Radaev S, Rostro B, Brooks AG, Colonna M, Sun PD (2001) Conformational plasticity revealed by the cocrystal structure of NKG2D and its class I MHC-like ligand ULBP3. Immunity 5:1039–1049

Radosavljevic M, Cuillerier B, Wilson MJ, Clement O, Wicker S, Gilfillan S, Beck S, Trowsdale J, Bahram S (2001) A cluster of ten novel MHC class I related genes on human chromosome 6q24.2-q25.3. Genomics 79:114–123

Raulet DH (2003) Roles of the NKG2D immunoreceptor and its ligands. Nat Rev Immunol 3:781–790

Romphruk AV, Romphruk A, Naruse TK, Raroengjai S, Puapairoj C, Inoko H, Leelayuwat C (2009) Polymorphisms of NKG2D ligands: diverse RAET1/ULBP genes in northeastern Thais. Immunogenetics 61:611–617

Seo JW, Bontrop R, Walter L, Günther E (1999) Major histocompatibility complex-linked MIC genes in rhesus macaques and other primates. Immunogenetics 50:358–362

Seo JW, Walter L, Günther E (2001) Genomic analysis of MIC genes in rhesus macaques. Tissue Antigens 58:159–165

Ward J, Bonaparte M, Sacks J, Guterman J, Fogli M, Mavilio D, Barker E (2007) HIV modulates the expression of ligands important in triggering natural killer cell cytotoxic responses on infected primary T-cell blasts. Blood 110:1207–1214

Wu J, Song Y, Bakker ABH, Bauer S, Spies T, Lanier LL, Phillips JH (1999) An activating immune receptor complex formed by NKG2D and DAP 10. Science 285:730–732

Immunogenetics (2012) 64:15–29
DOI 10.1007/s00251-011-0556-2

ORIGINAL PAPER

Positive selection of Toll-like receptor 2 polymorphisms in two closely related old world monkey species, rhesus and Japanese macaques

Akiko Takaki · Akiko Yamazaki · Tomoyuki Maekawa · Hiroki Shibata · Kenji Hirayama · Akinori Kimura · Hirohisa Hirai · Michio Yasunami

Received: 7 January 2011 / Accepted: 21 June 2011 / Published online: 9 July 2011

Abstract Toll-like receptor 2 (TLR2) plays an important role in the recognition of a variety of pathogenic microbes. In the present study, we compared polymorphisms of *TLR2* locus in two closely related old world monkey species, rhesus monkey (*Macaca mulatta*) and Japanese monkey (*Macaca fuscata*). By nucleotide sequencing of the third exon of *TLR2* gene from 21 to 35 respective individuals, we could assign 17 haplotype combinations of 17 coding SNPs of ten non-synonymous and seven synonymous substitutions. A non-synonymous substitution at codon position 326 appeared to be differentially fixed in each species, asparagine for *M. mulatta* whereas tyrosine for *M. fuscata*, and may contribute to certain functional properties because it locates in the region contributing to ligand binding and interaction with dimerization partner of TLR2-TLR1 heterodimeric complex. Although *TLR2* alleles have diverged to similar extent in both species, they have evolved in significantly different ways; *TLR2* of *M. fuscata* has undergone purifying selection while the membrane-proximal part of the extracellular domain of *M. mulatta TLR2* exhibits higher rates of non-synonymous substitutions, indicating a trace of Darwinian positive selection.

Keywords Innate immunity · TLR · Polymorphism · Nonhuman primate · Molecular evolution · Reporter gene assay

Electronic supplementary material The online version of this article (doi:10.1007/s00251-011-0556-2) contains supplementary material, which is available to authorized users.

A. Takaki · A. Yamazaki · H. Shibata · K. Hirayama · M. Yasunami (✉)
Department of Immunogenetics, Institute of Tropical Medicine, Nagasaki University,
1-12-4 Sakamoto,
Nagasaki 852-8523, Japan
e-mail: yasunami@nagasaki-u.ac.jp

A. Takaki · T. Maekawa · M. Yasunami
Department of Clinical Medicine, Institute of Tropical Medicine, Nagasaki University,
Nagasaki, Japan

A. Kimura
Department of Molecular Pathogenesis,
Medical Research Institute, Tokyo Medical and Dental University,
Tokyo, Japan

H. Hirai
Department of Cellular and Molecular Biology,
Primate Research Institute, Kyoto University,
Kyoto, Japan

Introduction

The rhesus macaque, *Macaca mulatta*, is one of the best known old world monkeys and has been used for various biomedical researches as a nonhuman primate model including infections of simian immunodeficiency virus (Ling et al. 2002; Matano et al. 2004) and *Mycobacterium tuberculosis* (McMurray 2000; Huang et al. 2007). *M. mulatta* belongs to the primate family Cercopithecidae that shares the last common ancestor of approximately 25 million years ago (Mya) with human and hominoids (Kumar and Hedges 1998). According to this fact, nucleotide sequence similarity between humans and *M. mulatta* has been maintained as high as 93% in average (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007). Analyses of molecular evolution of mitochondrial and nuclear DNA among species of genus *Macaca* estimate the divergence between rhesus and

cynomolgus macaque, *Macaca fascicularis* at 0.9–2.5 Mya (Hayasaka et al. 1996; Blancher et al. 2008; Osada et al. 2008; Stevison and Kohn 2009). Japanese monkey, *Macaca fuscata* and Taiwanese monkey, *Macaca cyclopis* had been derived from the rhesus lineage relatively recently, and the geographical isolation is fundamental for the diversification of these species (Smith et al. 2007). *M. mulatta* has a relatively broad geographical distribution from Afghanistan across Asia to the Chinese shore of the Pacific Ocean (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), where the tropical infectious agents such as *Plasmodium* species and arthropod-borne viruses are circulating and may influence the host genome as selection pressure. To date, a number of comparative studies on primate genome including *Macaca* species have been conducted. Among them, immune-related genes such as *TNFA* (Baena et al. 2007), eosinophil cationic protein (*ECP*), and eosinophil-derived neurotoxin (*EDN*) ribonuclease (Zhang et al. 1998) are shown to be evolved under positive Darwinian selection (or diversifying selection), suggesting the presence of a differential pressure between primate species.

Toll-like receptors (TLRs) are the first line of the host defense mechanisms against pathogenic microorganisms as pattern-recognition receptors (PRRs) (Akira et al. 2006). A number of single nucleotide polymorphisms (SNPs) have been found in the components of the TLR signaling pathway in humans. Alterations of the structure of these signaling molecules are often associated with susceptibility to various infectious diseases. In the case of human TLR2 that binds to peptidoglycan and lipoteichoic acid derived from Gram-positive bacteria as well as a variety of macromolecules from other microbes (Texereau et al. 2005), a non-synonymous substitution of glutamine for arginine at amino acid position 753 (Arg753Gln) was shown to be a risk factor of developing tuberculosis (Bochud et al. 2003; Ogus et al. 2004) and septic shock (Lorenz et al. 2000), whereas, have a protective effect against the development of Lyme disease (Schröder et al. 2005). It is hypothesized that TLR2 is also variable in *Macaca* species and exhibits a trace of molecular evolution events which is conferred by possible differences in individual response to various infectious agents. Previous reports on molecular evolution of TLRs of primates, which support the functional conservation of intra-cellular signal transduction machinery involving a common functional cytoplasmic Toll/Interleukin-1 receptor (TIR) domain during mammalian evolution, also suggested the presence of the positive Darwinian selection on the extracellular domain of TLR4 in *Macaca* species (Sanghavi et al. 2004; Nakajima et al. 2008).

In the present study, SNPs in the coding sequence of TLR2 from two closely related *Macaca* species, *M. mulatta* and *M. fuscata* were investigated. By nucleotide sequencing of 21 *M. mulatta* and 35 *M. fuscata* individuals, ten non-synonymous and seven synonymous substitutions were identified in the coding region. These 17 SNPs compose 17 haplotype combinations (or alleles) existing in the examined population. The ratio of rates of non-synonymous versus synonymous substitutions suggested the conservation of overall amino acid sequence except for a part of extracellular domain of *M. mulatta* TLR2, where multiple amino acid substitutions have taken place to give rise different alleles. Although the functional difference of TLR2 alleles was not evident by in vitro transfection study of expression vector to HEK293 cells, molecular modeling suggested that modified receptor-ligand interaction conferred by amino acid substitution in *M. mulatta* TLR2 is a driving force of diversifying evolution.

Materials and methods

Genomic DNA samples

Individual genomic DNA of *M. mulatta* were prepared from B lymphoblastoid cell lines which had been established from peripheral blood of mutually unrelated founders of breeding colony originated from wild population in Myanmar and Laos as described before (Takahashi-Tanaka et al. 2007). Samples of peripheral blood of *M. fuscata* were collected from three isolated colonies originated from different wild populations in Japan maintained at Primate Research Institute, Kyoto University, after the institutional review of experimental procedures. Genomic DNA of leukocytes was isolated using Wizard® Genomic DNA Purification Kit (Promega).

Nucleotide sequencing of coding region of TLR2

The coding sequence of TLR2 was amplified by PCR using KOD FX DNA polymerase (TOYOBO) with a pair of oligonucletides, TLR2 exon 3-forward (5′-ATTAGAAT TACGATATGCTGTC-3′) and TLR2 exon 3-reverse (5′-ATGACGGTACATCCACGTAG-3′) as primers, essentially according to manufacturer's recommendations. Sequencing reaction was performed using BigDye® Terminators v1.1 Cycle Sequencing Kit (Applied Biosystems) and analyzed by ABI3730xl automated DNA sequencer (Applied Biosystems).

Transfection of TLR2 expression vectors

PCR products including the entire coding sequence of TLR2 were ligated to pGEM®-TEasy (Promega) with T4 DNA ligase after treatment with A-attachment Mix (TOYOBO). After the confirmation of nucleotide sequence

of plasmid clones, the DNA fragment containing entire coding region for TLR2 was isolated by cleavage with restriction enzymes and inserted to pcDNA3.1/Hygro(+) expression vector (Invitrogen). Luciferase reporter for nuclear factor kappa B (NF-κB) activity, pGL4.22-6xκB was constructed by inserting DNA fragment containing 6 times tandem repeats of NF-κB binding sequence and TATA box (Shibata et al. 2006) in pGL4.22[luc2CP/Puro] promoterless firefly luciferase reporter vector (Promega). All expression plasmid DNA and reporter plasmid DNA were prepared by Illustra™ Plasmid Prep Midi Flow kit (GE Healthcare) and confirmed virtually endotoxin free.

HEK293 cells were cultured in Dulbecco's MEM (Wako) supplemented with 10% fetal bovine serum (GIBCO) and 100 U/ml penicillin–100 μg/ml streptomycin (Invitrogen) at 37°C in the presence of 5% CO_2. Synthetic bacterial lipoproteins as TLR2 agonists, Pam2CSK4 and Pam3CSK4 were purchased from InvivoGen and applied at 10 and 100 ng/ml, respectively. Transfection of DNA was performed as follows: HEK293 cells were inoculated in 35-mm tissue culture dishes at 4×10^5 cells/dish on the previous day, transfected with a mixture of 250 ng of each pcDNA3.1-TLR2 expression plasmid, 100 ng of pGL4.22-6xκB, and 20 ng pRL-TK internal control (Promega) with FuGENE® HD Transfection Reagent (Roche), then at 42 h after the addition of DNA, stimulated with TLR2 agonists for 6 h. Cells were washed with 1 ml of PBS and suspended in 100 μl of 1× Passive Lysis Buffer (Promega). Luciferase activities were determined using Dual-Luciferase® Reporter Assay System (Promega) by Wallac 1420 Multilabel Counter ARVO MA (PerkinElmer). Transfection was set up in triplicate to evaluate experimental variations and repeated at least twice to confirm the results. Difference in luciferase activity was statistically examined by ANOVA and Student's *t* test.

Analysis of molecular evolution

We estimate values of the number of segregating sites (S), number of haplotypes (H), haplotype diversity (Hd), nucleotide diversities (π), nucleotide polymorphism (θ), K_a, K_s, K_a/K_s ratio, and Tajima's D (Tajima 1989) by using DnaSP ver.5.10.00 (Librado and Rozas 2009). Genetic distances for non-synonymous substitution (d_N) and synonymous substitution (d_S) and the d_N-d_S difference were calculated according to Nei and Gojobori's method (Nei and Gojobori 1986) with Jukes and Canter's correction by MEGA4 (Kumar et al. 2008). Z tests for positive and purifying selections were also carried out on MEGA4, by which standard errors were produced with 500 bootstrap replications. Results were considered statistically significant when $p<0.05$. A neighbor-joining tree (Saitou and Nei 1987) was constructed on the basis of genetic distances from alignments of the coding sequence of TLR2 alleles, estimated by Kimura's two-parameter method (Kimura 1980). Difference in frequencies of non-synonymous and synonymous substitutions was evaluated by Fisher's exact test (Zhang et al. 1998). Further, codon-wise evaluation for positive selection was conducted by employing statistical test implemented in PAML 4 package (Yang 2007) based on maximum likelihood estimation method.

Homology modeling of molecular structure of *Macaca* TLR2-TLR1

The optimal structures of macaque TLR2-TLR1 variants were determined by means of homology modeling algorithm implemented in Molecular Operation Environment (MOE) 2010.10 (Chemical Computing Group Inc, http://www.chemcomp.com/) using a template structure of human TLR2-TLR1 heterodimer bound with Pam3CSK4 (PDB accession # 2Z7X). The PDB data contained structural data for human TLR2 (27–506) and human TLR1 (25–475) fused to 67 amino acid-long peptide derived from inshore hagfish *Eptatretus* VLRB.61. The structure of ternary complex of human TLR2 (27–506)-human TLR1 (25–475) dimer and a lipopeptide ligand Pam3CSK4 was optimized after trimming of C-terminal fusion peptides. Molecular stability was expressed by the sum of potential energy between atoms in kcal/mol. Amino acid sequence of macaque TLR1 (25–475) was retrieved from GenBank accession # NM_001130424. London dG score, an index for affinity between receptor and ligand within the binding pocket, was obtained for each optimized complex structure by Lig-X algorithm suite implemented in MOE 2010.10.

Results

Identification of SNPs in TLR2 coding sequence from *M. mulatta* and *M. fuscata*

In the present study, nucleotide sequence of the 2,355-bp long entire coding region of TLR2 encoded within the third exon was determined for of 21 *M. mulatta* and 35 *M. fuscata* individuals. By the comparison of *M. mulatta* sequences, six non-synonymous and three synonymous substitutions were identified, while three non-synonymous and five synonymous substitutions were found in *M. fuscata* (Table 1; Fig. 1). One synonymous substitution, GTG(Val)>GTT(Val) at codon 373 was common to these two species. Further, all 21 *M. mulatta* individuals were homozygous for AAT(Asn) while all *M. fuscata* examined were homozygous for TAT(Tyr) at codon 326. Collectively, ten non-synonymous and seven synonymous substitutions were found in this population.

Table 1 Haplotype of 17 SNPs found in 21 *Macaca mulatta* and 35 *Macaca fuscata* individuals in the present study

Haplotype	Nucleotide position																
	93 Codon	292	570	664	703[a]	976[a]	1119[a]	1176	1214	1246	1359[a]	1507	1633	1667	1743	1908	2344
	31	98	190	222	235[a]	326[a]	373[a]	392	405	416	453[a]	503	545	556	581	636	782
M. mulatta (frequency in 2*N*=42)																	
Mamu-Hap1 (17)	GAC (Asp)	TCC (Ser)	GAG (Glu)	AGT (Ser)	AAC (Asn)	AAT (Asn)	GTG (Val)	TTA (Leu)	ACC (Thr)	ACT (Thr)	GGC (Gly)	CTG (Leu)	ACT (Thr)	GCT (Ala)	TCA (Ser)	AAC (Asn)	ATA (Ile)
Mamu-Hap2 (10)	–	–	–	–	–	–	–	TTG (Leu)	–	–	–	GTG (Val)	–	–	–	–	–
Mamu-Hap3 (2)	–	–	–	–	–	–	–	–	–	GCT (Ala)	–	–	–	–	–	–	–
Mamu-Hap4 (2)	GAT (Asp)	–	–	–	–	–	–	TTG (Leu)	–	–	–	GTG (Val)	GCT (Ala)	–	–	–	–
Mamu-Hap5 (5)	–	–	–	–	–	–	–	–	–	–	–	GTG (Val)	–	–	–	–	–
Mamu-Hap6 (2)	–	–	–	–	–	–	–	–	–	–	–	GTG (Val)	–	GTT (Val)	–	–	–
Mamu-Hap7 (1)	–	–	–	–	–	–	GTT (Val)	–	ATC (Ile)	–	–	GTG (Val)	–	GTT (Val)	–	–	–
Mamu-Hap8 (1)	–	CCC (Pro)	–	–	–	–	–	–	–	–	–	GTG (Val)	–	–	–	–	–
Mamu-Hap9 (1)	–	–	–	–	–	–	–	–	–	–	–	GTG (Val)	GCT (Ala)	–	–	–	–
Mamu-Hap10 (1)	–	–	–	–	–	–	GTT (Val)	–	–	GCT (Ala)	–	–	–	–	–	–	–
M. fuscata (frequency in 2*N*=70)																	
Mafu-Hap1 (50)	GAC (Asp)	TCC (Ser)	GAG (Glu)	AGT (Ser)	GAC (Asp)	TAT (Tyr)	GTT (Val)	TTA (Leu)	ACC (Thr)	ACT (Thr)	GGT (Gly)	CTG (Leu)	ACT (Thr)	GCT (Ala)	TCA (Ser)	AAC (Asn)	ATA (Ile)
Mafu-Hap2 (4)	–	–	GAA (Glu)	–	AAC (Asn)	–	GTG (Val)	–	–	–	GGC (Gly)	–	–	–	–	–	–
Mafu-Hap3 (1)	–	–	–	–	AAC (Asn)	–	GTG (Val)	–	–	–	GGC (Gly)	–	–	–	–	–	–
Mafu-Hap4 (2)	–	–	–	GGT (Gly)	–	–	–	–	–	–	–	–	–	–	TCG (Ser)	–	–
Mafu-Hap5 (5)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	AAT (Asn)	–
Mafu-Hap6 (7)	–	–	–	GGT (Gly)	–	–	–	–	–	–	–	–	–	–	–	–	–
Mafu-Hap7 (1)	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	TTA (Leu)
Human[b]	GAC (Asp)	TCC (Ser)	GAG (Glu)	AGT (Ser)	GAC (Asp)	TAT (Tyr)	GTT (Val)	TTA (Leu)	ACC (Thr)	ACT (Thr)	GGT (Gly)	GTG (Leu)	ACT (Thr)	ATT (Ile)	TCA (Ser)	AGC (Ser)	ATA (Ile)

Polymorphic nucleotide in each codon is underscored. Frequency of haplotype appeared in the population is shown in parenthesis

[a] The positions where the major haplotypes for *M. mulatta* (Mamu-Hap1) and *M. fuscata* (Mafu-Hap1) are different from each other

[b] Only the human sequence (GenBank accession #NM_003264) corresponding to polymorphic codons among *Macaca* species are shown to tell whether is more probable to be ancestral between *Macaca* alleles

The extracellular domain of TLR2 is mainly composed with the leucine-rich repeat (LRR) modules and can be divided into three sub-domains; N-terminal, central, and C-terminal sub-domains with their unique β-sheet conformations (Fig. 1). The structure of extracellular domains of triacylated lipopeptide-bound TLR2-TLR1 heterodimer and those of diacylated lipopeptide-bound TLR2-TLR6 heterodimer has been revealed by X-ray diffraction analysis (Jin et al. 2007; Kang et al. 2009). Amino acid residues which contribute to the binding of ligands and the interaction to dimerization partner were mapped in the middle of extracellular domain. The distribution of polymorphic sites in terms of structural sub-domains was shown in Fig. 1. It is of note that the abovementioned inter-species substitution at codon 326, Asn in *M. mulatta* versus Tyr in *M. fuscata*, is located within ligand binding and dimerization domain for TLR2-TLR1 heterodimer which may contribute to certain functional properties (Fig. 1b).

Upon the inference of the combination of SNPs by PHASE program implemented in DnaSP package (see below), a major allele of *M. mulatta* population (Mamu-Hap1 in Table 1) was obtained. Mamu-Hap1 had four nucleotide substitutions, GTT(Val)>GTG(Val) at codon 373, ATC (Ile)>ACC(Thr) at codon 405, GTG(Val)>CTG (Leu) at codon 503 and GTT(Val)>GCT(Ala) at codon 556, from the *M. mulatta TLR2* sequence reported previously, GenBank accession #AB445630 (Nakajima et al. 2008), which is identical to Mamu-Hap7 in the present study. The homology between *M. mulatta* and humans is 97.2% (2,289/2,355 bp) at nucleotide level and 95.8% (751/784) at amino acid level. Because general homology in nucleotide sequences between *M. mulatta* and humans is 93.94% identical in average (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), TLR2 is one of the well-conserved genes during primate evolution. A major allele of *M. fuscata* (Mafu-Hap1 in Table 1) was also inferred (GenBank accession, #AB607964) and had four nucleotide substitutions from Mamu-Hap1: AAC(Asn)>GAC(Asp) at codon 235, AAT(Asn)>TAT(Tyr) at codon 326, GTG(Val)>GTT(Val) at codon 373 and GGC(Gly)>GGT(Gly) at codon 453. Another less-frequent Mafu-Hap3 had only one substitution, AAT(Asn)>TAT(Tyr) at codon 326 from Mamu-Hap1. Figure 2 explains genealogy of allelic variants. Intra-specific variants are connected by as many as five and six mutation events in *M. mulatta* and *M. fuscata*, respectively, while Mamu-Hap1 and Mafu-Hap3 differ by one nucleotide substitution as mentioned above. In addition, a synonymous substitution at codon 373 was

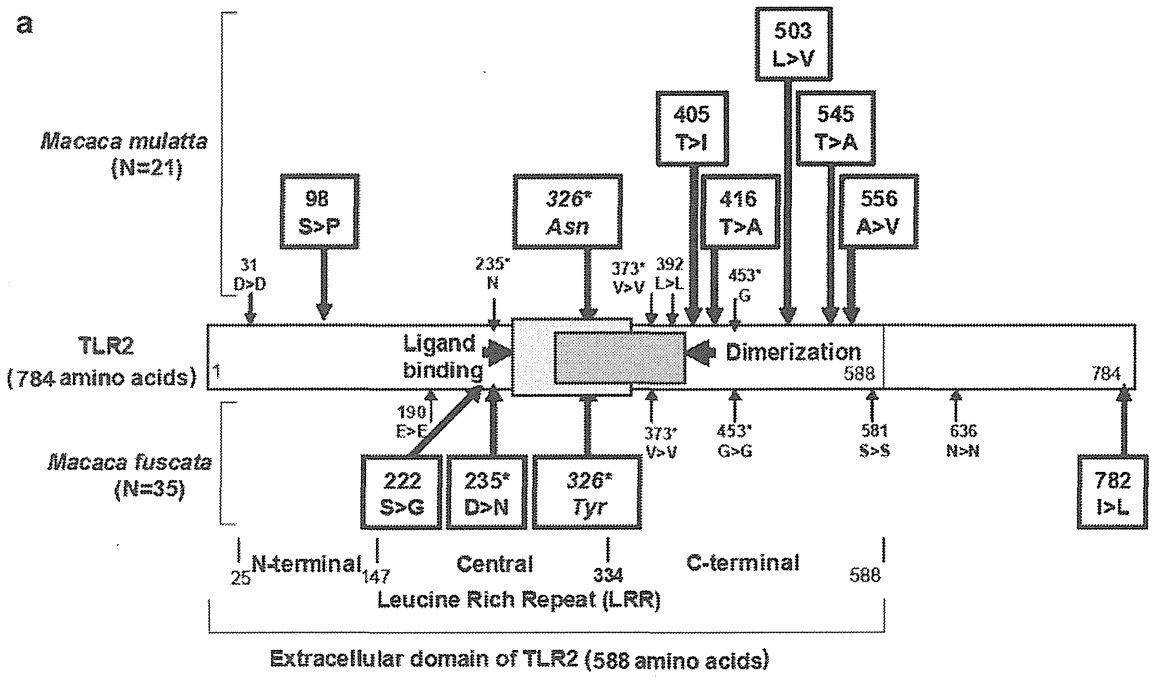

Fig. 1 The distribution of SNPs found in *M. mulatta* and *M. fuscata* populations. **a** The position of codons with synonymous (*plain*) and non-synonymous (*boxed*) SNPs is shown in schematic representation of TLR2; SNPs in *M. mulatta* and *M. fuscata* populations are the upper and the lower halves, respectively. Domain and sub-domain structures are also illustrated. The regions contributing to ligand binding (codon 266–359) and interaction with dimerization partner (codon 318–398 for TLR2-TLR1 and codon 318–404 for TLR2-TLR6) are depicted as overlapping *filled boxes* in the middle of extracellular domain (Jin et al. 2007; Kang et al. 2009). **b** Structure and amino acid sequence of TLR2. Leucine-rich domain (LRR)s at N terminus, #1 to #20 and at C terminus, and transmembrane and cytoplasmic regions (*TM+CY*) are shown above the sequence. Amino acid residues involved in ligand binding (*l*) and dimerization interface (*i*) are indicated. Sequence variations between human and *M. mulatta* (*asterisk*), among *M. mulatta* and among *M. fuscata* (var. site) are also indicated

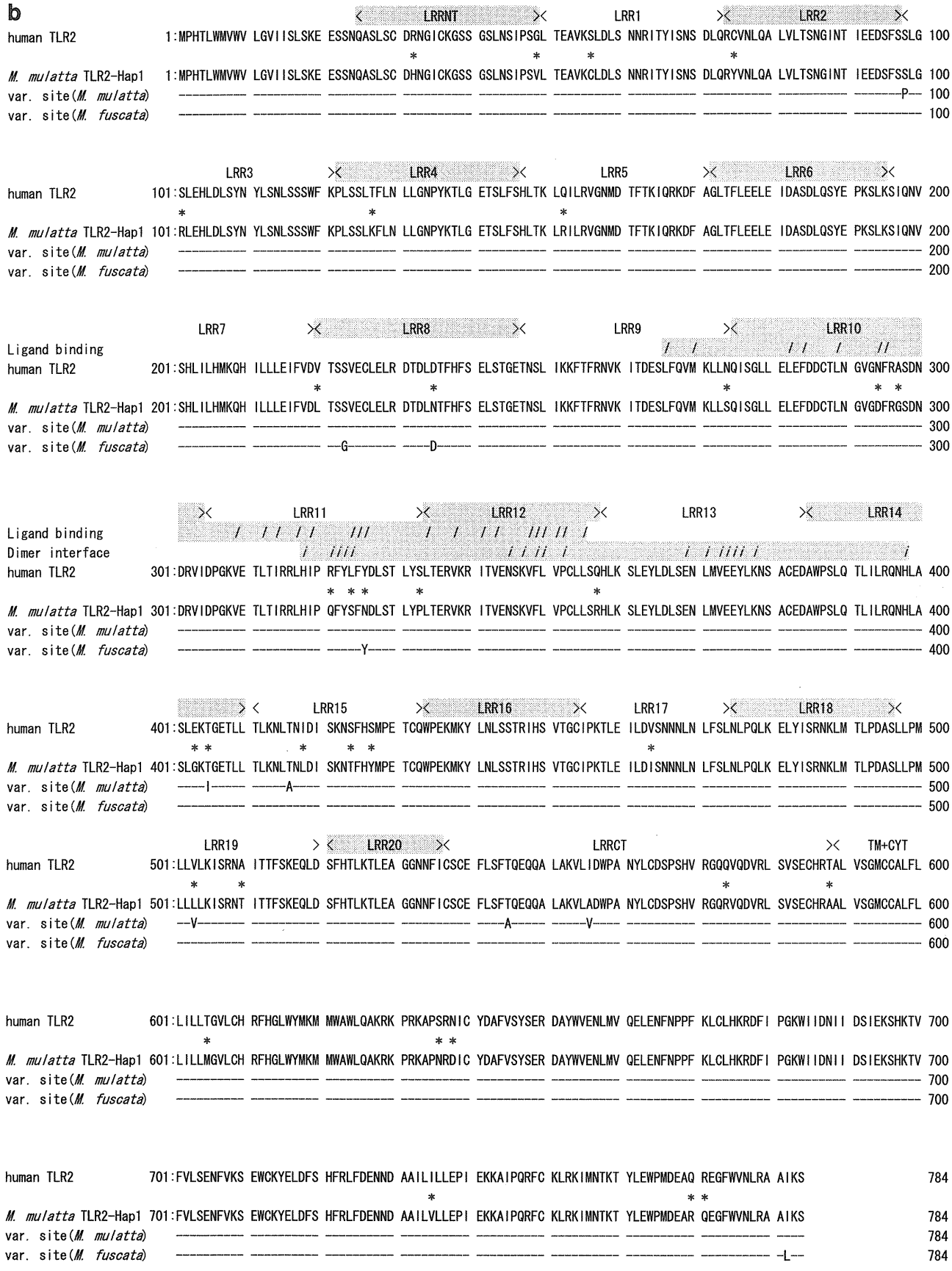

Fig. 1 (continued)

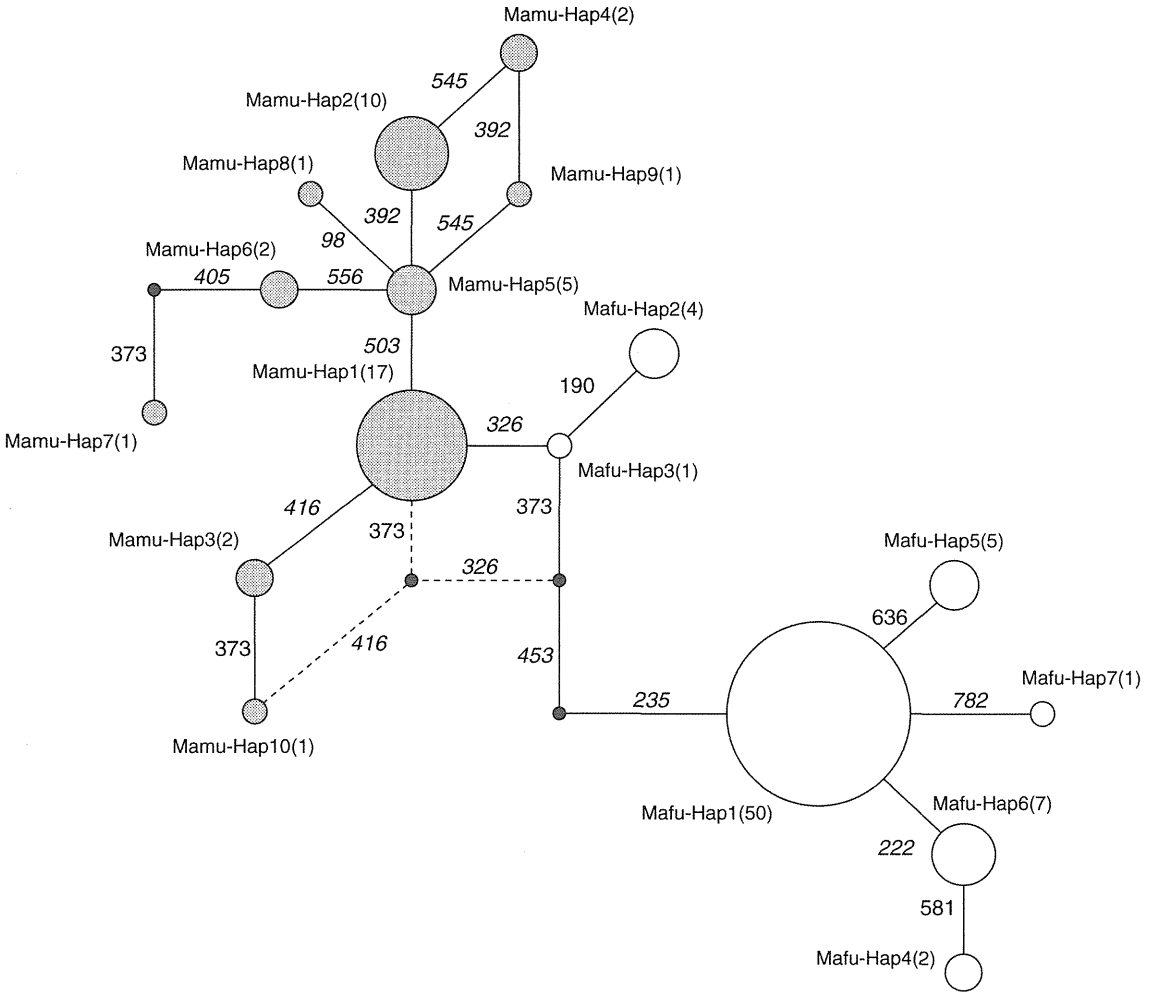

Fig. 2 Genealogical relationship of *Macaca* TLR2 alleles. A genealogical pathway of allelic variants is illustrated by network. *Gray* and *open circles* are allelic variants of *M. mulatta* and *M. fuscata* TLR2, respectively. Two variants differ by one nucleotide substitution are connected by *bar*. *Numbers near the bars* are codon positions; non-synonymous substitution is shown in *italic* style. Intra-specific variants are connected by as many as five and six mutation events in *M. mulatta* and *M. fuscata*, respectively, while Mamu-Hap1 and Mafu-Hap3 differ by one nucleotide substitution. It is of note that a synonymous substitution at codon 373 is commonly found in two species

commonly found in two species. Thus the difference between variants across the species is not more than the maximal difference within a species for certain combinations.

Molecular evolution of TLR2

In order to evaluate the allelic divergence of *Macaca TLR2* locus, we applied PHASE program to infer the haplotypes of 17 SNPs observed in two *Macaca* species. Consequently, ten haplotypes for *M. mulatta* and seven haplotypes for *M. fuscata* were identified (Table 1). Presence of these haplotypes was confirmed by nucleotide sequencing of plasmid clones. A phylogenetic tree (Fig. 3) was drawn by neighbor-joining method according to sequence alignments shown in Table 1. Because of amino acid substitution at codon 326, alleles of *M. mulatta* and *M. fuscata* were positioned in two separate lineages.

S, *H*, *Hd*, π, θ, average codon-based evolutionary divergence (d_S, d_N, and d_N, d_S), and Tajima's D were calculated for overall and sub-domains of *M. mulatta* and *M. fuscata* TLR2 (Table 2). Sliding window plots for Tajima's D value for each species and K_a/K_s ratio between two species are also presented (Figs. 4 and 5). When the statistical significance was examined by *Z* test, the entire TLR2 coding sequence of *M. fuscata* gave a significant result with a codon-based test (d_N, d_s) for purifying selection (p=0.0435), whereas became not significant if the extracellular domain was separately examined (p= 0.061). But further breaking down of the extracellular domain into three parts, C-terminal LRR domain underwent purifying selection by codon-based test (p=0.0406). In contrast, a part of the extracellular domain C-terminal to ligand binding and dimerization domain of *M. mulatta* TLR2 was rich in non-synonymous substitutions with a

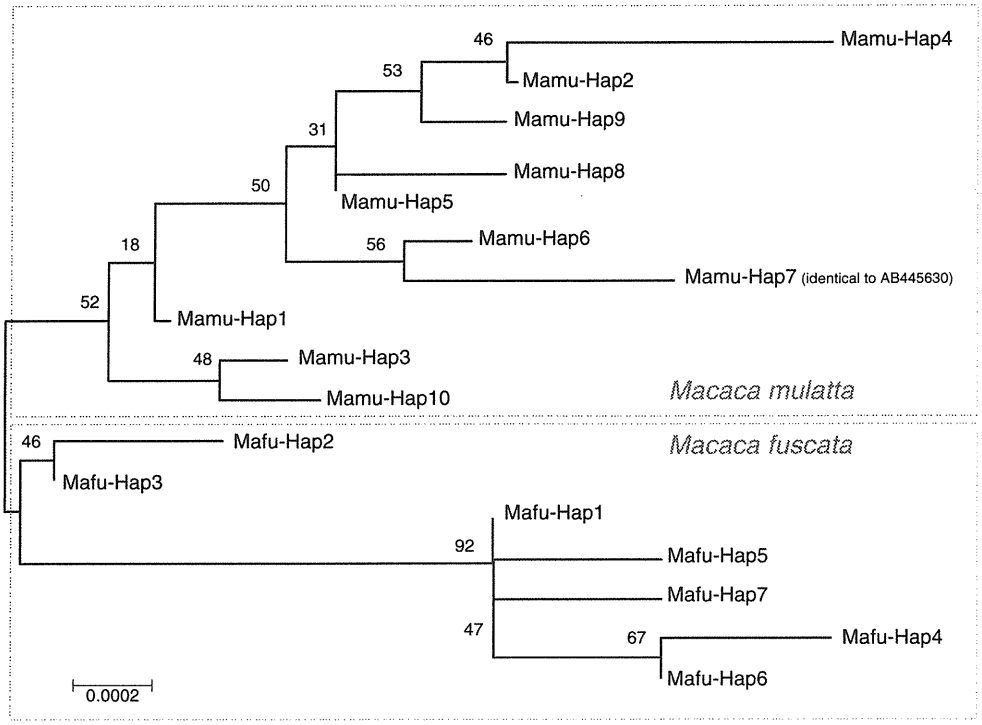

Fig. 3 Phylogenetic tree for *Macaca* TLR2 alleles. A phylogenetic tree is drawn by neighbor-joining method based on a distance matrix of Kimura's two-parameter evolutional model using MEGA4.0.2. Bootstrap percent probability for 500 replications is shown at each node. The *scale bar* represents a genetic distance of 0.0002

significant result for codon-based test for positive or diversifying selection (p=0.0362). Sliding window plot of Tajima's D for *M. mulatta* TLR2 (Fig. 4a) shows clearly different pattern from that of *M. fuscata* TLR2 (Fig. 4b): the N terminus of *M. mulatta* TLR2 and the membrane-proximal region of extracellular domain of *M. fuscata* TLR2 exhibit weak tendency of purifying selection because of the presence of synonymous substitutions, while the membrane-proximal region of extracellular domain of *M. mulatta* TLR2 shows positive D value suggesting the operation of positive selection. When all sequence of two species ($2N$=42 for ten *M. mulatta* variants and $2N$=70 for seven *M. fuscata* variants) were taken into account, average codon-based evolutionary divergence indicated by the value of K_a/K_s ratio is highest in the part of central LRR, mostly attributed to amino acid substitution at codon 326 (Table 3) while two peaks of K_a/K_s ratio corresponding to the central LRR and the amino acid substitutions in the membrane-proximal part of extracellular domain of *M. mulatta* TLR2 appear when the sliding window is narrowed to 350-bp range (Fig. 5). When the frequency of non-synonymous substitution in possible sites (five in 387) within in the membrane-proximal part of extracellular domain (codon 401 to 566) of *M. mulatta* TLR2 was compared with that of synonymous substitution (0 in 111), they were not significantly different (p=0.358; Table 4, comparison A), but the number of non-synonymous substitutions is over-represented in *M. mulatta* than *M. fuscata* (5 vs. 0) in comparison to that of synonymous substitutions (0 vs.2) within the same region (p=0.048; Table 4, comparison B), suggesting the presence of differential selection pressure on these two species. Further, codon-wise evaluation for non-synonymous and synonymous substitutions was conducted for extracellular domain of *M. mulatta* TLR2 by employing statistical test implemented in PAML 4 package based on maximum likelihood estimation method. Likelihood statistics under different molecular evolution models were presented in Table 5. By comparison of likelihoods between neutral and selection site models, we could obtain a weak but significant signal for positive selection at codon position 545. This position is not known to be involved in ligand binding or dimerization interface in the diacylated lipoprotein-bound structure of TLR2-TLR6 dimer or the triacylated lipoprotein-bound structure of TLR2-TLR1 dimer (Fig. 1b).

Functional analysis for TLR2 alleles

In order to clarify whether amino acid substitutions in *Macaca* TLR2 alleles can bring any functional alteration, an experimental system was designed to measure the downstream transcriptional activation of NF-κB-dependent synthetic enhancer-promoter upon the stimulation with TLR2 agonists. It is of importance to examine functional integrity of variants to rule out pseudogenization of TLR-2 in *Macaca* lineage as one of possible cause of rapid accumulation of non-synonymous mutations. We chose HEK293 cells as recipients of DNA transfection assay because it has been proven to be lower intrinsic TLR2

Table 2 Polymorphism and nucleotide diversity in the coding sequence of TLR2 locus

Domain[a]	Range (bp)	*S*	*H*	*Hd*	π ($\times 10^{-4}$)	θ ($\times 10^{-4}$)	d_S (±SE)	d_N (±SE)	d_N-d_S (±SE)	Tajima's D
Macaca mulatta (n=21)										
Total	1–2,355	9	10	0.775	6.9	8.9	0.0011±0.0008	0.0006±0.0003	−0.0006±0.0008	−0.65546
Extracellular	1–1,764	9	10	0.775	9.2	11.9	0.0015±0.0010	0.0007±0.0004	−0.0008±0.0011	−0.65546
N-terminal LRR	73–438	2	3	0.138	3.8	12.7	0.0010±0.0010	0.0002±0.0002	−0.0009±0.0011	−1.30048
Central LRR	439–999	0	1	0.000	0.0	n.d.	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	n.d.
C-terminal LRR	1,000–1,764	7	9	0.769	19.3	21.3	0.0029±0.0024	0.0016±0.0009	−0.0013±0.0027	−0.25176
N-terminal part	1–795	2	3	0.138	1.8	5.8	0.0005±0.0005	0.0001±0.0001	−0.0004±0.0005	−1.30048
Lig-bind/dimer	796–1,194	2	3	0.483	12.8	11.6	0.0057±0.0048	0.0000±0.0000	−0.0057±0.0046	0.18543
C-terminal part	1,195–1,764	5	6	0.695	17.0	20.4	0.0000±0.0000	0.0022±0.0012	0.0022±0.0012*	−0.42381
TM-Cyt	1,765–2,355	0	1	0.000	0.0	n.d.	0.0000±0.0000	0.0000±0.0000	0.0000±0.0000	n.d.
Macaca fuscata (n=35)										
Total	1–2,355	8	7	0.477	4.1	7.0	0.0011±0.0005	0.0002±0.0001	−0.0009±0.0005**	−1.08369
Extracellular	1–1,764	6	5	0.351	4.5	7.1	0.0011±0.0005	0.0003±0.0002	−0.0008±0.0006	−0.85999
N-terminal LRR	73–438	0	1	0.000	0.0	n.d.	00000±0.0000	0.0000±0.0000	0.0000±0.0000	n.d.
Central LRR	439–999	3	4	0.345	8.4	11.1	0.0009±0.0009	0.0008±0.0006	0.0000±0.0010	−0.46715
C-terminal LRR	1,000–1,764	3	3	0.187	4.3	8.1	0.0019±0.0011	0.0000±0.0000	−0.0019±0.0011***	−0.91696
N-terminal part	1–795	3	4	0.345	5.9	7.8	0.0006±0.0006	0.0006±0.0004	0.0000±0.0007	−0.46715
Lig-bind/dimer	796–1,194	1	2	0.135	3.4	5.2	0.0015±0.0015	0.0000±0.0000	−0.0015±0.0015	−0.43152
C-terminal part	1,195–1,764	2	3	0.187	3.3	7.3	0.0015±0.0011	0.0000±0.0000	−0.0015±0.0012	−0.89232
TM-Cyt	1,765–2,355	2	3	0.161	2.8	7.0	0.0011±0.0010	0.0001±0.0001	−0.0010±0.0011	−1.00276

S number of variable sites, *H* number of haplotypes, *Hd* haplotype diversity, π nucleotide diversity (per site), θ nucleotide polymorphism (per site), d_S number of synonymous substitutions per site, d_N number of non-synonymous substitutions per site, d_N-d_S difference between the non-synonymous and synonymous distances per site; Tajima's D: Tajima's D statistic

*p=0.0362 upon Z test for positive (or diversifying) selection; *p=0.0435; *p=0.0406 upon Z test for purifying selection

[a] 2,355-bp of total open reading frame for TLR2 is divided to extracellular domain and the rest (TM-Cyt: transmembrane and cytoplasmic). Extracellular domain is further broken down to three sub-domains in two ways: one according to the differential motifs of leucine-rich repeats (LRR) and the other according to structural analysis of TLR2-TLR1 heterodimeric complex (Jin et al. 2007) delineating the region contributes to ligand binding and interaction to dimerization partner (Lig-bind/Dimer)

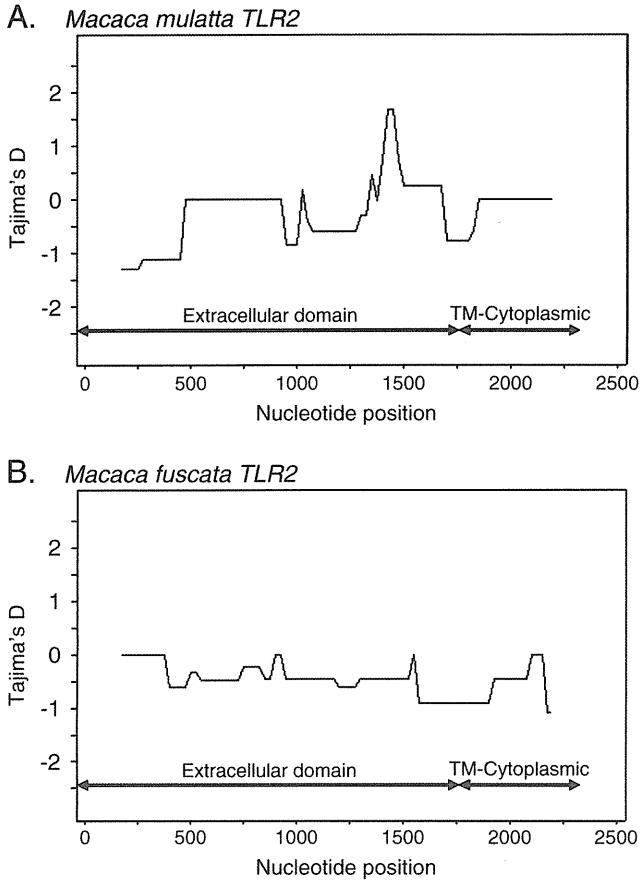

Fig. 4 Distribution of polymorphisms evaluated by Tajima's D statistic. Sliding window plots of Tajima's D with 350 bp window and 25-bp steps are drawn for *M. mulatta* (**a**) and *M. fuscata* (**b**) *TLR2* using DnaSP ver5.10.00. Although Tajima's D does not reach to significant levels, it tends to be below zero in general

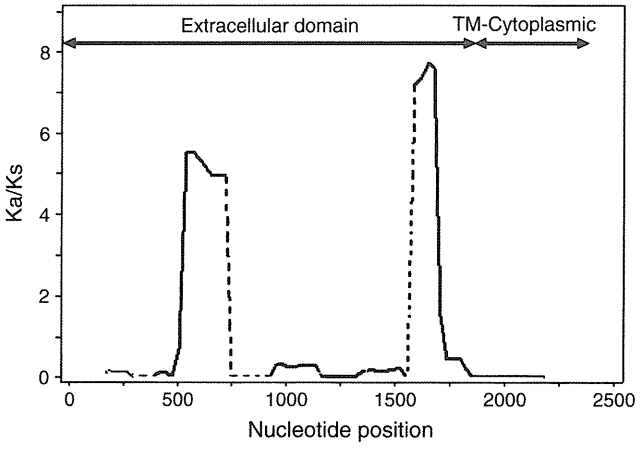

Fig. 5 Distribution of non-synonymous substitutions in two *Macaca* species. A sliding window plot of K_a/K_s ratio between *M. mulatta* and *M. fuscata* alleles with 350 bp window and 30-bp steps is drawn using DnaSP ver5.10.00. *Dotted lines* show the parts where the K_a/K_s ratio is not obtained because of the absence of synonymous substitutions

Table 3 Estimation of rate of synonymous and non-synonymous substitutions between *Macaca mulatta* and *Macaca fuscata*

Domain[a]	Range (bp)	K_s	K_a	K_a/K_s
Total	1–2,355	0.00435	0.00157	0.359
Extracellular	1–1,764	0.00553	0.00209	0.377
N-terminal LRR	73–438	0.00053	0.00009	0.164
Central LRR	439–999	0.00046	0.00471	10.312
C-terminal LRR	1,000–1,764	0.01221	0.00129	0.105
TM-Cyt	1,765–2,355	0.00057	0.00003	0.054

K_s synonymous nucleotide divergence, K_a non-synonymous nucleotide divergence, K_a/K_s the ratio of K_a to K_s

[a] See footnote of Table 2

activity and to possess the other components necessary for TLR2-mediated, ligand-induced transcriptional activation (Schwandner et al. 1999). When the expression plasmid for a major allele of *M. mulatta* TLR2, Mamu-Hap1 was introduced to HEK293 cells, two synthetic TLR2 ligands, Pam2CSK4 and Pam3CSK4, which are relatively specific to TLR2-TLR6 and TLR2-TLR1 heterodimers, respectively, could induce very strong luciferase activity to almost equivalent extent to human TLR2 expression plasmid (Fig. 6a). Therefore, the function of *M. mulatta* TLR2 in combination with HEK293-intrinsic TLR6 and TLR1 remains intact in the presence of mutations taken place after the species diversification, indicating the higher rate of amino acid substitutions observed in *M. mulatta* lineage are not attributable to pseudogenization. Further, the effect of amino acid substitution at codon 326 was evaluated by the comparison between Mamu-Hap1 and Mafu-Hap3, obtaining no evidence for functional relevance to the substitution by this experimental system (Fig. 6b). Further, the effect of all six non-synonymous substitutions found in *M. mulatta* population was examined by using four additional expression plasmids: Mamu-Hap3 carrying Ala416, Mamu-Hap7 carrying Ile405-Val503-Val556, Mamu-Hap8 carrying Pro98-Val503, and Mamu-Hap9 carrying Val503-Ala545 (Fig. 6c). Although Pam3CSK4-induced luciferase activity was higher for Mamu-Hap8 than Mamu-Hap7 and Mamu-Hap9, statistic obtained by ANOVA, by which the inflation of statistical error by multiplicity of comparison is taken into account, did not reach to significance. Thus, we could not demonstrate any functional augmentation or deterioration accompanied with seven non-synonymous amino acid substitutions identified in the present study.

Structural analysis by computational chemistry

The non-synonymous substitution at codon position 326 appeared to be differentially fixed in each species, asparagine for *M. mulatta* whereas tyrosine for *M. fuscata*.

Table 4 Comparison of frequencies of non-synonymous and synonymous substitutions in proximal region (amino acid positions, 401–566) of extracellular domain of TLR2

	Non-synonymous	Synonymous	Total	*P* value (Fisher's exact test)
Comparison A: test for difference in frequencies of non-synonymous and synonymous changes in *Macaca mulatta TLR2*				
Changes SNPs)	5	0	5	0.358
No changes	382	111	493	
Total	387	111	498 (nucleotide positions, 1,201–1,698)	
Comparison B: comparison of number of SNPs in *M. mulatta* and *Macaca fuscata TLR2*				
M. mulatta	5	0	5	0.048
M. fuscata	0	2	2	
Total	5	2	7	

The change may contribute to certain functional properties, because it locates in the region contributing to ligand binding and interaction with dimerization partner of TLR2-TLR1 heterodimeric complex (Fig. 1a). Homology modeling was applied to determine the most stabilized structure for macaque TLR2-TLR1 dimers composed of different TLR2 variants. Because of the sequence conservation across primates including human, Tyr326 (*M. fuscata* type) is assumed to be ancestral and Asn326 (*M. mulatta* type) is derivative. The effect of single amino acid substitution at position 326 was evaluated by the comparison between Mamu-Hap1 and Mafu-Hap3 (Table 6, Supplementary figures 1 and 2). Relative binding affinity to Pam3CSK4 is attenuated by approximately 5% by amino acid substitution Tyr>Asn (100.0>94.5 Table 6), which is elucidated by structural change in ligand binding pocket composed of TLR2-TLR1 dimer in part (Supplementary Figures 1 and 2). The affinity to ligand is partially recovered by second substitutions in the membrane-proximal region of extracellular domain of TLR2 (Table 6), although these substitutions were not directly involved in ligand binding or dimerization with the TLR1 partner but were supposed to induce subtle change in intra-chain interaction of TLR2 (Suppremental Figures 3 and 4).

Discussion

The rhesus monkey, *M. mulatta* has been applied to many biomedical researches for a long time, because it provides a better relevance of model organism to humans in many pathological and physiological aspects. While the lack of inbred strains or closed colony with great animals is a drawback for the use of experimental model because of variations in biological response based on underlying genetic diversity, it turns to an advantage once the genetic variations are fully understood and are controlled adequately (Matano et al. 2004). Recently, complete genome of *M. mulatta* has

Table 5 Log-likelihood values and parameter estimates under different codon substitution models for ten *Macaca mulatta* TLR2 allelic variants

Model	Number of parameters	*l*: ln(likelihood)	Average ω (d_N/d_S)	Estimates of parameters
One ratio (M0)	1	−2,461.85	0.777	ω=0.777
Site-specific models				
M1a: neutral	2	−2,461.42	0.530	p_0=0.47 (p_1=1–p_0=0.53), ω_0=0.00
M2a: selection	4	−2,458.36	0.781	p_0=0.963, p_1=0 (p_2=1–p_0–p_1=0.037), ω_0=0.00, ω_2=21.29
M3: discrete	5	−2,458.36	0.782	p_0=0.002, p_1=0.961 (p_2=1–p_0–p_1=0.037), ω_0=0.00, ω_1=0.00, ω_2=21.30
M7: beta	2	−2,460.40	0.500	p=0.0074, q=0.0073
M8: beta and ω	4	−2,458.36	0.782	p_0=0.963, p=0.0050, q=4.116 (p_1=1–p_0=0.037), ω_s=21.30
Comparisons				
One ratio (M0) vs. discrete (M3)	χ^2=−2Δ*l*=6.98	*df*=4	*p*=0.14	
Neutral (M1a) vs. selection (M2a)	χ^2=−2Δ*l*=6.12	*df*=2	*p*=0.047	
Beta (M7) vs. beta and ω (M8)	χ^2=−2Δ*l*=4.08	*df*=2	*p*=0.13	

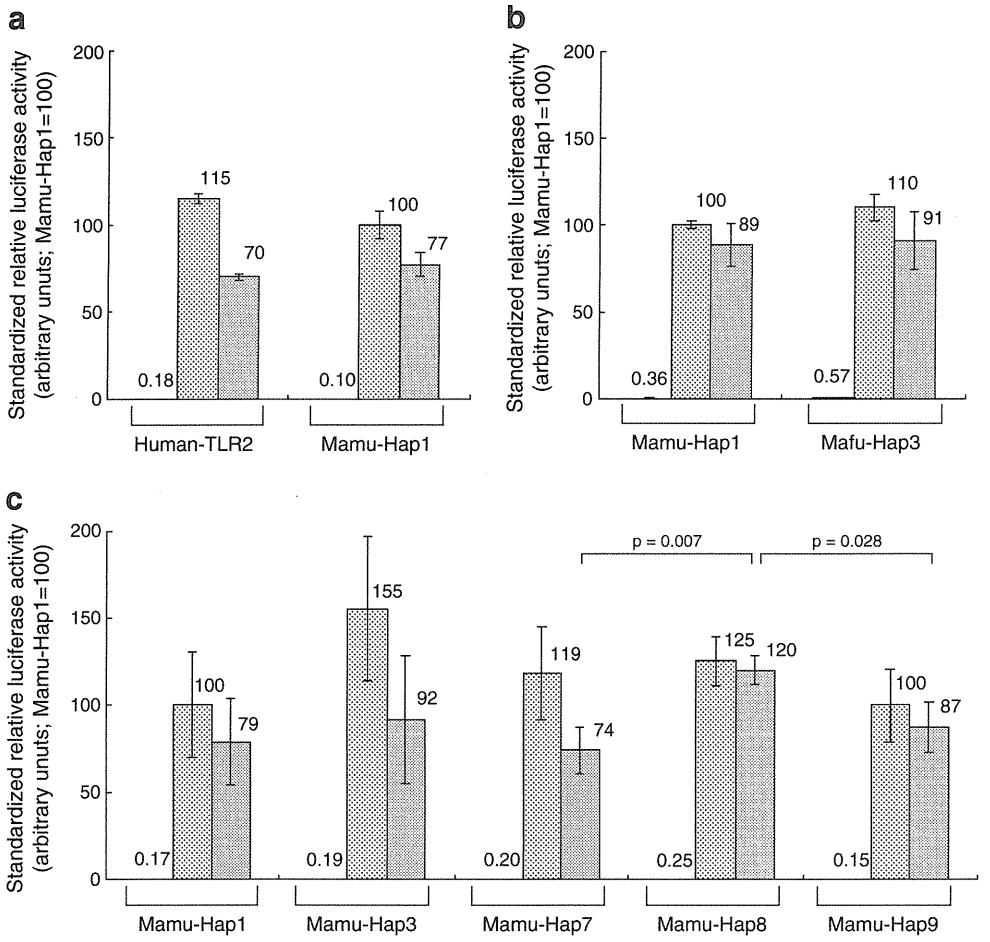

Fig. 6 Functional analysis of TLR2 alleles by luciferase reporter gene assay. Relative luciferase activity of each triplicate transfection is obtained as the ratio of relative light units (*RLUs*) of firefly luciferase assay (for NF-κB-dependent activity) to RLUs of *Renilla* luciferase assay (for constitutive basal transcriptional activity) and standardized by the mean value of those of triplicate Mamu-Hap1 transfections to be 100 for respective experiments. The results of representatives of repeated experiments at least two times are shown. Mean value of triplicate standardized relative luciferase activities without TLR2 ligand (*left*) that with 10 ng/ml of Pam2CSK4 (*middle*, *stippled bar*) and that with 100 ng/ml of Pam3CSK4 (*right*, *filled bar*) are depicted as *bar gragh with whiskers* of SEM. **a** A major allele of *M. mullata*, Mamu-Hap1, brought almost equivalent activities to the human TLR2. **b**, **c** *Macaca* TLR2 alleles carrying Tyr326 (Mafu-Hap3), Ala416 (Mamu-Hap3), Ile405-Val503-Val556 (Mamu-Hap7), Pro98-Val503 (Mamu-Hap8), and Val503-Ala545 (Mamu-Hap9) substitutions were compared

been uncovered (Rhesus Macaque Genome Sequencing and Analysis Consortium 2007), and efforts to reveal genetic polymorphism of this species have been carried out (Singh and Schmidtke 2005; Ferguson et al. 2007; Karl et al. 2008; Flynn et al. 2009; Blokhuis et al. 2009). On the other hand, the Japanese monkey, *M. fuscata*, an endemic species of Japan, also draws an attention of investigators as experimental models to human pathophysiology (Kawai et al. 1993; Isa et al. 2009). *M. mulatta* and *M. fuscata* are very close to each other in terms of species diversification (Smith et al. 2007), but geographic isolation and differential influence of dwelling environment cause species-specific characteristic through evolution. For example, *M. fuscata* is extremely labile to cerebral involvement of monkey malaria in comparison to tropical *Macaca* species (Kawai et al. 1993), presumably because of the absence of circulation of *Plasmodium* protozoa in wild *M. fuscata* population. It is suggested that the origin of current species belonging to genus *Macaca* had emerged in northern part of African continent, migrated through the Middle-East into India, and settled in Asian continent and islands (Hernandez et al. 2007). The population history of migration out of Africa and settlement in their destinations is very similar to that of divergent human ethnic groups. Infectious diseases such as malaria and arthropod-borne viral hemorrhagic fever would be a heavy burden to thrive in the tropical region in *Macaca* species as well as humans. Therefore, comparative studies of different *Macaca* species would provide valuable clues for the better understanding of host defense mechanisms of these species as well as those of humans against various infectious agents.

The previous studies on primate TLR genes revealed that the cytoplasmic TIR domains of *M. mulatta* TLRs shares high

Table 6 Molecular stability and interaction energy estimated for TLR ternary complex structure obtained by optimization through homology modeling

Ternary complex of TLR dimer and ligand	Total energy[a] (kcal/mol)	Ligand[b] (kcal/mol)	Interaction[c] (kcal/mol)	London dG score[d] (kcal/mol)	Relative affinity to ligand (macaque TLR2(Tyr326)=100)
Human TLR2 (27–506)-TLR1(25–475)-Pam3CSK4[e]	−11,680.30	−393.12	−110.59	−26.21	105.6
Macaque TLR2(Tyr326)-TLR1-Pam3CSK4 (*M. fuscata*, hap-3)	−9,101.94	−398.09	−110.41	−24.81	100.0 (reference)
Macaque TLR2(Asn326)-TLR1-Pam3CSK4 (*Macaca mulatta*, hap-1)	−9,114.51	−397.85	−109.13	−23.52	94.5
Macaque TLR2(Asn326-Ile405)-TLR1-Pam3CSK4	−9,808.11	−396.92	−110.80	−24.24	97.7
Macaque TLR2(Asn326-Ala416)-TLR1-Pam3CSK4	−9,524.80	−399.36	−109.04	−24.46	98.6

[a] Total potential energy between all the atoms of extracellular domains of TLR2-TLR1 heterodimeric receptor (position 27–506 of TLR2 and position 25–475 of TLR1) and a ligand Pam3CSK4 was calculated by Molecular Operating Environment (MOE) 2010.10(Chemical Computing Group Inc)

[b] Total potential energy between all the atoms of Pam3CSK4

[c] Total potential energy between the atoms of TLR2-TLR1 heterodimeric receptor and those of Pam3CSK4

[d] An index for affinity between receptor and ligand within the binding pocket obtained by Lig-X algorithm suite implemented in MOE 2010.10

[e] Molecular structure was optimized by MOE 2010.10 after retrieval of PDB data accession# 2Z7X (available at RCSB, http://www.rcsb.org/pdb/) and trimming of C-terminal 67 residues of fusion peptide derived from inshore hagfish Eptatretus VLRB.61

levels of homology with human counterparts (Sanghavi et al. 2004), and TIR domains of TLRs and other related genes had undergone purifying selection among seven primate species (Nakajima et al. 2008), by which the presence of strict functional constraint is suggested. On the other hand, it was also revealed that the extracellular domain of *Macaca* TLR4 have been uniquely evolved under the diversifying selection by comparison between primate species (Nakajima et al. 2008). While the coding polymorphisms in *M. mulatta TLR4* and *TLR5* (Ferguson et al. 2007), as well as microsatellite located in the second intron of *TLR2* (Yim et al. 2006), were reported, we investigated for the first time SNPs in the coding region of *TLR2* for *M. mulatta* and *M. fuscata* individuals and reach to the trace evidence for differential purifying and diversifying selections on certain parts of extracellular domain in the present study. It is of note that the shapes of sliding window plots of Tajima's D (Fig. 4) were quite different between *M. mulatta* and *M. fuscata* suggesting differential molecular evolution for these species after species diversification. Inter-specific differential molecular evolution is also prominent in the sliding window plot of K_a/K_s ratio for combined *M. mulatta* and *M. fuscata* allelic variants (Fig. 5). Further, we could obtain a trace of positive selection at one amino acid position 545 by application of maximal likelihood estimation for codon-wise evaluation of molecular evolution events. Certain polymorphisms located in extracellular domain of swine *TLR2* (Shinkai et al. 2006; Bergman et al. 2010) and those of bovine *TLR2* (Jann et al. 2008) were also suggested to have been under positive selection during the domestication and breeding of these mammalian species. It is probable that positive selection has been driven by the functional alteration adapting to the changes of dwelling environments. Further, one of several rare variants of human TLR2, Thr411Ile, which has been shown functionally deteriorated upon the Pam3CSK4-induced, NF-κB-driven reporter gene assay (Merx et al. 2007; Kormann et al. 2009), is located in the part C-terminal to the ligand binding and dimerization domain in the vicinity of membrane-proximal region where five of the non-synonymous substitutions were found in our *M. mulatta* population. Although we could not demonstrate apparent augmentation or deterioration accompanied with amino acid substitutions in *M. mulatta* alleles by luciferase reporter gene assay using HEK293 cells as recipients with saturating doses of lipopeptide ligands in the present study, the affinity to ligand appeared different and would be differentially tuned during the evolution of two *Macaca* species under distinct environmental selection pressure.

Acknowledgments This work was supported in part by Japan Initiative for Global Research Network on Infectious Diseases (J-GRID) and the Global Centers of Excellence Program of the Ministry of Education, Culture, Sports, Science and Technology of Japan and by the Cooperation Research Program of Primate Research Institute, Kyoto University, Japan

References

Akira S, Uematsu S, Takeuchi O (2006) Pathogen recognition and innate immunity. Cell 124(4):783–801

Baena A, Mootnick AR, Falvo JV, Tsytskova AV, Ligeiro F, Diop OM, Brieva C, Gagneux P, O'Brien SJ, Ryder OA, Goldfeld AE

(2007) Primate TNF promoters reveal markers of phylogeny and evolution of innate immunity. PLoS One 2:e621

Bergman IM, Rosengren JK, Edman K, Edfors I (2010) European wild boars and domestic pigs display different polymorphic patterns in the Toll-like receptor (TLR) 1, TLR2, and TLR6 genes. Immunogenetics 62:49–58

Blancher A, Bonhomme M, Crouau-Roy B, Terao K, Kitano T, Saitou N (2008) Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (*Macaca fascicularis fascicularis*). J Hered 99:254–264

Blokhuis JH, van der Wiel MK, Doxiadis GG, Bontrop RE (2009) Evidence for balancing selection acting on KIR2DL4 genotypes in rhesus macaques of Indian origin. Immunogenetics 61:503–512

Bochud PY, Hawn TR, Aderem A (2003) A Toll-like receptor 2 polymorphism that is associated with lepromatous leprosy is unable to mediate mycobacterial signaling. J Immunol 170:3451–3454

Ferguson B, Street SL, Wright H, Pearson C, Jia Y, Thompson SL, Allibone P, Dubay CJ, Spindel E, Norgren RB Jr (2007) Single nucleotide polymorphisms (SNPs) distinguish Indian-origin and Chinese-origin rhesus macaques (*Macaca mulatta*). BMC Genomics 8:43

Flynn S, Satkoski J, Lerche N, Kanthaswamy S, Smith DG (2009) Genetic variation at the TNF-alpha promoter and malaria susceptibility in rhesus (*Macaca mulatta*) and long-tailed (*Macaca fascicularis*) macaques. Infect Genet Evol 9:769–777

Hayasaka K, Fujii K, Horai S (1996) Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA. Mol Biol Evol 13:1044–1053

Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, Nazareth L, Indap A, Bourquin T, McPherson J, Muzny D, Gibbs R, Nielsen R, Bustamante CD (2007) Demographic histories and patterns of linkage disequilibrium in Chinese and Indian rhesus macaques. Science 316:240–243

Huang D, Qiu L, Wang R, Lai X, Du G, Seghal P, Shen Y, Shao L, Halliday L, Fortman J, Shen L, Letvin NL, Chen ZW (2007) Immune gene networks of mycobacterial vaccine-elicited cellular responses and immunity. J Infect Dis 195:55–69

Isa T, Yamane I, Hamai M, Inagaki H (2009) Japanese macaques as laboratory animals. Exp Anim 58:451–457

Jann OC, Werling D, Chang JS, Haig D, Glass EJ (2008) Molecular evolution of bovine Toll-like receptor 2 suggests substitutions of functional relevance. BMC Evol Biol 8:288

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

Kang JY, Nan X, Jin MS, Youn SJ, Ryu YH, Mah S, Han SH, Lee H, Paik SG, Lee JO (2009) Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. Immunity 31:873–884

Karl JA, Wiseman RW, Campbell KJ, Blasky AJ, Hughes AL, Ferguson B, Read DS, O'Connor DH (2008) Identification of MHC class I sequences in Chinese-origin rhesus macaques. Immunogenetics 60:37–46

Kawai S, Aikawa M, Kano S, Suzuki M (1993) A primate model for severe human malaria with cerebral involvement: *Plasmodium coatneyi*-infected *Macaca fuscata*. Am J Trop Med Hyg 48:630–636

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kormann MS, Ferstl R, Depner M, Klopp N, Spiller S, Illig T, Vogelberg C, von Mutius E, Kirschning CJ, Kabesch M (2009) Rare TLR2 mutations reduce TLR2 receptor function and can increase atopy risk. Allergy 64:636–642

Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. Nature 392:917–920

Kumar S, Nei M, Dudley J, Tamura K (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform 9:299–306

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

Ling B, Veazey RS, Luckay A, Penedo C, Xu K, Lifson JD, Marx PA (2002) SIVmac pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans. AIDS 16:1489–1496

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

Matano T, Kobayashi M, Igarashi H, Takeda A, Nakamura H, Kano M, Sugimoto C, Mori K, Iida A, Hirata T, Hasegawa M, Yuasa T, Miyazawa M, Takahashi Y, Yasunami M, Kimura A, O'Connor DH, Watkins DI, Nagai Y (2004) Cytotoxic T lymphocyte-based control of simian immunodeficiency virus replication in a preclinical AIDS vaccine trial. J Exp Med 199:1709–1718

McMurray DN (2000) A nonhuman primate model for preclinical testing of new tuberculosis vaccines. Clin Infect Dis 30:S210–S212

Merx S, Neumaier M, Wagner H, Kirschning CJ, Ahmad-Nejad P (2007) Characterization and investigation of single nucleotide polymorphisms and a novel TLR2 mutation in the human TLR2 gene. Hum Mol Genet 16:1225–1232

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

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Ogus AC, Yoldas B, Ozdemir T, Uguz A, Olcen S, Keser I, Coskun M, Cilli A, Yegin O (2004) The Arg753Gln polymorphism of the human toll-like receptor 2 gene in tuberculosis disease. Eur Respir J 23:219–223

Osada N, Hashimoto K, Kameoka Y, Hirata M, Tanuma R, Uno Y, Inoue I, Hida M, Suzuki Y, Sugano S, Terao K, Kusuda J, Takahashi I (2008) Large-scale analysis of *Macaca fascicularis* transcripts and inference of genetic divergence between *M. fascicularis* and *M. mulatta*. BMC Genomics 9:90

Rhesus Macaque Genome Sequencing and Analysis Consortium (2007) Evolutionaly and biomedical insights from the rhesus macaque genome. Science 316:222–234

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing Phylogenetic trees. Mol Biol Evol 4:406–425

Sanghavi SK, Shankarappa R, Reinhart TA (2004) Genetic analysis of Toll/Interleukin-1 receptor (TIR) domain sequences from rhesus macaque Toll-like receptors (TLRs) 1–10 reveals high homology to human TLR/TIR sequences. Immunogenetics 56:667–674

Schröder NW, Diterich I, Zinke A, Eckert J, Draing C, von Baehr V, Hassler D, Priem S, Hahn K, Michelsen KS, Hartung T, Burmester GR, Göbel UB, Hermann C, Schumann RR (2005) Heterozygous Arg753Gln polymorphism of human TLR-2 impairs immune activation by *Borrelia burgdorferi* and protects from late stage Lyme disease. J Immunol 175:2534–2540

Schwandner R, Dziarski R, Wesche H, Rothe M, Kirschning CJ (1999) Peptidoglycan- and lipoteichoic acid-induced cell activation is mediated by toll-like receptor 2. J Biol Chem 274:17406–17409

Shibata H, Yasunami M, Obuchi N, Takahashi M, Kobayashi Y, Numano F, Kimura A (2006) Direct determination of SNP haplotype of NFKBIL1 promoter polymorphism by DNA conformation analysis and its application to association study of chronic inflammatory diseases. Hum Immunol 67:363–373

Shinkai H, Tanaka M, Morozumi T, Eguchi-Ogawa T, Okumura N, Muneta Y, Awata T, Uenishi H (2006) Biased distribution of

single nucleotide polymorphisms (SNPs) in porcine Toll-like receptor 1 (TLR1), TLR2, TLR4, TLR5, and TLR6 genes. Immunogenetics 58:324–330

Singh KK, Schmidtke J (2005) Single nucleotide polymorphisms within the promoter region of the rhesus monkey tumor necrosis factor-alpha gene. Immunogenetics 57:289–292

Smith DG, McDonough JW, George DA (2007) Mitochondrial DNA variation within and among regional populations of longtail macaques (*Macaca fascicularis*) in relation to other species of the fascicularis group of macaques. Am J Primatol 69:182–198

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

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Takahashi-Tanaka Y, Yasunami M, Naruse T, Hinohara K, Matano T, Mori K, Miyazawa M, Honda M, Yasutomi Y, Nagai Y, Kimura A (2007) Reference strand-mediated conformation analysis (RSCA)-based typing of multiple alleles in the rhesus macaque MHC class I Mamu-A and Mamu-B loci. Electrophoresis 28:918–924

Texereau J, Chiche JD, Taylor W, Choukroun G, Comba B, Mira JP (2005) The importance of Toll-like receptor 2 polymorphisms in severe infections. Clin Infect Dis 41:S408–S415

Yang Z (2007) PAML 4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–1591

Yim JJ, Adams AA, Kim JH, Holland SM (2006) Evolution of an intronic microsatellite polymorphism in Toll-like receptor 2 among primates. Immunogenetics 58:740–745

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci U S A 95:3708–3713

Immunogenetics (2012) 64:131–141
DOI 10.1007/s00251-011-0568-y

ORIGINAL PAPER

Diversity of MHC class I haplotypes in cynomolgus macaques

Yusuke Saito · Taeko K. Naruse · Hirofumi Akari · Tetsuro Matano · Akinori Kimura

Received: 7 May 2011 / Accepted: 19 August 2011 / Published online: 1 September 2011

Abstract Cynomolgus macaques are widely used as a primate model for human diseases associated with an immunological process. Because there are individual differences in immune responsiveness, which are controlled by the polymorphic nature of the major histocompatibility (MHC) locus, it is important to reveal the diversity of MHC in the model animal. In this study, we analyzed 26 cynomolgus macaques from five families for MHC class I genes. We identified 32 *Mafa-A*, 46 *Mafa-B*, 6 *Mafa-I*, and 3 *Mafa-AG* alleles in which 14, 20, 3, and 3 alleles were novel. There were 23 MHC class I haplotypes and each haplotype was composed of one to three *Mafa-A* alleles and one to five *Mafa-B* alleles. Family studies revealed that there were two haplotypes which contained two *Mafa-A1* alleles. These observations demonstrated further the complexity of MHC class I locus in the Old World monkey.

Keywords Cynomolgus macaque · MHC · *Mafa* class I gene · Haplotype · Polymorphism

Yusuke Saito and Taeko K. Naruse contributed equally to this work.

Y. Saito · T. K. Naruse · A. Kimura (✉)
Department of Molecular Pathogenesis,
Medical Research Institute, Tokyo Medical and Dental University,
1-5-45 Yushima,
Bunkyo-ku, Tokyo 113-8510, Japan
e-mail: akitis@mri.tmd.ac.jp

Y. Saito · A. Kimura
Laboratory of Genome Diversity,
Graduate School of Biomedical Science,
Tokyo Medical and Dental University,
Tokyo, Japan

H. Akari
Primate Research Institute,
Kyoto University,
Inuyama, Japan

T. Matano
AIDS Research Center, National Institute of Infectious Diseases,
Tokyo, Japan

T. Matano
The Institute of Medical Science, The University of Tokyo,
Tokyo, Japan

Introduction

Non-human primates are widely used for immunological research because their immune system is similar to that of humans. In particular, the Old World monkeys such as cynomolgus macaques (crab-eating macaques, *Macaca fascicularis*) became a useful model for human infectious diseases including acquired immunodeficiency syndrome (AIDS) (Wiseman et al. 2007), severe acute respiratory syndrome (Lawler et al. 2006), and influenza (Kobasa et al. 2007) as well as in the transplantation field (Wiseman and O'Connor 2007). In the AIDS research, cynomolgus and rhesus macaques are important animal models for the development of vaccines against human immunodeficiency virus (HIV) or studies for susceptibility to HIV infection and/or development of AIDS (Matano et al. 2004; Loffredo et al. 2008; Tsukamoto et al. 2008; Burwitz et al. 2009; Mee et al. 2009; Aarnink et al. 2011a). To fully evaluate the results of immunological experiments in the macaque models, it is essential to characterize the genetic diversity of immune-related molecules which may control the individual differences in the immune response against foreign antigens and/or pathogens.

The major histocompatibility complex (MHC) is well known to control the immune-responsiveness to foreign

antigens. There are two classes of MHC molecules: one is the MHC class I molecule presenting peptides of intracellular origin to $CD8^+$ T cell and the other is the MHC class II molecule binding extracellular-derived antigenic peptides for presenting to $CD4^+$ T cell. It has been reported that the complexity of MHC genes in the rhesus and cynomolgus macaques is higher than that in humans (Kulski et al. 2004; Watanabe et al. 2006; Gibbs et al. 2007; Otting et al. 2007, 2008; Doxiadis et al. 2011). For example, *MHC class I* configurations in macaques are usually composed of one copy of highly transcribed major *MHC-A1*gene (*Mamu-A1*or *Mafa-A1*) and several other minor *MHC-A* genes (*Mamu-A2~A7* or *Mafa-A2~A6*) in addition to several *MHC-B* genes (*Mamu-B* or *Mafa-B*) (Watanabe et al. 2006; Otting et al. 2007, 2008, 2009; Naruse et al. 2010; Doxiadis et al. 2011), whereas each one copy of *MHC-A* and *-B* genes (*HLA-A* and *-B*) can be found in human *MHC class I* locus. In addition, other *MHC* loci showing lower expression levels, i.e., *HLA-B*-like gene (*Mamu-I* or *Mafa-I*) and *HLA-G*-like non-classical gene (*Mamu-AG* or *Mafa-AG*) have been identified (Slukvin et al. 2000; Urvater et al. 2000). The extent of genetic diversity is different, in part, depending on the geographic areas, as we have previously reported for *MHC class I* genes in rhesus macaque (Naruse et al. 2010). As for the cynomolgus macaques, *MHC class I* allelic diversity was reported for Indonesian (Pendley et al. 2008; Wu et al. 2008; Kita et al. 2009; Otting et al. 2009), Malaysian (Otting et al. 2009; Aarnink et al. 2011b), Mauritian (Budde et al. 2010), Vietnamese (Wu et al. 2008; Kita et al. 2009), and Philippino (Campbell et al. 2009; Kita et al. 2009) macaques, but information about the *MHC class I* haplotype remains insufficient.

In the present study, we have analyzed *MHC class I* loci in cynomolgus macaques originated from Indonesia, Malaysia, and the Philippines to obtain information on haplotype configuration. We report here further the complex nature of *MHC class I* loci in the Old World monkey, i.e., the presence of unique haplotypes carrying two *Mafa-A1* genes.

Materials and methods

Animals

A total of 26 cynomolgus macaques from five families were the subjects. Each family was composed of one or two males with one or two females and their offspring. They were maintained in the breeding colonies in Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Japan. The founders of the colonies were captured in Indonesia, Malaysia, and the Philippines. All care including blood sampling of animals were in accordance with the Guidelines for the Care and Use of Laboratory Animals published by the National Institute of Health (NIH Publication 85–23, revised 1985) and were subjected to prior approval by the local animal protection authority.

Sequencing analysis of cDNAs from *Mafa* class I genes

Total cellular RNA was extracted from whole blood by using RNAeasy (QIAGEN, Gmbh, Germany). Oligo(dT)-primed cDNA was synthesized using Transcriptor reverse transcriptase (Roche, Mannheim, Germany) according to the manufacturer's recommendations. Full-length cDNAs

Table 1 Primers used in PCR or sequencing of *Mafa* class I genes

Primer ID	Application	Direction	Sequence (5′–3′)	Position	Reference
5′ MHC_UTR	PCR	Sense	GGACTCAGAATCTCCCCAGACGCCGAG	5′ UTR	Karl et al. 2008
3′ MHC_UTR_A	PCR	Antisense	CAGGAACAYAGACACATTCAGG	3′ UTR	Karl et al. 2008
3′ MHC_UTR_B	PCR	Antisense	GTCTCTCCACCTCCTCAC	3′ UTR	Karl et al. 2008
5A long	PCR	Sense	ATGGCGCCCCGAACCCTCCTCCTG	Exon 1	Tanaka-Takahashi et al. 2007
3A	PCR	Antisense	TCACACTTTCAAGCCGTGAGAGA	Exon 7	Tanaka-Takahashi et al. 2007
5ASSP	PCR	Sense	ATGGCGCCCCGAACCCTCCTCCTGG	Exon 1	Tanaka-Takahashi et al. 2007
4R	PCR	Antisense	CCAGGTCAGTGTGATCTCCG	Exon 4	Tanaka-Takahashi et al. 2007
P000044	PCR	Sense	GATTCTCCGCAGACGCCCA	5′ UTR	Wu et al. 2008
P000023	PCR	Antisense	GGAGAACCAGGCCAGCAAT	Exon 5	Wu et al. 2008
P000076	Sequencing	Sense	GAGCAGCGACGGGACCGCA	Intron 1	Wu et al. 2008
P000060	Sequencing	Antisense	CCTGGGGCTCTCCCGGGTCA	Intron 2	Wu et al. 2008
P000096	Sequencing	Sense	TGTACTGAGTCTCCCTGATGG	Intron 2	Wu et al. 2008
P000098	Sequencing	Antisense	TTCATTCCCTCAGAGATTTT	Intron 3	Wu et al. 2008
P000055	Sequencing	Sense	CCCAGGTRCCTSTGTCCAGGA	Intron 3	Wu et al. 2008
P000281	Sequencing	Antisense	AGAGGGGAAAGTGAGGGGT	Intron 4	Wu et al. 2008

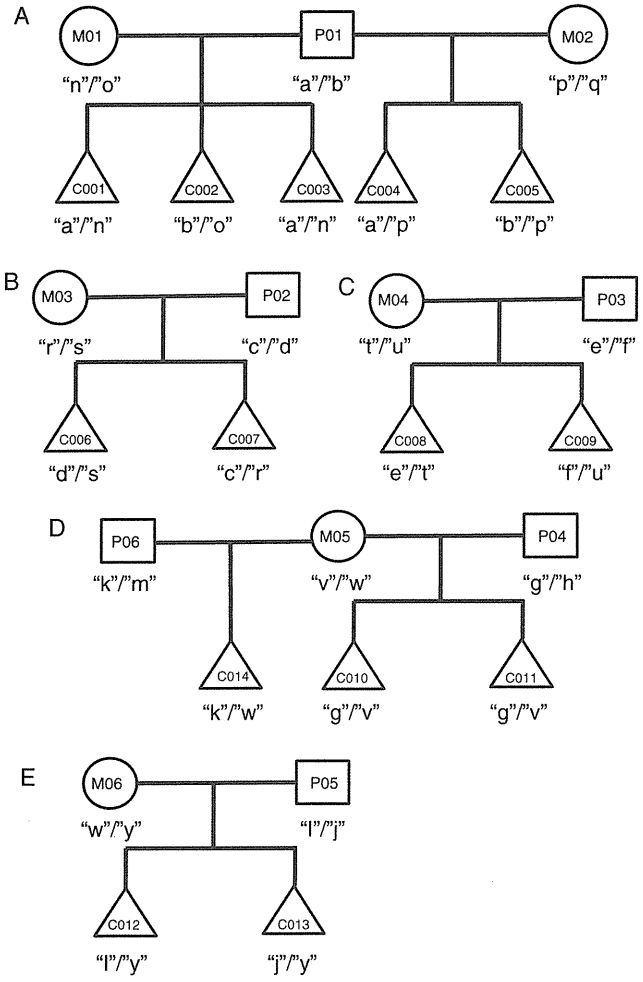

Fig. 1 Pedigree of cynomolgus macaques. The pedigrees of macaques analyzed in this study are shown. Founders were originated from Indonesia (**a**), Malaysia (**b**, **c**), and Philippines (**d**, **e**). *Open square*, *open circle*, and *open triangles* indicate father, mother, and offspring, respectively. The ID of each subject is noted in the symbol. *Mafa class I* haplotypes determined in this study are indicated under the subjects

for *Mafa* class I genes were amplified by polymerase chain reaction (PCR), as described previously (Tanaka-Takahashi et al. 2007; Naruse et al. 2010), by using locus-specific primer pairs as reported by Karl et al. (2008). Genomic gene and cDNA for *Mafa-A2* gene were analyzed according to the method described by Wu et al. (2008). The primers used in this study are listed in Table 1. To estimate the expression level of *Mafa-A* alleles, we also used an additional primer pair: MafaF (5′-TACGTGGACGA CACGCAGTT) and MafaR (5′-GGTGGGTCA CATGTGTCTTG). PCR was done under the condition of initial denaturation at 98°C for 10 s, 25 cycles of 98°C for 1 s, 64°C for 5 s, and 72°C for 20 s, followed by an additional extension at 72°C for 1 min, using Phusion Flash DNA polymerase (Finzymes, Espoo, Finland). The PCR products were cloned into pSTBlue-1 Perfectly Blunt vector (Novagen, WI, USA) according to the manufacturer's instructions and were transformed to NovaBlue Giga Singles™ competent cells (Merck Biosciences Japan, Tokyo, Japan). A total of 30 to 90 independent cDNA clones were obtained from each macaque for each locus and were sequenced on both strands by BigDye Terminator cycling system in an ABI 3730 automated sequence analyzer (Applied Biosystems, CA, USA).

Data analyses and nomenclature for *Mafa* class I allele

Nucleotide sequences of cDNA clones were aligned using the Genetyx software package (version 8.0, Genetyx Corp., Japan). When a cDNA sequence, which was represented by at least three clones, was independently obtained from at least two animals or repeatedly obtained from at least two independently prepared cDNAs from single animals, we considered it a real allele, not an artifact, and the sequences were submitted to the DNA Data Bank of Japan (DDBJ) database and to the Immuno Polymorphism Database for non-human primate MHC (http://www.ebi.ac.uk/ipd/mhc/sumit.html; Robinson et al. 2003) to obtain official nomenclature for the novel alleles of *Mafa-A* and *Mafa-B* genes. Neighbor-joining trees were constructed with Kimura's two-parameter method for a phylogenetic analysis of *Mafa-A* sequences spanning exons 2, 3, and a part of exon 4 obtained in this study by using the Genetyx software. Bootstrap values were based on 5,000 replications.

Table 2 *Mafa* class I alleles found in the cynomolgus macaques

Locus	Number of observed alleles	Number of novel alleles (%)	Number of observed alleles in macaques from different regions[a]		
			Indonesian	Malaysian	Philippino
Mafa-A	32	14 (43.7%)	9 (3), 33.3%	12 (8), 66.7%	11 (3), 27.3%
Mafa-B	46	20 (43.5%)	13 (5), 38.5%	20 (15), 75.0%	18 (1), 5.6%
Mafa-I	6	3 (50.0%)	2 (1), 50.0%	4 (3), 75.0%	2 (0), 0%
Mafa-AG	3	3 (100%)	0 (0), 0%	2 (2), 100%	1 (1), 100%
Total	87	40 (45.5%)	24 (9), 37.5%	38 (28), 73.7%	32 (5), 15.6%

[a] The number and frequency of novel alleles are indicated in parentheses

Results

Identification of *Mafa* class I alleles in cynomolgus macaques

We determined the nucleotide sequences of cDNA clones for *Mafa-A* and *-B* loci in 26 cynomolgus macaques from one family of Indonesian origin (six haplotypes), two families of Malaysian origin (eight haplotypes), and two families of Philippino origin (nine haplotypes) (Fig. 1). When the observed alleles were segregated in the family or when at least three clones with identical sequences were observed from two independent PCR for an individual, the nucleotide sequences were considered to be real and not artifacts. As shown in Table 2, 32 *Mafa-A*, 46 *Mafa-B*, 6 *Mafa-I*, and 3 *Mafa-AG* sequences were obtained in this study. Among them, 14 (43.7%), 20 (43.5%), 3 (50.0%), and 3 (100%) were novel alleles of *Mafa-A*, *Mafa-B*, *Mafa-I*, and *Mafa-AG* loci, respectively (Table 2).

Table 3 Alleles of *Mafa-A* locus identified in the cynomolgus macaques

Locus	Allele name	Novelty[a]	Accession number[b]	Origin[c]	Identical *Mamu* and/or *Mane* alleles[d]	Origin and reference of known alleles[e]
A1	*A1*001:01*		AM295828	Malaysian		Utrecht, Otting et al. 2007
A1	*A1*002:01:02*	Novel	AB569214	Indonesian		
A1	*A1*008:02*		EU392108	Philippino		Philippino, Campbell et al. 2009
A1	*A1*008:03-like*	Novel	AB647187	Philippino		
A1	*A1*018:06*		FM246489	Indonesian		Utrecht, Otting et al. 2007
A1	*A1*019:05*		AB447616	Indonesian		Indonesian, Kita et al. 2009
A1	*A1*023:01*	Novel	AB569216	Malaysian		
A1	*A1*032:05*	Novel	AB569215	Malaysian		
A1	*A1*052:02*		EU392105	Philippino	*Mamu-A1*052:01/03/06*	Philippino, Campbell et al. 2009
A1	*A1*054:01*		AB154771	Malaysian		Tsukuba, Uda et al. 2004
A1	*A1*056:02*	Novel	AB569218	Malaysian		
A1	*A1*062:05*	Novel	AB569219	Malaysian		
A1	*A1*068:02*	Novel	AB569217	Malaysian		
A1	*A1*074:02*		AB447606	Philippino		Philippino, Kita et al. 2009
A1	*A1*079:01*		AB154773	Malaysian		Tsukuba, Uda et al. 2004
A1	*A1*089:02*		EU392104	Philippino		Philippino, Campbell et al. 2009
A1	*A1*093:01*		EU392103	Philippino		Philippino, Campbell et al. 2009
A1	*A1*094:01*		EU392111	Philippino		Philippino, Campbell et al. 2009
A1	*A1*097:01*		AB447576	Indonesian	*Mamu-A1*109:01*	Indonesian, Kita et al. 2009
A1	*A1*103:01*	Novel	AB583236	Indonesian		
A1	*A1*124:01*	Novel	AB583237	Malaysian		
A2	*A2*05:13-like*	Novel	AB647189	Philippino		
A2	*A2*05:16*		AM295878	Indonesian		Utrecht, Otting et al. 2007
A2	*A2*05:34-like*	Novel	AB647190	Philippino		
A3	*A3*13:03*		EU392112	Philippino		Philippino, Campbell et al. 2009
A3	*A3*13:15*	Novel	AB583238	Malaysian		
A3	*A3*13:16*	Novel	AB583240	Indonesian		
A4	*A4*14:01*		AM295880	Indonesian		Utrecht, Otting et al. 2007
A4	*A4*14:02*		AM295881	Malaysian		Utrecht, Otting et al. 2007
A6	*A6*01:05*	Novel	AB583239	Malaysian		

[a] New alleles are indicated as novel

[b] Nucleotide sequences were submitted to a public database and given accession numbers

[c] Origin of cynomolgus macaques

[d] Identical sequences were found in *Mamu* or *Mane* alleles

[e] Origin and references in which each known allele was first reported. Utrecht and Tsukuba indicate that the alleles were found in colonies maintained in the University of Utrecht, The Netherlands, and Tsukuba primate center, Japan, respectively