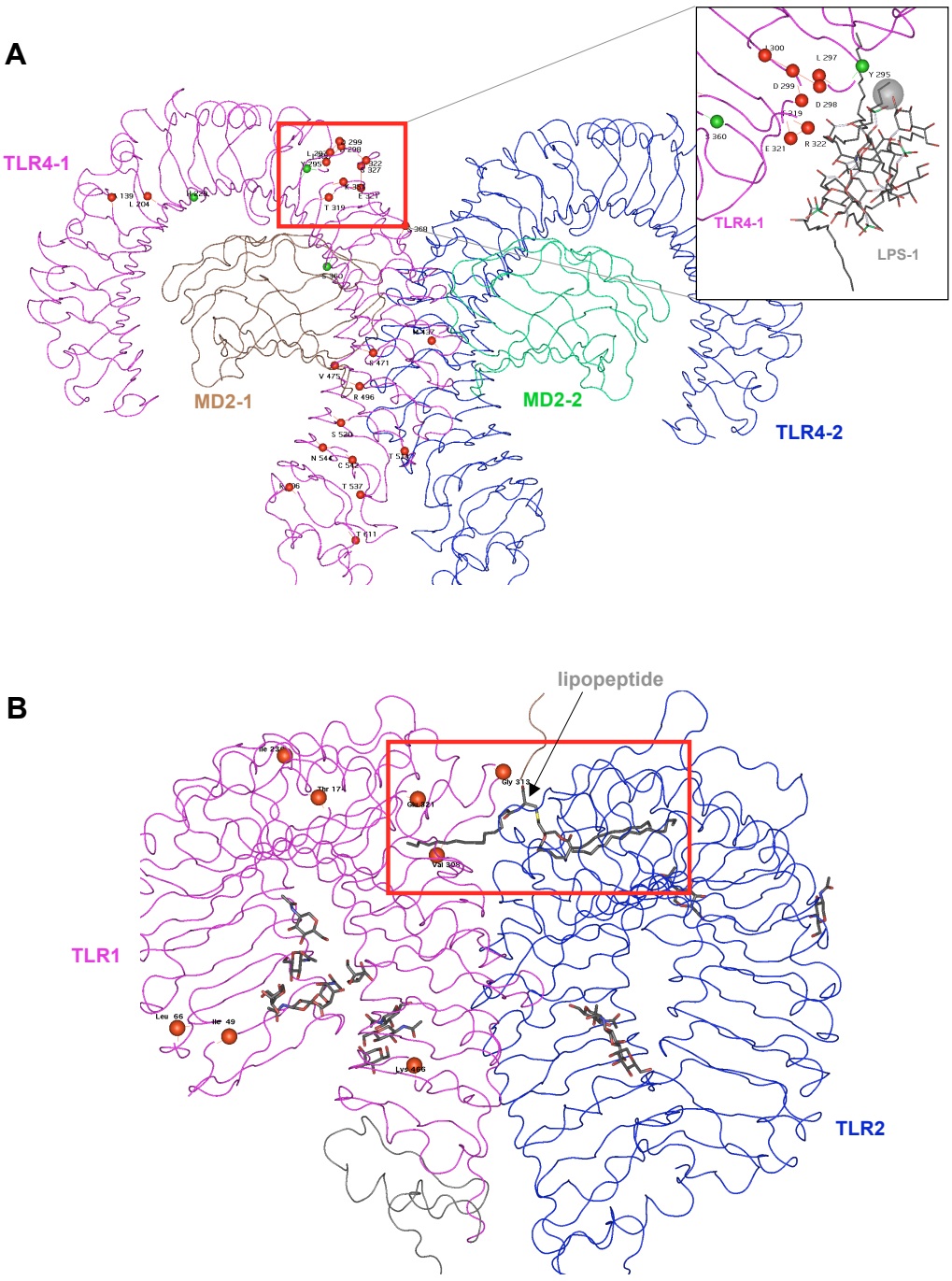
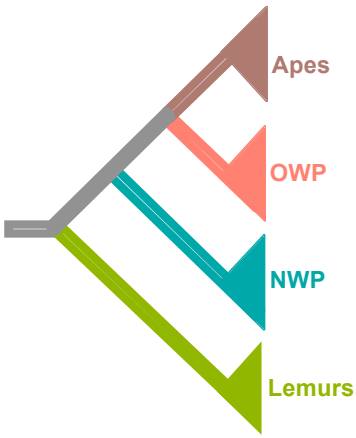


Figure 2



	TLR1	TLR4	TLR6	TLR7	TLR9
Apes	6 / 4.4	2 / 6.3	3 / 1.6	1 / 1.4	2 / 1.7
OWP	0 / 2.3	9 / 3.2 *	1 / 0.8	1 / 0.7	0 / 1.1
NWP	3 / 3.0	2 / 3.3	2 / 1.1	1 / 0.8	2 / 1.6
Lemurs	4 / 3.3	9 / 9.2	0 / 2.4	1 / 1.1	2 / 1.6

Figure 3

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

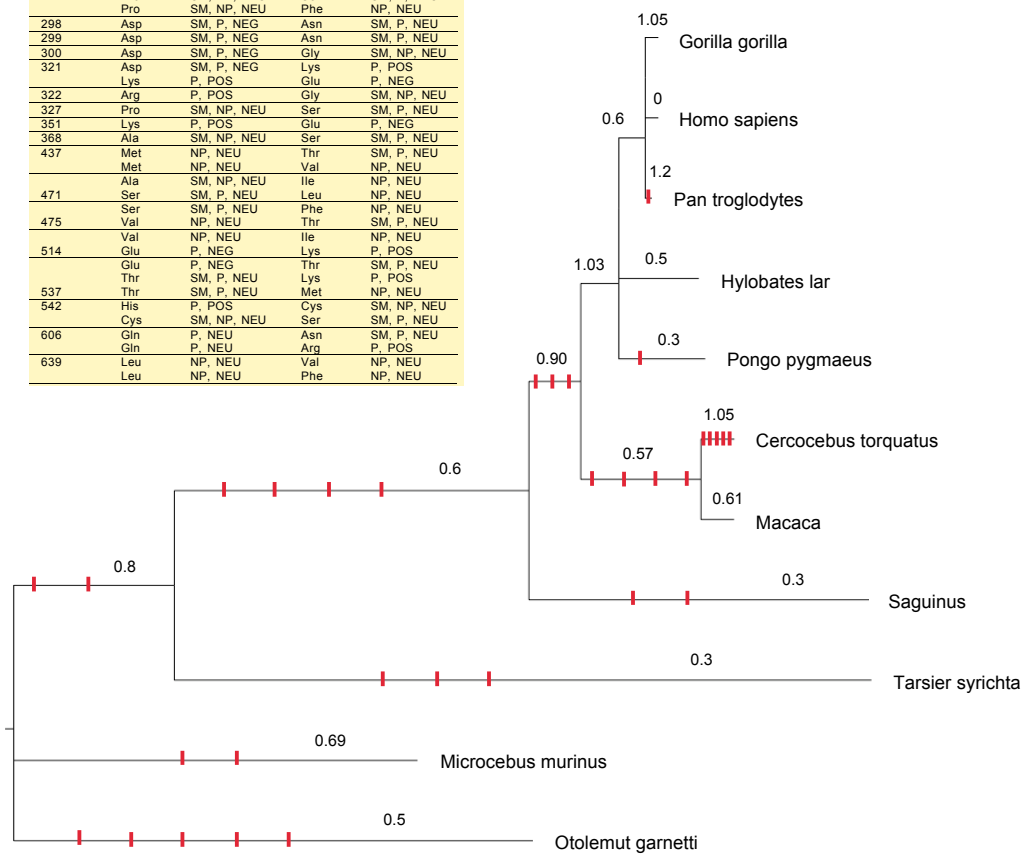


Figure 4

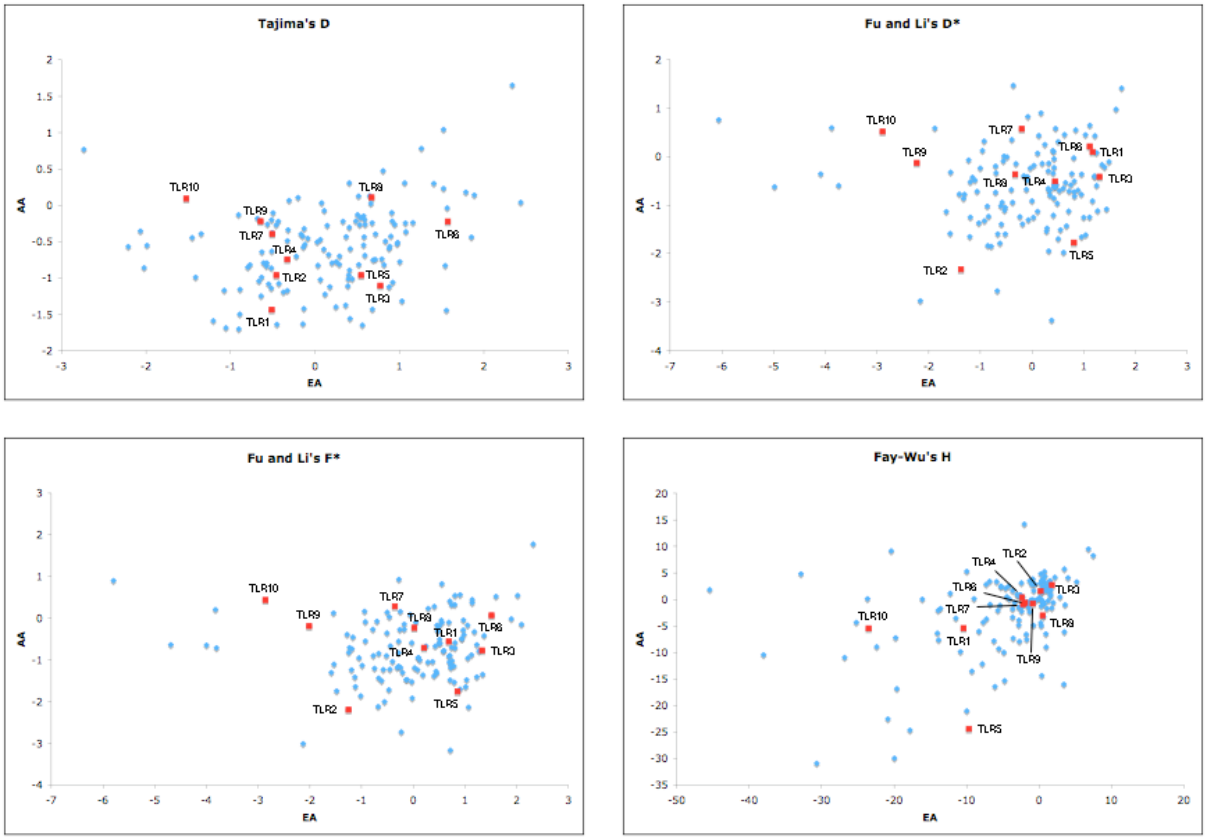

Tajima's D
AA
EA
TLR10
TLR9
TLR7
TLR4
TLR2
TLR5
TLR3
TLR1
TLR8
Fu and Li's D*
AA
EA
TLR10
TLR7
TLR9
TLR8
TLR1
TLR4
TLR3
TLR5
TLR2
Fu and Li's F*
AA
EA
TLR10
TLR7
TLR9
TLR8
TLR1
TLR8
TLR4
TLR3
TLR2
TLR5
Fay-Wu's H
AA
EA
TLR4
TLR2
TLR6
TLR3
TLR10
TLR7
TLR8
TLR1
TLR9
TLR5

Table S1. Sequence accession numbers of the sequences used in the phylogenetic analysis

Species	TLR1	TLR2	TLR3	TLR4	TLR5
Homo sapiens	ENSG00000174125	ENSG00000137462	ENSG00000164342	ENSG00000136869	ENSG00000187554
Pan troglodytes	ENSPTRG00000015985	ENSPTRG00000016528	ENSPTRG00000016658	ENSPTRG00000021299	FJ546367
Gorilla gorilla	ENSGGOG00000013191	ENSGGOG00000003066	AB445634	ENSGGOG00000007950	AB445648
Pongo pygmaeus	ENSPPYG00000014668	ENSPPYG00000015131	ENSPPYG00000015254	ENSPPYG00000019548	AB445649
Hylobates lar	EU488847	EU488848	EU488849	EU488850	
Hylobates pileatus					FJ542204
Cercocebus torquatus	EU204931	EU204932	EU204935	EU204937	EU204938
Macaca mulatta	ENSMMUG00000016754	ENSMMUG00000015847	ENSMMUG00000021762	ENSMMUG00000009112	ENSMMUG00000000870
Saguinus oedipus	EU488856	EU488857		EU488859	
Saguinus fuscicollis					FJ542216
Callithrix jacchus		*			FJ542218
Aotus nancymaae					
Tarsius syrichta	ENSTSYG00000005530	ENSTSYG00000011069	ENSTSYG00000009003	ENSTSYG00000003583	ENSTSYG00000011910
Microcebus murinus	**	ENSMICG00000003197		ENSMICG00000004341	ENSMICG00000003667

Species	TLR6	TLR7	TLR8	TLR9	TLR10
Homo sapiens	ENSG00000174130	ENSG00000196664	ENSG00000101916	ENSG00000173366	ENSG00000174123
Pan troglodytes	ENSPTRG00000015986	ENSPTRG00000028318	ENSPTRG00000021667	ENSPTRG00000014992	ENSPTRG00000015984
Gorilla gorilla	ENSGGOG00000008858	AB445662	ENSGGOG00000012324	ENSGGOG00000015810	AB445683
Pongo pygmaeus	ENSPPYG00000014669	AB445663	ENSPPYG00000020126	AB445677	ENSPPYG00000014667
Hylobates lar	EU488852	EU488853	EU488854		EU488855
Hylobates pileatus					
Cercocebus torquatus	EU204940	EU204942	EU204945	EU204946	
Macaca mulatta	ENSMMUG00000002418	ENSMMUG00000014254	ENSMMUG00000014255	ENSMMUG00000012263	ENSMMUG00000003373
Saguinus oedipus	EU488861				EU488864
Saguinus fuscicollis					
Callithrix jacchus	*	*		*	*
Aotus nancymaae				AY788894	
Tarsius syrichta					
Microcebus murinus	ENSMICG00000007905		ENSMICG00000014450		
Otolemur garnetti	ENSOGAG00000016497	ENSOGAG00000016092	ENSOGAG00000016094	ENSOGAG00000016022	

* = Sequences were obtained from the *C. jacchus* draft assembly at the UCSC genome browser using BLAT (with the human coding sequence).
** = Sequence obtained from the *M. murinus* draft assembly at the Ensembl browser using BLAT (with the human coding sequence).

Table S2. Primers used to PCR amplify and sequence chimpanzee TLRs

TLR1

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR01-080	TCTGGATTCAGGCATAGGAATAA	x	x	x	TLR01-081	AGCACCAAATTCCTCTTCACATA	x	x	x
TLR01-090	TCAGGATAACAGAGGCATATTGG	x	x	x	TLR01-091	GGCCTCTCAGAAGATCCATTTAT	x	x	x
TLR01-100	CCCAAGCACTGTCTCATACCTAC	x	x	x	TLR01-101	GTAATAGGTTGAATGGTGGCTC	x	x	x
TLR01-110	TCAGTTGCCTCTTCTCTTGTCTC	x	x	x	TLR01-111	GAACACAATGTGCAGACTCTCAG	x	x	x
TLR01-120	ATGCAAAGAGTTTGGCAATATGT	x	x	x	TLR01-121	AGTACCTTGATCCTGGGAGGTAA	x	x	x
TLR01-130	GCACCTTACTGAGTTGGAGACAC	x	x	x	TLR01-131	AAGAGATTGTGATGGGCAAAGTA	x	x	x
TLR01-140	TCATATAGTGGGCACGATTCTTT	x	x	x	TLR01-141	AGCTCTTTCTGACATTGGTTGAG	x	x	x
Amplification pairs		Tm							
TLR01-080 / TLR01-081		57							
TLR01-090 / TLR01-091		63							
TLR01-100 / TLR01-101		65							
TLR01-110 / TLR01-111		64							
TLR01-120 / TLR01-121		61							
TLR01-130 / TLR01-131		59							
TLR01-140 / TLR01-141		55							

TLR2

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR02-1FGW	CCTCAGGTATGGGATTAGCATC	x	x		TLR02-1RGW	TTTCTCTTTGCCTACCTTCTGG		x	
TLR02-2FGW	AGATTCTGTGAATGTGGGGAAT		x		TLR02-1RbGW	CGCAATCCAGGAAGTAAACATA		x	
TLR02-3FGW	AGTGATAGGTGTGAGGCAGGTT		x		TLR02-3RGW	TAGGTGATCCTGTTGTTGGACA	x		
TLR02-060	TGTATGTCTGTGTTAGCATTG	x	x	x	TLR02-3RbGW	TATGTGGCATTGTCCAGTGATT		x	
TLR02-070	ATTAGCAACAGTGACCTACAG		x	x	TLR02-061	CTTAGTGAAGGTGTCCATATT		x	x
TLR02-080	AAAGTTTGTTTCAGGTTATGA		x	x	TLR02-071	AACGTTTCCACTTTACCT		x	x
TLR02-090	TGAACTTATCCAGCACACGA		x	x	TLR02-081	CGTAGTTATTGCATTCCTACT		x	x
TLR02-100	CCTACTGGGTGGAGAACCTTA		x	x	TLR02-091	AGTCCAGTTCATACTTGCACC		x	x
TLR02-110	TCATTTGAGGACTTGCTTACT		x	x	TLR02-101	CGAGGATCCTGCAACCAAT		x	x
					TLR02-111	GAGAAACCTATCTAATAGCC	x	x	x
Amplification pairs		Tm							
TLR02-1FGW / TLR02-3RGW		62							
TLR02-060 / TLR02-111		52							

TLR3

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR03-050	TCCGTTTGATGTATGACTTGT	x	x	x	TLR03-051	GTGGCAAATTGGATTAGAT	x	x	x
TLR03-060	AGCCATGTGAAAATTTCTTG	x	x	x	TLR03-061	GCTCTTGACCATCGTACTCT	x	x	x
TLR03-070	CTTTCAACAGCCTTACAGAGT	x	x	x	TLR03-071	AGTCAGCTGCAGGTACTTGTT	x	x	x
TLR03-080	ATTCTCCCTTACACATACTCA	x	x	x	TLR03-081	AGTGTTGCAAAGGTAGTGG	x	x	x
TLR03-090	CAGGAACCTGACTGAGTTAGA	x	x	x	TLR03-092	TGAGTTAGATATGCGCTTT		x	x
TLR03-094	TTTCAGGGCTGGAGG		x	x	TLR03-093	GGGCAACAGAGCAGGAC	x		x
TLR03-100	CCAATATTACTCACAACTCAA	x	x	x	TLR03-091	AAAAAAAGAAGTTTAATTGAAAA		x	x
					TLR03-101	TTACAAAGGAAATAGCATTA	x		x
Amplification pairs		Tm			TLR03-103	TTTCTCCTTTGCTAATTGAAT		x	x
TLR03-050 / TLR03-051		59							
TLR03-060 / TLR03-061		54							
TLR03-070 / TLR03-071		60							
TLR03-080 / TLR03-081		53							
TLR03-090 / TLR03-093		63							
TLR03-100 / TLR01-103		52							

TLR4

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR04-040	AGCAAACTAATACAAATAGTGTT	x	x	x	TLR04-041	CACATAGCCACAAGGTAG	x		x
TLR04-050	ACTCTGCTCAAGGGTCAATGA	x		x	TLR04-045	AGGGCACACAGTGAGAAGT		x	x
TLR04-052	AAGGTCTGGACATCTTGACT		x	x	TLR04-051	GGAACCACCTGCAATGATTACAC	x		x
TLR04-060	TCTGCTTATCATGTATGCCTAAC	x		x	TLR04-053	GTATCTTTACTGCATGGTCT		x	x
TLR04-062	CCTGCTTGATGTCTTTGCCT		x	x	TLR04-061	CGGAATTCTTCCATGGTCAAA	x		x
TLR04-068A	CAAGCACGATATTGGAT	x		x	TLR04-063	CTAAACCAGCCAGACCTTGA		x	x
TLR04-070	TCAAGGTCTGGCTGGTTTAGA	x		x	TLR04-069A	CTGTCTTAAAAGAAAACCTAA	x		x
TLR04-072	AATTCCGATTAGCATACTTA		x	x	TLR04-071	TTGTGGCTCATATTTAGTACCTG	x		x
TLR04-080	TCAGAAACCTCATTTACCTTGAC	x		x	TLR04-073	AGTTAAATGCTGTTGGAGAC		x	x
TLR04-082	CTGGACCTCTCTCAGTGTCA		x	x	TLR04-081	GGGAATAAAGTCTCTGTAGTGA	x		x
TLR04-090	CGAATGGAATGTGCAACACC	x		x	TLR04-083	ACTAGCTCATTCCTTACCCA		x	x
TLR04-092	ACCTGTCAGATGAATAAGAC		x	x	TLR04-091	CACTGGAAGCTCCTTGAGATTAG	x		x
TLR04-100	GGCACATCTTCTGGAGACGAC	x		x	TLR04-093	ATTACTCACCCTTAGCATAA		x	x
TLR04-102	ATTGGCAGGAAGCAACATCT		x	x	TLR04-101	TTAGCTTATAGGCAAGACGTAAA	x		x
TLR04-110	ATTGTATTATGTTATAGCCA	x		x	TLR04-103	AGGAATTAGCCACTAGACTT		x	x
TLR04-112	AAGTCTAGTGGCTAATTCCT		x	x	TLR04-111	GAATCCCTTCACACATAGTTCTC	x		x
					TLR04-113	CTAAAGAGATTCAGAGGTCC		x	x

Amplification pairs	Tm
TLR04-040 / TLR04-041	48
TLR04-050 / TLR04-051	57
TLR04-060 / TLR04-061	50
TLR04-068A / TLR04-069A	48
TLR04-070 / TLR04-071	57
TLR04-080 / TLR04-081	48
TLR04-090 / TLR04-091	49
TLR04-100 / TLR04-101	49
TLR04-110 / TLR04-111	48

TLR6

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR06-050	CCATGCCCAGCTAAGTTT	x	x	x	TLR06-051	GCAATATAGGCCTTGTTATCA		x	x
TLR06-060	CCTCTGAACATGCGTCCAG		x	x	TLR06-061	GACCTGAAGCTCAGCTATGTA		x	x
TLR06-070	GAGTAACCATCTGATCTTTA		x	x	TLR06-071	TTTCTATGTGGTTGAGGGTAA		x	x
TLR06-080	CTGAATTTCTTGGGATTGAGT		x	x	TLR06-081	TCCAAAGAATTCCAGCTAACA	x	x	x
TLR06-090	CCTCATGCACCAAGCACATTC	x	x	x	TLR06-091	CAGGGCAGATCCAAGTAGATG		x	x
TLR06-100	ATAGAGGAACCCCACTAAAG		x	x	TLR06-101	TCTCTAACTGGCAGGTTGAGT		x	x
TLR06-110	CATGATGCAGCGGACTTA		x	x	TLR06-111	TTACATGCCGTAGCTGATCTA		x	x
TLR06-120	TGCCCTGTGAATCCTATTGGT		x	x	TLR06-121	AAGGGCAGGTTGCTATCAGT	x	x	x

Amplification pairs	Tm
TLR06-050 / TLR06-081	55
TLR06-090 / TLR06-121	60

TLR7

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR07-270	TAGCTGAATTACAGTTGTGTGCC	x	x	x	TLR07-261	ATCTTCCCAAGTGACTGAATCAA		x	x
TLR07-280	TCAAATGCCACTTACTACTGGGT	x	x	x	TLR07-271	TGGAGGGTGAAGAAATGTTAGAA	x	x	x
TLR07-290	TGGTAGAGATCGATTTCAGATGC	x	x	x	TLR07-281	TTTCTATGTTGGCCAGTTCTGTT	x	x	x
TLR07-300	CCAATTGGATCTGTCTTTCAATTT	x	x	x	TLR07-291	AGTGCCAAGATCAAGAACTTCAA	x	x	x
TLR07-310	GACAATGACATCTCTTCCTCCAC	x	x	x	TLR07-301	AAACTCCAGAAGGCAAGAAACTT	x	x	x
TLR07-330	CATGACATTGAGAAGAACTGCAT		x	x	TLR07-311	AATGTTTCTCTCTTGGGTCTTCC	x	x	x
TLR07-220GW	CCTCTATTTCTGGGATGTGTGG		x		TLR07-327	TATTTTTAGTACAGACGGGG	x	x	x
TLR07-350	CTACAAGATGCCTTCCAGTTGCGATA	x	x	x	TLR07-351	GCAAAGAAGGGCTAGACCGTTTC	x	x	x
TLR07-360	CAACAAACCCGCAAGC	x	x	x					

Amplification pairs	Tm
TLR07-270 / TLR07-271	54
TLR07-280 / TLR07-281	65
TLR07-290 / TLR07-291	65
TLR07-300 / TLR07-301	53
TLR07-310 / TLR07-311	64
TLR07-350 / TLR07-351	63
TLR07-360 / TLR07-327	57

TLR8

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR08-170	CAAAATTCATGGGGGTCATCTT	x	x	x	TLR08-171	TTCTTGCATGACTGGAAAGTGG	x	x	x
TLR08-180	CAGATTTGCTCAAGTCCCAGTG	x	x	x	TLR08-181	GCAGTATTTGCCAGGTGTTTTG	x	x	x
TLR08-190	CACGCAGCCCATCTTCAACTTC	x	x	x	TLR08-191	TCTGGGTGTTGCTCAGAAAAAG	x	x	x
TLR08-200	TTTGGCCTGGAACTGCTATTTT	x	x	x	TLR08-201	AAGGCTTTTCCATAAGCAGCAC	x	x	x
TLR08-210	CTGCCCCGCTTAGAAATACTTG	x	x	x	TLR08-211	AGGAATGCTTCATTTGGGATGT	x	x	x
TLR08-220	GGAAAGCAAGTCCCTGGTAGAA	x	x	x	TLR08-221	ACGTTTTTGTCTCGGCTCTCTT	x	x	x
TLR08-230	GGTCTCTTTCCACATCCCAAAC	x	x	x	TLR08-231	CTTCAGCTGTTTTCCCTGGACT	x	x	x

Amplification pairs	Tm
TLR08-170 / TLR08-171	65
TLR08-180 / TLR08-181	63
TLR08-190 / TLR08-191	65
TLR08-200 / TLR08-201	64
TLR08-210 / TLR08-211	65
TLR08-220 / TLR08-221	63
TLR08-230 / TLR08-231	65

TLR9

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR09-F1GW	CAAGGGCAGAAAAGGACAAG	x	x		TLR09-R1GW	AAGAGGCCCCTACTCTCTGG	x	x	
TLR09-082	CTGTTCTGACCCATAAGGCA		x	x	TLR09-R3GW	AGCTCCATCAGTGCCTCAGT		x	
TLR09-090	CTGGTTCTGAAGCCTAATTC		x	x	TLR09-R4GW	TGGGGAACCTGAACTCAGTC		x	
TLR09-100	CATGTCACATGACCATCGAG		x	x	TLR09-R5GW	CAGGTGGGCAAAGTCAGAAT		x	
TLR09-110	CACCTGAGCCGTCTTGA		x	x	TLR09-R6GW	GATGTAGCTGGGGGAAGTGA		x	
TLR09-120	CCTTGGATCTGTCACGGAACA		x	x	TLR09-R7GW	AGACGCAGAGTCTGGAGCAT		x	
TLR09-130	AAACTGGAAGTCCTCGACCTG		x	x	TLR09-R8GW	CTTGGCTGTGGATGTTGTTG		x	
TLR09-142	GACTGGGTGTACAACGAGCTT		x	x	TLR09-R9GW	CCTCCAGCAGGAAGTCCATA		x	
					TLR09-R10GW	CCCACAGGTTCTCAAAGAGG		x	

Amplification pairs	Tm
TLR09-F1GW / TLR09-R1GW	66

TLR10

Primer name	Forward Sequence 5'-3'	PCR	Seq.	IID	Primer name	Reverse Sequence 5'-3'	PCR	Seq.	IID
TLR10-400P	ACTTGCTGTTCTCTGTCTCTCCA	x	x	x	TLR10-401p	ATCTAAATGTCTGAAACCTGCCA	x	x	x
TLR10-504P	CTTCCCTCACAGAGTATGGTGAT	x	x	x	TLR10-503p	AAGTCTGCGGGAACCTTTCTTAG	x	x	x
TLR10-600P	TTTGAGCTCATCTTCATCATTCG	x	x	x	TLR10-601p	TCAAGATAAGCCTTACCACCAAA	x	x	x
TLR10-700P	CGTGATGGAATCAAGACTTCAAA	x	x	x	TLR10-701p	GATTTCTTCCCGCATTTAGAGTC	x	x	x
TLR10-800P	CTGATGGCCTTACGAGAACTAAA	x	x	x	TLR10-801p	TCCAGGAGAGCTTTCAGTTTATG	x	x	x
TLR10-900P	TATGAAAGCTACTTTGACCCTGG	x	x	x	TLR10-901p	AGATRGAGTCTTGCTCTGTCAC	x	x	x

Amplification pairs	Tm
TLR10-400P / TLR10-401p	64
TLR10-504P / TLR10-503p	62
TLR10-600P / TLR10-601p	64
TLR10-700P / TLR10-701p	64
TLR10-800P / TLR10-801p	65
TLR10-900P / TLR10-901p	64

Table S3. Table of polymorphism TLR1 chimpanzees

TLR1 CHIMPS	N	N	N	N	N	N	N	R	S	R	R	R	R	R	N	N	N	N
	112	197	313	524	1116	1513	2489	3201	3674	3813	3990	4452	5020	5044	5166	5204	5225	5261
consensus	G	G	C	A	T	C	C	A	G	C	A	G	A	C	G	G	G	A
CH09-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH29-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH29-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH85-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH87-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH88-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH88-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH94-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH96-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH96-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH97-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH106-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH107-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH108-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH108-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH110-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH112-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH123-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH98-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH109-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH110-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH112-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH123-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH09-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH85-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH106-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH95-B	A	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH87-B	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH95-A	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.	.
CH97-B	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH107-B	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH103-B	.	.	T	.	.	.	.	.	.	A	.	A	.	.	A	A	.	.
CH109-A	.	.	T	.	.	.	.	.	.	A	.	A	.	.	A	A	.	.
CH94-B	.	.	.	.	A	.	G	G	A	.	T	.	G	G	.	.	C	.
CH98-B	.	.	.	.	A	.	G	G	A	.	T	.	G	G	.	.	C	.
CH102-A	.	.	.	.	A	.	G	G	A	.	T	.	G	G	.	.	C	.
CH102-B	.	.	.	.	A	.	G	G	A	.	T	.	G	G	.	.	C	.
CH103-A	.	.	.	.	A	.	G	G	A	.	T	.	G	G	.	.	C	.

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S4. Table of polymorphism TLR2 chimpanzees

TLR2 CHIMPS	N	N	N	N	N	N	N	S	R	S	S	R	R	S	S	N	N	N
	657	673	773	859	1126	1319	1429	1930	2153	2269	2797	2893	3237	3331	3844	4299	4431	4492
consensus	C	T	C	A	T	T	G	G	C	G	C	G	C	A	A	C	A	T
CH09-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH85-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH88-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH88-B	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH94-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH96-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH97-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH106-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH109-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH109-B	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH110-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH112-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH123-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH98-A	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	T	.
CH110-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH87-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH95-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH09-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH29-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH102-A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH102-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G
CH29-B	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.	G
CH85-B	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.
CH87-B	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.
CH103-A	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.
CH103-B	.	.	.	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.
CH123-B	.	.	T	.	G	.	A	.	.	.	.	.	.	.	.	.	.	.
CH96-B	.	.	T	.	G	.	A	A	.	.	.	.	.	.	.	.	.	.
CH106-B	T	.	.	G	.	C	.	A	T	A	.	.	.	G	G	.	T	.
CH94-B	T	.	.	G	.	C	.	A	T	A	.	.	.	G	G	.	T	.
CH108-B	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH97-B	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH107-B	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH112-B	.	A	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
CH98-B	.	A	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.
CH108-A	.	A	.	.	.	.	.	.	.	.	T	.	A	.	.	.	T	.
CH95-B	.	A	.	.	.	.	.	.	.	.	T	.	A	.	.	.	T	.
CH107-A	.	A	.	.	.	.	.	.	.	.	T	.	A	.	.	.	T	.

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S5. Table of polymorphism TLR3 chimpanzees

TLR3 CHIMPS	R	N	S	N	N	R	N
	287	567	2452	2729	2771	6181	8813
consensus	A	G	G	A	G	A	T
CH09-A	.	.	.	.	.	.	.
CH09-B	.	.	.	.	.	.	.
CH29-A	.	.	.	.	.	.	.
CH87-A	.	.	.	.	.	.	.
CH94-A	.	.	.	.	.	.	.
CH94-B	.	.	.	.	.	.	.
CH95-A	.	.	.	.	.	.	.
CH95-B	.	.	.	.	.	.	.
CH96-A	.	.	.	.	.	.	.
CH98-A	.	.	.	.	.	.	.
CH98-B	.	.	.	.	.	.	.
CH102-A	.	.	.	.	.	.	.
CH102-B	.	.	.	.	.	.	.
CH103-A	.	.	.	.	.	.	.
CH107-A	.	.	.	.	.	.	.
CH109-A	.	.	.	.	.	.	.
CH109-B	.	.	.	.	.	.	.
CH110-A	.	.	.	.	.	.	.
CH110-B	.	.	.	.	.	.	.
CH112-A	.	.	.	.	.	.	.
CH112-B	.	.	.	.	.	.	.
CH123-A	.	.	.	.	.	.	.
CH123-B	.	.	.	.	.	.	.
CH85-A	.	.	.	.	.	.	.
CH88-A	.	.	.	.	.	.	.
CH106-A	.	.	.	.	.	.	.
CH108-A	.	.	.	.	.	.	.
CH87-B	.	T	.	.	.	.	.
CH29-B	G	.	.	C	A	.	.
CH106-B	G	.	.	C	A	.	.
CH107-B	G	.	.	C	A	.	.
CH96-B	G	.	.	C	A	.	.
CH103-B	.	.	.	C	A	.	.
CH108-B	.	.	.	C	A	.	.
CH85-B	.	.	.	C	A	T	.
CH97-A	.	.	.	C	A	T	.
CH88-B	.	.	.	C	A	T	G
CH97-B	.	.	A	C	A	T	G

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S6. Table of polymorphism TLR4 chimpanzees

TLR4 CHIMPS	N	N	N	N	N	N	N	N	R	S	N
	138	3955	4153	4163	4174	4338	4715	4743	8678	8847	11177
consensus	A	G	C	C	A	C	T	C	G	T	C
CH88-B	.	.	.	.	.	.	.	.	.	.	.
CH85-B	.	.	.	.	.	.	.	.	.	.	.
CH107-B	.	.	.	.	.	.	.	.	.	.	.
CH108-B	.	.	.	.	.	.	.	.	.	.	.
CH110-A	.	.	.	.	.	.	.	.	.	.	.
CH112-A	.	.	.	.	.	.	.	.	.	.	.
CH112-B	.	.	.	.	.	.	.	.	.	.	.
CH123-A	.	.	.	.	.	.	.	.	.	.	.
CH103-A	.	N	.	.	.	.	.	.	.	.	.
CH103-B	.	N	.	.	.	.	.	.	.	.	.
CH109-A	C	.	.	.	.	.	.	.	.	.	.
CH109-B	C	.	.	.	.	.	.	.	.	.	.
CH110-B	C	.	.	.	.	.	.	.	.	.	.
CH123-B	C	.	.	.	.	.	.	.	.	.	.
CH98-A	C	.	.	.	.	.	.	.	.	.	.
CH95-B	C	.	T	.	.	.	.	.	.	.	.
CH96-B	C	.	T	.	.	.	.	.	.	.	.
CH97-B	C	.	.	.	.	.	.	.	T	.	.
CH98-B	C	.	.	.	.	.	.	.	T	.	.
CH09-A	.	.	.	.	G	T	.	G	.	.	.
CH09-B	.	.	.	.	G	T	.	G	.	.	.
CH29-A	.	.	.	.	G	T	.	G	.	.	.
CH85-A	.	.	.	.	G	T	.	G	.	.	.
CH87-A	.	.	.	.	G	T	.	G	.	.	.
CH87-B	.	.	.	.	G	T	.	G	.	.	.
CH88-A	.	.	.	.	G	T	.	G	.	.	.
CH94-A	.	.	.	.	G	T	.	G	.	.	.
CH95-A	.	.	.	.	G	T	.	G	.	.	.
CH96-A	.	.	.	.	G	T	.	G	.	.	.
CH97-A	.	.	.	.	G	T	.	G	.	.	.
CH102-A	.	.	.	.	G	T	.	G	.	.	.
CH102-B	.	.	.	.	G	T	.	G	.	.	.
CH106-A	.	.	.	.	G	T	.	G	.	.	.
CH107-A	.	.	.	.	G	T	.	G	.	.	.
CH108-A	.	.	.	.	G	T	.	G	.	.	.
CH29-B	C	.	.	.	G	T	.	G	.	.	.
CH106-B	C	.	.	.	G	T	.	G	.	C	T
CH94-B	C	A	.	T	.	.	C	.	.	.	.

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S7. Table of polymorphism TLR5 chimpanzees

TLR5 CHIMPS	N	N	N	N	N	R	S	R	R
	203	480	492	562	1035	1724	1744	1954	3536
consensus	C	G	T	A	T	C	A	A	C
CH09-B	.	.	.	.	.	.	.	.	.
CH29-A	.	.	.	.	.	.	.	.	.
CH94-A	.	.	.	.	.	.	.	.	.
CH94-B	.	.	.	.	.	.	.	.	.
CH95-A	.	.	.	.	.	.	.	.	.
CH95-B	.	.	.	.	.	.	.	.	.
CH96-A	.	.	.	.	.	.	.	.	.
CH96-B	.	.	.	.	.	.	.	.	.
CH97-A	.	.	.	.	.	.	.	.	.
CH98-A	.	.	.	.	.	.	.	.	.
CH98-B	.	.	.	.	.	.	.	.	.
CH102-A	.	.	.	.	.	.	.	.	.
CH103-A	.	.	.	.	.	.	.	.	.
CH103-B	.	.	.	.	.	.	.	.	.
CH106-A	.	.	.	.	.	.	.	.	.
CH106-B	.	.	.	.	.	.	.	.	.
CH107-A	.	.	.	.	.	.	.	.	.
CH107-B	.	.	.	.	.	.	.	.	.
CH108-A	.	.	.	.	.	.	.	.	.
CH108-B	.	.	.	.	.	.	.	.	.
CH110-A	.	.	.	.	.	.	.	.	.
CH112-A	.	.	.	.	.	.	.	.	.
CH123-A	.	.	.	.	.	.	.	.	.
CH123-B	.	.	.	.	.	.	.	.	.
CH87-A	.	.	.	.	.	.	.	.	.
CH88-A	.	.	.	.	.	.	.	.	.
CH09-A	.	.	.	.	.	.	.	.	T
CH109-B	T	.	.	.	.	.	.	.	.
CH112-B	T	.	.	.	.	.	.	.	.
CH85-B	T	.	.	.	.	.	.	.	.
CH85-A	.	A	C	G	C	T	C	.	.
CH97-B	.	A	C	G	C	T	C	.	.
CH109-A	.	A	C	G	C	T	C	.	.
CH87-B	.	A	C	G	C	T	C	.	.
CH88-B	.	A	C	G	C	T	C	G	.
CH29-B	.	A	C	G	C	T	C	G	.
CH102-B	.	A	C	G	C	T	C	G	.
CH110-B	.	A	C	G	C	T	C	G	.

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S8. Table of polymorphism TLR6 chimpanzees

TLR6 CHIMPS	N	N	N	N	N	S	R	S	R	R
	17	349	401	848	975	1541	2493	2609	2618	2785
consensus	T	G	T	C	T	T	A	A	C	C
CH09-A	.	.	.	.	.	.	.	.	.	.
CH87-A	.	.	.	.	.	.	.	.	.	.
CH96-A	.	.	.	.	.	.	.	.	.	.
CH98-A	.	.	.	.	.	.	.	.	.	.
CH106-A	.	.	.	.	.	.	.	.	.	.
CH106-B	.	.	.	.	.	.	.	.	.	.
CH107-A	.	.	.	.	.	.	.	.	.	.
CH108-A	.	.	.	.	.	.	.	.	.	.
CH123-A	.	.	.	.	.	.	.	.	.	.
CH112-B	C	.	.	.	.	.	.	.	.	.
CH123-B	C	.	.	.	.	.	.	.	.	.
CH107-B	C	.	.	.	.	.	.	.	.	.
CH97-B	C	.	.	.	.	.	.	.	.	.
CH09-B	C	.	.	.	.	.	.	.	.	.
CH29-A	.	A	.	.	.	.	.	.	.	.
CH29-B	.	A	.	.	.	.	.	.	.	.
CH85-B	.	A	.	.	.	.	.	.	.	.
CH88-A	.	A	.	.	.	.	.	.	.	.
CH88-B	.	A	.	.	.	.	.	.	.	.
CH95-B	.	A	.	.	.	.	.	.	.	.
CH96-B	.	A	.	.	.	.	.	.	.	.
CH97-A	.	A	.	.	.	.	.	.	.	.
CH110-B	.	A	.	.	.	.	.	.	.	.
CH112-A	.	A	.	.	.	.	.	.	.	.
CH94-B	.	A	.	.	.	.	.	.	.	.
CH87-B	.	.	.	.	C	.	.	.	.	.
CH95-A	.	.	.	.	.	.	G	.	.	.
CH110-A	.	.	.	.	.	.	G	.	.	.
CH85-A	.	.	A	.	.	.	G	.	.	.
CH94-A	.	.	.	.	.	A	G	.	.	.
CH98-B	.	.	.	.	.	A	G	.	.	.
CH102-A	.	.	.	.	.	A	G	.	.	.
CH102-B	.	.	.	.	.	A	G	.	.	.
CH109-A	.	.	.	.	.	A	G	.	.	.
CH103-B	.	.	.	.	.	A	G	.	.	.
CH103-A	.	.	.	.	.	A	.	.	.	.
CH109-B	C	.	.	.	.	.	G	.	.	.
CH108-B	.	.	.	G	.	A	G	G	A	G

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S9. Table of polymorphism TLR7 chimpanzees

TLR7 CHIMPS	N	N	N	N	R
	190	246	361	406	2408
consensus	A	A	G	T	G
CH29-A	.	.	.	.	.
CH94-A	.	.	.	.	.
CH123-A	.	.	.	.	.
CH09-A	.	G	.	.	.
CH85-A	.	G	.	.	.
CH88-A	.	G	.	.	.
CH97-A	.	G	.	.	.
CH107-A	.	G	.	.	.
CH95-A	.	G	.	.	C
CH87-A	G	.	.	.	.
CH96-A	G	.	.	.	.
CH98-A	G	.	.	.	.
CH108-A	G	.	.	.	.
CH110-A	G	.	A	.	.
CH88-B	.	.	.	G	.
CH102-A	.	.	.	G	.
CH103-A	.	.	.	G	.
CH109-A	.	.	.	G	.
CH112-A	.	.	.	G	.
CH106-A	.	.	.	G	C

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S10. Table of polymorphism TLR8 chimpanzees

TLR8 CHIMPS	N	N	N	S	S	S	S	S	N
	204	686	1207	1576	1732	3007	3214	4288	4855
consensus	C	G	C	G	G	C	C	C	T
CH09-A	.	.	.	.	.	.	.	.	.
CH85-A	.	.	.	.	.	.	.	.	.
CH88-A	.	.	.	.	.	.	.	.	.
CH95-A	.	.	.	.	.	.	.	.	.
CH97-A	.	.	.	.	.	.	.	.	.
CH110-A	.	.	.	.	.	.	.	.	.
CH29-A	G	.	.	.	.	.	.	.	.
CH94-A	.	.	.	.	.	.	.	.	G
CH88-B	.	.	.	.	.	.	A	.	.
CH123-A	.	.	.	.	.	.	A	.	.
CH102-A	.	A	.	.	.	.	.	.	.
CH103-A	.	A	.	.	.	.	.	.	.
CH107-A	.	A	.	.	.	.	.	.	.
CH112-A	.	A	.	.	.	.	.	.	.
CH106-A	.	.	.	A	.	.	.	.	.
CH109-A	.	.	.	A	.	T	.	.	.
CH87-A	.	.	T	.	A	.	.	T	.
CH96-A	.	.	T	.	A	.	.	T	.
CH98-A	.	.	T	.	A	.	.	T	.
CH108-A	.	.	T	.	A	.	.	T	.

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S11. Table of polymorphism TLR9 chimpanzees

TLR9 CHIMPS	N	N	N	N	N	R	S	R	R	R	S
	176	246	470	719	831	1522	1734	2092	3022	3248	3381
consensus	G	G	G	C	G	A	C	G	C	C	G
CH85-A	.	.	.	.	.	.	.	.	.	.	.
CH88-A	.	.	.	.	.	.	.	.	.	.	.
CH102-A	.	.	.	.	.	.	.	.	.	.	.
CH106-A	.	.	.	.	.	.	.	.	.	.	.
CH107-A	.	.	.	.	.	.	.	.	.	.	.
CH98-B	.	A	.	.	.	.	.	.	.	.	.
CH29-B	.	A	.	.	.	.	.	.	.	.	.
CH95-A	.	A	.	.	.	.	.	.	.	.	.
CH107-B	.	A	.	.	.	.	.	.	.	.	.
CH109-A	.	A	.	.	.	.	.	.	.	.	.
CH110-B	.	A	.	.	.	.	.	.	.	.	.
CH87-B	.	A	.	.	.	.	.	.	.	.	.
CH94-B	.	A	.	.	.	.	.	.	.	.	.
CH112-B	.	A	.	.	.	.	.	.	.	.	.
CH108-A	.	A	.	.	.	.	.	A	.	.	.
CH96-B	.	A	.	.	.	.	.	A	.	.	.
CH88-B	.	A	.	.	.	.	.	A	.	.	.
CH123-B	.	A	.	.	.	.	.	A	.	.	.
CH96-A	.	.	.	.	.	.	.	.	.	T	.
CH85-B	.	.	.	.	.	.	.	.	.	T	.
CH87-A	.	.	.	.	.	G	.	.	.	.	.
CH98-A	.	.	.	.	.	G	.	.	.	T	.
CH106-B	.	.	.	.	.	G	.	.	.	T	.
CH29-A	.	.	.	.	.	G	.	.	.	T	.
CH102-B	.	.	T	.	.	G	.	.	.	T	.
CH108-B	A	.	.	.	.	.	.	.	.	T	.
CH109-B	A	.	.	.	.	.	.	.	T	.	.
CH95-B	.	A	.	.	A	.	.	.	.	T	.
CH09-A	.	.	.	T	.	.	T	.	.	.	A
CH09-B	.	.	.	T	.	.	T	.	.	.	A
CH110-A	.	.	.	T	.	.	T	.	.	.	A
CH112-A	.	.	.	T	.	.	T	.	.	.	A
CH123-A	.	.	.	T	.	.	T	.	.	.	A
CH94-A	.	.	.	T	.	.	T	.	.	.	A

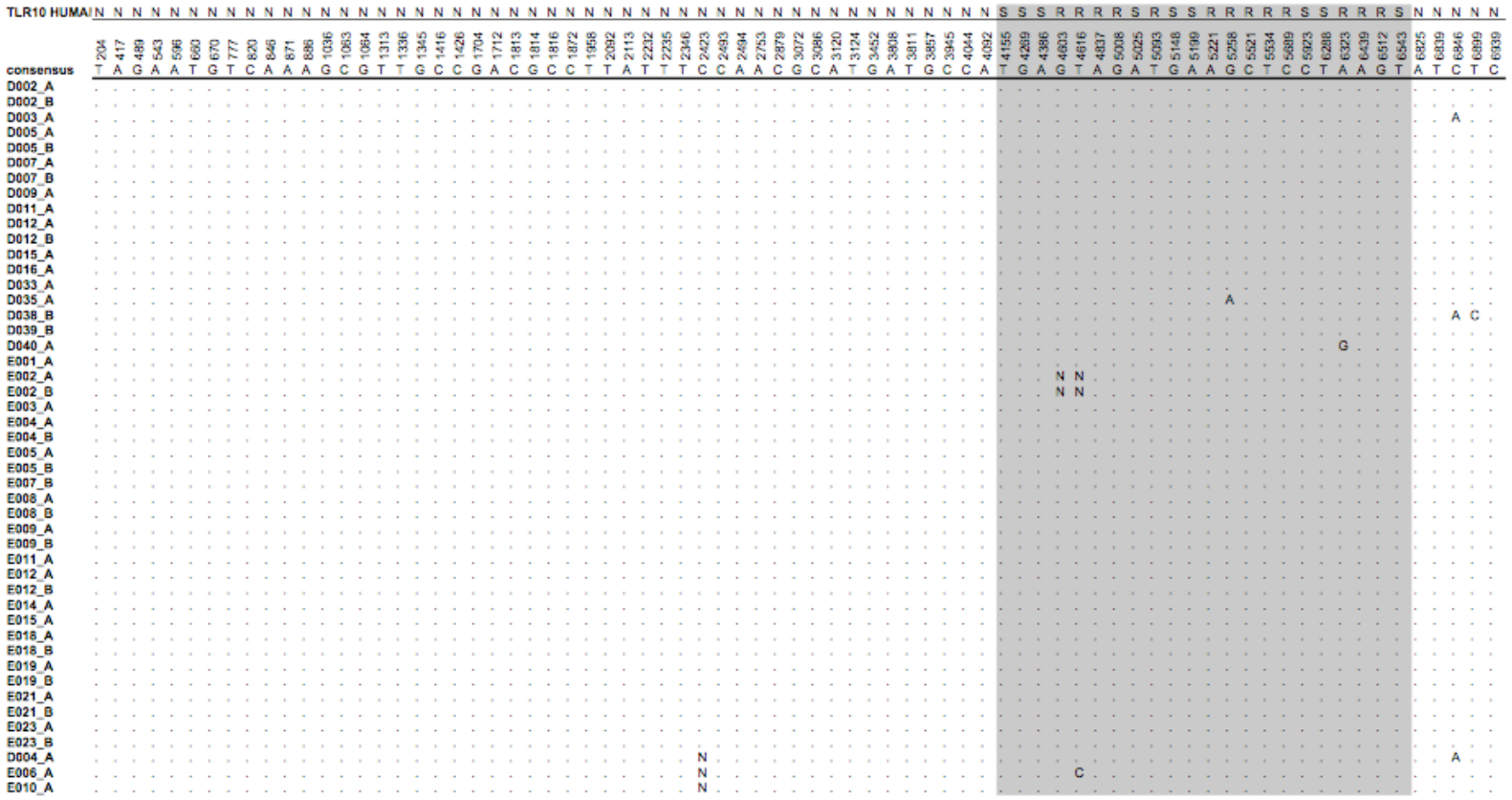
Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

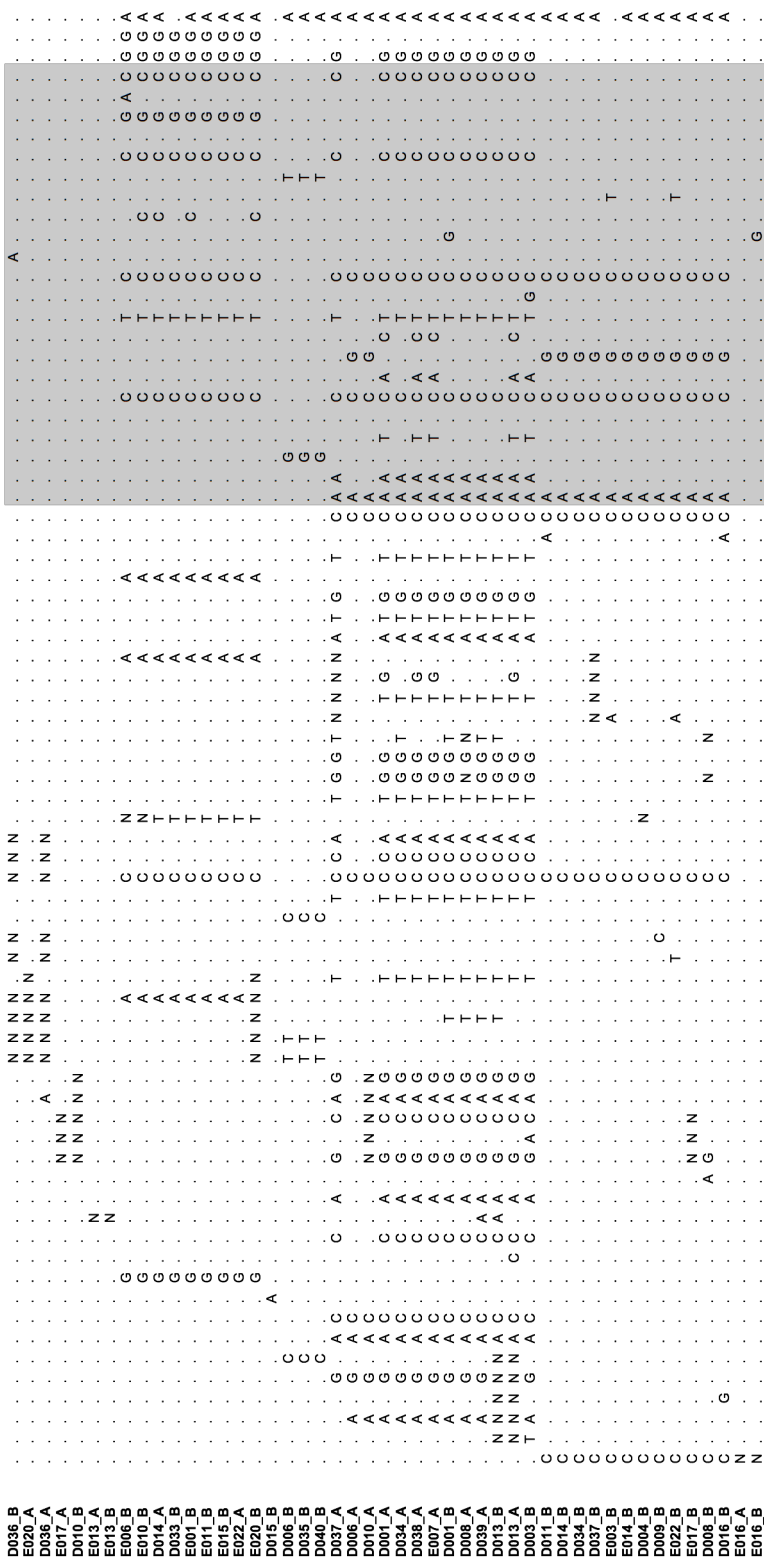
Table S12. Table of polymorphism TLR10 chimpanzees

TLR10 CHIMPS	N	N	N	N	N	R	S	R	N
	506	687	1050	1194	1349	2380	3023	3276	4026
	C	C	A	G	C	G	A	A	C
CH09-A	.	.	.	.	.	.	.	.	.
CH85-A	.	.	.	.	.	.	.	.	.
CH87-A	.	.	.	.	.	.	.	.	.
CH94-A	.	.	.	.	.	.	.	.	.
CH96-A	.	.	.	.	.	.	.	.	.
CH98-A	.	.	.	.	.	.	.	.	.
CH98-B	.	.	.	.	.	.	.	.	.
CH102-A	.	.	.	.	.	.	.	.	.
CH102-B	.	.	.	.	.	.	.	.	.
CH107-A	.	.	.	.	.	.	.	.	.
CH109-A	.	.	.	.	.	.	.	.	.
CH123-A	.	.	.	.	.	.	.	.	.
CH108-A	.	.	.	.	.	.	.	.	.
CH110-A	N	.	.	.	.	.	.	.	.
CH09-B	.	.	.	.	.	A	.	.	.
CH95-B	.	.	.	.	.	A	.	.	.
CH112-B	.	.	.	.	.	A	.	.	.
CH123-B	.	.	.	.	.	A	.	.	.
CH106-B	.	.	.	.	.	A	.	.	.
CH29-A	.	.	.	.	.	.	G	.	.
CH29-B	.	.	.	.	.	.	G	.	.
CH112-A	.	.	.	.	.	.	G	.	.
CH85-B	.	.	.	.	.	.	G	.	.
CH88-A	.	.	.	.	.	.	G	.	.
CH88-B	.	.	.	.	.	.	G	.	.
CH94-B	.	.	.	.	.	.	G	.	.
CH95-A	.	.	.	.	.	.	G	.	.
CH96-B	.	.	.	.	.	.	G	.	.
CH97-A	.	.	.	.	.	.	G	.	.
CH106-A	.	.	.	.	.	.	G	.	.
CH110-B	N	.	.	.	.	.	G	.	.
CH109-B	.	.	T	.	.	.	.	C	.
CH87-B	T	T	T	.	.	.	.	.	G
CH97-B	T	T	.	A	T	.	.	.	G
CH108-B	T	T	.	A	T	.	.	.	G
CH107-B	T	T	.	A	T	.	.	.	.

Shaded areas indicate coding positions. N=noncoding, R=replacement, S=silent.

Table S13. Table of polymorphism TLR10 humans




D036_B
E020_A
D036_A
E017_A
D010_B
E013_A
E013_B
E006_B
E010_B
D014_A
D033_B
E001_B
E011_B
E015_B
E022_A
E020_B
D015_B
D006_B
D035_B
D040_B
D037_A
D006_A
D010_A
D001_A
D034_A
D038_A
E007_A
D001_B
D008_A
D039_A
D013_B
D013_A
D003_B
D011_B
D014_B
D034_B
D037_B
E003_B
E014_B
D004_B
D009_B
E022_B
E017_B
D008_B
D016_B
E016_A
E016_B

APPENDIX D: PROMISCUITY, SOCIALITY AND THE RATE OF MOLECULAR EVOLUTION AT PRIMATE IMMUNITY GENES.

ABSTRACT

Recently, a positive correlation between basal leukocyte counts and mating system across primates suggested that sexual promiscuity could be an important determinant of the evolution of the immune system. Motivated by this idea, we examined the patterns of molecular evolution of 15 immune defense genes in primates in relation to promiscuity and other variables expected to affect disease risk (group size, density, diet and habitat). We obtained maximum likelihood estimates of the rate of protein evolution of terminal branches of the primate phylogeny at these genes. Using phylogenetically independent contrasts, we found that immunity genes evolve faster in more promiscuous species, but only for a subset of genes that interact closely with pathogens. We also observed a significantly greater proportion of branches under selection in the more promiscuous species. Analyses of independent contrasts also showed a positive effect of group size. However, the effect of group size was not restricted to genes that interact closely with pathogens, and no differences were observed in the proportion of branches under selection in species with small and large groups. Together, these results suggest that mating system has influenced the evolution of some immunity genes in primates, possibly due to the increased risk of acquiring sexually transmitted diseases in species with higher levels of female promiscuity.

INTRODUCTION

In recent years, a large body of work has been devoted to understanding the intricate relationship between immunity and reproduction (Schmid-Hempel 2003; Lawniczak et al. 2007). It has become increasingly clear that there are trade-offs between immune and reproductive functions due to the costly nature of both systems (Sheldon and Verhulst 1996; Lochmiller and Deerenberg 2000; Zuk and Stoehr 2002). For example, in flycatchers, infection rates (measured by serological parameters) increase when brood size is experimentally increased, and parasitized females lay smaller clutches than non-parasitized counterparts (Gustafsson et al. 1994). A more direct link between immunity and reproduction has been provided by the discovery that immune molecules are commonly expressed in reproductive tissues of vertebrates (Li et al. 2001; Com et al. 2003; Silphaduang et al. 2006) and invertebrates (Lung, Kuo, and Wolfner 2001). Moreover, there is now evidence that the female immune system can be modulated in response to mating. In *Drosophila*, for example, a large suite of immune related genes change their expression profiles following mating (Lawniczak and Begun 2004; McGraw et al. 2004).

Another connection between reproduction and immunity comes from the work of Hamilton and Zuk (1982), who first suggested a role for parasites in the context of sexual selection. At the precopulatory level, mate choice could be based on secondary sexual traits that indirectly reflect heritable variation in immune condition (Hamilton and Zuk 1982). At the postcopulatory level, it is possible that male ejaculates interfere with female immunity leading to sexual conflict (Fedorka and Zuk 2005), or that antimicrobial

peptides that inhibit sperm motility (Reddy, Yedery, and Aranha 2004) mediate cryptic female choice (Lawniczak et al. 2007).

Disease transmission during mating provides another connection between reproduction and immunity. Sexually transmitted diseases (STDs) are ubiquitous among animals and differ from infectious diseases in a number of important characteristics (Lockhart, Thrall, and Antonovics 1996). STDs affect fitness mostly by a negative effect on sterility rather than by inducing mortality. STDs also persist longer in their hosts and do not generally exhibit cyclic fluctuations compared with other infectious diseases. For these reasons, it is possible that STDs might impose different selective pressures on their hosts than other types of diseases (Lockhart, Thrall, and Antonovics 1996). Mating system might affect the evolution of immunity because species with higher levels of sexual promiscuity might experience increased risk of STDs. Alternatively mating behavior itself might evolve as a consequence of STDs (Immerman 1986; Loehle 1995; Thrall, Antonovics, and Bever 1997).

Using the baseline number of leucocytes (a common indicator of immunocompetence), Nunn and others (Nunn, Gittleman, and Antonovics 2000; Nunn 2002b; Anderson, Hessel, and Dixson 2004) found a positive correlation between levels of white blood cells and several proxies of female sexual promiscuity among species of primates with different mating systems. The lack of associations with several other social, ecological and life history variables led to the hypothesis that increased levels of transmission of STDs in promiscuous species have resulted in the evolution of a greater investment in immune function (Nunn, Gittleman, and Antonovics 2000). An alternative

interpretation for a positive correlation between female promiscuity and the strength of the immune system is based on antagonistic coevolution between male ejaculate and female immunity. Support for this idea comes from studies in crickets of the genus *Allonemobious*, in which multiple mating with diverse males results in suppression of female immunity (Fedorka and Zuk 2005). If such immunodepression is costly to the female, over evolutionary time, coevolutionary processes caused by conflicting male and female interests might result in the evolution of stronger immunity.

Here we sought to extend the disease risk/promiscuity hypothesis (Nunn, Gittleman, and Antonovics 2000) to the molecular level, by exploring the relationship between sexual promiscuity and the evolution of immunity genes in primates. If the increase in leukocyte levels in promiscuous species reported by Nunn and others truly reflects differences in disease risk among species, we might expect that natural selection will have shaped other aspects of the immune system in a similar way. In particular, natural selection on genes involved in immunity might be stronger or more frequent in species in which females routinely mate with multiple males than in species in which females mate with one male. We thus predict an acceleration of the rate of molecular evolution at immunity genes (particularly those that participate directly in host-pathogen interactions) in more promiscuous species. Primates constitute a good study-system for testing this hypothesis for several reasons. First, the original observations made by Nunn and colleagues were done on primates. Second, primates exhibit a diversity of social and mating systems, and several socio-ecological variables are available. Some of these variables are also expected to affect disease risk. Thus, we take advantage of this

information and also explore the effect of group size, density, diet and habitat on the rate of molecular evolution of immunity genes.

Using phylogenetically independent comparisons for a set of 15 genes related to immune defense in primates we found that both female promiscuity and group size show a weak but significant positive effect on the rate of protein evolution. The effect of mating system (female promiscuity) was stronger for a subset of genes that interact directly with pathogens, and this seems to be driven by positive selection. Mating system and group size, however, explain only a small fraction of the variation in the rate of protein evolution, emphasizing that factors related to the biology of particular species play a major role in the evolution of immune defense genes.

MATERIALS AND METHODS

Samples: DNA samples from 14 primate species were obtained from the following sources: *Cercopithecus mona, Theropithecus gelada* and *Mandrillus sphinx* from William Switzer; *Papio anubis, Callithrix jacchus* and *Macaca fascicularis* from the Southwest National Primate Research Center; *Chiropotes satanas* and *Saguinus midas* from Smithsonian Instituion; *Ateles geoffroyi, Allenopithecus nigrovirdis, Pithecia pithecia, Cercocebus agilis, Symphalangus syndactylus* and *Colobus guereza* from Coriell Cell Repositories.

Molecular data: We gathered published sequence data on 14 genes related to immune defense in several primate species. We made an effort to include genes for which there was previous evidence of positive selection in the patterns of protein evolution (Filip and Mundy 2004; Sawyer, Emerman, and Malik 2004; Sawyer et al. 2005; OhAinle et al. 2006; Zelezetsky et al. 2006; Osorio, Antunes, and Ramos 2007; Sawyer, Emerman, and Malik 2007; Kerns, Emerman, and Malik 2008; Zhang, Weinstock, and Gerstein 2008; Elde et al. 2009; Wlasiuk et al. 2009). By looking at a set of genes that in most cases have a recognized history of positive selection we sought to maximize the chances of uncovering a positive relationship between promiscuity and molecular evolution, if such a relationship exists. The sample includes three pattern recognition receptors (TLR1, TLR4, TLR5), one antimicrobial peptide (CAMP), other innate immunity genes (PKR, CCR5), a series of intrinsic immunity genes with antiviral function (APOBEC3G, APOBEC3H, TRIM5, TRIM22, ZAP), two adaptive immunity genes (CD4, CD45), and a gene with putative immune function (ANG). Some of these genes (CAMP,

APOBEC3G, APOBEC3H, PKR, TLR1, TLR4, TLR5, TRIM5, TRIM22, ZAP) interact directly with pathogens, while others do not. We will refer to these two classes as 'pathogen-interacting genes' and 'non pathogen-interacting genes' respectively. An average of 16 species (9-29) was used per gene. The complete list of species and accession numbers is presented in Supplementary Table 1a and 1b.

Additionally, we sequenced the DUFFY (DARC) gene, because of its recently reported association with differences in white-blood cell counts within and between human populations (Nalls et al. 2008; Reich et al. 2009). A fragment of ~1800 bp containing the entire coding region of the DUFFY gene was sequenced in the 14 species mentioned above. Together with eight additional sequences from GenBank (Supplementary Table 1a), the complete dataset for DUFFY consists of 22 primate species. PCR was performed in 50 μl reactions using Platinum *Taq* High Fidelity DNA Polymerase (Invitrogen, San Diego, CA), with primers F1-CTTTCTGGTCCCCACCTTTT and R1-TAAGAAACCACCCGCYTCAC. 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), using the following sequencing primers: F2-TAGTCCCRACCAGYCAAATC, F5-ATCGGCTTCCCCAGGA and R2-CGCTTCACAAARGCAKTGTA. Sequences were deposited in GenBank under the accession numbers: GU219517-30. Sequence editing and assembly were performed using SEQUENCHER (Gene Codes, Ann Arbor, MI).

Hypotheses and predictions for other socio-ecological variables: Many host traits and ecological factors have been proposed to influence disease risk in primates

(reviewed in Nunn and Altizer, 2006), and these hypotheses generate testable predictions. Aside from sexual promiscuity, we investigated the effect of group size, density, diet and habitat. Disease risk is expected to increase with group size and density, because more contacts among individuals should promote transmission of infectious diseases (Altizer et al. 2003). Disease risk is also expected to be higher in species that consume leaves (because folivorous primates consume larger volumes of food, and potentially more parasites) (Moore 2002) or insect prey (because insects can be intermediate hosts for trophically transmitted diseases) (Dunn 1968) than in frugivorous primates. Finally, disease risk is expected to be higher in terrestrial primates than in arboreal primates, because terrestrial species should be exposed to fecal contamination more than arboreal species (Nunn, Gittleman, and Antonovics 2000).

Primate variables: Data on mating system, testis size, group size, density, diet and habitat was obtained from several published compilations (Kenagy and Trombulak 1986; Harcourt 1991; Harcourt, Purvis, and Liles 1995; Rowe 1996; Lindenfors and Tullberg 1998; Nunn 2002b; Semple, Cowlishaw, and Bennett 2002; Nunn et al. 2003; Anderson, Hessel, and Dixson 2004). The values for these variables are presented in Supplementary Table 2.

Mating system was further categorized as unimale (UM-monogamous or polygynous) or multimale (MM-polyandrous or promiscuous). To deal with the problem of ambiguities in mating system we grouped mating system three ways. First, we assigned all ambiguous cases as UM (partition 1-MS1). Second, we assigned all ambiguous cases as MM (partition 2-MS2), and third, we excluded all species with

ambiguous mating system (partition 3-MS3). In primates, large testes are likely the result of selection for high sperm production due to sperm competition, in species in which females mate multiply (Harcourt 1991). Thus, the residuals of the regression of log testis size *vs* log body size (residual testis size-RTS) were also used as a proxy for female promiscuity.

Group size and density were log transformed to approach normality. For all the analyses described below, except for the linear regression models, RTS, group size, and density were transformed into discrete variables to capture the essence of the pattern of variation at these variables. For RTS, we coded as 0 the negative values and as 1 the positive values. For group size and density, taking an interval of one standard deviation centered on the mean, we coded as 1 all the values above this range, and as 0 all the values below it. Below we refer to the discrete categories of small group size (SG) and large group size (LG). Habitat and diet were treated as discrete variables with three states each. For habitat we used: strictly arboreal, terrestrial in wooded environments and terrestrial in open environments. For diet we used: insectivores, folivores and frugivores.

Phylogenetic reconstruction: Sequences were aligned in Revtrans (http://cbs.dtu.dk/services/RevTrans/) with manual adjustment of small indels. Sites with indels were removed from the alignments. Phylogenetic trees for each gene were constructed in PAUP (Swofford 1993) using parsimony, distance, and maximum likelihood methods (ML). In all cases there was overall general agreement between the accepted species tree (Purvis 1995) and the gene tree, with only a few branches in slightly different positions. Because the methods for detecting selection described below are

relatively robust to minor changes in the phylogeny and the inferences of trait evolution are based on the species relationships, for all the subsequent analyses we used the accepted species trees (Figure 1) (Purvis 1995).

Maximum likelihood estimate of evolutionary rates: We estimated dN/dS, the number of nonsynonymous substitutions per non-synonymous site (dN) divided by the number of synonymous substitutions per synonymous site (dS) in a maximum likelihood (ML) framework using Codeml, in the PAML ver 4.2 package (Yang 1997; Yang 2007). A dN/dS ratio >1 represents unambiguous evidence of positive selection while a value <1 indicates purifying or negative selection.

We ran a free-ratio model, in which dN/dS is estimated independently for each branch in the phylogeny (Nielsen and Yang 1998; Yang 1998). These estimates of the rate of protein evolution in terminal branches of the phylogeny were used to test for correlations with socio-ecological variables. For some genes, a few branches lacked synonymous substitutions, preventing the calculation of dN/dS. This is a common problem with short branches. However, in some of these cases, the number of nonsynonymous substitutions was high, arguing against low divergence. To minimize the amount of missing data in subsequent analyses that were based on these values, for the branches with dS=0 and more than two nonsynonymous substitutions, we calculated the dN/dS ratio assuming one synonymous substitution. To check for convergence, the free-ratio models were run twice, using initial ω values of 0.5 and 1.5. In all cases we used the F3x4 model of codon frequencies.

Statistical analyses: To minimize the problem of uncertainty in mating system reconstruction along long branches or uncertainty due to incomplete phylogenetic sampling, we restricted analyses to the terminal branches of the phylogeny. In all the analyses described below the rate of protein evolution, dN/dS, was treated as the dependent variable, while mating system, relative testis size, group size, density, diet and habitat were treated as independent variables. For the first analysis (independent contrasts; see below) we included all variables and, guided by the results, for additional analyses we used only female promiscuity and group size. An outline of this second set of analyses focusing exclusively on mating system and group size (with their specific predictions) is presented in Figure 2.

Although dN/dS was estimated independently for each branch of the phylogeny, mating system (or other traits) might be the same in closely related species due to shared ancestry, leading to non-independence. To take into account this potential problem, we first performed an analysis based on phylogenetically independent contrasts (Felsenstein 1985), in which variation in the set of independent variables was examined (separately) in relation to variation in dN/dS. Because branches of the phylogeny are used only once, these contrasts represent independent transitions in the predictor variables given a certain topology. An excess of positive contrasts (i.e. the independent variable and dN/dS vary in the direction predicted by the hypothesis) can be taken as evidence of correlated evolution. Phylogenetically independent contrasts were obtained using the BRUNCH algorithm implemented in the CAIC software (Purvis and Rambaut 1995). The number of contrasts per gene was generally low resulting in little statistical power. Thus, we

summed the number of positive contrasts across the 15 genes. To test for deviations from a null expectation of equal number of positive and negative contrasts, we performed sign tests on the number of positive contrasts.

Because of the inferred effect of sexual promiscuity and group size (but not other variables) from the previous analyses we focused on these two variables for all subsequent analyses. Some of these analyses do not explicitly correct for phylogenetic effects. However, because for each gene usually only one species per genus was included, we do not expect a high degree of phylogenetic correlation. First, the mean and variance in dN/dS of UM and MM species were compared with a t-test and a Z-test respectively. Similarly, means and variances were compared between SG and LG species. Second, we investigated multiple regression models including RTS or mating system and group size as predictor variables. Third, differences in the proportion of branches with dN/dS>1 between UM and MM, and between SG and LG species, were evaluated using a Z-test.

RESULTS

We used a combination of approaches to evaluate the effect of sexual promiscuity, group size, density, habitat and diet on the rate of molecular evolution of immunity genes. We began implementing free-ratio models, in which the dN/dS ratio can vary only among branches. The values of dN/dS of the terminal branches of the phylogeny (Figure 1) were used in the analyses described below. Panels 1-4 in Figure 2 summarize the results obtained in these analyses, which are presented in detail in Tables 1-4.

We first examined the direction of the change in dN/dS in relation to sexual promiscuity using phylogenetically independent contrasts. We used three mating system partitions and RTS as proxies for sexual promiscuity. When summed across genes, the number of positive contrasts in which an increase in promiscuity was accompanied by an increase in dN/dS showed a very slight trend in the predicted direction but was not significant (Sign test, MS2 $p=0.10$, MS3 $p=0.11$) (Table 1). When we separated the pathogen-interacting (PI) genes (APOBEC3G, APOBEC3H, CAMP, PKR, TLR1, TLR4, TLR5, TRIM5, TRIM22 and ZAP) from the rest (non PI genes), for the three mating system partitions the number of positive contrasts significantly exceeded the null expectation of 50% for the PI genes (Sign test $p=0.03$, 0.01 and 0.01 respectively). In contrast, none of the measures of sexual promiscuity deviated from the null expectation for the non-pathogen interacting genes (Table 1).

We repeated these analyses with group size, density, diet and habitat as independent variables. Only group size showed a significant or marginally significant excess of positive contrasts (Table 1). These results, summarized in the first panel of

Figure 2, suggest that both promiscuity and group size might influence the rate of evolution of immunity genes. Interestingly, genes that interact directly with pathogens showed a more pronounced effect of sexual promiscuity while all the genes showed a similar effect of group size.

Next, for each gene, we compared the mean and variance in dN/dS between UM and MM branches and between SG and LG branches. In 10 of the 15 genes (8/10 PI, 2/5 non-PI) we observed a higher mean dN/dS in MM branches than in UM branches in at least one of the three mating system partitions. Similarly, in 9 of the 15 genes (5/10 PI, 4/5 non-PI), LG branches had a higher mean dN/dS than SG branches. Only in a few cases, however, were these differences significant, with four genes showing a weak effect of promiscuity (t-test APOBEC3G p=0.06, CD45 p=0.07, PKR p=0.08, TLR4 p=0.09) and only one gene showing a significant effect of group size (t-test APOBEC3G p=0.04) (Table 2, Figure 2 panel 2)

The analyses based on independent contrasts (Figure 2 panel 1, Table 1) strongly suggest a link between promiscuity, group size and molecular evolution of immune genes. In these analyses however, all the variables were analyzed separately, precluding teasing apart potential correlations among them. In an attempt to disentangle the potentially confounding effects of sexual and social factors on dN/dS, for each gene we fit multiple regression models using dN/dS as the dependent variable and promiscuity (RTS or Mating system partition 3) and group size as independent factors. Table 3 shows the multiple regression models. In five of the 15 genes, variation in promiscuity (RTS or MS3), group size, or both, explain a significant (or close to significant) proportion of the

variance in dN/dS (Table 3, Figure 2 panel 3). In most of these cases group size showed a stronger effect than promiscuity. Also, in some cases the effect (slope) was negative, indicating, contrary to expectations, that for a given level of promiscuity species with smaller group sizes have higher dN/dS.

Because across species, RTS and group size show a significant positive correlation ($p<0.01$) (Figure 3), we also used the first axis derived from a principal component analysis of RTS and log group size (that captured ~78% of the variation in both variables) as a combined index of promiscuity and sociality. With a couple of exceptions, this resulted in a loss of significance and poorer fit with respect to the multiple regression models (data not shown). This reinforces the idea that, in spite of being positively correlated at a large taxonomic scale, mating system and group size might influence dN/dS independently and sometimes in opposite ways.

The previous analyses, particularly the independent contrasts, suggest that increases in dN/dS are associated with increases in promiscuity and increases in group size. An increase in dN/dS is suggestive of adaptive evolution, but only a dN/dS>1 constitutes unambiguous evidence of positive selection. Therefore, we compared the relative proportion of branches with dN/dS>1 between UM and MM branches and between SG and LG branches. When summed across genes, we found a significantly or marginally significant greater proportion of branches with dN/dS>1 among MM species, than among UM species for two of the three mating system partitions (Table 4). This pattern is driven by the pathogen-interacting genes (Table 4, Figure 2 panel 4). This indicates that MM species have on average, more instances of positive selection than UM

species and argues for a role of sexual promiscuity in the evolution of immunity genes. On the other hand, the proportion of branches with dN/dS>1 was the same among SG and LG species for the entire dataset as well as for the two groups of genes considered separately (Table 4, Figure 2 panel 4).

DISCUSSION

The main determinants of rates of protein evolution at immune loci other than adaptive immune receptors are largely unknown. Coevolution between host and pathogens is frequently invoked to explain the rapid evolution of immune loci (Holmes 2004), but certain host features such as promiscuous behavior or sociality might also have an effect by influencing disease risk. Here we investigated whether female mating promiscuity and other social and ecological variables have had a major effect on the rate of molecular evolution in functionally diverse immune defense genes. The underlying hypothesis is that the risk of STD (or more generally infectious diseases) should be higher in species with multiple mating, increasing pathogen exposure and/or diversity, and thus exerting stronger selective pressures on the host.

Using a comparative approach, we demonstrated a positive correlation between promiscuity and the rate of protein evolution at these genes across primates (Tables 1 and 4, Figure 2). We also found a positive correlation between group size and the rate of evolution (Table 1, Figure 2). The effect is weak, and only significant when combining data across genes. However, the noise introduced by trait measurement errors, unknown trait variation within species or over time, incomplete phylogenetic sampling and the fact that disease risk is likely influenced by other variables make this approach conservative. Given all the potential sources of variation, it is in fact remarkable to find a signal, suggesting that promiscuity, group size, or some other correlated variable, genuinely impacts immune protein evolution.

By controlling for phylogeny, we first found that transitions to higher promiscuity and larger group size were associated with increases in dN/dS (Table 1). In spite of the low statistical power to conduct tests on a gene-by-gene basis, most genes showed the same trend that emerged when we combined genes (data not shown), ruling out the possibility that one or a few outliers are driving the general pattern. Interestingly, we observed more transitions to higher dN/dS associated with higher promiscuity in the group of genes that directly interact with pathogens, such as the antiretroviral genes and the pattern recognition receptors. In line with theory, genes that lie at the host-pathogen interface should exhibit more evidence of selection due to coevolutionary arms races with pathogens. The increase in dN/dS associated with large groups, on the other hand, was not restricted to pathogen-interacting genes, but was instead distributed across the entire set of immunity genes.

The trends that emerged when comparing the mean dN/dS among species with low and high promiscuity or small and large groups, or when regressing dN/dS against the range of promiscuity and group size, were generally consistent with the independent contrasts but were largely not significant (Tables 2 and 3). Higher mean and variance in dN/dS were usually associated with more promiscuous mating systems or species with larger groups, as expected if more contacts increase the opportunities for disease transmission. However, the results of the multiple regression models showed that, at least in some cases, promiscuity and group size might have different effects on the rate of molecular evolution (Table 3).

An increase in dN/dS is suggestive of positive selection but might also reflect relaxed constraint. A more stringent analysis based on branches with unambiguous evidence of selection (dN/dS>1), also revealed more adaptive evolution in more promiscuous species, but not in species with larger groups (Table 4).

In spite of the overall pattern reported, a high degree of heterogeneity in dN/dS is evident from the branch-based analysis (Figure 1). This heterogeneity is not restricted to the promiscuous or large group branches but instead is distributed across the different gene phylogenies, and indicates that other lineage-specific factors might have similar importance. In light of such a high degree of heterogeneity, the comparison that focused on the proportion of branches with dN/dS>1 was more informative about the relative potential for natural selection. Similarly, the sign-test of positive contrasts (which focuses on the direction of the change but not the magnitude) resulted in more statistical power to expose the relationship between dN/dS, promiscuity and group size. Taken together, these two analyses suggest that higher levels of promiscuity and larger group size might underlie an increase in the rate of protein evolution. Nevertheless, only for promiscuity does this seems to be due to positive selection.

Interestingly, in spite of underlying differences in leukocyte levels in humans, the DUFFY gene did not exhibit patterns of substitution between species consistent with differences in promiscuity. At least three explanations can account for this result. i) It is possible that regulatory rather than coding variation at DUFFY is responsible for leukocyte differences between primates. In fact, in humans, that seems to be the case (Nalls et al. 2008; Reich et al. 2009). ii) Leukocyte levels might be under the genetic

control of loci other than DUFFY. iii) The pattern originally reported by Nunn, Gittleman, and Antonovics (2000) might not reflect evolved differences in leukocyte levels between species but instead might be caused by some other difference among promiscuous and monogamous primates in captivity, such as density, sex ratio or stress, as pointed out by Read and Allen (2000).

A potential problem with the interpretation of differences in evolutionary rates among species in the context of adaptation is that relaxation of purifying selection due to reduction in population size can also affect dN/dS (Ohta 1993b). In smaller populations, selection is less efficient at removing deleterious mutations, which should result in an increase in the rate of fixation of nonsynonymous changes, and a concomitant increase in dN/dS (Ohta 1993b). For example, primates have a higher dN/dS and lower effective population size than rodents (Ohta 1993a; Hughes and Friedman 2009). If promiscuity or group size are correlated with effective population size in primates, this might result in a spurious correlation between these variables and dN/dS.

A few lines of evidence seem to argue against this possibility in the data presented here. For a few species of primates, estimates of effective population sizes (N_e) are available [human: (Yu et al. 2004); chimpanzee: (Yu et al. 2004); bonobo: (Yu et al. 2004; Won and Hey 2005); gorilla: (Yu et al. 2004); rhesus and cynomolgus macaques: (Stevison and Kohn 2009)]. Additionally, for the orangutan, we estimated N_e based on available estimates of polymorphism and divergence (Fischer et al. 2006). We calculated the neutral mutation rate as $\mu=Da/2t$ (Kimura 1983) where Da is the net sequence divergence and t is the divergence time between the two species compared. We used a net

sequence divergence of 2.96 % between orangutans and humans, obtained by subtracting the average of the human and orangutan nucleotide diversity (π) from the raw sequence divergence between the species, and a divergence time of 13.5 MY (Goodman et al. 1998). Assuming a generation time of 15 years we obtained a mutation rate per site per generation of 1.65×10^{-8}. Then, using the nucleotide diversity estimated by Fischer et al. (2006) of 0.36 %, we calculated N_e as $\pi/4\mu$ and obtained an effective population size of 54,545. These few species of apes, human and macaques do not show any consistent relation between population size and dN/dS in the 11 genes for which at least four of these species were included (Figure 4). Similarly, no consistent pattern has been found in the rate of molecular evolution of social and non-social insects (Schmitz and Moritz 1998; Bromham and Leys 2005). Finally, as mentioned above, dN/dS values greater than one are only expected under positive selection. Our analyses based on branches with dN/dS>1 should then reflect patterns of adaptation and not simply relaxation of constraint.

However, it is still possible that at the larger scale of the primate radiation the accumulation of slightly deleterious substitutions in species with smaller population sizes has contributed to some extent to the pattern. If more social primates tend to have on average lower effective population sizes (as has been proposed for social insects; Crozier 1979), this might offer an explanation for the weaker effect of group size. The fact that in the analysis of independent contrast, species with larger groups (a proxy for more social species) had higher dN/dS at both the pathogen-interacting and non pathogen-interacting genes is consistent with this idea, because a population size effect is expected to affect all

genes equally. Also in line with this hypothesis, none of the genes or groups of genes showed an excess of branches with dN/dS>1 among the LG species, indicating that positive selection is not necessarily more prevalent in LG species. Thus, it is plausible that overall acceleration in dN/dS in LG species is due to relaxed purifying selection along these branches.

Many theoretical and empirical studies have suggested strong connections between social organization and the spread of horizontally transmitted parasites (e.g. Cote and Poulin 1995; reviewed in Altizer et al. 2003). In primates, somewhat contradictory results have been obtained when correlating direct and indirect measures of disease risk and sociality defined in a broad sense (components of mating and social systems). For example, in spite of the positive relationship between white blood cells and sexual promiscuity, spleen mass, another surrogate measure of disease risk, was not associated with measures of sociality or promiscuity (Nunn 2002a). Similarly, Nunn (2003) did not find support for the hypothesis that behaviors expected to reduce STD transmission are correlated with promiscuity. On the other hand, sociality measured as group size accounts for helminth diversity (Vitone, Altizer, and Nunn 2004), but population density (another measure of social contact) is the main predictor of parasite species richness in primates, including all the main classes of parasites (Nunn et al. 2003). Neither the previously mentioned studies nor ours found a strong effect of population density, although in some cases, the incorporation of density in our multiple regression models significantly improved the fit (data not shown). The integration of all

these results, however, is not straightforward because different aspects of immune defense might be characterized by different trade-offs and constraints.

Using the rate of molecular evolution at immunity genes as a surrogate of disease risk, our comparative data on 15 primate defense genes provides support for the idea that female promiscuity increases the potential for natural selection at the immune system level. The detected effect of promiscuity, to the exclusion of group size and density, is consistent with the idea that STDs might be important drivers of this pattern. This is an intriguing result, because even if they are expected to interact with sexually transmitted pathogens or participate in pathways that lead to their clearance, the genes included in this study are not specifically involved in immunity against STDs. Recently compiled information of primate parasites show that STDs are common in non-human primates and the documented STDs appear to be more frequent in promiscuous species (Nunn and Altizer 2006). Moreover, most of the known sexually transmitted pathogens in non-human primates are viruses, and among viruses, those transmitted by close contact (sexual or non-sexual) exhibit higher levels of host specificity (Pedersen et al. 2005). In virtue of this closer relationship with their hosts, it is in principle possible that sexually transmitted pathogens engage more often in arms races with their hosts than pathogens with other transmission modes.

The hypotheses tested here are not mutually exclusive, and the variables studied as well as other potentially confounding variables could interact in complicated ways. Importantly, focusing on the opportunities for disease transmission facilitated by social structure is only one of the possible theoretical frameworks in which to cast this problem.

Another equally valid approach would be to study how social behavior is shaped by disease risk over evolutionary time or as a plastic response. Our results provide another interesting piece of information linking promiscuity, STDs and the evolution of the immune system, but this complex relationship is far from being understood. Even if sexual promiscuity causally underlies the pattern of evolution of immunity genes, a large portion of the variance in dN/dS remains unexplained and suggests that the biological details of host-pathogen interactions in particular lineages play a large role in determining rates of evolution of immunity genes.

REFERENCES

Altizer, S, Nunn, CL, Thrall, PH et al. 2003. Social organization and parasite risk in mammals: Integrating theory and empirical studies. Annual Review of Ecology Evolution and Systematics 34:517-547.

Anderson, MJ, Hessel, JK, and Dixson, AF. 2004. Primate mating systems and the evolution of immune response. Journal of Reproductive Immunology 61:31-38.

Bromham, L, and Leys, R. 2005. Sociality and the rate of molecular evolution. Mol. Biol. Evol. 22:1393-1402.

Com, E, Bourgeon, F, Evrard, B, Ganz, T, Colleu, D, Jegou, B, and Pineau, C. 2003. Expression of antimicrobial defensins in the male reproductive tract of rats, mice, and humans. Biology of Reproduction 68:95-104.

Cote, IM, and Poulin, R. 1995. Parasitism and group size in social animals - A metaanalysis. Behavioral Ecology 6:159-165.

Crozier, R. 1979. Genetics of sociality. Pp. 223-286 in H. R. Hermann, ed. Social insects. Academic Press, New York.

Dunn, FL. 1968. The parasites of Saimiri: in the context of platyrrhine parasitism. Pp. 31-68 in Rosenblum, LA, and Cooper, RW, eds. The Squirrel Monkey. Academic Press, New York.

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

Fedorka, KM, and Zuk, M. 2005. Sexual conflict and female immune suppression in the cricket, Allonemobious socius. Journal of Evolutionary Biology 18:1515-1522.

Felsenstein, J. 1985. Phylogenies and the comparative method. Am. Nat. 125:1-15.

Filip, LC, and Mundy, NL. 2004. Rapid evolution by positive Darwinian selection in the extracellular domain of the abundant lymphocyte protein CD45 in primates. Mol. Biol. Evol. 21:1504-1511.

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

Goodman, M, Porter, CA, Czelusniak, J, Page, SL, Schneider, H, Shoshani, J, Gunnell, G, and Groves, CP. 1998. Toward a phylogenetic classification of primates based

on DNA evidence complemented by fossil evidence. Mol. Phylogenet. Evol. 9:585-598.

Gustafsson, L, Nordling, D, Andersson, MS, Sheldon, BC, and Qvarnstrom, A. 1994. Infectious diseases, reproductive effort and the cost of reproduction in birds. Pp. 323-331.

Hamilton, WD, and Zuk, M. 1982. Heritable true fitness and bright birds - a role for parasites. Science 218:384-387.

Harcourt, AH. 1991. Sperm competition and the evolution of nonfertilizing sperm in mammals. Evolution 45:314-328.

Harcourt, AH, Purvis, A, and Liles, L. 1995. Sperm competition - Mating system, not breeding season, affects testes of primates. Functional Ecology 9:468-476.

Holmes, EC. 2004. Adaptation and immunity. Plos Biology 2:1267-1269.

Hughes, AL, and Friedman, R. 2009. More radical amino acid replacements in primates than in rodents: Support for the evolutionary role of effective population size. Gene 440:50-56.

Immerman, RS. 1986. Sexually transmitted disease and human evolution: survival of the ugliest? Human Ethology Newsletter 4:6-7.

Kenagy, GJ, and Trombulak, SC. 1986. Size and function of mammalian testes in relation to body size. J. Mammal. 67:1-22.

Kerns, JA, Emerman, M, and Malik, HS. 2008. Positive selection and increased antiviral activity associated with the PARP-containing isoform of human zinc-finger antiviral protein. Plos Genetics 4.

Kimura, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge, United Kingdom.

Lawniczak, MKN, Barnes, AI, Linklater, JR, Boone, JM, Wigby, S, and Chapman, T. 2007. Mating and immunity in invertebrates. Trends in Ecology & Evolution 22:48-55.

Lawniczak, MKN, and Begun, DJ. 2004. A genome-wide analysis of courting and mating responses in Drosophila melanogaster females. Genome 47:900-910.

Li, P, Chan, HC, He, B, So, SC, Chung, YW, Shang, Q, Zhang, YD, and Zhang, YL. 2001. An antimicrobial peptide gene found in the male reproductive system of rats. Science 291:1783-1785.

Lindenfors, P, and Tullberg, BS. 1998. Phylogenetic analyses of primate size evolution: the consequences of sexual selection. Biol. J Linn. Soc. 64:413-447.

Lochmiller, RL, and Deerenberg, C. 2000. Trade-offs in evolutionary immunology: just what is the cost of immunity? Oikos 88:87-98.

Lockhart, AB, Thrall, PH, and Antonovics, J. 1996. Sexually transmitted diseases in animals: Ecological and evolutionary implications. Biological Reviews of the Cambridge Philosophical Society 71:415-471.

Loehle, C. 1995. Social barriers to pathogen transmission in wild animal populations. Ecology 76:326-335.

Lung, O, Kuo, L, and Wolfner, MF. 2001. Drosophila males transfer antibacterial proteins from their accessory gland and ejaculatory duct to their mates. Journal of Insect Physiology 47:617-622.

McGraw, LA, Gibson, G, Clark, AG, and Wolfner, MF. 2004. Genes regulated by mating, sperm, or seminal proteins in mated female Drosophila melanogaster. Curr. Biol. 14:1509-1514.

Moore, J. 2002. Parasites and the behavior of animals. Oxford University Press, Oxford.

Nalls, MA, Wilson, JG, Patterson, NJ et al. 2008. Admixture mapping of white cell count: Genetic locus responsible for lower white blood cell count in the health ABC and Jackson Heart Studies. Am. J. Hum. Genet. 82:81-87.

Nielsen, R, and Yang, ZH. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936.

Nunn, CL. 2002a. Spleen size, disease risk and sexual selection: a comparative study in primates. Evol Ecol Res 4:91-107.

Nunn, CL. 2002b. A comparative study of leukocyte counts and disease risk in primates. Evolution 56:177-190.

Nunn, CL. 2003. Behavioural defenses against sexually transmitted diseases in primates. Anim Behav 66:37-48.

Nunn, CL, Altizer, S, Jones, KE, and Sechrest, W. 2003. Comparative tests of parasite species richness in primates. Am. Nat. 162:597-614.

Nunn, CL, and Altizer, SM. 2006. Infectious diseases in primates. Behavior, Ecology and Evolution. Oxford University Press, New York.

Nunn, CL, Gittleman, JL, and Antonovics, J. 2000. Promiscuity and the primate immune system. Science 290:1168-1170.

OhAinle, M, Kerns, JA, Malik, HS, and Emerman, M. 2006. Adaptive evolution and antiviral activity of the conserved mammalian cytidine deaminase APOBEC3H. Journal of Virology 80:3853-3862.

Ohta, T. 1993a. An examination of the generation-time effect on molecular evolution. Proc. Natl. Acad. Sci. USA 90:10676-10680.

Ohta, T. 1993b. Amino-acid substitution at the Adh locus of Drosophila is facilitated by small population size. Proc. Natl. Acad. Sci. USA 90:4548-4551.

Osorio, DS, Antunes, A, and Ramos, MJ. 2007. Structural and functional implications of positive selection at the primate angiogenin gene. Bmc Evolutionary Biology 7.

Pedersen, AB, Altizer, S, Poss, M, Cunnighan, A, and Nunn, CL. 2005. Patterns of host specificity and transmission among parasites of wild primates. Int J Parasitol. 35:647-657.

Purvis, A. 1995. A Composite Estimate of Primate Phylogeny. Philos T Roy Soc B 348:405-421.

Purvis, A, and Rambaut, A. 1995. Comparative analysis by independent contrasts (CAIC) - an Apple-Macintosh application for analyzing comparative data. Comput. Appl. Biosci. 11:247-251.

Read, AF, and Allen, JE. 2000. Evolution and immunology - The economics of immunity. Science 290:1104-1105.

Reddy, KVR, Yedery, RD, and Aranha, C. 2004. Antimicrobial peptides: premises and promises. International Journal of Antimicrobial Agents 24:536-547.

Reich, D, Nalls, MA, Kao, WHL et al. 2009. Reduced Neutrophil Count in People of African Descent Is Due To a Regulatory Variant in the Duffy Antigen Receptor for Chemokines Gene. Plos Genetics 5.

Rowe, N. 1996. The pictorial guide of living primates. Pogonias Press.

Sawyer, SL, Emerman, M, and Malik, HS. 2004. Ancient adaptive evolution of the primate antiviral DNA-editing enzyme APOBEC3G. Plos Biology 2:1278-1285.

Sawyer, SL, Emerman, M, and Malik, HS. 2007. Discordant evolution of the adjacent antiretroviral genes TRIM22 and TRIM5 in mammals. Plos Pathogens 3:1918-1929.

Sawyer, SL, Wu, LI, Emerman, M, and Malik, HS. 2005. Positive selection of primate TRIM5 alpha identifies a critical species-specific retroviral restriction domain. Proc. Natl. Acad. Sci. USA 102:2832-2837.

Schmid-Hempel, P. 2003. Variation in immune defense as a question of evolutionary ecology. P. Roy. Soc. Lond. B Bio. 270:357-366.

Schmitz, J, and Moritz, RFA. 1998. Sociality and the rate of rDNA sequence evolution in wasps (Vespidae) and honeybees (Apis). J Mol Evol 47:606-612.

Semple, S, Cowlishaw, G, and Bennett, PM. 2002. Immune system evolution among anthropoid primates: parasites, injuries and predators. P. Roy. Soc. Lond. B Bio. 269:1031-1037.

Sheldon, BC, and Verhulst, S. 1996. Ecological immunology: Costly parasite defenses and trade-offs in evolutionary ecology. Trends in Ecology & Evolution 11:317-321.

Silphaduang, U, Hincke, MT, Nys, Y, and Mine, Y. 2006. Antimicrobial proteins in chicken reproductive system. Biochemical and Biophysical Research Communications 340:648-655.

Stevison, LS, and Kohn, MH. 2009. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. Mol. Ecol. 18:2457-2475.

Swanson, WJ, Nielsen, R, and Yang, QF. 2003. Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. 20:18-20.

Swofford, DL. 1993. PAUP - A computer program for phylogenetic inference using Maximum Parsimony. Journal of General Physiology 102:A9-A9.

Thrall, PH, Antonovics, J, and Bever, JD. 1997. Sexual transmission of disease and host mating systems: Within-season reproductive success. Am. Nat. 149:485-506.

Vitone, ND, Altizer, S, and Nunn, CL. 2004. Body size, diet and sociality influence the species richness of parasitic worms in anthropoid primates. Evol Ecol Res 6:183-199.

Wlasiuk, G, Khan, S, Switzer, WM, and 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, and 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, and Pedersen, AMK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.

Yu, N, Jensen-Seaman, MI, Chemnick, L, Ryder, O, and Li, WH. 2004. Nucleotide diversity in gorillas. Genetics 166:1375-1383.

Zelezetsky, I, Pontillo, A, Puzzi, L, Antcheva, N, Segat, L, Pacor, S, Crovella, S, and Tossi, A. 2006. Evolution of the primate cathelicidin - Correlation between structural variations and antimicrobial activity. Journal of Biological Chemistry 281:19861-19871.

Zhang, ZD, Weinstock, G, and Gerstein, M. 2008. Rapid evolution by positive darwinian selection in t-cell antigen CD4 in primates. J Mol Evol 66:446-456.

Zuk, M, and Stoehr, AM. 2002. Immune defense and host life history. Am. Nat. 160:S9-S22.

Table 1. Phylogenetically independent contrasts of dN/dS and sexual, social and ecological variables across genes.

	Independent variable	Positive contrasts [a]	Total contrasts	Sign-test [b]
All genes	Habitat	20	44	p>0.50
	Diet	23	51	p>0.50
	Mating System (1)	41	73	p=0.17
	Mating System (2)	43	74	p=0.10#
	Mating System (3)	39	67	p=0.11
	Residual Testis size	19	36	p=0.43
	Group size	24	38	p=0.07#
	Density	13	27	p>0.50
Non pathogen interacting genes	Habitat	8	15	p>0.50
	Diet	6	14	p>0.50
	Mating System (1)	10	25	p>0.50
ANG, CCR5, CD4, CD45,	Mating System (2)	11	27	p>0.50
DUFFY	Mating System (3)	10	25	p>0.50
	Residual Testis size	3	11	p>0.50
	Group size	8	10	p=0.05*
	Density	7	13	p>0.50
Pathogen interacting genes	Habitat	12	29	p>0.50
	Diet	17	37	p>0.50
APOBEC3G, APOBEC3H,	Mating System (1)	31	48	p=0.03*
CAMP, PKR, TLR1, TLR4,	Mating System (2)	32	47	p=0.01*
TLR5, TRIM5, TRIM22,	Mating System (3)	29	42	p=0.01*
ZAP	Residual Testis size	16	25	p=0.11
	Group size	18	27	p=0.06#
	Density	6	14	p>0.50

[a] A contrast is positive when both variables vary in the direction predicted by the hypothesis.
[b] One-tailed test

#=p<0.1, *=p<0.05

Table 2. Mean and variance in dN/dS of immunity genes in relation to mating system and group size.

		Mating system partition 1				Mating system partition 2				Mating system partition 3				Group Size			
		UM[a]		MM[b]		UM[a]		MM[b]		UM[a]		UM[a]		SG[c]		LG[d]	
Class	Gene	mean[e]	var[f]	mean[e]	var[f]	mean[e]	var[f]	mean[e]	var[f]	mean[e]	var[f]	mean[e]	var[f]	mean[e]	var[f]	mean[e]	var[f]
Non pathogen-interacting genes	ANG	1.39	1.09	0.97	0.63	1.50	1.28	0.96	0.50	1.50	1.28	0.97	0.63	**0.68**	**0.04**	**1.36**	**1.10 ***
	CCR5	0.38	0.19	0.21	0.04	0.37	0.19	0.25	0.06	0.37	0.19	0.21	0.04	**0.29**	0.23	**0.36**	0.16
	CD4	0.71	0.32	0.72	0.11	**0.48**	0.17	**0.80**	0.15	**0.48**	0.17	**0.72**	0.11	**0.70**	**0.00**	**0.78**	**0.19 #**
	CD45	**2.69**	**3.48**	**5.71 #**	**11.86**	**2.69**	**3.48**	**5.71 #**	**11.86**	**2.69**	**3.48**	**5.71 #**	**11.86**	4.55	8.23	3.27	-
	DUFFY	0.56	0.09	0.39	0.06	0.57	0.10	0.40	0.06	0.57	0.10	0.39	0.06	**0.34**	**0.01**	**0.45**	**0.07 ***
Pathogen-interacting genes	APOBEC3G	**1.35**	1.95	**1.59**	0.35	**0.88**	**0.84**	**1.83 #**	**1.07**	**0.88**	0.84	**1.59 #**	0.35	**0.87**	**0.21**	**2.04 ***	**1.28 #**
	APOBEC3H	**0.97**	**0.09**	**1.11**	**0.52 #**	**0.86**	**0.07**	**1.13**	**0.46**	**0.86**	**0.07**	**1.11**	**0.52**	1.24	**0.41**	0.90	**0.51**
	CAMP	1.47	2.21	0.88	0.29	1.88	2.54	0.73	0.20	1.88	2.54	0.88	0.29	0.93	0.41	0.65	0.16
	PKR	**0.79**	**0.51**	**1.42 #**	**1.14**	**0.95**	**0.54**	**1.18**	**1.15**	**0.95**	**0.54**	**1.42**	**1.14**	**0.90**	**0.19**	**1.33**	**1.33 ***
	TLR1	**0.50**	**0.04**	**0.65**	**0.20 ***	0.57	**0.06**	0.55	**0.11**	**0.57**	**0.06**	**0.65**	**0.20**	**0.43**	**0.02**	**0.65**	**0.30 ***
	TLR4	**0.64**	**0.43**	**1.18 #**	**1.17 #**	**0.54**	**0.45**	**1.14 #**	**0.99**	**0.54**	**0.45**	**1.11**	**1.08**	1.27	1.74	0.69	0.23
	TLR5	**0.43**	**0.02**	**0.74**	**0.65 ***	**0.44**	**0.02**	**0.67**	**0.53 ***	**0.44**	**0.02**	**0.74**	**0.65 ***	**0.35**	**0.02**	**0.49**	**0.08 #**
	TRIM5	1.28	0.88	1.22	0.26	1.38	1.21	1.17	0.23	1.38	1.21	1.22	0.26	1.56	0.84	0.95	0.26
	TRIM22	**0.36**	**0.06**	**0.52**	**0.29**	**0.34**	**0.07**	**0.47**	**0.20**	**0.34**	**0.07**	**0.52**	**0.29**	**0.35**	**0.02**	**0.50**	**0.30 ***
	ZAP	**0.61**	**0.08**	**0.76**	**0.26**	0.76	**0.07**	0.62	**0.19**	0.76	**0.07**	0.76	**0.26**	0.82	0.31	0.55	0.09

[a] Unimale species
[b] Multimale species
[c] Small group size
[d] Large group size
[e] t-test between average dN/dS of MM branches and average dN/dS of UM branches (1 tail test)
[f] z-test of differences in variance of dN/dS between MM and UM branches (1 tail test)
*=$p<0.05$
#=$p<0.1$
Shaded cells indicate cases in which the mean or variance in dN/dS are higher in MM or LG species.

Table 3. Relative effects of promiscuity and group size on the rate of evolution of immunity genes.

Class	Gene	Model 1 [a] fit		Individual factor effects [b]			Model 2 [c] fit		Individual factor effects [b]		
		R^2	p-value[d]	RTS [e]	log GS [f]	RTS*log GS [g]	R^2	p-value[d]	MS3 [h]	log GS [f]	M3*log GS [i]
Non pathogen-interacting genes	ANG	0.52	0.07 (#)	- (*)	+ (*)		0.36	0.42	-	+	+
	CCR5	0.16	0.33	-	+	+	0.12	0.49	-	+	-
	CD4	0.24	0.67	-	+	-	0.02	0.99	+	-	+
	CD45	0.4	0.51	+	-	+	0.63	0.08 (#)	+	- (#)	
	DUFFY	0.24	0.29	-	+	-	0.16	0.49	-	+	-
Pathogen-interacting genes	APOBEC3G	0.34	0.04 (*)		+ (*)		0.58	0.03 (*)	+	+ (*)	
	APOBEC3H	0.07	0.88	-	+	+	0.13	0.8	+	-	-
	CAMP	0.53	0.34	-	-	+	0.23	0.75	-		-
	PKR	0.16	0.53	+	-	+	0.11	0.73	+	-	+
	TRL1	0.35	0.42	-	+	+	0.22	0.77	+	+	-
	TLR4	0.23	0.35	+	-	+	0.46	0.03 (*)	+ (*)	- (*)	
	TLR5	0.14	0.56	+	+	+	0.15	0.72	+	+	+
	TRIM5	0.06	0.85	+	-	-	0.09	0.82	-	-	-
	TRIM22	0.18	0.64	+	-	+	0.29	0.52	+	-	-
	ZAP	0.31	0.05 (*)		+ (*)		0.59	0.03 (*)	+	+ (*)	

[a] Model includes residual testis size and group size
[b] Contribution of individual factors to the model based on the direction of the slope
[c] Model includes mating system partition 3 and group size
[d] Significance of the model
[e] Residual testis size
[f] Group size
[g] Interaction term between RTS and log GS
[h] Mating system partition 3 (excludes species with unambiguous mating systems)
[i] Interaction term between MS3 and log GS
(*) = $p<0.05$, (#) = $p<0.10$

Table 4. Adaptive evolution at immunity genes along lineages of multimale species or species with large groups.

		Mating system partition 1				Mating system partition 2				Mating system partition 3				Group Size			
		UM[a] branches		MM[b] branches		UM[a] branches		MM[b] branches		UM[a] branches		MM[b] branches		SG[c] branches		LG branches[d]	
Class	Gene	ω[e]>1	Total	ω[e]>1	Total	ω[e]>1	Total	ω[e]>1	Total	ω[e]>1	Total	ω[e]>1	Total	ω[e]>1	Total	ω[e]>1	Total
Non pathogen-interacting genes	**ANG**	3	6	2	5	3	5	2	6	2	5	2	5	0	3	3	5
	CCR5	2	15	0	10	2	13	0	12	2	13	0	10	1	7	1	10
	CD4	1	4	3	7	0	3	4	8	0	3	3	7	0	2	3	6
	CD45	5	6	2	2	5	6	2	2	5	6	2	2	4	4	1	1
	DUFFY	0	10	0	9	0	9	0	10	0	9	0	9	0	5	0	6
Pathogen-interacting genes	**APOBEC3G**	3	7	5	6	1	5	7	8	1	5	5	6	2	4	5	6
	APOBEC3H	2	4	4	8	1	3	5	9	1	3	4	8	2	3	2	6
	CAMP	4	7	2	4	4	5	2	6	4	5	2	4	2	4	1	3
	PKR	3	9	4	8	3	7	4	10	3	7	4	8	1	5	4	8
	TLR1	0	7	1	4	0	4	1	7	0	4	1	4	0	4	1	3
	TLR4	2	9	4	9	1	7	5	11	1	7	4	9	2	4	2	6
	TLR5	0	11	1	8	0	9	1	10	0	9	1	8	0	5	0	7
	TRIM5	5	10	6	8	4	7	7	11	4	7	6	8	5	7	3	6
	TRIM22	0	11	1	9	0	6	1	14	0	5	1	9	0	7	1	9
	ZAP	1	7	2	5	1	4	2	8	1	4	2	5	2	4	1	5
	ALL GENES	31	123	37	102	25	93	43	132	24	92	37	102	21	68	28	87
	p-value[f]		p=0.05*				n.s.				p=0.09#				n.s.		
	PATHOGEN INTERACTING	20	82	30	69	15	57	35	94	15	56	30	69	16	47	20	59
	p-value[f]		p=0.01*				p=0.12				p=0.04*				n.s.		
	Non PATHOGEN INTERACTING	11	41	7	33	10	36	8	38	9	36	7	33	5	21	8	28
	p-value[f]		n.s				n.s				n.s				n.s.		

[a] Unimale
[b] Multimale
[c] Below one standard deviation around the mean log group size
[d] Above one standard deviation around the mean log group size
[e] ω=dN/dS
[f] 1-tailed Z-test of the difference between the proportions of branches with dN/dS>1 between UM and MM classes or (SG and LG), across the 15 genes.
*=p<0.05, #=p<0.1

FIGURE LEGENDS

Figure 1. Phylogenies of the 15 genes [species trees according to Purvis (1995)] showing dN/dS values (from free-ratio ML models), mating system, and group size (discretized). Branches are not to scale. Red branches indicate multimale species, blue branches indicate unimale species, and black branches represent species with ambiguous mating system. Yellow squares indicate species with large groups and green squares indicate species with small groups. l.d.=low divergence, 'dS=0'=no synonynonymous substitutions. For the terminal branches with 0 sysnonymous substitutions and more than 2 nonsynonymous substitutions, dN/dS was conservatively calculated assuming 1 synonymous substitution. These cases are indicated with an asterix.

Figure 2. Summary of tests using sexual promiscuity (residual testis size or mating system) and group size as independent variables. Red branches indicate multimale species or species with large groups, while blue branches represent unimale species or species with small groups. Grey branches indicate internal branches, whose dN/dS values were not used in the analyses. In panel 3 log group size was used as a continuous variable while in the rest of the analyses group size was discretized (see methods for details). Each of the analyses is shown in detail in Tables 1-4, corresponding to panels 1-4 in this figure. MS=Mating system, GS=Group size, UM=Unimale mating system, MM=Multimale mating system, SG=Small Group size, LG=Large group size**,** n.s.= not significant. **Panel 1**. Sign tests of the number of positive contrasts [in which an increase in dN/dS is

accompanied by an increase in promiscuity (MS3) or group size]. **Panel 2.** T-tests of differences in mean dN/dS between unimale and multimale species (or species with small groups and large groups). The right half of the figure (under OBSERVATIONS) shows the number of genes with the predicted pattern, followed by the number that are significant or marginally significant in parentheses. For example, for pathogen-interacting genes, 8/10 had higher dN/dS in MM branches than in UM branches, and 3/10 of these were marginally significant at $p<0.10$. **Panel 3**. Multiple regressions of dN/dS as dependent variable, with promiscuity (RTS or MS3) and log group size as independent variables. Only the effects of individual variables are shown. **Panel 4.** Z-test of differences in the proportion of branches with dN/dS>1 between unimale and multimale species (MS3), or species with small groups and large groups.

Figure 3. Positive correlation between residual testis size and log group size ($R^2=0.33$, $p=0.0001$).

Figure 4. Relationship between population size and dN/dS for a sample of primates that includes human, apes, and macaques. Effective population sizes were taken from the literature or calculated based on multi-locus polymorphism and divergence estimates (see Discussion). Only genes with a minimum of four species with available dN/dS values were included. Regression lines are shown. None of these were significant in the expected direction ($p>0.05$ for all genes except for PKR). PKR showed a positive correlation between N_e and dN/dS ($R^2=0.84$, $p=0.03$).

Figure 1

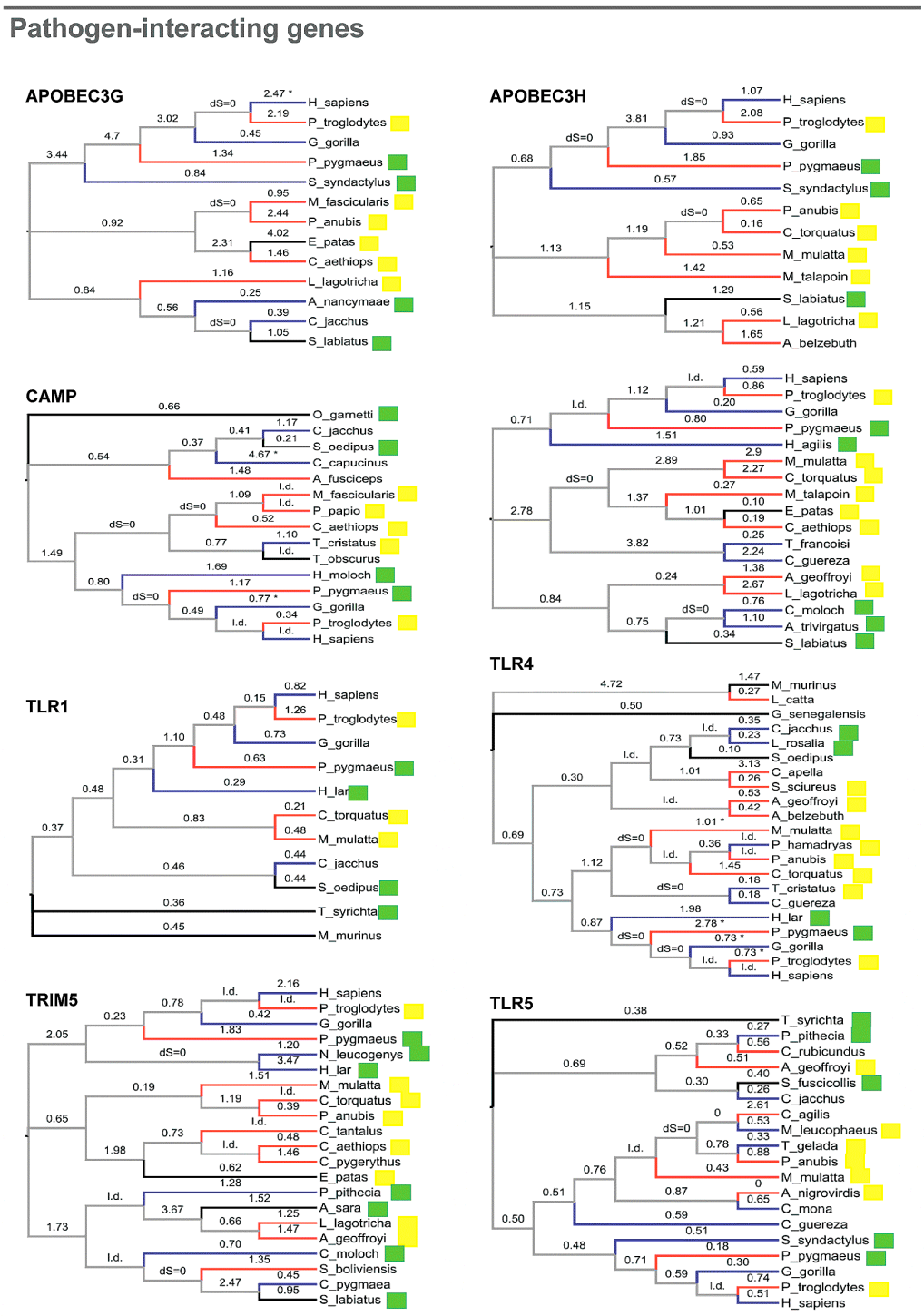

Pathogen-interacting genes
APOBEC3G
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
S_syndactylus
M_fascicularis
P_anubis
E_patas
C_aethiops
L_lagotricha
A_nancymaae
C_jacchus
S_labiatus
APOBEC3H
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
S_syndactylus
P_anubis
C_torquatus
M_mulatta
M_talapoin
S_labiatus
L_lagotricha
A_belzebuth
CAMP
O_garnetti
C_jacchus
S_oedipus
C_capucinus
A_fusciceps
M_fascicularis
P_papio
C_aethiops
T_cristatus
T_obscurus
H_moloch
P_pygmaeus
G_gorilla
P_troglodytes
H_sapiens
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
M_mulatta
C_torquatus
M_talapoin
E_patas
C_aethiops
T_francoisi
C_guereza
A_geoffroyi
L_lagotricha
C_moloch
A_trivirgatus
S_labiatus
TLR1
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
H_lar
C_torquatus
M_mulatta
C_jacchus
S_oedipus
T_syrichta
M_murinus
TLR4
M_murinus
L_catta
G_senegalensis
C_jacchus
L_rosalia
S_oedipus
C_apella
S_sciureus
A_geoffroyi
A_belzebuth
M_mulatta
P_hamadryas
P_anubis
C_torquatus
T_cristatus
C_guereza
H_lar
P_pygmaeus
G_gorilla
P_troglodytes
H_sapiens
TRIM5
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
N_leucogenys
H_lar
M_mulatta
C_torquatus
P_anubis
C_tantalus
C_aethiops
C_pygerythus
E_patas
P_pithecia
A_sara
L_lagotricha
A_geoffroyi
C_moloch
S_boliviensis
C_pygmaea
S_labiatus
TLR5
T_syrichta
P_pithecia
C_rubicundus
A_geoffroyi
S_fuscicollis
C_jacchus
C_agilis
M_leucophaeus
T_gelada
P_anubis
M_mulatta
A_nigrovirdis
C_mona
C_guereza
S_syndactylus
P_pygmaeus
G_gorilla
P_troglodytes
H_sapiens

Figure 1. Ctd.

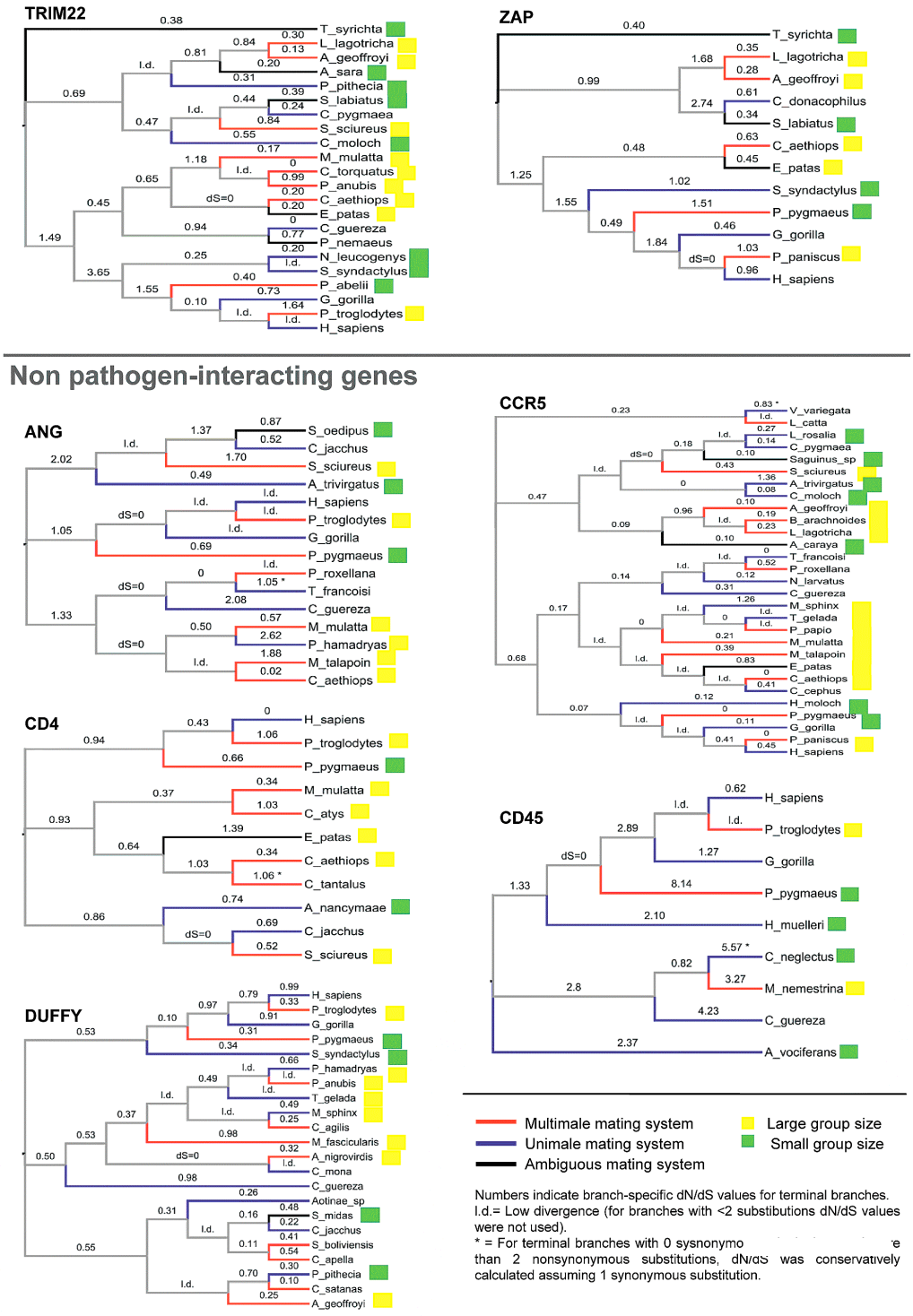

TRIM22
T_syrichta
L_lagotricha
A_geoffroyi
A_sara
P_pithecia
S_labiatus
C_pygmaea
S_sciureus
C_moloch
M_mulatta
C_torquatus
P_anubis
C_aethiops
E_patas
C_guereza
P_nemaeus
N_leucogenys
S_syndactylus
P_abelii
G_gorilla
P_troglodytes
H_sapiens
ZAP
T_syrichta
L_lagotricha
A_geoffroyi
C_donacophilus
S_labiatus
C_aethiops
E_patas
S_syndactylus
P_pygmaeus
G_gorilla
P_paniscus
H_sapiens
Non pathogen-interacting genes
ANG
S_oedipus
C_jacchus
S_sciureus
A_trivirgatus
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
P_roxellana
T_francoisi
C_guereza
M_mulatta
P_hamadryas
M_talapoin
C_aethiops
CCR5
V_variegata
L_catta
L_rosalia
C_pygmaea
Saguinus_sp
S_sciureus
A_trivirgatus
C_moloch
A_geoffroyi
B_arachnoides
L_lagotricha
A_caraya
T_francoisi
P_roxellana
N_larvatus
C_guereza
M_sphinx
T_gelada
P_papio
M_mulatta
M_talapoin
E_patas
C_aethiops
C_cephus
H_moloch
P_pygmaeus
G_gorilla
P_paniscus
H_sapiens
CD4
H_sapiens
P_troglodytes
P_pygmaeus
M_mulatta
C_atys
E_patas
C_aethiops
C_tantalus
A_nancymaae
C_jacchus
S_sciureus
CD45
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
H_muelleri
C_neglectus
M_nemestrina
C_guereza
A_vociferans
DUFFY
H_sapiens
P_troglodytes
G_gorilla
P_pygmaeus
S_syndactylus
P_hamadryas
P_anubis
T_gelada
M_sphinx
C_agilis
M_fascicularis
A_nigrovirdis
C_mona
C_guereza
Aotinae_sp
S_midas
C_jacchus
S_boliviensis
C_apella
P_pithecia
C_satanas
A_geoffroyi
Multimale mating system
Unimale mating system
Ambiguous mating system
Large group size
Small group size
Numbers indicate branch-specific dN/dS values for terminal branches.
l.d.= Low divergence (for branches with <2 substibutions dN/dS values were not used).
* = For terminal branches with 0 sysnonymo e than 2 nonsynonymous substitutions, dN/dS was conservatively calculated assuming 1 synonymous substitution.

Figure 2

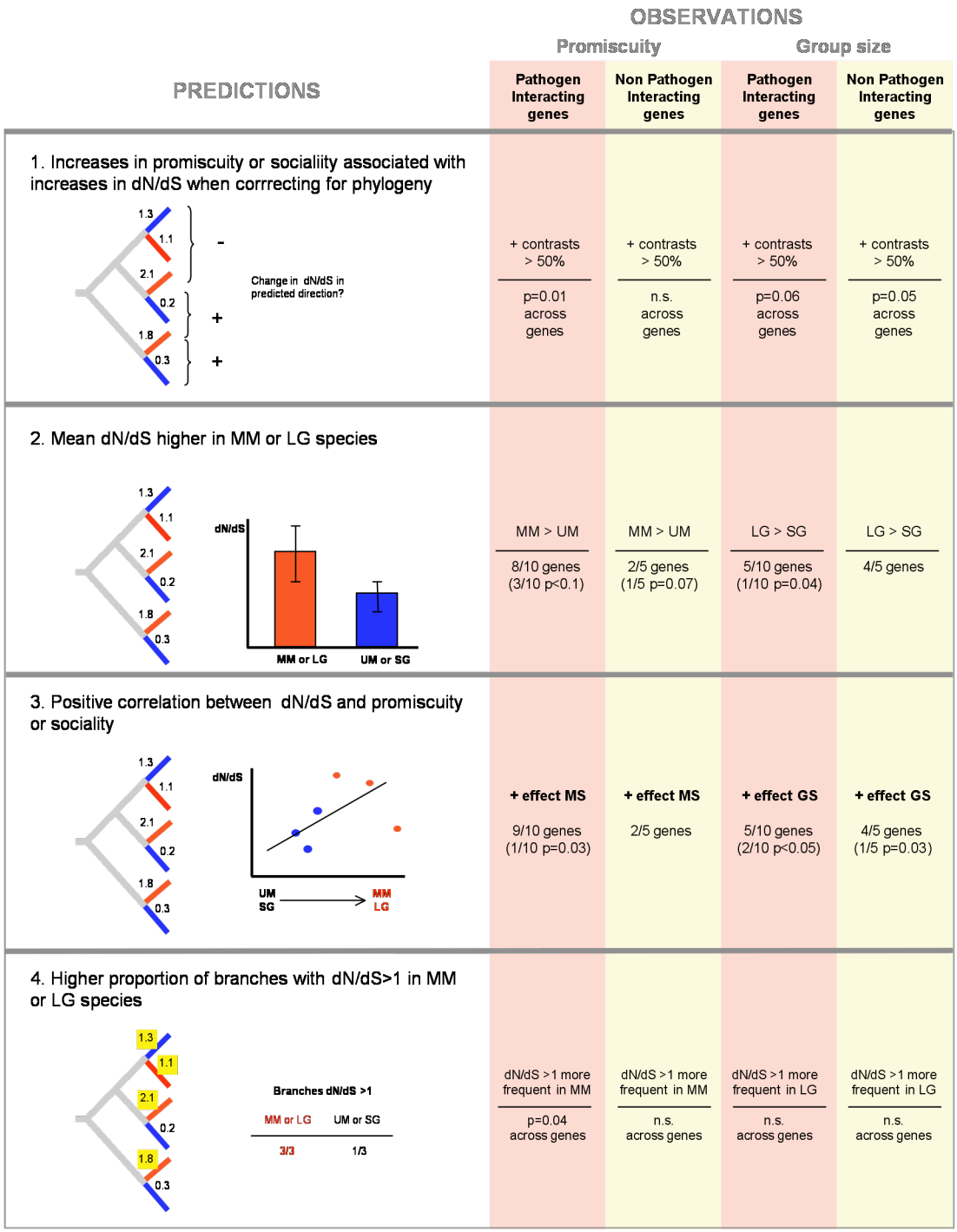

OBSERVATIONS
Promiscuity
Group size
PREDICTIONS
Pathogen Interacting genes
Non Pathogen Interacting genes
Pathogen Interacting genes
Non Pathogen Interacting genes
1. Increases in promiscuity or sociality associated with increases in dN/dS when corrrecting for phylogeny
1.3
1.1
2.1
0.2
1.8
0.3
-
+
+
Change in dN/dS in predicted direction?
+ contrasts > 50%
p=0.01 across genes
+ contrasts > 50%
n.s. across genes
+ contrasts > 50%
p=0.06 across genes
+ contrasts > 50%
p=0.05 across genes
2. Mean dN/dS higher in MM or LG species
dN/dS
MM or LG
UM or SG
MM > UM
8/10 genes (3/10 p<0.1)
MM > UM
2/5 genes (1/5 p=0.07)
LG > SG
5/10 genes (1/10 p=0.04)
LG > SG
4/5 genes
3. Positive correlation between dN/dS and promiscuity or sociality
dN/dS
UM SG
MM LG
+ effect MS
9/10 genes (1/10 p=0.03)
+ effect MS
2/5 genes
+ effect GS
5/10 genes (2/10 p<0.05)
+ effect GS
4/5 genes (1/5 p=0.03)
4. Higher proportion of branches with dN/dS>1 in MM or LG species
Branches dN/dS >1
MM or LG
UM or SG
3/3
1/3
dN/dS >1 more frequent in MM
p=0.04 across genes
dN/dS >1 more frequent in MM
n.s. across genes
dN/dS >1 more frequent in LG
n.s. across genes
dN/dS >1 more frequent in LG
n.s. across genes

Figure 3

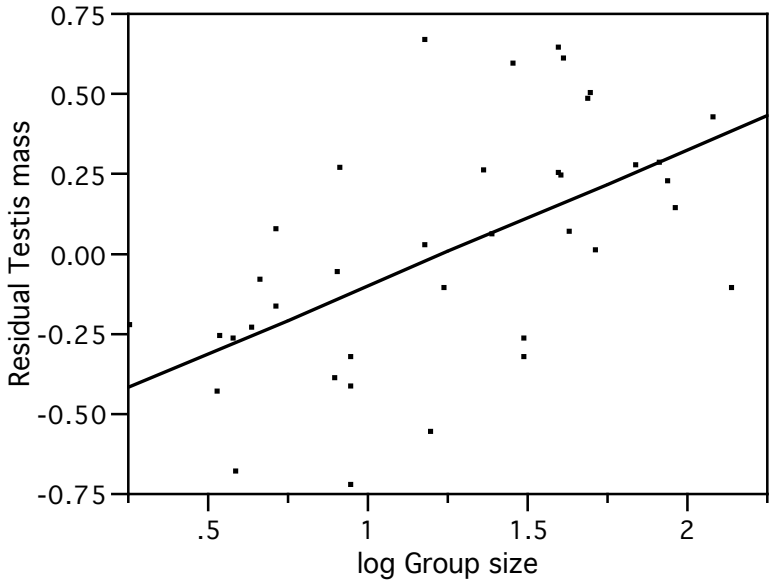

0.75
0.50
0.25
0.00
-0.25
-0.50
-0.75
Residual Testis mass
.5
1
1.5
2
log Group size

Figure 4

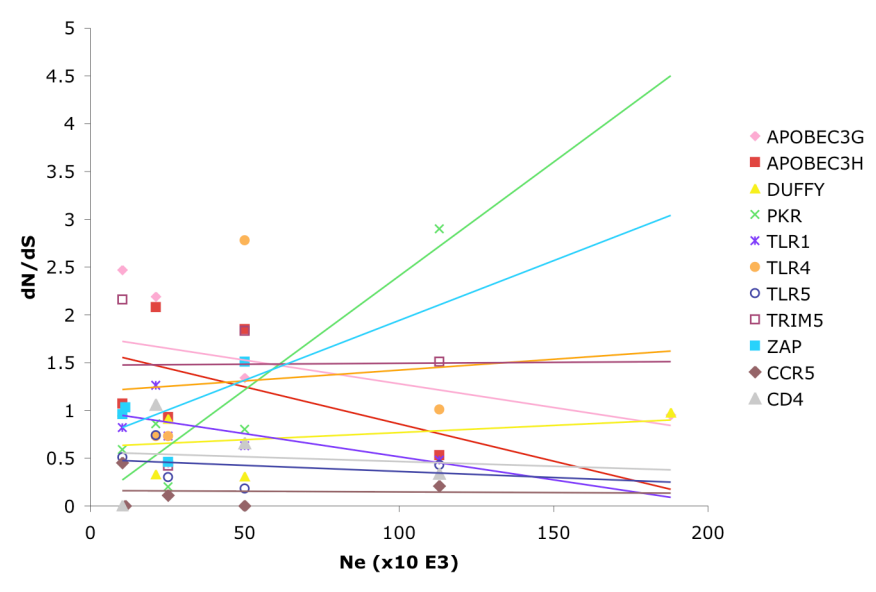

dN/dS
Ne (x10 E3)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
0
50
100
150
200
APOBEC3G
APOBEC3H
DUFFY
PKR
TLR1
TLR4
TLR5
TRIM5
ZAP
CCR5
CD4

Table S1a. Species and sequence accession numbers used in this study for ANG, APOBEC3G, APOBEC3H, CAMP, CCR5, CD4 and CD45.

Species	ANG	APOBEC3G	APOBEC3H	CAMP	CCR5	CD4	CD45
Allenopithecus nigroviridis							
Allouata caraya					AF161945		
Alouatta sara							
Aotinae sp.							
Aotus nancymaae		FJ638412				FJ623078	
Aotus trivirgatus	AF441669				AF161947		
Aotus vociferans							AY445818
Ateles belzebuth			DQ408612				
Ateles fusciceps				DQ471355			
Ateles geoffroyi					AF177885		
Brachyteles aracnoides					AY278752		
Cacajao rubicundus							
Callicebus donacophilus							
Callicebus moloch					AF177887		
Callithrix jacchus	*	FJ638413		DQ471358		AF452616	
Callithrix pygmaea					AY278746		
Cebus apella							
Cebus capucinus				DQ471357			
Cercocebus agilis							
Cercocebus torquatus			DQ408611			X73327	
Cercopithecus cephus					AF035217		
Cercopithecus mona							
Cercopithecus neglectus							AY539665, AY539673, AY539681, AY539689, AY539697, AY539705
Cercopithecus pygerythrus							
Cercopithecus tantalus						AF001222	
Chiropotes satanas							
Chlorocebus aethiops	AF441664	AB266486		DQ471356	AF01937	D86589	
Colobus guereza	AY221128				AF141639		AY539666, AY539674, AY539682, AY539690, AY539698, AY539706
Erythrocebus patas		AH013824			AF177879	X73324	
Galago senegalensis							

Gorilla gorilla	AF441662	AY639868	EU861358	DQ471359	AF005659		AY539661, AY539669, AY539677, AY539685, AY539693, AY539701
Homo sapiens	NM_001145	NM_021822	NM_181773	BC055089	NM_000579	NM_000616	AY539659, AY539667, AY539675, AY539683, AY539691, AY539699
Hylobates agilis							
Hylobates lar							
Hylobates moloch				DQ471364	AF177899		
Hylobates mulleri							AY539663, AY539671, AY539679, AY539687, AY539695, AY539703,
Lagothrix lagothricha		AH013831	DQ408615		AF252553		
Lemur catta					AF162010		
Leontopithecus rosalia					AY278751		
Macaca fascicularis		AB266488		DQ471365			
Macaca mulatta	AF441667		DQ507277		DQ499968	D63347	
Macaca nemestrina							AY539664, AY539672, AY539680, AY539688, AY539696, AY539704
Mandrillus leucophaeus							
Mandrillus sphinx					AF177877		
Microcebus murinus							
Miopithecus talapoin	AF441665		DQ408613		AF177886		
Nasalis larvatus					AF177882		
Nomascus leucogenys							
Otolemur garnettii				ENSOGAG00000002998			
Pan paniscus					AF177893		
Pan troglodytes	AF441661	NM_001009001	EU861357	DQ471371		EF437442	AY539660, AY539668, AY539676, AY539684, AY539692, AY539700
Papio anubis		AH013832	DQ408605				
Papio hamadryas	AF441666						

Papio papio				DQ471369	AF161994		
Pithecia pithecia							
Pongo abelii							
Pongo pygmaeus	AF441663	AY639869	EU861359	DQ471370	AF075446	ENSPPYT00000004978	AY539662, AY539670, AY539678, AY539686, AY539694, AY539702
Pygathrix nemaeus							
Pygathrix roxellana	AY221130				AF075444		
Saguinus fuscicollis							
Saguinus labiatus		AH013830	DQ408614				
Saguinus midas							
Saguinus oedipus	AF441668			DQ471372			
Saguinus sp					AF161926		
Saimiri boliviensis							
Saimiri sciureus	AF441670				AF45261	D86588	
Symphalangus syndactylus		DQ251286	EU861360				
Tarsius syrichta							
Theropithecus gelada					AF177891		
Trachypithecus cristatus				DQ471367			
Trachypithecus francoisi	AY221129				AF075442		
Trachypithecus obscurus				DQ471368			
Varecia variegata					AF162013		

* = Sequence obtained from the C_jacchus draft assembly at the UCSC genome browser

Table S1b. Species and sequence accession numbers used in this study for DUFFY, PKR, TLR1, TLR4, TLR5, TRIM5, TRIM22 and ZAP.

Species	DUFFY	PKR	TLR1	TLR4	TLR5	TRIM5	TRIM22	ZAP
Allenopithecus nigroviridis	this study				FJ542207			
Allouata caraya								
Alouatta sara						AY843511	EU124713	
Aotinae sp.	AF311917							
Aotus nancymaae								
Aotus trivirgatus		FJ374685						
Aotus vociferans								
Ateles belzebuth				AB446521				
Ateles fusciceps								
Ateles geoffroyi	this study	EU733263		AB446522	FJ542219	AY740616	EU124691	EF494431
Brachyteles aracnoides								
Cacajao rubicundus					FJ542215			
Callicebus donacophilus								EF494434
Callicebus moloch		EU733265				EU124690	EU124692	
Callithrix jacchus	this study			AB446516	FJ542218			
Callithrix pygmaea						AY843512	EU124714	
Cebus apella	AF311919			AB446520				
Cebus capucinus								
Cercocebus agilis	this study				FJ542212			
Cercocebus torquatus		EU733262	EU204931	EU204937		EF113923	EU124695	
Cercopithecus cephus								
Cercopithecus mona	this study				FJ542206			
Cercopithecus neglectus								
Cercopithecus pygerythrus						AY740612		
Cercopithecus tantalus						AY593973		
Chiropotes satanas	this study							
Chlorocebus aethiops		AY623897				AY669399	EU124709	EF494430
Colobus guereza	this study	EU733267		AB446515	FJ542205		EU124712	
Erythrocebus patas		EU733260				AY740619	EU12469	EF494429
Galago senegalensis				AB446523				
Gorilla gorilla	AF311914	EU733258	ENSGGOG00000013191	AB445641	FJ542201	AY923178	EU124706	EF494426
Homo sapiens	NM_002036	NM_002759	ENSG00000174125	NM_138554	NM_003268	NM_033034	NM_006074	NM_020119
Hylobates agilis		EU733270						
Hylobates lar			EU488847	EU488850		AY923180		
Hylobates moloch								
Hylobates mulleri								

Lagothrix lagothricha		EU733266				AY843520	EU124693	EF494432
Lemur catta				AB446524				
Leontopithecus rosalia				AB446518				
Macaca fascicularis	this study							
Macaca mulatta		EU733261	ENSMMUG00000016754	AB445644	XM_001099501	DQ842020	EU124697	
Macaca nemestrina								
Mandrillus leucophaeus					FJ542210			
Mandrillus sphinx	this study							
Microcebus murinus			**	ENSMICG00000004341				
Miopithecus talapoin		EU733269						
Nasalis larvatus								
Nomascus leucogenys						EF551343	EU124698	
Otolemur garnettii								
Pan paniscus								EF494425
Pan troglodytes	AF311920	EU733256	ENSPTRG00000015985	AB445639	FJ542200	AY923177	EU124699	
Papio anubis	this study			AH008378	FJ542209	EF551342	EU124711	
Papio hamadryas	AF303532			AB446513				
Papio papio								
Pithecia pithecia	this study				FJ542214	AY843515	EU124715	
Pongo abelii							EU124707	
Pongo pygmaeus	ENSPPYG00000000662	EU733259	ENSPPYG00000014668	AB445642	FJ542202	AY923179		EF494427
Pygathrix nemaeus							EU124710	
Pygathrix roxellana								
Saguinus fuscicollis					FJ542216			
Saguinus labiatus		EU733264				AY740615	EU124694	EF494433
Saguinus midas	this study							
Saguinus oedipus			EU488856	AB446517				
Saguinus sp								
Saimiri boliviensis	AF311918					AY740614		
Saimiri sciureus				AB446519			EU124716	
Symphalangus syndactylus	this study				FJ542203		EU124708	EF494428
Tarsius syrichta			ENSTSYG00000005530		ENSTSYG00000011910		ENSTSYT00000004368	ENSTSYT00000006602
Theropithecus gelada	this study				FJ542208			
Trachypithecus cristatus				AB446514				
Trachypithecus francoisi		EU733268						
Trachypithecus obscurus								
Varecia variegata								

** = Sequence obtained from the M. murinus draft assembly at the Ensembl genome browser

Table S2. Primate Variables.

Species	Body mass (g)	Testis mass (g)	Residual testis mass	Mating system	Recoded Mating system	Group size	Population Density (individuals per km2)	Habitat	Diet
Allenopithecus nigroviridis	5000	16.96	0.248790398	MM	MM	40		1	1
Allouata caraya			0.271692455	UM and MM	amb	7.3	159	0	2
Alouatta sara			0.271692455	UM and MM	amb	4.6 (4-17)	44.8	0	2
Aotus nancymaae			-0.435145691	Mon	UM	3			
Aotus trivirgatus	1020	1.2	-0.435145691	S(m)	UM	3.4	29.7	0	1
Aotus vociferans			-0.435145691	Mon	UM	3.3	20.5	0	
Ateles belzebuth			0.0108054	MM	MM	20.8	14.6	0	1
Ateles fusciceps			0.0108054	MM	MM			0	1
Ateles geoffroyi	7940	13.4	0.0108054	M	MM	52.3	14.4	0	1
Brachyteles aracnoides				MM	MM	26	11.8	0	2
Cacajao rubicundus	3450	5.8	-0.108358493	MM	MM	(5-30) (to100)			
Callicebus donacophilus				Mon	UM			0	1
Callicebus moloch				Mon	UM	3.7	31.3	0	1
Callithrix jacchus	320	1.3	-0.060329458	S(m)	UM	8.2	1030	0	1
Callithrix pygmaea	130	0.33	-0.391519369	P?	UM	7.9	37.5	0	1
Cebus apella	2600 3000	9.1 4.64	0.026407031	M	MM	15.1	22.9	0	1
Cebus capucinus				UM	UM	16.6	10.7	0	1
Cercocebus agilis			0.257235034		MM				
Cercocebus torquatus	8680	25.1	0.257235034	M	MM	23.2	52.2	1	1
Cercopithecus cephus			-0.523909185	SP	UM	(3-35)	21.3	0	1
Cercopithecus mona			-0.523909185	UM	UM	11.4		0	1
Cercopithecus neglectus			-0.523909185	Mon/UM	UM	6.7	112	1	1
Cercopithecus pygerythrus			0.237716753	MM	MM				
Cercopithecus tantalus			0.237716753	MM	MM				
Chiropotes satanas				MM	MM	(10-30)	7.5	0	1
Chlorocebus aethiops	5290 4950	20.6 13	0.237716753	M	MM	(5-76)	66.4	2	1
Colobus guereza	10400	2.98	-0.721254241	S	UM	8.9	209	0	2
Erythrocebus patas	10000	7.2	-0.326632884	S,M UM	amb	31	0.7	2	1
Galago senegalensis	220	1.66	0.155749009	UM or MM D	amb				10
Gorilla gorilla	134000 169000	23.2 29.6	-0.559668104	S(p)	UM	15.8	1	1	2
Homo sapiens	63540 65650	50.2 40.5	-0.074634836	S(m, p) S	UM	148			
Hylobates agilis	6000	6.32	-0.233401139	P	UM	4.4		0	1
Hylobates lar	5500	5.5	-0.268231343	S(m)	UM	3.8	8.3	0	1
Hylobates moloch	6510 5440	5.7 6.1	-0.26204139	S(m) P	UM	(3-4)	7	0	1
Hylobates mulleri	5470	5.8	-0.254557957	Mon	UM	3.4	12.4	0	1
Lagothrix lagothricha	5220	11.2	0.055951369	M	MM	24.4	7	0	1
Lemur catta				MM	MM	15.9	168	1	1
Leontopithecus rosalia	550	1.48	-0.16288491	Mon P	UM	5.3		0	1
Macaca fascicularis	4787 4420	35.7 35.2	0.593217096	M	MM	(10-48)	58.5	0	1
Macaca mulatta	10430 9200	76 46.2	0.607553503	M	MM	42.1	113	1	2
Macaca nemestrina	9980	66.7	0.640747728	M	MM	40.2	34.8	1	1
Mandrillus leucophaeus	20000	41.05	0.226017852	UM	UM	87.8		1	1
Mandrillus sphinx	35000	68	0.281054183	S? UM	UM	84	5	1	1
Microcebus murinus	70	2.49	0.667756907	Um or MM D	amb	(to15)			
Miopithecus talapoin	1250	5.2	0.142027753	M	MM	92.5	71.3	0	1
Nasalis larvatus	20640	11.8	-0.324653219	ma	UM	9			2
Nomascus leucogenys			0.391609647	Mon	UM	5.2	2.9	0	1
Otolemur garnettii				UM or MM	amb	3	35	0	10
Pan paniscus			0.487235948	MM	MM	53.6	1.9	1	1
Pan troglodytes	44340 45000	118.8 139	0.487235948	M M	MM	49.2	3.5	1	1
Papio anubis	26400	93.5	0.502074952	MM	MM	50	15	2	2
Papio hamadryas	24200 20170	72.3 27.1	0.278646062	S(p)	UM	69	1.8	2	2
Papio papio	31980	88.9	0.423917535	MM	MM	(40-200)	10.9	2	1
Pithecia pithecia	1600	0.92	-0.682602459	Mon	UM	3.9	3.6	0	1
Pongo abelii			-0.221359306	S(p) D	MM	(1-3)			
Pongo pygmaeus	74640 69000	35.3 34.2	-0.221359306	S(p) D	MM	1.8	3.1	0	1
Pygathrix nemaeus				UM/MM	amb	9.8		0	2
Pygathrix roxellana				MM	MM	20			
Saguinus fuscicollis			-0.081173188	Mon (PA)	amb	7	17.6	0	1
Saguinus labiatus			-0.081173188	Mon (PA)	amb	5.8	10.4	0	1
Saguinus midas	570	1.83	-0.081173188	Mon (PA)	amb	4.7	32.9	0	1
Saguinus oedipus	501 520	1.48 3.4	0.076105368	S(m) M	amb	5.3	35.8	0	1
Saimiri boliviensis			0.064928262	MM	MM	20		0	10
Saimiri sciureus	779 780	3.13 3.2	0.064928262	M	MM	43.1	86.7	0	1
Symphalangus syndactylus			0.019953481	Mon	UM	4	1.4	0	2
Tarsius syrichta				UM and MM	amb	>2			10
Theropithecus gelada	20400	21.5 17.1	-0.107546965	S(p)	UM	137.1	78.6	2	2
Trachypithecus cristatus	6580 6700 6520	6.2 6.3 6.2	-0.268794842	S(p) UM	UM	30.6	145	0	2
Trachypithecus francoisi			-0.268794842	SP	UM	14.6	5	1	2
Trachypithecus obscurus	7450	4.8	-0.416372429	S(p) UM and MM	amb	9	31	0	2
Varecia variegata	4750	17.17	0.269181378	Mon P	UM	8.3	175	0	1

Codes for body mass, testis mass and mating system

red text	M=multimale, promiscuous or polygynous/promiscuous; (M)=expectation based on closely rel. sp.; S=single-male; (m)=monogamous, (p)=extremely polygynous (Kenagy and Trombulak 1986)
black text	pa=polyandrous; ma=monoandrous (Harcourt et al. 1991)
green text	P=paired; M=multi-male; S=single-male; D=dispersed (Harcourt et al 1995)
blue text	UM=unimale; MM=multimale, Mon=monogamous, PA=facultative polyandrous (Linderfors and Tullberg 1998)
black text	SP=single partner, MP=multiple partner (Anderson et al. 2004)
black text	inferred based in clade

Codes for group size, density, terrestriality and diet

black text	(Nunn 2002, Nunn et al. 2003)
black text	(Rowe 1996)
black text	(Semple at al. 2002)
Habitat codes	0= arboreal, 1=terrestrial in wooded habitat, 2=terrestrial in open habitat
Diet codes	0= frugivorous, 1 folivorous, 10=insectivorous

Other codes

	LH variables: For Allouata sara, info from A. seniculus was used; For Cacajao rubicundus, info from C. calvus was used.
	testis size from other species in the genus or average if available

Notes

Testis size and body size: If more than one measurement was available, the average was used.
Mating system: Was defined as ambiguous if different studies reported differences.
Group size: If more than one value was reported, the average was used; if a range was reported, the middle point of the range was used.