NIH-PA Author Manuscript

Characterization of Mauritian Cynomolgus Macaque Major Histocompatibility Complex Class I Haplotypes by High Resolution Pyrosequencing

Melisa L. Budde[1], **Roger W. Wiseman**[2], **Julie A. Karl**[2], **Bozena Hanczaruk**[3], **Birgitte B. Simen**[3], and **David H. O'Connor**[1,2,*]

[1]Department of Pathology and Laboratory Medicine, University of Wisconsin-Madison, Madison, Wisconsin, USA 53705

[2]Wisconsin National Primate Research Center, University of Wisconsin-Madison, Madison, Wisconsin, USA 53715

[3]454 Life Sciences, A Roche Company, Branford, CT

Abstract

Major histocompatibility complex (MHC) class I alleles of nonhuman primates (NHP) have been associated with disease susceptibility, resistance, and resolution. Here, using high-resolution pyrosequencing, we characterized MHC class I transcripts expressed in Mauritian cynomolgus macaques (MCM), a nonhuman primate population with restricted MHC diversity. Using this approach we identified 67 distinct MHC class I transcripts encoded by the seven most frequent MCM MHC class I haplotypes, 40 (60%) of which span the complete open reading frames. These results double the number of MHC class I sequences previously defined by cloning and Sanger sequencing of cDNA-PCR products and provide a rapid, high-throughput, and economical method for MHC characterization. Overall, this approach significantly expanded our knowledge of MCM haplotypes and will facilitate future studies on disease pathogenesis and protective cellular immunity.

Keywords

Mauritian cynomolgus macaque; MHC; pyrosequencing; *Macaca fasicularis*

Major histocompatibility complex (MHC) class I proteins are present on all nucleated cells and present peptides to circulating CD8+ T cells (Parham 1994). In humans, MHC class I loci are well defined with only three classical polymorphic genes per haplotype (HLA-A, HLA-B, and HLA-C). In contrast, macaque MHC class I genes have undergone complex duplications and deletions, resulting in a heterogeneous assortment of up to 20 MHC class I loci per haplotype (Daza-Vamenta et al. 2004; Otting et al. 2005). Indeed, MHC polymorphism has a profound impact on disease susceptibility and resolution as well as cellular immune responses (Wiseman et al. 2007; Blackwell et al. 2009; O'Connor et al. 2010). Thus, comprehensively understanding MHC class I genetics is important for defining correlates of disease susceptibility and protection in nonhuman primates (NHP).

*Corresponding Author: Address: 555 Science Drive, Madison, WI 53711, Phone: (608) 890-0845, Fax: (608) 265-8084, doconnor@primate.wisc.edu.

Mauritian cynomolgus macaques (MCM) are widely used as models for infectious disease, vaccine, and transplantation research. MCM are unique because they descended from a small founder population on the geographically isolated island of Mauritius (Lawler et al. 1995). Consequently, MCM have extremely restricted MHC genetic diversity consisting of seven common haplotypes, termed M1 to M7. These MHC haplotypes vary in their gene content and population frequency; however, 19% of MCM carry the most frequent MHC class I M1 haplotype (Figure 1) and more than 4% of MCM are homozygous for this haplotype (O'Connor et al. 2010). In contrast, shared MHC class I haplotypes are more rarely observed in unrelated, wild NHP populations.

High-resolution pyrosequencing with a 190-base pair (bp) cDNA-PCR amplicon revealed that the number of MHC class I transcripts per MCM haplotype is significantly greater than previously recognized by cDNA-PCR cloning and Sanger sequencing efforts (Wiseman et al. 2009) (Krebs et al. 2005) (Wiseman et al. 2007). Unfortunately, the short length of this cDNA-PCR amplicon hindered the ability to resolve closely related MHC class I variants. To unambiguously determine the major and minor MHC class I transcripts important to MCM functional studies we developed a novel, high-resolution method for full-length MHC class I sequencing that takes advantage of Roche/454 Titanium amplicon pyrosequencing technology. Ultimately, this approach provides a more comprehensive representation of MHC class I diversity in MCM and is applicable to other NHP used in biomedical research.

Identification of novel MCM MHC class I transcripts by pyrosequencing

To identify full-length MHC class I transcripts we designed three overlapping cDNA-PCR amplicons averaging 576-bp with a series of primers anchored in highly conserved sequences defined by a database of nearly 1,200 previously described macaque class I transcripts (Figure 2). The amplicons span the entire ~1100-bp MHC class I open reading frames and together allow for unambiguous resolution of NHP MHC class I sequences. Due to limited availability of 5′ UTR information, two separate forward primers were used in independent cDNA-PCR reactions to generate 5′ amplicons that were then pooled prior to sequencing. Likewise, three separate 3′ reverse primers were synthesized and pooled for a single 3′ amplicon reaction in attempt to overcome the sequence diversity in the 3′ UTR (Table 1).

To demonstrate the utility and sensitivity of this approach for NHP class I allele discovery, we examined MHC class I transcripts associated with the seven most frequent MCM haplotypes. Initially, we pyrosequenced overlapping amplicons from seven MCM including homozygous M1, M2, M3, and M4 MCM as well as a heterozygous M5, M6, and M7 MCM. Briefly, primary cDNA-PCR products for each amplicon were generated and purified as previously described (Wiseman et al. 2009). Primary amplicons from 12 (including 5 unrelated non-MCM) macaques were normalized and pooled in an equimolar ratio for pyrosequencing in one quarter of a 70×75 PicoTiterPlate on a Roche/454 Genome Sequencer FLX instrument with Titanium amplicon chemistry. The resulting sequences were identified by 10-bp multiplex identifiers incorporated during the primary PCR (Wiseman et al. 2009). For these seven MCM we obtained 72,000 high quality sequence reads with an average of 10,306 reads per individual (range 6,430-13,508). To supplement this we performed a second pyrosequencing experiment with amplicons from three additional MCM that were homozygous for the M5, M6, and M7 haplotypes in a full PicoTiterPlate on a Roche/454 Genome Sequencer Junior (GS Junior) instrument as described above. The GS Junior run resulted in 44,500 high quality sequence reads with an average of 14,589 reads per individual (range 12,245 – 18,349). There were no significant differences between the transcripts obtained from heterozygotes or homozygotes.

As with our 190-bp amplicon pyrosequencing study, we identified all of the major MCM transcripts described previously (Wiseman et al. 2007) (Krebs et al. 2005). In addition, we identified an average of one MHC class I A (*Mafa-A*), three-five MHC class I B (*Mafa-B*) and one MHC class I *I* (*Mafa-I*) transcripts per haplotype that were not previously detected by cloning and Sanger sequencing. All novel sequences were also detected previously with the 190-bp amplicon, albeit with greater ambiguity. Indeed, we doubled the number of MHC class I sequences associated with the previously described MCM haplotypes (Krebs et al. 2005)(Wiseman et al. 2007). In summary, each of the MCM haplotypes carries an average of three *Mafa-A* transcripts, seven *Mafa-B* transcripts, and one *Mafa-I* transcript (Figure 3).

The use of longer overlapping amplicons greatly improved our ability to unambiguously resolve MCM MHC class I allelic variants. We identified 67 distinct MCM MHC class I sequences from total PBMC that ranged in expression from 70% to less than 0.1% transcript levels of total MHC class I sequence reads per individual (Table 2). Full-length sequences were recovered from 40 (60%) of the 67 transcripts identified. Interestingly, we obtained full-length sequence for all but five transcripts (of 36, 14%) expressed at >1% of the total MHC class I transcripts (*Mafa-A1*047:01, -B*05702, -B*079:01, -B*079:03, and -B*151:01*) while only nine transcripts (of 31, 29%) expressed by <1% of the total MHC class I transcripts had full-length sequence recovered (Table 2). An additional 21 sequences (31%) extended through the alpha 1 and alpha 2 domains (peptide binding domain, >600 nucleotides). Further, 32 transcripts identified represent novel MHC class I sequence not previously reported in the Immuno Polymorphism Database (Robinson et al. 2010). Thus, we believe that this approach should be practical for discovery of novel, full-length MHC class I transcripts, present in at least 1% abundance, in a wide variety of NHP populations.

Further examination of the MCM MHC class I haplotypes revealed several interesting phenomena. First, as previously described, the three most common haplotypes, M1, M2, and M3, contain exceptionally similar MHC class I A sequences (Burwitz et al. 2009). Indeed, all three haplotypes transcribe an identical *Mafa-A4*01:01* allele and highly similar *Mafa-A1*063* alleles with a single isoleucine-methionine substitution in the signal peptide of *Mafa-A1*063:02* on the M3 haplotype. Further, the M1 and M2 haplotypes encode an identical minor allele, *Mafa-A2*05:01*, while the M3 haplotype encodes an allelic variant *Mafa-A2*05:11* which differs by seven nucleotides. Likewise, the M4 and M7 haplotypes share *Mafa-A5*30:01* despite carrying *Mafa-A1* alleles that are only 92% identical at the nucleotide level. In contrast, there are no identical *Mafa-B* or *-I* transcripts shared between MCM haplotypes. However, there are several lineages including *Mafa-B*045, -B*046, -B*051, -B*060, -B*079, -B*098, -B11L*01, and -I*01*, that are transcribed from multiple haplotypes and may arise from common ancestral class I loci (Figure 3 and Table 2).

We also identified several transcripts with unusual sequences. First, *Mafa-B*144:02*, a major M1 transcript, lacks a canonical start codon, but has an in-frame start codon 85 nucleotides downstream resulting in complete loss of the leader peptide. Next, *Mafa-B*027:02*, a major M4 transcript, lacks an in-frame stop codon before our 3′ primer site. Finally, *Mafa-B*167:01*, a minor M5 transcript, appears to be a non-functional pseudogene containing a premature stop codon in exon 3. Further work is necessary to determine if these unique sequences restrict functional CD8+ T cell responses.

Currently, very few specific MCM CD8+ T cell responses and their restriction have been identified and published (Burwitz et al. 2009). Resolution of the MHC class I transcripts associated with each MCM haplotype will allow us to evaluate the breadth and functionality of CD8+ T cell responses elicited by an individual gene products as well as entire haplotype. Also, characterization of MHC class I transcripts facilitates generation of peptide binding motifs which can be used to predict epitopes that may be bound and presented by specific

MHC class I gene products. Thus, MCM represent a powerful model for correlation of MHC genetics with successful immune responses.

Each MCM haplotype contains an average of 10 MHC class I transcripts (range 9-12), more than triple the number expressed by humans (Daza-Vamenta et al. 2004). Given our current understanding of cellular immunity and immunological tolerance, it is difficult to reconcile the expression of so many distinct MHC class I transcripts. Several lines of evidence, however, suggest that more total MHC class I transcripts may be beneficial in host pathogen interactions (O'Connor et al. 2010). The relatively simple, solved MHC class I genetics of MCM may allow dissection of the advantages and disadvantages of expressing an increased number of distinct MHC class I gene products.

Intriguingly, nearly half of the MHC class I transcripts identified in this study are expressed at less than 1% of total MHC class I transcripts identified on total PBMC. In contrast, human cells express relatively equal distributions of MHC class I transcripts (Lank et al. in press). Recently, we have discovered that specific MHC class I transcripts that are expressed at low levels in total PBMC can be abundant on distinct cell subsets (J. Greene manuscript submitted). One such MCM transcript, from the MCM M2 haplotype, *Mafa-B*098:01*, was nearly undetectable on all cell subsets except CD14+ monocytes. As shown in Figure 3, variants of the *Mafa-B*098* lineage are also expressed on the M3, M5, and M6 haplotypes. It is likely that these variants are also differentially expressed on CD14+ monocytes since the Mamu-B*098 lineage is also differentially expressed on the Indian rhesus macaque background (J. Greene manuscript submitted). Taken together, this suggests an additional mechanism for molding immune responses and tolerance in the presence of a complex number of MHC class I transcripts.

In this study, we sequenced transcripts from 10 MCM resulting in 67 distinct sequences, more than doubling the number identified by previous cloning and Sanger sequencing. Indeed, MCM are a well-characterized NHP population with exceptionally low genetic diversity. Applying this new technique for full-length MHC class I allele discovery to unique, uncharacterized NHP populations will yield a plethora of novel MHC class I transcripts which may guide current and retrospective immunological studies. Of note, if full-length sequences are unnecessary a shorter amplicon could be used for allele discovery and genotyping. Additionally, this approach could accelerate characterization of other highly polymorphic loci including MHC class II alleles, killer immunoglobulin receptors, and T cell receptors, further advancing immunological studies.

Overall, these data demonstrate that pyrosequencing can be used to successfully reconstruct full-length MHC class I sequences and solve MHC class I genotypes of individuals. There are, however, several improvements that could be made. First, pyrosequencing technology is continually improving and read lengths are expected to soon approach 800-bp. These longer read lengths will decrease the number of amplicons required while facilitating analysis and assembly of full-length transcripts. As new MHC class I transcripts are deposited into the Immuno Polymorphism Database improved primers may be designed to enhance the number of NHP sequences identified. Combined, these approaches will allow for full-length assembly and unambiguous resolution of MHC class I transcripts from numerous NHP populations.

In summary, pyrosequencing of the MCM MHC class I region using three overlapping cDNA-PCR amplicons has doubled the number of transcripts characterized on MCM haplotypes. Approximately five novel transcripts per MCM haplotype were identified that had not been detected previously by traditional cloning and Sanger sequencing of cDNA-PCR products. Thirty-two transcripts contain novel sequences not reported previously in the

Immuno Polymorphism Database. Additionally, our semi-quantitative results demonstrate that the transcript levels on total PBMC can vary greatly ranging from 70% to less than 0.1%, further emphasizing the level of allelic variation within each individual. Overall, this significantly expands our knowledge of MCM MHC class I haplotypes and will facilitate future biomedical studies with this important animal model.

Acknowledgments

References

1. Blackwell JM, Jamieson SE, Burgner D. HLA and infectious diseases. Clin Microbiol Rev. 2009; 22:370–85. [PubMed: 19366919]

2. Burwitz BJ, Pendley CJ, Greene JM, Detmer AM, Lhost JJ, Karl JA, Piaskowski SM, Rudersdorf RA, Wallace LT, Bimber BN, Loffredo JT, Cox DG, Bardet W, Hildebrand W, Wiseman RW, O'Connor SL, O'Connor DH. Mauritian cynomolgus macaques share two exceptionally common major histocompatibility complex class I alleles that restrict simian immunodeficiency virus-specific CD8+ T cells. J Virol. 2009; 83:6011–6019. [PubMed: 19339351]

3. Daza-Vamenta R, Glusman G, Rowen L, Guthrie B, Geraghty DE. Genetic divergence of the rhesus macaque major histocompatibility complex. Genome Res. 2004; 14:1501–1515. [PubMed: 15289473]

4. Krebs KC, Jin Z, Rudersdorf R, Hughes AL, O'Connor DH. Unusually high frequency MHC class I alleles in Mauritian origin cynomolgus macaques. J Immunol. 2005; 175:5230–5239. [PubMed: 16210628]

5. Lawler SH, Sussman RW, Taylor LL. Mitochondrial DNA of the Mauritian macaques (Macaca fascicularis): an example of the founder effect. Am J Phys Anthropol. 1995; 96:133–141. [PubMed: 7755104]

6. O'Connor SL, Lhost JJ, Becker EA, Detmer AM, Johnson RC, Macnair CE, Wiseman RW, Karl JA, Greene JM, Burwitz BJ, Bimber BN, Lank SM, Tuscher JJ, Mee ET, Rose NJ, Desrosiers RC, Hughes AL, Friedrich TC, Carrington M, O'Connor DH. MHC heterozygote advantage in simian immunodeficiency virus-infected mauritian cynomolgus macaques. Sci Transl Med. 2010; 2:22ra18.

7. Otting N, Heijmans CM, Noort RC, de Groot NG, Doxiadis GG, van Rood JJ, Watkins DI, Bontrop RE. Unparalleled complexity of the MHC class I region in rhesus macaques. Proc Natl Acad Sci U S A. 2005; 102:1626–1631. [PubMed: 15665097]

8. Parham P. The rise and fall of great class I genes. Semin Immunol. 1994; 6:373–382. [PubMed: 7654994]

9. Robinson J, Mistry K, McWilliam H, Lopez R, Marsh SG. IPD--the Immuno Polymorphism Database. Nucleic Acids Res. 2010; 38:D863–9. [PubMed: 19875415]

10. Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, Detmer AM, Bouffard P, Levenkova N, Turcotte CL, Szekeres EJ, Wright C, Harkins T, O'Connor DH. Major histocompatibility complex genotyping with massively parallel pyrosequencing. Nat Med. 2009; 15:1322–1326. [PubMed: 19820716]

11. Wiseman RW, Wojcechowskyj JA, Greene JM, Blasky AJ, Gopon T, Soma T, Friedrich TC, O'Connor SL, O'Connor DH. Simian immunodeficiency virus SIVmac239 infection of major histocompatibility complex-identical cynomolgus macaques from Mauritius. J Virol. 2007; 81:349–361. [PubMed: 17035320]

Figure 1. Frequencies of MHC class I haplotypes in MCM
Microsatellite analysis was used to determine the MCM MHC class I haplotype frequencies from 425 feral MCM (850 chromosomes). The seven most frequent haplotypes were designated M1 to M7 and assigned colors (M1, black; M2, red; M3, blue; M4, green; M5, yellow; M6, gray; and M7, purple) throughout the figures for clarification. *We have detected an eighth haplotype by microsatellite analysis in two wild-caught MCMs, as well as a parent and offspring in an independent captive breeding facility. Due to the limited number of individuals with this putative haplotype we were unable to characterize it.
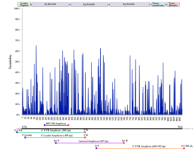
Figure 2. Polymorphic variation of known MCM MHC class I transcripts
Distribution of nucleotide variability for MCM MHC class I products. We aligned all 67 distinct MCM MHC class I sequences identified by pyrosequencing and plotted the frequency of the nucleotide differences. Below the variability plot is a depiction of the cDNA-PCR amplicons used in this study as well as the location of the 190-base pair amplicon used previously (Wiseman et al. 2009). Above the variability plot is a depiction of the macaque MHC class I domain structure. The alpha 1 and alpha 2 domains form the highly polymorphic peptide-binding pocket of the MHC class I transcript.
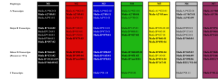
Figure 3. MCM MHC class I haplotypes
MHC class I A, B, and I transcripts associated with each MCM haplotype were identified by pyrosequencing of three overlapping cDNA-PCR amplicons. Our cohort for this experiment included ten animals, seven homozygotes and three heterozygotes. The transcripts that were not previously detected by cloning and Sanger sequencing are denoted in bold. Further, the *Mafa-B* transcripts are divided into major and minor transcripts. We defined minor transcripts as those present at less than 1% of the total MHC class I reads identified in total PBMC. Two putative minor transcripts, denoted by a +, were identified using the 190-bp amplicon in Wiseman et al. 2009 but not in this experiment.

Table 1
cDNA–PCR for full-length pyrosequencing with overlapping amplicons

Amplicon	Primer Name	Gene-Specific Sequence[a]	Product Size (bp)[b]
5′ UTR	5′ UTR	5′-AGAGTCTCCTCAGACGCCGAG	546
5′ Leader	5′ Leader	5′-CCCCGAACCCTCCTCCTG	509
	5′ R	5′-CGTTCAGGGCGATGTAATCC	
Internal	Int F	5′-AGGGKCCGGAGTATTGGGA	547
	Int R	5′-TGATCTCCGCAGGGTAGAAGC	
3′ UTR	3′ F	5′-ACCCAGCGSAAGTGGGA	698-702
	3′UTRa	5′-CCTCGCAGTCCCACACAAG	
	3′ UTRb	5′-CCTGCTTCTCAGTTCCACACAAG	
	3′ UTRc	5′-CTGCATCTCAGTCCCACACAAG	

Primer sequences were adjusted to obtain an average T_m of 63°C so that all PCR reactions could be performed under a uniform set of thermal cycling conditions as described previously (Wiseman et al. 2009).

[a]In addition to the gene-specific sequences shown here, the 5′ end of each forward and reverse primer contains a GS FLX Titanium (Lib-A) adaptor A or B, respectively as well as a 10 bp Multiplex Identifier (MID) sequence

[b]Amplicon lengths include 70 bp for GS FLX Titanium (Lib-A) adaptors and MID's

Table 2

Summary of MCM Transcripts

New Nomenclature*	Previous Nomenclature	Haplotype	Relative Transcript Abundance (%)	Full-Length Sequence Available	Length Recovered from Pyrosequencing (bp)	Immuno Polymorphism Database ID	Genbank ID
Mafa-A1*031:01	Mafa-A*310101	M4	20.9	Yes	1098	MHC01735	AM295836, AY958107
Mafa-A1*032:01	Mafa-A*320101	M6	26.9	Yes	1098	MHC01736	AM295824, AY958108
Mafa-A1*033:01	Mafa-A*330101	M5	19.8	Yes	1098	MHC01737	AY958109
Mafa-A1*047:01	Mafa-A*470101	M6	15.5	No	1090	MHC01802	EF028175, HM581955
Mafa-A1*060:05		M7	40.9	Yes	1098	MHC02610	FJ430071
Mafa-A1*063:01	Mafa-A*250301	M1, M2	8.1	Yes	1098	MHC01769	AM295848, DQ979878
Mafa-A1*063:02	Mafa-A*250201	M3	15.4	Yes	1098	MHC01728	AY958100
Mafa-A2*05:01		M1, M2	0.1	No	951	MHC01781	AM295861
Mafa-A2*05:04		M4	0.4	No	1065	MHC01223	AM295864, EF550527
Mafa-A2*05:11		M3	0.4	No	1065	MHC01789	AM295871
Mafa-A2*05:nov		M5	0.7	No	1065	#	HQ230581
Mafa-A4*01:01	Mafa-A*290101	M1, M2, M3	1.8	Yes	1098	MHC01733	AY958105
Mafa-A4*0109		M6	0.6	No	1074	#	HM581956
Mafa-A5*30:01	Mafa-A*300101	M4, M7	3.0	Yes	1098	MHC01734	AM295882, AY958106
Mafa-B*011:01	Mafa-B*04501	M3	25.9	Yes	1089	MHC01648	AY958143, EU203717
Mafa-B*019:01	Mafa-B*06001	M2	12.0	Yes	1086	MHC02700	DQ979879
Mafa-B*027:02	Mafa-B*09402	M4	6.1	Yes	1096	MHC02730	FM212816, HM581963
Mafa-B*033:01	Mafa-B*06501	M6	3.1	Yes	1086	MHC02704	DQ979883
Mafa-B*036:01:02	Mafa-B*0110102	M5	10.1	Yes	1080	MHC02707	EU606046, FM212796
Mafa-B*037:01	Mafa-B*05001	M5	13.3	Yes	1089	MHC02695	AY958149
Mafa-B*044:04		M7	10.7	Yes	1080	#	HM581964
Mafa-B*045:01	Mafa-B*01201	M5	12.1	Yes	1089	MHC01680	AB195442, EU203690
Mafa-B*045:03	Mf-B*nov011	M6	0.5	Yes	1089	#	GQ153330
Mafa-B*046:01	Mafa-B*0550101	M1	0.9	Yes	1080	MHC02699	EF442021
Mafa-B*046:09		M6	0.5	Yes	1080	#	HM581959
Mafa-B*050:04	Mf-B*nov020	M5	29.1	Yes	1080	#	GQ153337
Mafa-B*051:03	Mf-B*nov018	M4	<0.1	No	425	#	GQ153335

New Nomenclature*	Previous Nomenclature	Haplotype	Relative Transcript Abundance (%)	Full-Length Sequence Available	Length Recovered from Pyrosequencing (bp)	Immuno Polymorphism Database ID	Genbank ID
Mafa-B*051:04	Mafa-B*05402	M5	0.2	No	425	#	HM581967
Mafa-B*057:02		M1	4.3	Yes	691	MHC02699	FM212808, HM581958
Mafa-B*060:01	Mafa-B*07701	M5	4.1	Yes	1080	MHC01682	EU203692
Mafa-B*060:04	Mf-B*nov003	M6	7.2	Yes	1080	#	HM581968
Mafa-B*060:05:01	Mf-B*nov003	M4	0.5	No	682	#	HM581965
Mafa-B*060:05:02	Mf-B*nov003	M1	0.3	No	682	#	GQ153323
Mafa-B*060:06		M7	0.6	No	682	#	HM581966
Mafa-B*064:01	Mafa-B*04601	M1	7.8	Yes	1077	MHC02691	AY958144
Mafa-B*070:02	Mf-B*nov008	M3	0.1	No	1065	#	GQ153328
Mafa-B*072:02	Mafa-B*09001	M7	4.6	Yes	1089	#	HM581961
Mafa-B*075:01	Mafa-B*05101	M3	46.1	Yes	1089	MHC01649	AY958150, EU203718
Mafa-B*079:01	Mafa-B*09301	M3	4.5	Yes	683	MHC02349	AM943362
Mafa-B*079:03		M2	1.5	No	683	#	HM581962
Mafa-B*088:01	Mafa-B*06201	M4	1.0	Yes	1089	MHC02701	DQ979880
Mafa-B*095:01	Mafa-B*0490101	M6	40.0	Yes	1089	MHC02028	AY958148, EU392113
Mafa-B*098:01	Mafa-B*06601	M2	0.5	No	966	MHC02705	EF028176, HM581960
Mafa-B*098:04		M5	1.0	No	981	#	HM581969
Mafa-B*098:05	Mf-B*nov007	M3	0.3	Yes	1089	#	GQ153327
Mafa-B*098:06	Mf-B*nov014	M6	0.1	No	981	#	GQ153332
Mafa-B*104:01	Mafa-B*0440101	M1	64.6	Yes	1089	MHC02689	AY958141
Mafa-B*109:04	Mf-B*nov005	M2	0.5	No	434	#	GQ153325
Mafa-B*131:02	Mf-B*nov001	M1	0.3	No	844	#	GQ153321
Mafa-B*134:01	Mafa-B*04301	M1	5.3	Yes	1089	MHC02688	AY958140
Mafa-B*144:02	Mafa-B*02702	M1	5.6	Yes	1089	MHC02698	EF442022, HM581957
Mafa-B*147:01	Mafa-B*04701	M4	70.1	Yes	1086	MHC02692	AY958145
Mafa-B*148:01	Mafa-B*0480101	M2	19.5	Yes	1080	MHC02694	AY958147
Mafa-B*149:01	Mafa-B*09101	M5	0.9	Yes	1089	MHC02708	EU606047
Mafa-B*150:01	Mafa-B*0630101	M2	52.1	Yes	1089	MHC02702	DQ979881
Mafa-B*151:01	Mafa-B*06401	M6	2.2	Yes	968	MHC02703	DQ979882
Mafa-B*152:01:01N	Mf-B*nov002	M1	0.4	No	434	#	GQ153322

New Nomenclature*	Previous Nomenclature	Haplotype	Relative Transcript Abundance (%)	Full-Length Sequence Available	Length Recovered from Pyrosequencing (bp)	Immuno Polymorphism Database ID	Genbank ID
Mafa-B*164:02	Mf-B*nov015	M7	35.2	Yes	1089	#	GQ153333
Mafa-B*165:01	Mf-B*nov006	M3	0.1	No	667	#	GQ153326
Mafa-B*166:01	Mf-B*nov019	M7	<0.1	No	425	#	GQ153336
Mafa-B*167:01	Mafa-I*nov002	M5	0.6	Yes	1098	#	HM584118
Mafa-B11L*01:04	Mf-B*nov012	M6	0.1	No	948	#	GQ153331
Mafa-B11L*01:05	Mf-B*nov017	M7	0.1	No	658	#	GQ153334
Mafa-I*01:10	Mafa-I*100101	M3	0.8	Yes	1089	MHC02496	DQ979884
Mafa-I*01:11	Mafa-I*100201	M6	0.7	Yes	972	MHC02054	DQ979885, EU392142
Mafa-I*01:22	Mafa-I*01012	M7	0.6	Yes	1089	#	AF161864, HM581970
Mafa-I*03:01	Mafa-I*110101	M4	2.5	Yes	1071	MHC02497	DQ979886

#
represents transcripts that are awaiting an ID from the Immuno Polymorphism Database

*
Recently, there have been several updates to the MCM MHC nomenclature which are reflected in this paper. First, the nomenclature has undergone a transition to include colons in the names (for instance Mafa-A4*0101 is now Mafa-A4*01:01). Additionally, MCM nomenclature has updated the lineage names to better correspond to Indian rhesus macaque lineage names so that comparisons can be made across species.

This table is intended to be a summary of all transcript data available from both this manuscript and previous cloning and Sanger sequencing experiments. Thus, the columns denoting previous nomenclature, full-length sequence availability, IPD ID, and Genbank ID may contain information not directly from this manuscript.