Foster City, CA, USA). Editing and assembly of sequences were performed using SEQUENCHER (Gene Codes, Ann Arbor, MI, USA). The *CD3E* and *CD3G* sequences determined in this study were deposited in DNA Data Bank of Japan under the following accession numbers: AB583139-AB583171 (ESM Table 2).

Statistical analyses

In this study, we used both Bn-Bs program and PAML program to reduce the chance of false-positive findings. A criterion for the gene under positive selection pressure was that the p values obtained by both the Bn-Bs and PAML programs were less than 0.05. The first screening of IgSF genes under the selection pressure was performed by using the Bn-Bs program. The genes showing p values less than 0.05 in the first screening were further analyzed by using the PAML program.

The Bn-Bs program estimates the values of non-synonymous substitution rate (dn) and synonymous substitution rate (ds) based on the modified Nei–Gojobori method (Nei and Gojobori 1986), where a phylogenetic tree is given (Zhang et al. 1998). The value of ω, an abbreviation for the value of dn/ds, is a criterion of natural selective pressure acting on the gene. Statistical significance of the difference between dn and ds were examined by Z test (Chatterjee et al. 2009). An ordinary least-squares method was used to estimate the branch lengths and variances for the evolutionary distances between two sequences (Rzhetsky and Nei 1993).

Investigation on the presence of branch-specific positive selection (the branch model) and site-specific positive selection (the site model) were performed by using CODEML, an application from PAML version 4.7 (Yang 2007). The branch model is used for evaluation of difference in the value of ω for each branch, and it is useful for detecting a positive selection acting on particular branch by using the likelihood ratio tests (Yang and Nielsen 2000). The site model treats ω allowing the variance among codons (Yang 2005; Yang and Nielsen 2000). Bayes empirical Bayes (BEB) method was used to detect the sites under the positive selection (Yang and Nielsen 2000; Wong et al. 2004; Yang et al. 2005; Yang 2005, 2007).

Results

Non-synonymous/synonymous substitution ratio of IgSF genes

Four hundred sixty-one IgSF genes were selected from the human genome, based on the Conserved Domain Database v2.22 at NCBI (http://www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml). Among them, 47 genes composed of MHC, KIR, and PSG genes (ESM Table 3) were excluded from the phylogenetic analysis, because there are many paralogous genes with high similarity in the sequences, which may lead to uncertainty to identify the reliable orthologous genes. Thus, a total of 414 IgSF genes were subjected to the following analysis. By using the UCSC Genome Browser (http://genome.ucsc.edu/), we attempted to identify orthologous genes for these 414 human IgSF genes in genomes from chimpanzee, orangutan, rhesus macaque, and common marmoset. We were unable to identify orthologs for 53, 55, 52, and 81 genes from the genome of chimpanzee, orangutan, rhesus macaque, and common marmoset, respectively, due to the alignment incompleteness (sequence identity of less than 80%), insertion/deletions accompanied by frameshift, or nucleotide substitutions resulting in a premature stop codon. After removing the IgSF genes of which the reliable orthologous genes were not identified in the non-human primates, remaining 249 IgSF genes were used in the study of positive selection.

The Bn-Bs program was applied to evaluate the non-synonymous/synonymous substitution ratio (Larkin et al. 2007), and the value of Σdn and Σds, which were the sum values of dn and ds, respectively, in seven primate lineages, human, chimpanzee, human-chimpanzee ancestor, orangutan, human–chimpanzee–orangutan ancestor, rhesus macaque, and common marmoset were calculated. The IgSF genes were classified into 11 functional categories based on the Gene Ontology database (http://www.geneontology.org/); GO:0002376: immune system process, GO:0006952: defense response, GO:0051704: multi-organism process, GO:0007166: cell surface receptor linked signaling pathway, GO:0007155: cell adhesion, GO:0007165: signal transduction, GO:0008219: cell death, GO:0030154 : cell differentiation, GO:0008283: cell proliferation, GO:0019222: regulation of metabolic process and GO:0050794: regulation of cellular process. When the Σdn/Σds ratios were calculated for the entire coding sequences, there was no evidence to support the presence of positive natural selection. The Σdn/Σds ratios from the analyzed genes, except for *LAIR1* (Σdn/Σds ratio= 1.00, statistically not significant), were lower than 1.0, implying that most of the IgSF genes had been under the pressure of negative selection in the course of primate evolution. Among the functional categories, GO:0002376: immune system process (median Σdn/Σds ratio=0.407; interquartile range (IQR), 0.285–0.632, $p=1.40\times10^{-6}$), GO:0006952: defense response (median Σdn/Σds ratio= 0.394; IQR, 0.287–0.506, $p=6.73\times10^{-3}$), and GO:0051704: multi-organism process (median Σdn/Σds ratio=0.400, IQR, 0.302–0.475, $p=2.46\times10^{-2}$) showed much higher values of Σdn/Σds ratio than the tested IgSF genes (median Σdn/Σds ratio=0.208, IQR, 0.107–0.440; Fig. 1a).

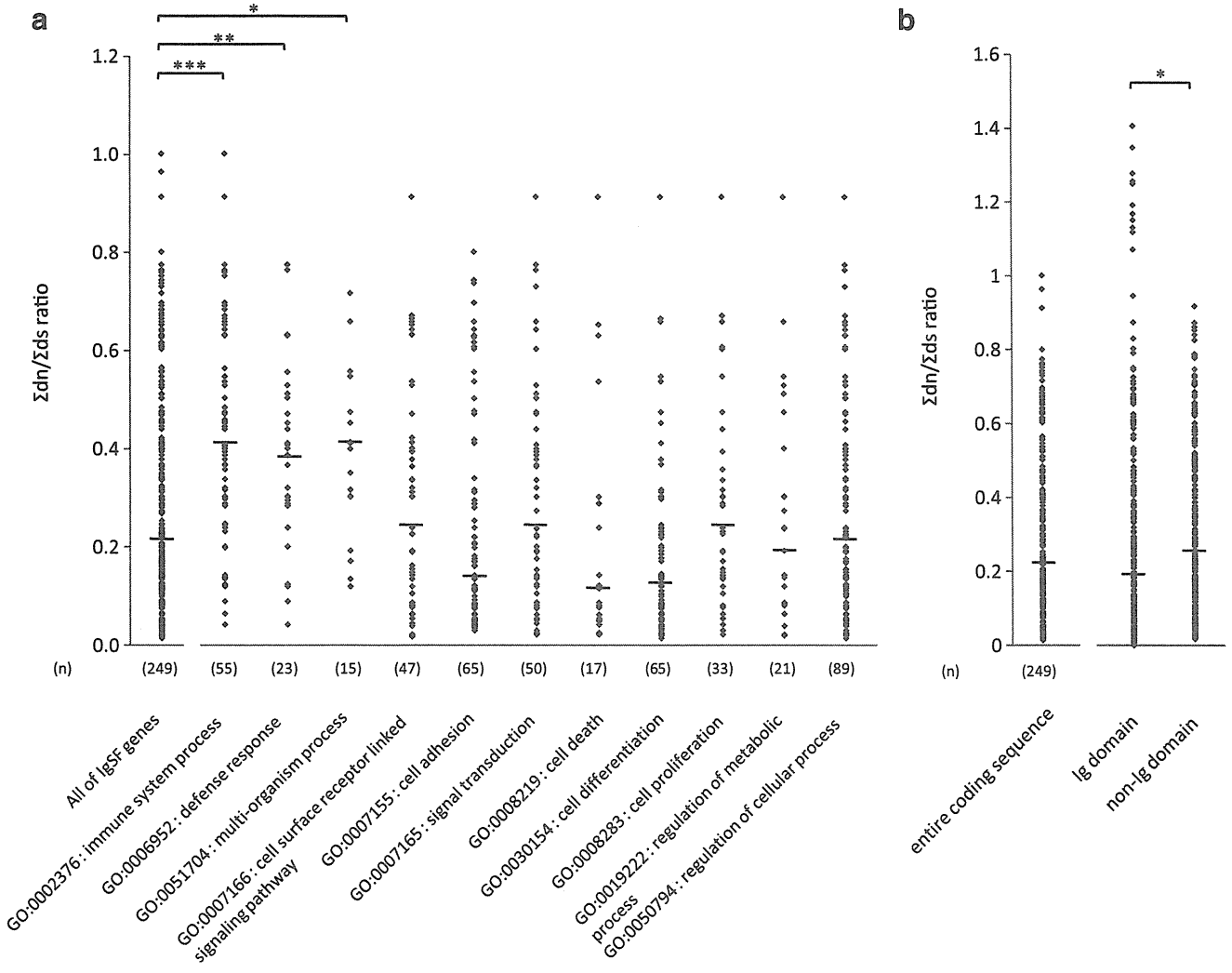

Fig. 1 Σdn/Σds ratio of IgSF genes. **a** The IgSF genes were categorized by gene ontology. **b** Σdn/Σds ratios for the entire coding region, Ig domain, and non-Ig domain. The Σdn/Σds ratios were calculated by Bn-Bs program. *Bars* indicate median values of Σdn/Σds ratio for each group. An *asterisk* indicates that there was significant difference between two groups (*$p<0.05$, **$p<0.01$, ***$p<0.001$)

The coding segments of IgSF genes were divided into two segments in each gene; one was the segment encoding the Ig domain, whereas the other was the coding region other than the Ig domain (non-Ig domain). The Σdn/Σds ratios were also separately calculated for the Ig and non-Ig domains in the IgSF genes. As shown in Fig. 1b, the Σdn/Σds ratios for the Ig domains (median Σdn/Σds ratio=0.198; IQR, 0.070–0.420) were significantly lower than the Σdn/Σds ratios for the non-Ig domains (median Σdn/Σds ratio=0.242; IQR, 0.125–0.456, $p=2.10\times10^{-2}$). Interestingly, despite the lower levels of Σdn/Σds ratio for the Ig domains, the Σdn/Σds ratios of Ig domains from 11 genes, *LAIR1*, *CD3G*, *CD3E*, *CEACAM7*, *ICAM4*, *CD244*, *CD4*, *CD3D*, *CD7*, *SLAMF6*, and *BTLA*, were over 1.0 and higher than mean Σdn/Σds of the non-Ig domains, although none of them was statistically significant.

Non-synonymous/synonymous substitution ratios of IgSF genes in primate lineages

Average values of ω for different lineages, including human, chimpanzee, orangutan, rhesus macaque, and common marmoset lineages, were calculated. First, we made a long sequence by connecting the coding sequences from all IgSF genes to calculate the average ω value, because there were many IgSF genes in which the ds was 0, which made it impossible to determine the exact value of ω. Then, the values of ω at intervals of approximately 20,000 bases were calculated. The average values of ω for each branch of the five-specie phylogeny were also calculated by using the Bn-Bs program. It was found that the average values of ω for the entire coding region was 0.241 in the human lineage, 0.277 in the chimpanzee lineage, 0.225 in the orangutan lineage, 0.234 in the rhesus

lineage, and 0.281 in the marmoset lineage. We also estimated the average value for the immune-related IgSF genes, i.e. the IgSF genes categorized into GO:0002376, GO:0006952, and GO:0051704, and the average value for the other IgSF genes; 0.285 and 0.225 in the human lineage, 0.381 and 0.236 in the chimpanzee lineage, 0.307 and 0.193 in the orangutan lineage, 0.370 and 0.181 in the rhesus lineage, and 0.473 and 0.206 in the marmoset lineage, respectively. In addition, essentially identical results were obtained by using the PAML program (ESM Figs. 1 and 2).

IgSF genes under the pressure of positive natural selection

The dn and ds values for the entire coding regions of IgSF genes in each lineage were calculated by using the Bn-Bs program and plotted in Fig. 2, where the genes with dn values higher than ds values were distributed in the upper diagonal portion. We performed statistical tests by using both the Bn-Bs and PAML programs. When a statistically significant level (p value less than 0.05) was obtained by the Bn-Bs program, further analyses by using the PAML program were done. As the results, five IgSF genes were suggested to have been under the significant positive selection; *SIGLEC5* [Z score=2.70 (p=0.003), chi-square value=5.32 (p=0.021)], and *SLAMF6* [Z score=1.69 (p=0.046), chi-square value=3.93 (p=0.048)] in the human lineage, *CD33* [Z score=2.43 (p=0.008), chi-square value=4.90 (p=0.027)] in the chimpanzee lineage, and *CD3E* [Z score=2.67 (p=0.004), chi-square value=9.04 (p=0.003)] and *CEACAM8* [Z score=2.08 (p=0.019), chi-square value=6.52 (p=0.011)] in the human–chimpanzee–orangutan ancestor lineage (Table 1). No gene under the significant control of positive selection was identified in the human–chimpanzee ancestor lineage, the orangutan lineage, the rhesus lineage, or the marmoset lineage. We also calculated the dn and ds values in the seven lineages for the Ig and non-Ig domains in the IgSF genes. Five genes, *CD3E*, *CD3G*, *FCER1A*, *CD48*, and *CD4*, and four genes, *SIGLEC5*, *TIM4*, *FCGR2A*, and *CD3E*, were suggested to be under the positive natural selection in the Ig and non-Ig domains, respectively (Table 1, ESM Figs. 3 and 4).

It should be noted here that we did not perform a multiple-test adjustment, such as a strict Bonferroni correction, and thus the levels of statistical significance were marginal. Nevertheless, we obtained significant results in the analyses done by two different programs, the Bn-Bs and PAML programs, instead of performing the multiple-test adjustment. In the Bn-Bs program, statistical significance of the difference between the dn and ds values were examined by Z test. On the other hand, in the PAML program, detection of positive selection acting on particular branch was based on the likelihood ratio test.

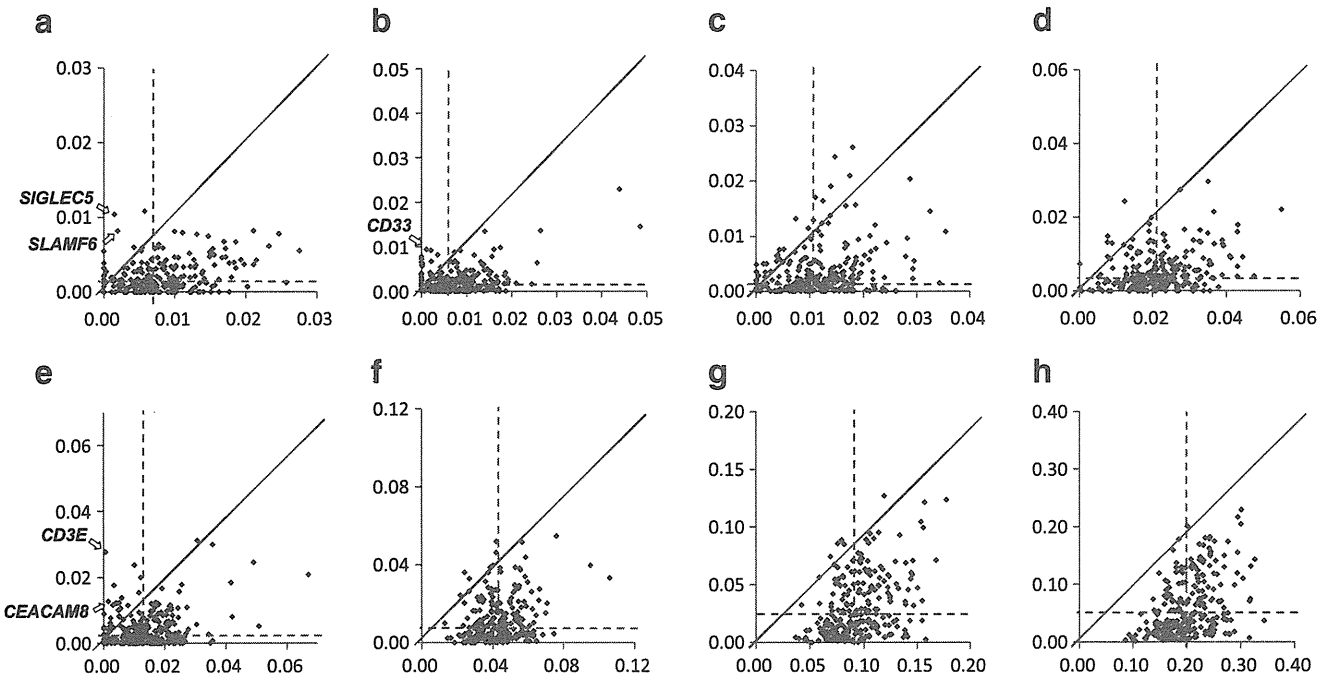

Fig. 2 Pairwise comparison plots of dn and ds values for the entire coding regions of IgSF genes in primate lineages. The values of dn (*vertical axis*) and ds (*horizontal axis*) for each primate lineage and their summation (Σ) were calculated by Bn-Bs program. *Dotted lines* indicate the average values of dn or ds. Arrows indicate the IgSF genes which were identified as to be under a positive selection by the analysis of both Bn-Bs and PAML program analyses. **a** Human lineage, **b** chimpanzee lineage, **c** human–chimpanzee ancestor lineage, **d** orangutan lineage, **e** human–chimpanzee–orangutan ancestor lineage, **f** rhesus lineage, **g** marmoset lineage, **h** Σ

Table 1 IgSF genes suggested to be under the positive selection in the course of primate evolution

Region	Gene name	Accession	BnBs			PAML			Lineage[c]
			ω (dn, ds)	*Z* score	*p* Value	ω	Chi-square	*p* Value	
Entire coding region	*SIGLEC5*	NM_003830	6.90 (0.010, 0.002)	2.70	0.003	nc	5.32	0.021	H
	SLAMF6	NM_052931	4.19 (0.008, 0.002)	1.68	0.046	nc	3.93	0.048	H
	FCGR3A	NM_000569	nc[a] (0.008, 0.000)	2.56	0.005	nc	2.80	ns[b]	C
	CD33	NM_001772	8.40 (0.009, 0.001)	2.43	0.008	nc	4.90	0.027	C
	TIM4	NM_138379	29.43 (0.006, 0.000)	2.12	0.017	nc	3.81	ns	C
	IL11RA	NM_001142784	nc (0.004, 0.000)	2.01	0.022	nc	2.24	ns	C
	FCGR2A	NM_021642	nc (0.004, 0.000)	1.99	0.023	nc	1.27	ns	C
	ICAM2	NM_000873	65.00 (0.006, 0.001>)	1.93	0.027	nc	2.16	ns	C
	AMICA1	NM_001098526	nc (0.004, 0.000)	1.77	0.039	nc	2.33	ns	C
	CD244	NM_001166663	4.34 (0.009, 0.002)	1.70	0.044	3.06	1.27	ns	C
	CD3E	NM_000733	48.67 (0.028, 0.001)	2.67	0.004	nc	9.03	0.003	HCO
	CEACAM8	NM_001816	8.73 (0.013, 0.001)	2.08	0.019	nc	6.52	0.011	HCO
	BTLA	NM_181780	5.30 (0.018, 0.003)	1.69	0.046	4.74	2.13	ns	HCO
Ig domain	*CD244*	NM_001166663	6.20 (0.022, 0.004)	1.96	0.025	nc	3.77	ns	H
	FCGR3A	NM_000569	19.98 (0.009, 0.001>)	1.89	0.029	nc	1.98	ns	H
	SLAMF6	NM_001184714	49.70 (0.011, 0.001>)	1.89	0.029	nc	1.95	ns	H
	CD244	NM_001166663	nc (0.017, 0.000)	2.51	0.006	nc	2.08	ns	C
	FCGR3A	NM_000569	nc (0.008, 0.000)	2.03	0.021	nc	1.69	ns	C
	BTN2A2	NM_006995	nc (0.007, 0.000)	1.77	0.038	nc	1.47	ns	C
	IGSF2	NM_004258	341.00 (0.007, 0.001>)	1.75	0.040	nc	2.35	ns	C
	VSIG10L	NM_001163922	nc (0.017, 0.000)	1.66	0.048	nc	0.63	ns	C
	LILRA4	NM_012276	nc (0.035, 0.000)	2.64	0.004	nc	2.50	ns	HC
	SLAMF6	NM_052931	nc (0.013, 0.000)	1.80	0.036	nc	2.58	ns	HC
	TIM4	NM_138379	206.00 (0.014, 0.000)	1.71	0.043	nc	1.11	ns	HC
	PECAM1	NM_000442	nc (0.009, 0.000)	2.31	0.010	nc	2.10	ns	O
	CEACAM7	NM_006890	292.40 (0.015, 0.000)	1.79	0.036	nc	1.75	ns	O
	BTLA	NM_181780	nc (0.025, 0.000)	2.43	0.008	nc	1.99	ns	HCO
	CD3E	NM_000733	16.07 (0.055, 0.003)	2.16	0.015	nc	4.43	0.035	HCO
	IGSF2	NM_004258	nc (0.009, 0.000)	1.76	0.039	nc	2.94	ns	HCO
	CD3G	NM_000073	nc (0.068, 0.000)	3.42	0.000	nc	4.74	0.029	R
	FCER1A	NM_002001	77.38 (0.029, 0.000)	2.86	0.002	nc	5.35	0.021	R
	ICAM4	NM_001544	nc (0.049, 0.000)	2.71	0.003	nc	2.89	ns	R
	CD3E	NM_000733	7.77 (0.085, 0.011)	2.62	0.004	nc	5.32	0.021	R
	SIGLEC11	NM_052884	2.08 (0.032, 0.015)	1.70	0.044	2.70	1.73	ns	R
	CD48	NM_001778	2.32 (0.197, 0.085)	2.31	0.010	2.65	4.23	0.040	M
	LRRC4	NM_022143	6.61 (0.051, 0.008)	1.79	0.037	3.59	1.69	ns	M
	CD4	NM_000616	1.92 (0.199, 0.104)	1.75	0.040	3.45	6.34	0.012	M
non-Ig domain	*SIGLEC5*	NM_003830	9.49 (0.012, 0.001)	2.66	0.004	nc	4.82	0.028	H
	TREML1	NM_178174	nc (0.005, 0.000)	1.75	0.040	nc	2.38	ns	H
	KAZALD1	NM_030929	nc (0.007, 0.000)	1.72	0.042	nc	1.72	ns	H
	CD33	NM_001772	24.03 (0.009, 0.001>)	2.17	0.015	nc	2.89	ns	C
	TIM4	NM_138379	27.66 (0.008, 0.001>)	2.09	0.018	nc	4.01	0.045	C
	IL11RA	NM_001142784	nc (0.005, 0.000)	1.79	0.037	nc	2.22	ns	C
	BTN1A1	NM_001732	52.71 (0.004, 0.001>)	1.72	0.043	nc	2.19	ns	C
	CD3E	NM_000733	154.60 (0.008, 0.001>)	1.67	0.047	nc	1.76	ns	HC
	SIRPB2	NM_001122962	10.29 (0.012, 0.001)	1.79	0.037	nc	2.89	ns	O
	FCGR2A	NM_021642	nc (0.017, 0.000)	2.64	0.004	nc	5.00	0.025	HCO
	CD3E	NM_000733	nc (0.017, 0.000)	1.66	0.048	nc	3.84	0.050	HCO

[a] *nc* Not calculated (Bs=0)

[b] *ns* Not significant ($p>0.05$)

[c] *H* human, *C* chimpanzee, *R* rhesus, *M* marmoset, *HC* human–chimpanzee ancestor, *HCO* human–chimpanzee–orangutan ancestor

Natural selection of *CD3E* and *CD3G* in primates

We further analyzed two genes, *CD3E* and *CD3G*, which were suggested to be under the positive natural selection. *CD3G* was the gene giving the lowest *p* value, and the Ig domain of CD3E underwent the positive selection. CD3E and CD3G tightly bound to each other (Xu et al. 2006). Because it is known that interacting protein pairs, such as receptor and its ligand, exhibit higher level of co-evolution than non-interacting protein pairs (Goh et al. 2000; Jothi et al. 2006; Li et al. 2005), we hypothesized that co-evolution might occur between *CD3E* and *CD3G*. To investigate the natural selection operated on these genes in the course of primate evolution, we determined protein coding sequences of *CD3E* and *CD3G* from 23 different primate species, including eight hominoids (human, chimpanzee, bonobo, gorilla, orangutan, black gibbon, white-handed gibbon, and siamang), six Old World monkeys (rhesus macaque, crab-eating macaque, hamadryas baboon, black and white colobus, silvered lutong, and dusky lutong), eight New World monkeys (common marmoset, cotton-top tamarin, red-handed tamarin, golden lion tamarin, common squirrel monkey, tufted capuchin, long-haired spider monkey, and Central American spider monkey), and one prosimian (lesser galago). After the alignment of nucleotide sequences and removal of alignment gaps, the values of dn and ds for the entire region, Ig domain, and non-Ig domain were calculated by using Bn-Bs program in each lineage of the phylogenic tree of primates.

The dn and ds values for *CD3E* in each primate lineage are indicated in Fig. 3. The dn values were larger than the ds values in several lineages which might be underwent positive selection pressure in the primate evolution. In particular, the dn values for the Ig domain were significantly

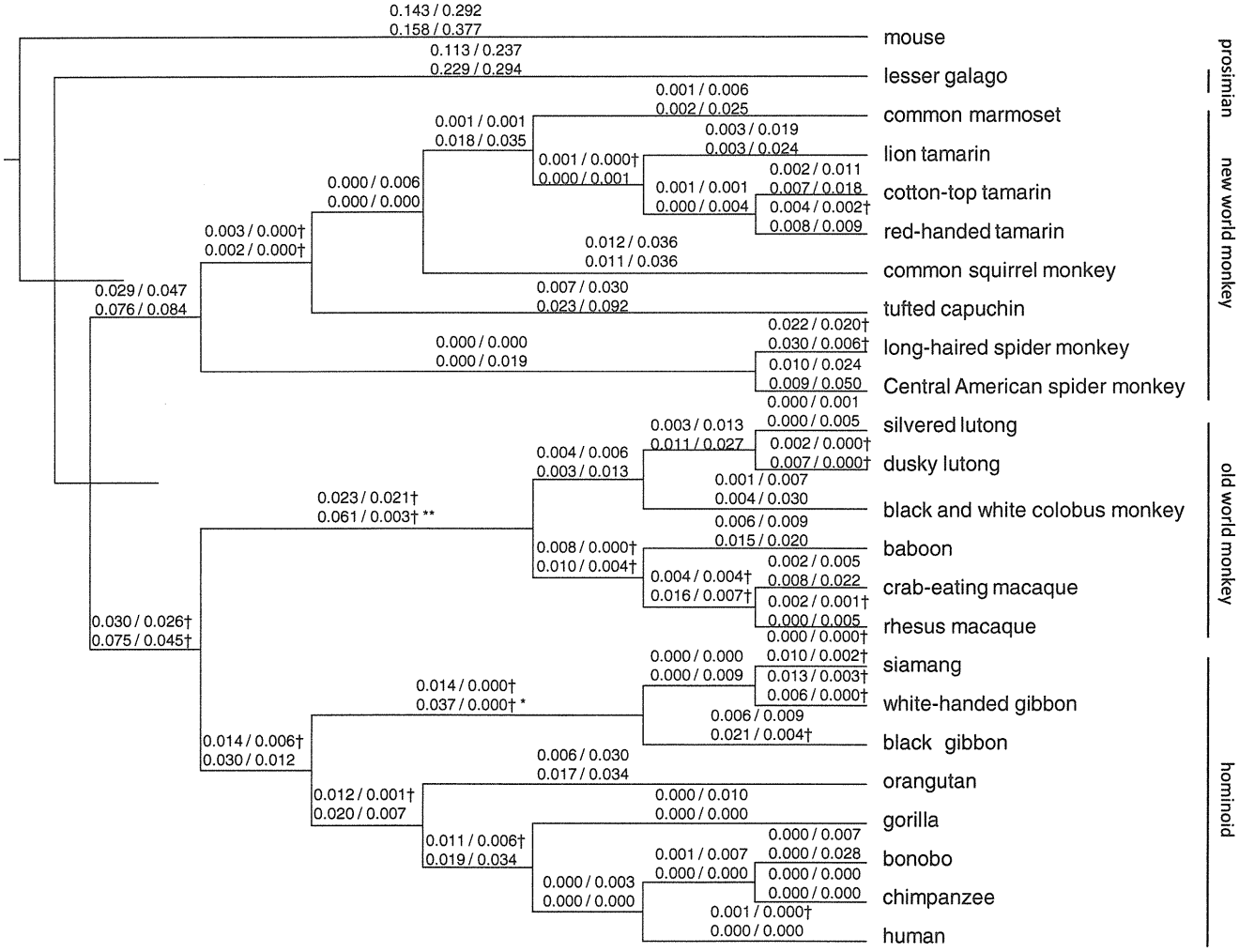

Fig. 3 Phylogenetic trees of *CD3E* in the primate evolution. *Values above branches* indicate estimated values of dn and ds per lineage by using Bn-Bs program. Values indicated in the upper parts are for the entire coding region, while values in the lower parts are for the Ig domain. *Daggers* indicate the value of dn higher than ds. An *asterisk* indicates that there was a significant difference between the dn and ds values (*$p<0.05$, **$p<0.01$, *Z* test)

larger than the ds values in two lineages: gibbon ancestor lineage [dn=0.061, ds=0.003, Z score=2.37 (p=0.009)] and Old World monkey ancestor lineage [dn=0.037, ds=0.000, Z score=1.94 (p=0.026)]. The significant positive selection on the Ig domains in the Old World monkey ancestor lineage [chi-square value=4.45 (p=0.035)] was confirmed by the PAML program, whereas it was not significant in the gibbon ancestor lineage [chi-square value=0.58 (p=0.446)]. Amino acid (AA) sequence alignment of CD3E in the primates is shown in Fig. 4. We identified three alignment gaps, all of which were in the Ig domain. Of 53 AAs of the Ig domain, approximately 30% (16/53) were evolutionary conserved among the primate species. On the other hand, approximately 70% (97/143) of AAs were conserved in the non-Ig domain, demonstrating that significantly more AA substitutions were distributed in the Ig domain ($p=2.16\times10^{-6}$). In addition, five AA sites in the Ig domain (positions at 51, 53, 72, 80, and 105 in the human sequence) were identified as possible target sites for the positive selection by the BEB method using the PAML program.

The dn and ds values for *CD3G* in each primate lineage were also measured by using the Bn-Bs program (ESM Fig. 5). The dn values for the Ig domain were significantly larger than the ds values in two lineages; Old World monkey ancestor lineage [dn=0.060, ds=0.000, Z score=3.30 (p=0.0005)] and hominoid and Old World monkey ancestor lineage [dn=0.042, ds=0.007, Z score=1.65 (p=0.049)]. The positive selection was confirmed by the PAML program in both lineages; Old World monkey ancestor lineage [chi-square value=5.32 (p=0.021)] and hominoid and Old World monkey ancestor lineage [chi-square value=4.17 (p=0.041)]. The AA alignment of CD3G from 23 primate species is shown in Fig. 4. Approximately 40% (25/56) of AAs in the Ig domain were conserved in the primate evolution, whereas

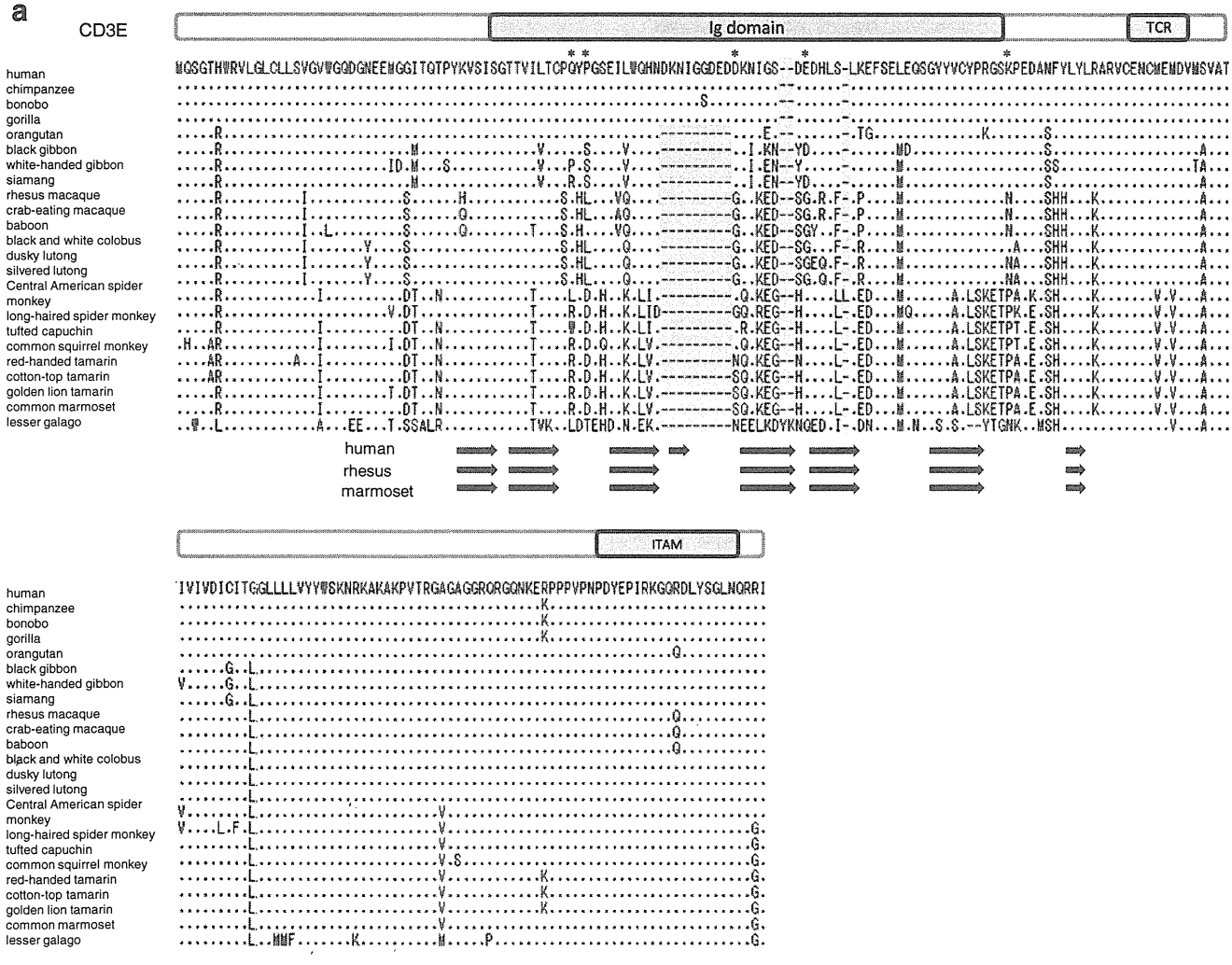

Fig. 4 Alignments of CD3E (**a**) and CD3G (**b**) amino acid sequences from 23 primate species. *Dots* indicate the identities to the human reference sequence, while hyphens indicate alignment gaps. TCR indicates amino acid sites at which the CD3 molecules interact with T cell receptor. ITAM represents the immune-tyrosine activation motif. *Arrows* indicated under the amino acid sequences are β-strand structures modeled by the SWISS-MODEL program. *Asterisks* indicate AA sites identified as being under the significant positive selection (p<0.05)

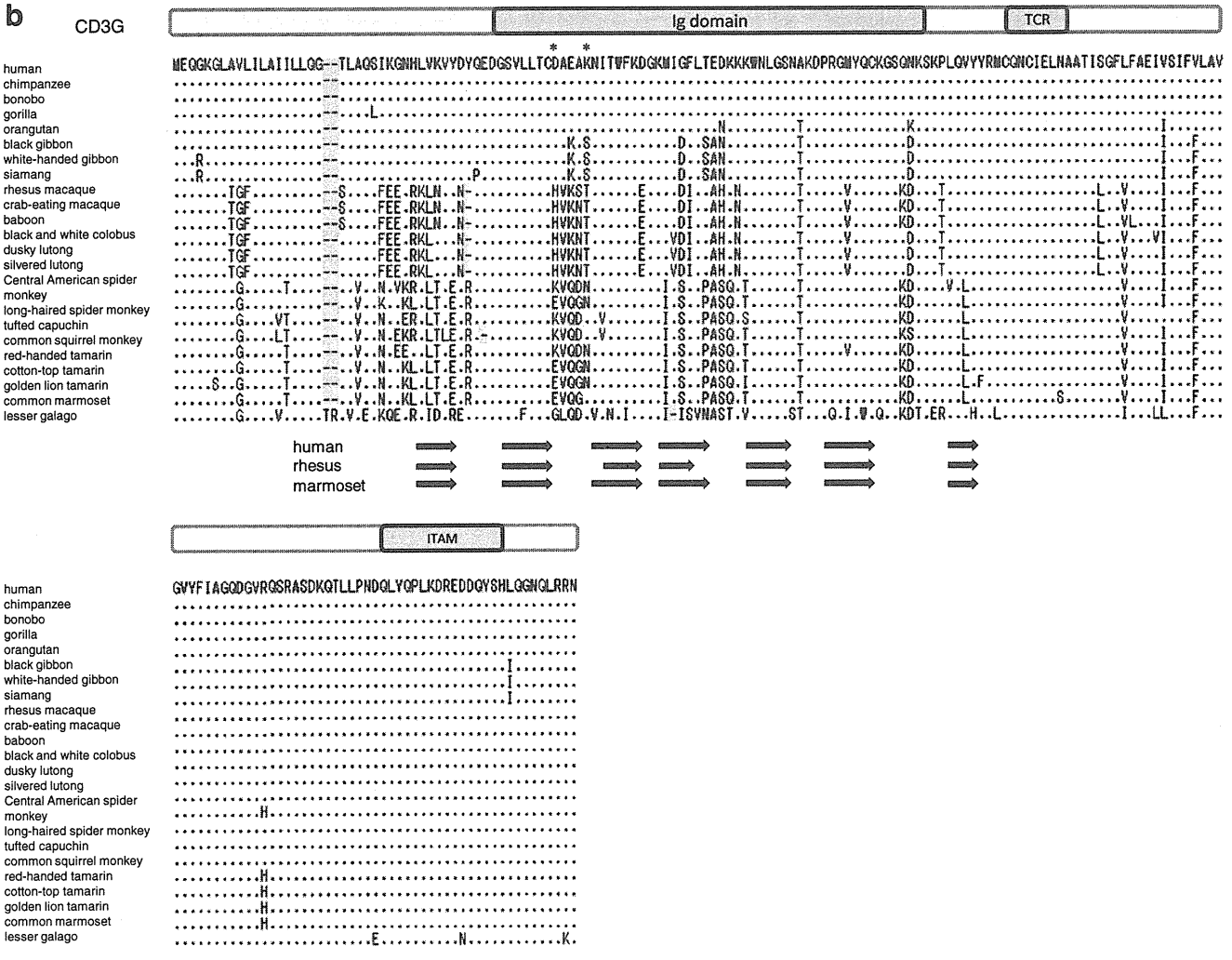

Fig. 4 (continued)

approximately 60% (81/128) of AAs in the non-Ig domain were conserved, demonstrating that the AA changes were significantly more frequent in the Ig domain (p=0.019). Two AA sites in the Ig domain (positions at 47 and 51 in the human sequence) were identified as significant target sites for the positive selection. These lines of evidence suggested that the pressure of positive Darwinian selection had shaped the structure of Ig domains in CD3E and CD3G during the course of primate evolution.

Discussion

Members of the IgSF have a wide variety of cellular activities and were classified into 11 functional categories based on the Gene Ontology database (http://www.geneontology.org/). When the association between the IgSF functional categories and the Σdn/Σds ratios were analyzed, three GO categories tightly linked to the immune system, i.e., GO:0002376: immune system process, GO:0006952: defense response, and GO:0051704: multi-organism process, showed much higher values for the Σdn/Σds ratio than the average value of the IgSF genes. It has been reported that the evolutionary rate of immune-related genes is higher than the other genes (Gibbs et al. 2007; Kosiol et al. 2008; Nielsen et al. 2005; Yu et al. 2006). The rapid evolution of immune-related genes might be a direct consequence of a complex selection pressure exerted by infectious diseases, autoimmunity, and tumors (Barreiro and Quintana-Murci 2010). On the other hand, as shown in Fig. 1a, the Σdn/Σds ratios of genes linked to three functional categories, GO:0007155: cell adhesion, GO:0007165: signal transduction, and GO:0008219: cell death, were comparable to those for the functional categories other than the immune-related genes. Genes in these three categories have also been reported to have a higher non-synonymous/synonymous substitution ratio than the genes in the other categories (Clark et al. 2003; Gibbs et al. 2007; Nielsen et al. 2005).

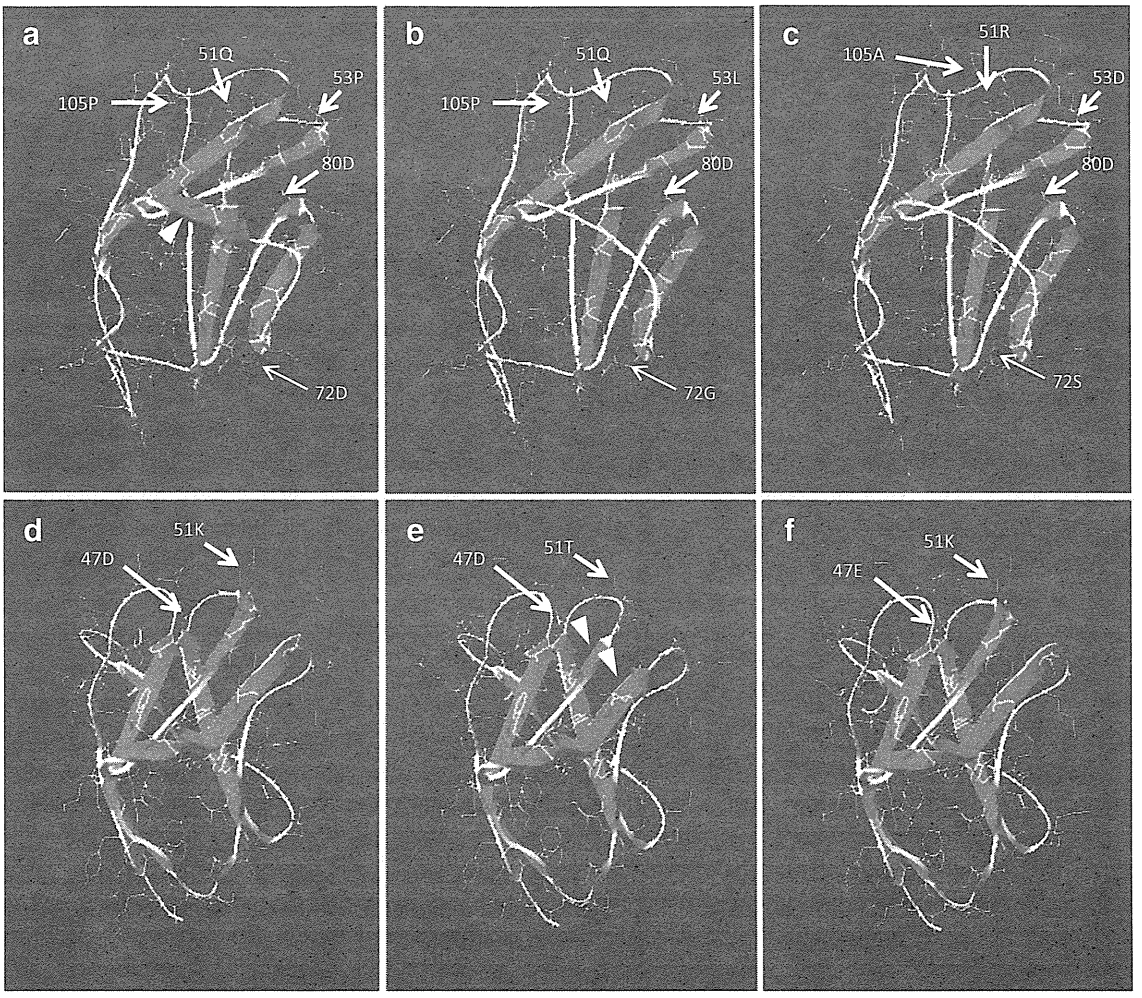

Fig. 5 Three-dimensional structures of CD3E and CD3G modeled by SWISS-MODEL. Arrows indicate amino acid sites identified as being under positive selection by using the BEB method in the PAML program. **a** Human CD3E. An *arrowhead* indicates a β-strand which is unique to human CD3E. **b** rhesus macaque CD3E, **c** marmoset CD3E, **d** human CD3G, **e** rhesus macaque CD3G. *Arrowheads* indicate short strands of β-strand which are unique to rhesus CD3G. **f** marmoset CD3G

It has been reported that the average value of ω in the human lineage is higher than that in the other primate lineages (Ellegren 2008; Gibbs et al. 2007; Kosiol et al. 2008). The differences in the ω among the primate lineages may be attributable to the differences in the effective population size during the course of evolution (Bakewell et al. 2007). Interestingly, in our study, the average value of ω for immune-related genes in the human lineage was the lowest among the primate lineages. Because previous studies suggest that the rapid evolution of the immune-related genes may be due to a direct consequence of complex selection pressure exerted by infectious reagents including microbes and viruses (Barreiro and Quintana-Murci 2010), the observation in our study led us to a hypothesis that in the course of human evolution there might be fewer challenges from pathogens than the other primates, in part due to a shorter course of human evolution. In support of this, it was reported that humans had faced relatively fewer challenges from retroviruses and that humans were consequently at present more susceptible to retrovirus infections than the other primates (Sawyer et al. 2006). However, such a slow evolution of human lineage might also be caused by other factors such as long generation time and small population size.

We identified 11 genes possibly having undergone positive selective pressure (Table 1). Among them, *SIGLEC5*, *CD33*, *CD4*, and *CD3E* have been reported to be genes under the pressure of positive selection in the primate evolution (Angata et al. 2004; Gibbs et al. 2007; Zhang et al. 2008). These genes play crucial roles in the innate and adaptive immune systems, and infectious pathogens might have exerted selective pressure on them (Angata 2006; Crocker et al. 2007), because SIGLEC5, CD33 (SIGLEC3), and CD4 are cell surface receptors for microorganisms.

CD3E and *CD3G* encode the components of T cell antigen receptor (TCR) complex, TCR–CD3 complex. The

TCR–CD3 complex plays a key role in the regulation of immune system through the recognition of antigenic peptides presented by MHC molecules, and mutations in either *CD3E* or *CD3G* are known to cause primary immunodeficiency in humans (Buckley 2004; de Saint et al. 2004; Recio et al. 2007; Sun et al. 2001). Furthermore, previous studies have revealed the role of Ig domains of CD3E and CD3G. For example, cell surface expression of the stable TCR on mature T cell, assembly of the components of T cell antigen receptor complex and T cell activity are regulated by the Ig domains of CD3E and CD3G (Dietrich et al. 1996; Guy and Vignali 2009; Sun et al. 2001).

Because we hypothesized that co-evolution might occur between *CD3E* and *CD3G*, protein coding sequences of these *CD3* genes from 23 primate species were determined, and it was suggested that the Ig domains of both CD3E and CD3G have undergone positive Darwinian selection pressure in the primate evolution, especially in the Old World monkey ancestor lineage. However, the role of Ig domains in the direct interaction of CD3E and CD3G has not been reported. Although the direct demonstration of functional interaction are needed to clarify the impacts of AA substitutions on the function of CD3E and CD3G, we modeled three-dimensional structures of the Ig domains of CD3E using SWISS-MODEL, an Automated Comparative Protein Modeling Server (http://swissmodel.expasy.org/SWISS-MODEL.html; Bordoli et al. 2009). As shown in Fig. 5, eight, seven, and seven β-strands were found in the Ig domains of CD3E from human, rhesus macaque, and common marmoset, respectively, and the three AA insertion/deletion observed in the Ig domain appeared to have a strong impact on the modeled structure. In addition, it was possible that five AA sites in the Ig domain, which were identified as target sites for the positive selection, would change the Ig domain structure. It has been reported that highly conserved CXXCXEXXX motifs in the CD3 family play an important role in the molecular interactions among components of the TCR–CD3 complex (Borroto et al. 1998; Xu et al. 2006). Because the Ig domains are localized just upstream of the N-terminal of CXXCXEXXX motifs, drastic structural changes in the Ig domains might affect the functional properties of CD3E and CD3G. It is likely that such structural changes would affect the stability of the TCR–CD3 complex and the expression level in mature T cells (Call and Wucherpfennig 2004; Guy and Vignali 2009; Wang et al. 1998).

What was the extent of selective pressure exerted on *CD3E* and *CD3G* in the course of primate evolution? Given that the TCR–CD3 complex plays a crucial role in the regulation of immune system, infectious diseases and autoimmunity have been postulated to be the strongest selective pressures (Robins et al. 2009; Sun et al. 2001). It is widely accepted that the susceptibility to infectious pathogens, such as *Mycobacterium tuberculosis bacilli* and HIV-1, are different among primate species (Lyashchenko et al. 2008; Song et al. 2005). Because the dn values for the Ig domains of *CD3E* and *CD3G* are significantly greater than the ds values in the Old World monkey ancestor lineage, their ancestors might have been exposed to powerful selective pressure. To clarify the selective pressure exerted on *CD3E* and *CD3G*, further study on phenotypic differences, such as the relative susceptibilities to infectious pathogens and/or autoimmune disease, among various primate species is needed.

In conclusion, we investigated the molecular evolution of IgSF genes in primates. The study has demonstrated that the immune-related IgSF genes have high non-synonymous/synonymous substitution rates, and those 11 IgSF genes, namely *SIGLEC5*, *SLAMF6*, *CD33*, *CD3E*, *CEACAM8*, *CD3G*, *FCER1A*, *CD48*, *CD4*, *TIM4*, and *FCGR2A*, may undergo the positive selective pressure in the primate evolution.

Acknowledgments This work was supported in part by research grants from the Ministry of Health, Labor and Welfare, Japan, the Japan Health Science Foundation, the program of Founding Research Centers for Emerging and Reemerging Infection Disease, the program of Research on Publicly Essential Drugs and Medical Devices, by grant-in-aids for Scientific research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan, by a grant from the Life Science Institute Foundation, Japan, and by grants for India–Japan Cooperative Science Program from Japan Society for the Promotion of Science (JSPS), Japan and Department of Science and Technology (DST), India.

References

Angata T, Margulies EH, Green ED, Varki A (2004) Large-scale sequencing of the CD33-related Siglec gene cluster in five mammalian species reveals rapid evolution by multiple mechanisms. Proc Natl Acad Sci USA 101:13251–6

Angata T (2006) Molecular diversity and evolution of the Siglec family of cell-surface lectins. Mol Divers 10:555–66

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC, Richardson JE, Ringwald M, Rubin GM, Sherlock G (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet 25:25–9

Bakewell MA, Shi P, Zhang J (2007) More genes underwent positive selection in chimpanzee evolution than in human evolution. Proc Natl Acad Sci USA 104:7489–94

Barclay AN (2003) Membrane proteins with immunoglobulin-like domains—a master superfamily of interaction molecules. Semin Immunol 15:215–23

Barreiro LB, Quintana-Murci L (2010) From evolutionary genetics to human immunology: how selection shapes host defence genes. Nat Rev Genet 11:17–30

Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, Haussler D, Miller W (2004) Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res 14:708–15

Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T (2009) Protein structure homology modeling using SWISS-MODEL workspace. Nat Protoc 4:1–13

Borroto A, Mallabiabarrena A, Albar JP, Martinez AC, Alarcon B (1998) Characterization of the region involved in CD3 pairwise interactions within the T cell receptor complex. J Biol Chem 273:12807–16

Buckley RH (2004) The multiple causes of human SCID. J Clin Invest 114:1409–11

Call ME, Wucherpfennig KW (2004) Molecular mechanisms for the assembly of the T cell receptor-CD3 complex. Mol Immunol 40:1295–305

Chatterjee HJ, Ho SY, Barnes I, Groves C (2009) Estimating the phylogeny and divergence times of primates using a supermatrix approach. BMC Evol Biol 9:259

Clark AG, Glanowski S, Nielsen R, Thomas PD, Kejariwal A, Todd MA, Tanenbaum DM, Civello D, Lu F, Murphy B, Ferriera S, Wang G, Zheng X, White TJ, Sninsky JJ, Adams MD, Cargill M (2003) Inferring nonneutral evolution from human–chimp–mouse orthologous gene trios. Science 302:1960–3

Consortium CSaA (2005) Initial sequence of the chimpanzee genome and comparison with the human genome. Nature 437:69–87

Crocker PR, Paulson JC, Varki A (2007) Siglecs and their roles in the immune system. Nat Rev Immunol 7:255–66

de Saint BG, Geissmann F, Flori E, Uring-Lambert B, Soudais C, Cavazzana-Calvo M, Durandy A, Jabado N, Fischer A, Le Deist F (2004) Severe combined immunodeficiency caused by deficiency in either the delta or the epsilon subunit of CD3. J Clin Invest 114:1512–7

Dietrich J, Neisig A, Hou X, Wegener AM, Gajhede M, Geisler C (1996) Role of CD3 gamma in T cell receptor assembly. J Cell Biol 132:299–310

Ellegren H (2008) Comparative genomics and the study of evolution by natural selection. Mol Ecol 17:4586–96

Gibbs RA et al (2007) Evolutionary and biomedical insights from the rhesus macaque genome. Science 316:222–34

Goh CS, Bogan AA, Joachimiak M, Walther D, Cohen FE (2000) Co-evolution of proteins with their interaction partners. J Mol Biol 299:283–93

Guy CS, Vignali DA (2009) Organization of proximal signal initiation at the TCR:CD3 complex. Immunol Rev 232:7–21

Halaby DM, Mornon JP (1998) The immunoglobulin superfamily: an insight on its tissular, species, and functional diversity. J Mol Evol 46:389–400

Jothi R, Cherukuri PF, Tasneem A, Przytycka TM (2006) Co-evolutionary analysis of domains in interacting proteins reveals insights into domain–domain interactions mediating protein-protein interactions. J Mol Biol 362:861–75

Kent WJ, Baertsch R, Hinrichs A, Miller W, Haussler D (2003) Evolution's cauldron: duplication, deletion, and rearrangement in the mouse and human genomes. Proc Natl Acad Sci USA 100:11484–9

Kosiol C, Vinar T, da Fonseca RR, Hubisz MJ, Bustamante CD, Nielsen R, Siepel A (2008) Patterns of positive selection in six mammalian genomes. PLoS Genet 4:e1000144

Lander ES et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23:2947–8

Li Y, Wallis M, Zhang YP (2005) Episodic evolution of prolactin receptor gene in mammals: coevolution with its ligand. J Mol Endocrinol 35:411–9

Lyashchenko KP, Greenwald R, Esfandiari J, Chambers MA, Vicente J, Gortazar C, Santos N, Correia-Neves M, Buddle BM, Jackson R, O'Brien DJ, Schmitt S, Palmer MV, Delahay RJ, Waters WR (2008) Animal-side serologic assay for rapid detection of *Mycobacterium bovis* infection in multiple species of free-ranging wildlife. Vet Microbiol 132:283–92

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–26

Nielsen R, Bustamante C, Clark AG, Glanowski S, Sackton TB, Hubisz MJ, Fledel-Alon A, Tanenbaum DM, Civello D, White TJ, Sninsky JJ, Adams MD, Cargill M (2005) A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol 3:e170

Otey CA, Dixon R, Stack C, Goicoechea SM (2009) Cytoplasmic Ig-domain proteins: cytoskeletal regulators with a role in human disease. Cell Motil Cytoskeleton 66:618–34

Recio MJ, Moreno-Pelayo MA, Kilic SS, Guardo AC, Sanal O, Allende LM, Perez-Flores V, Mencia A, Modamio-Hoybjor S, Seoane E, Regueiro JR (2007) Differential biological role of CD3 chains revealed by human immunodeficiencies. J Immunol 178:2556–64

Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, Riddell SR, Warren EH, Carlson CS (2009) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. Blood 114:4099–107

Rzhetsky A, Nei M (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. Mol Biol Evol 10:1073–95

Sawyer SL, Wu LI, Akey JM, Emerman M, Malik HS (2006) High-frequency persistence of an impaired allele of the retroviral defense gene TRIM5alpha in humans. Curr Biol 16:95–100

Song B, Javanbakht H, Perron M, Park DH, Stremlau M, Sodroski J (2005) Retrovirus restriction by TRIM5alpha variants from Old World and New World primates. J Virol 79:3930–7

Sun ZJ, Kim KS, Wagner G, Reinherz EL (2001) Mechanisms contributing to T cell receptor signaling and assembly revealed by the solution structure of an ectodomain fragment of the CD3 epsilon gamma heterodimer. Cell 105:913–23

Wang B, Wang N, Salio M, Sharpe A, Allen D, She J, Terhorst C (1998) Essential and partially overlapping role of CD3gamma and CD3delta for development of alphabeta and gammadelta T lymphocytes. J Exp Med 188:1375–80

Wong WS, Yang Z, Goldman N, Nielsen R (2004) Accuracy and power of statistical methods for detecting adaptive evolution in protein coding sequences and for identifying positively selected sites. Genetics 168:1041–51

Xu C, Call ME, Wucherpfennig KW (2006) A membrane-proximal tetracysteine motif contributes to assembly of CD3deltaepsilon and CD3gammaepsilon dimers with the T cell receptor. J Biol Chem 281:36977–84

Yang Z (2005) The power of phylogenetic comparison in revealing protein function. Proc Natl Acad Sci USA 102:3179–80

Yang Z (2007) PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol 24:1586–91

Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol Biol Evol 17:32–43

Yang Z, Wong WS, Nielsen R (2005) Bayes empirical bayes inference of amino acid sites under positive selection. Mol Biol Evol 22:1107–18

Yu XJ, Zheng HK, Wang J, Wang W, Su B (2006) Detecting lineage-specific adaptive evolution of brain-expressed genes in human using rhesus macaque as outgroup. Genomics 88:745–51

Zhang J, Rosenberg HF, Nei M (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. Proc Natl Acad Sci USA 95:3708–13

Zhang ZD, Weinstock G, Gerstein M (2008) Rapid evolution by positive Darwinian selection in T-cell antigen CD4 in primates. J Mol Evol 66:446–56

Immunogenetics (2011) 63:501–509
DOI 10.1007/s00251-011-0531-y

ORIGINAL PAPER

ULBP4/RAET1E is highly polymorphic in the Old World monkey

Taeko K. Naruse · Yukiko Okuda · Kazuyasu Mori · Hirofumi Akari · Tetsuro Matano · Akinori Kimura

Received: 22 February 2011 / Accepted: 21 April 2011 / Published online: 7 May 2011

Abstract Natural-killer group 2 member D (NKG2D) is an activating receptor that plays an important role in the immune response mediated by NK cells, $\gamma\delta^+$ T cells, and $CD8^+$ T cells. In humans, MHC class I chain-related genes and UL-16 binding protein (ULBP)/retinoic acid early transcript 1 (REAT1) gene family encode ligands for NKG2D. The rhesus and crab-eating macaques, which belong to the Old World monkeys, are widely used as non-human primate models in medical researches on the immunological process. In the present study, we investigated the polymorphisms of *ULBP4/RAET1E*, a member of the *ULBP/RAET1* family, and found 25 and 14 alleles from the rhesus and crab-eating macaques, respectively, of which diversities were far more extended than in humans. A phylogenetic study suggested that the allelic diversification of ULBP4/RAET1E predated the divergence of rhesus and crab-eating macaques.

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

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

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

T. Matano
International Research Center for Infectious Diseases,
The Institute of Medical Science, The University of Tokyo,
Tokyo, Japan

Keywords Rhesus macaque · Crab-eating macaque · ULBP4/RAET1E · NKG2D · Polymorphism

Introduction

Non-human primates, such as rhesus and crab-eating macaques, are important animal models for the study of infectious diseases, autoimmune diseases, and organ transplantation. These macaques are members of the Old World monkeys, and it has been reported that the genetic diversity in the rhesus macaque is quite unique, that is, more than 60% of the rhesus macaque-specific expansions are found in the protein coding sequences (Gibbs et al. 2007). To 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. It has been reported that the gene copy number in the major histocompatibility complex (MHC) loci in the rhesus and crab-eating macaques is higher than that in humans (Kulski et al. 2004; Gibbs et al. 2007; Otting et al. 2007). In addition, the extent of genetic diversity differed, in part, depending on the geographic areas, and we have reported that the diversity of MHC class I genes in the rhesus macaque is considerably different depending on habitat (Naruse et al. 2010).

Because the innate immune system is involved in the response to environmental pathogens, it is necessary to consider the function of natural killer (NK) cells in the experimental animal models. Natural-killer group 2 member D (NKG2D), a C-type lectin molecule, is an activating receptor expressed on the cell surface of NK, $\gamma\delta^+$, and $CD8^+$ T cells, which plays an important role in the immune response (Wu et al. 1999; Raulet 2003). In humans, MHC class I chain-related genes (MIC) and UL-16 binding protein (ULBP)/retinoic acid early transcript 1 (REAT1)

gene family are known to encode ligands for NKG2D (Bauer et al. 1999; Cosman et al. 2001; Chalupny et al. 2003; Bacon et al. 2004). These ligand molecules are usually stress-inducible, and their recognition by NKG2D can lead to the activation of NK cells, consequently killing virus-infected and tumor cells (Pende et al. 2002; Eagle et al. 2006; Pappworth et al. 2007; Ward et al. 2007).

The human *ULBP/RAET1* gene family is located on chromosome 6q24.2, which is composed of ten members including six functional genes, *ULBP1*, *2*, *3*, *4*, *5*, and *6*, corresponding to *RAET1I*, *H*, *N*, *E*, *G*, and *L*, respectively (Radosavljevic et al. 2001; Chalupny et al. 2003; Eagle et al. 2009a, b). In addition, several sequence polymorphisms in each *ULBP* gene have been identified (Romphruk et al. 2009; Antoun et al. 2010). Although it is evident that the cell surface expression of the ligand molecules on target cells is differentially regulated (Eagle et al. 2006), genetic polymorphisms in the coding regions might have a functional impact. We have previously investigated the genetic polymorphisms of *ULBP/RAET1* genes and have found that the *ULBP4/RAET1E* gene is the most polymorphic, with the allelic distribution differing among ethnic groups (Romphruk et al. 2009).

On the other hand, rhesus macaque *ULBP4/RAET1E* (GenBank: NW_001116520) is mapped on the long arm of chromosome 4 (i.e., positions from 31, 164, 822 to 31, 175, 032 of chromosome 4 in the rhesus genome; data obtained from the UCSC Genome Browser at http://genome.ucsc.edu/cgi-bin/hgGateway; Gibbs et al. 2007). However, its genetic polymorphisms are poorly characterized, although the MIC gene polymorphisms are well studied in the rhesus macaque (Seo et al. 1999, 2001; Doxiadis et al. 2007; Averdam et al. 2007). In the present study, we investigated the polymorphisms of *ULBP4/RAET1E* in rhesus and crab-eating macaques. This is the first report demonstrating the extreme diversity of the NKG2D ligand in the Old World monkey.

Materials and methods

Animals

A total of 38 rhesus macaques from seven lineages previously analyzed for the MHC polymorphisms (Naruse et al. 2010) and 24 crab-eating macaques from five lineages were the subjects. They were maintained in the breeding colonies in Japan. The founders of the rhesus macaque colonies were captured in Myanmar and Laos, whereas the founders of crab-eating macaque 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 Institutes of Health (NIH publication 85–23, revised 1985) and were subjected to prior approval by the local animal protection authority.

DNA extraction and sequencing analysis

Genomic DNAs from B lymphoblastoid cell lines of the rhesus macaque (Naruse et al. 2010) and from whole blood sample of the crab-eating macaque were extracted by using the QuickGene DNA kit (Fujifilm, Tokyo, Japan) according to the manufacturer's instructions. The genomic gene for *ULBP4/RAET1E* of rhesus and crab-eating macaques was amplified by polymerase chain reaction (PCR) with a primer pair designed for the region spanning from introns 1 to 3 of the rhesus gene (NC007861), ULBP4F (5′-TGGGCCTCTTCCCCTGTCC) and ULBP4R (5′-GTGGGAGGTGGGATGGG), using FastStart Taq DNA polymerase (Roche, Mannheim, Germany). The PCR condition was composed of the following steps: denaturation at 95°C for 4 min; 30 cycles of 95°C for 30 s, 63°C for 30 s, and 72°C for 45 s; and additional extension at 72°C for 7 min. The PCR products of about 1,200 bp in length were cloned into pSTBlue-1 AccepTer vector (Novagen, WI, USA) according to the manufacturer's instructions and were transformed to Nova Blue Single™ competent cells (Merck4Biosciences Japan, Tokyo, Japan). Ten to 20 independent transformant colonies were picked up for each sample and subjected to sequencing on both strands by using a BigDye Terminator cycling system and an ABI 3730 automated sequence analyzer (Applied Biosystems, CA, USA).

Data analyses

Nucleotide sequences of *ULBP4/RAET1E* from cloned DNAs were aligned using the Genetyx software package (version 8.0, Genetyx Corp., Japan). If at least three clones from independent PCR or from different individuals showed identical sequences, the sequences were submitted to the DNA Data Bank of Japan (DDBJ). Neighbor-joining trees were constructed with Kimura's 2-parameter method for a phylogenetic analysis of *ULBP4/RAET1E* sequences spanning from exons 2 to 3 including intron 2 by using the Genetyx software. Bootstrap values were based on 5,000 replications. The *ULBP4/RAET1E* sequences from humans (GenBank accession number AY252119), chimpanzees (AY032638), and rhesus (NC007861) were included in the analysis as references.

Structure model analysis

A three-dimensional (3-D) structure model of rhesus ULBP4/RAET1E, with amino acid positions from 1 to 178, was created by a molecular visualization software RasTop2.2 (http://sourceforge.net/projects/rastop/), and the

human RAET1B in complex with NKG2D (Radaev et al. 2001) from the Molecular Modeling Database (MMCB No. 18231) was used as the reference. Polymorphic sites were mapped on the 3-D structure model of macaque RAET1E by using the Cn3D 4.1 program (http://www.ncbi.nlm.nih.gov/Structure/CN3D/cn3d.shtml).

Table 1 Identified alleles of the ULBP4 gene in rhesus and cynomolgus

Species	Allele name	Accession no.	Reference animal	Identical sequence
Rhesus macaque	*Mamu-ULBP4*1.1*	AB568525	R228, R367	
	*Mamu-ULBP4*1.2*	AB568533	R492, R396, R465	
	*Mamu-ULBP4*2*	AB568526	R283, R384, R328, R337	
	*Mamu-ULBP4*3*	AB568527	R346, R361, R396, R379, R408	
	*Mmau-ULBP4*4*	AB568528	R320, R490, R321, R465, R367, R446, R328, R234, R237, R314	
	*Mamu-ULBP4*5*	AB568529	R430, R453, R325, R477, R439, R360, R379, R446, R355	
	*Mamu-ULBP4*6*	AB568530	R437, R350,	
	*Mamu-ULBP4*7.1*	AB568531	R325, R384, R491, R333, R337	
	*Mamu-ULBP4*7.2*	AB568544	R477	
	*Mamu-ULBP4*8*	AB568532	R408, R454, R241, R342, R316	
	*Mamu-ULBP4*9.1*	AB568534	R312, R314	
	*Mamu-ULBP4*9.2*	AB568535	R333	
	*Mamu-ULBP4*10*	AB568536	R316	
	*Mamu-ULBP4*11*	AB568537	R241	
	*Mamu-ULBP4*12*	AB568538	R342	
	*Mamu-ULBP4*13*	AB568539	R491	
	*Mamu-ULBP4*14*	AB568540	R495	*Mafa-ULBP4*1.1*
	*Mamu-ULBP4*15*	AB568541	R350	
	*Mamu-ULBP4*16*	AB568542	R492	
	*Mamu-ULBP4*17*	AB568543	R495	
	*Mamu-ULBP4*18*	AB568545	R454	
	*Mamu-ULBP4*19*	AB568546	R321	
	*Mamu-ULBP4*20*	AB568547	R355	
	*Mamu-ULBP4*21*	AB571025	R437	
	*Mamu-ULBP4*22*	AB571026	R439	
Crab-eating macaque	*Mafa-ULBP4*1.1*	AB578934	M01, P01, P02, C001, C003, C004, C005, C006	*Mamu-ULBP4*14*
	*Mafa-ULBP4*1.2*	AB578935	M02, C004	
	*Mafa-ULBP4*2*	AB578936	P04, M06, C010, C011, C013	
	*Mafa-ULBP4*3*	AB578938	M03, C007	
	*Mafa-ULBP4*4*	AB578939	M03, C006	
	*Mafa-ULBP4*5*	AB578940	P04, P05, M05, M06, C012, C013	
	*Mafa-ULBP4*6*	AB578941	M05, C010, C011	
	*Mafa-ULBP4*7.1*	AB578942	M01, C002	
	*Mafa-ULBP4*7.2*	AB578943	P03, C008	
	*Mafa-ULBP4*8*	AB578944	P03, M04, C008, C009	
	*Mafa-ULBP4*9*	AB578945	P01, C001, C002	
	*Mafa-ULBP4*10*	AB578946	M04, C009	
	*Mafa-ULBP4*11*	AB578947	P02, C007	
	*Mafa-ULBP4*12*	AB578948	M02, C005	

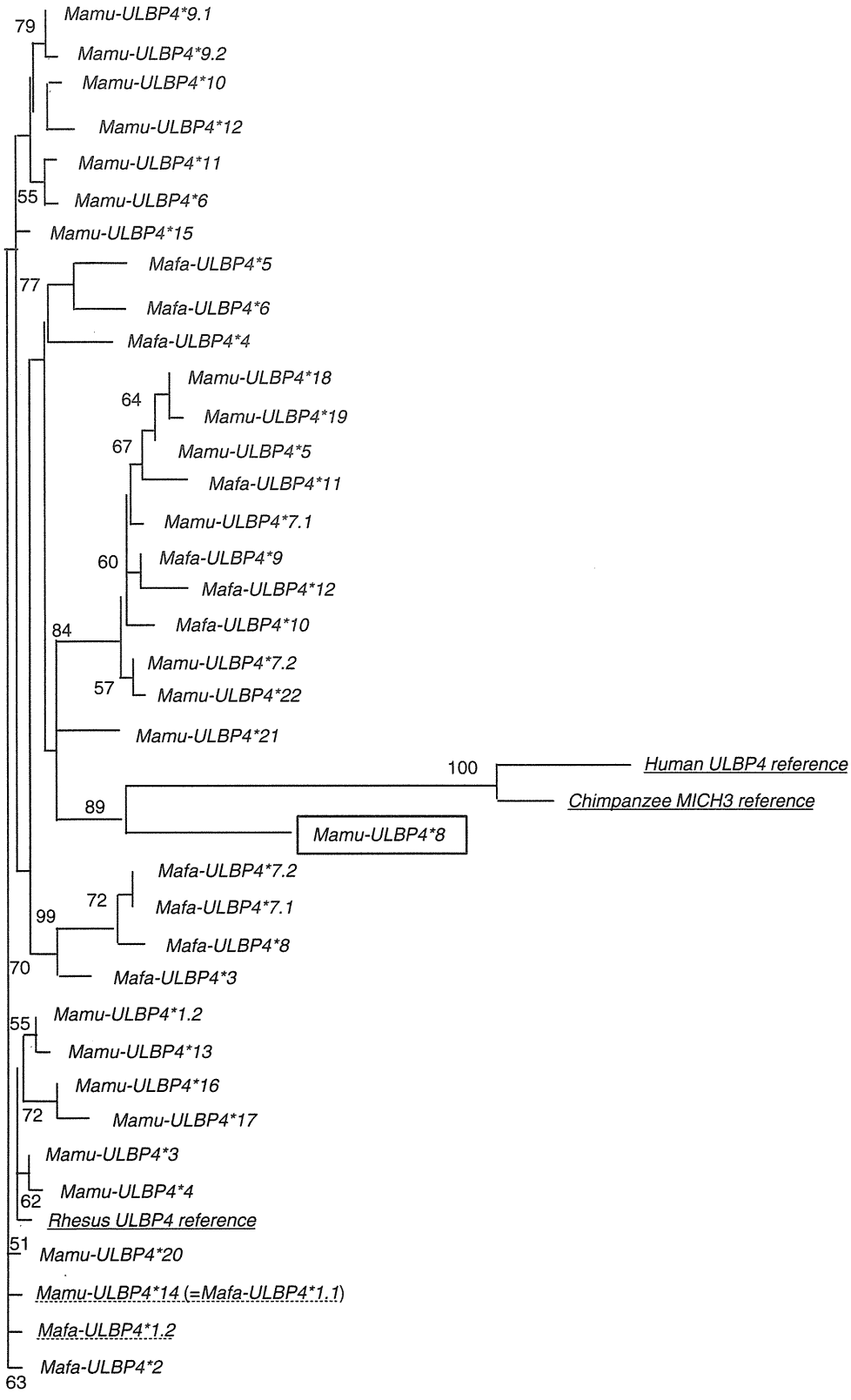

Fig. 1 Phylogenetic tree of *Mamu-* and *Mafa-ULBP4/RAET1E* alleles. A phylogenetic tree of *ULBP4/RAET1E* sequences spanning from exons 2 to 3, obtained in this study, was constructed by using the neighbor-joining method with bootstrap values of 5,000 replications. Values are indicated as percentages, and only those with more than 50% are shown. Sequences of human *ULBP4/RAET1E* (AY252119), chimpanzee *MICH3* (AY032638), and rhesus *ULBP4/RAET1E* (NC007861) were *underlined* and included in the analysis as reference sequences. Alleles represented with *broken underlines* had identical amino acid sequences predicted from the nucleotide sequences. The allele containing an in-frame termination codon was *boxed*

Results

ULBP4/RAET1E polymorphisms in the rhesus macaque

In the rhesus macaque genome (Gibbs et al. 2007), there are two paralogous genes for *ULBP4/RAET1E*, one of which appears to be functional, whereas the other is a pseudogene because it contains a large deletion containing the most part of exons 2, 3, and 4. Therefore, we designed primer pairs to amplify the region containing exons 2 and 3, which encode for α1 and α2 domains of ULBP4/RAET1E molecule, respectively, from the functional *ULBP4/RAET1E*. By using the primer pair, we obtained *ULBP4/RAET1E* sequences from 38 individuals of rhesus macaque. Because one or two sequences were obtained from each individual, the sequences were considered to be alleles of *ULBP4/RAET1E*. They were classified into 25 different alleles, designated as *Mamu-ULBP4*1.1* to *Mamu-ULBP4*22*, submitted to DDBJ, and given accession numbers (Table 1). The allele names with different numbers indicate that they are different in predicted amino acid sequences, whereas the alleles with the same deduced amino acid sequences but different nucleotide sequences, such as *Mamu-ULBP4*1.1* and *Mamu-ULBP4*1.2*, are designated as subtypes. None of the sequences obtained in this study was identical to the reference sequence, NC007861, which was previously deposited to the GenBank database as the rhesus *ULBP4/RAET1E*. On the other hand, when the sequences were aligned referring the human *ULBP4/RAET1E*, one rhesus allele (*Mamu-ULBP4*8*) was found to contain a nonsense mutation at codon 29, which would make the ULBP4/RAET1E molecule non-functional.

ULBP4/RAET1E polymorphisms in the crab-eating macaque

By using the primer pair designed for the rhesus *ULBP4/RAET1E*, we could amplify the *ULBP4/RAET1E* sequences from 24 individuals of the crab-eating macaque. Sequencing analysis revealed 14 different *ULBP4/RAET1E* alleles, and inheritance of each allele was confirmed by family studies. The identified allele sequences were submitted to DDBJ, given accession numbers, and designated as *Mafa-ULBP4*1.1* to *Mafa-ULBP4*12* (Table 1). The nucleotide sequences from exons 2 to 3 of *Mamu-ULBP4*14* were identical to those of *Mafa-ULBP4*1.1* and differed by only one nucleotide in intron 2 from those of *Mafa-ULBP4*1.2*. In addition, a neighbor-joining analysis performed by using nucleotide sequences spanning from exons 2 to 3 showed that the alleles of rhesus and crab-eating macaques were not separately clustered from each other (Fig. 1).

Comparative analysis of *ULBP4/RAET1E*

The alignment of *ULBP4/RAET1E* sequences from rhesus and crab-eating macaques with those from humans and chimpanzees showed that the macaque genes were homologous to the human gene by more than 90% and were equally diverged (Fig. 2). In addition, rhesus and crab-

```
                                                                                50
                  Human :GHSLCFNFTI KSLSRPGQPW CEAQVFLNKN LFLQYNSDNN MVKPLGLLGK KVYATSTwge
              Chimpanzee :---------- -L-------- ---------- ---------- ---------- --N-------
          Rhesus macaque :A--------- --W------- ------M-K- -----D—S- ---------- --N-----Q-
      Crab-eating macaque :A--------- --W------- ------M--- -F---D—SN ---------- --N-----Q-

                                                                    100
                  Human :ltqtlgevgr dlrmllcdIK PQIKTSDPST LQVEMFCQRE AERCTGASWQ FATNGEKSLL
              Chimpanzee :---M------ ------L--- ------G--- ---------- ---------- --I-------
          Rhesus macaque :---K------ ------L-V- ------GP-- -----L---- ---------- -TI----C--
      Crab-eating macaque :---------- -LM---L-V- P-----G--- L----L---- ---------- -TI----C--

                                                          150
                  Human :FDAMNMTWTV INHEAskike twkkdrglek yfrklskgdc dhwlreflgh weampEPTV
              Chimpanzee :---------- ---------- ---------- ---------- ---------- --------Q
          Rhesus macaque :F----M---- ----AS---- ---K------ ------M--- N------L-- Q-A-----Q
      Crab-eating macaque :---------- -----S---- ---------- ------M--- N-WLR----- Q------TQ
```

Fig. 2 Alignment of deduced amino acid sequences of α1 and α2 domains of ULBP4/RAET1E. Amino acid sequences were deduced from the nucleotide sequences of *ULBP4/RAET1E* or *MICH3* from humans (AY252119), chimpanzees (AY032638), rhesus macaques (NC007861), and crab-eating macaques (AY032639). *Numbers* above the sequences represent the amino acid positions in mature protein. *Dashes* indicate identical sequences. Sequences for the predicted α helix structure were indicated by *small italicized characters*. Positions of polymorphic sites in the *human*, *rhesus macaque*, and *crab-eating macaque* were *underlined*

Table 2 Single nucleotide polymorphisms of ULBP4 gene among human and Old World monkeys

	Number of alleles	Exon 2		Intron 2	Exon 3	
		Polymorphism	Non-synonymous change (%)	Polymorphism	Polymorphism	Non-synonymous change (%)
Human	5	2	2 (100%)	3	3	3 (100%)
Rhesus macaque	25	9	5 (55.6%)	22	22	14 (63.6%)
Crab-eating macaque	14	17	9 (52.9%)	18	16	9 (56.3%)

eating macaques showed a higher degree of polymorphism in the analyzed region, namely, exon 2, intron 2, and exon 3, than in humans (Table 2). All polymorphisms found in exons of human *ULBP4/RAET1E* were non-synonymous, whereas a considerable part of the polymorphisms were synonymous in the Old World monkeys. On the other hand, the polymorphic sites in the rhesus macaque (positions 29, 46, 59, 64, 79, 88, 112, 121, 126, 135, 136, 144, 157, 158, 161, 168, 171, and 173) and the crab-eating macaque (positions 32, 39, 40, 59, 72, 73, 79, 91, 112, 136, 163, 164, 165, 171, 178, and 179) were shared at five positions (59, 79, 112, 136, and 171) by each other, whereas only one position (position 112) was shared with polymorphic sites in humans (positions 53, 99, 112, and 113) (Fig. 2). In addition, a termination at position 29 was found in a rhesus macaque allele *Mamu-ULBP4*8*; a single amino acid deletion caused by deletions of a total of three nucleotides was found in a crab-eating macaque allele *Mafa-ULBP4*6* [i.e., TGGCTCAGG sequences corresponding to codons 163–165 were changed to TGCTCA, which may be due to two different deletions at codons 163 (from TGG to TG) and 165 (from AGG to A)], whereas such polymorphisms were not observed in humans. These findings suggest that a selection pressure to generate and maintain the polymorphic sites might be considerably different between the lineages of humans and the Old World monkeys.

Discussion

It has been suggested that the ancestral gene for the ULBP/REAT molecule of placental mammals was originally diverged and duplicated in each species after an emigration from the MHC region (Kondo et al. 2010). In humans, MHC genes (*HLA* genes) are clustered and mapped on the short arm of chromosome 6, 6p21.3, whereas the *ULBP/RAET1* genes are located on the long arm of chromosome 6, 6q25.1. As for the *MHC* genes in the macaque, it was previously reported that rhesus macaque MHC, e.g., BAT1 gene, was localized to chromosome 6q24 by using fiber-fluorescence in situ hybridization (Huber et al. 2003) and cynomolgus (crab-eating) macaque MHC, e.g., Mafa-A and Mafa-B genes, was cytogenetically mapped to chromosome 6p13 (Liu et al. 2007), although the rhesus macaque MHC is mapped on the short arm of chromosome 4 in the draft genome sequence database of rhesus macaques (Gibbs et al. 2007); e.g., Mamu-A and BAT1 were mapped from positions 29, 517, 308 to 29, 520, 221 and from 31, 164, 822 to 31, 175, 032, respectively, on chromosome 4 (data were obtained from the UCSC Genome Browser at http://genome.ucsc.edu/cgi-bin/hgGateway). The discrepancy between the cytogenetic mapping and the assignment in draft genome sequence should be resolved in the future. On the other hand, it is interesting to note that each member of the *ULBP/RAET1* gene family, except for *ULBP6*, is completely or partially duplicated in the rhesus genome. As for the *ULBP4/RAET1E*, two related sequences, LOC695031 (NC007861) and LOC694265, have been identified as orthologs of human *ULBP4/RAET1E*. On the other hand, the configuration of *ULBP/RAET1* loci in the crab-eating macaque genome remained unknown. Because LOC694265 was a pseudogene lacking most part of the coding exons, we designed PCR primers by referring the NC007861 sequence. By using the designed primers, we could successfully amplify *ULBP4/RAET1E* alleles from both rhesus and crab-eating macaques.

In this study, we identified a total of 25 and 14 alleles from rhesus and crab-eating macaques, respectively. One of the rhesus macaque alleles had identical sequences to one of the crab-eating macaque alleles, and the phylogenetic analysis demonstrated that the *ULBP4/RAET1E* alleles were widely diverged. None of the alleles identified in this study were identical to the previously reported sequence NC007861, which was derived from an individual of Indian rhesus macaque. Given that we analyzed rhesus macaques of Burmese origin in this study, and allele distribution of MHC-related polymorphic genes are well known to be largely dependent on the habitat regions, the extent of diversity and variation in *ULBP4/RAET1E* may be further expanded.

It was demonstrated that the diversity of *ULBP4/RAET1E* in the Old World monkeys was much higher than that of human *ULBP4/RAET1E*. It is possible that the genes in the *ULBP/RAET1* locus, in particular, *ULBP4/RAET1E* and *ULBP/RAET1*s, might be highly polymorphic in the

Old World monkeys. We therefore investigated ten unrelated rhesus macaque subjects, in which we had detected 16 *ULBP4/RAET1E* alleles for polymorphisms in the adjacent *ULBP/RAET1* genes. We found one *ULBP1/RAET1I* allele, seven *ULBP2/RAET1H* alleles, and one *ULBP3/RAET1N* allele in these subjects. The observation suggested that *ULBP4/RAET1E* was highly polymorphic as compared to the adjacent *ULBP/RAET1* genes.

We revealed a high degree of polymorphism in the *ULBP4/RAET1E* of the rhesus and crab-eating macaques, although about half of the polymorphisms were synonymous changes (Table 2). Albeit the expression of the ULBP4/RAET1E molecule is known to be involved in the recognition of tumor cells by the NKG2D receptor (Cao et al. 2008; Kong et al. 2009), the functional significance of the polymorphisms in the extracellular domain of the ULBP4/RAET1E molecules remained unknown. To investigate a possible role of the polymorphisms, we have created a 3-D structure model of rhesus ULBP4/RAET1E molecule by using the structure data of human ULBP3/RAET1N in complex with NKG2D (Radaev et al. 2001) as the reference. As shown in Fig. 3, only one polymorphic site at 173 was on the surface of the α helix pointing to the NKG2D receptor, five sites at 59, 136, 144, 161, and 165 were positioned outside the α helix, and only two sites at 32 and 91 were mapped on the β sheet in the groove. The other polymorphic sites were on the β sheet outside of the groove or were not on the surface of the α helix. In addition, expression of *ULBP4/RAET1E* is predominantly found in the skin and tumor tissues and not induced by viral infection in normal cells (Chalupny et al. 2003; Eagle et al. 2006). These observations suggest that the polymorphisms are unlikely to be involved in the differential presentation of characteristic small molecules bound by the ULBP4/RAET1E molecules, as found in the presentation of antigenic peptides by the MHC molecules. Nevertheless, highly prevalent polymorphisms leading to amino acid replacements suggest that a selection pressure had operated on the configuration of diversity in *ULBP4/RAET1E*.

Of particular interest in this study was the rhesus macaque allele *Mamu-ULBP4*8*, which was supposed to contain a stop codon in the exon 2 coding sequence that would truncate the most part of the molecule. This is the first report of a non-functional *ULBP/RAET1* allele in primates; however, a similar situation was reported for another NKG2D ligand gene, *MIC*. For example, a specific human *MIC* haplotype linked to HLA-B*048 consists of non-functional *MIC* genes, in which *MICA* was deleted and *MICB* contained a termination codon (Ota et al. 2000); the non-functional *MIC* haplotype is widely distributed in the East Asian populations (Komatsu-Wakui et al. 2001). It is interesting to note that there are two distinct and polymorphic genes for *MIC* in the rhesus macaque, *MICA* (previously designated as *MIC1* and *MIC3*) and *MICB* (previous *MIC2*); however, they are not considered to be orthologous to the human *MICA* and *MICB* genes, respectively (Seo et al. 1999, 2001; Doxiadis et al. 2007; Averdam et al. 2007). Because members of the MIC and ULBP/RAET1 molecules are structurally related (Li et al. 2002), there is a functional redundancy in the recognition by NKG2D, and thus, the presence of a null allele had been allowed during the evolution of primates.

In the present study, we demonstrated the *ULBP4/RAET1E* allelic polymorphisms not only in the rhesus macaque but also in the crab-eating macaque. Although the localization of *ULBP4/RAET1E* in the crab-eating macaque

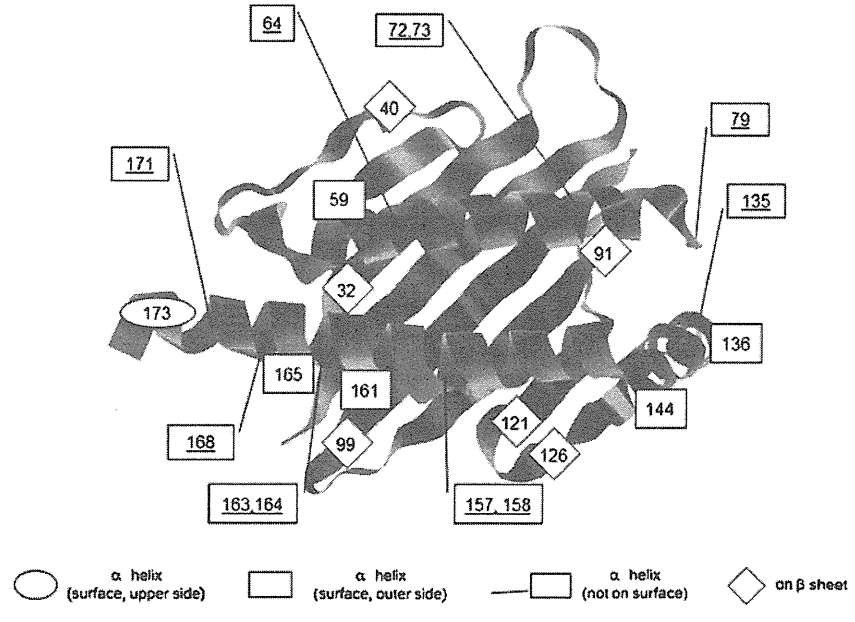

Fig. 3 Mapping of polymorphic sites on the structure model of the macaque ULBP4/RAET1E molecule. Polymorphic sites found in the Old World monkeys were mapped on the 3-D structure model of ULBP4/RAET1E. Residues on the upper and outer sides of the α helix structure were indicated by a *circle* and *squares*, respectively. Residues not found on the surface of the α helix were *underlined*, and those on the β sheet structure were represented by *rhombi*

genome is unknown, a homology search showed that a *Mafa-MICH3* gene (AY032639) was homologous to *Mafa-ULBP4/RAET1E* because the nucleotide sequences of *Mafa-ULBP4*1.1* showed a 96% homology to *Mafa-MICH3*. Similarly, nucleotide sequences of a chimpanzee gene, *Patr-MICH3* (AY032638), showed a 94% homology to the rhesus *ULBP4/RAET1E*. These findings strongly suggest that *MICH3* in the crab-eating macaque and chimpanzee is orthologous to *ULBP4/RAET1E* in the human and rhesus macaque.

In conclusion, we revealed a large diversity of *ULBP4/RAET1E* in two related species of the Old World monkey. Because there were extremely large polymorphisms in the extracellular domain of the ULBP4/RAET1E molecule in the Old World monkey, which was larger than that in the human, the functional impact of the polymorphisms and its significance in the evolution of primates should be investigated in future studies.

Acknowledgments We thank Ms. Yukiko Ueda for her technical assistance. This work was supported, in part, by research grants from the Ministry of Health, Labor and Welfare, Japan; the Japan Health Science Foundation; the program of Founding Research Centers for Emerging and Reemerging Infection Disease; the program of Research on Publicly Essential Drugs and Medical Devices; grants-in-aid for scientific research from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT), Japan; a support for women researchers from the Tokyo Medical and Dental University; and a grant from the Life Science Institute Foundation.

References

Antoun A, Jobson S, Cook M, O'Callaghan CA, Moss P, Briggs DC (2010) Single nucleotide polymorphism analysis of the NKG2D ligand cluster on the long arm of chromosome 6: extensive polymorphisms and evidence of diversity between human populations. Hum Immunol 71:610–620

Averdam A, Seelke S, Grützner I, Ronser C, Roos C, Westphal N, Stahl-Hennig C, Muppala V, Schrod A, Sauermann U, Dressel R, Walter L (2007) Genotyping and segregation analyses indicate the presence of only two functional MIC genes in rhesus macaques. Immunogenetics 59:247–251

Bacon L, Eagle RA, Meyer M, Easom N, Young NT, Trowsdale J (2004) Two human ULBP/RAET1 molecules with transmembrane region are ligands for NKG2D. J Immunol 173:1078–1084

Bauer S, Groh V, Wu J, Steinle A, Phillips JH, Lanier LL, Spies T (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. Science 285:727–729

Cao W, Xi X, Wang Z, Dong L, Hao Z, Cui L, Ma C, He W (2008) Four novel ULBP splice variants are ligands for human NKG2D. Int Immunol 20:981–991

Chalupny NJ, Sutherland CL, Lawrence WA, Rein-Weston A, Cosman D (2003) ULBP4 is a novel ligand for human NKG2D. Biochem Biophys Res Commun 305:129–135

Cosman D, Müllberg J, Sutherland CL, Chin W, Armitage R, Fanslow R, Kubin M, Chalupny NJ (2001) ULBPs, novel MHC class I–related molecules, bind to CMV glycoprotein UL16 and stimulate NK cytotoxicity through the NKG2D receptor. Immunity 14:123–133

Doxiadis GGM, Heijmans CM, Otting N, Bontrop RE (2007) MIC gene polymorphism and haplotype diversity in rhesus macaques. Tissue Antigens 69:212–219

Eagle RA, Traherne JA, Ashiru O, Wills MR, Trowsdale J (2006) Regulation of NKG2D ligand gene expression. Hum Immunol 67:1159–1169

Eagle RA, Flack G, Warford A, Martinez-Borra J, Jafferji I, Traherne JA, Ohashi M, Boyle LH, Barrow AD, Caillat-Zucman S, Young NT, Trowsdale J (2009a) Celllar expression, trafficking, and function of two isoforms of human ULBP5/RAET1G. Pros ONE 4:e4503

Eagle RA, Traherne JA, Hair JR, Jafferji I, Trowsdale J (2009b) ULBP6/RAET1L is an additional human NKG2D ligand. Eur J Immunol 39:3207–3216

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

Huber I, Walter L, Wimmer R, Pasantes JJ, Günther E, Schempp W (2003) Cytogenetic mapping and orientation of the rhesus macaque MHC. Cytogenet Genome Res 103:144–1449

Komatsu-Wakui M, Tokunaga K, Ishikawa Y, Leelayuwat C, Kashiwase K, Tanaka H, Moriyama S, Nakajima F, Park MH, Jia GJ, Chimge NO, Sideltseva EW, Juji T (2001) Wide distribution of the MICA-MICB null haplotype in East Asians. Tissue Antigens 57:1–8

Kondo M, Maruoka T, Otsuka N, Kasamatsu J, Fugo K, Hanzawa N, Kasahara M (2010) Comparative genomic analysis of mammalian NKG2D ligand family genes provides insights into their origin and evolution. Immunogenetics 62:441–450

Kong Y, Cao W, Xi X, Ma C, Cui L, He W (2009) The NKG2D ligand ULBP4 binds to TCRgamma9/delta2 and induces cytotoxicity to tumor cells through both TCRgammadelta and NKG2D. Blood 114:310–317

Kulski JK, Anzai T, Shiina T, Inoko H (2004) Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. Mol Biol Evol 21:2079–2091

Li P, McDermott G, Strong RK (2002) Crystal structures of RAE-1β and its complex with the activating immunoreceptor NKG2D. Immunity 16:77–86

Liu QY, Wang XX, Zhang JZ, Chen WH, He XW, Lin Y, Wang JF, Zhu Y, Hu SN, Wang XN (2007) Mapping cynomolgus monkey MHC class I district on chromosome 6p13 using pooled cDNAs. Biotech Histochem 82:267–272

Naruse KT, Chen Z, Yanagida R, Yamashita T, Saito Y, Mori K, Akari H, Yasutomi Y, Miyazawa M, Matano T, Kimura A (2010) Diversity of MHC class I genes in Burmese-origin rhesus macaque. Immunogenetics 62:601–611

Ota M, Bahram S, Katsuyama Y, Saito S, Nose Y, Sada M, Ando H, Inoko H (2000) On the MICA deleted-MICB null, HLA-B*4801 haplotype. Tissue Antigens 56:268–271

Otting N, Otting N, deVos-Rouweler AJM, Heijmans CMC, de Groot NG, Doxiadis GGM, Bontrop RE (2007) MHC class I A region diversity and polymorphism in macaque species. Immunogenetics 59:367–375

Pappworth IY, Wang EC, Rowe M (2007) The switch from latent to productive infection in Epstein-Barr virus-infected B cell is associated with sensitization to NK cell killing. J Virol 1:474–482

Pende D, Rivera P, Marcenaro S, Chang CC, Biassoni R, Conte R, Kubin M, Cosman D, Ferrone S, Moretta L, Moretta A (2002) Major histocompatibility complex class I-related chain A and UL16-binding protein expression on tumor cell lines of different histotypes: analysis of tumor susceptibility to NKG2D-dependent natural killer cell cytotoxicity. Cancer Res 62:6178–6186

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