Diversity of *ULBP5* in Old-World monkeys (Cercopithecidae) and divergence of the *ULBP* gene family in primates

By Taeko K. NARUSE,^{*1} Hirofumi AKARI,^{*2} Tetsuro MATANO^{*3,*4} and Akinori KIMURA^{*1,†}

(Communicated by Shizuo AKIRA, M.J.A.)

Abstract: Non-human primates such as rhesus macaque and cynomolgus macaque are important animals for medical research. These species are classified as Old-World monkeys (Cercopithecidae), in which the immune-related genome structure is characterized by gene duplications. In the present study, we investigated polymorphisms in two genes for *ULBP5* encoding ligands for NKG2D. We found 18 and 11 *ULBP5.1* alleles and 11 and 13 *ULBP5.2* alleles in rhesus macaques and cynomolgus macaques, respectively. In addition, phylogenetic analyses revealed that *ULBP5.2* diverged from a branch of *ULBP5.1*. These data suggested that human *ULBP* genes diverged from an ancestral gene of *ULBP2-ULBP5* and that *ULBP6/RAET1L*, specifically identified in human, diverged from an ancestral *ULBP2* by a recent gene duplication after the diversification of homininae (human and other higher great apes), which were consistent with the findings in our previous analysis of *ULBP2* genes in rhesus and cynomolgus macaques.

Keywords: rhesus macaque, cynomolgus macaque, *ULBP5/RAET1G* gene, NKG2D, polymorphism

Introduction

Human UL16 binding protein (*ULBP*)/retinoic acid early transcript 1 (*RAET1*) molecules are encoded by the *ULBP/RAET1* gene family located on chromosome 6q24.2, which is composed of 10 members, including six functional genes, *ULBP1*, 2, 3, 4, 5, and 6, corresponding to *RAET1I*, *H*, *N*, *E*, *G*, and *L*, respectively.^{1)–4)} There are several major histocompatibility complex (MHC) class I-like molecules, including *ULBP/RAET1* and MHC class I chain-related chain (MIC), which are known as ligands for a C-type lectin molecule, natural-killer group 2 member D (NKG2D).^{2),5)–7)} NKG2D is an activating receptor expressed on the surface of NK,

$\gamma\delta^+$ and $CD8^+$ T cells that plays an important role in the immune response system.^{8),9)} These ligands are usually stress-inducible, and their recognition by NKG2D can lead to the activation of NK cells, resulting in the killing of virus-infected cells and tumor cells.^{10)–13)} It has been reported that each *ULBP* gene has several sequence variations in humans.^{14),15)} Although the cell surface expression of ligand molecules on target cells is differentially regulated,¹¹⁾ genetic variations or polymorphisms in the coding region may also have a functional impact.

In the medical field, non-human primates including rhesus and cynomolgus macaques are used as animal models in immunological studies for infectious diseases, autoimmune diseases, organ transplantation, and the development of vaccines. These macaques are Old-World monkeys, and it has been reported that the genetic diversity of immune-related genes in the rhesus macaque is unique, *i.e.*, more than 60% of rhesus macaque-specific expansions of immune-related genes are found in the protein coding sequences.¹⁶⁾ To fully evaluate the results of immunological experiments using macaque models, it is essential to characterize the genetic diversity of immune-related molecules, which may shape the basis of individual differences in the immune response

^{*1} Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University (TMDU), Tokyo, Japan.

^{*2} Primate Research Institute, Kyoto University, Inuyama, Japan.

^{*3} AIDS Research Center, National Institute of Infectious Diseases, Tokyo, Japan.

^{*4} The Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[†] Correspondence should be addressed: A. Kimura, Department of Molecular Pathogenesis, Medical Research Institute, Tokyo Medical and Dental University (TMDU), 1-5-45 Yushima, Bunkyo-ku, Tokyo 113-8510, Japan (e-mail: akitis@mri.tmd.ac.jp).

against foreign antigens and/or pathogens. It has been reported that the copy numbers of genes in the *MHC* loci in Old-World monkeys are higher than those in humans.^{16)–18)}

In addition, the extent of genetic diversity in the *MHC* loci of Old-World monkeys differs in part depending on their origin, such that the diversity of *MHC* class I genes is considerably different depending on their habitat in both rhesus and cynomolgus macaques.^{19),20)} In our previous study, we demonstrated that *ULBP2* and *ULBP4* are more polymorphic in Old-World monkeys compared with in humans.^{21),22)} It was also revealed that each member of the *ULBP/RAET1* gene family, except for *ULBP6*, had been duplicated in the rhesus genome.^{21),22)}

In the present study, we focused on sequence variations of the *ULBP5* genes in rhesus and cynomolgus macaques, to further delineate the divergence and diversity of the *ULBP* gene family in macaques and humans.

Materials and methods

Subjects. The study subjects were 38 rhesus macaques from seven colonies and 24 cynomolgus macaques from five colonies, previously analyzed for polymorphisms in *MHC* class I genes.^{19),20)} They were maintained in breeding colonies at Tsukuba Primate Research Center, National Institute of Biomedical Innovation, Japan or the Primate Research Institute, Kyoto University, Japan. The founders of the rhesus macaque colonies were captured in Myanmar and Laos,¹⁹⁾ whereas the founders of cynomolgus macaque colonies were captured in Indonesia, Malaysia, and the Philippines.²⁰⁾ All care, including blood sampling of animals, were in accordance with the guidelines for the Guide for the Care and Use of Laboratory Animals (National Research Council, 8th ed., 2010) and were subjected to prior approval by the Animal Experimentation Committees of The University of Tokyo and Kyoto University.

DNA extraction and sequencing analysis. Genomic DNA from B lymphoblastoid cell lines from rhesus macaques and whole blood samples from cynomolgus macaques were extracted using a Quick-Gene DNA kit (FUJIFILM Corp., Tokyo, Japan), as described previously.^{21),22)} Amplification of each rhesus *ULBP5* was performed with two primer pairs, which were designed for the region spanning from intron 1 to intron 3 of two rhesus *ULBP5*, LOC694076 (designated *ULBP5.1*) and LOC694265 (designated *ULBP5.2*), using FastStart Taq DNA polymerase (Roche, Mannheim, Germany). Se-

quences of the primer pairs were as follows: *UL5.1F* (5'-TCCTACCTCCCACAATCCACA) and *UL5.1R* (5'-AGGAGCGCCCCATG) for *ULBP5.1* (LOC694076), and *UL5.2F* (5'-CCTTTGCCATC-CAGACC) and *UL5.2R* (5'-AGAAGCAAGTG-CAGTGAGGA) for *ULBP5.2* (LOC694265). These primers were also used for the amplification of cynomolgus *ULBP5*. The polymerase chain reaction (PCR) program was composed of the following steps: denaturation at 95 °C for 4 min; 30 cycles of 95 °C for 30 sec, 63 °C for 30 sec for *ULBP5.1* or 56 °C for 30 sec for *ULBP5.2*, 72 °C for 45 sec; and additional extension at 72 °C for 7 min. Expected sizes of the amplified fragments were about 1,270 bp and 1,307 bp for *ULBP5.1* and *ULBP5.2*, respectively, and the PCR products were cloned into pSTBlue-1 AccepTer vector (Novagen, WI, U.S.A.) in accordance with the manufacturer's instructions and transformed into Nova Blue SingleTM competent cells (Merck Biosciences Japan, Tokyo, Japan). Ten to 20 independent transformed colonies were picked up for each sample and subjected to sequencing on both strands using a BigDye Terminator cycling system and an ABI 3730 automated sequence analyzer (Applied Biosystems, CA, U.S.A.).

Data analysis. Nucleotide sequences from cloned DNAs were aligned using the Genetyx software package (version 8.0, Genetyx Corp., Japan). When 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 using Kimura's 2-parameter method for phylogenetic analysis of the sequences of exon 2 and exon 3 of *ULBP5*, as well as intron 2 sequences, using Genetyx software. Bootstrap values were based on 5,000 replications. The *ULBP1* to *ULBP5* sequences from rhesus monkeys: LOC694216, LOC694341 (*ULBP1*), LOC694466, LOC694600 (*ULBP2*), LOC694135, LOC694525 (*ULBP3*), LOC695031 (*ULBP4*), LOC694076, LOC694265 (*ULBP5*), and AM418584 (*MICAB*); *ULBP1* to *ULBP6* sequences from humans, GenBank accession numbers AL355497 (*ULBP1*, *ULBP3*, *ULBP6*), AL583835 (*ULBP2* and *ULBP5*) and AL355312 (*ULBP4*), were adopted as references. Similarly, cynomolgus (LOC102142491 as *RAET1L*), human (NG034139 as *MICA*, NG21405 as *MICB* and AB003618 as *MICB* variant, *MICB*0107N*), chimpanzee (LOC463066 as *ULBP2*, LOC748488 as *ULBP5* and NC 006473 as *MICB*), western gorilla (LOC101140391 as *ULBP2*), and another mem-

bers of the Old-World monkeys, olive baboon (LOC101017701 as *ULBP2*), as well as a member of the New-World monkeys, marmoset (LOC100390841 as *MICA*), and mouse (AC159330 as *Rae1- α*) were included in the phylogenetic analysis as references. Rates of synonymous change in exons and variation in introns were calculated for *ULBP5* genes in rhesus and cynomolgus macaques and indicated as average $\pm$ SD. Statistical significance of differences among mutation rates was examined using Welch's t-test. Statistical significance was defined as $p < 0.05$.

Nomenclature. Alleles of *ULBP5.1* and *ULBP5.2* are designated using a 6-digit system such as *Mamu-ULBP5.1*01.01.01*, in which the first, second, and third pairs of digits indicate alleles that differ in their predicted amino acid sequences, synonymous sites in exons, and variation in introns, respectively. In addition, as for the nomenclature for the alleles at the predicted amino acid sequences, we placed “-” instead of “*” after the protein name, such as *Mamu-ULBP5.1-01.01.01*, to distinguish it from the nomenclature for nucleotide sequences.

Results

Identification of alleles for the *ULBP5* gene, *ULBP5.1*. There are two paralogous genes for *ULBP5*, LOC694076 and LOC694265, in the rhesus macaque genome. In the present study, we designated LOC694076 and LOC694265 as *ULBP5.1* and *ULBP5.2*, respectively, and designed primer pairs in which the deletion site of nucleotide sequences in LOC694265 was used to separately amplify *ULBP5.1* and *ULBP5.2*. As expected, PCR products from each gene could be obtained and distinguished by their length, although minor length differences due to a CCT repeat number polymorphism in intron 1 of *ULBP5.1* and several deletion sites in exons and introns of *ULBP5.2* were observed.

We obtained nucleotide sequences for the region from part of intron 1 to part of intron 3 of *ULBP5.1* from 38 rhesus macaques and 24 cynomolgus macaques by the sequencing of cloned PCR products. The *ULBP5.1* sequences from rhesus macaques were classified into 18 different alleles (Table 1), designated as *Mamu-ULBP5.1*01.01.01* to *-ULBP5.1*08.01.01*. The LOC4964076 sequence was given with the allele name *Mamu-ULBP5.1*01.01.01*. In cynomolgus macaques, 11 different alleles, *Mafa-ULBP5.1*01.01.01* to *-ULBP5.1*09.01.01*, were identified (Table 1).

Identification of alleles for another *ULBP5* gene, *ULBP5.2*. PCR products for *ULBP5.2*

could be obtained from genomic DNA of both rhesus and cynomolgus macaques. Sequencing data from the cloned PCR products were classified into 11 different alleles, *Mamu-ULBP5.2*01.01.01* to *-ULBP5.2*09.01.01*, from rhesus macaques and 13 alleles, *Mafa-ULBP5.2*01.01.01* to *-ULBP5.2*10.01.01*, from cynomolgus macaques (Table 1). Nucleotide sequences of *Mamu-ULBP5.2*01.01.01* and *Mafa-ULBP5.2*01.01.01* were identical to each other and to LOC694265.

Diversity of *ULBP5* genes in Old-World monkeys. As for the diversification of *ULBP5* in macaques, several non-synonymous substitutions were observed, and amino acid sequences (AASs) were deduced from the nucleotide sequences. As a result, *ULBP5* polymorphisms were classified into 11, 9, 7, and 7 AAS alleles in *Mamu-ULBP5.1*, *Mafa-ULBP5.1*, *Mamu-ULBP5.2*, and *Mafa-ULBP5.2*, respectively (Table 2, Supplementary Table 1). There were several AAS alleles with truncating mutations in both rhesus and cynomolgus *ULBP5.2* (Fig. 1). In particular, the *Mamu-ULBP5.2-04.01.01* AAS allele was predicted to encode a very short form due to a stop codon caused by a single nucleotide substitution early in exon 2. In addition, *Mamu-ULBP5.2-05.01.01*, *-ULBP5.2-05.02.01*, and *-ULBP5.2-06.01.01* AAS alleles were also predicted to be truncated at amino acid position 15, as a result of a frameshift by two nucleotide deletions at exon 2 in *Mamu-ULBP5.2*05.01.01*, *-ULBP5.2*05.02.01*, and *-ULBP5.2*06.01.01* alleles, respectively. Furthermore, *Mamu-ULBP5.2-09.01.01*, *Mafa-ULBP5.2-08.01.01*, *-ULBP5.2-09.01.01*, and *-ULBP5.2-10.01.01* AAS alleles carried several amino acid substitutions and truncations, as shown in Fig. 1. These changes were generated by a single nucleotide “C” insertion in exon 3. Notably, the nucleotide position of each insertion was conserved among the *Mamu* and *Mafa* *ULBP5.2* alleles, suggesting that this insertion occurred in an ancestral population of rhesus and cynomolgus macaques.

Phylogenetic tree analysis of the *ULBP5* gene. To determine out the evolutionary divergence and diversity of *ULBP5* genes, we conducted a neighbor-joining analysis using nucleotide sequences, from intron 1 to intron 3, from *ULBP5.1* and *ULBP5.2* in macaques. As shown in Fig. 2, *ULBP5.2* alleles appeared to be generated from a branch of *ULBP5.1*, and several *ULBP5.1* and *ULBP5.2* alleles from both rhesus and cynomolgus macaques were found in the same cluster.

Table 1. Alleles of *ULBP5.1* or *ULBP5.2* genes in rhesus and cynomolgus macaques

Gene	Species	Allele name	Accession No	ID of reference animal	Clone name	Remarks
<i>ULBP5.1</i>	<i>Macaca mulatta</i>	<i>Mamu-ULBP5.1*01.01.01</i>	LC322170	R473		Identical to <i>LOC694076</i>
		<i>Mamu-ULBP5.1*02.01.01</i>	LC322171	R320, R321, R396	R321-10F	
		<i>Mamu-ULBP5.1*02.02.01</i>	LC322173	R314, R429, R434, R455	R429-1F	
		<i>Mamu-ULBP5.1*02.02.02</i>	LC322172	R333, R360, R496	R333-5F	
		<i>Mamu-ULBP5.1*02.03.01</i>	LC322174	R492	R492-3F	
		<i>Mamu-ULBP5.1*02.04.01</i>	LC322175	R312, R350, R437	R312-2F	
		<i>Mamu-ULBP5.1*03.01.01</i>	LC322176	R396	R396-16F	
		<i>Mamu-ULBP5.1*04.01.01</i>	LC322177	R328, R337, R384, R495	R337-2F	
		<i>Mamu-ULBP5.1*05.01.01</i>	LC322178	R237, R333, R337, R346, R360, R361, R379, R384, R408, R429, R430, R439, R446, R446, R491, R492, R495	R333-1F	
		<i>Mamu-ULBP5.1*05.01.02</i>	LC322179	R434	R434-3F	
		<i>Mamu-ULBP5.1*05.01.03</i>	LC322180	R350, R477	R350-2F	
		<i>Mamu-ULBP5.1*05.01.04</i>	LC322181	R477	R477-9F	
		<i>Mamu-ULBP5.1*05.01.05</i>	LC322182	R234, R237, R321, R325, R337, R439, R446, R465, R473, R490, R496	R234-3F	
		<i>Mamu-ULBP5.1*05.01.06</i>	LC322187	R312, R342, R437	R312-1F	
		<i>Mamu-ULBP5.1*06.01.01</i>	LC322183	R314	R314-4F	
		<i>Mamu-ULBP5.1*06.01.02</i>	LC322184	R237, R312	R237-1F	
		<i>Mamu-ULBP5.1*07.01.01</i>	LC322185	R491	R491-7F	
		<i>Mamu-ULBP5.1*08.01.01</i>	LC322186	R342, R408	R342-8F	
	<i>Macaca fascicularis</i>	<i>Maifa-ULBP5.1*01.01.01</i>	LC322188	P01, P02, C003, C004, C005, C007	P01-F-2	
		<i>Maifa-ULBP5.1*02.01.01</i>	LC322189	M06, C012, C013	M06-F-11	
		<i>Maifa-ULBP5.1*03.01.01</i>	LC322190	M02, C004	M02-F-6	
		<i>Maifa-ULBP5.1*03.01.02</i>	LC322191	M02	M02-F-6 (2)	
		<i>Maifa-ULBP5.1*04.01.01</i>	LC322192	M03, C006	M03-F-5	
		<i>Maifa-ULBP5.1*05.01.01</i>	LC322193	P04, M03, M04, M06, C007, C008, C009, C010, C013	M06-F-13	
		<i>Maifa-ULBP5.1*06.01.01</i>	LC322194	M05, C010	M05-F-6	
		<i>Maifa-ULBP5.1*07.01.01</i>	LC322195	P03, C008, C009	P03-F-3	
		<i>Maifa-ULBP5.1*08.01.01</i>	LC322196	P02, C006	P02-F-8	
		<i>Maifa-ULBP5.1*08.01.02</i>	LC322197	P04, M05	P04-F-7	
		<i>Maifa-ULBP5.1*09.01.01</i>	LC322198	P01, M01, C001, C002, C003	M01-F-6	

Continued on next page.

To further delineate the phylogeny of *ULBP* and *MIC*, genes for ligands of NKG2D, we used to conduct a phylogenetic analysis using the mouse *rae1-a* sequence as an outgroup. Because the sequences of intron 1 and intron 3 were too diverged to be aligned among the *ULBP* and *MIC*, we used sequences from exon 2 to exon 3 in this analysis. By trimming sequences of intron 1 and intron 3, 15

and 11 *ULBP5.1* and 10 and 11 *ULBP5.2* nucleotide sequences of rhesus and cynomolgus macaques, respectively, were aligned. As shown in Fig. 3, alleles of *ULBP5.1* and those of *ULBP5.2* in rhesus and cynomolgus macaques were clustered separately, but both rhesus and cynomolgus alleles were found in the same cluster. To our surprise, *ULBP5.1* alleles were generated from a branch of *ULBP5.2*, which

Continued.

Gene	Species	Allele name	Accession No	ID of reference animal	Clone name	Remarks
<i>ULBP5.2</i>	<i>Macaca mulatta</i>	<i>Mamu-ULBP5.2*01.01.01</i>	LC322199	R234, R237, R333, R349, R350, R360, R361, R429, R434, R473, R492, R495	R333F-9	Identical to <i>LOC694265</i> , <i>Mafa-ULBP5.2*01.01.01</i>
		<i>Mamu-ULBP5.2*01.01.02</i>	LC322200	R312	R312F5	
		<i>Mamu-ULBP5.2*02.01.01</i>	LC322201	R316, R437	R437-4F	
		<i>Mamu-ULBP5.2*03.01.01</i>	LC322202	R325, R333, R337, R349, R360, R379, R384, R430, R477	R325F9	
		<i>Mamu-ULBP5.2*04.01.01</i>	LC322203	R328, R337, R350, R384, R396, R437	R328F-8	Truncation mutation
		<i>Mamu-ULBP5.2*05.01.01</i>	LC322204	R342, R495	R342F-1	Truncation mutation
		<i>Mamu-ULBP5.2*05.02.01</i>	LC322205	R465	R465-1F	Truncation mutation
		<i>Mamu-ULBP5.2*06.01.01</i>	LC322206	R321	R321-F5	Truncation mutation
		<i>Mamu-ULBP5.2*07.01.01</i>	LC322207	R314	R314-F1	
		<i>Mamu-ULBP5.2*08.01.01</i>	LC322208	R312	R312-F7	
	<i>Macaca fascicularis</i>	<i>Mamu-ULBP5.2*09.01.01</i>	LC322209	R234, R237, R314, R320, R321, R328, R346, R361, R396, R408, R429, R434, R455, R465, R473, R490, R492, R494	R328F-4	Truncation mutation
		<i>Mafa-ULBP5.2*01.01.01</i>	LC322210	M04, C009	M04-F-3	Identical to <i>LOC694265</i> , <i>Mamu-ULBP5.2*01.01.01</i>
		<i>Mafa-ULBP5.2*02.01.01</i>	LC322211	M05, P04, C010, C011	M05-F-12	
		<i>Mafa-ULBP5.2*02.01.02</i>	LC322212	M06, C013	M06-F-6	
		<i>Mafa-ULBP5.2*03.01.01</i>	LC322213	M03, C006, C007	M03-F-6	
		<i>Mafa-ULBP5.2*04.01.01</i>	LC322214	P03, C008, C009	P03-F-1	
		<i>Mafa-ULBP5.2*05.01.01</i>	LC322215	P01, M03, C007	P01-F-2	
		<i>Mafa-ULBP5.2*06.01.01</i>	LC322216	P01, M002, C001, C002, C003, C004, C005	P01-F-5	
		<i>Mafa-ULBP5.2*07.01.01</i>	LC322217	P02, C007	P02-F-7	
		<i>Mafa-ULBP5.2*07.01.02</i>	LC322218	M04, C008	M04-F-1	
		<i>Mafa-ULBP5.2*07.02.01</i>	LC322219	M02, C004, C005	M02-F-1	
		<i>Mafa-ULBP5.2*08.01.01</i>	LC322220	P04, P05, P06, M05, M06, C012, C013	P04-F-7	Truncation mutation
		<i>Mafa-ULBP5.2*09.01.01</i>	LC322221	P02, C006	P02-F-6	Truncation mutation
		<i>Mafa-ULBP5.2*10.01.01</i>	LC322222	P03	P03-F-4	Truncation mutation

was contradictory to the data in Fig. 2. This might be caused by the addition of the outgroup.

To obtain additional data for which of *ULBP5.1* and *ULBP5.2* had been generated earlier, the rate of variations at synonymous sites in exons (exons 2 and 3) and introns (introns 1, 2, and 3) was calculated for *ULBP5.1* and *ULBP5.2* in rhesus and cynomolgus macaques (Table 2). It was revealed that the variation rate of *ULBP5.1* was significantly higher than that of *ULBP5.2* in both rhesus and

cynomolgus macaques. Because synonymous sites and introns are not under natural selection, these findings supported that *ULBP5.2* was generated from a branch of *ULBP5.1*, if the mutation rate at the neutral sites was similar for both *ULBP5.1* and *ULBP5.2*.

Phylogenetic data in Fig. 3 demonstrated that *ULBP* apparently diverged from *MIC*, and each *ULBP* showed specific diversification except that *ULBP2* and *ULBP5* diverged from the same ances-

Table 2. Number of alleles in *ULBP5.1* and *5.2* genes in rhesus and cynomolgus macaques

Macaque species	Locus	Allele	AAS allele* (truncated allele)	Intron 1		Exon 2		Intron 2		Exon 3		Intron 3	Rate of synonymous change in exons and variation in introns
				Polymorphic sites	CCT ins/del	Polymorphic sites** (TT ins)	Nonsynonymous sites	Polymorphic sites	CCT ins/del	Polymorphic sites*** (C ins)	Nonsynonymous sites	Polymorphic sites (25 bp ins/del)****	
Rhesus (n = 38)	<i>Mamu</i> <i>ULBP5.1</i>	18	11	10	4	11	5	11	0	18	8	1	0.0121 ± 0.0077
	<i>Mamu</i> <i>ULBP5.2</i>	11	7 (3)	4	0	18 (1)	10	6	1	19 (1)	10	4 (1)	0.0051 ± 0.0072
Cynomolgus (n = 24)	<i>Mafa</i> <i>ULBP5.1</i>	11	9	9	4	11	5	9	0	16	10	2	0.0184 ± 0.0049
	<i>Mafa</i> <i>ULBP5.2</i>	13	7 (2)	4	0	15	8	4	1	23 (1)	15	5 (1)	0.0117 ± 0.0070

* ; number of alleles at amino acid sequences (AAS) and number of truncated alleles are in parentheses.

** ; polymorphic sites in exon 2 including a frameshift mutation indicated in parentheses.

*** ; polymorphic sites in exon 3 including a frameshift mutation indicated in parentheses.

**** ; polymorphic sites in intron 3 including a 25 bp ins/del polymorphism indicated in parentheses.

tral gene. Moreover, when only exon 2 and exon 3 sequences were used in the phylogenetic analysis, similar results were obtained except that *ULBP5.1* and *ULBP5.2* were clustered separately (Supplementary Fig. 1). The difference in the phylogenetic aspects may be because the polymorphic exons of *ULBP* and *MIC* had been under natural selection.

Discussion

In humans, *ULBP5/RAET1G* transcripts are detected in most tissues, with an overall expression pattern similar to *ULBP2*. Nucleotide sequences for extracellular domains, $\alpha 1$ and $\alpha 2$, of *ULBP5* showed a high homology with those of *ULBP2* and *ULBP6*, *i.e.*, 94% and 96% nucleotide sequence identity, respectively.¹⁾ On the other hand, we reported previously that each member of the *ULBP/RAET1* gene family was duplicated in the rhesus genome, and any orthologous genes to human *ULBP6/RAET1L* were not detected in macaques.^{21),22)} When we obtained *ULBP2.1* and *ULBP2.2* sequences from macaques, we considered that *ULBP2.1* or *ULBP2.2* might be orthologous to human *ULBP6*. However, our observations including the results of phylogenetic analyses have suggested that *ULBP2.2* diverged from *ULBP2.1*, and both are orthologs of human *ULBP2*.²²⁾ This is why we focused on *ULBP5* in this study, in order to verify any possible orthologous genes to human *ULBP6* in Old-World monkeys.

We identified 18 and 11 *ULBP5.1* alleles and 12 and 13 *ULBP5.2* alleles in rhesus and cynomolgus macaques, respectively. The average sequence homologies with the $\alpha 1$ and $\alpha 2$ domains of human *ULBP5/RAET1G* were 92% for *ULBP5.1* and 93% for *ULBP5.2*, and these results were consistent with our previous findings from the analysis of *ULBP2* and *ULBP4*, *i.e.*, both *ULBP2* and *ULBP4* were duplicated and highly polymorphic even at the amino acid sequence level.^{21),22)} Phylogenetic analyses indicated that the clustering of *ULBP5.1* alleles and *ULBP5.2* alleles did not depend on the species, *i.e.*, alleles from both rhesus macaque and cynomolgus macaques were found in the same clusters, supporting a trans-species evolution possibly caused by either balancing selection or incomplete sorting of alleles (Fig. 2), and the clustering pattern was similar to those for *ULBP2.1* and *ULBP2.2* in our previous study.²²⁾ These observations imply that *ULBP5.1* and *ULBP5.2* are orthologous to human *ULBP5/RAET1G*, but not to *ULBP6/RAET1L*. As shown in Fig. 3, a phylogenetic analysis using sequences of corresponding regions from *ULBP1* to *ULBP6*,

On the other hand, it has been reported that ULBP5/RAET1G only weakly interacts with NKG2D and does not bind cytomegalovirus glycoprotein UL16. It exhibits an apparently lower avidity for NKG2D due to substitution of the amino acid at position 122 (133, position in the mature form) in the $\alpha 2$ domain (Fig. 1), in the center of the MHC-like fold, whereas ULBP2 strongly binds NKG2D.²⁴⁾ In

Fig. 2. Phylogenetic tree of *ULBP5.1* and *ULBP5.2* alleles. The tree was constructed using the neighbor-joining method with bootstrap values of 5,000 replications. The values are indicated as percentages, but those less than 50% are not shown. The underlined alleles contained truncation mutations. Stars indicate alleles carrying Leu at position 122 (113), which is a contact site with NKG2D.

the 3-D structure of *ULBP3*, phenylalanine 122 (113) is located in the middle of the β -pleated sheet of the $\alpha 2$ domain and contributes to the hydrophobic core of the $\alpha 1$ - $\alpha 2$ domain,²⁵⁾ where Phe122 (113) is conserved among all human *ULBP*, except *ULBP5/RAET1G*. On the contrary, *ULBP5/RAET1G* has Leu instead of Phe at this site and no polymorphism

was reported so far.^{14),15)} Nevertheless, amino acid polymorphism of Phe or Leu at 122 (113) was found only in the cynomolgus *ULBP5.2*; Mafa-*ULBP5.2**1 had Phe at 122 (113), and the remaining nine alleles had Leu, whereas Phe122 (113) was conserved in *ULBP5.1* of both rhesus and cynomolgus macaques as well as in *ULBP5.2* of rhesus macaque. Therefore,

Fig. 3. Phylogenetic tree of alleles of *ULBP5* and other *ULBP* genes among Old-World monkeys and hominines. The tree was constructed using the neighbor-joining method with bootstrap values of 5,000 replications. Nucleotide sequences from exon 2, intron 2, and exon 3 were used in the analysis. The bootstrap values are indicated as percentages, but those less than 50% are not shown. There were several alleles with the same nucleotide sequences in the analyzed region, which are grouped into 'alleles'. Truncation alleles were underlined. The *ULBP1* to *ULBP5* sequences from rhesus; LOC694216, LOC694341 (*ULBP1*), LOC694466, LOC694600 (*ULBP2*), LOC694135, LOC694525 (*ULBP3*), LOC695031 (*ULBP4*), LOC694076, LOC694265 (*ULBP5*), and AM418584 (*MICAB*), and *ULBP1* to *ULBP6* sequences from human; GenBank accession numbers AL355497 (*ULBP1*, *ULBP3*, *ULBP6*), AL583835 (*ULBP2* and *ULBP5*), and AL355312 (*ULBP4*). Cynomolgus (LOC102142491 as *RAET1L*), human (NG034139 as *MICA*, NG21405 as *MICB* and AB003618 as *MICB* variant, *MICB*0107N*), chimpanzee (LOC463066 as *ULBP2*, LOC748488 as *ULBP5* and NC 006473 as *MICB*), western gorilla (LOC101140391 as *ULBP2*), olive baboon (LOC101017701 as *ULBP2*), marmoset (LOC100390841 as *MICA*) and mouse (AC159330 as *Rae1-α*).

all 13 cynomolgus *ULBP5.2* alleles containing a truncation and/or nine substituted alleles may result in weaker interaction with NKG2D. To clarify the biological significance of duplication and polymorphisms in macaque *ULBP* genes, the functional impact of the polymorphisms should be investigated in future studies.

Acknowledgements

We thank Yukiko Ueda, Yasuko Saida, and Nana Ohkubo for technical assistance. We also thank Junji Iizuka for his contributions in the initial step of this study. This work was supported in part by research grants from the Japan Agency for Medical Research and Development (AMED, grant no. JP17fk0410203h0302 and JP18fk04100013) and a Grant-in-Aid for Scientific Research from Japan Society for the Promotion of Science (JSPS, grant no. 15K08328). This work was also supported by a supporting program for women researchers from the Tokyo Medical and Dental University (TMDU).

References

- 1) Radosavljevic, M., Cuillerier, B., Wilson, M.J., Clement, O., Wicker, S., Gilfillan, S. *et al.* (2001) A cluster of ten novel MHC class I related genes on human chromosome 6q24.2-q25.3. *Genomics* **79**, 114–123.
- 2) Chalupny, N.J., Sutherland, C.L., Lawrence, W.A., Rein-Weston, A. and Cosman, D. (2003) ULBP4 is a novel ligand for human NKG2D. *Biochem. Biophys. Res. Commun.* **305**, 129–135.
- 3) Eagle, R.A., Flack, G., Warford, A., Martinez-Borra, J., Jafferji, I., Traherne, J.A. *et al.* (2009) Cellular expression, trafficking, and function of two isoforms of human ULBP5/RAET1G. *PLoS One* **4**, e4503.
- 4) Eagle, R.A., Traherne, J.A., Hair, J.R., Jafferji, I. and Trowsdale, J. (2009) ULBP6/RAET1L is an additional human NKG2D ligand. *Eur. J. Immunol.* **39**, 3207–3216.
- 5) Bauer, S., Groh, V., Wu, J., Steinle, A., Phillips, J.H., Lanier, L.L. *et al.* (1999) Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science* **285**, 727–729.
- 6) Cosman, D., Mullberg, J., Sutherland, C.L., Chin, W., Armitage, R., Fanslow, R. *et al.* (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.
- 7) Bacon, L., Eagle, R.A., Meyer, M., Easom, N., Young, N.T. and Trowsdale, J. (2004) Two human ULBP/RAET1 molecules with transmembrane region are ligands for NKG2D. *J. Immunol.* **173**, 1078–1084.
- 8) Wu, J., Song, Y., Bakker, A.B.H., Bauer, S., Spies, T., Lanier, L.L. *et al.* (1999) An activating immunoreceptor complex formed by NKG2D and DAP10. *Science* **285**, 730–732.
- 9) Raulet, D.H. (2003) Roles of the NKG2D immunoreceptor and its ligands. *Nat. Rev. Immunol.* **3**, 781–790.
- 10) Pende, D., Rivera, P., Marcenaro, S., Chang, C.C., Biassoni, R., Conte, R. *et al.* (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.
- 11) Eagle, R.A., Traherne, J.A., Ashiru, O., Wills, M.R. and Trowsdale, J. (2006) Regulation of NKG2D ligand gene expression. *Hum. Immunol.* **67**, 1159–1169.
- 12) Pappworth, I.Y., Wang, E.C. and 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.* **81**, 474–482.
- 13) Ward, J., Bonaparte, M., Sacks, J., Guterman, J., Fogli, M., Mavilio, D. *et al.* (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.
- 14) Romphruk, A.V., Romphruk, A., Naruse, T.K., Raroengjai, S., Puapairoj, C., Inoko, H. *et al.* (2009) Polymorphisms of NKG2D ligands: Diverse RAET1/ULBP genes in Northeastern Thais. *Immunogenetics* **61**, 611–617.
- 15) Antoun, A., Jobson, S., Cook, M., O'Callaghan, C.A., Moss, P. and Briggs, D.C. (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.
- 16) Gibbs, R.A., Rogers, J., Katze, M.G., Bumgarner, R., Weinstock, G.M., Mardis, E.R. *et al.* (2007) Evolutionary and biomedical insights from the rhesus macaque genome. *Science* **316**, 222–234.
- 17) Kulski, J.K., Anzai, T., Shiina, T. and 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.
- 18) Otting, N., Otting, N., de Vos-Rouweler, A.J.M., Heijmans, C.M.C., de Groot, N.G., Doxiadis, G.G.M. *et al.* (2007) MHC class I A region diversity and polymorphism in macaque species. *Immunogenetics* **59**, 367–375.
- 19) Naruse, T.K., Chen, Z., Yanagida, R., Yamashita, T., Saito, Y., Mori, K. *et al.* (2010) Diversity of MHC class I genes in Burmese-origin rhesus macaque. *Immunogenetics* **62**, 601–611.
- 20) Saito, Y., Naruse, T.K., Akari, H., Matano, T. and Kimura, A. (2012) Diversity of MHC class I haplotypes in cynomolgus macaques. *Immunogenetics* **64**, 131–141.
- 21) Naruse, T.K., Okuda, Y., Mori, K., Akari, H., Matano, T. and Kimura, A. (2011) *ULBP4/*

- RAET1E* is highly polymorphic in the Old World monkey. Immunogenetics **63**, 501–509.
- 22) Naruse, T.K., Akari, H., Matano, T. and Kimura, A. (2014) Divergence and diversity of *ULBP2* genes in rhesus and cynomolgus macaques. Immunogenetics **66**, 161–170.
- 23) Kondo, M., Maruoka, T., Otsuka, N., Kasamatsu, J., Fugo, K., Hanzawa, N. *et al.* (2010) Comparative genomic analysis of mammalian NKG2D ligand family genes provides insights into their origin and evolution. Immunogenetics **62**, 441–450.
- 24) Wittenbrink, M., Spreu, J. and Steinle, A. (2009) Differential NKG2D binding to highly related human NKG2D ligands ULBP2 and RAET1G is determined by a single amino acid in the $\alpha 2$ domain. Eur. J. Immunol. **39**, 1642–1651.
- 25) Radaev, S., Rostro, B., Brooks, A.G., Colonna, M. and Sun, P.D. (2001) Conformational plasticity revealed by the cocrystal structure of NKG2D and its class I MHC-like ligand ULBP3. Immunity **5**, 1039–1049.

(Received Sep. 20, 2017; accepted Oct. 10, 2018)

Supplementary Fig. 1. Phylogenetic tree of alleles of *ULBP5* and other *ULBP* genes among Old-World monkeys and hominines. The tree was constructed using the neighbor-joining method with bootstrap values of 5,000 replications. In this analysis, only the sequences of exon 2 and exon 3 were used. The bootstrap values are indicated as percentages, but those less than 50% are not shown. There were several alleles with the same nucleotide sequences in the analyzed region, which are grouped into 'alleles'. Truncation alleles are underlined. The *ULBP1* to *ULBP5* sequences from rhesus; LOC694216, LOC694341 (*ULBP1*), LOC694466, LOC694600 (*ULBP2*), LOC694135, LOC694525 (*ULBP3*), LOC695031 (*ULBP4*), LOC694076, LOC694265 (*ULBP5*), and AM418584 (*MICAB*), and *ULBP1* to *ULBP6* sequences from humans; GenBank accession numbers AL355497 (*ULBP1*, *ULBP3*, *ULBP6*), AL583835 (*ULBP2* and *ULBP5*) and AL355312 (*ULBP4*). *Cynomolgus* (LOC102142491 as *RAET1L*), human (NG034139 as *MICA*, NG21405 as *MICB* and AB003618 as *MICB* variant, *MICB**0107N), chimpanzee (LOC463066 as *ULBP2*, LOC748488 as *ULBP5* and NC 006473 as *MICB*), western gorilla (LOC101140391 as *ULBP2*), olive baboon (LOC101017701 as *ULBP2*), marmoset (LOC100390841 as *MICA*), and mouse (AC159330 as *Rae1-α*).

Supplementary Table 1. List of *ULBP5.1* and *ULBP5.2* allele-pairs in rhesus and cynomolgus macaques, which differed only by synonymous or intron polymorphisms

Gene	Species	Allele pair	Location and numbers of polymorphic sites					Remarks	
			Intron 1	(CCT)n in Intron 1	Exon 2	Intron 2	Exon 3		Intron 3
ULBP5.1	Macaca mulatta	Mamu-ULBP5.1*02.01.01 and Mamu-ULBP5.1*02.02.01					1		
		Mamu-ULBP5.1*02.01.01 and Mamu-ULBP5.1*02.02.02	1				1		
		Mamu-ULBP5.1*02.02.01 and Mamu-ULBP5.1*02.02.02	1						
		Mamu-ULBP5.1*05.01.01, Mamu-ULBP5.1*05.01.02, and Mamu-ULBP5.1*05.01.03		1					
		Mamu-ULBP5.1*05.01.03 and Mamu-ULBP5.1*05.01.04				1			
		Mamu-ULBP5.1*05.01.04 and Mamu-ULBP5.1*05.01.05	2			1			
		Mamu-ULBP5.1*05.01.05 and Mamu-ULBP5.1*05.01.06	1			1			
		Mamu-ULBP5.1*06.01.01 and Mamu-ULBP5.1*06.01.02	1			1			
		Mafa-ULBP5.1*03.01.01 and Mafa-ULBP5.1*03.01.02		1					
		Mafa-ULBP5.1*08.01.01 and Mafa-ULBP5.1*08.01.02		1		1			
		Macaca fascicularis	Mamu-ULBP5.2*01.01.01 and Mamu-ULBP5.2*01.01.02	3					
			Mamu-ULBP5.2*05.01.01 and Mamu-ULBP5.2*05.02.01			1 (truncated)			Truncation mutation
			Mafa-ULBP5.2*02.01.01 and Mafa-ULBP5.2*02.01.02						1
			Mafa-ULBP5.2*07.01.01 and Mafa-ULBP5.2*07.01.02						1
	Mafa-ULBP5.2*07.01.01 and Mafa-ULBP5.2*07.02.01				1				
Macaca fascicularis	Mafa-ULBP5.2*07.01.02 and Mafa-ULBP5.2*07.02.01			1			1		