Immunogenetics (2009) 61:635–648
DOI 10.1007/s00251-009-0390-y

ORIGINAL PAPER

MHC class I A loci polymorphism and diversity in three Southeast Asian populations of cynomolgus macaque

Yuki F. Kita · Kazuyoshi Hosomichi · Sakae Kohara · Yasushi Itoh · Kazumasa Ogasawara · Hideaki Tsuchiya · Ryuzo Torii · Hidetoshi Inoko · Antoine Blancher · Jerzy K. Kulski · Takashi Shiina

Received: 21 April 2009 / Accepted: 22 July 2009 / Published online: 1 August 2009

Abstract Cynomolgus macaques (*Macaca fascicularis*, *Mafa*) have emerged as important animal models for biomedical research, necessitating a more extensive characterization of their *major histocompatibility complex* polymorphic regions. The current information on the polymorphism or diversity of the polygenetic *Mafa class I A* loci is limited in comparison to the more commonly studied rhesus macaque *Mafa class I A* loci. Therefore, in this paper, to better elucidate the degree and types of polymorphisms and genetic differences of *Mafa-A1* among three native Southeast Asian populations (Indonesian, Vietnamese, and Filipino) and to investigate how the allele differences between macaques and humans might have evolved to affect their respective immune responses, we identified 83 *Mafa-A* loci-derived alleles by DNA sequencing of which 66 are newly described. Most alleles are unique to each population, but seven of the most frequent alleles were identical in sequence to some alleles in other macaque species. We also revealed (1) the large and dynamic genetic and structural differences and similarities in allelic variation by analyzing the population allele frequencies, Hardy-Weinberg's equilibrium, heterozygosity, nucleotide diversity profiles, and phylogeny, (2) the difference in genetic structure of populations by Wright's FST statistic and hierarchical analysis of molecular variance, and (3) the different demographic and selection pressures on the three populations by performing Tajima's *D* test of neutrality. The large level of diversity and polymorphism at the *Mafa-A1* was less evident in the Filipino than in the Vietnam or the Indonesian populations, which may have important implications in animal capture, selection, and breeding for medical research.

Electronic supplementary material The online version of this article (doi:10.1007/s00251-009-0390-y) contains supplementary material, which is available to authorized users.

Y. F. Kita · K. Hosomichi · H. Inoko · J. K. Kulski · T. Shiina (✉)
Department of Molecular Life Science,
Division of Basic Medical Science and Molecular Medicine,
Tokai University School of Medicine,
143 Shimokasuya,
Isehara, Kanagawa 259-1143, Japan
e-mail: tshiina@is.icc.u-tokai.ac.jp

S. Kohara
Shin Nippon Biomedical Laboratories, Ltd.,
16-1 Minamiakasaka,
Kainan, Wakayama 642-0017, Japan

Y. Itoh · K. Ogasawara
Department of Pathology, School of Medicine,
Shiga University of Medical Science,
Seta Tsukinowa-cho,
Ōtsu, Shiga 520-2192, Japan

H. Tsuchiya · R. Torii
Research Center for Animal Life Science, School of Medicine,
Shiga University of Medical Science,
Seta Tsukinowa-cho,
Ōtsu, Shiga 520-2192, Japan

A. Blancher
Laboratoire d'Immunogenetique moleculaire,
Universite Paul Sabatier, Faculte de Medecine de Rangueil,
Batiment A2, 133, Route de Narbonne,
31062 Toulouse, France

J. K. Kulski
Center for Forensic Science,
The University of Western Australia, Nedlands,
Perth 6009 Western Australia, Australia

Keywords Cynomolgus macaque · MHC · Allele · Diversity · Population genetics

Introduction

Cynomolgus macaques (*Macaca fascicularis*, *Mafa*), alias the crab-eating monkeys or long-tailed macaques, are widely dispersed in nature, and they live in a vast range of Southeast Asia including the Philippines (Phi), Indonesia (Ind), Vietnam (Vie), Malaysia, Thailand, Cambodia, Mauritius, and Brunei. This species breeds seasonally in captivity and is smaller and easier to manage and more economical to maintain than the more commonly studied rhesus macaque species, which is now difficult to access for breeding and medical research purposes because of bans and restrictions on exportation and trading (Southwick and Siddiqi 1994). Therefore, like the rhesus macaque and because of their close genetic similarity, the cynomolgus macaques are often used for biomedical research into infectious diseases including AIDS, tuberculosis, and severe acute respiratory syndrome, neurological diseases including Alzheimer's disease and Parkinson's disease, reproduction, regenerative medicine, transplantation, and immunotherapy (Capuano et al. 2003; Conlee et al. 2004; Lawler et al. 2006; Emborg 2007; Shively et al. 2007; Wang et al. 2007; Wiseman et al. 2007; Liu et al. 2007; Wiseman and O'Connor 2007).

In order to effectively use the cynomolgus macaques for medical research, it is necessary to better understand the genetic diversities between and within the different populations and their association with infections and various acute and chronic diseases. The highly polymorphic *major histocompatibility complex* (*MHC*) region that encodes the *MHC* transplantation and immunoregulatory molecules is one of the medically important genomic regions that warrants special attention for genetic investigation because it has been implicated in the generation or mediation of many hundreds of infectious and/or autoimmune diseases (Shiina et al. 2009). The primate *MHC* molecules play an important role in the immune response, and the class I molecules are expressed on all nucleated cells and present peptides of intracellular origin to CD8+ cells, whereas the class II molecules are expressed on immune cells, B cells, helper T cells or CD4+ T cells, macrophages, and other antigen-presenting cells. Of the human class I loci, *HLA-A*, *HLA-B*, and *HLA-C* loci have 2,384 alleles in total, many of which have been implicated in disease resistance and/or susceptibility (IMGT/HLA Database, http://www.ebi.ac.uk/imgt/hla/; Shiina et al. 2004, 2009). On the other hand, only 140 cynomolgus macaque *MHC* (*Mafa*) class I *A* (*Mafa-A*) and 35 class I *B* (*Mafa-B*) alleles have been identified so far, although numerous numbers of polygenic variations are observed in the *Mafa* class I region (Uda et al. 2004, 2005; Krebs et al. 2005; Otting et al. 2007; Wiseman et al. 2007; Pendley et al. 2008). Since 88 novel *Mafa-A* alleles were identified from only 123 cynomolgus macaques that were investigated recently (Otting et al. 2007; Pendley et al. 2008; Campbell et al. 2009) and there are at least seven possible *Mafa-A* loci combinations (Otting et al. 2007), it is highly likely that comprehensive allele information is still lacking for the *Mafa-A* in the cynomolgus macaques and that many more new alleles will be identified in future from different population groups.

Extensive macaque *Mafa* polymorphisms are evident among geographically separated macaque populations (Kondo et al. 1993; Leuchte et al. 2004; Sano et al. 2006; Bonhomme et al. 2007; Smith et al. 2007). In addition, the *Mafa* and rhesus macaque (*Macaca mulatta*, *Mamu*) *MHC* class I alleles that originated from different geographic populations have a small subset of allele sharing or transspecies polymorphisms (van Oosterhout 2009), although most of the alleles are unique to the species and the population (Krebs et al. 2005; Otting et al. 2007; Karl et al. 2008). Therefore, a detailed comparison of the *Mafa-A* allele sequences among three South Asian populations such as Phi, Ind, and Vie may lead to a better understanding of allele population distribution and the biogeography of the cynomolgus macaques.

In this paper, we determined the number and types of polymorphisms and genetic differences at the *Mafa-A* loci for three populations (Phi, Ind, and Vie). To investigate how the allele differences between cynomolgus macaques and humans might affect their respective immune responses, we identified 66 novel alleles within the *Mafa-A* loci using a total of 83 nonrelated individuals and five families. We analyzed the genetic differences among the *Mafa* populations and between the macaque and human by performing both population genetic and amino acid sequence analyses using the coding regions (exon 1 to exon 7) of the complementary DNA (cDNA) nucleotide sequences that we obtained from the peripheral white blood cells of the macaques in this study and human DNA from previous studies.

Materials and methods

Animals

A total of 88 cynomolgus macaques were used for this study with five infants bred and provided by the Shiga University of Medical Science (Shiga, Japan) and 83 unrelated individuals obtained from the Philippines (28 individuals), Indonesia (27 individuals), and Vietnam (28 individuals) and imported into Japan from INA Research

Philippines INC (Laguna, Philippines), CV Universal Fauna Breeder and Exporter of Nonhuman Primates for Laboratories (Jakarta, Indonesia), and Nafovanny (Dong Nai Province, Vietnam). Before starting this study, we first confirmed that the macaques came from the expected export origins by sequencing and phylogenetic analyses of mtDNA D-loop region (data not shown; Blancher et al. 2008). The blood collection and animal studies were conducted in accordance with the Guidelines for Animal Experiments at the Shiga University of Medical Science, Shiga Japan.

Human DNA samples

A reference set of 30 Japanese DNA samples genotyped for *HLA* alleles at the *HLA-A*, *HLA-B*, and *HLA-DR* loci by DNA sequencing was obtained from the Department of Legal Medicine, Shinshu University School of Medicine, Matsumoto, Nagano, Japan. This reference set of DNA samples represents a Japanese population of registered donors from the Nagano region in the Japanese unrelated bone marrow donor registry (Moriyama et al. 2006). A reference set of 30 Australian–Caucasian DNA samples genotyped for *HLA* alleles at the *HLA* class I gene loci by DNA sequencing was obtained from The Department of Clinical Immunology and Biochemical Genetics, Royal Perth Hospital, Perth, Western Australia (Kulski et al. 2008). This reference set of samples represents a predominantly Caucasian (99.6%) population from the seaside town of Busselton in Western Australia (http://www.busseltonhealthstudy.com/).

RT-PCR amplification of the *Mafa-A* genes

Total RNA was directly isolated from the 88 individual peripheral white blood cells using the TRIzol reagent (Invitrogen, CA, USA). cDNA was synthesized by oligo d (T) primer using the ReverTra Ace for reverse-transcriptase (RT) reaction (TOYOBO, Japan). A new set of *Mafa-A* gene-specific primers was designed, incorporating sequences from exon 1 and exon 7 of the gene, and was used for RT-polymerase chain reaction (PCR) amplification with the sense primer (Mafa-A_F1.1: 5′-AACCCTCCTCCTGGTG CTCT-3′ and Mafa-A_F1.2: 5′-AACCCTCCTCCT GCTGCTCT-3′) and the antisense primer (Mafa-A_R: 5′-CCTGGGCACTGTCACTGCTT-3′). In brief, the 20-μl amplification reaction contained 10 ng of cDNA, 1.0 unit of TaKaRa LA Taq™ polymerase (TaKaRa Shuzo, Japan), 1× PCR buffer, 2.5 mM $MgCl_2$, 400 μM of each dNTP, and 0.5 μM of each primer. The cycling parameters were as follows: an initial denaturation of 98°C/1 min followed by 30 cycles of 98°C/10 s and 68°C/1 min. Alternatively, the 20-μl amplification reaction contained 10 ng of cDNA, 0.4 units of KOD FX polymerase (TOYOBO, Japan), 1× PCR buffer, 2 mM of each dNTP, and 0.5 μM of each primer. The cycling parameters were as follows: an initial denaturation of 94°C/2 min followed by 30 cycles of 98°C/10 s and 68°C/1 min. PCR reactions were performed by using the thermal cycler GeneAmp PCR system 9700 (Applied Biosystems, CA, USA).

Nucleotide sequencing

RT-PCR products of the 88 cynomolgus macaques were cloned into the pGEM-T Easy vector and TArget vector with the TA cloning kit according to the protocol provided by the manufacturer (Promega, Madison, WI, USA, or TOYOBO, Japan) and sequenced by using the ABI3130 genetic analyzer (Applied Biosystems, CA, USA) in accordance with the protocol of Big Dye terminator method. To avoid PCR and sequencing artifacts generated by polymerase errors, 32 clones per individual were sequenced. The nucleotide sequences of all individuals were also determined by direct sequencing of the RT-PCR products using PCR primers as sequencing primers. Allele types were determined by comparing them with known *Mafa-A* and *Mamu-A* allele sequences in the GenBank/European Molecular Biology Laboratory (EMBL)/DNA Databank of Japan (DDBJ) databases.

HLA-A genotyping

The Japanese and Australian–Caucasian DNA samples were previously genotyped for *HLA-A* alleles to two or four digits by direct sequencing (Moriyama et al. 2006; Kulski et al. 2008).

Sequence analyses

The sequences were analyzed using the GENETYX software (Software Development Co. Ltd., Japan). Nucleotide similarities between sequences were calculated by Sequencher 4.1 (Gene Codes Co., MI, USA) and BLAST (http://www.ncbi.nlm.nih.gov/BLAST/). Allele numbers, allele frequencies, and heterozygosities were calculated using the software program CERVUS 3.0 (http://en.bio-soft.net/other/CERVUS.html). The allele diversity percentage per population was calculated as the number of alleles (nA) times 100% divided by the number of individuals in the population. Polymorphic information content was calculated using the online software Probability Calculator for Paternity Likelihood and Exclusion Ver. 2.5 at http://www3.kmu.ac.jp/legalmed/DNA/ppl3.html. Hardy-Weinberg's equilibrium was calculated using ARLEQUIN Ver. 3.11 (http://cmpg.unibe.ch/software/arlequin3/, Excoffier et al. 2005). Multiple sequence alignments were created using the ClustalW

Sequence Alignment program of the Molecular Evolution Genetics Analysis software (MEGA4.1, http://www.megasoftware.net/, Tamura et al. 2007); the phylogenetic trees were constructed by the neighbor-joining method provided in the MEGA4.1 software and Neighbor-Net (Bryant and Moulton 2004), as implemented in SplitsTree4 (Huson and Bryant 2006) using uncorrected *p* distances. Synonymous and nonsynonymous substitution rates (dN/dS) were calculated by the modified Nei and Gojobori method (Nei and Gojobori 1986) with the *p* distance parameter in the MEGA4.1 software. Proportion of variation attributable to differences among the populations was estimated by Wright's FST statistic (Wright 1965), and the amount of population genetic structure according to different hierarchical analyses of molecular variance (AMOVA) was evaluated using the ARLEQUIN 3.1 software (Excoffier et al. 2005) with 100 random permutations for the significance tests. Nucleotide diversity profile and neutrality test by Tajima's *D* test (Tajima 1989) were calculated using DnaSP version 4.90 (http://www.ub.edu/dnasp/, Rozas and Rozas 1999).

Nomenclature

Novel *Mafa-A* allele sequences were submitted first to the GenBank/EMBL/DDBJ databases for assignment of accession numbers and then to the IMGT/MHC-NHP database for naming of new alleles according to the nomenclature proposal guidelines currently in use (Robinson et al. 2003; Otting et al. 2005). An example of the allele nomenclature is as follows: *Mafa-A1*01002* is the *MHC* allele of the cynomolgus macaque (*M. fascicularis*, *Mafa*) encoded by the class I locus *A1*. The first three digits after the asterisk define the lineage *010*, whereas the fourth to fifth digits define the allele number *02*. These allele numbers are arbitrary and were numbered in the order in which they were identified. A sixth and seventh digit may be used to describe a synonymous base pair difference between two sequences. The *Mafa-A* alleles identified in this paper and given numbers at the five- or seven-digit level are summarized in Table 1 and Supplementary Fig. 1.

Results

Allele numbers of *Mafa-A* and transspecies polymorphisms

A total of 83 distinct *Mafa-A* alleles at the five- to seven-digit level of classification was identified using RT-PCR products from 83 unrelated macaques from the Vie, Ind, and Phi populations by using the subcloning sequencing method and direct-sequencing method (Table 1). The newly designated *Mafa-A* PCR primer set had extremely high *Mafa-A* specificity and detected only the *Mafa-A* loci sequences without detection of the other *Mafa* class I sequences such as *Mafa-B*, *Mafa-AG*, *Mafa-E*, and *Mafa-F*. The number of *Mafa-A* alleles in the 28 Vie (36 alleles) and 27 Ind (34 alleles) was more than twice the number detected in the 28 Phi (16 alleles) macaques. Three of these alleles were shared between the Ind and Vie populations.

Of the 83 distinct *Mafa-A* alleles, 17 were previously reported in GenBank/EMBL/DDBJ databases while the other 66 alleles at the five- to seven-digit level of classification are new. All of the macaques had one or two transcribed *Mafa-A* alleles except for three individuals from the Phi population (CE0299M, CE0316F, and CE0321F) who each had three *Mafa-A1* alleles defined by the IPD-MHC Nomenclature Committee (Supplementary Fig. 1). One of these alleles, *Mafa-A1*07402*, was shared between the three individuals, and the subclone numbers derived for this allele were fewer than for the other *Mafa-A1* alleles (data not shown). Therefore, the *Mafa-A1*07402* allele is thought to be a novel *Mafa-A* locus generated by either a genetic recombination or a gene duplication event in Phi population in relatively recent times.

Seven alleles (*Mafa-A1*00102*, *Mafa-A1*00402*, *Mafa-A1*00403*, *Mafa-A1*01903*, *A1*06504*, *Mafa-A1*8903*, and *Mafa-A2*2401*) were perfectly matched with previously reported alleles of the rhesus macaque (*M. mulatta*, *Mamu*) *Mamu-A* and the southern pig-tailed macaque (*Macaca nemestrina*, *Mane*) *Mane-A* genes (Table 1). These transspecies polymorphisms were not only observed in the Vie population but also in the Ind and Phi populations, which are separated from each other by vast expanses of ocean, suggesting that the transspecies *MHC-A* alleles were generated before speciation of cynomolgus macaques 1.8–2.0 million years ago (Hayasaka et al. 1996). Although the possibility of generating the occasional coincidental allelic matching between species after speciation cannot be ignored, it is noteworthy that whereas seven *Mafa* alleles were shared between different macaque species only three *Mafa* alleles, *Mafa-A1*01002*, *Mafa-A1*01005*, and *Mafa-A1*0630302*, were shared between the same species from three different populations (Table 1), suggesting little gene flow between the populations. In this regard, the seven *Mafa-A1* transspecies polymorphisms support the many previous reports of shared *Mafa-A1* alleles between different macaque species (Krebs et al. 2005; Otting et al. 2007; Karl et al. 2008).

Correlation between *Mafa-A* alleles and loci

Of the 83 *Mafa-A* alleles, 76 were linked to the *Mafa-A1* locus, three to *Mafa-A2*, four to *Mafa-A3*, one to *Mafa-A4*, and one to *Mafa-A5* (Table 1). Therefore, our findings are consistent to the previous reports that most *Mafa-A* alleles are expressed by the *Mafa-A1* locus (Pendley et al. 2008;

Table 1 *Mafa-A* sequence frequencies identified in three different population groups

Official name	Accession numbers	Macaque counterpart	Total (n=83)		Vietnam (n=28)		Indonesia (n=27)		Philippine (n=28)	
			Seq. num.	Seq. freq.	Seq. num.	Seq. freq.	Seq. num.	Seq. freq.	Seq. num.	Seq. Freq.
*Mafa-A1*00102*	AY958101	*Mamu-A1*1001*	2	0.012	2	0.036				
*Mafa-A1*00104*	AB447618		1	0.006			1	0.019		
*Mafa-A1*00202*	AB447609		1	0.006	1	0.018				
*Mafa-A1*00401*	AB447581		6	0.036					6	0.102
*Mafa-A1*00402*	AB448751	*Mamu-A*04*	2	0.012	2	0.036				
*Mafa-A1*00403*	AB448750	*Mamu-A1*0403*	5	0.030			5	0.093		
*Mafa-A1*00404*	AB447575		1	0.006			1	0.019		
*Mafa-A1*00703*	AB447561		1	0.006	1	0.018				
*Mafa-A1*00704*	AB447578		2	0.012			2	0.037		
*Mafa-A1*00802*	AB447588		6	0.036					6	0.102
*Mafa-A1*01002*	EU203687		3	0.018	1	0.018	2	0.037		
*Mafa-A1*0100202*	AB447601		1	0.006			1	0.019		
*Mafa-A1*01005*	EU203707		2	0.012	1	0.018	1	0.019		
*Mafa-A1*01006*	EU203699		1	0.006			1	0.019		
*Mafa-A1*01103*	AB447595		1	0.006	1	0.018				
*Mafa-A1*01501*	AY958087		3	0.018	3	0.054				
*Mafa-A1*01803*	EU203709		2	0.012			2	0.037		
*Mafa-A1*01804*	AB447598		3	0.018					3	0.051
*Mafa-A1*01902*	AB447567		1	0.006			1	0.019		
*Mafa-A1*01903*	AB448747	*Mane-A*12*	1	0.006	1	0.018				
*Mafa-A1*01904*	AB447582		1	0.006			1	0.019		
*Mafa-A1*01905*	AB447616		1	0.006			1	0.019		
*Mafa-A1*02203*	AB447602		1	0.006	1	0.018				
*Mafa-A1*02703*	AB447615		1	0.006	1	0.018				
*Mafa-A1*03203*	AB447566		1	0.006	1	0.018				
*Mafa-A1*03204*	AB447599		1	0.006	1	0.018				
*Mafa-A1*03801*	AY958114		1	0.006	1	0.018				
*Mafa-A1*03802*	AB447619		1	0.006					1	0.017
*Mafa-A1*04101*	AY958116		4	0.024	4	0.071				
*Mafa-A1*04103*	AY958118		2	0.012			2	0.037		
*Mafa-A1*04201*	AY958117		1	0.006	1	0.018				
*Mafa-A1*04301*	AY958118		1	0.006	1	0.018				
*Mafa-A1*04303*	AB447579		1	0.006			1	0.019		
*mafa-A1*05202*	AM943361		3	0.018					3	0.051
*Mafa-A1*05301*	AB447562		4	0.024	4	0.071				
*Mafa-A1*05401*	AB154771		2	0.012	2	0.036				
*Mafa-A1*05902*	AB447620		1	0.006			1	0.019		
*Mafa-A1*06003*	EU203698		2	0.012			2	0.037		
*Mafa-A1*06004*	AB447574		2	0.012			2	0.037		
*Mafa-A1*06102*	AB447617		1	0.006			1	0.019		
*Mafa-A1*06203*	AB447587		1	0.006			1	0.019		
*Mafa-A1*0630302*	AB447592		2	0.012	1	0.018	1	0.019		
*Mafa-A1*06403*	AB447608		1	0.006	1	0.018				
*Mafa-A1*06502*	AB447611		1	0.006	1	0.018				
*Mafa-A1*06503*	AB447614		1	0.006	1	0.018				
*Mafa-A1*06504*	AB448748	*Mamu-A1*6501*	6	0.036	6	0.107				

Table 1 (continued)

Official name	Accession numbers	Macaque counterpart	Total (n=83)		Vietnam (n=28)		Indonesia (n=27)		Philippine (n=28)	
			Seq. num.	Seq. freq.	Seq. num.	Seq. freq.	Seq. num.	Seq. freq.	Seq. num.	Seq. Freq.
*Mafa-A1*06603*	EU203712		4	0.024			4	0.074		
*Mafa-A1*06604*	AB447564		1	0.006	1	0.018				
*Mafa-A1*06605*	AB447589		1	0.006			1	0.019		
*Mafa-A1*07001*	EU203708		2	0.012	2	0.036				
*Mafa-A1*07401*	AB447596		1	0.006	1	0.018				
*Mafa-A1*07402*	AB447606		3						3	0.051
*Mafa-A1*07602*	AB447571		1	0.006			1	0.019		
*Mafa-A1*0770102*	AB447613		2	0.012			2	0.037		
*Mafa-A1*07902*	AB447568		3	0.018			3	0.056		
*Mafa-A1*08501*	AB447577		1	0.006			1	0.019		
*Mafa-A1*08702*	AB447593		2	0.012	2	0.036				
*Mafa-A1*08802*	AB447572		2	0.012			2	0.037		
*Mafa-A1*08903*	AM943360	*Mamu-A1*8903*	19	0.115					19	0.322
*Mafa-A1*08906*	AB447610		1	0.006	1	0.018				
*Mafa-A1*08907*	AB474960		1	0.006					1	0.017
*Mafa-A1*09201*	AB447594		2	0.012	2	0.036				
*Mafa-A1*09301*	AB447597		2	0.012					2	0.034
*Mafa-A1*09302*	AB447605		1	0.006					1	0.017
*Mafa-A1*09401*	AB447584		7	0.042					7	0.119
*Mafa-A1*09402*	AB447570		2	0.012			2	0.037		
*Mafa-A1*09501*	AB447565		1	0.006	1	0.018				
*Mafa-A1*09601*	AB447569		2	0.012	2	0.036				
*Mafa-A1*09701*	AB447576		3	0.018			3	0.056		
*Mafa-A1*09801*	AB447580		1	0.006					1	0.017
*Mafa-A1*09901*	AB447586		1	0.006			1	0.019		
*Mafa-A1*10001*	AB447591		1	0.006	1	0.018				
*Mafa-A1*10101*	AB447600		1	0.006			1	0.019		
*Mafa-A1*10102*	AB447607		1	0.006	1	0.018				
*Mafa-A1*10201*	AB447603		1	0.006					1	0.017
*Mafa-A1*10301*	AB447612		1	0.006			1	0.019		
*Mafa-A2*2401*	AB448752	*Mamu-A*24*	1	0.006	1	0.018				
*Mafa-A2*2402*	AB447585		2	0.012					2	0.034
*Mafa-A3*1304*	AB447604		2	0.012					2	0.034
*Mafa-A3*1305*	AB447573		1	0.006			1	0.019		
*Mafa-A3*1306*	AB447563		1	0.006	1	0.018				
*Mafa-A4*0105*	AB447583		1	0.006					1	0.017
*Mafa-A5*3003*	AB447590		1	0.006			1	0.019		
Total			169	1.000	56	1.000	54	1.000	59	1.0000

Underlines indicate newly identified allele sequences

Campbell et al. 2009). In regard to the *Mafa-A* haplotype classifications (Otting et al. 2007), we identified only four multiloci *Mafa-A* allelic haplotypes in nine individuals, three *A1–A2* (CE0398F, CE0300M, and CE0314F), four *A1–A3* (CE0013F, CE197F, CE0308M, and CE0313F), one *A1–A4* (CE303M), and one *A1–A5* (CE0094F). Surprisingly, the *Mafa-A1* alleles were homozygous in the nine individuals with the *Mafa-A* allelic haplotypes because only two kinds of *Mafa-A* sequences were detected in each individual (Supplementary Fig. 1). Of the nine individuals, CE0013F in Vie, CE197F and CE0094F in Ind and CE303M, and CE0313F in Phi had the most frequent

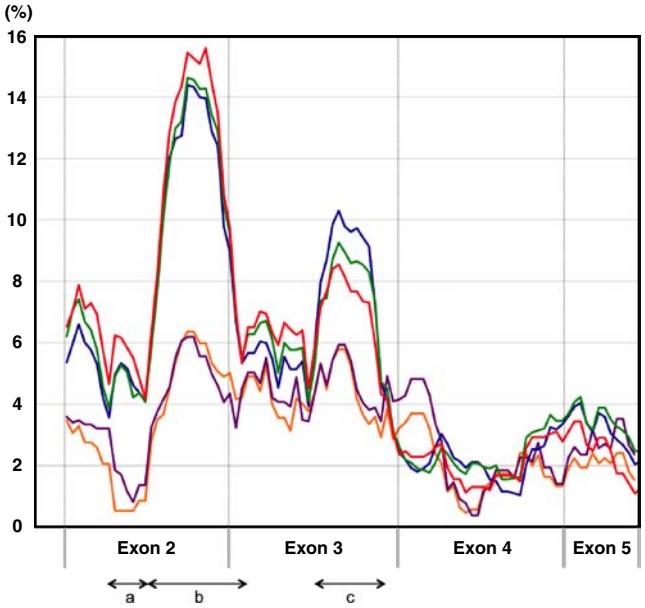

Fig. 1 Nucleotide diversity profile of three cynomolgus macaque populations and two human populations. *Blue*, *green*, and *red lines* show the three cynomolgus macaque populations Vie, Ind, and Phi, respectively. The *orange* and *purple lines* show the two human populations Japanese and Caucasian, respectively. *Vertical axis* shows nucleotide differences (%) per 100 bp. The *horizontal axis* shows the continuous nucleotide sequence length of exon 2 to exon 5 in the aligned cDNA sequences from each species and population group. The *horizontal arrows* labeled *a*, *b*, and *c* show the regions harboring the peptide-binding domains of the amino acid translated sequences

alleles, *Mafa-A1*06504*, *Mafa-A1*00403*, and *Mafa-A1*08903*, respectively, whereas the *Mafa-A2* to *Mafa-A5* alleles were not detected in all individuals with *Mafa-A1*06504*, *Mafa-A1*00403*, and *Mafa-A1*08903*.

In a family study (Table 2), the *Mafa-A* allele typing results for three families (families 1–3) were consistent with the expected allele segregations, but those for another two families (families 4 and 5) were unexpected. Namely, whereas the mother, CE0013F, of families 4 and 5 had the *Mafa-A1*06504* and *Mafa-A3*1306* alleles representing an *A1–A3* haplotype, her offspring, CE0325F in family 4 and CE0112F in family 5, had the *Mafa-A1*06504* allele but not the *Mafa-A3*1306* allele (Table 2). The reason for the absence of the *Mafa-A3*1306* allele in the offspring of CE0013F was not determined but might be due to gene mutations, deletions, rearrangements, or regulated suppression of *Mafa-A3* gene expression.

Genetic and statistical features of the *Mafa-A1* locus

On the basis of the results described in the previous sections, we excluded the nine individuals with the seven *Mafa-A2* to *Mafa-A5* alleles (*Mafa-A2*2401*, *Mafa-A2*2402*, *Mafa-A3*1304*, *Mafa-A3*1305*, *Mafa-A3*1306*, *Mafa-A4*0105*, and *Mafa-A5*3003*) and *Mafa-A1*07402* from the data set of the 83 macaques and only analyzed the remaining 75 *Mafa-A1* sequences in the following studies. Of the 75 *Mafa-A1* allele sequences, there were novel alleles for 26 of the 34 Vie sequences, 26 of the 34 Ind sequences, and 13 of the 16 Phi sequences (Table 1). The most frequent allele shown in Table 1 is *Mafa-A1*08903*, and it was found in 19 Filipino macaques at a frequency of 0.322. This allele like the other most frequent alleles, *Mafa-A1*06504* in Vie (allelic frequency 0.107) and *Mafa-A1*00403* in Ind (0.093), shared identity with *Mamu-A* and *Mane-A* counterparts as transspecies polymorphisms. Therefore, most *Mafa-A1* alleles in the three populations probably evolved as new alleles by positive selection and continuing adaptation to environmental pathogens with a small subset inherited as transspecies polymorphisms (Krebs et al. 2005; Otting et al. 2007; Karl et al. 2008).

Genetic structure, nucleotide diversity profiles, and phylogenetic analyses of the *Mafa-A1* locus

In order to elucidate the differences in the genetic structure of the *Mafa-A1* polymorphism in the three populations, we estimated the proportion of variation attributable to differences among the populations by Wright's FST statistic (Wright 1965) and different hierarchical AMOVA to evaluate the amount of population genetic structure by ARLEQUIN 3.1 (Excoffier et al. 2005) with 100 random permutations for the significance tests. From these analyses, significant differences were not observed among the three populations (P=0.99 in Vie vs Ind, P=0.99 in Vie vs Phi and P=0.99 in Ind vs Phi), although allele sharing between populations was observed for only three alleles *Mafa-A1*01002*, *Mafa-A1*01005*, and *Mafa-A1*0630302* and only between the Vie and Ind populations (Table 1). This suggests that the three populations have similar genetic structures.

To further evaluate the genetic similarity of the populations, we constructed a nucleotide diversity plot, phylogenetic trees, and phylogenetic networks of the *Mafa-A1* alleles from the three macaque populations, Ind, Phi, and Vie. The nucleotide diversity plot of the *Mafa-A1* polymorphic sequences (Fig. 1) showed similar overlapping profiles across exon 2 to exon 5 between the three macaque populations and a much greater percentage nucleotide difference for the macaque gene than the human *HLA-A* polymorphic sequences of Japanese and Australian Caucasians. The nucleotide diversity on average for the *Mafa-A* gene in the three macaque populations was about 5.25%, whereas the 3′ ends ("b" and "c" segments in Fig 1) of exon 2 and exon 3 (peptide-binding sites) showed much higher diversities at 14~16% and 8~10%, and exon 4 was the most conserved at 2%. By contrast, the nucleotide diversity on average for the *HLA-A* gene in Japanese and Caucasians

Table 2 *Mafa-A* sequence typing for five families

Familial number	Individual number	Relationship	Population	*Mafa-A* sequence	
Family 1	CE0069F	Mother	Vietnam	*Mafa-A1*06604*	***Mafa-A1*09501***
	CE0118M	Father	Vietnam	***Mafa-A1*01903***	*Mafa-A1*04101*
	CE0324F	Offspring	Vietnam	***Mafa-A1*01903***	***Mafa-A1*09501***
Family 2	CE0111F	Mother	Vietnam	*Mafa-A1*04301*	***Mafa-A1*06504***
	CE0117M	Father	Vietnam	*Mafa-A1*05301*	***Mafa-A1*01501***
	CE0204F	Offspring	Vietnam	***Mafa-A1*06504***	***Mafa-A1*01501***
Family 3	CE0108F	Mother	Indonesia	*Mafa-A1*04103*	***Mafa-A1*01803***
	CE0117M	Father	Vietnam	***Mafa-A1*05301***	*Mafa-A1*01501*
	CE0169F	Offspring	Vietnam	***Mafa-A1*05301***	***Mafa-A1*01803***
Family 4	CE0013F	Mother	Vietnam	***Mafa-A1*06504***	*Mafa-A3*1306*
	CE0118M	Father	Vietnam	***Mafa-A1*01903***	*Mafa-A1*04101*
	CE0325F	Offspring	Vietnam	***Mafa-A1*01903***	***Mafa-A1*06504***
Family 5	CE0013F	Mother	Vietnam	***Mafa-A1*06504***	*Mafa-A3*1306*
	CE0049M	Father	Vietnam	***Mafa-A1*07001***	*Mafa-A1*00703*
	CE0112F	Offspring	Vietnam	***Mafa-A1*07001***	***Mafa-A1*06504***

Bold letters indicate the inherited Mafa-A sequences from parents to infants

was about 3.2%; the 3′ ends of exon 2 and exon 3 ("b" and "c" segments) showed moderately highly diversity at 6%, and exon 4 was also well conserved at 1~2%. Overall, the nucleotide diversities of the cynomolgus macaque populations were about two times higher than the two human populations in this study, but with the "a" segment five times higher, the "b" segment 2.5 times higher, and the "c" segment 1.5 times higher in the macaque than the human.

We reconstructed phylogenetic trees and networks using the neighbor-joining method to examine the interrelationship of the nucleotide *Mafa-A1* allelic sequences obtained from the three macaque populations, Ind, Phi, and Vie, along with previously detected Mauritian (Mau) *Mafa-A1* sequences (Krebs et al. 2005). The phylogenetic analyses of the 76 aligned *Mafa-A1* nucleotide sequences using reconstructed phylogenetic trees by the NJ method showed intermingled allelic clusters across the three macaque populations, similar to the example shown in Supplementary Fig. 2. The *Mafa-A1* nucleotide sequences could be separated into four major sequence groups A, B, C, and D after divergence of human *HLA-A* sequences.

To better visualize the complexity of the *Mafa-A1* nucleotide sequence interrelationships, we used the Neighbor-Net, NJ, and Splits Tree4 methods to construct phylogenetic networks of the 80 *Mafa-A1* alleles. Figure 2 shows that the nucleotide-based phylogenetic network of the almost full-length *Mafa-A1* sequences (exon 1 to exon 7) constructed by the Neighbor-Net method formed highly complex networks of intermingled allelic clusters across the four populations rather than a simple branching tree. The boxed networks seen in the figure represent conflicting signals from a complex interrelationship between the sequences that might imply the possibility of recombination, gene conversions, or selection pressure. To show that the conflicting signals from this analysis were generated mostly by the presence of the high-diversity peptide-binding sequences in the regions of exon 2 and exon 3 of the *Mafa-A1* sequences (Fig. 1), we also reconstructed a phylogenetic network of only the *Mafa-A1* exon 2 to exon 3 nucleotide sequences, and the phylogenetic tree had similar pattern with the tree of almost complete *Mafa-A1* (data not shown). This analysis helps to confirm that the complexity of the sequence diversity within the restricted regions of peptide binding contributes to the difficulty of interpreting the evolutionary interrelationships of the *Mafa-A1* sequences from different macaque populations. Nevertheless, the phylogenetic analyses, together with the nucleotide diversity plots, also help to confirm the results of the Wright's FST statistic and AMOVA analyses that the macaque Vie, Ind, and Phi populations have similar genetic structures for the *Mafa-A1* gene.

The effect of selection and/or change in population size on *Mafa-A1* polymorphism

In order to further evaluate whether there may have been affects of selection pressure and/or demographic change on the sequence variation of the *Mafa-A1* locus in the three cynomolgus macaque populations, we performed a neutrality test by using the Tajima's *D* test on the aligned *Mafa-A1* allelic sequences for each of the populations. Figure 3 shows the results of Tajima's *D* test. The null hypothesis of Tajima's *D* test is that neutral evolution (genetic drift or neutrality) occurs in an equilibrium population, which

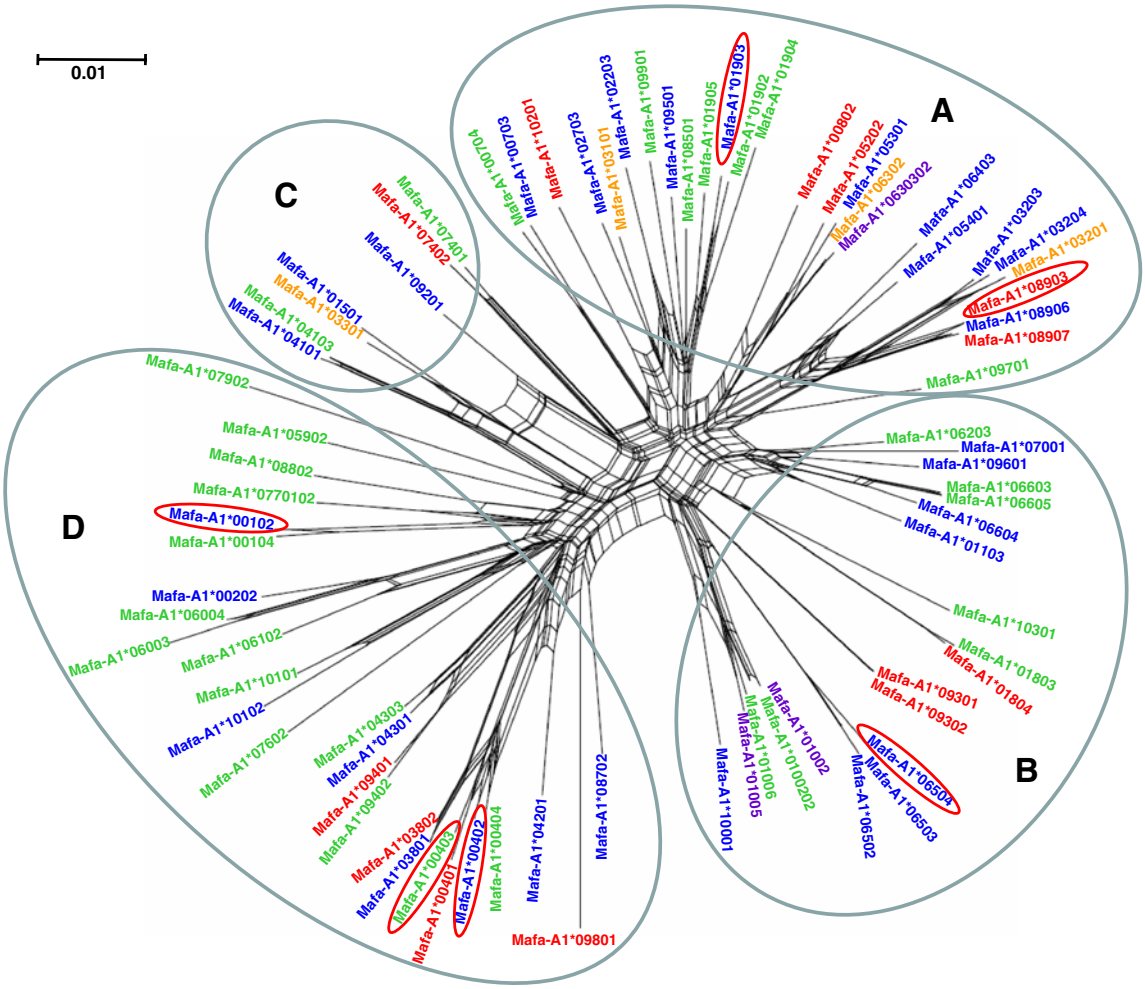

Fig. 2 Nucleotide-sequence-based phylogenetic tree of *Mafa-A1* alleles constructed by neighbor-net and splitsTree4 method using uncorrected *p* distances. *Blue*, *green*, and *red rectangles* indicate *Mafa-A1* alleles, which were identified in this study, derived from Vie, Ind, and Phi, respectively. The *orange rectangles* indicate the *Mafa-A1* alleles derived from Mau by other researchers. *Purple rectangles* indicate the *Mafa-A1* alleles shared between Vie and Ind. *Red circles* show transspecies *Mafa-A1* alleles with the other macaque species as indicated in Table 1

implies that no selection is acting at the locus and that the population has not experienced any recent growth or contraction (Tajima 1989). A negative Tajima's *D* value indicates an excess of rare alleles due to purifying (negative) selection and/or population size expansion, whereas a positive *D* value indicates an intermediate frequency of alleles due to balancing (positive) selection and/or a decrease in population size. As seen in Fig. 3, the average Tajima's *D* values were negative at −0.43 and −0.38 for the Vie and Ind macaque populations, respectively, but positive at 0.72 for the Phi macaque population. In the case of the *D* value variations across the exonic nucleotide sequences, exon 2 (α1 domain) showed a positive Tajima's *D* (*D*>0) in all populations, whereas exon 3 (α2 domain) and exon 4 (α3 domain) showed strong negative Tajima's *D* (*D*<0) values in the Vie and Ind. On the other hand, exon 3 and exon 4 showed positive Tajima's *D* (*D*>0) in the Phi. Overall, this analysis suggests that the three populations have a similar distribution of *Mafa-A1* polymorphisms that have been affected differently by natural selection such as overdominant selection on exon 2 in all populations, purifying selection on exon 3, and exon 4 in the Vie and the Ind and a recent bottleneck or founder effect on exon 3 and exon 4 in the Phi. An increase in out-group breeding was not observed in any of the three population groups.

Discussion

Mafa-A loci diversity and polymorphism

In this study, we identified 66 novel *Mafa-A* alleles expressed by different *Mafa-A* loci, of which 59 were *Mafa-A1* alleles to the five- or seven-digit level of classification. Our collection of *Mafa-A1* alleles has added

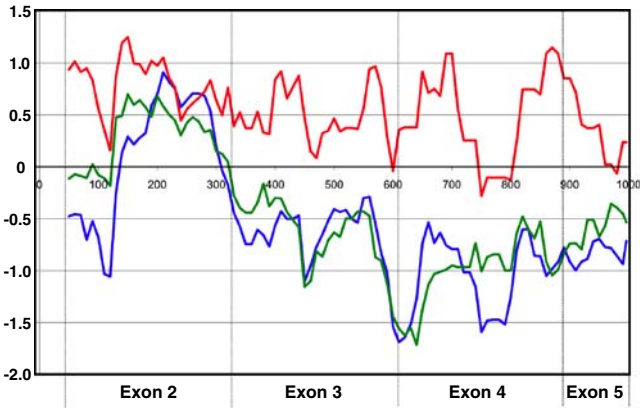

Fig. 3 Neutrality test by Tajima's D test of three cynomolgus macaque populations on *Mafa-A1* locus. *Blue*, *green*, and *red letters* indicate *Mafa-A1* alleles, which were identified in this study, derived from Vie, Ind, and Phi, respectively. The *vertical axis* shows the computed positive and negative D value above and below the 0 value, respectively. The *horizontal axis* shows the continuous nucleotide sequence length of exon 2 to exon 5 in the aligned cDNA sequences from each population group

to at least another 100 *Mafa-A1* alleles previously published and deposited in the IPD-MHC database (http://www.ebi.ac.uk/ipd/mhc/nhp/intro.html). Therefore, at least 160 *Mafa-A1* alleles have now been reported for the cynomolgus macaques and about 94 *Mamu-A1* alleles for the rhesus macaque, including some shared alleles, which may be the result of transspecies polymorphic inheritance. In this regard, it is believed that the cynomolgus macaque has undergone an ancient male introgression with a rhesus macaque in the Indochinese peninsula with a subsequent gene flow in the direction from *M. mulatta* to *M. fascicularis* (Tosi et al. 2002; Bonhomme et al. 2009).

The occurrence of multiple *Mafa-A* loci as a consequence of gene duplication events and a large number of allele variations at different loci is clear evidence for *Mafa-A* gene divergence evolution. The number of *Mafa-A* gene copies is not known exactly, but previous allele and phylogenetic analysis suggests at least six possible loci (Otting et al. 2007). We have identified five loci in our study, with allelic sequences possibly expressed by the gene loci *Mafa-A1* to *Mafa-A5*. In contrast to the detection of 83 *Mafa-A1* alleles, only seven *Mafa-A2* to *Mafa-A5* alleles were identified in this study. These results are similar to those reported by others (Otting et al. 2007; Campbell et al. 2009). On the basis of the familial study (Supplementary Fig. 1 and Table 2) and our ability to clone or detect only a few *Mafa-A2*, *Mafa-A3*, *Mafa-A4*, or *Mafa-A5* alleles, we assume that the *Mafa-A2* to *Mafa-A5* gene products were poorly expressed or suppressed in the peripheral blood cells. This view is consistent with a report by Otting et al. (2007) that transcription levels of the *Mafa-A2* to *Mafa-A5* genes were lower than that of the *Mafa-A1* gene in the B cell lines that were used in their study. Since we and others have used peripheral white blood cells for cloning and sequence analysis of the *Mafa-A* gene transcripts (Pendley et al. 2008; Campbell et al. 2009), it is likely that the *Mafa-A2* to *Mafa-A5* gene expression activities are cell or tissue specific and that some of the *Mafa-A* loci might produce RNA null alleles. Alternatively, null alleles from the non-A1 loci have not been amplified because of the failure of cDNA primers to hybridize efficiently due to substitutions and/or indels in the sequence of the primer-binding sites. However, since there are at least 53 *Mafa-A2* to *Mafa-A5* alleles (Uda et al. 2004; Krebs et al. 2005; Otting et al. 2007) including the seven alleles that we identified in this study, additional polymorphism, genome structure, and expression analyses using different tissue sources, cDNA loci-specific primers, and comparative genetic methods are warranted to obtain a better perspective of the *Mafa-A* loci polymorphism and expression profiles.

The *Mafa-A1* gene, like the *HLA-A* gene of humans, has high gene diversity as seen by the results of the large number of polymorphic sites across the gene sequence, the large level of heterozygosity that reveals a large number of individuals with polymorphic loci, and the high percentage (57% to 129%) of allele diversity per population (Supplementary Table 2). This extraordinarily large level of diversity in the *Mafa-A1* gene, which is caused by point mutation turnover and/or gene conversions (recombinations), appears to be an intrinsic feature of the structure and function of the antigen sequence with pathogen-mediated balancing selection conferring a survival advantage of the gene's heterozygosity and polymorphism onto the breeding animals. The greatest sequence variability was found in the peptide-binding domains of the *Mafa-A1* exons 2 and 3, as previously seen for the *MHC* classical class I genes of many other mammalian species including *HLA-A1* of humans.

The effect of selection and demographic factors on *Mafa-A1* diversity

Although *MHC* genetic diversity may be generated by pathogen-mediated balancing selection where the prevalence of genetic variants confer pathogen resistance, demographic factors such as the level of gene flow, bottleneck, and expansion events also may have had an effect. Tests for intrapopulation diversity and interpopulation differentiation (*F* statistic) and phylogenetic analysis of the *Mafa-A1* sequences in our study did not identify any significant differences between the three macaque populations nor differentiate between demographic and selective forces on *Mafa-A1* genetic variation. On the other hand, Tajimi's *D* test did reveal a major difference between the Phi and other populations with the *Mafa-A1* gene variants

of the Phi population affected by either balancing (positive) selection and/or a decrease in population size. The demographic factors seem a more appropriate explanation than selection in the Phi population because the diversity profile, which is a strong indicator of the effects of balancing selection, revealed essentially that there was little or no difference between the three populations in terms of selection pressure. In comparison to the macaques, the shape of the nucleotide diversity profiles were the same for the *HLA-A* polymorphisms of Japanese and Caucasians but at a much lower level of intensity than those for the macaque *Mafa-A1* polymorphisms. The higher degree of polymorphisms in macaques than in humans corresponds to the view that macaques are much older as a species than humans and that the macaque polymorphisms were generated over a longer evolutionary time period.

Both positive selection and demographic history might explain the lower levels of *Mafa-A1* allele diversity (57%) in the Phi population in comparison to the Vie (129%) and Ind (126%) populations. The lower levels of *Mafa-A1* variability in the Phi is consistent with a previous study of *Mafa-DPB* variability in the Phi (Sano et al. 2006), which suggested a population founder effect and a bottleneck. Recently, Blancher et al. (2008) showed a low nucleotide diversity for a mitochondrial DNA sequence in the Phi macaque population and suggested that a bottleneck occurred following colonization by Ind individuals, around 110,000 years before present (BP). However, the lower levels of *Mafa-A1* variability might also correlate with the occurrence of low parasitemia in Phi individuals, in contrast to the greater number of fatal parasite infections observed in other populations (Schmidt et al. 1977; Bonhomme et al. 2007). This low parasitemia may be due in part to the efficiency of the *MHC* antigens in immune protection or to an as yet unspecified cleaner or pathogen-freer environment resulting in fewer infections.

The high frequency (0.322) of the allele *Mafa-A1*08903* in the Phi but not in the Vie or Ind populations might have been selected for in favor of its protective role against parasites and other infectious agents. The selection value of this allele, previously reported by Campbell et al. (2009), is emphasized by its presence also in the rhesus macaque species as *Mamu-A*8903*. Alternatively, the *Mafa-A1*08903* may have increased in frequency because of low environmental selection pressure. In this regard, a reduction in environmental selection forces might have contributed to the lower levels of *Mafa-A1* variability that we have observed in the Phi population.

Interestingly, only three of 66 *Mafa-A1* alleles were shared as identical sequences between the Ind and Vie populations, which is indicative of a low level of gene flow between these two populations. In addition, there was no exact match of the *Mafa-A1* alleles between the cynomolgus macaques of the Phi and those of Vie or Ind. These results support those of previous studies that the *Mafa-A1* alleles are mostly population specific, probably as a result of divergence evolution and balancing selection. However, the phylogenetic analysis of *Mafa-A1* in this and other studies (Campbell et al. 2009; Karl et al. 2008; Otting et al. 2007; Pendley et al. 2008) has shown that the population specificities are not absolute, and there is considerable overlap between different populations for some *Mafa-A1* allelic lineages at the three-digit level of classification such as between the Phi, Vie, or Ind for the *Mafa-A*004* and *Mafa-A*038*, *Mafa-A*010*, or *Mafa-A*066* clusters.

The evolutionary history of *Mafa* population origins

The cynomolgus macaques are believed to have originated and dispersed throughout Southeast Asia after the divergence of the genus *Macaca* approximately two million years ago (Abegg and Thierry 2002). The migration of animals across land bridges between continental Asia and islands of Indonesia is believed to have coincided with the glaciation events during the late Pleistocene epoch (~550,000 years BP) when the sea levels had lowered to expose the Sunda Shelf (Voris 2000). The Indonesian cynomolgus macaque were then later introduced onto other islands such as the Philippines and Mauritius via sea rafting or by human seafarers (Abegg and Thierry 2002). Identical or highly similar mitochondrial DNA, MHC microsatellite DNA, and class I cDNA sequences in Indonesian and Mauritian or Filipino cynomolgus macaques suggest that these populations probably arose from the same founding populations of Indonesian macaques (Blancher et al. 2008; Bonhomme et al. 2007; Kawamoto et al. 2008; Kondo et al. 1993; Smith et al. 2007; Tosi and Coke 2007). The low nucleotide diversity in the Phi macaque population is believed to have arisen from a bottleneck after colonization by Ind individuals around 110,000 years BP, well before the foundation of the Mauritius macaque population a few hundred years ago (Blancher et al. 2008; Bonhomme et al. 2008).

However, whereas the mitochondrial and microsatellite DNA markers were useful for separating the cynomolgus macaques into continental and insular subgroups and for inferring demographic histories, the *Mafa-A1* cDNA sequences from different populations often grouped together into ancient lineages or generally overlapped into less well-defined clusters. Although our phylogenetic results in this study are limited to only three populations, they do not obviously support the previously suggested ancient differentiation of cynomolgus macaques into the continental (Indochinese [Vie], Malaysian, Thai) and insular (Phi, Ind, Mau) subgroups. In our study, we obtained a major difference in allele percentage between the Phi (insular

group) and the Vie (continental) or Ind (insular) populations but not between the Vie (continental) and Ind (insular). Therefore, the *Mafa-A1* alleles alone appear to have limited value as phylogeographic markers for differentiating between continental and insular regional populations, which for more reliable and informative phylogeographic tree reconstructions seem to require neutral sequence markers that have not undergone a long and strong history of environmental selection pressures.

The implications of *Mafa-A* gene loci diversity in medical research

Because cynomolgus macaques are closely related genetically to humans, they are often used as experimental animals for biomedical research of human pathogenic diseases, reproduction, regenerative medicine, drug discovery, transplantation, and immunotherapy. The development of *MHC* homozygote macaques is considered necessary for vaccine development and for evaluating and validating the use of human regenerated cells originating from induced pluripotent stem and/or embryonic stem cells in transplantation medicine (Klimanskaya et al. 2006; Takahashi et al. 2007). In this regard, the cynomolgus macaque from Mauritius (Mau) and the Phi are thought to be the most suitable populations for use in biomedical research (Tosi and Coke 2007; Wiseman et al. 2007; Kawamoto et al. 2008) mainly because they have a low degree of polymorphisms in the *MHC* genes (Leuchte et al. 2004; Krebs et al. 2005; Blancher et al. 2006). The low degree of polymorphisms in the *MHC* genes of the Mau cynomolgus macaque population is believed to stem from their recent colonization of Mau, probably by human travelers during the Dutch occupation, or even preceding the Portuguese occupation of the island in the late sixteenth or early seventeenth century (Sussman and Tattersall 1981, 1986). Although the history of the cynomolgus macaques' origin in the Phi is not known exactly, like the Mau population, they also show a relatively high genetic diversity at the *Mafa-A1* locus but a smaller number of *Mafa-A1* alleles than the animals in Ind and Vie.

Extensive gene diversity might also be of value in medical research using cynomolgus macaques from different population groups as animal models especially in understanding diseases and developing *HLA*-associated disease models such as the rheumatoid arthritis collagen-induced model. It has been proposed that highly heterozygous macaques will be needed to investigate the immune responses and safety of idiosyncratic drugs that are associated with *MHC* molecules (Uetrecht 2007). In such cases, the diverged Vie, Ind, and Phi populations may have greater value for biomedical research than Mau because the Mau has a small allele repertoire originating from an Ind population (Pendley et al. 2008). The origin of the individuals and the genetic polymorphism of the macaque species need to be considered carefully at the population level because the results of biomedical experiments strongly depend on the immunogenetic background of animals conditioned by various environmental selective factors such as pathogens.

Conclusion

We have identified and reported on a large number of novel alleles for the *Mafa-A* genes and provided some further insights into the role of demographic and selection factors on the genetic structure, nucleotide diversity, and phylogenetic relationships of *Mafa-A1* alleles in three contrasting Southeast Asian populations. This provides researchers with an added basis for a more comprehensive analysis of the distribution and variation of *Mafa-A1* alleles in various other continental and insular populations and for investigating the adaptation of macaque populations to autoimmune disease and infection.

Acknowledgments We thank C. Iwatani and J. Yamasaki for helping with the cynomolgus macaque blood collection. We thank T. Takada, S. Kanazashi, T. Bando, A. Matsushita, and Y. Fukiya for helpful discussion. This work was supported by Grant-in-Aid for Scientific Research (B) (21300155) from Japan Society for the Promotion of Science (JSPS) and for Scientific Research on Priority Areas "Comparative Genomics" (20017023) from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (MEXT).

References

Abegg C, Thierry B (2002) Macaque evolution and dispersal in insular south-east Asia. Biol J Linn Soc Lond 75:555–576

Blancher A, Tisseyre P, Dutaur M, Apoil PA, Maurer C, Quesniaux V, Raulf F, Bigaud M, Abal M (2006) Study of cynomolgus monkey (*Macaca fascicularis*) *MhcDRB* (*Mafa-DRB*) polymorphism in two populations. Immunogenetics 58:269–282

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

Bonhomme M, Blancher A, Jalil MF, Crouau-Roy B (2007) Factors shaping genetic variation in the MHC of natural non-human primate populations. Tissue Antigens 70:398–411

Bonhomme M, Blancher A, Cuartero S, Chikhi L, Crouau-Roy B (2008) Origin and number of founders in an introduced insular primate: estimation from nuclear genetic data. Mol Ecol 17:1009–1019

Bonhomme M, Cuartero S, Blancher A, Crouau-Roy B (2009) Assessing natural introgression in 2 biomedical model species, the rhesus macaque (*Macaca mulatta*) and the long-tailed macaque (*Macaca fascicularis*). J Hered 100:158–169

Bryant D, Moulton V (2004) Neighbor-Net: an agglomerative method for the construction of phylogenetic networks. Mol Biol Evol 21:255–265

Campbell KJ, Detmer AM, Karl JA, Wiseman RW, Blasky AJ, Hughes AL, Bimber BN, O'Connor SL, O'Connor DH (2009) Characterization of 47 MHC class I sequences in Filipino cynomolgus macaques. Immunogenetics 61:177–187

Capuano SVr, Croix DA, Pawar S, Zinovik A, Myers A, Lin PL, Bissel S, Fuhrman C, Klein E, Flynn JL (2003) Experimental *Mycobacterium tuberculosis* infection of cynomolgus macaques closely resembles the various manifestations of human *M. tuberculosis* infection. Infect Immun 71:5831–5844

Conlee KM, Hoffeld EH, Stephens ML (2004) A demographic analysis of primate research in the United States. ATLA Supplement 1:315–322

Emborg ME (2007) Nonhuman primate models of Parkinson's disease. ILAR J 48:339–355

Excoffier L, Laval G, Schneider S (2005) Arlequin ver. 3.0: an integrated software package for population genetics data analysis. Evolutionary Bioinformatics Online 1:47–50

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

Huson DH, Bryant D (2006) Application of phylogenetic networks in evolutionary studies. Mol Biol Evol 23:254–267

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

Kawamoto Y, Kawamoto S, Matsubayashi K, Nozawa K, Watanabe T, Stanley MA, Perwitasari-Farajallah D (2008) Genetic diversity of longtail macaques (*Macaca fascicularis*) on the island of Mauritius: an assessment of nuclear and mitochondrial DNA polymorphisms. J Med Primatol 37:45–54

Klimanskaya I, Chung Y, Becker S, Lu J, Lanza R (2006) Human embryonic stem cell lines derived from single blastomeres. Nature 444:481–485

Kondo M, Kawamoto Y, Nozawa K, Matsybayashi K, Watanabe T, Griffiths O, Stanley M (1993) Population genetics of crab-eating macaques (*Macaca fascicularis*) on the island of Mauritius. Am J Primatol 29:167–182

Krebs KC, Jin Z, Rudersdorf R, Hughes AL, O'Connor DH (2005) Unusually high frequency MHC class I alleles in Mauritian origin cynomolgus macaques. J Immunol 175:5230–5239

Kulski JK, Shigenari A, Shiina T, Ota M, Hosomichi K, James I, Inoko H (2008) Human endogenous retrovirus (HERVK9) structural polymorphism with haplotypic HLA-A allelic associations. Genetics 180:445–457

Lawler JV, Endy TP, Hensley LE, Garrison A, Fritz EA, Lesar M, Baric RS, Kulesh DA, Norwood DA, Wasieloski LP, Ulrich MP, Slezak TR, Vitalis E, Huggins JW, Jahrling PB, Paragas J (2006) Cynomolgus macaque as an animal model for severe acute respiratory syndrome. PLoS Med 3:677–686

Leuchte N, Berry N, Kohler B, Almond N, LeGrand R, Thorstensson R, Titti F, Sauermann U (2004) *MhcDRB*-sequences from cynomolgus macaques (*Macaca fascicularis*) of different origin. Tissue Antigens 63:529–537

Liu C, Noorchashm H, Sutter JA, Naji M, Prak EL, Boyer J, Green T, Rickels MR, Tomaszewski JE, Koeberlein B, Wang Z, Paessler ME, Velidedeoglu E, Rostami SY, Yu M, Barker CF, Naji A (2007) B lymphocyte-directed immunotherapy promotes long term islet allograft survival in nonhuman primates. Nat Med 13:1295–1298

Moriyama Y, Kato K, Mura T, Juji T (2006) Analysis of HLA gene frequencies and HLA haplotype frequencies for bone marrow donors in Japan. MHC 12:183–201 (in Japanese)

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

Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE (2005) Unparalleled complexity of the MHC class I region in rhesus macaques. Proc Natl Acad Sci U S A 102:1626–1631

Otting N, deVos-Rouweler AJ, Heijmans CM, deGroot NG, Doxiadis GG, Bontrop RE (2007) MHC class I A region diversity and polymorphism in macaque species. Immunogenetics 59:367–375

Pendley CJ, Becker EA, Karl JA, Blasky AJ, Wiseman RW, Hughes AL, O'Connor SL, O'Connor DH (2008) MHC class I characterization of Indonesian cynomolgus macaques. Immunogenetics 60:339–351

Robinson J, Waller MJ, Parham P, de Groot N, Bontrop R, Kennedy LJ, Stoehr P, Marsh SGE (2003) IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. Nucleic Acids Research 31:311–314

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

Sano K, Shiina T, Kohara S, Yanagiya K, Hosomichi K, Shimizu S, Anzai T, Watanabe A, Ogasawara K, Torii R, Kulski JK, Inoko H (2006) Novel cynomolgus macaque *MHC-DPB1* polymorphisms in three South-East Asian populations. Tissue Antigens 67:297–306

Schmidt LH, Fradkin R, Harrison J, Rossan RN (1977) Differences in the virulence of *Plasmodium knowlesi* for *Macaca irus (fascicularis)* of Philippine and Malayan origins. Am J Trop Med Hyg 26:612–622

Shiina T, Inoko H, Kulski JK (2004) An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens 64:631–649

Shiina T, Hosomichi K, Inoko H, Kulski JK (2009) The HLA genomic loci map: expression, interaction, diversity and disease. J Hum Genet 54:15–39

Shively CA, Wood CE, Register TC, Willard SL, Lees CJ, Chen H, Sitruk-Ware RL, Tsong YY, Cline JM (2007) Hormone therapy effects on social behavior and activity levels of surgically postmenopausal cynomolgus monkeys. Psychoneuroendocrinology 32:981–990

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

Southwick CH, Siddiqi MF (1994) Population status of nonhuman primates in Asia, with emphasis on rhesus macaques in India. Am J Primatol 34:51–59

Sussman RW, Tattersall I (1981) Behavior and ecology of *Macaca fascicularis* in Mauritius: a preliminary study. Primates 22:192–205

Sussman RW, Tattersall I (1986) Distribution, abundance, and putative ecological strategy of *Macaca fascicularis* on the island of Mauritius, Southwestern Indian Ocean. Folia Primatology 46:28–43

Tajima F (1989) The effect of change in population size on DNA polymorphism. Genetics 123:597–601

Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, Yamanaka S (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131:861–872

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tosi AJ, Coke CS (2007) Comparative phylogenetics offer new insights into the biogeographic history of *Macaca fascicularis* and the origin of the Mauritian macaques. Mol Phylogenet Evol 42:498–504

Tosi AJ, Morales JC, Melnick DJ (2002) Y-chromosome and mitochondrial markers in Macaca fascicularis indicate introgression with Indochinese *M. mulatta* and a biogeographic barrier in the isthmus of Kra. Int J Primatol 23:161–178

Uda A, Tanabayashi K, Yamada YK, Akari H, Lee YJ, Mukai R, Terao K, Yamada A (2004) Detection of 14 alleles derived from the MHC class I A locus in cynomolgus monkeys. Immunogenetics 56:155–163

Uda A, Tanabayashi K, Fujita O, Hotta A, Terao K, Yamada A (2005) Identification of the MHC class I B locus in cynomolgus monkeys. Immunogenetics 57:189–197

Uetrecht J (2007) Idiosyncratic drug reactions: current understanding. Annu Rev Pharmacol Toxico 47:513–539

van Oosterhout C (2009) A new theory of MHC evolution: beyond selection on the immune genes. Proc Biol Sci 276:657–665

Voris H (2000) Maps of Pleistocene sea levels in Southeast Asia: shorelines, river systems and time durations. J Biogeogr 27:1153–1167

Wang CY, Finstad CL, Walfield AM, Sia C, Sokoll KK, Chang TY, Fang XD, Hung CH, Hutter-Paier B, Windisch M (2007) Site specific UBITh amyloid-beta vaccine for immunotherapy of Alzheimer's disease. Vaccine 25:3041–3052

Wiseman RW, O'Connor DH (2007) Major histocompatibility complex-defined macaques in transplantation research. Transplant Rev 21:17–25

Wiseman RW, Wojcechowskyj JA, Greene JM, Blasky AJ, Gopon T, Soma T, Friedrich TC, O'Connor SL, O'Connor DH (2007) Simian immunodeficiency virus SIVmac239 infection of major histocompatibility complex-identical cynomolgus macaques from Mauritius. J Virol 81:349–361

Wright S (1965) The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420

Footnote

The nucleotide sequence data reported in this paper have been submitted to the DDBJ, EMBL, and GenBank nucleotide sequence databases and have been assigned accession numbers ranging from AB447561 to 447620, from AB448747 to AB448752 and AB474960, and also have been submitted to the IPD-MHC database and have been assigned accession numbers ranging from 70004335 to 70004408.