MHC class I allele diversity in cynomolgusmacaques of Vietnamese origin (#37873)

First submission

Guidance from your Editor

Please submit by **11 Jul 2019** for the benefit of the authors (and your $200 publishing discount).

Structure and Criteria
Please read the 'Structure and Criteria' page for general guidance.

Custom checks
Make sure you include the custom checks shown below, in your review.

Raw data check
Review the raw data. Download from the materials page.

Image check
Check that figures and images have not been inappropriately manipulated.

Privacy reminder: If uploading an annotated PDF, remove identifiable information to remain anonymous.

Files

Download and review all files from the materials page.

5 Figure file(s)
6 Table file(s)
1 Other file(s)

Custom checks

DNA data checks

- Have you checked the authors data deposition statement?
- Can you access the deposited data?
- Has the data been deposited correctly?
- Is the deposition information noted in the manuscript?

Field study

- Have you checked the authors field study permits?
- Are the field study permits appropriate?

Structure and Criteria

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- Clear, unambiguous, professional English language used throughout.
- Intro & background to show context. Literature well referenced & relevant.
- Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- Figures are relevant, high quality, well labelled & described.
- Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- Original primary research within Scope of the journal.
- Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- Rigorous investigation performed to a high technical & ethical standard.
- Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- Impact and novelty not assessed. Negative/inconclusive results accepted. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- All underlying data have been provided; they are robust, statistically sound, & controlled.
- Speculation is welcome, but should be identified as such.
- Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	*Example*
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

MHC class I allele diversity in cynomolgusmacaques of Vietnamese origin

Shuting Huang [Equal first author, 1], **Xia Huang** [Equal first author, 1], **Shuang Li** [1], **Mingjun Zhu** [1], **Min Zhuo** [Corresp. 1]

[1] School of Biology and Biological Engineering, South China University of Technology, Guangzhou, Guangdong, China

Corresponding Author: Min Zhuo
Email address: zhuomin@scut.edu.cn

Many human diseases and treatment outcomes are influenced by variable immune responses caused by the diversity of major histocompatibility complex (MHC) alleles. Thanks to their genetic similarity tohumans, cynomolgus macaque (*Macaca fascicularis*, *Mafa*) have been used as an important experimental animal model for carrying out biomedical research es. However, there is much less information available on the polymorphism of MHC class I genes in cynomolgus macaques, than is currently available for humans. Consequently, we have identified 40 *Mafa-A* and 60 *Mafa-B* exons 2 and 3 sequences from 30 unrelated cynomolgus macaques of Vietnamese origin. Among these alleles, 28 are novel. As for the remaining 72 known alleles, 15 alleles are shared with other cynomolgus macaque populations and 32are identical to alleles previously reported in other macaque species. A potential recombination event was observed between *Mafa-A1*091:02* and *Mafa-A1*057:01*. In addition, the *Mafa-A1* genes was foundto be more diverse than human *HLA-A* and the functional residues for peptide binding sites (PBS) or TCR binding sites (TBS) in *Mafa-A1* have greater variability than that for non-PBS or non-TBS regions. Overall, this study provides important information on the diversity of *Mafa-A* and *Mafa-B* alleles from Vietnamese origin, which may benefit biomedical researches using cynomolgus macaque.

Article category：Original Manuscript

Title：**MHC class I allele diversity in cynomolgus macaques of Vietnamese origin**

Running Head：MHC class I allele diversity in macaques

Authors：

Shuting Huang[1, §], Xia Huang[1, §], [1], Shuang Li[1], Mingjun Zhu[1], Min Zhuo[1, *]

[1] School of Biology and Biological Engineering, South China University of Technology, Guangzhou 510006, China

* Corresponding author.

Phone number: +86-13560121763

Fax:86-20-39380601

E-mail address: zhuomin@scut.edu.cn

[§] Shuting Huang and Xia Huang contributed equally to this article.

Abstract

Many human diseases and treatment outcomes are influenced by variable immune responses caused by the diversity of major histocompatibility complex (MHC) alleles. Thanks to their genetic similarity to humans, cynomolgus macaque (*Macaca fascicularis*, *Mafa*) have been used as an important experimental animal model for carrying out biomedical researches. However, there is much less information available on the polymorphism of MHC class I genes in cynomolgus macaques, than is currently available for humans. Consequently, we have identified 40 *Mafa-A* and 60 *Mafa-B* exons 2 and 3 sequences from 30 unrelated cynomolgus macaques of Vietnamese origin. Among these alleles, 28 are novel. As for the remaining 72 known alleles, 15 alleles are shared with other cynomolgus macaque populations and 32 are identical to alleles previously reported in other macaque species. A potential recombination event was observed between *Mafa-A1*091:02* and *Mafa-A1*057:01*. In addition, the *Mafa-A1* genes was found to be more diverse than human *HLA-A* and the functional residues for peptide binding sites (PBS) or TCR binding sites (TBS) in *Mafa-A1* have greater variability than that for non-PBS or non-TBS regions. Overall, this study provides important information on the diversity of *Mafa-A* and *Mafa-B* alleles from Vietnamese origin, which may benefit biomedical researches using cynomolgus macaque.

Keywords: diversity; major histocompatibility complex; cynomolgus macaque

Introduction

The MHC glycoproteins, usually known as MHC class I and class II molecules, play important roles in the regulation of innate and adaptive immune response. The MHC classical class I molecules contribute both to innate immunity, by engaging Natural Killer (NK) cell receptors, and to adaptive immunity, by presenting antigens to $CD8^+$ T cells to induce their activation and cytotoxicity. Correlating with these functions, the antigen-binding sites, which are mostly located throughout the α1 and α2 domains encoded by the exons 2 and 3, exhibit the highest polymorphism within the full length of MHC class I gene sequences. Scholars found that the polymorphism of MHC genes is generated by a combination of mutation, recombination, and gene duplication and loss, and is maintained over time by selection [1-3]. Due to the fact that numerous infectious and autoimmune diseases are strongly associated with particular MHC alleles and haplotypes, the diversity of MHC class I genes have received particular attention [4-8].

The cynomolgus macaque and the rhesus macaque (*Macaca mulatta*, *Mamu*), are both important nonhuman primate animal models for the study of various human diseases such as acquired immunodeficiency syndrome, tuberculosis, Alzheimer's disease, Parkinson's disease, diabetes, as well as transplantation researches and pharmacodynamic evaluation [9-14]. Since the export of rhesus macaques from India was restricted in 1978, the use of cynomolgus macaques for biomedical research has become increasingly prevalent. This species inhabit widely throughout Southeast Asia, including the Philippines, Indonesia, Vietnam, Malaysia, Thailand, Cambodia, and Brunei [15]. In addition, it also had been introduced to Mauritius island located in the western Indian Ocean about 400 years ago, where the Mauritius cynomolgus macaque had become an insular population [16]. Previous studies have demonstrated that most MHC class I alleles found in cynomolgus macaques are unique to animals from particular regions [17-19]. Hence, the information on MHC diversity from different regions as well as their association with various disease susceptibility need to be concerned, when cynomolgus macaques from different regions were used in some similar biomedical studies.

The human classical MHC class I genes, *HLA-A*, *HLA-B*, and *HLA-C*, exhibit high polymorphism. 5018 *HLA-A* and 6096 *HLA-B* alleles have been included in the IMGT/HLA database (Release 3.36.0, 2019-04-17), which is a module of The Immuno Polymorphism Database (IPD) [20]. In comparison, the orthologues of the *HLA-C* gene have not been identified so far in macaques and the *MHC-A* and *MHC-B* genes in macaques have more complex organization than the human genes [21-23]. Only one copy of the *HLA-A* and *HLA-B* genes are present in humans, whilst seven A-like genes and up to nine B-like genes are present in macaques. Furthermore, a novel gene locus, *Mafa-A8*01:01*, was discovered recently in cynomolgus macaque of Filipino origin. Both the *MHC-A* and *MHC-B* loci are duplicated in cynomolgus macaque during evolution. Nevertheless, only 375 *Mamu-A* and 513 *Mamu-B* alleles, along with 494 *Mafa-A* and 717 *Mafa-B* alleles, have been deposited in the IPD-MHC database (Release 3.2.0.0.) [22,24]. In comparison to their human counterparts, the number and the detailed information on the polymorphism of MHC classic class I genes in macaques are still lacking. The cynomolgus macaques bred in South China mainly originated from Vietnam and have been exported to various places in the world for biomedical research. To better understand the characteristics of their MHC class I alleles, the polymorphism analysis of both the *Mafa-A* and the *Mafa-B* exons 2 and 3 sequences were carried out simultaneously in 30 unrelated animals.

Materials and methods

Animals

All cynomolgus monkeys were housed in the South China Primate Research & Development Center (Guangdong, China) and were clinically asymptomatic for known diseases. Peripheral blood samples were collected from over 30 unrelated Vietnamese-origin cynomolgus macaques. The experiments were reviewed and approved by the Institutional Animal Care and Use

Committee (IACUC) of Guangdong Landau Biotechnology Co. Ltd. (project number: IACUC-003).

RNA extraction, cDNA cloning and sequencing of MHC class I genes

Total RNA was extracted from peripheral blood mononuclear cell samples of 30 animals using E.Z.N.A.™ Blood RNA Kits (OMEGA bio-tek). cDNA was synthesized using a PrimeScript™ II 1st Strand cDNA Synthesis Kit (TaKaRa). Amplification of the full length of exons 2 and 3 sequences was investigated using specific primer pairs (*Mafa-A*: 'A-F' 5'-AACCCTCCTCCTGGTGCTCT-3', and 'A-R' 5'-GGAAGGTTCCATCTCCTGCAG-3', *Mafa-B*: 'B-F' 5'- AACCCTCCTCCTGCTGCT-3', and 'B-R' 5'- TGGACTGGGAAGATGGCT-3') The two upstream primers are both located in exon 1 and the two downstream primers are located in exon 4. PCR employed a denaturation process for 3 min at 94 ℃, followed by 32 cycles at 94 ℃ for 30 s, 58 ℃ (*Mafa-A*) or 56 ℃ (*Mafa-B*) for 30 s, 72 ℃ for 1 min, with a final process at 72 ℃ for 8 min. *Ex Taq* DNA polymerase (TaKaRa) was used in this reaction. PCR products were purified and cloned into the pMD18-T vector (TaKaRa). For each animal, about 30 clones were selected for *Mafa-A* and *Mafa-B* respectively and then were sequenced bidirectionally by the service provider (Beijing Genomics Institute, Shenzhen). Nucleotide sequences of cDNAs were assembled and processed using SeqMan (DNASTAR [25]) and aligned using the Clustal W program (BioEdit [26]). To ensure authenticity, each sequence was uniquely named if three or more identical clones were observed from at least two individuals, or from two independent PCR for an individual. These sequences were then submitted to the GenBank for accession numbers and to the IPD–MHC database for allele nomenclature [27,28].

Phylogenetic analysis

Recombination analysis was performed using the Recombination Detection Program version 4 (RDP4; [29]) with a window size of 20 nucleotides and P value less than 0.000005, and using the Recombination Identification Program (RIP) with a window size of 200 and 99% confidence

intervals (http://www.hiv.lanl.gov/) [30]. A phylogenetic tree was constructed using the neighbor-joining (NJ) method [31] in MEGA7 [32] using exons 2 and 3. Evolutionary distances were computed using the Kimura 2-parameter model [33] and assessed using 1000 bootstrap replicates. Values greater than 50% were used as data-points to construct the tree. The nucleotide polymorphic sites were analyzed by DnaSP. The frequency for the second-most common nucleotide at each position was calculated by the number of occurrences of a nucleotide divided by the number of *Mafa-A1* sequences used in this analysis. The frequency for the second-most common amino acid at each position was also calculated by the number of occurrences of an amino acid divided by the number of *Mafa-A1* amino acid sequences used in this analysis[1].

Results and discussion

Summary of the identified MHC class I alleles

cDNA clones were obtained by RT-PCR using *Mafa-A* and *Mafa-B* specific primer pairs. A total of 1965 clones were sequenced, and 882 *Mafa-A* and 859 *Mafa-B* cDNA sequences were acquired. After sequences alignment and filtering out the sequences detected identical in less than three clones, we identified 100 MHC class I alleles from 30 cynomolgus macaques of Vietnamese origin, including 40 *Mafa-A* and 60 *Mafa-B* genes, of which 28 alleles (11 *Mafa-A*, 17 *Mafa-B*) were identified as new ones. These alleles contain the full length of exons 2 and 3, along with partial sequences of exons 1 and 4. Their allele names, accession numbers, shared alleles in other cynomolgus macaque populations and counterparts in other macaque species are listed in Tables 1 and 2, respectively. Among the 28 novel MHC class I alleles, 5 of them, namely *Mafa-A1*048:01* (KT907348), *Mafa-B*006:01:01* (KT895494), *Mafa-B*112:01* (KT895480), *Mafa-B*180:01* (KT895475) and *Mafa-B*202:01* (KT895441), were newly detected in cynomolgus macaque. The other 23 alleles are new at five- to seven-digit levels of classification. The remaining 72 alleles have been reported previously in the IPD-MHC database, with 41 of them (17 *Mafa-A*, 24 *Mafa-B*) being identified in our laboratory [34-36].

Among the 40 *Mafa-A* alleles, 35 sequences originated from the *Mafa-A1* locus, with the other 2, 1, and 2 alleles originating from *Mafa-A2*, *-A3*, and *-A4* loci, respectively. This indicates that most *Mafa-A* alleles are expressed in the *Mafa-A1* locus [18,19]. As for *Mafa-B* alleles, the locus number designation has not yet been introduced for them because the macaque *MHC-B* genes greatly differ in number between haplotypes [22,28]. Amongst the 30 animals analyzed, each individual expressed 1 to 5 *Mafa-A* genes and 2 to 7 *Mafa-B* genes. On average each monkey expressed 6.8 *Mafa-A*/*-B* genes. Especially, 11 animals were found to express more than 2 *Mafa-A* sequences. And 14 macaques were found to have more than 4 *Mafa-B* alleles. These data showed that both *MHC-A* and *-B*, especially *-B* genes, were duplicated in cynomolgus macaque of Vietnamese origin. Many of the macaques may contain at least two *Mafa-A* and three *Mafa-B* genes loci. This is similar to the finding of previous articles, which proved that the most frequent *Mafa* haplotype in the Filipino macaque population contains two *MHC-A* and three *MHC-B* loci [22,37].

Among the 100 MHC class I alleles, 50 alleles were shared between two or more individuals. The most frequently shared *Mafa-A* molecules, containing the same amino acid sequences in exons 2 and 3 with distribution frequency greater than 10%, were *Mafa-A1*007:01* (8/30, 26.7%), *Mafa-A1*056:03* (7/30, 23.3%) and *Mafa-A1*040:03* (4/30, 13.3%). Similarly, the most frequently shared *Mafa-B* alleles with distribution frequency greater than 10% were *Mafa-B*007:01* (12/30, 40%), *Mafa-B*039:01* (8/30, 26.7%), *Mafa-B*060:13* (7/30, 23.3%), *Mafa-B*093:02* (5/30, 16.7%), *Mafa-B*101:02* (5/30, 16.7%) , *Mafa-B*030:17* (4/30, 13.3%), *Mafa-B*144:01* (4/30, 13.3%) and *Mafa-B*145:01* (4/30, 13.3%). A summary of the shared alleles and the number of allele clones identified in each macaque are shown in Figure 1. All of the *Mafa-A1*007:01*, *Mafa-B*007:01:01*, *Mafa-B*039:01* and *Mafa-B*060:13* were detected to express in individuals 4, 11, and 29, respectively, which indicates that some of them may segregate on one haplotype. Macaque MHC class I allele haplotypes contain variable numbers of loci, which makes them more difficult to characterize than their human counterparts. Although the next-generation sequencing (NGS) techniques have been reported to be effective for high-throughput

genotyping of MHC genes and for the detection of low-level-expressed MHC alleles [38,39], the new technologies are error-prone because it can be more difficult to discriminate between sequencing errors and true rare alleles. Nevertheless, this problem can be overcome by applying the conventional Sanger sequencing methods. The combined use of the conventional Sanger sequencing methods and the NGS techniques can make the characterization of the highly duplicated macaque *MHC-A/-B* alleles easier to perform [22].

Analysis of alleles shared with other populations or with other species

The 72 known alleles identified in this study were compared with other populations from Filipino, Indonesian, Malaysian and Mauritian origin. We found that the majority of them (57) were reported previously in Vietnamese origin population [34-36,40], while the remaining 15 alleles share the same exons 2 and 3 sequences with other populations. One of the 15 alleles, namely *Mafa-B*137:03*, was identical to sequences previously described in both Filipino and Indonesian origin populations, which indicates that this allele is an old entity generated before the divergence of these three populations[18,22]. For the other 14 alleles, 6 of them were found identical to Indonesian-origin counterparts [18,37,41], 4 shared with Filipino-origin population [22]and 4 with Malaysian-origin cynomolgus macaque [41]. None allele was found similar to Mauritian-origin population. Interestingly, two sequences of the *Mafa-A4*14* lineage, *Mafa-A4*14:03* and *Mafa-A4*14:17* identified in this study exhibit identical exon 2 and 3 sequences to *Mamu-A4*14:03*. Meanwhile, the same exon 2 and 3 sequences were shared with Filipino-origin cynomolgus macaque and Malaysian-origin population [42] (Table 1), which indicates that this fragment is conserved in macaque during evolution. Surprisingly, the *Mamu-A4*14:03* allele was reported to be expressed mainly inside the cell, in contrast to *Mamu-A*-encoded molecules which are mostly found on the cell surface. The different expression patterns were assigned to the antigen-binding α1 and α2 domains [43]. It is possible that the two *Mafa-A4*14* alleles take the same expression pattern in cynomolgus macaque as those for the *Mamu-A4*14:03* and they have some important functions in the cell rather than presenting peptides on the cell surface to T cells. Meanwhile,

194 there are also 5 alleles possessing the same deduced amino acid sequences encoding α1 and α2
195 domains as their counterparts from other populations, including 2 from Malaysia, 1 from
196 Philippines, 1 from Indonesia and the last one shared with Philippines and Indonesia (Tables 1
197 and 2). No Mauritian origin sequences were matched.

198 Hence, in this study, we discovered 15 sequences with perfect identity and 6 sequences with
199 identical amino acid sequences encoding α1 and α2 domains to previously defined MHC class I
200 alleles from Indonesian, Filipino, or Malaysian populations. The sharing of alleles between these
201 geographically distinct populations was consistent with the findings of previous studies, i.e. ,
202 there is considerable overlap between different populations for some *Mafa-A* or *-B* lineages at the
203 three-digit level of classification, despite the fact that most *Mafa-A* or *-B* alleles are population
204 specific[37]. Therefore, the majority of *Mafa-A* or *-B* alleles in distinct populations probably fine-
205 tuned their sequences to coping with environmental pathogens, along with a few parts inherited
206 conservatively. It is believed that these shared alleles between continental (Vietnamese,
207 Malaysian) and insular (Filipino, Indonesian) subgroups had been generated before the migration
208 of cynomolgus macaques across land bridges between continental Asia and islands of Indonesia
209 during the late Pleistocene epoch [37].

210 On the other hand, of the 100 alleles identified, 13 *Mafa-A* and 19 *Mafa-B* sequences were
211 identical to previously reported alleles from other macaque species (Tables 1 and 2). These
212 included the rhesus macaque, the southern pig-tailed macaque (*Macaca nemestrina*, *Mane*), the
213 Northern pig-tailed macaque (*Macaca leonina*, *Malo*) and the Assam Macaque (*Macaca*
214 *assamensis*, *Maas*). *M.assamensis* inhabits the southern region of Yunnan province, China [44] and
215 this is the first report of a shared allele (*MHC-B*039:01*) expressed in cynomolgus, rhesus and
216 assamensis macaques. Another 3 lineages were also shared among at least three macaque
217 species, such as *MHC-A1*130:01*, *MHC-B*018:01* and *MHC-B*028*. Interestingly, the shared
218 *MHC-B*039:01*, was reported in rhesus macaque that it contained a specific B pocket structural
219 motif and had a unique peptide-binding preference consisting of glycine at the second position.

This pocket structure was reported in about 6% of rhesus macaque sequences but absent in human *HLA* genes [45]. Our data showed for the first time that the unique B pocket structural motif also occurred in *Mafa-B*039:01* of Vietnamese origin with appreciable frequency. The biological significance of this molecule needs to be further analyzed in the future and needs to be concerned when using Vietnamese-origin cynomolgus macaques in biomedical researches.

As presented in tables 1 and 2, five alleles were shared with Indian-origin rhesus macaque, which indicates that they are old entities generated before the divergence of these species and may be related with pivotal functions in immune responses. Meanwhile, 19 alleles were shared with Chinese-origin rhesus macaque. It can be easily noticed that cynomolgus macaque shared more alleles with Chinese-origin than with Indian-origin rhesus macaque. This may be explained by the overlap in the geographical areas inhabited by both species in eastern Asia, where the two species are likely to hybridize with extensive ancient introgression from Chinese rhesus macaque into the Vietnamese-origin cynomolgus macaque population, as reported previously [46-48]. This theory, i.e. ancient hybridization and admixture in macaques, can be also used to explain the fact that the Vietnamese-origin cynomolgus macaque shared more alleles with other macaque species than with other cynomolgus macaque populations [49]. Regarding the above analysis, cynomolgus macaque of Vietnamese origin may be the best replacer for Chinese rhesus macaque when used in biomedical researches.

Analysis of recombination in *Mafa-A* and *Mafa-B* alleles

Recombination events are one of the proposed mechanisms to explain the diversity of MHC alleles. It has been reported that the *Mafa-B*099* allele lineage was generated by the combination of the *Mafa-B*054* and the *Mafa-B*095* allele lineages in cynomolgus macaque mostly originated from the Philippines [50]. In order to detect whether some other recombination events exist in Vietnamese origin cynomolgus macaque, we analyzed 77 *Mafa-A* and 99 *Mafa-B* sequences discovered in our laboratory, including data in this study and those previously

reported [34-36]. Using the RDP program, four possible recombination events in *Mafa-A* and *Mafa-B* alleles, as shown in Table S1, were detected simultaneously by at least four different recombination detection methods. Among them, three recombinants showed lower sequence similarity with their parents hence require further investigation. Only *Mafa-A1*036:04* was detected as a potential recombinant by five recombination detection methods and was exhibited over 98% sequence similarity with the counterparts from its parents. As shown in Figure 2, the *Mafa-A1*036:04* and the *Mafa-A1*091:02* contain very similar sequences in exon 3, with minor variation of four nucleotides in their 5' regions. However, the exon 2 sequences of these two alleles differ considerably, with 24 nucleotides being mismatched. We also found that the exon 2 sequence of *Mafa-A1*036:04* is very similar to the *Mafa-A1*057:01* allele, with only two nucleotide differences present in their 5' regions. To determine whether the *Mafa-A1*036:04* was created by a recombination event, we further conducted analysis using RIP (Figure S1). The result indicates that the *Mafa-A1*036:04* allele was possibly generated by a crossover event between *Mafa-A1*091:02* and *Mafa-A1*057:01*. The exact breakpoint cannot be defined because their intron sequences are not available in this study. Additionally, we further performed phylogenetic analysis of all reported sequences belonging to the *A1*036*, *A1*057* and *A1*091* lineages from cynomolgus and rhesus macaques. The phylogeny map of exon 2 sequences presented in Figure 3A showed that *Mafa-A1*036* cluster more closely to Mafa-*A1*057* than the *Mafa-A1*091*. While in the phylogenetic tree of exon 3 presented in Figure 3B, *Mafa-A1*036* separates *Mafa-A1*057* in different branches and groups *Mafa-A1*091* in a cluster. We also found that the *-A1*057* and *-A1*091* lineages are grouped together in these two trees regardless of the species, while the *-A1*036* lineages are separated according to species. *Mamu-A1*036* exhibits higher sequence similarity with *Mamu-A1*091* than with *Mamu-A1*057*. It is possible that the *-A1*036* allele generated by crossover recombination between *-A1*091* and *-A1*057* just occurred recently in cynomolgus macaque, but not yet in rhesus macaque.

Analysis of the diversity in *Mafa-A1* locus

We have obtained 77 *Mafa-A* and 99 *Mafa-B* sequences in our laboratory, including 67 *Mafa-A1*. Since *Mafa-A1* was the highest polymorphic gene than other *Mafa-A* loci and the locus number designation for *Mafa-B* was not yet clear, here we only analyze the diversity in *Mafa-A1* locus of Vietnamese origin. The sequences of the 67 *Mafa-A1* exons 2 and 3, encoding residues 2-182 of the MHC class I protein, were aligned and a total of 157 nucleotide polymorphic sites (28.8%) were discovered by DnaSP (Table 3). To distinguish positions presenting two or more nucleotides from sites dominated by one nucleotide, we calculated the incidence for the second-most common nucleotide at each position (Table S2). 97 variable sites were considered highly polymorphic with the incidence greater than 5%, while the remaining 60 with the incidence less than 5% were considered to exhibit rare variation. Analysis on the variability index of the second-most common amino acid residue showed that 48 out of 181 sites (26.5%) were defined as highly polymorphic, where the distribution frequency of the second-most common amino acid is greater than 5% (Table S3). In comparison, only 70 nucleotide positions and 45 amino acid residues in *HLA-A* were considered highly polymorphic [1]. These data showed that cynomolgus macaque *Mafa-A1* exhibit higher polymorphism than human *HLA-A* and several polymorphic sites are macaque-specific[37]. The α1 and α2 domains of MHC class I glycoproteins contains many functional sites that bind peptide antigens and engage T cell receptors. According to previous studies [51], the deduced 36 peptide binding sites (PBS) and 26 TCR binding sites (TBS) were determined. Among these binding sites, 8 residues were both involved in the interaction with the peptides and the receptors. The diversity of nucleotide sequences encoding PBS or TBS (Pi=0.162) in the *Mafa-A1* were predominantly higher than that in non-PBS or non-TBS coding regions (Pi=0.035) of the corresponding alleles (Table 3). This is consistent with the observation that highly polymorphism at these functional residues is significant to increase the depth and breadth of the weaponry to cope with variant pathogens during evolution [1,52]. All of these 54 functional residues were listed in Figure 4, including 34 positions with high polymorphism. On the other hand, 16 out of the 54 functional residues are completely conserved in the 67 *Mafa-A1* sequences. 8 of the 16 residues are also

298 conserved in human, including Y7, Y59, and Y159. The three tyrosine residues were located at
299 an end of the peptide binding groove and may contribute to the recognition of a constant
300 feature of processed antigens [53] which indicates that these conserved residues are also
301 important to maintain some constant features for presenting peptide and for lymphocyte
302 recognition during evolution.

303 **Conclusion**

304 In this study, we have identified 40 *Mafa-A* and 60 *Mafa-B* alleles from 30 unrelated cynomolgus
305 macaques of Vietnamese origin. 28 of these alleles were found to be novel ones. Each monkey
306 expressed 1 to 5 *Mafa-A* genes and 2 to 7 *Mafa-B* genes. These data showed that both *MHC-A*
307 and *-B*, especially *-B* genes, were duplicated in cynomolgus macaque of Vietnamese origin.
308 Many of the macaques may contain at least two *Mafa-A* and three *Mafa-B* genes loci, which is
309 similar to that in the Filipino-origin population. We also identified some alleles with distribution
310 frequency greater than 10% and four alleles (*Mafa-A1*007:01*, *Mafa-B*007:01:01*, *Mafa-*
311 *B*039:01* and *Mafa-B*060:13*) were detected to express simultaneously in three individuals.
312 Whether these four alleles segregate on one haplotype need to be verified in future study. Among
313 the 72 known alleles, 15 alleles shared the same exons 2 and 3 sequences with other populations,
314 including Filipino, Indonesian and Malaysian origin populations. The sharing of alleles between
315 these geographically distinct populations indicates that a few alleles preserved conservatively in
316 evolution may exercise vital immune functions, and many of the *Mafa-A* or *-B* alleles in distinct
317 populations probably fine-tuned their sequences to coping with environmental pathogens. On the
318 other hand, 32 sequences were identical to previously reported alleles from other macaque
319 species, including 19 shared with Chinese-origin rhesus macaque. The fact that the Vietnamese-
320 origin cynomolgus macaque shared more alleles with Chinese-origin rhesus macaque than with
321 other cynomolgus macaque populations may be explained by ancient hybridization and
322 admixture in macaques. In this regard, cynomolgus macaque of Vietnamese origin may be the
323 best replacer for Chinese rhesus macaque when used in biomedical researches. To further explain
324 the diversity of *Mafa-A and -B* genes, recombination events and the variability in *Mafa-A1* locus

were analyzed, the *Mafa-A1*036:04* allele was possibly generated by a crossover event between *Mafa-A1*091:02* and *Mafa-A1*057:01*, which occurred recently in cynomolgus macaque, but not in rhesus macaque yet. 97 variable nucleotide positions in *Mafa-A1* exons 2 and 3 sequences and 48 amino acid sites in the α1 and α2 domains were considered highly polymorphic. In comparison to human, the exons 2 and 3 sequences from cynomolgus macaque exhibit higher polymorphism. The diversity of the functional residues for PBS or TBS in the *Mafa-A1* molecules was predominantly higher than that for non-PBS or non-TBS regions. The high polymorphism at these functional residues is presumably vital not only to increase the width of an individual's immune response to a pathogen, but also to help families or populations survive various infections.

Acknowledgments

We are grateful for the Primate Research Center of South China for helping with the cynomolgus macaque blood collection, and for IPD-MHC Nomenclature Committee for providing the designations for the MHC class I alleles.

Ethics:

The methods of using animals in this study had been reviewed and approved by the institutional Animal Care and Use Committees (IACUC), Guangdong Landau Biotechnology Co. Ltd. The date of approval was Oct.8,2015. The approval code was IACUC-003.

Funding：

This study was funded by the National Natural Science Foundation of China (31401088) and supported by the Fundamental Research Funds for the Central Universities of South China University of Technology (2014ZZ0056).

Disclosures:

There are no conflicts of interest among all authors in this article.

Reference

1. Robinson J, Guethlein LA, Cereb N, et al. Distinguishing functional polymorphism from random variation in the sequences of >10,000 HLA-A, -B and -C alleles. *PLoS Genet.* 2017;13(6):e1006862.
2. Steiner NK, Baldassarre L, Koester R, Ng J, Hartzman RJ, Hurley CK. Novel HLA-B alleles formed by an inter-locus recombination with HLA-C, HLA-B*0713 and B*6702. *Tissue Antigens.* 2002;59(2):148–150.
3. Bahr A, Wilson AB. The evolution of MHC diversity: evidence of intralocus gene conversion and recombination in a single-locus system. *Gene.* 2012;497(1):52-57.
4. Nomura T, Matano T. Association of MHC-I genotypes with disease progression in HIV/SIV infections. *Frontiers in microbiology.* 2012;3:234.
5. Nomura T, Yamamoto H, Shiino T, et al. Association of major histocompatibility complex class I haplotypes with disease progression after simian immunodeficiency virus challenge in burmese rhesus macaques. *J Virol.* 2012;86(12):6481-6490.
6. Kuniholm MH, Xiaojiang G, Xiaonan X, et al. Human leukocyte antigen genotype and risk of HIV disease progression before and after initiation of antiretroviral therapy. *Journal of Virology.* 2011;85(20):10826-10833.
7. Carrington M, O'Brien SJ. The influence of hla genotype on AIDS. *Annual Review of Medicine.* 2003;54(1):535-551.
8. Peter P. MHC class I molecules and KIRs in human history, health and survival. *Nature Reviews Immunology.* 2005;5(3):201-214.
9. Vierboom M, Breedveld E, Kondova I, t Hart BA. The significance of non-human primates as preclinical models of human arthritic disease. *Expert Opin Drug Discov.* 2008;3(3):299-310.
10. Vallender EJ. Bringing non-human primate research into the post-genomic era: how monkeys are teaching us about elite controllers of HIV/AIDS. *Genome Biology.* 2014;15(11):1-3.
11. Smits SL, van den Brand JM, de Lang A, et al. Distinct severe acute respiratory syndrome coronavirus-induced acute lung injury pathways in two different nonhuman primate species. *J Virol.* 2011;85(9):4234-4245.
12. Lin PL, Rodgers M, Smith L, et al. Quantitative Comparison of Active and Latent Tuberculosis in the Cynomolgus Macaque Model. *Infection and Immunity.* 2009;77(10):4631-4642.
13. Kisu I, Mihara M, Banno K, et al. Uterus allotransplantation in cynomolgus macaque: a preliminary experience with non-human primate models. *J Obstet Gynaecol Res.* 2014;40(4):907-918.
14. Walter L, Ansari AA. MHC and KIR Polymorphisms in Rhesus Macaque SIV Infection. *Front Immunol.* 2015;6:540.
15. Gumert MD. The common monkey of Southeast Asia: long-tailed macaque populations, ethnophoresy, and their occurrence in human environments. *Book Chapter.* 2011.
16. Sussman RW, Tattersall I. Distribution, abundance, and putative ecological strategy of Macaca fascicularis on the island of Mauritius, southwestern Indian Ocean. *Folia primatologica.* 1986;46 (1):28-43.
17. Otting N, de Vos-Rouweler AJ, Heijmans CM, de Groot NG, Doxiadis GG, Bontrop RE. MHC class I A region diversity and polymorphism in macaque species. *Immunogenetics.* 2007;59(5):367-375.
18. Pendley CJ, Becker EA, Karl JA, et al. MHC class I characterization of Indonesian cynomolgus macaques. *Immunogenetics.* 2008;60(7):339-351.

19. Campbell KJ, Detmer AM, Karl JA, et al. Characterization of 47 MHC class I sequences in Filipino cynomolgus macaques. *Immunogenetics.* 2009;61(3):177-187.
20. Robinson J, Mistry K, Mcwilliam H, Lopez R, Parham P, Marsh SGE. The IMGT/HLA database. *Methods in Molecular Biology.* 2011;39(Database issue):1171-1176.
21. Shiina T, Tamiya G, Oka A, Takishima N, Inoko H. Genome sequencing analysis of the 1.8 Mb entire human MHC class I region. *Immunological Reviews.* 1999;167(1):193–199.
22. Shiina T, Yamada Y, Aarnink A, et al. Discovery of novel MHC-class I alleles and haplotypes in Filipino cynomolgus macaques (Macaca fascicularis) by pyrosequencing and Sanger sequencing: Mafa-class I polymorphism. *Immunogenetics.* 2015;67(10):563-578.
23. Wiseman RW, Karl JA, Bohn PS, Nimityongskul FA, Starrett GJ, O'Connor DH. Haplessly hoping: macaque major histocompatibility complex made easy. *ILAR J.* 2013;54(2):196-210.
24. Maccari G, Robinson J, Ballingall K, et al. IPD-MHC 2.0: an improved inter-species database for the study of the major histocompatibility complex. *Nucleic acids research.* 2016;45(D1):D860-D864.
25. Burland TG. DNASTAR's Lasergene sequence analysis software. *Methods in Molecular Biology.* 2000;132:71-91.
26. Hall BTA. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98. Paper presented at: Nucl. Acids. Symp. Ser2015.
27. Robinson J, Halliwell JA, McWilliam H, Lopez R, Marsh SG. IPD-the Immuno Polymorphism Database. *Nucleic Acids Res.* 2013;41(Database issue):D1234-1240.
28. de Groot NG, Otting N, Robinson J, et al. Nomenclature report on the major histocompatibility complex genes and alleles of Great Ape, Old and New World monkey species. *Immunogenetics.* 2012;64(8):615-631.
29. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. *Virus Evol.* 2015;1(1):vev003.
30. Zhao M, Wang Y, Shen H, et al. Evolution by selection, recombination, and gene duplication in MHC class I genes of two Rhacophoridae species. *Bmc Evolutionary Biology.* 2013;13(1):1-13.
31. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molbiolevol.* 1987;4(6):406-425.
32. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol Biol Evol.* 2016;33(7):1870-1874.
33. Kimura M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution.* 1981;16(2):111-120.
34. Zhang GQ, Ni C, Ling F, et al. Characterization of the major histocompatibility complex class I A alleles in cynomolgus macaques of Vietnamese origin. *Tissue Antigens.* 2012;80(6):494-501.
35. Wang HB, Ling F, Zhuo M, Wang JF, Wang XN. Eighteen novel MHC class I A alleles identified in Vietnamese-origin cynomolgus macaques. *Tissue Antigens.* 2011;78(2):139-142.
36. Wang HB, Ling F, Zhuo M, Wang JF, Wang XN. Twenty-three novel major histocompatibility complex class I B alleles identified in cynomolgus macaques of Vietnamese origin. *Tissue Antigens.* 2011;77(4):346-348.
37. Kita YF, Hosomichi K, Kohara S, et al. MHC class I A loci polymorphism and diversity in three Southeast Asian populations of cynomolgus macaque. *Immunogenetics.* 2009;61(9):635-648.

38. Wiseman RW, Karl JA, Bimber BN, et al. Major histocompatibility complex genotyping with massively parallel pyrosequencing. *Nat Med.* 2009;15(11):1322-1326.
39. Budde ML, Wiseman RW, Karl JA, Hanczaruk B, Simen BB, O'Connor DH. Characterization of Mauritian cynomolgus macaque major histocompatibility complex class I haplotypes by high-resolution pyrosequencing. *Immunogenetics.* 2010;62(11-12):773-780.
40. Krebs KC, Jin Z, Rudersdorf R, Hughes AL, O'Connor DH. Unusually High Frequency MHC Class I Alleles in Mauritian Origin Cynomolgus Macaques. *The Journal of Immunology.* 2005;175(8):5230-5239.
41. Saito Y, Naruse TK, Akari H, Matano T, Kimura A. Diversity of MHC class I haplotypes in cynomolgus macaques. *Immunogenetics.* 2012;64(2):131-141.
42. Aarnink A, Apoil PA, Takahashi I, Osada N, Blancher A. Characterization of MHC class I transcripts of a Malaysian cynomolgus macaque by high-throughput pyrosequencing and EST libraries. *Immunogenetics.* 2011;63(11):703-713.
43. Rosner C, Kruse PH, Lubke T, Walter L. Rhesus macaque MHC class I molecules show differential subcellular localizations. *Immunogenetics.* 2010;62(3):149-158.
44. Yan X, Li A, Zeng L, et al. Identification of MHC class I sequences in four species of Macaca of China. *Immunogenetics.* 2013;65(12):851-859.
45. Sette A, Sidney J, Southwood S, et al. A shared MHC supertype motif emerges by convergent evolution in macaques and mice, but is totally absent in human MHC molecules. *Immunogenetics.* 2012;64(6):421-434.
46. Bonhomme M, Cuartero S, Blancher A, Crouauroy B. Assessing Natural Introgression in 2 Biomedical Model Species, the Rhesus Macaque (Macaca mulatta) and the Long-Tailed Macaque (Macaca fascicularis). *Journal of Heredity.* 2009;100(2):158.
47. Stevison LS, Kohn MH. Divergence population genetic analysis of hybridization between rhesus and cynomolgus macaques. *Molecular Ecology.* 2009;18(11):2457-2475.
48. Kanthaswamy S, Satkoski J, George D, Kou A, Erickson BJ, Smith DG. Interspecies Hybridization and the Stratification of Nuclear Genetic Variation of Rhesus (Macaca Mulatta) and Long-Tailed Macaques (Macaca Fascicularis). *Int J Primatol.* 2008;29(5):1295-1311.
49. Fan Z, Zhou A, Osada N, et al. Ancient hybridization and admixture in macaques (genus Macaca) inferred from whole genome sequences. *Molecular Phylogenetics & Evolution.* 2018:S105579031730489X.
50. Orysiuk D, Lawrence J, Prashar T, et al. Evidence of recombination producing allelic diversity in MHC class I Mafa-B and -A alleles in cynomolgus macaques. *Tissue Antigens.* 2012;79(5):351-358.
51. Lafont BAP, Buckler-White A, Plishka R, Buckler C, Martin MA. Characterization of Pig-Tailed Macaque Classical MHC Class I Genes: Implications for MHC Evolution and Antigen Presentation in Macaques. *The Journal of Immunology.* 2003;171(2):875-885.
52. Lian XD, Zhang XH, Dai ZX, Zheng YT. Cloning, sequencing, and polymorphism analysis of novel classical MHC class I alleles in northern pig-tailed macaques (Macaca leonina). *Immunogenetics.* 2016;68(4):261-274.
53. Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature.* 1987;329(6139):512-518.

Figure 1(on next page)

Figure 1. Summary of MHC class *I-A* and *-B* alleles identified from 30 cynomolgusmacaques of Vietnamese origin.

Figure 1

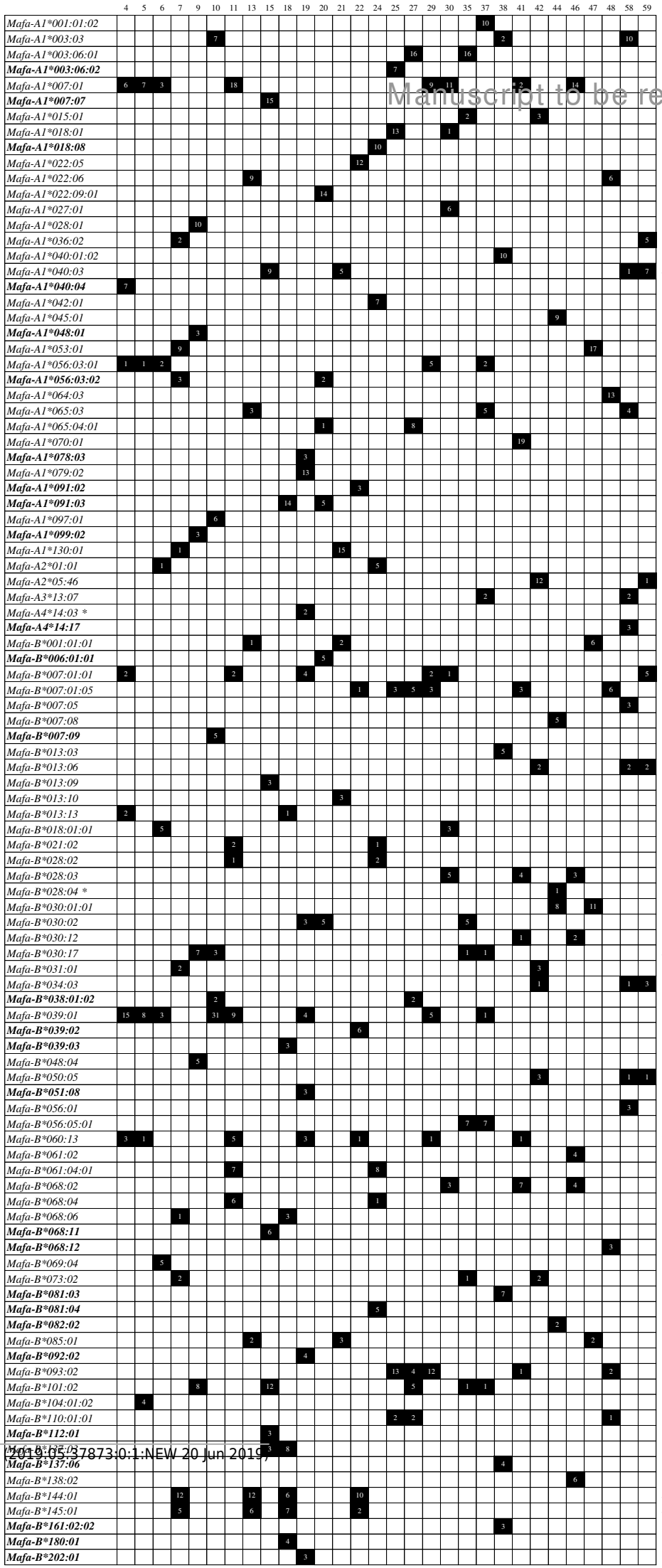

Figure 1. Summary of MHC class *I-A* and *-B* alleles identified from 30 cynomolgus macaques of Vietnamese origin. The novel alleles are represented in bold. The number of cDNA clones for every allele detected in each animal are shown in the cells and highlighted in black. The numbers of individuals sharing the particular allele are shown at the end of each line. The two asterisk-labeled alleles (*Mafa-A4*14:03* and *Mafa-B*028:04*) are also identified in other individuals which are not in these 30 animals, respectively.

Figure 2(on next page)

Figure 2. The nucleotide sequences alignment of *Mafa-A1*091:02*,*Mafa-A1*036:04* and *Mafa-A1*057:01*.

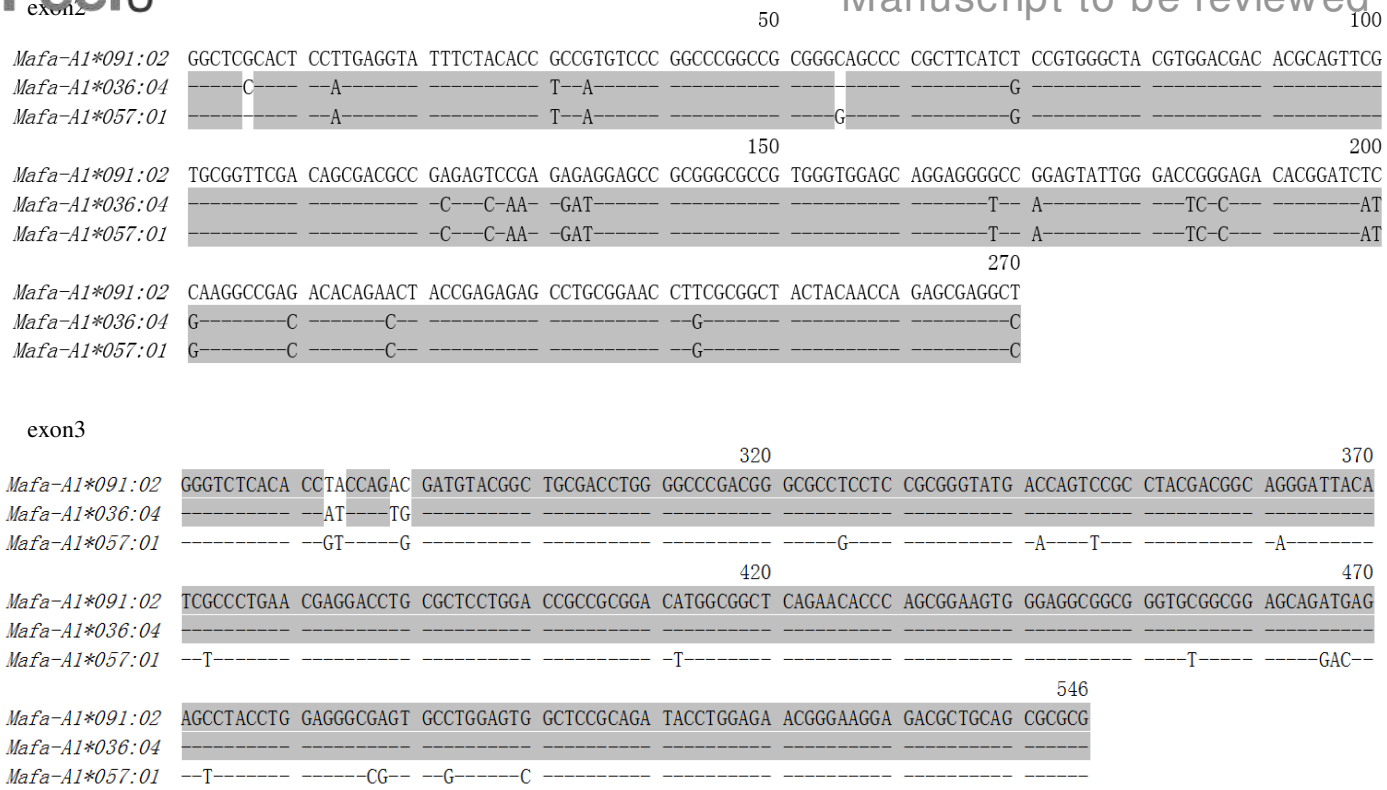

Figure 2. The nucleotide sequences alignment of *Mafa-A1*091:02*, *Mafa-A1*036:04* and *Mafa-A1*057:01*. Nucleotide identical to the top sequence is indicated by a hyphen. Nucleotide numbers are shown above. *Mafa-A1*036:04* appears to be a result of a crossover between *Mafa-A1*057:01* (exon 2, grey marked) and *Mafa-A1*091:02* (exon 3, grey marked).

Figure 3(on next page)

Figure 3. Phylogenetic analysisof exon 2 (A) and exon 3 (B) of *A1*036,A1*57* and *A1*091* alleles from cynomolgus and rhesusmacaques.

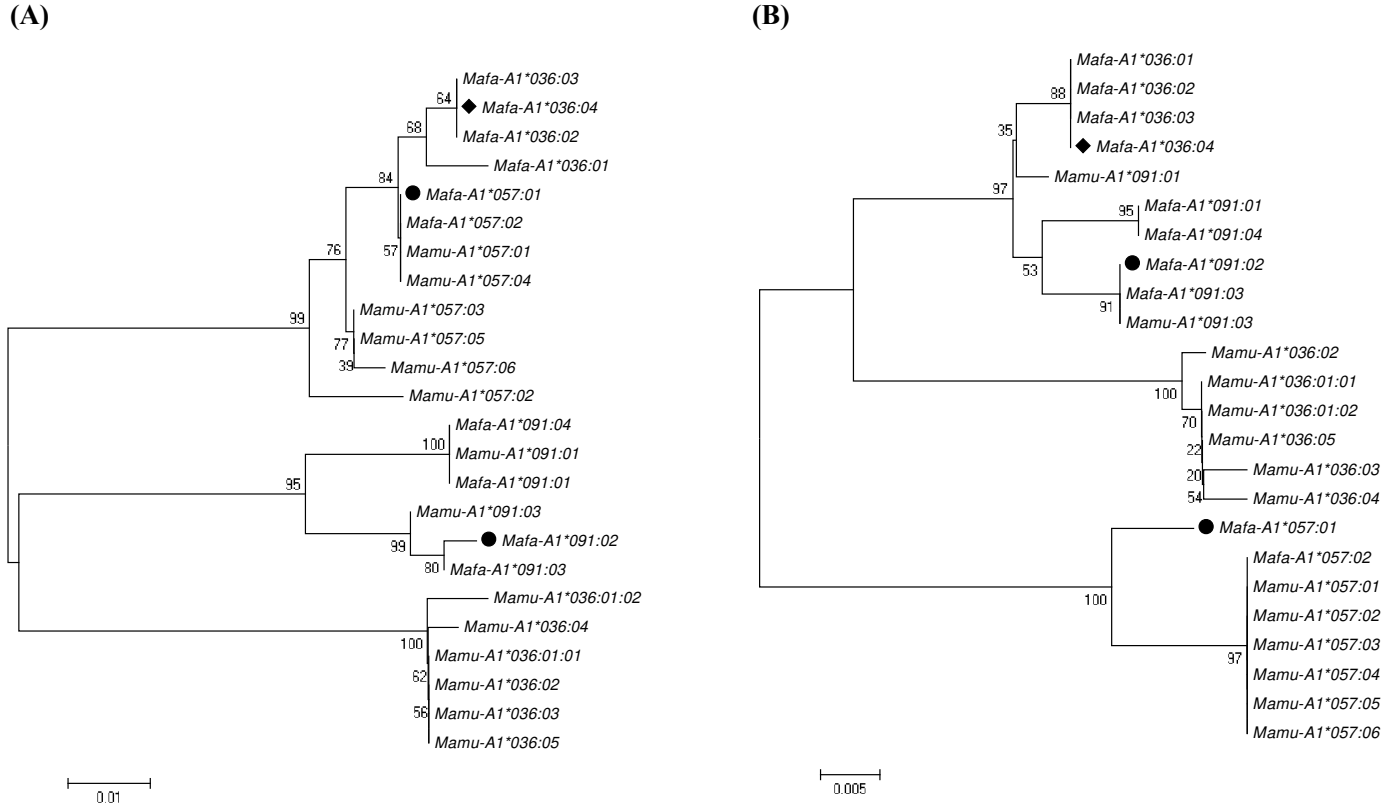

Figure 3. Phylogenetic analysis of exon 2 (A) and exon 3 (B) of *A1*036, A1*57* and *A1*091* alleles from cynomolgus and rhesus macaques. Solid circles represent alleles for recombinant parents and the solid diamonds indicate the recombinant.

Figure 4(on next page)

Figure 4. The diversity of amino acid residues at 54 functional positions relativeto PBS or TBS in 67 *Mafa-A1* ofVietnamese origin.

Figure 4

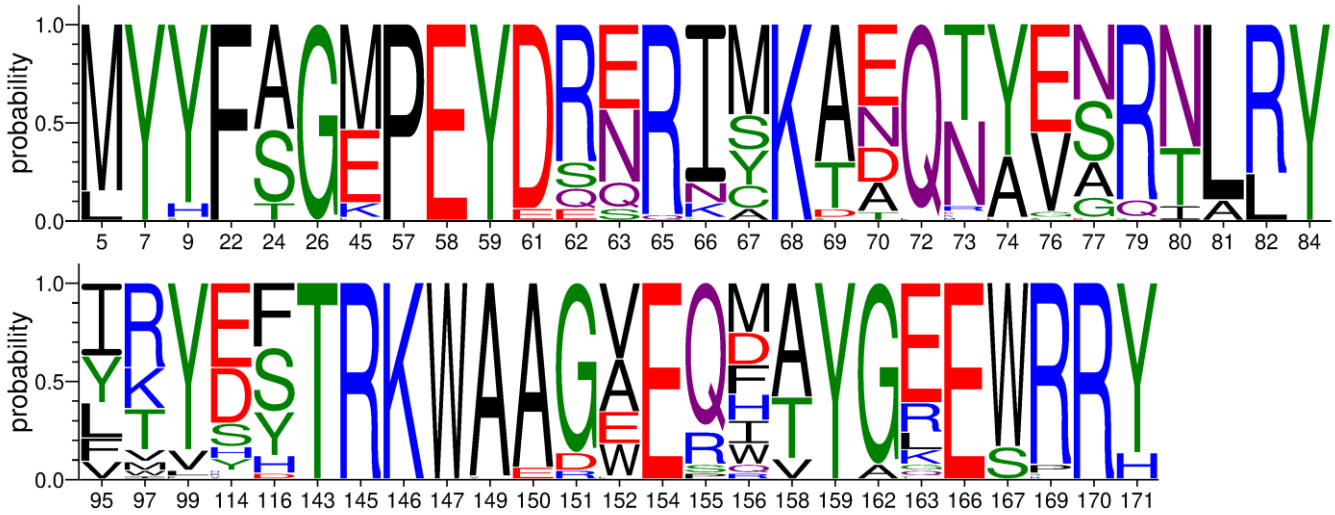

Figure 4. The diversity of amino acid residues at 54 functional positions relative to PBS or TBS in 67 *Mafa-A1* of Vietnamese origin. The height of symbols within the stack reflects the distribution frequency of each amino acid residue. The 54 functional residues were determined according to Ref.51 and listed as follows: 28 PBS (5, 7, 9, 22, 24, 26, 45, 59, 63, 66, 67, 70, 73, 74, 77, 80, 81, 84, 95, 97, 99, 114, 116, 143, 147, 156, 159, 171), 18 TBS (57, 61, 65, 68, 69, 72, 76, 79, 82, 145, 149, 150, 151, 154, 158, 162, 166, 169) and 8 both PBS and TBS (58, 62, 146, 152, 155, 163, 167, 170).

Table 1(on next page)

Table1 40 *Mafa-A* alleles detected inVietnamese-origin cynomolgus macaques

Table1 40 *Mafa-A* alleles detected in Vietnamese-origin cynomolgus macaques

Allele name	Accession number	Other origin	Macaque counterparts [a]
*Mafa-A1*001:01:02*	KT907313	*A1*001:01:01*(MacM)[c]	*Mamu-A1*001:01*(U50836-I)
*Mafa-A1*003:03*	KT907312		*Mamu-A1*003:01:01*(U41379-Unk)
*Mafa-A1*003:06:01*	KT907326		
Mafa-A1*003:06:02	KT907327		
*Mafa-A1*007:01*	KT907328		
Mafa-A1*007:07	KT907316		*Mamu-A1*007:02*(AF157397-Unk)
*Mafa-A1*015:01*	KT907351		*Mamu-A1*015:01*(AB551785-Bu)
*Mafa-A1*018:01*	KT907329		
Mafa-A1*018:08	KT907330		
*Mafa-A1*022:05*	KT907331		
*Mafa-A1*022:06*	KT907309		
*Mafa-A1*022:09:01*	KT907332		
*Mafa-A1*027:01*	KT907333		
*Mafa-A1*028:01*	KT907334		
*Mafa-A1*036:02*	KY073130		
*Mafa-A1*040:01:02*	KT907315		
*Mafa-A1*040:03*	KT907321		
Mafa-A1*040:04	KT907322		
*Mafa-A1*042:01*	KT907324		
*Mafa-A1*045:01*	KT907335		*Mamu-A1*045:01*(EU262741-Ch)
Mafa-A1*048:01	KT907348		
*Mafa-A1*053:01*	KT907336		*Mamu-A1*053:02*(EU551177-Ch)
*Mafa-A1*056:03:01*	KT907337		*Mamu-A1*056:02:01*(AM295922-Ch)
Mafa-A1*056:03:02	KT907338		
*Mafa-A1*064:03*	KT907325		
*Mafa-A1*065:03*	KT907339		
*Mafa-A1*065:04:01*	KT907340		*Mamu-A1*065:01*(AB430441-Bu, EU418506-Ch)
*Mafa-A1*070:01*	KT907341	ICM [b]	
Mafa-A1*078:03	KT907344		
*Mafa-A1*079:02*	KT907342	ICM [b]	
Mafa-A1*091:02	KT907319		
Mafa-A1*091:03	KT907320		
*Mafa-A1*097:01*	KT907318	ICM [b]	*Mamu-A1*109:01*(AB444902-Bu)
Mafa-A1*099:02	KT907323		
*Mafa-A1*130:01*	KT907343		*Mane-A1*130:01*(LN875412-Unk), *Mamu-A1*130:01*(HG813262-Unk)
*Mafa-A2*01:01*	KT907314		*Mamu-A2*01:03*(AB444917-Bu, GQ902066-Ch)

*Mafa-A2*05:46*	KT907345		*Mamu-A2*05:21*(AM295935-Ch)
*Mafa-A3*13:07*	KT907347		
*Mafa-A4*14:03*	KT907349	PCM [b]	*Mamu-A4*14:03:01*(AB444876-Bu/I, GU080236-Ch)
Mafa-A4*14:17	KT907350	*A4*14:04* (MaCM) [b]	

The 40 *Mafa-A* alleles identified from Vietnamese-origin cynomolgus monkeys are listed. The bold and underlined ones indicate newly identified alleles. IPD name, GenBank accession number, other origin and counterpart(s) in other macaque species are listed for each allele.

[a] For alleles shared with other macaque species, the names of their counterparts, accession numbers, as well as regional populations are also listed. I, Indian rhesus macaque; Bu, Burmese rhesus macaque; Ch, Chinese rhesus macaque; Unk, Unknown-origin rhesus macaque.

[b] For alleles shared identical exons 2 and 3 nucleotide sequences with other populations, ICM: Indonesian origin; PCM: filipino origin; MaCM: Malaysian origin.

[c] For alleles shared identical deduced amino acid sequences encoding α1 and α2 domains with other populations, ICM: Indonesian origin; PCM: filipino origin; MaCM: Malaysian origin.

Table 2(on next page)

Table2 60 *Mafa-B* alleles detected in Vietnamese-origin cynomolgus macaques

Table2 60 *Mafa-B* alleles detected in Vietnamese-origin cynomolgus macaques

Allele name	Accession number	Other origin	Macaque counterparts [a]
*Mafa-B*001:01:01*	KT895485		*Mamu-B*001:01:01*(AB477408- Bu, U42837-I)
Mafa-B*006:01:01	KT895494		*Mamu B*006:01*(U41828-Unk)
*Mafa-B*007:01:01*	KT895444	PCM [b]	*Mamu-B*007:03*(AB477412-Bu, EU682528-Ch, AJ556876-I)
*Mafa-B*007:01:05*	KT895442		
*Mafa-B*007:05*	KT895443		
*Mafa-B*007:08*	KT895446		*Mamu-B*007:04:01*(GQ902078-Ch)
Mafa-B*007:09	KT895445		
*Mafa-B*013:03*	KT895451		
*Mafa-B*013:06*	KT895447		
*Mafa-B*013:09*	KT895448	*B*013:08* (PCM & ICM) [c]	
*Mafa-B*013:10*	KT895449		
*Mafa-B*013:13*	KT895450		
*Mafa-B*018:01:01*	KT895490	ICM [b]	*Mamu-B*018:01*(AM902534-Ch) *Malo-B*018:01*(KT214460-Unk)
*Mafa-B*021:02*	KT895452		*Mamu-B*021:02*(AM902536-Bu/Ch)
*Mafa-B*028:02*	KT895487		*Mane-B*028:01* (FJ875264.1-Unk) *Mamu-B*028:02:01* (AM902532.1-Ch)
*Mafa-B*028:03*	KT895486	PCM [b]	
*Mafa-B*028:04*	KY131948	ICM [b]	
*Mafa-B*030:01:01*	KT895454		*Mamu-B*030:03:02*(AM902546- Ch)
*Mafa-B*030:02*	KT895489	MaCM [b]	*Mamu-B*030:03:03*(AM902547- Ch)
*Mafa-B*030:12*	KT895438		*Mane-B*030:04*(FJ875259-Unk)
*Mafa-B*030:17*	KT895453		
*Mafa-B*031:01*	KT895491		
*Mafa-B*034:03*	KT895455		
Mafa-B*038:01:02	KT895456		*Mamu-B*038:02*(AB477391-Bu) *Maas-B*039:01*(KF012951-Ch)
*Mafa-B*039:01*	KT895457		*Mamu-B*039:01*(AB477411-Bu, EF580146-Ch, AJ556890-I)
Mafa-B*039:02	KT895436		
Mafa-B*039:03	KT895437		
*Mafa-B*048:04*	KT895458		
*Mafa-B*050:05*	KT895459		
Mafa-B*051:08	KT895460		
*Mafa-B*056:01*	KT895488	ICM [b]	*Mamu-B*056:01*(GQ902079-Ch)
*Mafa-B*056:05:01*	KT895461		
*Mafa-B*060:13*	KT895462		

*Mafa-B*061:02*	KT895464	MaCM [b]	
*Mafa-B*061:04:01*	KT895463		*Mamu-B*061:02*(AM902564-Bu/Ch)
*Mafa-B*068:02*	KT895466		
*Mafa-B*068:04*	KT895468	MaCM [b]	*Mamu-B*068:04*(AM902571-Bu/Ch)
*Mafa-B*068:06*	KT895467		*Mamu-B*068:02*(EF219482-Unk)
Mafa-B*068:11	KT895465		
Mafa-B*068:12	KT895469	*B*068:08*(PCM) [c]	
*Mafa-B*069:04*	KT895470		
*Mafa-B*073:02*	KT895472		*Mamu-B*073:01*(AB477404-Bu, AM902578-Ch)
Mafa-B*081:03	KT895473		
Mafa-B*081:04	KT895474	*B*081:01*(ICM) [c]	
Mafa-B*082:02	KT895495		
*Mafa-B*085:01*	KT895484	PCM [b]	
Mafa-B*092:02	KT895471	*B*092:01*(MaCM) [c]	*Mamu-B*092:02*(AB477386-Bu)
*Mafa-B*093:02*	KT895476		
*Mafa-B*101:02*	KT895493		
*Mafa-B*104:01:02*	KT895477		*Mane-B*104:02*(FJ875231-Unk)
*Mafa-B*110:01:01*	KT895478		
Mafa-B*112:01	KT895480		
*Mafa-B*137:03*	KT895439	PCM, ICM [b]	
Mafa-B*137:06	KT895440		
*Mafa-B*138:02*	KT895479	MaCM [b]	
*Mafa-B*144:01*	KT895482		
*Mafa-B*145:01*	KT895483		
Mafa-B*161:02:02	KT895481		
Mafa-B*180:01	KT895475		
Mafa-B*202:01	KT895441		

The 60 *Mafa-B* alleles identified from Vietnamese-origin cynomolgus monkeys are listed. The bold and underlined ones indicate newly identified alleles. IPD name, GenBank accession number, other origin and counterpart(s) in other macaque species are listed for each allele.

[a] For alleles shared with other macaques species, the name of their counterparts, the accession numbers, as well as regional populations are also listed. I, Indian rhesus macaque; Bu, Burmese rhesus macaque; Ch, Chinese rhesus macaque; Unk, Unknown-origin rhesus macaque.

[b] For alleles shared identical exons 2 and 3 nucleotide sequences with other populations, ICM: Indonesian origin; PCM: filipino origin; MaCM: Malaysian origin.

[c] For alleles shared identical deduced amino acid sequences encoding α1 and α2 domains with other populations, ICM: Indonesian origin; PCM: filipino origin; MaCM: Malaysian origin.

Table 3(on next page)

Table 3 Polymorphism of exons 2 and 3 sequences for *Mafa-A1* of Vietnamese origin

Table 3 Polymorphism of exons 2 and 3 sequences for *Mafa-A1* of Vietnamese origin

Site	Ns	S	N	Pi	K
All	546	157	218	0.073	39.886
PBS or TBS	162	81	127	0.162	26.316
Non-PBS or TBS	384	76	91	0.035	13.570

Ns = the number of nucleotides, S = the number of polymorphic sites, N= the number of mutations, Pi = the nucleotide diversity, K = the average number of nucleotide differences