Immunogenetics (2015) 67:563–578
DOI 10.1007/s00251-015-0867-9

ORIGINAL PAPER

Discovery of novel *MHC-class I* alleles and haplotypes in Filipino cynomolgus macaques (*Macaca fascicularis*) by pyrosequencing and Sanger sequencing

Mafa-class I polymorphism

Takashi Shiina[1] · Yukiho Yamada[1,2] · Alice Aarnink[3] · Shingo Suzuki[1] · Anri Masuya[1] · Sayaka Ito[1] · Daisuke Ido[2] · Hisashi Yamanaka[2] · Chizuru Iwatani[4] · Hideaki Tsuchiya[4] · Hirohito Ishigaki[5] · Yasushi Itoh[5] · Kazumasa Ogasawara[5] · Jerzy K. Kulski[1,6] · Antoine Blancher[3]

Received: 25 June 2015 / Accepted: 27 August 2015 / Published online: 8 September 2015

Abstract Although the low polymorphism of the major histocompatibility complex (*MHC*) transplantation genes in the Filipino cynomolgus macaque (*Macaca fascicularis*) is expected to have important implications in the selection and breeding of animals for medical research, detailed polymorphism information is still lacking for many of the duplicated class I genes. To better elucidate the degree and types of *MHC* polymorphisms and haplotypes in the Filipino macaque population, we genotyped 127 unrelated animals by the Sanger sequencing method and high-resolution pyrosequencing and identified 112 different alleles, 28 at cynomolgus macaque *MHC (Mafa)-A*, 54 at *Mafa-B*, 12 at *Mafa-I*, 11 at *Mafa-E*, and seven at *Mafa-F* alleles, of which 56 were newly described. Of them, the newly discovered *Mafa-A8*01:01* lineage allele had low nucleotide similarities (<86 %) with primate *MHC* class I genes, and it was also conserved in the Vietnamese and Indonesian populations. In addition, haplotype estimations revealed 17 *Mafa-A*, 23 *Mafa-B*, and 12 *Mafa-E* haplotypes integrated with 84 *Mafa-class I* haplotypes and *Mafa-F* alleles. Of these, the two *Mafa-class I* haplotypes, F/A/E/B-Hp1 and F/A/E/B-Hp2, had the highest haplotype frequencies at 10.6 and 10.2 %, respectively. This suggests that large scale genetic screening of the Filipino macaque population would identify these and other high-frequency *Mafa-class I* haplotypes that could be used as *MHC* control animals for the benefit of biomedical research.

Keywords Cynomolgus macaque · Major histocompatibility complex · MHC · Polymorphism · Haplotype · Next generation sequencing · NGS · Sanger sequencing

Electronic supplementary material The online version of this article (doi:10.1007/s00251-015-0867-9) contains supplementary material, which is available to authorized users.

✉ Takashi Shiina
tshiina@is.icc.u-tokai.ac.jp

[1] 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

[2] Ina Research Inc., 2148-188 Nishiminowa, Ina, Nagano 399-4501, Japan

[3] Laboratoire d'Immunogénétique moléculaire, Université Toulouse III, Faculte de Medecine Purpan, Laboratoire d'immunologie, 1 Avenue Jean Poulhes, Hôpital Rangueil, TSA50032, 31059 ToulouseCedex 9, France

[4] Research Center for Animal Life Science, Shiga University of Medical Science, School of Medicine, Setatsukinowa, Ohtsu, Shiga 520-2192, Japan

[5] Department of Pathology, Shiga University of Medical Science, Setatsukinowa, Ohtsu, Shiga 520-2192, Japan

[6] Centre for Forensic Science, The University of Western Australia, Nedlands,, WA 6008, Australia

Introduction

The cynomolgus macaques (*Macaca fascicularis*), alias the crab-eating monkeys or long-tailed macaques, are widely dispersed in nature, inhabiting a vast range of Southeast Asia including the Philippines, Indonesia, Vietnam, Malaysia, Thailand, Cambodia, Mauritius, and Brunei (Fuentes et al. 2011). This species along with other macaque species such

as the rhesus macaque are often used for biomedical research on infectious and chronic diseases, neurology, immunology, reproduction, regenerative medicine, transplantation, vaccination, and immunotherapy (Gardner and Luciw 2008; Vallender and Miller 2013). Many of these diseases and treatment outcomes are influenced by polymorphisms of the major histocompatibility complex (MHC), a genomic region that encodes the MHC transplantation and immune regulatory molecules (Shiina et al. 2009; Shiina et al. 2004). In this regard, information on the diversity of the MHC alleles and haplotypes may provide an internal genetic control to better define the factors of disease susceptibility and immunity and help to maximize the power of the small study sizes and minimize a priori bias and the effect of heterozygosity in the macaque disease study groups (Arikata et al. 2012; Ericsen et al. 2014; Vallender 2014).

The human classical *MHC* class I genes, *Human Leukocyte Antigen (HLA)-A*, *HLA-B*, and *HLA-C*, are distinguished by their extraordinary polymorphisms with over 9,600 alleles implicated in disease resistance or susceptibility; whereas the human non-classical *MHC* class I genes, *HLA-E*, *HLA-F*, and *HLA-G*, are more limited in diversity by their relative monomorphism (IMGT/HLA Database release 3.20.0, http://www.ebi.ac.uk/imgt/hla/) (Robinson et al. 2011). In comparison, the overall structure of the *MHC* class I regions in the rhesus macaques (Daza-Vamenta et al. 2004) is more complex than the *HLA* (Shiina et al. 1999) and the common chimpanzee *MHC* (*Patr*) (Anzai et al. 2003) because the macaques carry many more duplicated *MHC* class I genes than humans and chimpanzees. For example, the *Mafa-class I* region, like the *HLA* region, is divided into three segments, whereby up to nine duplicated *Mafa-B* and *Mafa-I* (*Mafa-B/I*) genes correspond to *HLA-B/C* in the beta block, *Mafa-E* corresponds to *HLA-E* in the kappa block, and numerous duplicated *Mafa-A*, *Mafa-AG*, and *Mafa-F* genes correspond to *HLA-A/G/F* in the alpha block (Kulski et al. 2002; Watanabe et al. 2007) and Fig. 1. Overall, the homologous duplications appear to be fewer in number in the *Mafa-class I* region (Campbell et al. 2009; Otting et al. 2007; Pendley et al. 2008; Uda et al. 2004; Watanabe et al. 2007) than in the *Mamu-class I* region (Shiina et al. 2006).

The detection and breeding of *MHC* homozygous primate species are considered necessary for the development of effective vaccine and immunosuppressive drug protocols, evaluation, and validation of organ and regenerated cells originating from induced pluripotent stem (iPS) cells and/or embryonic stem (ES) cells in transplantation medicine (Klimanskaya et al. 2006; Takahashi et al. 2007). In this regard, the cynomolgus macaques from Mauritius and the Philippines are considered the most suitable populations for use in biomedical research because they are thought to be genetically less diverse than those from other geographic locations (Kawamoto et al. 2008; Tosi and Coke 2007; Wiseman et al. 2007). The Mauritian and Filipino macaque populations are believed to have originated from a colonization of Indonesian/continental animals followed by a bottleneck approximately 110,000 years BP (Blancher et al. 2008). In fact, the Mauritian and Filipino macaques show a relatively high genetic diversity, but a smaller number of *MHC* class I and class II alleles than the Indonesian and Vietnamese populations (Blancher et al. 2012; Blancher et al. 2014; Blancher et al. 2006; Kita et al. 2009; Krebs et al. 2005; Leuchte et al. 2004; Sano et al. 2006). Although the seven most frequent *Mafa-class I* haplotypes were characterized from 67 distinct *Mafa-class I* transcripts in the Mauritian population (Budde et al. 2010), detailed polymorphic information for the *Mafa-class I* genes is still lacking in comparison to the *Mafa-class II* genes (Blancher et al. 2014) in the Filipino population. Hence, in order to effectively use the Filipino population for medical research based on discovery of *MHC* homozygous animals, it is necessary to better understand the allele and haplotype diversity of the *Mafa-class I* genes.

Massively parallel pyrosequencing is one of the next generation sequencing (NGS) techniques and more effective methodologies for high-throughput genotyping of *MHC* genes and the detection of low-level-expressed *MHC* alleles (Babik et al. 2009; Wegner 2009; Wiseman et al. 2009). Pyrosequencing has been used successfully to discover and accumulate a large number of *MHC* alleles in some macaque species (Budde et al. 2010; Karl et al. 2009; O'Leary et al. 2009; Wiseman et al. 2009). This NGS method is precise and speedy for genotyping the highly duplicated macaque *MHC* class I genes from tissue RNA than the conventional *MHC* genotyping methods such as microsatellite, sub-cloning, and Sanger sequencing after reverse transcriptase-polymerase chain reaction (RT-PCR).

In this study, to better elucidate the degree and types of allele and haplotype diversity of the *Mafa-class I* genes in the Filipino population, we identified 56 novel alleles from 112 alleles obtained for the *Mafa-A*, *Mafa-B/I*, *Mafa-E*, and *Mafa-F* genes in 127 unrelated cynomolgus macaques by using both the conventional and the NGS genotyping methods. We analyzed the allele sequences by performing both population genetic and phylogenetic analyses on the coding regions (exons 1 to 7) of the complementary DNA (cDNA) nucleotide sequences that we obtained from the peripheral white blood cells of the macaques.

Material and methods

Animals

In this study, we used the RNA from the white blood cells of 127 unrelated cynomolgus macaques from the Philippines archipelago (112 from INA Research Philippines INC. and 15

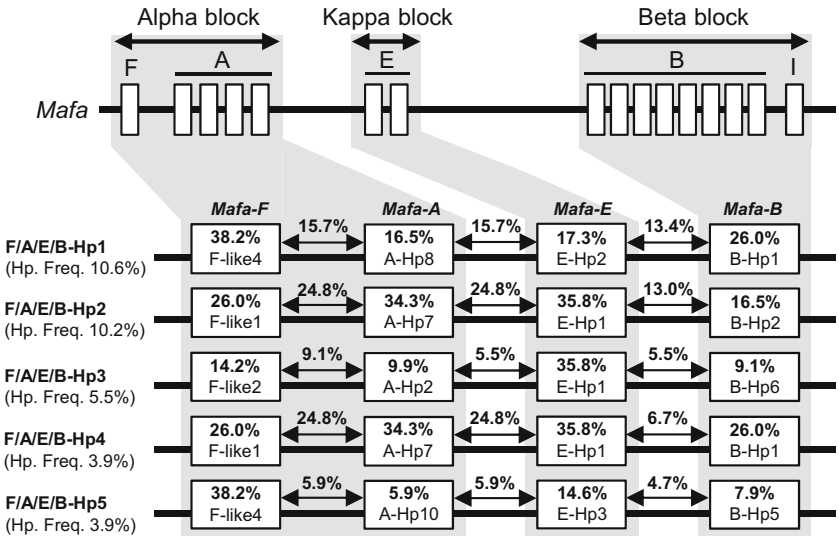


Fig. 1 The *Mafa-class I* duplicated genes within the *Mafa-class I* genomic structure and the representative *Mafa-class I* haplotypes that have high frequencies in the Filipino cynomolgus macaque population. The number and distribution of the *Mafa-class I* genes *A*, *B/I*, *E*, and *F* within each of the *shaded segmental block structures* were taken from previous publications (de Groot et al. 2012; Kulski et al. 2002; Urvater et al. 2000; Watanabe et al. 2007). *Names and numbers in the boxes* show *Mafa-F* allele or A-Hp, B-Hp and E-Hp names and their frequencies in the 127 Filipino cynomolgus macaques. *Numbers between* boxes show haplotype frequencies of F/A-Hp, A/E-Hp, and E/B-Hp. Hp.Freq. means haplotype frequency

from Sicombrec Co.). The *Mafa-DR—Mafa-DQ—Mafa-DP* haplotypes of these animals have been previously characterized (Blancher et al. 2012; Blancher et al. 2014). We also used the genomic DNA from the white blood cells of 20 unrelated cynomolgus macaques from Vietnam and Indonesia (ten from Nafovanny and ten from CV Universal Fauna Breeder and Exporter of Non-human Primates for Laboratories) for genotyping of a novel *MHC* class I gene, *Mafa-A8*01:01*. The blood collection and animal studies were conducted in accordance with the guidelines for animal experiments specific to each location.

RNA extraction and reverse-transcriptase (RT) reaction

Total RNA was extracted directly from the peripheral white blood cell samples of each of the 127 animals using the TRIzol reagent (Invitrogen/Life Technologies/Thermo Fisher Scientific, Carlsbad, CA). cDNA was synthesized by oligo d(T) primer using the ReverTraAce for reverse transcriptase reaction (TOYOBO, Osaka, Japan) after treatment of the isolated RNA with DNase I (Invitrogen/Life Technologies/Thermo Fisher Scientific, Carlsbad, CA).

Sub-cloning and Sanger sequencing of Mafa-A, Mafa-B, and Mafa-I genes

The Sanger sequencing method genotyped 127 animals using the *Mafa-A*, *Mafa-B*, and *Mafa-I* cDNA sequences, ranging from exon 1 to exon 7 (PCR product size: 1,037~1,061 bp). The sequencing primers used for the initial genotyping procedure were one common sense primer (Mafa-A/B/I_F) and the two locus-specific anti-sense primers, Mafa-A_R for *Mafa-A* (Kita et al. 2009) and Mafa-B/I_R for *Mafa-B* and *Mafa-I* (Table 1). The Sanger sequencing method also was used to confirm 37 novel cDNA sequences of *Mafa-A*, *Mafa-B*, and *Mafa-I* in 26 animals that were identified by the pyrosequencing method. In this case, confirmatory genotyping by the Sanger method was performed with six allele-specific primers (underlined in Table 1) that were specifically based on the newly identified sequences. *Mafa-G* and *Mafa-AG* were not examined in this study because *Mafa-G* is not expressed by white blood cells and the *Mafa-G* copies are unexpressed pseudogenes. In brief, the 20 μL amplification-reaction-volume contained 10 ng of cDNA, 0.4 units of KOD FX polymerase (TOYOBO), 2× 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, 60°C/30 s, and 68°C/1 min. PCR reactions were performed using the thermal cycler GeneAmp PCR system 9700 (Applied Biosystems/Life Technologies/Thermo Fisher Scientific, Foster City, CA). After PCR amplification, PCR products were cloned into the PGEM-T Easy vector with the pGEM®-T Easy Vector System according to the protocol provided by the manufacturer (Promega, Madison, WI) and sequenced by using the ABI3130 genetic analyzer (Applied Biosystems/Life Technologies/Thermo Fisher Scientific) in accordance with the protocol of Big Dye terminator method. To avoid PCR and sequencing artifacts generated by polymerase errors, 48 clones per animal were sequenced.

Table 1 PCR primers used for this study

Primer name	Direction	Primer sequence (5′ to 3′)	Product size
For full-length cDNA sequence determination of *Mafa-A* gene by Sanger method			
Mafa-A/B/I_F	Sense	AACCCTCCTCCTGSTGCTCT	1037 bp or 1043 bp
Mafa-A_R	Anti-sense	CCTGGGCACTGTCACTGCTT	
Mafa-A/B_F(1402)	Sense	AACCCTTCTCCTGGTGCTCT	
Mafa-A/B_F(7102)	Sense	AACTCTCCTTCTGGTG2CTCT	
Mafa-A_R(7102)	Anti-sense	CCTGGGCACTGTCATTGCTT	
For full-length cDNA sequence determination of *Mafa-B/I* gene by Sanger method			
Mafa-A/B/I_F	Sense	AACCCTCCTCCTGSTGCTCT	1040 bp, 1043 bp or 1061 bp
Mafa-B/I_R	Anti-sense	CCTGGGCACTGTCGYTGGAC	
Mafa-B_R(4101)	Anti-sense	CCTGGGGACTGTCCTTGGAT	
Mafa-B_R(890101)	Anti-sense	TCTGGGCACTGTTGCTACAT	
Mafa-B_R(9802)	Anti-sense	CCTGGGCACTGTCGCTGTAC	
For full-length cDNA sequence determination of *Mafa-A8* gene by Sanger method			
MHCI-5.1	Sense	GGACTCAGAATCTCCCCAGA	1264 bp
339-F	Sense	AACTCTCCTCCTGCTGCTCT	
339-R	Anti-sense	CCTGGGCACTGTCACTCCTT	
MHCI-3.1	Anti-sense	GTCTCTCCACCTCCTCACAT	
For genotyping by pyrosequencing			
Class_I_F	Sense	CGTGCGGTTYGAYAGCGACG	514 bp or 517 bp
Class_I_R	Anti-sense	CCAGCAYCTCAGGGTGGCCTC	
For qPCR analysis of *Mafa-A8*			
339genome_F	Sense	GAACTCTGACGAGACAATTGCCCA	857 bp
339genome_R2	Anti-sense	CAGGTCTGGTGTCCACAAGATCCA	

Genotyping of Mafa-class I genes by pyrosequencing

A single *Mafa-class I*-specific primer set (Class_I_F and Class_I_R) was designed based on generic sequences in exon 2 and exon 4 (PCR product size: 514 or 517 bp) that could amplify all known *Mafa-class I* alleles for massively parallel pyrosequencing (Table 1) of 127 animals. The 20 μL RT-PCR amplification-reaction-volume contained 10 ng of cDNA, 0.4 units of high-fidelity KOD FX polymerase (TOYOBO), 2× PCR buffer, 2 mM of each dNTP and 0.5 μM of each primer. The cycling parameter was as follows: 25 cycles of 98°C/10 s, 58°C/30 s, and 68°C/30 s. The PCR products were purified by the QIAquick PCR Purification Kit (QIAGEN, Hiden, Germany) and quantified by the Picogreen assay (Invitrogen) with a Fluoroskan Ascent micro-plate fluorometer (Thermo Fisher Scientific, Waltham, MA).

Titanium rapid libraries of fragmented DNA linked to AMPure beads (Beckman Coulter Genomics, Danvers, MA) were prepared for the Roche Genome Sequencer 454 FLX system by nebulization, fragment end repair, and multiple identifier (MID)-labeled adaptor ligation, emulsion PCR (emPCR), and emulsion breaking were performed according to the manufacturer's protocol (Roche, Basel, Switzerland) (Kita et al. 2012). After the emulsion breaking step, the beads carrying the single-stranded DNA templates were enriched, counted, and deposited into a PicoTiterPlate to obtain sequence reads (Kita et al. 2012). After the sequencing run, image processing, signal correction, and base-calling were performed by the GS Run Processor Ver. 3.0 (Roche) with full processing for shotgun or paired-end filter analysis. Quality-filter sequence reads that were passed by the assembler software (single sff file) were binned on the basis of the MID labels into each separate sequence sff file using the sff file software (Roche). These files were further quality trimmed to remove poor sequence at the end of the reads with quality values (QVs) of less than 20. The *Mafa-class I* alleles were mainly assigned by matching the trimmed and MID labeled sequence reads as 99 and 100 % matching, 200 of minimum overlap length and 10 alignment identity score parameters with all the known *Mafa-class I* allele sequences released in the IMGT/MHC-NHP database (Robinson et al. 2013) using the GS Reference Mapper Ver. 3.0. On the other hand, to discover novel *Mafa-class I* sequences, the trimmed and MID-binned sequences were de novo assembled as >85 % matched parameters after the outputs were converted to ace files for the Sequencher Ver. 5.01 DNA sequence assembly software (Gene Code Co., Ann Arbor, MI). A defined consensus sequence obtained from the de novo assembly was used as

a reference sequence to identify and map the correct allele sequence in this study.

Nomenclature of novel Mafa-class I alleles

All newly determined *Mafa-class I* sequences were deposited in the Genbank database (accession references are given in Table 2), and the novel *Mafa-A*, *Mafa-B*, and *Mafa-I* sequences also were submitted to the IMGT/MHC-NHP database (Robinson et al. 2013) for allele nomenclature. The *Mafa-class I* allele names were assigned by the IPD-MHC NHP database following classical rules (de Groot et al. 2012). An example of the allele nomenclature is *Mafa-A1*001:02:01* where *Mafa-A1* 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 *001*, whereas the two digits after the first colon define the allele number *02*. These allele numbers are arbitrary and are numbered in the order in which they were identified. The two digits *01* after the second colon describe a synonymous base pair difference between two sequences. Whereas the *Mafa-A* alleles may be grouped into at least four lineages or loci, no locus number designation has been introduced for most of the *Mafa-B* loci because the macaque *B* genes largely differ in number between haplotypes (de Groot et al. 2012). An exception to the rule is the oligomorphic locus *B3*, which was previously named as the locus *Mamu-I* in the rhesus macaque (Urvater et al. 2000), and is found on each haplotype (de Groot et al. 2012). In this study, we refer to *Mafa-B3* locus as *Mafa-I.*

Haplotype estimation

The *Mafa-A*, *Mafa-B*, and *Mafa-E* haplotypes were characterized by manually sorting of *Mafa-A*, *Mafa-B*, *Mafa-I*, and *Mafa-E* alleles based on *Mafa-A*, *Mafa-B/I*, and/or *Mafa-E* homozygous animals. In contrast, we estimated *Mafa-A*, *Mafa-B/I*, and *Mafa-E* haplotypes by comparison between the homozygous animals and heterozygous animals that have the identical *MHC-class I* alleles with the homozygous animals as shown in Supplementary figure 1. Estimation of the *Mafa-class I* haplotypes and haplotype frequencies were performed by the PHASE 2.1.1 program (Stephens et al. 2001) using *Mafa-F* allele and *Mafa-A*, *Mafa-B*, and *Mafa-E* haplotype data.

Characterization of a novel Mafa-class I locus Mafa-A8*01:01

The full coding sequence of *Mafa-A8*01:01* was determined by the conventional method as described above using two pairs of *Mafa-A8*01:01* gene-specific primers (MHCI-5.1 and 339-R, and 339-F, and MHCI-3.1, Table 1).

One *Mafa-A8*01:01*-specific primer set (339genome_F and 339genome_R2) was used for detection of the gene in Vietnamese and Indonesian animals by PCR analysis and for detection of the copy numbers of *Mafa-A8*01:01* in allele-positive Filipino animals by quantitative PCR (qPCR) method (Table 1). The primers were designed to amplify the allele sequence in exons 3 and 4 (PCR product size: 857 bp). Twenty genomic DNAs from ten Vietnamese and ten Indonesian animals were extracted from peripheral blood cells using the QIA amp Blood Kit (QIAGEN). The 20-μL amplification-reaction-volume contained 10 ng of genomic DNA, 0.4 units of KOD FX polymerase (TOYOBO), 2× 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, 65 °C/30 s, and 68 °C/1 min.

Copy numbers of the gene was detected by qPCR method using the StepOnePlus™ Real-Time PCR System (Applied Biosystems/Life Technologies/Thermo Fisher Scientific) with KOD SYBR® qPCR Mix (TOYOBO). Melting curve analysis showed that there was no primer dimer formation. The relative quantitative values were calculated by the comparative C(T) method also referred to as the 2 (−DeltaDeltaC(T)) method (Schmittgen and Livak 2008).

Phylogenetic analysis

Multiple sequence alignment was created using the ClustalW Sequence Alignment program of the Molecular Evolution Genetics Analysis software 5 (MEGA5) (Tamura et al. 2011). Phylogenetic trees of the *MHC-class I* genes were constructed by the neighbor-joining (NJ) method in MEGA5 (Saitou and Nei 1987) using exons 2 to 4 (alignment length: 822 bp excluding gap sites). A NJ tree was constructed by the Maximum Composite Likelihood model and assessed using 10,000 bootstrap replicates. We used the following *MHC-I* sequences (DNA accession numbers) for phylogenetic analyses: *HLA-A* (Accession Number: NM_002116), *HLA-B* (NM_005514), *HLA-C* (NM_002117), *HLA-E* (NM_005516), and *HLA-F* (NM_018950) and *HLA-G* (NM_002127), *Mafa-A1* (LC043310), *Mafa-A2* (LC053847), *Mafa-A3* (LC043318), *Mafa-A4* (LC043319), *Mafa-B* (LC043350), *Mafa-I* (LC043377), *Mafa-E* (U02976), *Mafa-F* (DQ367725), *Mafa-AG* (HQ992797), and mouse H2-K1 (NM_001001892).

Table 2 112 distinct *Mafa-class I* alleles identified in 127 Filipino population samples

IMGT/NPH allele names	GenBankaccession number	Related MHC haplotype	Number of alleles[a]	Allele freq.	Ref.[b]
(A) *Mafa-A* alleles	*Mafa-A* haplotype				
A1*004:01	AB447581	A-Hp1	16	6.3 %	1, 2
A1*008:02	AB447588	A-Hp2.1	22	8.7 %	1, 2, 3
A1*008:05	LC043309	A-Hp2.2	3	1.2 %	
A1*018:04	AB447598	A-Hp6	14	5.5 %	2, 3
A1*038:02	AB447619	A-Hp9	8	3.1 %	2
A1*052:02	LC043310	A-Hp8.1	41	16.1 %	1
A1*052:04	LC043311	A-Hp8.2	1	0.4 %	
A1*071:02	LC053844	A-Hp10	15	5.9 %	4
A1*074:02	AB447606	A-Hp3.1, A-Hp3.3	14	5.5 %	2, 3
A1*074:03	LC043312	A-Hp3.2	2	0.8 %	
A1*086:02	LC053845	A-Hp2	25	9.8 %	Unpub.
A1*089:02	LC043313	A-Hp7.1	19	7.5 %	1, 3
A1*089:03	LC043314	A-Hp7.2	68	26.8 %	1, 3
A1*093:01	AB447597	A-Hp3.1, A-Hp3.2	13	5.1 %	1, 2
A1*093:02	AB447605	A-Hp3.3	3	1.2 %	2
A1*094:01	AB447584	A-Hp4	24	9.4 %	1, 2
A1*098:02	LC043315	A-Hp5.1, A-Hp5.2	6	2.4 %	Unpub
A1*098:03	LC043316	A-Hp5.3	1	0.4 %	
A2*05:04	LC043317	A-Hp4	24	9.4 %	5, 6
A2*05:56	LC053846	A-Hp7.1	19	7.5 %	
A2*05:50	LC053847	A-Hp7.2	67	26.4 %	Unpub
A2*24:02	AB447585	A-Hp5.1, A-Hp5.3	4	1.6 %	2
A2*24:06	LC053848	A-Hp5.2	3	1.2 %	
A3*13:03:01	LC043318	A-Hp6, A-Hp7	101	39.8 %	1, 3
A4*01:04	LC043319	A-Hp8	42	16.5 %	1
A4*14:03	LC043320	A-Hp9	8	3.1 %	Unpub
A4*14:14	LC053849	A-Hp10	14	5.5 %	
(B) A new *Mafa-A8* lineage allele					
A8*01:01	LC031882		39	15.4 %	
(C) *Mafa-B* alleles	*Mafa-B/I* haplotype				
B*007:01:01	LC043321	B-Hp4	30	11.8 %	7
B*007:01:02	LC043322	B-Hp5	20	7.9 %	1
B*017:01	LC043323	B-Hp6	23	9.1 %	1
B*021:03	LC043324	B-Hp15	1	0.4 %	
B*028:03	LC043325	B-Hp15	1	0.4 %	7
B*030:05:03	LC043326	B-Hp12	3	1.2 %	
B*030:14	LC043327	B-Hp11	3	1.2 %	
B*033:02	LC043328	B-Hp1, B-Hp13	69	27.2 %	1, 3
B*041:01	LC043330	B-Hp8.1, B-Hp9.1	10	3.9 %	1
B*041:01:02	LC043332	B-Hp8.2	1	0.4 %	
B*041:02	LC053850	B-Hp9.2	1	0.4 %	
B*044:03	LC043333	B-Hp12	3	1.2 %	Unpub
B*044:09	LC043334	B-Hp14	1	0.4 %	
B*045:05	LC043329	B-Hp3	28	11.0 %	
B*046:01:02	LC043335	B-Hp2	40	15.7 %	Unpub
B*048:03	LC043336	B-Hp9	5	2.0 %	1
B*050:08	LC043337	B-Hp2, B-Hp6	62	24.4 %	
B*056:02:01	LC043338	B-Hp6	23	9.1 %	1

Table 2 (continued)

IMGT/NPH allele names	GenBankaccession number	Related MHC haplotype	Number of alleles[a]	Allele freq.	Ref.[b]
B*057:04	LC043339	B-Hp2	40	15.7 %	
B*060:02	LC043340	B-Hp2, B-Hp6.3, B-Hp9	49	19.3 %	1, 3
B*060:03:01	LC043341	B-Hp6.1, B-Hp6.2	19	7.5 %	Unpub
B*060:03:03	LC043342	B-Hp15	1	0.4 %	
B*065:03	LC043331	B-Hp11	3	1.2 %	
B*068:08	LC043343	B-Hp10	5	2.0 %	
B*068:09	LC043344	B-Hp15	1	0.4 %	
B*070:05	LC043345	B-Hp5	20	7.9 %	
B*072:01	LC043346	B-Hp2	36	14.2 %	1
B*076:02	LC043347	B-Hp12	3	1.2 %	Unpub
B*079:02:02	LC043348	B-Hp4	18	7.1 %	
B*085:01	LC043363	B-Hp4	30	11.8 %	7
B*089:01:01	LC053851	B-Hp9	5	2.0 %	1
B*089:01:02	LC043349	B-Hp6, B-Hp11	25	9.8 %	1, 3
B*095:01	LC043350	B-Hp1	66	26.0 %	1, 3, 7
B*098:08	LC043352	B-Hp3.1, B-Hp4.1, B-Hp8, B-Hp13.1	47	18.5 %	
B*098:09	LC043353	B-Hp3.3, B-Hp4.2, B-Hp13.2	16	6.3 %	
B*098:11	LC053852	B-Hp3.2	1	0.4 %	
B*098:10	LC043354	B-Hp1	59	23.2 %	
B*091:02	LC043355	B-Hp10	5	2.0 %	
B*099:01	LC043356	B-Hp3, B-Hp13	31	12.2 %	1
B*101:01:02	LC043357	B-Hp8.1, B-Hp8.2	7	2.8 %	1
B*104:03	LC043358	B-Hp2	40	15.7 %	1, 3, 8
B*108:01	LC043359	B-Hp3	28	11.0 %	1
B*114:02	LC043360	B-Hp2	39	15.3 %	
B*115:04:02	LC043361	B-Hp5	20	7.9 %	
B*116:01	LC043362	B-Hp6	23	9.1 %	1
B*137:03	LC043364	B-Hp7	14	5.5 %	1, 3, 8
B*144:03N	LC043365	B-Hp2	40	15.7 %	1
B*157:01	LC043366	B-Hp6.1, B-Hp6.3	22	8.7 %	1, 3
B*157:03	LC043367	B-Hp6.2	1	0.4 %	
B*158:01	LC043368	B-Hp4	30	11.8 %	1
B*159:01	LC043369	B-Hp4	30	11.8 %	1
B*160:01	LC043370	B-Hp5	20	7.9 %	Unpub
B*161:04	LC043351	B-Hp11	3	1.2 %	
B*164:02	LC043371	B-Hp14	1	0.4 %	Unpub
(D) *Mafa-I* alleles	*Mafa-B/I* haplotype				
I*01:01:01	LC043378	B-Hp6	24	9.4 %	1, 9
I*01:27	LC043375	B-Hp8	7	2.8 %	
I*01:15:02	LC043376	B-Hp11	123	48.4 %	
I*01:10:02	LC043372	B-Hp3	123	48.4 %	
I*01:11	LC043377	B-Hp1, B-Hp5	123	48.4 %	1, 10
I*01:12:01	LC043379	B-Hp2	38	15.0 %	1
I*01:13:01	LC043380	B-Hp4	30	11.8 %	1
I*01:14	LC043381	B-Hp10	123	48.4 %	1, 9
I*01:28	LC053853	B-Hp14	1	0.4 %	
I*01:26	LC043373	B-Hp12	3	1.2 %	
I*01:18:02	LC043382	B-Hp9	14	5.5 %	
I*01:19	LC043383	B-Hp15	1	0.4 %	1

Table 2 (continued)

IMGT/NPH allele names	GenBankaccession number	Related MHC haplotype	Number of alleles[a]	Allele freq.	Ref.[b]
(E) *Mafa-E* alleles	*Mafa-E* haplotype				
E-like1	LC043384	E-Hp4	25	9.8 %	
E-like2	LC043385	E-Hp7	14	5.5 %	
E-like3	LC043386	E-Hp1	91	35.8 %	
E-like4	LC043387	E-Hp5.1, E-Hp8	22	8.7 %	
E-like5	LC043388	E-Hp2, E-Hp9	46	18.1 %	
E-like6	LC043389	E-Hp5.2	1	0.4 %	
E-like7	LC043390	E-Hp5.1	18	7.1 %	
E-like8	LC043391	E-Hp6	18	7.1 %	
E-like9	LC043392	E-Hp3, E-Hp6, E-Hp7.1, E-Hp9	66	26.0 %	
E-like10	LC043393	E-Hp1.1, E-Hp8	82	32.3 %	
E-like11	LC043394	E-Hp2	43	16.9 %	
(F) *Mafa-F* alleles					
F-like1	LC043395		66	26.0 %	
F-like2	LC043396		36	14.2 %	
F-like3	LC043397		19	7.5 %	
F-like4	LC043398		97	38.2 %	
F-like5	LC043399		1	0.4 %	
F-like6	LC043400		19	7.5 %	
F-like7	LC043401		16	6.3 %	

Bold letters indicate newly identified *Mafa-class I* allele sequences. The numbers of *Mafa-A8*01:01* was obtained by qPCR analysis using *Mafa-A8*01:01*-specific primers

[a] Numbers of *Mafa-class I* alleles after haplotyping of the 127 animals (Table 3)

[b] References quoted in the table: (1) Campbell et al. 2009, (2) Kita et al. 2009, (3) Lawrence et al. 2012, (4) Uda et al. 2004, (5) Otting et al. 2007, (6) Wu et al. 2008, (7) Krebs et al. 2005, (8) Pendley et al. 2008, (9) Urvater et al. 2000 and (10) Wiseman et al. 2007

Results

Sequence read information from 127 Filipino cynomolgus macaques

Mafa-class I cDNA amplified by using the specific primer set Class_I_F and Class_I_R (PCR product size: 514 bp or 517 bp) were sequenced by parallel pyrosequencing. Supplementary Table 1 shows the number of draft sequences reads (sequences passing quality control (QC) criteria) after base calling, generated by the pyrosequencing method using the manufacturer specifications for 127 Filipino population samples. Draft read numbers in total were 3,328,273 reads with a range of reads from 3,872 in the animal ID189 to 103,942 reads in the animal ID248 that were high-quality reads with more than 20 quality values (QVs) and an average QV of 33.6±3.0 in the high-quality sequence reads. The draft read bases in total were 1,217 Mb with a range between 1.4 Mb in ID189 and 36.2 Mb in ID248 (9.6±6.7 Mb on average), with an overall average read length of 365.7±24 bases and an overall median read length of 410±32 bases (ESM Table 1). Therefore, the sequence reads had sufficient high-quality and sequence volume for further genotyping analysis.

Mafa-class I *cDNA alleles*

From our initial characterization of the *Mafa-A*, *Mafa-B*, and *Mafa-I* allele sequences using 127 animals (exon 1 to exon 7) by the sub-cloning and Sanger sequencing methods, we detected in total 58 *Mafa-A*, *Mafa-B*, and *Mafa-I* alleles (39 known alleles and 19 novel alleles). After genotyping of *Mafa-class I* genes using 127 animals by pyrosequencing, we detected an additional 54 *Mafa-class I* alleles (17 known alleles and 37 novel alleles). Of the 37 novel alleles, 19 were *Mafa-A/B/I* alleles and the other 18 were *Mafa-E* or *Mafa-F* alleles. Finally, we characterized the 19 novel *Mafa-A/B/I* allele sequences using 26 animals (exon 1 to exon 7) by the sub-cloning and Sanger sequencing methods. Therefore, to summarize this procedure, we detected in total 112 alleles (56 known alleles and 56 novel alleles) using 127 animals, and of the 56 novel alleles we characterized 38 *Mafa-A*, *Mafa-B*, and *Mafa- I* allele sequences by the sub-cloning and Sanger sequencing methods. Of the 38 alleles, 19 were detected by the Sanger method before pyrosequencing and the other 19 were detected by the Sanger method after pyrosequencing.

Table 2 shows the number and frequencies of 112 distinct *Mafa-class I* alleles (28 *Mafa-A*, 54 *Mafa-B*, 12 *Mafa-I*, 11

Mafa-E, and 7 *Mafa-F*) that were identified in 127 Filipino cynomolgus macaques by mapping the sequence reads as 99 and 100 % matching parameters with all the known *Mafa-class I* allele sequences released in the IMGT/MHC-NHP database using the GS Reference Mapper Ver. 3.0. All of the *Mafa-A*, *Mafa-B*, and *Mafa-I* allele sequences were supported by conventional analytical methods, 56 were previously reported in IMGT/MHC-NHP database (Campbell et al. 2009; Kita et al. 2009; Krebs et al. 2005; Lawrence et al. 2012; Otting et al. 2007; Pendley et al. 2008; Uda et al. 2004; Urvater et al. 2000; Wiseman et al. 2007; Wu et al. 2008) and the other 56 (eight *Mafa-A*, 24 *Mafa-B*, six *Mafa-I*, 11 *Mafa-E*, and seven *Mafa-F*) were newly identified in this study (Table 2). All of the newly identified *Mafa-E* and *Mafa-F* alleles (tentatively named as *Mafa-E-like1~Mafa-E-like11* and *Mafa-F-like1~Mafa-F-like7*) have 94.7 to 99.6 % and 98.7 to 99.6 % nucleotide similarities among the 473-bp peptide-binding region, respectively (ESM Table 2). Overall, the number of *Mafa-class I* allele sequences per animal ranged from 8 to 27 in total with an average 18.9, and a range of alleles of 1 to 7 for *Mafa-A*, 1 to 14 for *Mafa-B*, 1 to 3 for *Mafa-I*, 1 to 4 for *Mafa-E*, and 1 to 2 for *Mafa-F*. The six most frequent alleles in the 127 animals were *Mafa-I*01:10/11/14/15* (48.4 %), *Mafa-A3*13:03:01* (39.8 %), *Mafa-F-like4* (38.2 %), *Mafa-E-like3* (35.8 %) and *Mafa-E-like10* (32.3 %) (Table 2).

Of the 112 alleles, 13 alleles were perfectly matched with previously reported alleles of the rhesus macaque (*Macaca mulatta*, *Mamu*) (Boyson et al. 1995; Boyson et al. 1996; Campbell et al. 2009; Karl et al. 2008; Muhl et al. 2002; Otting et al. 2007; Otting et al. 2005), the southern pig-tailed macaque (*Macaca nemestrina*, *Mane*) (Lafont et al. 2003; Lafont et al. 2004) and/or stump-tailed macaque (*Macaca arctoides*, *Maar*) (Urvater et al. 2000). These trans-species polymorphisms (Table 3) were probably already generated before speciation of cynomolgus macaques 2.4~4.2 million years ago (Hedges et al. 2006).

Haplotypes

Table 4 shows the numbers and frequencies of the *Mafa-A*, *-B/I*, and *-E* haplotypes estimated for 127 cynomolgus macaques at their different MHC loci. We sorted 17 *Mafa-A* haplotypes to four *Mafa-A* loci, 23 *Mafa-B/I* haplotypes to nine loci including locus *Mafa-I*, and 12 *Mafa-E* haplotypes to two loci. In regard to homozygous haplotypes, eight *Mafa-A* haplotypes (A-Hp1, A-Hp2.1, A-Hp2.2, A-Hp4, A-Hp6, A-Hp7.1, A-Hp7.2, and A-Hp8.1) were supported by 23 *Mafa-A* homozygous animals, seven *Mafa-B/I* haplotypes (B/I-Hp1, B/I-Hp2, B/I-Hp3.1, B/I-Hp4.1, B/I-Hp4.2, B/I-Hp5, and B/I-Hp7) were supported by 20 *Mafa-B/I* homozygous animals, and eight *Mafa-E* haplotypes (E-Hp1.1, E-Hp1.2, E-Hp2, E-Hp3, E-Hp4, E-Hp5.1, E-Hp6, and E-Hp7.1) were supported by 29 *Mafa-E* homozygous animals.

The three most frequent *Mafa-A* haplotypes were A-Hp7.2 (26.8 %), A-Hp8.1 (16.1 %), and A-Hp4 (9.4 %), while the least frequent haplotypes were A-Hp5.3 and A-Hp8.2 (0.4 %), haplotypes that are likely variants of A-Hp5.1 or A-Hp5.2, and A-Hp8.1, respectively (Table 4 (A)). In comparison, the four most frequent *Mafa-B/I* haplotypes were B/I-Hp1 (26.0 %), B/I-Hp2 (16.5 %), and B/I-Hp3.1 (8.3 %) and B/I-Hp4.1 (8.3 %), while the least frequent were B/I-Hp3.2, B/I-Hp6.2, B/I-Hp8.2, B/I-Hp9.2, B/I-Hp13.2, B/I-Hp14, and B/I-Hp15 (0.4 %). Of the least frequent, at least B/I-Hp3.2, B/I-Hp6.2, and B/I-Hp8.2 are possible variants of B/I-Hp3.1, B/I-Hp6.1, and B/I-Hp8.1, respectively, and B-Hp14 and B-Hp15 are rare haplotypes (Table 4 (B)). The two kinds of *Mafa-I* sequences observed in B/I-Hp6 were not observed in B/I-Hp9 and B/I-Hp13. The six *Mafa-B/I* haplotypes B-Hp1, B-Hp2, B-Hp4.1, B-Hp4.2, B-Hp6.3, and B-Hp8.1 were identified as consensus *Mafa-B/I* haplotypes, although some alleles such as *Mafa-B*072:01*, *Mafa-B*079:02:02*, *Mafa-B*089:01:02*, *Mafa-B*098:08*, *Mafa-B*098:10*, and *Mafa-I*01:12:01* were not observed in these six haplotypes in some animals. The absence of these alleles in some animals might be due to extremely low gene expression or the real absence of the actual allele. Of B/I-Hp2, a part of the *Mafa-B/I* alleles of the B/I-Hp2 composition were detected such as "*B*104:03 - B*144:03N - B*057:04 - B*060:02 - B*46:01:02*" in three animals and "*B*050:08 - B*114:02 - B*072:01*" in two animals. In addition, three recombinants between B/I-Hp3 and B/I-Hp4 and between B/I-Hp2 and B/I-Hp6 were observed in three animals. These structural variants may have been generated by either a genetic recombination or a gene duplication event in the Filipino population in relatively recent times.

The three most frequent *Mafa-E* haplotypes were E-Hp1.1 (31.9 %), E-Hp2 (17.3 %), and E-Hp3 (14.6 %), while the least frequent haplotype was E-Hp5.2 (0.4 %), a haplotype that is a possible variant of E-Hp5.1 (Table 4c). The *Mafa-E* genotyping identified one to four distinct *Mafa-E* alleles per animal. Thus, one or two *Mafa-E* loci are involved in the gene expression. Although there was no previous information for the *Mafa-E* haplotype (E-Hp) classification in cynomolgus macaques, we estimated 12 *Mafa-E* haplotypes by manually sorting of the observed alleles based on the *Mafa-A*, *Mafa-B/I*, and/or *Mafa-E* haplotype homozygous animals.

The number and frequency of the *Mafa* paired haplotypes estimated by the PHASE 2.1.1 program were 26 *Mafa-F—Mafa-A* haplotypes (F/A-Hp), 30 *Mafa-A—Mafa-E* haplotypes (A/E-Hp) and 49 *Mafa-E—Mafa-B* haplotypes (E/B-Hp), are shown in ESM Table 3. The six-haplotype pairs over 10 % frequencies were F/A-Hp1 (24.8 %), A/E-Hp1 (24.8 %), F/A-Hp2 (15.7 %), A/E-Hp2 (15.7 %), E/B-Hp1 (13.4 %), and E/B-Hp2 (13.0 %). Also, 84 *Mafa-F—Mafa-A—Mafa-E—Mafa-B* haplotypes (F/A/E/B-Hp) were estimated using

Table 3 *Mafa-class I* allele supporting the trans-species polymorphism

Mafa-class I allele	Identical macaque allele	Accession number	Ref. [a]
*Mafa-A1*052:02*	*Mamu-A1*052:02*	AM295917, EF580143	1, 2
*Mafa-A4*14:03*	*Mamu-A4*14:03:01*	AF519897, AY707077, EF580138, GU080236	2, 3
*Mafa-B*007:01:01*	*Mamu-B*007:03*	AJ556876, EF580149, EU682528	2, 4
*Mafa-B*041:02*	*Mamu-B*041:01*	AJ556892	4
*Mafa-B*044:03*	*Mamu-B*044:04*	AM902558	Unpub
*Mafa-B*050:08*	*Mamu-B*050:02*	AJ620415, FN396411	4
*Mafa-B*089:01:02*	*Mane-B*089:02*	GQ274890	Unpub
*Mafa-B*098:08*	*Mamu-B*098:02*	JX442253	Unpub
*Mafa-B*099:01*	*Mamu-B*099:01*	FN396416	Unpub
*Mafa-I*01:11*	*Mamu-I*01:08*	FJ009194	Unpub
	*Mamu-I*01:06:07*	FN396419, GQ471888	Unpub
	*Mamu-I*01:06:01*	EF580176	2
	*Maar-I*01:01*	AF161860	5
Mafa-E-like1	*Mamu-E*02:02*	FM986653, U02979	6
	*Mane-E*02:07*	AY204721	7, 8
Mafa-E-like2	*Mamu-E*02:05*	U41837	9
Mafa-E-like3	*Mane-E*02:01*	AY204714	7, 8
	*Mane-E*02:02*	AY204717	7, 8

[a] References quoted in the table: (1) Otting et al. 2007, (2) Karl et al. 2008, (3) Muhl et al. 2002, (4) Otting et al. 2005, (5) Urvater et al. 2000, (6) Boyson et al. 1995, (7) Lafont et al. 2003, (8) Lafont et al. 2004 and (9) Boyson et al. 1996

Mafa-F allele and *Mafa-A*, *Mafa-B*, and *Mafa-E* haplotype data (Supplementary Table 3). The probability of estimated *Mafa-class I* haplotypes in each animal ranged from 0.35 to 1.00 with average of 0.91. Figure 1 shows the five most frequent F/A/E/B haplotypes, F/A/E/B-Hp1 (10.6 %), F/A/E/B-Hp2 (10.2 %), F/A/E/B-Hp3 (5.5 %), F/A/E/B-Hp4 (3.9 %), and F/A/E/B-Hp5 (3.9 %) (Supplementary Table 4).

Characterization of a novel Mafa-class I locus, Mafa-A8*01:01

We first identified the novel *Mafa-class I* sequence *Mafa-A8*01:01*, which we categorized as a minor allele, during the de novo assembly of the sequence reads. The entire coding sequence was composed of 1,089 bp and 362 deduced amino acids. Although the amino acid sequence contains the structurally essential N-glycosylation sites (amino acid positions 86–88), it lacks one cysteine residue involved in disulfide bonding (amino acid position 101) while the three other cysteine residues are conserved.

We later confirmed the presence of this sequence in 37 of the 127 cynomolgus macaques by PCR based genotyping using a *Mafa-A8*01:01*-specific primer set (data not shown). Polymorphism of the gene was not detected in the 473-bp peptide-binding region (exons 2–4) of the 37 positive animals or in the 857 bp PCR region (exons 3–4) of nine Vietnamese and Indonesian cynomolgus macaques that were also found to have this sequence (Fig. 2a). Ten Vietnamese and ten Indonesian DNA samples were genotyped with the *Mafa-A8*01:01*-specific primer set, and PCR products corresponding to the gene were observed in three and five animals, respectively, suggesting that the gene has diffused among some different cynomolgus macaque species (Fig. 2a). The highest nucleotide and amino acid similarities of the coding sequences of *Mafa-A8*01:01* with known nonhuman primate *MHC* class I genes were 85.5 % with *Mamu-A1*011:03* (EF580154) and 76.9 % with *Mamu-A* (D5KRD5), respectively. In comparison, there was 83 to 84 % nucleotide identity with each of the *HLA* class I genes. However, although we tried the same primer set and PCR conditions in a PCR experiment using human DNA samples, no PCR products were obtained (data not shown). In addition, a phylogenetic tree supports the divergence of the gene from the other *Mafa* and *HLA* class I genes (Fig. 3).

In an attempt to determine the copy number of the *Mafa-A8*01:01* gene by qPCR analysis of 37 allele-positive cynomolgus macaques, two animals (ID165 and ID189) were detected to have two copies in contrast to only one copy in the other 35 animals. From the detailed linkage analysis of the copy numbers with *Mafa-F* alleles and all their haplotypes, *Mafa-A8*01:01* was linked to a genomic region between *Mafa-F* and *Mafa-A* (Fig. 2b, c). The gene was detected in 14 animals that have A-Hp6, and in 14 of 15 animals that have A-Hp10. In fact, A-Hp6 and A-Hp10 were observed in animal ID189, but only A-Hp6 was present in ID165 (Fig. 2b, c). However, *Mafa-A8*01:01* and *Mafa-A4*14:14* were missing

Table 4 *Mafa-A*, *Mafa-B/I*, and *Mafa-E* haplotypes estimated using 127 Filipino population samples

(A) Description of *Mafa-A* haplotypes

Haplotype name	1st *A1* locus	2nd *A1* locus	*A2* locus	*A3* locus	*A4* locus	Haplotype freq.
A-Hp1	A1*004:01					6.3 %
A-Hp2.1	A1*008:02	A1*086:02				8.7 %
A-Hp2.2	A1*008:05	A1*086:02				1.2 %
A-Hp3.1	A1*093:01	A1*074:02				4.3 %
A-Hp3.2	A1*093:01	A1*074:03				0.8 %
A-Hp3.3	A1*093:02	A1*074:02				1.2 %
A-Hp4	A1*094:01		A2*05:04			9.4 %
A-Hp5.1	A1*098:02		A2*24:02			1.2 %
A-Hp5.2	A1*098:02		A2*24:06			1.2 %
A-Hp5.3	A1*098:03		A2*24:02			0.4 %
A-Hp6	A1*018:04					5.5 %
A-Hp7.1	A1*089:02		A2*05:56	A3*13:03:01		7.5 %
A-Hp7.2	A1*089:03		A2*05:50	A3*13:03:01		26.8 %
A-Hp8.1	A1*052:02				A4*01:04	16.1 %
A-Hp8.2	A1*052:04				A4*01:04	0.4 %
A-Hp9	A1*038:02				A4*14:03	3.1 %
A-Hp10	A1*071:02				A4*14:14	5.9 %

(B) Description of *Mafa-B/I* haplotypes

Haplotype name	1st q*B* locus	2nd *B* locus	3rd *B* locus	4th *B* locus	5th *B* locus	6th *B* locus	7th *B* locus	8th *B* locus	*I* locus	Haplotype freq.
B/I-Hp1	B*095:01	B*033:02	B*098:10						I*01:11	26.0 %
B/I-Hp2	B*104:03	B*144:03N	B*057:04	B*060:02	B*046:01:02	B*050:08	B*114:02	B*072:01	I*01:12:01	16.5 %
B/I-Hp3.1	B*099:01	B*108:01	B*045:05	B*098:08					I*01:10:02	8.3 %
B/I-Hp3.2	B*099:01	B*108:01	B*045:05	B*098:11					I*01:10:02	0.4 %
B/I-Hp3.3	B*099:01	B*108:01	B*045:05	B*098:09					I*01:10:02	2.4 %
B/I-Hp4.1	B*159:01	B*007:01:01	B*158:01	B*085:01	B*079:02:02	B*098:08			I*01:13:01	8.3 %
B/I-Hp4.2	B*159:01	B*007:01:01	B*158:01	B*085:01	B*079:02:02	B*098:09			I*01:13:01	3.5 %
B/I-Hp5	B*160:01	B*007:01:02	B*070:05	B*115:04:02					I*01:11	7.9 %
B/I-Hp6.1	B*056:02:01	B*017:01	B*157:01	B*050:08	B*060:03:01	B*089:01:02	B*116:01		I*01:11 I*01:01:01	7.1 %
B/I-Hp6.2	B*056:02:01	B*017:01	B*157:03	B*050:08	B*060:03:01	B*089:01:02	B*116:01		I*01:11 I*01:01:01	0.4 %
B/I-Hp6.3	B*056:02:01	B*017:01	B*157:01	B*050:08	B*060:02	B*089:01:02	B*116:01		I*01:11 I*01:01:01	1.6 %
B/I-Hp7	B*137:03								I*01:18:02	5.5 %
B/I-Hp8.1	B*041:01	B*101:01:02	B*098:08						I*01:27	2.4 %
B/I-Hp8.2	B*041:01:02	B*101:01:02	B*098:08						I*01:27	0.4 %
B/I-Hp9.1	B*048:03	B*041:01	B*060:02	B*089:01:01					ND	1.6 %
B/I-Hp9.2	B*048:03	B*041:02	B*060:02	B*089:01:01					ND	0.4 %
B/I-Hp10	B*091:02	B*068:08							I*01:14	2.0 %
B/I-Hp11	B*065:03	B*161:04	B*030:14	B*089:01:02					I*01:15:02	1.2 %
B/I-Hp12	B*076:02	B*044:03	B*030:05:03						I*01:26	1.2 %
B/I-Hp13.1	B*099:01	B*033:02	B*098:08						ND	0.8$
B/I-Hp13.2	B*099:01	B*033:02	B*098:09						ND	0.4 %
B/I-Hp14	B*164:02	B*044:09							I*01:28	0.4 %
B/I-Hp15	B*028:03	B*068:09	B*021:03	B*060:03:03					I*01:19	0.4 %
recombinant										1.2 %

(C) Description of *Mafa-E* haplotypes

Haplotype name	1st *E* locus	2nd *E* locus	Haplotype freq.
E-Hp1.1	E-like3	E-like10	31.9 %
E-Hp1.2	E-like3		3.9 %
E-Hp2	E-like5	E-like11	17.3 %

Table 4 (continued)

E-Hp3	E-like9		14.6%
E-Hp4	E-like1		9.8%
E-Hp5.1	E-like4	E-like7	7.9%
E-Hp5.2	E-like6		0.4%
E-Hp6	E-like8	E-like9	7.1%
E-Hp7.1	E-like2	E-like9	3.5%
E-Hp7.2	E-like2		2.0%
E-Hp8	E-like4	E-like10	0.8%
E-Hp9	E-like5	E-like9	0.8%

The *Mafa-A*, *Mafa-B*, and *Mafa-E* haplotypes were characterized by manually sorting of *Mafa-A*, *Mafa-B*, *Mafa-I*, and *Mafa-E* alleles. Bold letters indicate *Mafa-class I* haplotypes observed in *Mafa-A*, *Mafa-B*, and/or *Mafa-E* homozygous animals. All other haplotypes presented here were deduced from heterozygous animals. All of the 254 haplotypes derived from the 127 animals, excluding three *Mafa-B/I* recombinants, were classified to one each in the *Mafa-A*, *Mafa-B*, and *Mafa-E* haplotypes. *ND* indicates "not detected". A newly identified *Mafa-A8*01:01* allele is not shown in this table

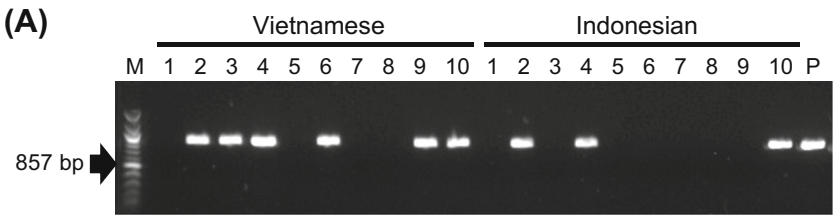
from one animal (ID181) with A-Hp10 (Fig. 2c). Therefore, *Mafa-A8*01:01* and *Mafa-A4*14:14* may have been segmentally deleted from A-Hp10 in relatively recent times.

Discussion

One major advantage of genotyping the macaque MHC by the NGS pyrosequencing method over the conventional Sanger sequencing method was the use of MID-tagged adaptors or PCR primers to identify individual amplicons that can be sequenced simultaneously with an arbitrarily chosen coverage. However, despite its increase in speed and sensitivity, this new sequencing technology is error-prone and poses considerable challenges because it can be more difficult to discriminate between sequencing errors and true rare alleles due to the complex nature of the generated artifacts and errors that require an efficient and accurate quality control (Babik et al. 2009). In fact, 42.0 % of the 3,328,276 draft sequence reads were unavailable for analysis due to various reasons such as short read lengths and incomplete extension, mixed reads, artificial substitutions, insertion, and deletion (indel) errors, including homopolymer errors (ESM Table 1). Nevertheless, the current pyrosequencing method produced 15,205 mapped sequence reads per animal (range 2,683 to 65,382) that were sufficient for effective genotyping and enabled the detection of all of the *Mafa-class I* alleles in 127 animals including the alleles expressed at extremely low levels. However, all of the known and novel *Mafa-class I* alleles detected in this study depended on using two different matching parameters (99 and 100 %) as part of an efficient mapping protocol. Both of these matching parameters were necessary to complete the allele assignment and consequently for genotyping and haplotyping of the *Mafa-class I* alleles with precision.

In this study, 1 to 3 *Mafa-A* and 2 to 7 *Mafa-B* classical class I alleles were detected per haplotype. These haplotypic allele numbers are consistent with a previous report that 5 of 13 classical class I alleles in homozygous Mauritian macaque were expressed at relatively high levels (Aarnink et al. 2011). In addition, we found that the *Mafa-E* region is organized and expressed in the peripheral white blood cells by at least three major *Mafa-E* alleles and eight minor *Mafa-E* alleles that have different amino acid sequences and variable *Mafa-E* haplotype structures (Tables 2 and 4). Although *HLA-E* has very limited polymorphism, it serves as the ligand for the inhibitory NKG2A receptor expressed by NK cells (Lee et al. 1998). Therefore, *Mafa-E* and *Mamu-E*, because of their increased polymorphism and haplotype diversity, may have a different function to *HLA-E*. On the other hand, the seven *Mafa-F* alleles were well conserved in each other. This characterization is similar to *HLA-F*, which was recently identified to be one of the ligands for NK cell Ig-like receptors (Goodridge et al. 2013). These between species comparisons suggest that more comprehensive *MHC* genotyping and functional studies are still required for cynomolgus macaques to provide better insight into the diversity and the role of the MHC genes in this species.

As part of this study, we discovered a novel *Mafa-class I* lineage, *Mafa-A8*01:01*, and linked it to the *Mafa-A* locus by comparative analysis of the different *Mafa-class I* haplotypes. *Mafa-A8*01:01* was linked most strongly with A-Hp6 and A-Hp10, but it was also detected in 5.7 % of A-Hp7 and 9.5 % of A-Hp8 (Fig. 2b and c). This allele was also observed in 30 and 50 % of Vietnamese and Indonesian cynomolgus macaques, respectively, that we examined (Fig. 2a). The *MHC-A* region was probably formed by birth and death evolution (Piontkivska and Nei 2003) such as by repeated gene duplication and deletion events in human and non-human primates (Kono et al. 2014; Shiina et al. 2006) as shown in the detailed segmental and phylogenetic sequence analysis of the rhesus macaque *MHC* alpha block (Kulski et al. 2004). In the *HLA-A* region, an approximately 70 kb insertion and 50 kb deletion near the *HLA-A* gene were detected in some *HLA-A* allele lineages (Watanabe et al. 1997). The genomic rearrangements seen in our study are the likely segmental deletion of *Mafa-A8*01:01* and *Mafa-A4*14:14* in A-Hp10 (Fig. 2), and the presence of *Mafa-A8*01:01* in A-Hp7 and A-Hp8 possibly could have been generated by recombinational crossing over



(B)

Sample ID	Mafa-F	Mafa-A8	Mafa-A		Mafa-E	Mafa-B
148	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp13
155	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp9	B-Hp4
163	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp2	B-Hp4
165	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp4
165	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp4
175	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp4
182	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp5
188	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp13
189	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp6	B-Hp4
203	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp4
217	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp2	B-Hp1
242	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp6	B-Hp4
249	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp5
B354	F-like4	A8*01:01	A3*13:03:01	A1*018:04	E-Hp1	B-Hp1
			A-Hp6			

(C)

Sample ID	Mafa-F	Mafa-A8	Mafa-A		Mafa-E	Mafa-B
167	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
170	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp4
176	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp1
189	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp6
198	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
202	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
209	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
212	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
213	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
230	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
231	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp1
239	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
246	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp6
B610	F-like4	A8*01:01	A4*14:14	A1*071:02	E-Hp3	B-Hp5
181	F-like4	ND	ND	A1*071:02	E-Hp3	B-Hp5
			A-Hp10			

Fig. 2 Detection of *Mafa-A8*01:01* locus in Vietnamese and Indonesian macaques (**a**) and in the *Mafa-class I* haplotype structures A-Hp6 (**b**) and A-Hp10 (**c**). The electrophoresis images in (**a**) show the PCR products from ten Vietnamese and ten Indonesian genomic DNA samples using *Mafa-A8*01:01*-specific primers. Numbers on upper side indicate genomic DNA samples. *M* and *P* indicate bands of the 100 bp DNA size marker ladder and a Filipino population DNA sample (ID152) used as a positive control, respectively. The *dark background* in (**b**) and (**c**) indicates the *Mafa-A* alleles that are composed of A-Hp6 and A-Hp10, respectively. *Light gray background* indicates *Mafa-F* alleles, *Mafa-E*, and *Mafa-B* haplotypes that link to A-Hp6 or A-Hp10. The gene order of the *Mafa-A* alleles and the exact location of *Mafa-A8*01:01* are estimated by haplotype sorting. *ND* means "not detected".

and gene conversion. Phylogenetic analysis suggests that the *Mafa-A8*01:01* has diverged from the *Mafa-A1-A4* lineage (Fig. 3), and that, in comparison to the *Mafa-A1-A4* allele lineage, one of the structurally essential cysteine residues of *Mafa-A8*01:01* was replaced by a serine at amino acid position 101. In addition, polymorphism analysis of *Mafa-A8*01:01* suggests that this allele is well conserved in various haplotypes among different macaque populations, and that it might have a function different to those of other classical class I genes. However, because the *Mafa-A8*01:01* allele does not express a classical α2 domain it might be considered to be a new type of non-classical *MHC* gene whose function has yet to be determined. Therefore, this allele warrants functional analysis and consideration when matching MHC alleles in macaque transplantation studies.

Because the results of biomedical experiments strongly depend on the immunogenetic background of animals conditioned by various environmental selective factors such as pathogens, *MHC* homozygous macaques are preferred for use in biomedical research (Vallender 2014). In recruiting macaques for biomedical studies, the origin of the animals and their genetic polymorphisms need to be considered carefully

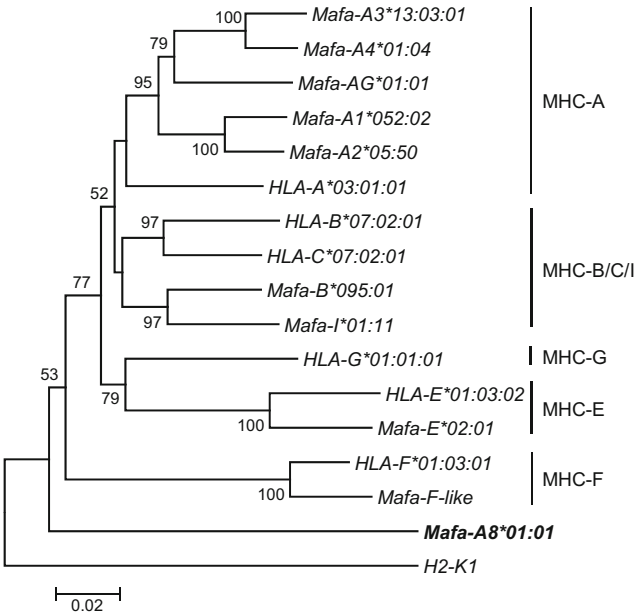


Fig. 3 Nucleotide sequence based phylogenetic tree using representative *Mafa-class I* alleles constructed by the neighbor-joining method. *Numbers at branches* indicate bootstrap values.

at the population level. In this regard, one of our Filipino animals (ID152) was strictly a "near-homozygote" that has the F/A/E/B-Hp1 in the *Mafa-class I* region and the #7 haplotype in the *Mafa-class II* region (Blancher et al. 2014) on both chromosomes (tentatively named "HT1") (Fig. 4). The HT1 potentially has only five classical class I alleles in the Filipino *Mafa-A* and *Mafa-B* genes which is half the number of alleles previously observed in the highly frequent classical class I haplotypes of Mauritian macaques (Aarnink et al. 2011; Budde et al. 2010). However, since only one copy of *Mafa-A8*01:01* was detected by qPCR analysis (data not shown) it is likely that this allele is located on only one of the two HT1 chromosomes. Nevertheless, this suggests that it is possible to detect *MHC* homozygotes in the Filipino macaque population as well as the Mauritius population, which is the preferred population group for recruitment of animals into biomedical studies. In the class II region, the #7 and #11 are highly frequent haplotypes in Filipino macaques with 30.6 and 13.9 %, respectively (Blancher et al. 2014). This suggests that *MHC* homozygous animals might be easily detected by large scale screening by focusing on F/A/E/B-Hp1, F/A/E/B-Hp2, #7 and #11 in the Filipino macaque population. In this regard, it can be expected that the *Mafa-A8*01:01* negative F/A/E/B-Hp1 and F/A/E/B-Hp2 will be preferentially selected for the simple reason that 90~95 % of these haplotypes are *Mafa-A8*01:01* negative.

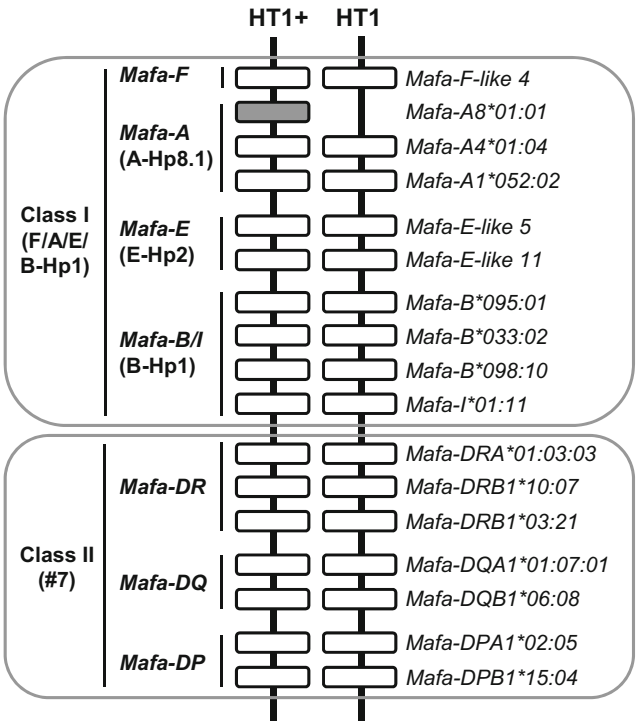


Fig. 4 Gene organization of the *MHC* homozygous-like animal ID152. This animal is a *MHC* homozygous-like animal because it has both of F/A/E/B-Hp1 and #7 haplotype (HT1) in one chromosome and F/A/E/B-Hp1, #7 and *Mafa-A8*01:01* haplotype (HT1+) in the other chromosome. *White and gray boxes* show common alleles between HT1 and HT1+, and *Mafa-A8*01:01*, respectively.

Acknowledgments We are grateful to Dr. Atsushi Toyoda at Center for Information Biology, Comparative Genomics Laboratory, National Institute of Genetics for making sequence data and Mr. Masayuki Tanaka, Mr. Hideki Hayashi, and Mr. Tadayuki Sato at the Education and Research Support Center, Research and Development Division, Tokai University for their technical support. This work was supported by MEXT KAKENHI (no. 221S0002), JSPS KAKENHI (no. 21300155), and was partially supported by the program "Research Center Network for Realization of Regenerative Medicine" in Japan Agency for Medical Research and Development (AMED) and from Adaptable and Seamless Technology transfer Program in AMED.

Conflict of interest The authors declare that they have no conflict of interest.

References

Aarnink A, Apoil PA, Takahashi I, Osada N, Blancher A (2011) Characterization of MHC class I transcripts of a Malaysian cynomolgus macaque by high-throughput pyrosequencing and EST libraries. Immunogenetics 63:703–13

Anzai T, Shiina T, Kimura N, Yanagiya K, Kohara S, Shigenari A, Yamagata T, Kulski JK, Naruse TK, Fujimori Y, Fukuzumi Y, Yamazaki M, Tashiro H, Iwamoto C, Umehara Y, Imanishi T, Meyer A, Ikeo K, Gojobori T, Bahram S, Inoko H (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. Proc Natl Acad Sci U S A 100:7708–13

Arikata M, Itoh Y, Okamatsu M, Maeda T, Shiina T, Tanaka K, Suzuki S, Nakayama M, Sakoda Y, Ishigaki H, Takada A, Ishida H, Soda K, Pham VL, Tsuchiya H, Nakamura S, Torii R, Shimizu T, Inoko H, Ohkubo I, Kida H, Ogasawara K (2012) Memory immune responses against pandemic (H1N1) 2009 influenza virus induced by a whole

particle vaccine in cynomolgus monkeys carrying Mafa-A1*052:02. PLoS One 7:e37220

Babik W, Taberlet P, Ejsmond MJ, Radwan J (2009) New generation sequencers as a tool for genotyping of highly polymorphic multilocus MHC system. Mol Ecol Resour 9:713–9

Blancher A, Tisseyre P, Dutaur M, Apoil PA, Maurer C, Quesniaux V, Raulf F, Bigaud M, Abbal M (2006) Study of Cynomolgus monkey (*Macaca fascicularis*) MhcDRB (Mafa-DRB) polymorphism in two populations. Immunogenetics 58:269–82

Blancher A, Bonhomme M, Crouau-Roy B, Terao K, Kitano T, Saitou N (2008) Mitochondrial DNA sequence phylogeny of 4 populations of the widely distributed cynomolgus macaque (*Macaca fascicularis fascicularis*). J Hered 99:254–64

Blancher A, Aarnink A, Tanaka K, Ota M, Inoko H, Yamanaka H, Nakagawa H, Apoil PA, Shiina T (2012) Study of cynomolgus monkey (*Macaca fascicularis*) Mhc DRB gene polymorphism in four populations. Immunogenetics 64:605–14

Blancher A, Aarnink A, Yamada Y, Tanaka K, Yamanaka H, Shiina T (2014) Study of MHC class II region polymorphism in the Filipino cynomolgus macaque population. Immunogenetics 66:219–30

Boyson JE, McAdam SN, Gallimore A, Golos TG, Liu X, Gotch FM, Hughes AL, Watkins DI (1995) The MHC E locus in macaques is polymorphic and is conserved between macaques and humans. Immunogenetics 41:59–68

Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI (1996) The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. J Immunol 156:4656–65

Budde ML, Wiseman RW, Karl JA, Hanczaruk B, Simen BB, O'Connor DH (2010) Characterization of Mauritian cynomolgus macaque major histocompatibility complex class I haplotypes by high-resolution pyrosequencing. Immunogenetics 62:773–80

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 61:177–87

Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE (2004) Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Res 14:1501–15

de Groot NG, Otting N, Robinson J, Blancher A, Lafont BA, Marsh SG, O'Connor DH, Shiina T, Walter L, Watkins DI, Bontrop RE (2012) Nomenclature report on the major histocompatibility complex genes and alleles of Great Ape, Old and New World monkey species. Immunogenetics 64:615–31

Ericsen AJ, Starrett GJ, Greene JM, Lauck M, Raveendran M, Deiros DR, Mohns MS, Vince N, Cain BT, Pham NH, Weinfurter JT, Bailey AL, Budde ML, Wiseman RW, Gibbs R, Muzny D, Friedrich TC, Rogers J, O'Connor DH (2014) Whole genome sequencing of SIV-infected macaques identifies candidate loci that may contribute to host control of virus replication. Genome Biol 15:478

Fuentes A, Gumert MD, Jones-Engel L (2011) Monkeys on the edge : ecology and management of long-tailed macaques and their interface with humans. Cambridge University Press, Cambridge, UK ; New York

Gardner MB, Luciw PA (2008) Macaque models of human infectious disease. ILAR J 49:220–55

Goodridge JP, Burian A, Lee N, Geraghty DE (2013) HLA-F and MHC class I open conformers are ligands for NK cell Ig-like receptors. J Immunol 191:3553–62

Hedges SB, Dudley J, Kumar S (2006) TimeTree: a public knowledge-base of divergence times among organisms. Bioinformatics 22:2971–2

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

Karl JA, Wiseman RW, O'Connor DH (2009) Cost-effective sequence-based nonhuman primate MHC class I genotyping from RNA. Methods 49:11–7

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

Kita YF, Hosomichi K, Kohara S, Itoh Y, Ogasawara K, Tsuchiya H, Torii R, Inoko H, Blancher A, Kulski JK, Shiina T (2009) MHC class I A loci polymorphism and diversity in three Southeast Asian populations of cynomolgus macaque. Immunogenetics 61:635–48

Kita YF, Ando A, Tanaka K, Suzuki S, Ozaki Y, Uenishi H, Inoko H, Kulski JK, Shiina T (2012) Application of high-resolution, massively parallel pyrosequencing for estimation of haplotypes and gene expression levels of swine leukocyte antigen (SLA) class I genes. Immunogenetics 64:187–99

Klimanskaya I, Chung Y, Becker S, Lu SJ, Lanza R (2006) Human embryonic stem cell lines derived from single blastomeres. Nature 444:481–5

Kono A, Brameier M, Roos C, Suzuki S, Shigenari A, Kametani Y, Kitaura K, Matsutani T, Suzuki R, Inoko H, Walter L, Shiina T (2014) Genomic sequence analysis of the MHC class I G/F segment in common marmoset (*Callithrix jacchus*). J Immunol 192:3239–46

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 175:5230–9

Kulski JK, Shiina T, Anzai T, Kohara S, Inoko H (2002) Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man. Immunol Rev 190:95–122

Kulski JK, Anzai T, Shiina T, Inoko H (2004) Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. Mol Biol Evol 21:2079–91

Lafont BA, Buckler-White A, Plishka R, Buckler C, Martin MA (2003) Characterization of pig-tailed macaque classical MHC class I genes: implications for MHC evolution and antigen presentation in macaques. J Immunol 171:875–85

Lafont BA, Buckler-White A, Plishka R, Buckler C, Martin MA (2004) Pig-tailed macaques (*Macaca nemestrina*) possess six MHC-E families that are conserved among macaque species: implication for their binding to natural killer receptor variants. Immunogenetics 56:142–54

Lawrence J, Orysiuk D, Prashar T, Pilon R, Fournier J, Rud E, Sandstrom P, Plummer FA, Luo M (2012) Identification of 23 novel MHC class I alleles in cynomolgus macaques of Philippine and Philippine/Mauritius origins. Tissue Antigens 79:306–7

Lee N, Llano M, Carretero M, Ishitani A, Navarro F, Lopez-Botet M, Geraghty DE (1998) HLA-E is a major ligand for the natural killer inhibitory receptor CD94/NKG2A. Proc Natl Acad Sci U S A 95:5199–204

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 63:529–37

Muhl T, Krawczak M, Ten Haaft P, Hunsmann G, Sauermann U (2002) MHC class I alleles influence set-point viral load and survival time in simian immunodeficiency virus-infected rhesus monkeys. J Immunol 169:3438–46

O'Leary CE, Wiseman RW, Karl JA, Bimber BN, Lank SM, Tuscher JJ, O'Connor DH (2009) Identification of novel MHC class I sequences in pig-tailed macaques by amplicon pyrosequencing and full-length cDNA cloning and sequencing. Immunogenetics 61:689–701

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 102:1626–31

Otting N, de Vos-Rouweler AJ, Heijmans CM, de Groot NG, Doxiadis GG, Bontrop RE (2007) MHC class I A region diversity and polymorphism in macaque species. Immunogenetics 59:367–75

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 60:339–51

Piontkivska H, Nei M (2003) Birth-and-death evolution in primate MHC class I genes: divergence time estimates. Mol Biol Evol 20:601–9

Robinson J, Mistry K, McWilliam H, Lopez R, Parham P, Marsh SG (2011) The IMGT/HLA database. Nucleic Acids Res 39:D1171–6

Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG (2013) IPD—the Immuno Polymorphism Database. Nucleic Acids Res 41:D1234–40

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–25

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

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3:1101–8

Shiina T, Tamiya G, Oka A, Takishima N, Inoko H (1999) Genome sequencing analysis of the 1.8 Mb entire human MHC class I region. Immunol Rev 167:193–9

Shiina T, Inoko H, Kulski JK (2004) An update of the HLA genomic region, locus information and disease associations: 2004. Tissue Antigens 64:631–49

Shiina T, Ota M, Shimizu S, Katsuyama Y, Hashimoto N, Takasu M, Anzai T, Kulski JK, Kikkawa E, Naruse T, Kimura N, Yanagiya K, Watanabe A, Hosomichi K, Kohara S, Iwamoto C, Umehara Y, Meyer A, Wanner V, Sano K, Macquin C, Ikeo K, Tokunaga K, Gojobori T, Inoko H, Bahram S (2006) Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. Genetics 173: 1555–70

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

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–89

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 131:861–72

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–9

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

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 56:155–63

Urvater JA, Otting N, Loehrke JH, Rudersdorf R, Slukvin II, Piekarczyk MS, Golos TG, Hughes AL, Bontrop RE, Watkins DI (2000) Mamu-I: a novel primate MHC class I B-related locus with unusually low variability. J Immunol 164:1386–98

Vallender EJ (2014) Bringing non-human primate research into the post-genomic era: how monkeys are teaching us about elite controllers of HIV/AIDS. Genome Biol 15:507

Vallender EJ, Miller GM (2013) Nonhuman primate models in the genomic era: a paradigm shift. ILAR J 54:154–65

Watanabe Y, Tokunaga K, Geraghty DE, Tadokoro K, Juji T (1997) Large-scale comparative mapping of the MHC class I region of predominant haplotypes in Japanese. Immunogenetics 46:135–41

Watanabe A, Shiina T, Shimizu S, Hosomichi K, Yanagiya K, Kita YF, Kimura T, Soeda E, Torii R, Ogasawara K, Kulski JK, Inoko H (2007) A BAC-based contig map of the cynomolgus macaque (*Macaca fascicularis*) major histocompatibility complex genomic region. Genomics 89:402–12

Wegner KM (2009) Massive parallel MHC genotyping: titanium that shines. Mol Ecol 18:1818–20

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 81:349–61

Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, Detmer AM, Bouffard P, Levenkova N, Turcotte CL, Szekeres E Jr, Wright C, Harkins T, O'Connor DH (2009) Major histocompatibility complex genotyping with massively parallel pyrosequencing. Nat Med 15:1322–6

Wu J, Bassinger S, Montoya GD, Chavez L, Jones CE, Holder-Lockyer B, Masten B, Williams TM, Prilliman KR (2008) Allelic diversity within the high frequency Mamu-A2*05/Mane-A2*05 (Mane-A*06)/Mafa-A2*05 family of macaque MHC-A loci. Tissue Antigens 72:29–38