through a HiTrap benzamidine column, and the protein was further purified by gel filtration on a Superdex200 column. The purity and homogeneity of the rVDR–LBD were assessed by SDS-PAGE.

Purified rVDR–LBD solution was concentrated to about 0.75 mg/mL by ultrafiltration. To an aliquot (800 μL) of the protein solution was added a ligand (ca. 10 equiv), then the solution was further concentrated to about 1/8, and a solution (25 mM Tris–HCl, pH 8.0; 50 mM NaCl; 10 mM DTT; 0.02% NaN_3) of coactivator peptide (H_2N-KNHPMLMNLLKDN-$CONH_2$) derived from DRIP205 was added. This solution of VDR/ligand/peptide was allowed to crystallize by the vapor diffusion method using a series of precipitant solutions containing 0.1 M MOPS–NaOH (pH 7.0), 0.1–0.4 M sodium formate, 12–22% (w/v) PEG4000, and 5% (v/v) ethylene glycol. Droplets for crystallization were prepared by mixing 2 μL of complex solution and 1 μL precipitant solution, and droplets were equilibrated against 500 μL precipitant solution at 20 °C.

Prior to diffraction data collection, crystals were soaked in a cryoprotectant solution containing 0.1 M MOPS–NaOH (pH 7.0), 0.1–0.4 M sodium formate, 15–20% PEG4000, and 17–20% ethylene glycol. Diffraction data sets were collected at 100 K in a stream of nitrogen gas at beamline BL-6A of KEK-PF (Tsukuba, Japan). Reflections were recorded with an oscillation range per image of 1.0°. Diffraction data were indexed, integrated, and scaled using the program HKL2000 (HKL Research Inc.). The structures were solved by molecular replacement with the program CNS[25] using rat VDR-LBD coordinates (PDB code: 1RK3), and finalized sets of atomic coordinates were obtained after iterative rounds of model modification and refinement with CNS by simulated annealing, positional minimization, water molecule identification, and restrained individual isotropic B-value refinement.

■ ASSOCIATED CONTENT

Supporting Information. ^{1}H NMR spectra and ^{13}C NMR spectra of synthesized compounds and details of preparation of 3-ethyl-3-triethylsilyloxy-7-bromoheptane, 4-(*p*-toluenesulfonyloxymethyl)-2-phenyl-1,3-dioxane, and 4-(*p*-toluenesulfonyloxymethyl)-2,2-dimethyl-1,3-dioxolane. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author
kage.omc@tmd.ac.jp

■ ACKNOWLEDGMENT

This study was partially supported by Grants-in-Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology, Japan (grant nos 21790110 to S.F., 22790106 to T.H., 19201044 to H.K., 19689004 to A.T.). A.T. thanks the Takeda Science Foundation and the Mitsubishi Foundation for financial support, and T.H. also thanks the Astellas Foundation and the Takeda Science Foundation.

■ REFERENCES

(1) For reviews of carboranes: (a) Bregadze, V. I. *Chem. Rev.* **1992**, *92*, 209. (b) Hawthorne, M. F. *Angew. Chem., Int. Ed.* **1993**, *32*, 950. (c) Soloway, A. H.; Tjarks, W.; Barnum, J. G. *Chem. Rev.* **1998**, *98*, 1515. (d) *The borane, carborane, carbocation continuum*; Casanova, J., Ed.; Wiley Interscience: New York, 1998.

(2) (a) Locher, G. L. *Am. J. Roentgenol., Radium Ther. Nucl. Med.* **1936**, *36*, 1. (b) Barth, R. F.; Coderre, J. A.; Vicente, M. G. H.; Blue, T. E. *Clin. Cancer Res.* **2005**, *11*, 3987. (c) Sivaev, I. B.; Bregadze, V. V. *Eur. J. Inorg. Chem.* **2009**, 1433.

(3) Recent examples of development of BNCT agents: (a) Altieri, S.; Balzi, M.; Bortolussi, S.; Bruschi, P.; Ciani, L.; Clerici, A. M.; Faraoni, P.; Ferrari, C.; Gadan, M. A.; Panza, L.; Pietrangeli, D.; Ricciardi, G.; Ristori, S. *J. Med. Chem.* **2009**, *52*, 7829. (b) Bonjoch, J.; Drew, M. G. B.; Gonzlez, A.; Greco, F.; Jawaid, S.; Osborn, H. M. I.; Williams, N. A. O.; Yaqoob, P. *J. Med. Chem.* **2008**, *51*, 6604. (c) Byun, Y.; Thirumamagal, B. T. S.; Yang, W.; Eriksson, S.; Barth, R. F.; Tjarks, W. *J. Med. Chem.* **2006**, *49*, 5513. (d) Dozzo, P.; Koo, M.-S.; Berger, S.; Forte, T. M.; Kahl, S. B. *J. Med. Chem.* **2005**, *48*, 357.

(4) Examples of application of carborane as a hydrophobic structure: (a) Cígler, P.; Kožíšek, M.; Řezáčová, P.; Brynda, J.; Otwinowski, Z.; Pokorná, J.; Plešek, J.; Grüner, B.; Dolečková-Marešová, L.; Máša, M.; Sedláček, J.; Bodem, J.; Kräusslich, H.-G.; Král, V.; Konvalinka, J. *Proc. Natl. Acad. Soc. U.S.A.* **2005**, *102*, 15394. (b) Beer, M. L.; Lemon, J.; Valliant, J. F. *J. Med. Chem.* **2010**, *53*, 8012. (c) Meggers, E. *Angew. Chem., Int. Ed.* **2011**, *50*, 2442.

(5) (a) Evans, R. M. *Science* **1988**, *240*, 889. (b) Mangelsdorf, D. J.; Thummel, C.; Beato, M.; Herrlich, P.; Schütz, G.; Umesono, K.; Blumberg, B.; Kastner, P.; Mark, M.; Chambon, P.; Evans, R. M. *Cell* **1995**, *83*, 835.

(6) (a) Fujii, S.; Yamada, A.; Tomita, K.; Nagano, M.; Goto, T.; Ohta, K.; Harayama, T.; Endo, Y.; Kagechika, H. *Med. Chem. Commun.* **2011**, *2*, 877. (b) Goto, T.; Ohta, K.; Fujii, S.; Ohta, S.; Endo, Y. *J. Med. Chem.* **2010**, *53*, 4917. (c) Fujii, S.; Goto, T.; Ohta, K.; Hashimoto, Y.; Suzuki, T.; Ohta, S.; Endo, Y. *J. Med. Chem.* **2005**, *48*, 4654. (d) Fujii, S.; Hashimoto, Y.; Suzuki, T.; Ohta, S.; Endo, Y. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 227.

(7) (a) Endo, Y.; Iijima, T.; Yamakoshi, Y.; Yamaguchi, M.; Fukasawa, H.; Shudo, K. *J. Med. Chem.* **1999**, *42*, 1501. (b) Endo, Y.; Iijima, T.; Yamakoshi, Y.; Fukasawa, H.; Miyaura, C.; Inada, M.; Kubo, A.; Itai, A. *Chem. Biol.* **2001**, *8*, 341. (c) Endo, Y.; Yoshimi, T.; Ohta, K.; Suzuki, T.; Ohta, S. *J. Med. Chem.* **2005**, *48*, 3941.

(8) (a) Iijima, T.; Endo, Y.; Tsuji, M.; Kawachi, E.; Kagechika, H.; Shudo, K. *Chem. Pharm. Bull.* **1999**, *47*, 398. (b) Ohta, K.; Iijima, T.; Kawachi, E.; Kagechika, H.; Endo, Y. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 5913.

(9) *Vitamin D*, 2nd ed.; Feldman, D., Pike, J. W., Glorieux, F. H., Eds.; Elsevier Academic Press: New York, 2005.

(10) Kubodera, N. Development of OCT and ED-71. In *Vitamin D*, 2nd ed.; Feldman, D., Pike, J. W., Glorieux, F. H., Eds.; Elsevier Academic Press: New York, 2005, pp 1525–1541.

(11) (a) Verstuyf, A.; Verlinden, L.; Van Baelen, H.; Sabbe, K.; d'Hallewyn, C.; De Clercq, P.; Vandewalle, M.; Bouillon, R. *J. Bone Miner. Res.* **1998**, *13*, 549. (b) Plonska-Ocypa, K.; Sicinski., R. R.; Plum, L. A.; Grzywacz, P.; Frelek, J.; Clagett-Dame, M.; DeLuca, H. F. *Bioorg. Med. Chem.* **2009**, *17*, 1747.

(12) Boehm, M. F.; Fitzgerald, P.; Zou, A.; Elgort, M. G.; Bischoff, E. D.; Mere, L.; Mais, D. E.; Bissonnette, R. P.; Heyman, R. A.; Nadzan, A. M.; Reichman, M.; Allegretto, E. A. *Chem. Biol.* **1999**, *6*, 265.

(13) (a) Jordan, V. C. *Cancer* **1992**, *70*, 977. (b) Delmas, P. D.; Bjarnason, N. H.; Mitlak, B. H.; Ravoux, A. C.; Shah, A. S.; Huster, W. J.; Draper, M.; Christiansen, C. *N. Engl. J. Med.* **1997**, *337*, 1641.

(14) Fujii, S.; Kano, A.; Sekine, R.; Kawachi, E.; Masuno, H.; Hirano, T.; Tanatani, A.; Kagechika, H. Abstract of 14th Vitamin D Workshop, Brugge, Belgium, October 4–8, 2009, p 6.

(15) Bouillon, R.; Okamura, W. H.; Norman, A. W. *Endocr. Rev.* **1995**, *16*, 200.

(16) Yamamoto, K.; Endo, Y. *Bioorg. Med. Chem. Lett.* **2001**, *11*, 2389.

(17) (a) Sai, H.; Takatsuto, S.; Hara, N.; Ikekawa, N. *Chem. Pharm. Bull.* **1985**, *33*, 878. (b) Gill, H. S.; Londowski, J. M.; Corradino, R. A.; Zinsmeister, A. R.; Kumar, R. *J. Steroid Biochem.* **1988**, *31*, 147.

(18) (a) Yamada, S.; Yamamoto, K.; Masuno, H.; Ohta, M. J. *Med. Chem.* **1998**, *41*, 1467. (b) Sussman, F.; Rumbo, A.; Villaverde, M. C.; Mouriño., A. *J. Med. Chem.* **2004**, *47*, 1613.

(19) Review of reactions of carboranes: Kaszynski, P. *Collect. Czech. Chem. Commun.* **1999**, *64*, 895.

(20) Mangelsdorf, D. J.; Koeffler, H. P.; Donaldson, C. A.; Pike, J. W.; Haussler, M. R. *J. Cell Biol.* **1984**, *98*, 391.

(21) Vanhooke, J. L.; Benning, M. M.; Bauer, C. B.; Pike, J. W.; DeLuca, H. F. *Biochemistry* **2004**, *43*, 4101.

(22) In addition to reports on specific receptor ligands, an enzyme inhibitor has been reported: Reynolds, R. C.; Campbell, S. R.; Fairchild, R. G.; Kisliuk, R. L.; Micca, P. L.; Queener, S. F.; Riordan, J. M.; Sedwick, W. D.; Waud, W. R.; Leung, A. K. W.; Dixon, R. W.; Suling, W. J.; Borhani, D. W. *J. Med. Chem.* **2007**, *50*, 3283.

(23) Kakuda, S.; Okada, K.; Eguchi, H.; Takenouchi, K.; Hakamata, W.; Kurihara, M.; Takimoto-Kamimura, M. *Acta Crystallogr., Sect. F: Struct. Biol. Cryst. Commun.* **2008**, *64*, 970.

(24) Hakamata, W.; Sato, Y.; Okuda, H.; Honzawa, S.; Saito, N.; Kishimoto, S.; Yamashita, A.; Sugiura, T.; Kittaka, A.; Kurihara, M. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 120.

(25) Brünger, A. T.; Adams, P. D.; Clore, G. M.; DeLano, W. L.; Gros, P.; Grosse-Kunstleve, R. W.; Jiang, J.-S.; Kuszewski, J.; Nilges, M.; Pannu, N. S.; Read, R. J.; Rice, L. M.; Simonson, T.; Warren, G. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **1998**, *54*, 905.

Immunogenetics (2011) 63:417–428
DOI 10.1007/s00251-011-0519-7

ORIGINAL PAPER

Molecular evolution of immunoglobulin superfamily genes in primates

Hitoshi Ohtani · Toshiaki Nakajima · Hirofumi Akari · Takafumi Ishida · Akinori Kimura

Received: 1 September 2010 / Accepted: 17 February 2011 / Published online: 10 March 2011

Abstract Genes of the immunoglobulin superfamily (IgSF) have a wide variety of cellular activities. In this study, we investigated molecular evolution of IgSF genes in primates by comparing orthologous sequences of 249 IgSF genes among human, chimpanzee, orangutan, rhesus macaque, and common marmoset. To evaluate the non-synonymous/synonymous substitution ratio (ω), we applied Bn-Bs program and PAML program. IgSF genes were classified into 11 functional categories based on the Gene Ontology (GO) database. Among them, IgSF genes in three functional categories, immune system process (GO:0002376), defense response (GO:0006952), and multi-organism process (GO:0051704), which are tightly linked to the regulation of immune system had much higher values of ω than genes in the other GO categories. In addition, we estimated the average values of ω for each primate lineage. Although each primate lineage had comparable average values of ω, the human lineage showed the lowest ω value for the immune-related genes. Furthermore, 11 IgSF genes, *SIGLEC5*, *SLAMF6*, *CD33*, *CD3E*, *CEACAM8*, *CD3G*, *FCER1A*, *CD48*, *CD4*, *TIM4*, and *FCGR2A*, were implied to have been under positive selective pressure during the course of primate evolution. Further sequence analyses of *CD3E* and *CD3G* from 23 primate species suggested that the Ig domains of *CD3E* and *CD3G* underwent the positive Darwinian selection.

Keywords Natural selection · Immune system · Immunoglobulin domain · Comparative genomics · CD3 complex

Introduction

Comparative genomics is a promising approach for studying the biological development of the genome from the view point of evolution. Recently, large-scale genome sequences of human, chimpanzee, orangutan, rhesus macaque, and common marmoset have been made available (Consortium CSaA 2005; Gibbs et al. 2007), and the comparative genomic analyses among primates are crucial for addressing the issue of which genetic changes have made us uniquely human. In addition, such analyses are also useful for identifying the susceptibility genes for human diseases and for understanding the pathophysiological mechanisms of the diseases, because the biological differences among primates, such as differences in the disease susceptibility, have been reported (Lyashchenko et al. 2008; Song et al. 2005).

To identify the genes that have come under the pressure of natural selection in the course of primate evolution is of critical importance, because such genes would very likely be linked to biological function involved in the human

Electronic supplementary material The online version of this article (doi:10.1007/s00251-011-0519-7) contains supplementary material, which is available to authorized users.

H. Ohtani · T. Nakajima · A. Kimura (✉)
Department of Molecular Pathogenesis, Medical Research Institute, and Laboratory of Genome Diversity, Graduate School of Biomedical Science, Tokyo Medical and Dental University, 1-5-45 Yushima, Bunkyo-ku, Tokyo 113–8510, Japan
e-mail: akitis@mri.tmd.ac.jp

H. Akari
Center for Human Evolution Modeling Research, Primate Research Institute, Kyoto University, Inuyama, Japan

T. Ishida
Unit of Human Biology and Genetics, Graduate School of Science, The University of Tokyo, Tokyo, Japan

diseases. In fact, comparisons of genomes between the human and chimpanzee and between the human and rhesus macaque have suggested that dozens of genes have emerged under the pressure of natural selection in the course of primate evolution, in particular, those which are involved in the host–pathogen interactions, reproduction, and sensory systems (Clark et al. 2003; Gibbs et al. 2007; Nielsen et al. 2005). These studies have also reported that the immunoglobulin superfamily (IgSF) genes are commonly observed among the genes in primates, which had come under the pressure of natural selection.

Members of the IgSF are defined by the presence of one or more regions homologous to the basic structural unit of immunoglobin (Ig) molecules. The Ig domain possesses a characteristic Ig fold, which is composed of two opposing anti-parallel beta-strands connected by disulfide bonds between cysteine residues (Halaby and Mornon 1998). The IgSF is a large group of cell surface, cytoplasmic, and serum proteins involved in the recognition, binding, and/or adhesion processes of cells (Lander et al. 2001). Members of the IgSF have a wide variety of cellular functions acting as cell surface antigen receptors, co-receptors and co-stimulatory molecules of the immune system, molecules involved in antigen presentation to lymphocytes, cell adhesion molecules, certain cytokine receptors, and intracellular muscle proteins. They are commonly ascribed to a role in molecular–molecular interactions (Barclay 2003; Lander et al. 2001; Otey et al. 2009).

Although the IgSF genes have been reported as showing evidence of positive selection, the phylogenetic analyses focused on the Ig domains of IgSF genes have not been conducted. The purpose of present study is to provide insights into the overview of the molecular evolution of the IgSF genes and to identify the IgSF genes under the positive selection in the course of five primate species.

Materials and methods

Sequence data collection

Selection of the IgSF genes was done by using the Conserved Domain Database v2.22 at the National Center for Biotechnology Information (NCBI) (http://www.ncbi.nlm.nih.gov/sites/entrez). As in the previous studies (Gibbs et al. 2007; Kosiol et al. 2008), we identified orthologous genes for human IgSF genes from chimpanzee, orangutan, rhesus macaque, and common marmoset by using the UCSC/MULTIZ alignment program which is constructed by the synteny-based genome-wide multiple alignments (Blanchette et al. 2004; Kent et al. 2003). Sequence alignment was done by using the Clustal X program (Larkin et al. 2007). IgSF genes were classified based on the Gene Ontology (GO) database (http://www.geneontology.org/) (Ashburner et al. 2000).

Primate genomic DNA samples

DNA samples from 23 primate species including human (*Homo sapiens*), chimpanzee (*Pan troglodytes*), bonobo (*Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), black gibbon (*Hylobates concolor*), white-handed gibbon (*Hylobates lar*), siamang (*Symphalangus syndactylus*), rhesus macaque (*Macaca mulatta*), crab-eating macaque (*Macaca fascicularis*), baboon (*Papio hamadryas*), black and white colobus (*Colobus guereza*), dusky lutong (*Trachypithecus obscurus*), silvered lutong (*Trachypithecus cristatus*), Central American spider monkey (*Ateles geoffroyi*), long-haired spider monkey (*Ateles belzebuth*), tufted capuchin (*Cebus apella*), common squirrel monkey (*Saimiri sciureus*), red-handed tamarin (*Saguinus midas*), cotton-top tamarin (*Saguinus oedipus*), golden lion tamarin (*Leonthopithecus rosalia*), common marmoset (*Callithrix jacchus*), and lesser galago (*Galago senegalensis*) were analyzed for *CD3E* and *CD3G* sequences.

PCR and sequencing analysis of *CD3E* and *CD3G*

Sequence information for coding regions of *CD3E* and *CD3G* were obtained by direct sequencing of gene segments amplified by polymerase chain reaction (PCR) from the genomic DNA samples. Primers for PCR were designed in the highly conserved non-coding regions among the genes from human, chimpanzee, orangutan, rhesus macaque, and common marmoset, referring the genomic sequences deposited in the UCSC Genome Browser (electronic supplementary material (ESM) Table 1). Primers for prosimian were designed by referencing the common marmoset sequences and whole-genome shotgun sequences from prosimians in NCBI BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The primers were used for both PCR and direct sequencing analyses of the genes. When sequence variations (heterozygous sequences) in specific species were detected, the sequences which were conserved among 23 primate species were considered as ancestral sequences and used in the statistical analyses.

PCR was performed in a reaction mixture of 15 μL containing 0.1 μL Taq DNA polymerase (Takara Bio Inc., Shiga, Japan), 1 μL of 50 ng/μL DNA template, 1.5 μL of 2.0 mM dNTPs, 0.5 μL of 10 μM each primer, 1.5 μL reaction buffer containing 20 mM $MgCl_2$, and sterile water. PCR condition was as follows: 94°C for 2 min, 35 cycles (94°C for 30 s, 55–60°C for 30 s, 72°C for 1 min), and 72°C for 5 min. PCR products were then purified and sequenced by the BigDye Terminator cycling system using an ABI3130× automated DNA sequencer (Applied Biosystems,

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).

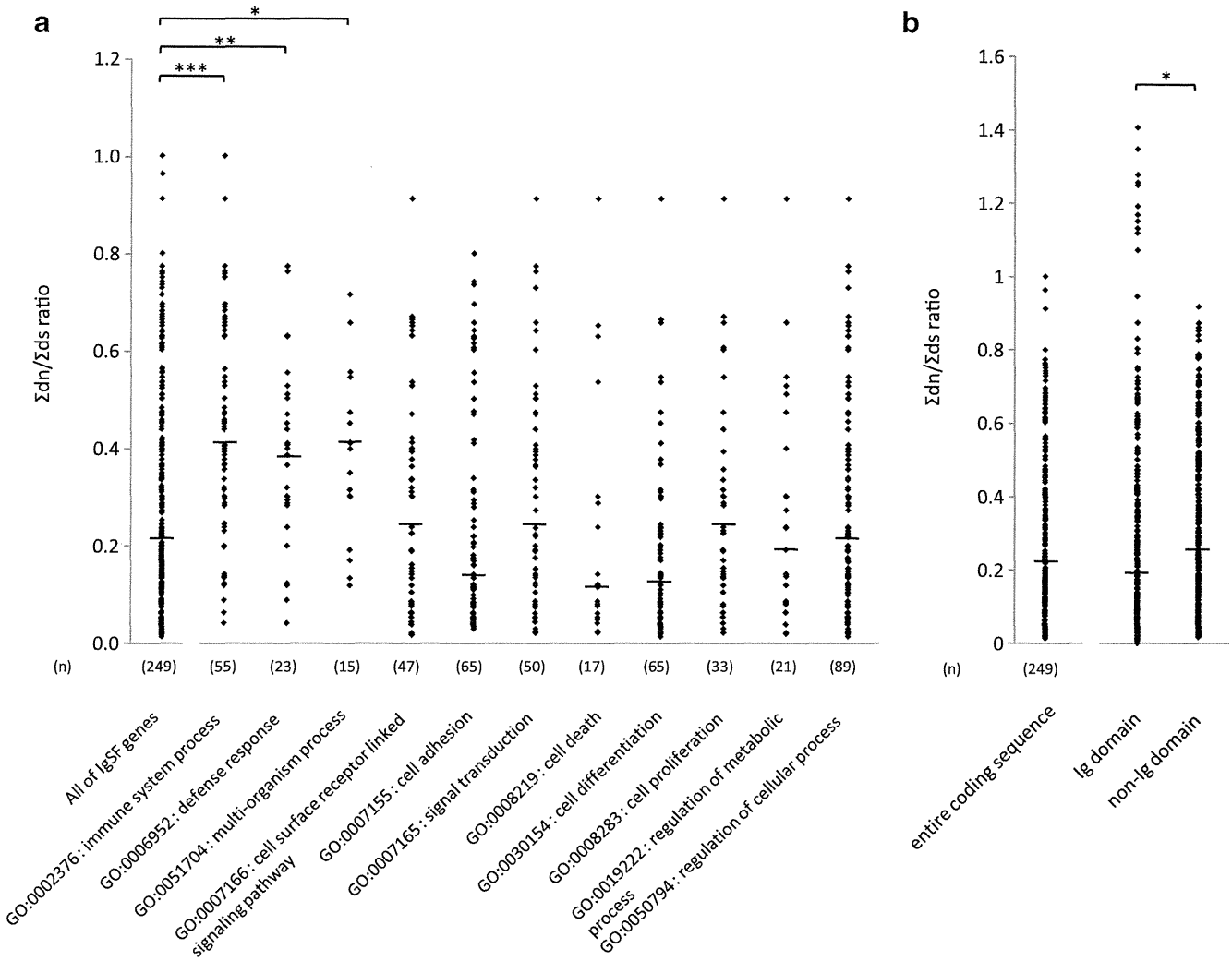

Fig. 1 Σdn/Σds ratio of IgSF genes. **a** The IgSF genes were categorized by gene ontology. **b** Σdn/Σds ratios for the entire coding region, Ig domain, and non-Ig domain. The Σdn/Σds ratios were calculated by Bn-Bs program. *Bars* indicate median values of Σdn/Σds ratio for each group. An *asterisk* 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.

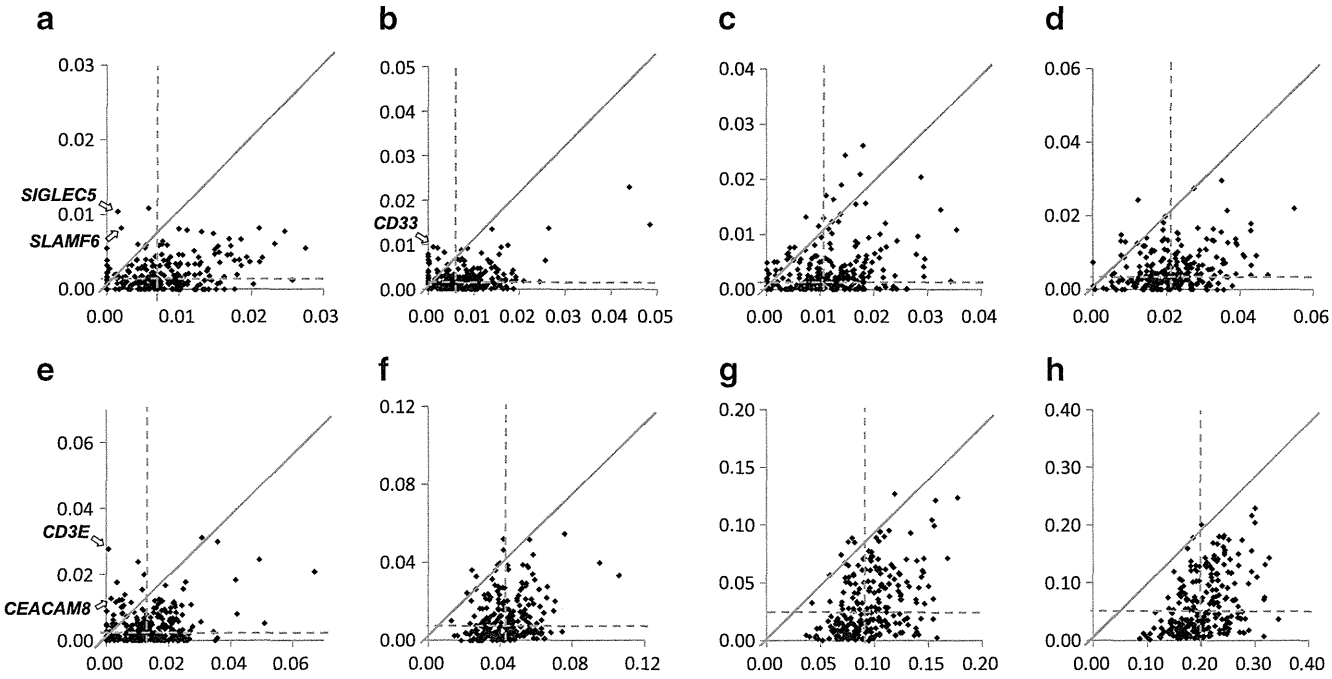

Fig. 2 Pairwise comparison plots of dn and ds values for the entire coding regions of IgSF genes in primate lineages. The values of dn (*vertical axis*) and ds (*horizontal axis*) for each primate lineage and their summation (Σ) were calculated by Bn-Bs program. *Dotted lines* indicate the average values 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
	IL1IRA	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
	IL1IRA	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

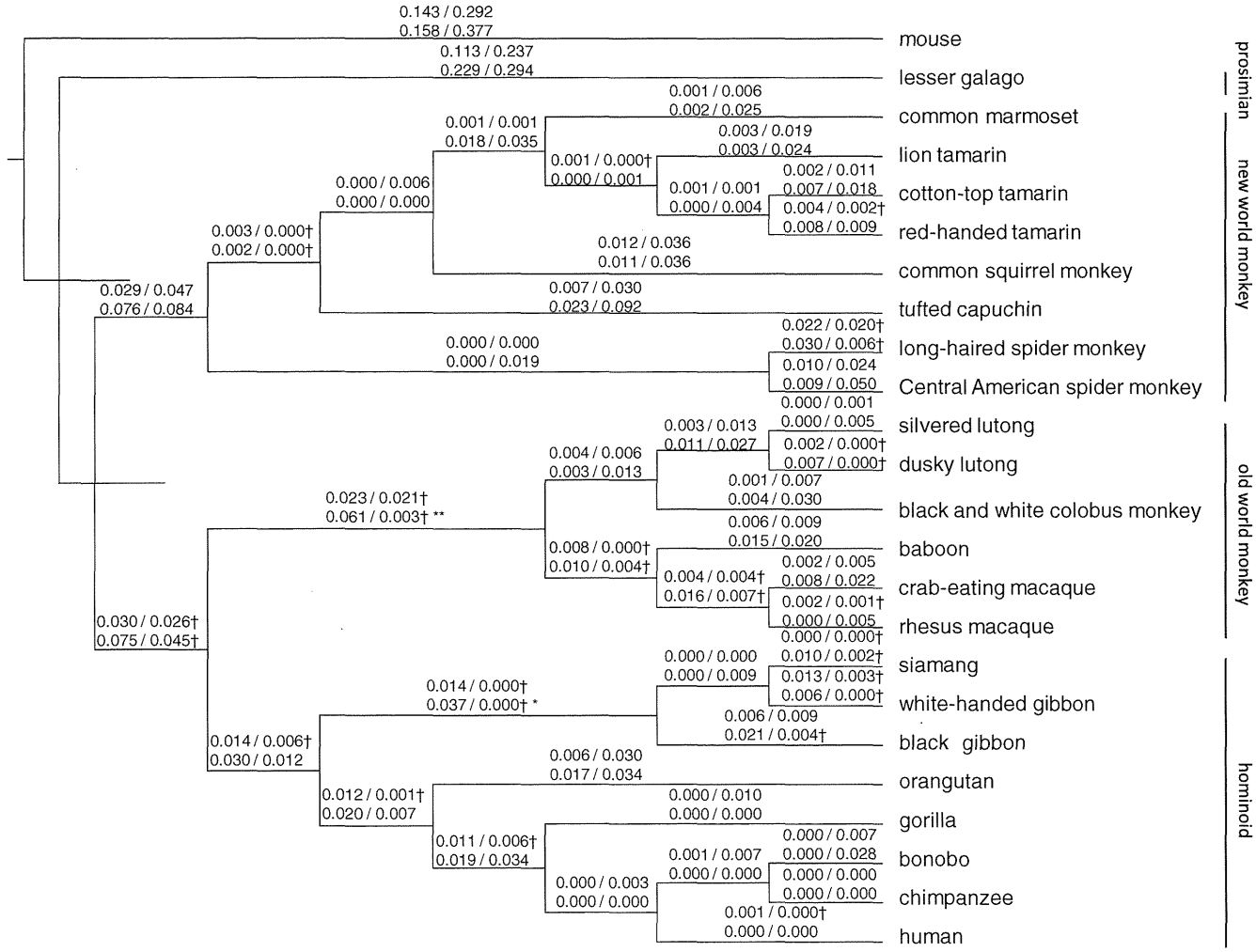

Fig. 3 Phylogenetic trees of *CD3E* in the primate evolution. *Values above branches* indicate estimated values of dn and ds per lineage by using Bn-Bs program. Values indicated in the upper parts are for the entire coding region, while values in the lower parts are for the Ig domain. *Daggers* indicate 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

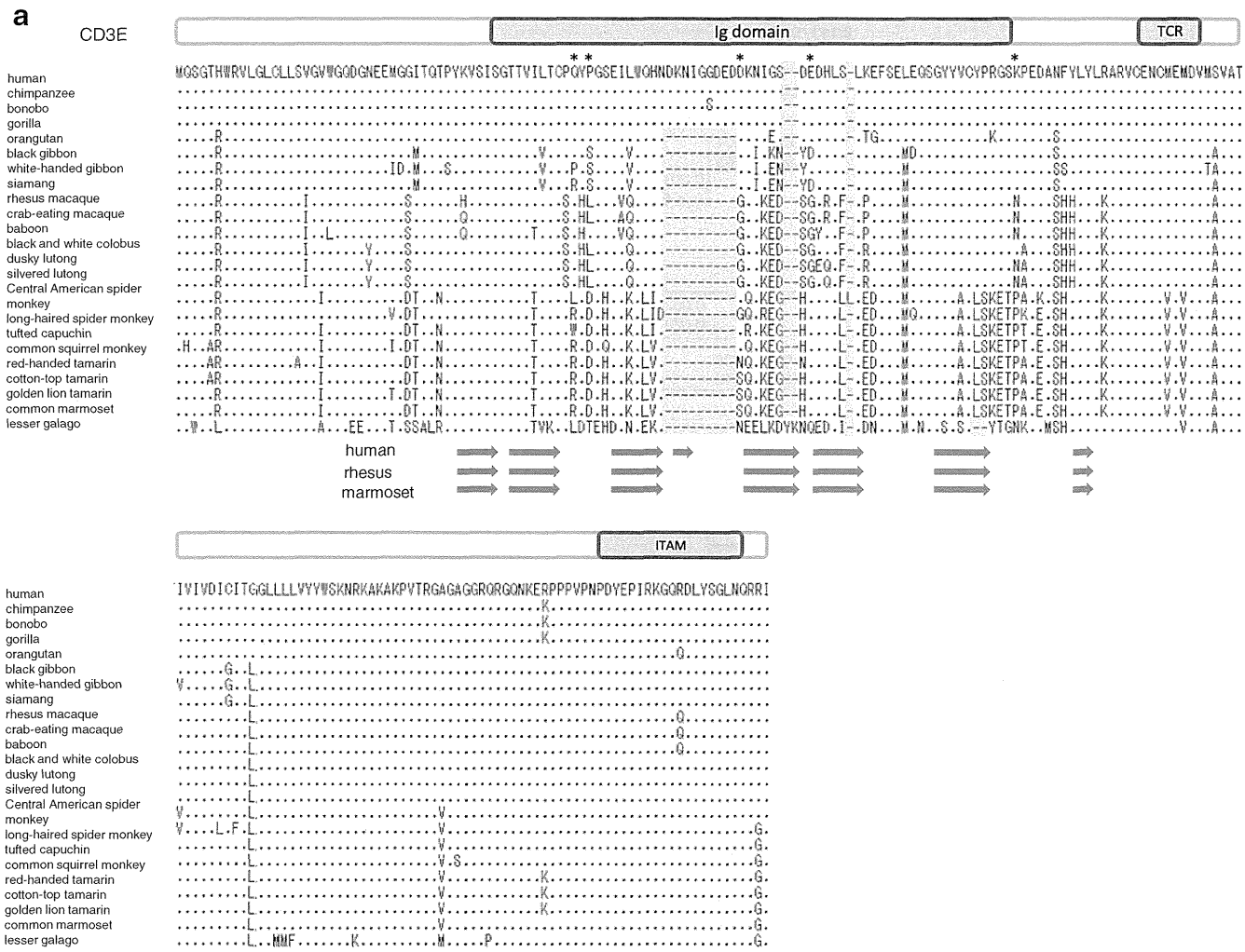

Fig. 4 Alignments of CD3E (**a**) and CD3G (**b**) amino acid sequences from 23 primate species. *Dots* indicate the identities to the human reference sequence, while hyphens indicate alignment gaps. TCR indicates amino acid sites at which the CD3 molecules interact with T cell receptor. ITAM represents 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 ($p<0.05$)

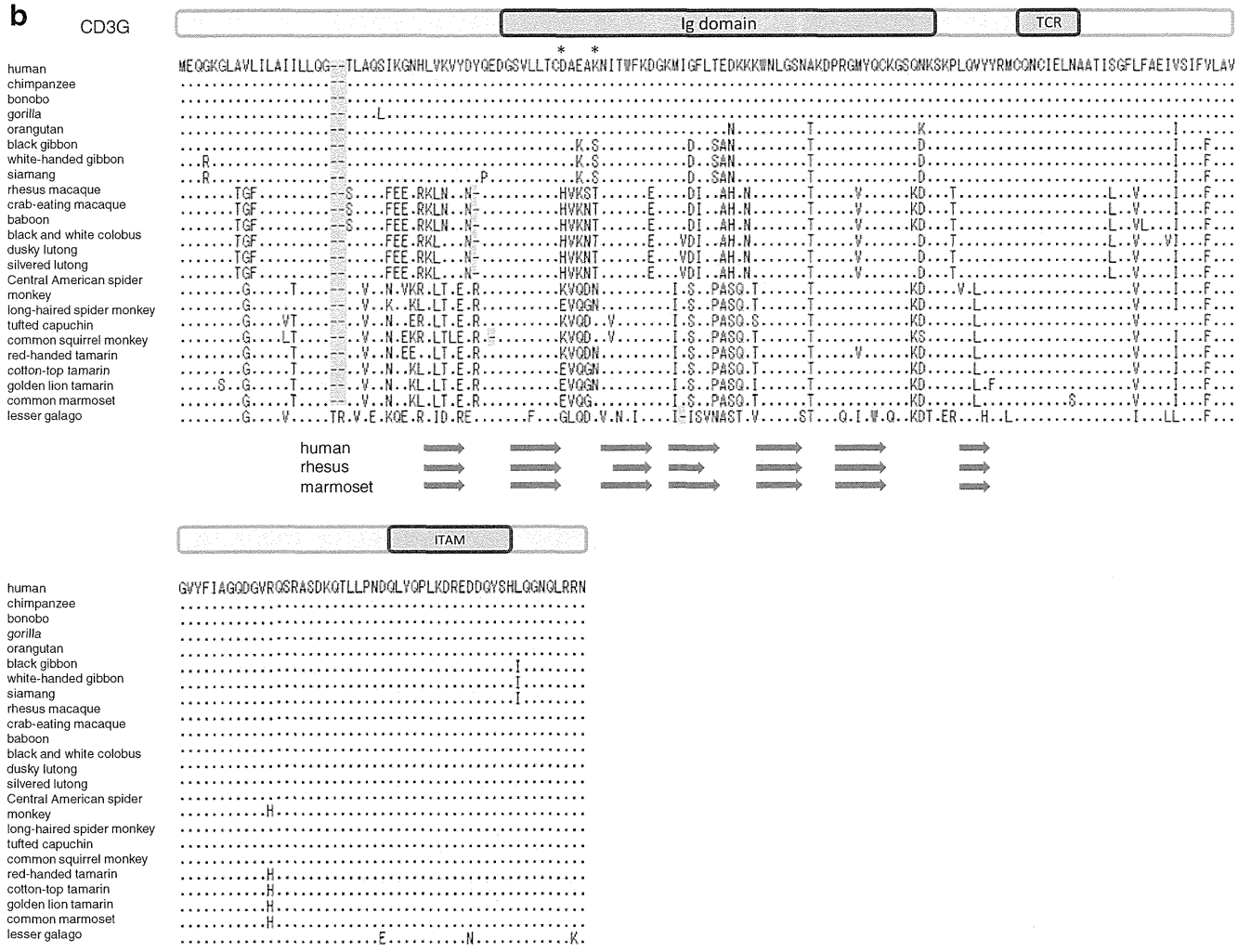

Fig. 4 (continued)

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).

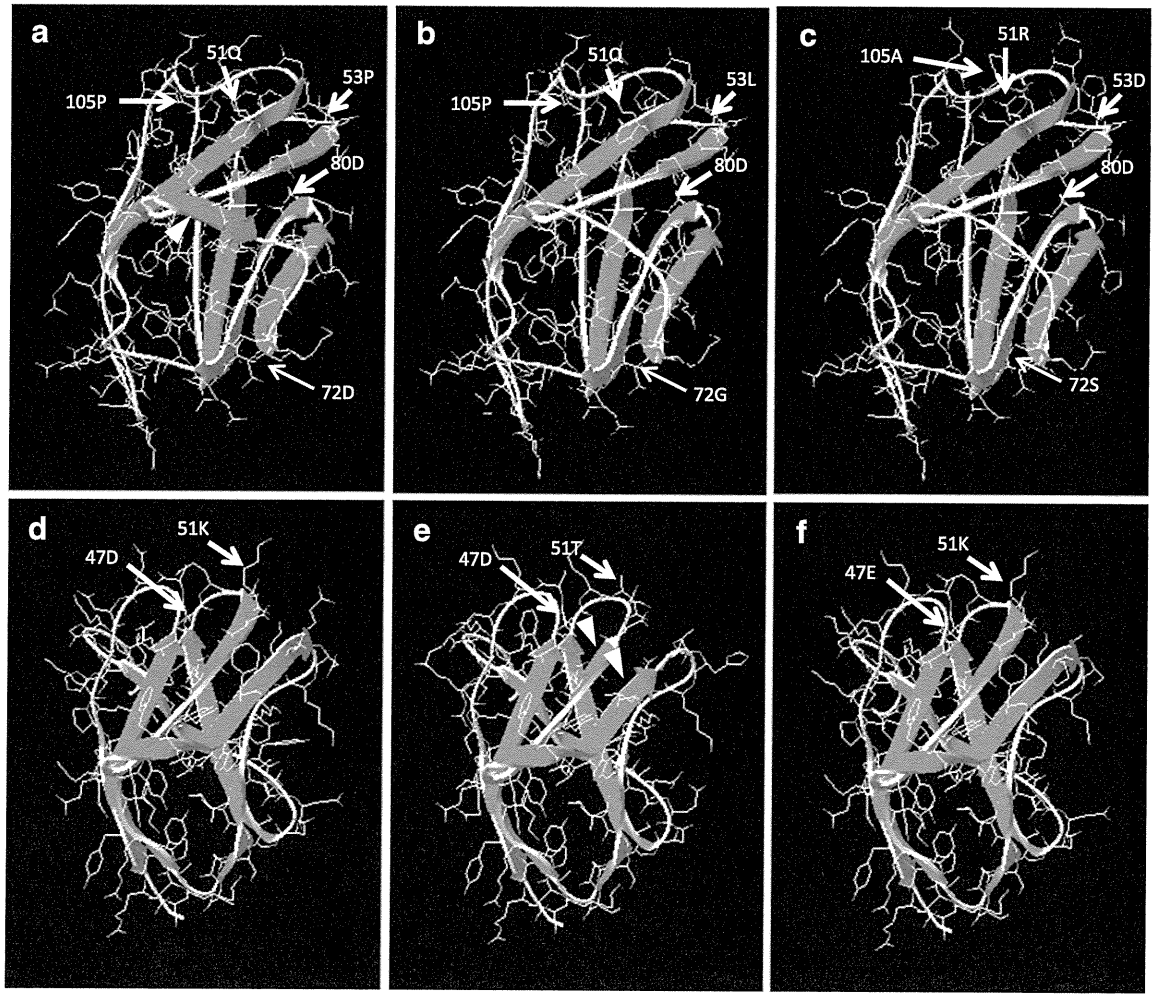

Fig. 5 Three-dimensional structures of CD3E and CD3G modeled by SWISS-MODEL. Arrows indicate amino acid sites identified as being under positive selection by using the BEB method in the PAML program. **a** Human CD3E. An *arrowhead* indicates a β-strand which is unique to human CD3E. **b** rhesus macaque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

Journal of Human Genetics (2011), 1–3

www.nature.com/jhg

SHORT COMMUNICATION

A single nucleotide polymorphism in the 3′-untranslated region of *MyD88* gene is associated with Buerger disease but not with Takayasu arteritis in Japanese

Zhiyong Chen[1,2,5], Toshiaki Nakajima[1,3], Yoshinori Inoue[2], Toshifumi Kudo[2], Masatoshi Jibiki[2], Takehisa Iwai[4] and Akinori Kimura[1,3]

Buerger disease (BD) and Takayasu arteritis (TA) are rare vascular disorders. Although their etiology and pathogenesis have not been elucidated, several studies have suggested the involvement of innate immunity. Myeloid differentiation primary-response protein 88 (MyD88) is a key signaling adaptor for all Toll-like receptors, which have a central role in innate immunity. In the present study, we evaluated the association of MyD88 with BD and TA. We conducted case–control studies in Japanese populations composing of 131 BD cases, 90 TA cases and 270 healthy controls to be genotyped for a single nucleotide polymorphism rs7744 A>G in the 3′-untranslated region of *MyD88* gene. The frequency of GG genotype was significantly lower in the BD patients than in the controls (6.9 vs 15.9%, *P*=0.011, odds ratio=0.39, 95% confidence interval; 0.19, 0.81), although there was no significant difference in the genotype frequencies between the TA patients and controls. It was suggested that MyD88 may confer resistance to BD in Japanese. Because this is the first report of the association between MyD88 and BD, replication studies in other cohorts are required.
Journal of Human Genetics advance online publication, 28 April 2011; doi:10.1038/jhg.2011.44

Keywords: Buerger disease; MyD88; single nucleotide polymorphism; Takayasu arteritis

Buerger disease (BD) and Takayasu arteritis (TA) are rare vascular diseases. BD affects small and medium sized arteries and veins. Smoking and periodontitis are well-known risk factors of BD, but the etiology and pathogenesis of BD have not been elucidated.[1] In its acute phase lesion, BD represents pathological findings of infectious characters such as occlusive, highly cellular, inflammatory thrombi and microabscesses.[1,2] We have reported that several oral microorganisms were frequently observed in the occluded arteries from patients with characteristic BD.[3] In addition, polymorphisms in genes for human leukocyte antigen and CD14, the main receptor for lipopolysaccharide derived from gram negative bacteria, are associated with BD in Japan.[4] These observations imply that bacterial infections and innate immunity might have crucial roles in the pathogenesis of BD. In contrast, TA belongs to a category of systemic vasculitis primarily involving large arteries including the aorta and its main branches as well as the coronary and pulmonary arteries.[5] Histopathological analysis of the acute phase lesion revealed massive cell infiltration mainly consisted of γδ T cells, as well as αβ T cells and NK cells around the vasa vasorum.[6] Genetic backgrounds for the susceptibility to TA have been identified in the human leukocyte antigen region.[7,8] From these observations, it was speculated that infection, which can trigger the autoimmune process, might contribute to the etiology of TA.[6]

Toll-like receptors (TLRs) have a crucial role in the early detection of pathogen-associated molecular patterns and subsequent activation of the adaptive immune response. Their ligands include a wide range of molecules, such as lipopolysaccharide, lipopeptides, bacterial DNA, double- or single-stranded RNA from virus.[9] Downstream signaling from TLRs involves five adaptor proteins known as myeloid differentiation primary-response protein 88 (MyD88), MyD88-adaptor-like, TIR-domain-containing adaptor protein inducing IFNβ, TRIF-related adaptor molecule, and sterile α- and armadillo-motif containing protein.[10]

[1]Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University, Tokyo, Japan; [2]Department of Vascular and Applied Surgery, Graduate School of Medical and Dental Sciences, Tokyo Medical and Dental University, Tokyo, Japan; [3]Laboratory of Genome Diversity, Graduate School of Biomedical Science, Tokyo Medical and Dental University, Tokyo, Japan and [4]Tsukuba Vascular Center and Buerger Disease Research Institute, Moriya, Japan
[5]Current address: Department of Rheumatology and Immunology, The Affiliated Drum Tower Hospital, Nanjing University Medical School, Nanjing, China
Correspondence: Professor 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
Received 28 January 2011; revised 26 March 2011; accepted 1 April 2011

Table 1 Primers for searching sequence variations in MyD88

Region	*Forward primer (5′–3′)*	*Reverse primer (5′–3′)*	*Length (bp)*
Promoter	GCAGCCAACCAGAAGGTGTA	GGCCTTGCCCTTTAGGTTTA	633
Promoter	TGCGACACCCTTGACACTTA	CCCTGCCCTACAATCTGGA	726
Exon 1	TGGACCTCTCCAGATCTCAAA	ATGGGAGACAGGATGCTGAG	656
Exon 2	TAGAACAACCCAGCCAGAGG	GCTTCAAACACCCATGCTC	315
Exon 3 and exon 4	TAGGCAGGGGACTCTTGG	ATCCACAGTCCTTGGGGAG	630
Exon 5	TGCCAGGGGTACTTAGATGG	GCATGTAGTCCAGCAACAGC	722
Exon 5	GCATAGCTCTGGGTCTCCTG	GTATGCTGGTGCCTGTGTGT	682
Exon 5	CTGGCCTCTGGCATATTCAT	CACCTAAGACCATGGCACCT	598

Table 2 Case–control association study of MyD88 SNP rs7744 in BD and TA patients

	Control (n=270)	*BD (n=131)*	P *for BD*	*OR for BD (95% CI)*	*TA (n=90)*	P *for TA*	*OR for TA (95% CI)*
Genotype frequency							
AA	114 (42.2%)	59 (45.0%)	NS*	1.12 (0.74, 1.71)	42 (46.7%)	NS	1.20 (0.74, 1.93)
AG	113 (41.9%)	63 (48.1%)	NS	1.29 (0.85, 1.96)	38 (42.2%)	NS	1.02 (0.63, 1.65)
GG	43 (15.9%)	9 (6.9%)	0.011	0.39 (0.19, 0.81)	10 (11.1%)	NS	0.66 (0.32, 1.37)
Allele frequency							
A	341 (63.2%)	181 (69.1%)	NS	1.30 (0.95, 1.79)	122 (67.8%)	NS	1.23 (0.86, 1.76)
G	199 (36.8%)	81 (30.9%)	NS	0.77 (0.56, 1.05)	58 (32.2%)	NS	0.84 (0.57, 1.17)
HWE	*P*=NS	*P*=NS			*P*=NS		

Abbreviations: BD, Buerger disease; CI, confidence interval; HWE, Hardy–Weinberg equilibrium; NS: not significant; OR, odds ratio; TA, Takayasu arteritis.
*$P \geqq 0.05$.

A total of 131 BD cases, 90 TA cases and 270 healthy controls were enrolled in this study. Diagnosis of BD was based on the Shionoya's criteria.[3] TA cases were diagnosed using the criteria reported by the American College of Rheumatology.[11] An informed consent for genetic studies was obtained from each subject before blood sampling and the study protocol was approved by the Ethics Review Committees of Graduate School and Medical Research Institute, Tokyo Medical and Dental University. We first searched for sequence variations in all exons and promoter region of MyD88 from 16 Japanese subjects by PCR-direct sequencing using specific primers (Table 1). Although several polymorphisms are registered in the HapMap database (http://hapmap.ncbi.nlm.nih.gov/index.html.en), we could not confirm them in our samples except for rs7744, a single nucleotide polymorphism located in the 3′-untranslated region (UTR) of *MyD88* gene. The rs7744 genotyping was done by TaqMan single nucleotide polymorphism Genotyping method (Assay ID C_3094830_10, Applied Biosystems, Life Technologies Japan, Tokyo, Japan) according to the manufacturer's instructions. As shown in Table 2, genotype distribution was not departed from Hardy–Weinberg equilibrium in both patients and controls. The frequency of GG genotype was significantly lower in the BD patients than in the controls (6.9 vs 15.9%, P=0.011, odds ratio=0.39, 95% confidence interval; 0.19, 0.81). In contrast, the genotype distribution was not significantly different between the TA patients and controls. The TA subjects also showed a similar tendency toward the lower frequency of rs7744 GG homozygote, but it was not statistically significant. The number of TA subjects analyzed in this study (n=90) was not large enough, and three times more TA and control subjects would be required, to detect the association with a significance level of $P<0.05$ at a power of 0.80.

It was difficult in this study to demonstrate certain positive evidence that rs7744 links to the expression of MyD88 or functional difference of the molecule. Because rs7744 is located at the 3′ untranslated region of *MyD88* gene, it is possible that rs7744 could be involved in the pathogenesis of BD via influencing the mRNA stability or protein expression of MyD88. We investigated whether rs7744 is located at the target sites of microRNAs by using public database, miRBase.[12] This effort failed to identify any microRNAs targeting to rs7744 region. On the other hand, we searched for sequence variations in the promoter region and all exons with ~60 bp of flanking introns, but we could identify no sequence variations, which are in tight linkage disequilibrium with rs7744. However, our single nucleotide polymorphism search did not cover the entire genomic region of MyD88, and it remains possible that rs7744 might be tightly linked to sequence variations in the intron region where we did not investigate. It has been reported that an alternatively spliced form of MyD88 acts as a negative regulator of IL-1R/TLR/MyD88-triggered signals, which lead to a negative regulation of innate immune responses,[13] and a possibility could not be excluded that an unidentified intron polymorphisms might influence the alternative splicing of *MyD88* gene. Further studies will be required to clarify the mechanisms how rs7744 is linked to the pathogenesis of BD.

Recently, two polymorphisms in MyD88 genetic region, rs4988453 and rs4988457, were reported to be associate with Hodgkin's lymphoma in Grecian.[14] The minor allele frequencies of both rs4988453 and rs4988457 were about 5% for Europe population in the HapMap database. The sites of these polymorphisms were contained in the region for screening, but we failed to detect them in our Japanese samples. As for polymorphisms of TLR genes, TLR4

(Asp299Gly; Thr399Ile) and TLR2 (Arg677Trp, Arg753Gln) had been well investigated in European populations, but we could not detect these variations in our screening samples (data not shown). These genetic discrepancies might be generated by different selective pressure, which might be in relation to the different incidences of several diseases between Asian and European populations.

To understand the pathogenesis of BD and TA, it is important to investigate the interaction between the genetic factors and environmental factors such as smoking and periodontitis in BD. We have tried to investigate a possible interaction between the rs7744 genotype and other background factors including gender, smoking history and status of periodontitis at the diagnosis, but we could not obtain any statistically definite conclusions due to the small sample size. A large cohort study will be required to demonstrate the interaction.

In conclusion, we demonstrated that a common polymorphism in *MyD88* gene was associated with BD but not with TA in Japanese. Because this is the first report, the association should be tested in other cohorts to confirm the finding. In addition, the functional impacts of rs7744 in the disease susceptibility and/or resistance should be investigated in the future.

ACKNOWLEDGEMENTS

This study was supported in part by Grant-in-Aids from the Ministry of Education, Science, Sports, Culture and Technology (MEXT) of Japan, research grants from the Ministry of Health, Labor and Welfare, Japan, grants for Japan-Korea collaboration research program from Japan Society for the Promotion of Science, research grants from The Institute of Seizon and Life Sciences and the follow-up grants provided from the Tokyo Medical and Dental University.

1 Olin, J. W. Current concepts: thromboangiitis obliterans (Buerger's disease). *N. Engl. J. Med.* **343,** 864–869 (2000).
2 Tanaka, K. Pathology and pathogenesis of Buerger's disease. *Int. J. Cardiol.* **66** (Suppl 1), S237–S242 (1998).
3 Iwai, T., Inoue, Y., Umeda, M., Huang, Y., Kurihara, N., Koike, M. *et al.* Oral bacteria in the occluded arteries of patients with Buerger disease. *J. Vasc. Surg.* **42,** 107–115 (2005).
4 Chen, Z., Takahashi, M., Naruse, T., Nakajima, T., Chen, Y. W., Inoue, Y. *et al.* Synergistic contribution of CD14 and HLA loci in the susceptibility to Buerger disease. *Hum Genet.* **122,** 367–372 (2007).
5 Numano, F., Okawara, M., Inomata, H. & Kobayashi, Y. Takayasu's arteritis. *Lancet* **356,** 1023–1025 (2000).
6 Seko, Y. Takayasu arteritis: insights into immunopathology. *Jpn. Heart J.* **41,** 15–26 (2000).
7 Kimura, A., Kitamura, H., Date, Y. & Numano, F. Comprehensive analysis of HLA genes in Takayasu arteritis in Japan. *Int. J Cardiol.* **54**(Suppl), S61–S69 (1996).
8 Kimura, A., Kobayashi, Y., Takahashi, M., Ohbuchi, N., Kitamura, H., Nakamura, T. *et al.* MICA gene polymorphism in Takayasu's arteritis and Buerger's disease. *Int. J. Cardiol.* **66**(Suppl 1), S107–S113; discussion S115 (1998).
9 Akira, S. & Takeda, K. Toll-like receptor signalling. *Nat. Rev. Immunol.* **4,** 499–511 (2004).
10 O'Neill, L. A. & Bowie, A. G. The family of five: TIR-domain-containing adaptors in Toll-like receptor signalling. *Nat. Rev. Immunol.* **7,** 353–364 (2007).
11 Arend, W. P., Michel, B. A., Bloch, D. A., Hunder, G. G., Calabrese, L. H., Edworthy, S. M. *et al.* The American College of Rheumatology 1990 criteria for the classification of Takayasu arteritis. *Arthritis Rheum.* **33,** 1129–1134 (1990).
12 Griffiths-Jones, S., Saini, H. K., van Dongen, S. & Enright, A. J. miRBase: tools for microRNA genomics. *Nucleic Acids Res.* **36,** D154–D158 (2008).
13 Burns, K., Janssens, S., Brissoni, B., Olivos, N., Beyaert, R. & Tschopp, J. Inhibition of interleukin 1 receptor/Toll-like receptor signaling through the alternatively spliced, short form of MyD88 is due to its failure to recruit IRAK-4. *J. Exp. Med.* **197,** 263–268 (2003).
14 Mollaki, V., Georgiadis, T., Tassidou, A., Ioannou, M., Daniil, Z., Koutsokera, A. *et al.* Polymorphisms and haplotypes in TLR9 and MYD88 are associated with the development of Hodgkin's lymphoma: a candidate-gene association study. *J. Hum. Genet.* **54,** 655–659 (2009).

Biochemical and Biophysical Research Communications xxx (2011) xxx–xxx

Contents lists available at ScienceDirect

Biochemical and Biophysical Research Communications

journal homepage: www.elsevier.com/locate/ybbrc

Dominant induction of vaccine antigen-specific cytotoxic T lymphocyte responses after simian immunodeficiency virus challenge

Yusuke Takahara [a,b], Saori Matsuoka [b], Tetsuya Kuwano [a], Tetsuo Tsukamoto [a], Hiroyuki Yamamoto [b], Hiroshi Ishii [a,b], Tadashi Nakasone [b], Akiko Takeda [b], Makoto Inoue [c], Akihiro Iida [c], Hiroto Hara [c], Tsugumine Shu [c], Mamoru Hasegawa [c], Hiromi Sakawaki [d], Mariko Horiike [d], Tomoyuki Miura [d], Tatsuhiko Igarashi [d], Taeko K. Naruse [e], Akinori Kimura [e], Tetsuro Matano [a,b,*]

[a] Division for AIDS Vaccine Development, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
[b] AIDS Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan
[c] DNAVEC Corporation, 6 Ohkubo, Tsukuba, Ibaraki 300-2611, Japan
[d] Institute for Virus Research, Kyoto University, 53 Kawahara-cho, Shogoin, Sakyo-ku, Kyoto 606-8507, Japan
[e] Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University, 2-3-10 Kandasurugadai, Chiyoda-ku, Tokyo 101-0062, Japan

ARTICLE INFO

Article history:
Received 12 April 2011
Available online xxxx

Keywords:
AIDS vaccine
HIV
SIV
CTL
Immunodominance

ABSTRACT

Cytotoxic T lymphocyte (CTL) responses are crucial for the control of human and simian immunodeficiency virus (HIV and SIV) replication. A promising AIDS vaccine strategy is to induce CTL memory resulting in more effective CTL responses post-viral exposure compared to those in natural HIV infections. We previously developed a CTL-inducing vaccine and showed SIV control in some vaccinated rhesus macaques. These vaccine-based SIV controllers elicited vaccine antigen-specific CTL responses dominantly in the acute phase post-challenge. Here, we examined CTL responses post-challenge in those vaccinated animals that failed to control SIV replication. Unvaccinated rhesus macaques possessing the major histocompatibility complex class I haplotype *90-088-Ij* dominantly elicited SIV non-Gag antigen-specific CTL responses after SIV challenge, while those induced with Gag-specific CTL memory by prophylactic vaccination failed to control SIV replication with dominant Gag-specific CTL responses in the acute phase, indicating dominant induction of vaccine antigen-specific CTL responses post-challenge even in non-controllers. Further analysis suggested that prophylactic vaccination results in dominant induction of vaccine antigen-specific CTL responses post-viral exposure but delays SIV non-vaccine antigen-specific CTL responses. These results imply a significant influence of prophylactic vaccination on CTL immunodominance post-viral exposure, providing insights into antigen design in development of a CTL-inducing AIDS vaccine.

1. Introduction

In human and simian immunodeficiency virus (HIV and SIV) infections, cytotoxic T lymphocyte (CTL) responses exert strong suppressive pressure on viral replication but fail to control viremia leading to AIDS progression [1–5]. A promising AIDS vaccine strategy is to induce CTL memory resulting in more effective CTL responses post-viral exposure compared to those in natural HIV infections. It is important to determine how prophylactic CTL memory induction affects CTL responses in the acute phase post-viral exposure.

We previously developed a prophylactic AIDS vaccine (referred to as DNA/SeV-Gag vaccine) consisting of DNA priming followed by boosting with a recombinant Sendai virus (SeV) vector expressing SIVmac239 Gag [6]. Evaluation of this vaccine's efficacy against a SIVmac239 challenge in Burmese rhesus macaques showed that some vaccinees contained SIV replication [7]. In particular, vaccination consistently resulted in SIV control in those animals possessing the major histocompatibility complex class I (MHC-I) haplotype *90-120-Ia* [8]; $Gag_{206-216}$ (IINEEAADWDL) and $Gag_{241-249}$ (SSVDEQIQW) epitope-specific CTL responses were shown to be responsible for this vaccine-based SIV control [9]. Furthermore, in a SIVmac239 challenge experiment of *90-120-Ia*-positive macaques that received a prophylactic DNA/SeV vaccine expressing the $Gag_{241-249}$ epitope fused with enhanced green fluorescent protein (EGFP), all the vaccinees controlled SIV replication [10]. This single epitope vaccination resulted in dominant $Gag_{241-249}$-specific CTL responses with delayed $Gag_{206-216}$-specific

* Corresponding author at: AIDS Research Center, National Institute of Infectious Diseases, 1-23-1 Toyama, Shinjuku-ku, Tokyo 162-8640, Japan. Fax: +81 3 5285 1165.
E-mail address: tmatano@nih.go.jp (T. Matano).

0006-291X/$ - see front matter © 2011 Published by Elsevier Inc.
doi:10.1016/j.bbrc.2011.04.071

CTL induction after SIV challenge, whereas $Gag_{206-216}$-specific and $Gag_{241-249}$-specific CTL responses were detected equivalently in unvaccinated *90-120-Ia*-positive animals.

These previous results in vaccine-based SIV controllers indicate dominant induction of vaccine antigen-specific CTL responses post-challenge, implying that prophylactic vaccination inducing vaccine antigen-specific CTL memory may delay CTL responses specific for viral antigens other than vaccine antigens (referred to as non-vaccine antigens) post-viral exposure. In these SIV controllers, the reduction of viral loads could be involved in delay of SIV non-vaccine antigen-specific CTL responses. Then, in the present study, we examined the influence of prophylactic vaccination on immunodominance post-challenge in those vaccinees that failed to control SIV replication. Our results showed dominant induction of vaccine antigen-specific CTL responses post-challenge even in these SIV non-controllers.

2. Materials and methods

2.1. Animal experiments

The first set of experiment used samples in our previous experiments of six Burmese rhesus macaques (Macaca mulatta) possessing the MHC-I haplotype *90-088-Ij* (macaques R02-004, R02-001, and R03-015, previously reported [7,11]; R04-014, R06-022, and R04-011, unpublished). Three of them, R02-001, R04-011, and R03-015, received a prophylactic DNA/SeV-Gag vaccine [7]. The DNA used for the vaccination, CMV-SHIVdEN, was constructed from *env*-deleted and *nef*-deleted simian–human immunodeficiency virus $SHIV_{MD14YE}$ [12] molecular clone DNA (SIVGP1) and has the genes encoding SIVmac239 Gag, Pol, Vif, and Vpx, SIVmac239-HIV chimeric Vpr, and HIV Tat and Rev. At the DNA vaccination, animals received 5 mg of CMV-SHIVdEN DNA intramuscularly. Six weeks after the DNA prime, animals received a single boost intranasally with 6×10^9 cell infectious units (CIUs) of F-deleted replication-defective SeV-Gag [13,14]. All six *90-088-Ij*-positive animals including three unvaccinated and three vaccinated were challenged intravenously with 1000 50% tissue culture infective doses (TCID50) of SIVmac239 [15] approximately 3 months after the boost. At week 1 after SIV challenge, macaque R03-015 was inoculated with nonspecific immunoglobulin G as previously described [11].

In the second set of experiment, unvaccinated (R06-001) and vaccinated (R05-028) rhesus macaques possessing the MHC-I haplotype *90-120-Ib* were challenged intravenously with 1000 TCID50 of SIVmac239. The latter R05-028 were immunized intranasally with F-deleted SeV-Gag approximately 3 months before the challenge.

In the third, three rhesus macaques received FMSIV plus mCAT1-expressing DNA vaccination three times with intervals of 4 weeks. The FMSIV DNA was constructed by replacing *nef*-deleted $SHIV_{MD14YE}$ with Friend murine leukemia virus (FMLV) *env*, carrying the same SIVmac239-derived antigen-coding regions with SIVGP1, as described before [16]. Vaccination of macaques with FMSIV and a DNA expressing the FMLV receptor (mCAT1) [17] three times with intervals of a week was previously shown to induce mCAT1-dependent confined FMSIV replication resulting in efficient CTL induction while vaccination three times with intervals of 4 weeks in the present study resulted in marginal levels of responses (data not shown). These three DNA-vaccinated animals were challenged intravenously with 1000 TCID50 of SIVmac239 approximately 2 months after the last vaccination.

Some animal experiments were conducted in the Tsukuba Primate Research Center, National Institute of Biomedical Innovation, with the help of the Corporation for Production and Research of Laboratory Primates, and others were in Institute for Virus Research, Kyoto University. All animals were maintained in accordance with the guidelines for animal experiments at the National Institute of Infectious Diseases.

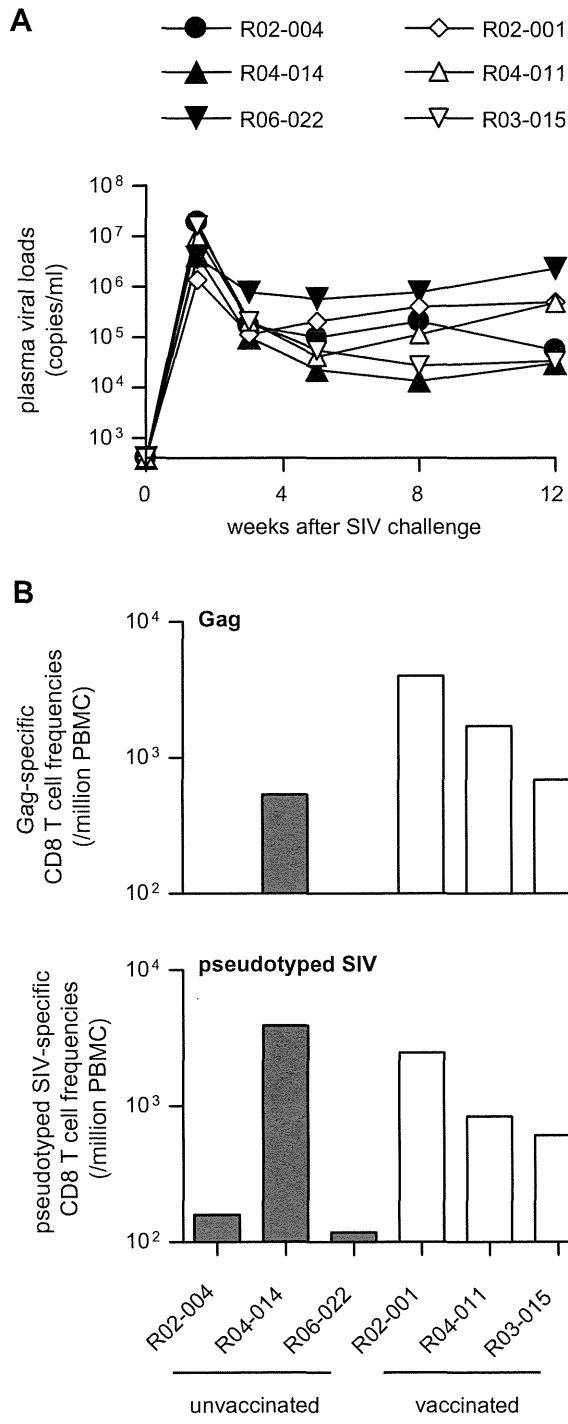

Fig. 1. CTL responses after SIVmac239 challenge in *90-088-Ij*-positive macaques. (A) Plasma viral loads after SIV challenge in unvaccinated (R02-004, R04-014, and R06-022) and DNA/SeV-Gag vaccinated animals (R02-001, R04-011, and R03-015). The viral loads (SIV gag RNA copies/ml) were determined as described previously [7]. (B) Vaccine antigen Gag-specific (upper panel) and pseudotyped SIV-specific $CD8^+$ T cell frequencies (lower panel) at week 2 after SIV challenge.

Please cite this article in press as: Y. Takahara et al., Dominant induction of vaccine antigen-specific cytotoxic T lymphocyte responses after simian immunodeficiency virus challenge, Biochem. Biophys. Res. Commun. (2011), doi:10.1016/j.bbrc.2011.04.071

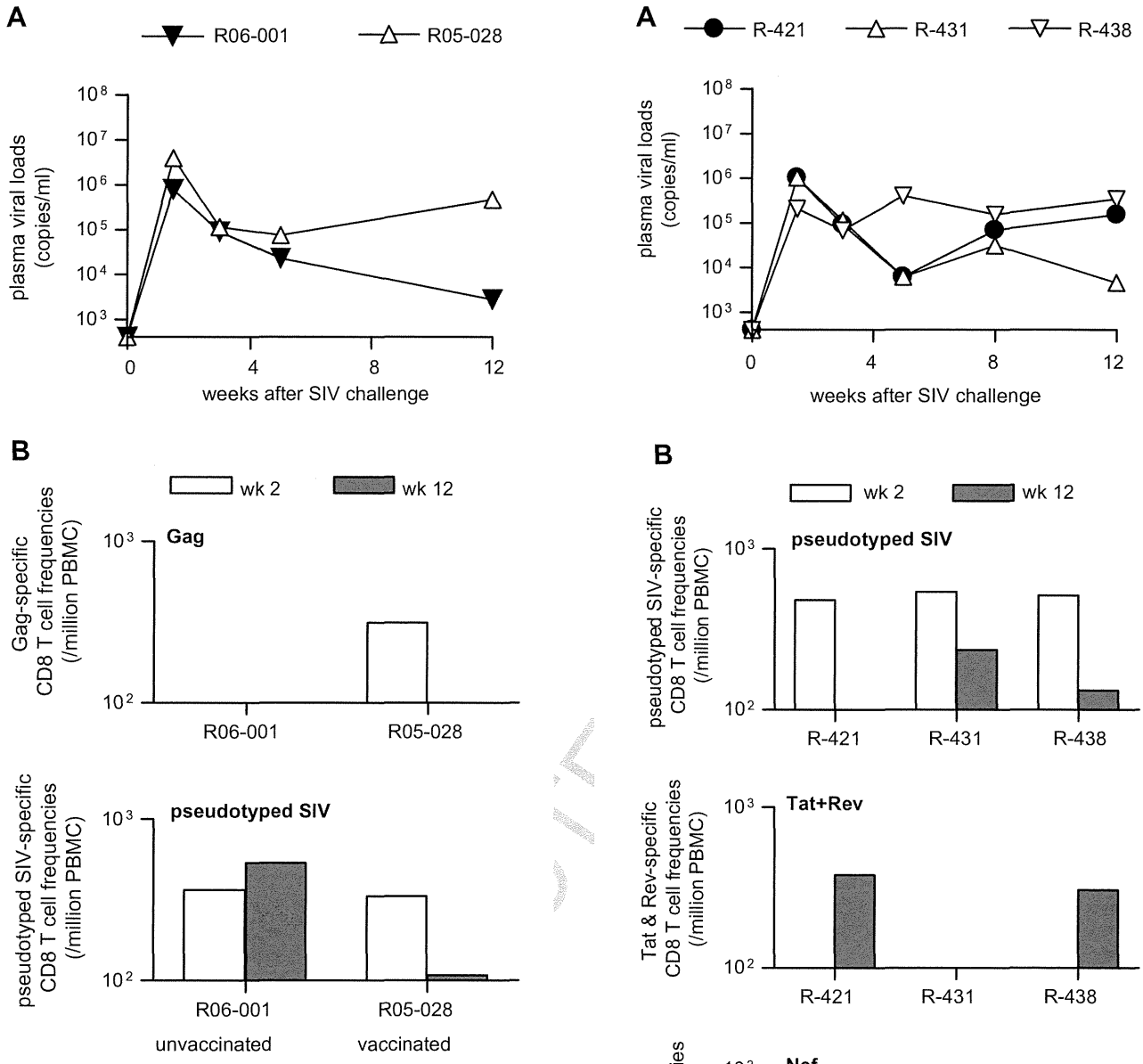

Fig. 2. CTL responses after SIVmac239 challenge in *90–120-Ib*-positive macaques. (A) Plasma viral loads after SIV challenge in unvaccinated R06–001 and SeV-Gag-vaccinated macaque R05-028. (B) Vaccine antigen Gag-specific (upper panel) and pseudotyped SIV-specific $CD8^+$ T cell frequencies (lower panel) at weeks 2 (white bars) and 12 (black bars) after SIV challenge.

Fig. 3. CTL responses after SIVmac239 challenge in DNA-vaccinated macaques. (A) Plasma viral loads after SIV challenge in DNA vaccinated macaques R-421, R-431, and R-438. (B) Vaccine antigen (pseudotyped SIV)-specific (top panel), Tat-plus-Rev-specific (middle panel), and Nef-specific $CD8^+$ T cell frequencies (bottom panel) at weeks 2 (white bars) and 12 (black bars) after SIV challenge. In macaque R-438, CTL responses at week 5 instead of week 12 are shown.

2.2. Analysis of virus-specific CTL responses

We measured virus-specific $CD8^+$ T-cell levels by flow cytometric analysis of gamma interferon (IFN-γ) induction after specific stimulation as described previously [18,19]. Peripheral blood mononuclear cells (PBMCs) were cocultured with autologous herpesvirus papio-immortalized B-lymphoblastoid cell lines (B-LCLs) infected with a vaccinia virus vector expressing SIVmac239 Gag for Gag-specific stimulation or a vesicular stomatitis virus G protein (VSV-G)-pseudotyped SIV for pseudotyped SIV-specific stimulation. The pseudotyped SIV was obtained by cotransfection of COS-1 cells with a VSV-G-expression plasmid and SIVGP1 DNA. Alternatively, PBMCs were cocultured with B-LCLs pulsed with peptide pools using panels of overlapping peptides spanning the entire SIVmac239 Tat, Rev, and Nef amino acid sequences. Intracellular IFN-γ staining was performed with a CytofixCytoperm kit (Becton Dickinson, Tokyo, Japan) and fluorescein isothiocyanate-conjugated anti-human CD4, peridinin chlorophyll protein-conjugated anti-human CD8, allophycocyanin-conjugated anti-human CD3, and phycoerythrin-conjugated anti-human

Please cite this article in press as: Y. Takahara et al., Dominant induction of vaccine antigen-specific cytotoxic T lymphocyte responses after simian immunodeficiency virus challenge, Biochem. Biophys. Res. Commun. (2011), doi:10.1016/j.bbrc.2011.04.071