

Akihiko Uda · Kiyoshi Tanabayashi · Osamu Fujita ·  
Akitoyo Hotta · Keiji Terao · Akio Yamada

## Identification of the MHC class I *B* locus in cynomolgus monkeys

Received: 29 August 2004 / Revised: 13 January 2005 / Published online: 3 March 2005  
© Springer-Verlag 2005

**Abstract** By determining the nucleotide sequences of more than 700 cDNA clones isolated from 16 cynomolgus monkeys, we identified 26 *Mafa-B* alleles. In addition, nine sequences with similarity to *Mamu-I* alleles were identified. Since multiple *Mafa-B* alleles were found in each individual, it was strongly suggested that the cynomolgus MHC class I *B* locus might be duplicated and that the *Mafa-I* locus was derived from the *B* locus by gene duplication, as in the case of the *Mamu-I* locus of rhesus monkeys.

**Keywords** Cynomolgus · MHC · *Mafa* · Allele

### Introduction

It is well established that CD8<sup>+</sup> T-cell activation is triggered through recognition of the MHC class I molecule loaded with an antigenic peptide by an antigen-specific T-cell receptor. The MHC molecules of the mammals including primates are known to influence the outcome of many diseases such as infectious diseases, cancer, and metabolic disorders. HLA class I genes are divided into three different categories, classical (*HLA-A*, *-B*, and *-C*), non-classical (*HLA-E*, *-F*, and *-G*), and pseudogene (*HLA-H*, *-J*, *-K*, and *-L*), according to their degree of polymorphism and cell surface expression, and the presence of orthologues of the human *HLA-A*, *-B*, *-E*, *-F*, and *-G* genes were identified in

several species of the Old World monkeys (Alvarez et al. 1997; Boyson et al. 1996a,b; Evans et al. 2000; Lafont et al. 2004; Otting and Bontrop 1993; Prilliman et al. 1996; Sidebottom et al. 2001; Uda et al. 2004). Cynomolgus monkeys as well as rhesus monkeys are preferentially used for biomedical research; however, cynomolgus MHC class I was not extensively studied compared with those in rhesus monkeys. We have previously reported the nucleotide sequences of cynomolgus MHC class I *A* locus and have shown that at least 14 *Mafa-A* alleles were present in cynomolgus monkeys (Uda et al. 2004). Although the MHC class I *B* locus is the most polymorphic MHC locus in primates, little information is available concerning the MHC class I *B* locus of cynomolgus monkeys. In this study, therefore, we have expanded our analysis on cynomolgus MHC class I genes and identified 26 *B* locus alleles by analyzing 16 monkeys. We have also found the presence of a novel locus that is very similar to MHC class I *I* locus recently identified in rhesus monkeys.

### Materials and methods

#### Animals

All cynomolgus monkeys were raised and reared in the Tsukuba Primate Center for Medical Science, the National Institute of Infectious Diseases (NIID). Both genders were involved, and the cynomolgus monkeys were between 5 and 24 years old. This study was conducted in accordance with the Guides for Animal Experiments Performed at the NIID.

#### RT-PCR and nucleotide sequencing

Preparation of mRNA from peripheral blood mononuclear cells (PBMC) and RT-PCR were performed as described before (Uda et al. 2004). Primers used in this study are listed in Table 1. 5' MBS and 3' MBS primers designed to amplify the gene products of the rhesus MHC class I *B*

A. Uda · K. Tanabayashi · O. Fujita · A. Hotta ·  
A. Yamada (✉)  
Department of Veterinary Science,  
National Institute of Infectious Diseases,  
1-23-1 Toyama, Shinjuku,  
Tokyo, 162-8640, Japan  
e-mail: yamada@nih.go.jp  
Tel.: +81-3-52851111  
Fax: +81-3-52851150

K. Terao  
Tsukuba Primate Center for Medical Science,  
National Institute of Infectious Diseases,  
1 Hachimandai, Tsukuba,  
Ibaraki, 305-0843, Japan

**Table 1** Primers used for the amplifications and sequencing of MHC class I cDNA from cynomolgus monkeys

Primers	Binding region (position)	Sequence <sup>a</sup>
Primers for amplification		
5' Beta 3 XMO <sup>b</sup>	All loci exon 1 (-27-1)	5'-CGC <u>TCG AGG</u> ACT CAG AAT CTC CCC AGA CGC CGA G-3'
Mafa-B1a	B and I loci exon 8 (1089-1117)	5'-CCA CTT AAG ACA GTT TCA GGC TTT T-3'
5' MBS <sup>b</sup>	B and I loci exon 1 (10-34)	5'-GCC <u>TCG AGA</u> ATT CAT GGC GCC CCG AAC CCT CCT CCT GC-3'
3' MBS <sup>b</sup>	B and I loci exon 8 (1095-1116)	5'-GCA <u>AGC TTC</u> TAG ACC ACA CAA GAC AGT TGT CTC AG-3'
Primers for sequencing		
T7 primer	pCR4Blunt-TOPO vector (328-347)	5'-TAA TAC GAC TCA CTA TAG GG-3'
T3 primer	pCR4Blunt-TOPO vector (243-262)	5'-AAT TAA CCC TCA CTA AAG-3'
Ia698	All loci exon 4 (680-698)	5'-TAG AAG CCC AGG GCC CAG C-3'
Is437	All loci exon 3 (437-456)	5'-ATT ACA TCG CCC TGA ACG AG-3'

<sup>a</sup>*Xho*I, *Sal*I, and *Hind*III sites of 5' beta 3 XHO, 5' MBS, and 3' MBS primers, respectively, are *underlined*

<sup>b</sup>From Boyson et al. (1996b)

locus by Boyson et al. (1996b) were also used to amplify the cynomolgus MHC class I *B* locus. PCR amplification was performed at least twice for each animal. PCR products were cloned into pCR4Blunt-TOPO plasmids (Invitrogen, Carlsbad, Calif., USA) and 48 clones were sequenced by 310 Capillary DNA Sequencer (Applied Biosystems, Foster City, Calif., USA) or 3100-Avant Capillary DNA Sequencer (Applied Biosystems). The *Mafa-B* nucleotide sequences were assembled with the Contig Manager of the DNASIS pro (Hitachi Software, Yokohama, Japan). The Clustal W algorithm provided in DNASIS PRO was used to align sequences.

### Phylogenetic analysis

The full-length nucleotide sequences of *Mafa-B*, *Mafa-I*, *Mafa-A*, *Mamu-A*, *Mamu-B*, *Mamu-I*, *HLA-A*, and *HLA-B* were aligned using Clustal W provided online by the DNA Data Bank of Japan [(DDBJ) <http://www.ddbj.nig.ac.jp>]. A phylogenetic tree of these nucleotide sequences was constructed by the neighbor-joining method of the Molecular Evolution Genetics Analysis, version 2.1 (MEGA 2.1). Genetic distances were estimated using the method of Jules-Canter. At the sites in which alignment indicated a gap, nucleotides at this position in all the sequences were deleted. The reliability of the tree topology was tested by the bootstrap method. Thousand relationships and 64,238 random seeds were used for determining bootstrap values (Fig. 2a, b). Since the bootstrap values of less than 50% were unreliable, the values of less than 50% were not shown in Fig. 2a, b.

**Fig. 1** Deduced amino acid sequences of *Mafa-B* and *Mafa-I* alleles. ▶ Amino acid sequences of *HLA-A*, *HLA-B*, *Mamu-A*, *Mamu-B*, *Mamu-I*, *Mafa-A*, and *Mafa-E* alleles were also included. Amino acids identical to those of *HLA-B\*2702* are indicated by *dots*. The deletions of amino acids are indicated by *hyphens*. The total numbers of clones obtained and the numbers of animals having the allele were indicated *after the allele name*

Clone /animal	Leader peptide	
	-20	-10
HLA-B*2702	MRVT	APRTL L L L L L W GAVALTETWA
HLA-B*5701	.....	V.....
Mamu-B*02	..M	.....S..L.....
Mamu-B*03	..M	...F...S..L.....
Mamu-B*04	..M	...F...S..L.....
Mafa-B*01	7/2	..M.....S..L.....
Mafa-B*02	3/1	.Q.M.....S..L.....
Mafa-B*03	13/3	..M.....S..L.....
Mafa-B*04	9/2	..M.....S..T.LS.....
Mafa-B*05	36/2	..M...I...S..T.LS.....
Mafa-B*06	6/1	..M...I...S..T.LS.....
Mafa-B*07	28/1	..M.....S..L.....
Mafa-B*08	2/1	..M.....S..T.LV.....
Mafa-B*09	21/2	..M.....S..L.....
Mafa-B*10	4/1	----S..L.....
Mafa-B*11	16/1	----S..L.....
Mafa-B*12	17/2	..M.....S..L.....
Mafa-B*13	4/1	..M.....S..L.....
Mafa-B*14	33/2	..M.....S..P.L.....
Mafa-B*15	10/1	----S..L.....
Mafa-B*16	25/4	..M.....S..A.L...K...
Mafa-B*17	45/3	.Q.M.....S..L.S.....
Mafa-B*18	3/1	..M.....S..L...Q...
Mafa-B*19	20/2	.QIM.....S..L.....
Mafa-B*20	6/2	..DM.....S..L.....
Mafa-B*21	22/2	.QIM.....S..L.....
Mafa-B*22	36/2	..M.....S..L.....
Mafa-B*23	22/2	..M.....S..L...R.
Mafa-B*24	10/3	..M..G.....S..L.....
Mafa-B*25	7/3	..FM.....S..L...Q...
Mafa-B*26	5/1	..M..G.....S..L.....
Mamu-I*04		----HS..L.....
Mamu-I*07		..M..G.....S..L.....
Mamu-I*08		----S..L.....
Mafa-I*01013	13/1	..M.....S..L.....
Mafa-I*02	12/1	..M..G.....S..L.....
Mafa-I*03	33/3	..M..G.....S..T.L.....
Mafa-I*04	2/1	..M..G.....S..L.....
Mafa-I*05	11/1	..M..G.....S..L.....
Mafa-I*06	16/2	..M..G.....S..L.....
Mafa-I*07	3/1	..M..G.....S..L.....
Mafa-I*08	5/1	..M..G.....S..L.....
Mafa-I*09	51/3	..M..G.....S..L.....
HLA-A*0201		.A.M.....V...S..L...Q...
Mamu-A*01		----V...S..L.V...Q.R.
Mamu-A*02		----V...S..L...Q.R.
Mafa-A*01		.A.M.....V...S..L...Q.R.
Mafa-A*02		.A.M.....V...S..V.L...Q.R.
Mafa-A*06		.A.M.....V...S..F.L...Q.L.

	Alpha 1 domain									Alpha 2 domain
	10	20	30	40	50	60	70	80	90	100
HLA-B*2702	GSHSMRYFHT	SVSRPGRGEP	RFITVGYVDD	TLFVRFSDA	ASPREEPRAP	WIEQEGPEYW	DRETQICKAK	AQTDRENLR	ALRYYNQSEA	GSHTLQNMYG
HLA-B*5701	Y. AM.	A.	Q.	MA.			G. RNM. S	Y.		II. V.
Mamu-B*02	F. S. A.	R. WYLE.	Q.	E. M.	V.		N. RNS. VT	F. VG. GN	LRG.	K.
Mamu-B*03	S.	S.	Q.	E.	M.		EE. RNA. GH	AD. GN	LRG.	T.
Mamu-B*04	SA A.	YLE.	Q.	M.	V.		EE. RRA. GN	F. VG. GN	LRG.	Y. W.
Mafa-B*01	T. A.	V.	Q.	E. M.	T. M.		EEQ. R. V. DN	F. VD. GT	LRG.	I. T.
Mafa-B*02	T. A.	V.	Q.	E. M.	T. M.		EEQ. R. V. DN	F. VD. GT	LRG.	I. T.
Mafa-B*03	F. S. A.	R. S.	Q.	E. M.	V.		N. RNS. VT	F. VS. GN	LRG.	K.
Mafa-B*04	S. A.	R. WYLE.	Q.	E. M.	V.		EE. RRA. N	VS. GN	LR.	V. I.
Mafa-B*05	L. A.	R. WYLE.	Q.	E. M.			N. RNA. H	VD. GT	LRG.	G
Mafa-B*06	L. AL.	W. Y.	Q.	E. M.	M.		EE. R. A. N	VD. GT	LRG.	G
Mafa-B*07	T. AL.	A.	Q.	E. M.	R.		EEQ. R. A. DA	F. VG. G.	LRG.	Y. W. S.
Mafa-B*08	S.	W. A.	P.	E. M.	V.		EEQ. R. A. DV	F. VG. GT	LRG.	F. R. S.
Mafa-B*09	L. Y. T.	A.	Q.	E. M.	R.		EEQ. RRV. R	QVD. GT	LRG.	G
Mafa-B*10	G. T.	Y.	Q. M.	E. M.	V.		EDV. RRA. R	VD. GT	LRG.	G
Mafa-B*11	Y.	A.	Q.	E. M.	V.		Q. NM. TA	T. AD. GT	L.	RG
Mafa-B*12	L. A.	W. S.	Q. Y.	E.	M.		EEH. R. A. N	H. G. T	LRG.	G
Mafa-B*13	L. A.	W. S.	Q.	E.	M.		R. A. DA	H. G. T	LRG.	D
Mafa-B*14	L. A.	Y.	Q.	M.	M.		N. RKA. DN	VD. GT		G
Mafa-B*15	L. S. T. Q.	W. A.	Q.	E. M.	M.		RNA. N	V. T	L.	K.
Mafa-B*16	L. S. A.	R. WYVE.	Q.	E. M.	M.		N. RRA. GH	H. G. T	L.	G
Mafa-B*17	T. A.	R. WYLE.	Q. V.	E. M.	V.		N. RRA. GN	E. G. T	L.	G
Mafa-B*18	S. A.	R. WYLE.	Q. W. A.	E. M.	V.		N. RRA. GN	F. VD. GN	LRG.	G
Mafa-B*19	S. A.	R. YLE.	Q. W.	E. M.	V.		N. RNA. GH	F. G. T		G
Mafa-B*20	T. A.	R. WYLE.	Q. V.	E. L.	M.		EE. RRA. ET	F. G. T		DG
Mafa-B*21	L. A.	W. S.	Q.	E. M.	M.		EE. R. A. N	H. VD. T		G
Mafa-B*22	T. VM.	D. R. A WYLE.	Q. V.	E. M.	V.		EEQ. RNS. N	H. VD. T		G
Mafa-B*23	L. A.	A.	Q.	E.	M.		EEQ. R. A. N	H. VD. GT	L.	C. T.
Mafa-B*24	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. WT		G
Mafa-B*25	S. A.	A.	Q.	E. M.			RRV. GN			G
Mafa-B*26	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. D. T		G
Mamu-I*04	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. T		G
Mamu-I*07	L. G. T.	A.	Q.	E. M.	M.		EEQ. R. A. R	E. G. T		G
Mamu-I*08	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. T		G
Mafa-I*01013	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. T		G
Mafa-I*02	L. G. T. Q.	A. N.	Q.	E. M.	M.		EE. R. A. R	GT. T	L.	G
Mafa-I*03	L. G. T. Q.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. T		G
Mafa-I*04	L. G. T. Q.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. T		G
Mafa-I*05	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. T		G
Mafa-I*06	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. T		G
Mafa-I*07	L. G. T.	A.	Q.	E. KM.	M.		EE. R. A. R	E. G. T		G
Mafa-I*08	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. WT		G
Mafa-I*09	L. G. T.	A.	Q.	E. M.	M.		EE. R. A. R	E. G. WT		G
HLA-A*0201	F.	A.	Q.	Q. M.			G. RKV. H	S. H. VD. GT	LRG.	V. R.
Mamu-A*01	K. Y. M.	Q. A.	Q.	Q. M.	V.		RNM. TE	T. NAPV.	T. L.	R. V.
Mamu-A*02	Y. M.	W. A.	Q.	Q. M.	V.		RNM. E	T. NAPV.	N. LRG.	I. R.
Mafa-A*01	S. Y.	Q. A.	Q.	Q. M.	V.		RNM. TE	T. MAPVD. QN	LRG.	F. T.
Mafa-A*02	S. Y. YM.	VA.	Q.	Q. M.	V.		N. R. M. E	T. NAPV.	N. LRG.	Y. M.
Mafa-A*06	Y. A.	A.	Q.	Q. M.	V.		RNM. TA	T. NAPV.	N. LRG.	R. V.

Fig. 1 (continued)

## GenBank accession numbers

The *Mafa-B* and *Mafa-I* sequences described in this manuscript had been deposited in the DDBJ and were assigned accession numbers AB195431 to AB195465. We previously deposited *Mafa-A* alleles in the DDBJ, and these alleles were assigned accession numbers AB154760 to AB154773. The GenBank accession numbers for other sequences used in this study are as follows: *HLA-A\*0201*, U07161; *HLA-B\*2702*, L38504; *HLA-B\*5701*, AJ458991;

*Mafa-E\*01*, U02976; *Mamu-A\*01*, U50836; *Mamu-A\*02*, U50837; *Mamu-A\*03*, U41379; *Mamu-A\*04*, U41380; *Mamu-B\*02*, U41833; *Mamu-B\*03*, U41825; *Mamu-B\*04*, U41826; *Mamu-B\*05*, U41827; *Mamu-B\*06*, U41828; *Mamu-B\*07*, U41829; *Mamu-B\*08*, U41830; *Mamu-B\*36*, AJ556886; *Mamu-I\*01011*, AF161865; *Mamu-I\*02012*, AF161869; *Mamu-I\*04*, AF4161874; *Mamu-I\*07*, AF161875; *Mamu-I\*08*, AF161876; *Mamu-I\*09*, AF161877; *Mamu-I\*10*, AF161878; and *Mamu-I\*11*, AF161879.

	110	120	130	140	150	160	170	180	Alpha 3 domain		
									190	200	
HLA-B*2702	CDVGPDGRLL	RGYHQDAYDG	KDYIALNEDL	SSWTAADTAA	QITQRKWEAA	RVAEQLRAYL	EGECVEWLR	YLENGKETLQ	RA	DPPKTHVT	HHPISDHEAT
HLA-B*5701	..L..	..HD.S..	..	..	..	..	..L..	..	..	..	..
Mamu-B*02	..L..	..F..	..	R..M..N..	GE..M..T..	H..	..	..	..	..V..	..
Mamu-B*03	..L..	..Y..	..F..	R..V..	..E..V..T..	..	..	..	..	..	..
Mamu-B*04	..	..D.F..	..Q..	R..V..N..	GE..Q..T..	..	..	..	..	..KR..D..V..	..
Mafa-B*01	..L..	..Y..	R..	R..E..N..	G..W..K..C..M..	..	..	..	..	..V..	..
Mafa-B*02	..L..	..Y..	R..	R..E..N..	..E..M..L..H..	..	..	..	..	..VP..	..
Mafa-B*03	..L..	..D.S..	..	R..M..N..	GE..M..T..	H..	..	..	..	..V..	..
Mafa-B*04	..L..	..S..	..	R..VM..N..GD	..Y..RF..R..H..	..	..	..	..	..V.N..	..
Mafa-B*05	..L..	..R..	..	R..V..N..GD	..Y..RF..T..	..	..	..	..	..V..	..
Mafa-B*06	..L..	..R..	..	R..I..N..T	..Y..RF..T..	..	..	..	..	..V..	..
Mafa-B*07	..N..	..H..	..F..	R..G.M..N..V	GE..RF..R..	..	..	..	..	..Y..VF..	..
Mafa-B*08	..L..	..H..	..E.T..	R..M..N..D	..Y..RF..T..L..	..	..	..	..	..V..	..
Mafa-B*09	..L..	..R..	..	R..V..N..K..	G..R..T..	..	..	..	..	..V..	..
Mafa-B*10	..LE..	..R..	..	R..M..N..	G..M..	..	..	..	..	..V..	..
Mafa-B*11	..L..	..E.F..	R..	R..L..N..	GE..W..	..	..	..	..	..	..
Mafa-B*12	..L..	..D.Y..	..V..	R..M..N..	..A..RQ..L..M..	..	..	..	..	..V..	..
Mafa-B*13	..L..	..H..	..D.Y..	..V..	R..M..N..	..A..RQ..L..	..	..	..	..R..	..
Mafa-B*14	..Y.E..	..R..	..Y..	R..M..N..	G..RV..P..M..	..	..	..	..	..V..	..
Mafa-B*15	..L..N..	..Q..	..	R..M..N..K..GD	..Y..RF..L..K..Q..	..	..	..	..	..V..	..
Mafa-B*16	..L..	..Y.H..	..	R..M..N..	..E..W..G.L..	..	..	..	..	..Y..V..	..
Mafa-B*17	..L..	..S..	..	R..M..RF..	..E..M..L..H..	..	..	..	..	..V..	..
Mafa-B*18	..L..	..F..	..	R..M..RF..	..E..Q..L..H..	..	..	..	..	..V..	..
Mafa-B*19	..L..	..F..	..	R..M..RF..	..E..Q..L..	..	..	..	..	..V..	..
Mafa-B*20	..L..E..	..D.H..	..	R..M..N..E..	..E..M..R.L..	..	..	..	..	..	..
Mafa-B*21	..L..	..Y..	..Q..	R..M..N..	GE..R..R..	..	..	..E	..	..F..	..
Mafa-B*22	..L..	..Y..	..Q..	R..M..N..	GE..R..R..	..	..	..E	..	..V..	..
Mafa-B*23	..L..	..S..	..	R..R..HN..	..A..LQ..R.L..	..	..	..	..	..V..	..
Mafa-B*24	..L..	..Y.R..	..	H..L..N..	G..R..R.L..	..	..	..	..	..V..TI	..
Mafa-B*25	..L..	..Y.R..	..	H..L..N..	G..R..R.L..	..	..	..	..	..V..TI	..
Mafa-B*26	..L..	..Y.S..	R..	R..GK..N..	G..R..L.L.S..A..	..	..	..	..	..	..
Mamu-I*04	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mamu-I*07	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mamu-I*08	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*01013	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*02	..L..	..Y..	R..	R..E..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*03	..L..R..	..Y.S..	R..	R..GE..HN..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*04	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*05	..L..	..Y.S..	R..	R..GV..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*06	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*07	..L..	..Y.S..	R..	R..GE..N..	GE..W..R..K..	..	..	..	..	..V..P..	..
Mafa-I*08	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
Mafa-I*09	..L..	..Y.S..	R..	R..GE..N..	GE..R..R..K..	..	..	..	..	..V..	..
HLA-A*0201	..S.W.F..	..Y..	..K..	R..M..T.KH..	H..	..	..	..T	..A..M..AV..	..	..
Mamu-A*01	..L..	..E.Y..	..	R..V..N..D..	D..SM..Q..P..K..T	..	..	..	..	..V..	..
Mamu-A*02	..L..	..S..	..	R..M..N..	GE..H.T..L..	..	..	..	..	..V..Q..	..
Mafa-A*01	..L..	..E.F..	R..	R..M..N..	G..M.V..R.L..	..	..	..	..	..V..Y..	..
Mafa-A*02	..L..	..D.F..	..D..	R..L..N..	G..XII.T..L..	..	..	..	..	..V..Y..	..
Mafa-A*06	..L..	..E.Y..	..F..	R..L..N..	G..I..L..S..	..	..	..	..	..V..	..

Fig. 1 (continued)

## Results

### Detection of 26 MHC class I *B* locus alleles and nine *I* locus alleles in cynomolgus monkeys

To amplify cynomolgus MHC class I *B* locus genes, PCR was carried out using primers that were successfully used for amplification of rhesus MHC class I *B* locus genes along with newly designed ones (Table 1; Boyson et al. 1996b). We obtained 48 clones from each animal. The nucleotide sequences that were found in just one clone

were excluded from the subsequent analyses to avoid incorporation of artificial sequences generated by PCR error or during the cloning procedure into public databases. Ambiguous sequences were also excluded. When the nucleotide sequence was shared by more than two clones, regardless of whether they were derived from one animal or multiple animals, the sequences were regarded as a consensus sequences representing a particular alleles of each animal. Eventually, 43 candidate alleles were obtained, and 34 of 43 were found to have substantial homology with *Mamu-B* alleles. Amino acid sequences deduced from the nucleo-

	210	220	230	240	250	260	270	Transmembrane domain			
	LRCWALGFYP	AEITLTWQRD	GEDQTQDTEL	VETRPAGDRT	FQKWAAVVVP	SGEEQRYTCH	VQHEGLPKPL	TLRW	280	290	300
HLA-B*2702									EPSSQS	TPVIVGIVAG	LAVLAVVVIG
HLA-B*5701											
Mamu-B*02	V			G	G		E		I		T
Mamu-B*03			E	G	G	H	E		I		T
Mamu-B*04				G	G		Q	E	I	—	T
Mafa-B*01		S RQ	E	G	G			LE	SI		T
Mafa-B*02				G	G			E	I	M	T
Mafa-B*03	V			G	G			E	I		T
Mafa-B*04	V			G	G			E	I	M	T
Mafa-B*05			I F	G	G			E	I	V	T
Mafa-B*06			I F	G	G			E	I	A V	T
Mafa-B*07				G	G			Q	I		T
Mafa-B*08	R			G	G			RE	I		T
Mafa-B*09	V			G	G			E	I		T
Mafa-B*10			E	G	G			E	I	S	T
Mafa-B*11				G	G	H		E	I		I T
Mafa-B*12			E	G	G			E V	I		T
Mafa-B*13				G	G			LE	S	I	T
Mafa-B*14			I F	G	G			LE	I	V	T
Mafa-B*15			E	G	NG			E	I		T
Mafa-B*16			E	G	G			RE	I	G	T
Mafa-B*17				G	G			E	I	M	—T
Mafa-B*18				G	G			E	I		—T
Mafa-B*19				G	G			E	IA		—T
Mafa-B*20				G				E	E	I	T
Mafa-B*21			E	G	G	H		E	I	V	T
Mafa-B*22			E	G	G	H		E	I	V	T
Mafa-B*23			E	G	G			E	I	M V	—T
Mafa-B*24	D			G	G			E	I		T
Mafa-B*25	D			G	G			E	I		P T
Mafa-B*26			E F	G	G			E	IA	V	I T
Mamu-I*04			E	G	GN			E	I	M	T
Mamu-I*07				G	GN			E	I	M	T
Mamu-I*08				G	GN			E	I	M	T
Mafa-I*01013				G	GN			E	I	M	T
Mafa-I*02				G	GN			E	I	M	T
Mafa-I*03			E	G	GN			E	I	M	T
Mafa-I*04			E	G	GN			E	I	M	T
Mafa-I*05			E	G	GN			E	I	M	T
Mafa-I*06			E	G	GN			E	I	M	T
Mafa-I*07			E	G				E	I	M	T
Mafa-I*08			E	G	GN			E	I	M	T
Mafa-I*09			E	G	GN			E	I	M	T
HLA-A*0201	S			G		Q			P	I	I VLFGA IT
Mamu-A*01				G			H	K	F	I	M I VL GA T
Mamu-A*02				G		K	RE		IL	I	VL GI
Mafa-A*01		G	E	G				E	I	I	VL GA T
Mafa-A*02				G		K H		E	I	I	VL GA T
Mafa-A*06	V			G		K		K	I	I	VL GA T

Fig. 1 (continued)

tide sequences of these 34 candidate *B* alleles were further subjected to phylogenetic analysis using the neighbor-joining method (Saitou and Nei 1987; data not shown). When the predicted amino acid sequence variation between two candidates was negligible ( $d < 0.025$ ), the amino acid sequence shared by a majority of the clones was regarded as representing a particular allele. The other sequence shared by a minority of clones was excluded from the subsequent analyses. As the result of the analysis, 26 *Mafa-B* alleles were identified. It was found that the remaining nine candidate alleles were closely related to those of *Mamu-I* locus

reported by Urvater et al. (2000b). Since Urvater et al. also identified two *Mafa-I* alleles (*Mafa-I\*01011* and *Mafa-I\*01012*), we named tentatively alleles identified here *Mafa-I\*01013* through *Mafa-I\*09*. The *Mafa-I\*01013* allele was identical in amino acid sequence with *Mafa-I\*01011* and *Mafa-I\*01012*, but there were several synonymous nucleotide changes scattered around the sequence. We therefore considered that this alleles was a variant of *Mafa-I\*01*, although reported sequences of *Mafa-I\*01011* and *Mafa-I\*01012* were incomplete. The deduced amino acid sequences of *Mafa-B* and *Mafa-I* alleles were shown in Fig. 1



		Cytoplasmic domain			
		310	320	330	340
HLA-B*2702	AVVAAVMC	RR	KSSGGKGGSY	SQAACSDSAQ	GSDVSLTA*--
HLA-B*5701	.....	..	.....	.....	.....
Mamu-B*02	..... W	..	..... S.	.....	.....
Mamu-B*03	..... W	..	..... SN.	.....	.....
Mamu-B*04	..... W	..	..... S.	.....	.....
Mafa-B*01	..... R	..	..... S.	.....	.....
Mafa-B*02	..... M.	. K . T . . . . .	F . . SK . . . . .	M . . . . .	.....
Mafa-B*03	..... W	..	..... S.	.....	.....
Mafa-B*04	..... W	..	..... S. N.	.....	.....
Mafa-B*05	..... W	..	..... SN.	.....	.....
Mafa-B*06	..... W	..	..... SN.	.....	.....
Mafa-B*07	..... W	..	..... S.	.....	.....
Mafa-B*08	..... W	..	..... S.	.....	.....
Mafa-B*09	..... W	..	..... SN.	.....	.....
Mafa-B*10	..... W	..	..... S.	.....	.....
Mafa-B*11	..... W	..	..... S.	.....	.....
Mafa-B*12	..... W	..	..... S.	.....	.....
Mafa-B*13	..... W	..	..... S.	.....	.....
Mafa-B*14	..... W	..	..... SN.	.....	.....
Mafa-B*15	..... W	..	..... SN.	.....	.....
Mafa-B*16	..... W	. K . . S . . R . . . . .	S . . . . .	M . . . . .	.....
Mafa-B*17	..... M . . . . .	. . . T . . . R . . . . .	F . . SK . . P . . . . .	E . . M . . . . .	.....
Mafa-B*18	..... M . R . . . . .	. . . . .	F . . SK . . P . . . . .	E . . RS . . . . .	.....
Mafa-B*19	..... W	..	..... F . . SK . . P . . . . .	E . . M . . . . .	.....
Mafa-B*20	..... W	..	..... S.	.....	.....
Mafa-B*21	..... W	..	..... S.	.....	.....
Mafa-B*22	..... W	..	..... S.	.....	.....
Mafa-B*23	..... W	. K RT . . R . . . . .	F . . S . . . . .	.....	.....
Mafa-B*24	..... W	..	..... S.	.....	.....
Mafa-B*25	..... W	..	..... S.	.....	.....
Mafa-B*26	..... W	..	..... SN.	.....	.....
Mamu-I*04	..... W	..	..... S.	.....	.....
Mamu-I*07	..... W	..	..... S.	.....	.....
Mamu-I*08	..... W	..	..... S. N.	.....	.....
Mafa-I*01013	P . . . . . W	..	..... S. N.	.....	.....
Mafa-I*02	P . . . . . W	..	..... S. N.	.....	.....
Mafa-I*03	..... W	..	..... S. N.	.....	.....
Mafa-I*04	..... W	..	..... S. N.	.....	.....
Mafa-I*05	..... W	..	..... S. N.	.....	.....
Mafa-I*06	..... W	..	..... S.	.....	.....
Mafa-I*07	..... W	..	..... S. N.	.....	.....
Mafa-I*08	..... W	..	..... S.	.....	.....
Mafa-I*09	..... W	..	..... S. N.	.....	.....
HLA-A*0201	..... W	..... DR . . . . .	..... S . . . . .	..... CK V*	.....
Mamu-A*01	..... W	..... DR . . . . .	..... S . . . . .	..... CK V*	.....
Mamu-A*02	.. I . . IW . . . . .	..... DR . . . . .	..... S . . . . .	..... CK V*	.....
Mafa-A*01	..... W	..... DR . . . . .	..... SN . . . . .	..... CK V*	.....
Mafa-A*02	... V . . W . . . . .	..... DR . . . . .	..... S . . . . .	..... CK V*	.....
Mafa-A*06	... T . . W . . . . .	..... DR . . . . .	..... S . . . . .	..... CK V*	.....

Fig. 1 (continued)

along with those of alleles reported for other primates. The total numbers of clones obtained and the numbers of animals having the allele were shown in the figure. The putative glycosylation site was located at residue 86, and the conserved cysteine residues occurred at positions 101 and 164 in  $\alpha 2$  and at positions 203 and 259 in  $\alpha 3$ . To evaluate whether the nucleotide sequences of *Mafa-B* and *Mafa-I* alleles established in this study were gene products of class I *B* and *I* loci, respectively, *Mafa-B* and *Mafa-I* alleles were phylogenetically analyzed (Fig. 2a). The full-length nucleotide sequences of *Mafa-B*, *Mafa-I*, *Mafa-A*, *Mamu-A*, *Mamu-B*, *Mamu-I*, *HLA-A*, and *HLA-B* were aligned by Clustal W. A

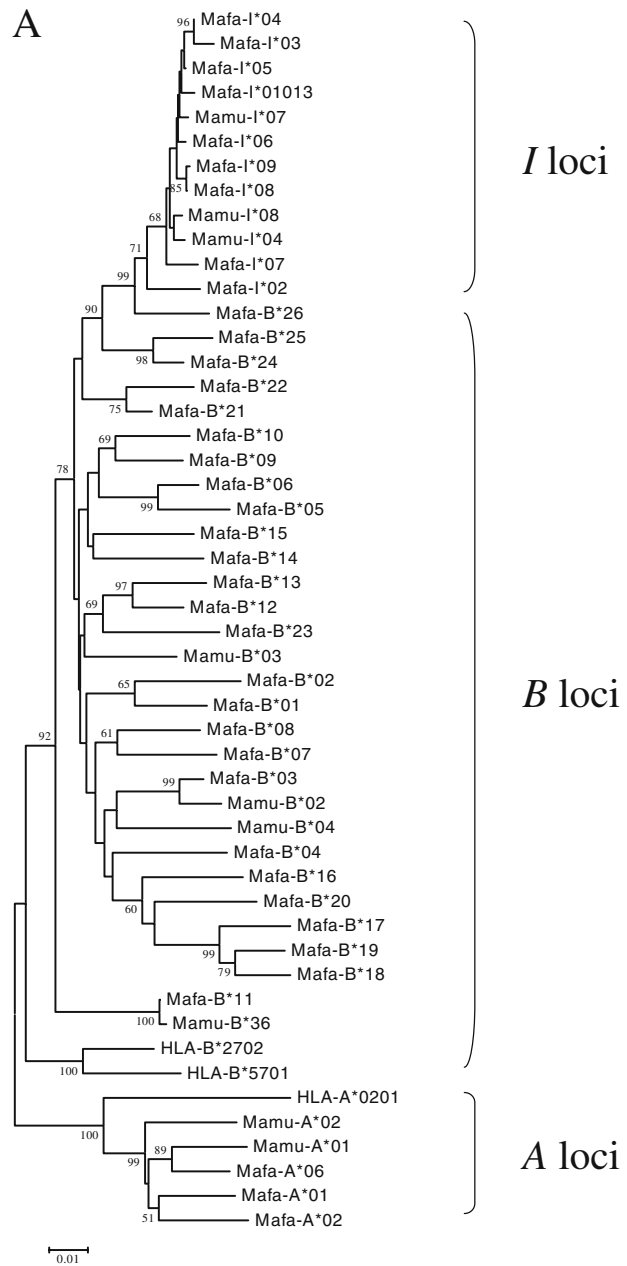


Fig. 2 Phylogenetic analysis of primate class I MHC molecules. The phylogenetic tree was constructed using a full-length and b exon five to eight nucleotide sequences by neighbor-joining method with MEGA2.1. The bootstrap values of more than 50% were shown

phylogenetic tree was constructed by the neighbor-joining method of MEGA2.1 software. The reliability of the tree topology was tested by the bootstrap method, and the bootstrap values are shown in Fig. 2a. Since the bootstrap values of less than 50% were unreliable, the bootstrap values of greater than 50% are shown in Fig. 2a. Several *Mafa-B* alleles (*Mafa-B*\*21, 22, 24, 25, and 26) appeared to cluster with *Mamu-I* or *Mafa-I* allele rather than *B* locus alleles. Since amino acid difference between alleles of *I* and *B* loci were more apparent at the carboxy half of the protein, we recon-

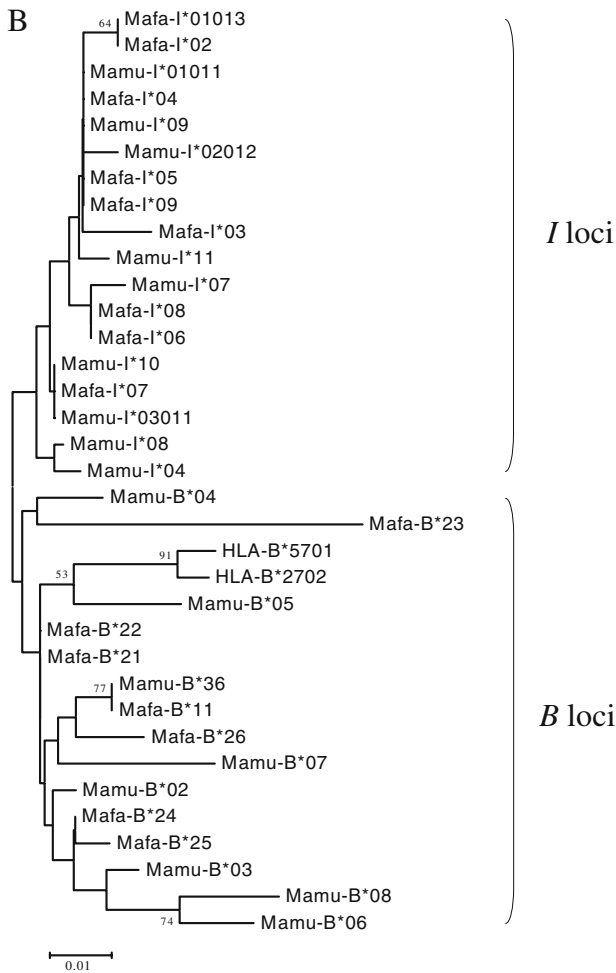


Fig. 2 (continued)

Table 2 Segregation of alleles with haplotypes

Haplotypes	Alleles	Animal no.						
		3032	3028	1159	1113	0079	7071	0068
A	<i>Mafa-B*03</i>							○
	<i>Mafa-I*09</i>				○			○
B	<i>Mafa-B*24</i>		○			○	○	
	<i>Mafa-B*25</i>		○			○	○	
C	<i>Mafa-B*17</i>			○				○
	<i>Mafa-B*20</i>			○				○
	<i>Mafa-I*06</i>			○				○
D	<i>Mafa-B*01</i>			○				○
	<i>Mafa-B*04</i>			○				○
E	<i>Mafa-B*16</i>	○			○	○		
	<i>Mafa-I*03</i>	○			○	○		
F	<i>Mafa-B*06</i>	○						
	<i>Mafa-B*23</i>	○						
	<i>Mafa-I*02</i>	○						

○ :Alleles were detected in each individual

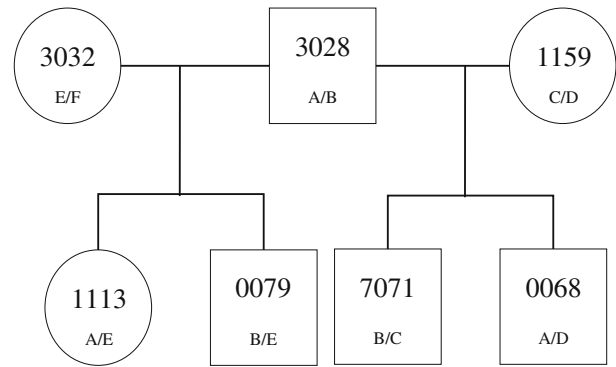


Fig. 3 The family pedigree demonstrating the inheritance of alleles of the MHC class I *B* and *I* loci of cynomolgus monkeys. Male and female are denoted by squares and circles, respectively. The animal number assigned to the animal is shown. The haplotypes of each animal are given by capital letters

structured the phylogenetic tree using the amino acid sequences of the exons 5 to 8. The result clearly showed that these nine alleles clustered with *Mamu-I* alleles (Fig. 2b). These results strongly suggested that these cDNA clones were derived from distinct alleles on MHC class I *B* and *I* loci of cynomolgus monkeys.

Inheritance of *Mafa-B* and *Mafa-I* in a family of cynomolgus monkeys

A family consisting of three parents (one sire and two dams) and four offspring was subjected to genetic analysis to study inheritance of *Mafa-B* and *Mafa-I* alleles. By nucleotide sequence analysis, ten *Mafa-B* alleles and four *Mafa-I* alleles were detected in this family as shown in Table 2. Since certain alleles appeared to be inherited in this family as a complex, we considered those gene complexes as haplotypes and assigned letters A through F to those combinations of alleles (Table 2). Haplotype A (*Mafa-B\*03* and *Mafa-I\*09*)

Table 3 The presence of multiple *Mafa-B* alleles in cynomolgus monkeys

Animal no.	Alleles	Number of copies	Primers
2010	<i>Mafa-B*09</i>	5	5'MBS/3'MBS, 5'Beta 3 XHO/Mafa-Bla
	<i>Mafa-B*11</i>	16	5'MBS/3'MBS, 5'Beta 3 XHO/Mafa-Bla
	<i>Mafa-B*12</i>	7	5'MBS/3'MBS, 5'Beta 3 XHO/Mafa-Bla
	<i>Mafa-B*19</i>	2	5'MBS/3'MBS, 5'Beta 3 XHO/Mafa-Bla
5076	<i>Mafa-B*10</i>	4	5'MBS/3'MBS
	<i>Mafa-B*14</i>	17	5'MBS/3'MBS, 5'Beta 3 XHO/Mafa-Bla
	<i>Mafa-B*15</i>	10	5'MBS/3'MBS
	<i>Mafa-I*010103</i>	13	5'MBS/3'MBS, 5'Beta 3 XHO/Mafa-Bla

was detected in 3028, 1113, and 0068, whereas haplotype B (*Mafa-B\*24* and *Mafa-B\*25*) was carried by 3028, 0079, and 7071 (Fig. 3). Haplotype C (*Mafa-B\*17*, *Mafa-B\*20*, and *Mafa-I\*06*) was found in 1159 and 7071, haplotype D (*Mafa-B\*01* and *Mafa-B\*04*) in 1159 and 0068, haplotype E (*Mafa-B\*16*, *Mafa-I\*03*) in 3032, 0079, and 1113, and haplotype F (*Mafa-B\*06*, *Mafa-B\*23*, and *Mafa-I\*02*) in 3032 (Fig. 3). We could not detect *Mafa-I* alleles in monkeys bearing haplotypes B and D. It was evident that *Mafa-B* alleles were inherited in a Mendelian fashion. Moreover, cynomolgus monkeys in this family were shown to have two to four *Mafa-B* alleles. The presence of multiple *Mafa-B* alleles was confirmed by nucleotide sequences analysis of two additional cynomolgus monkeys unrelated to this family. Table 3 showed that 2010 had four *Mafa-B* alleles and 5076 had three *Mafa-B* alleles. These results indicated that MHC class I *B* locus of cynomolgus monkeys was duplicated as in the case of rhesus monkeys (Boyson et al. 1996b).

## Discussion

Although cynomolgus monkeys are widely used as animal models in a variety of biomedical researches, there are no nucleotide sequence data on cynomolgus MHC class I *B* locus. In this study, we tried to identify the alleles of cynomolgus MHC class I *B* locus, using PBMC cDNA from 16 cynomolgus monkeys.

Nucleotide sequence analyses and following phylogenetic analysis identified 26 *Mafa-B* alleles (Figs. 1, 2a). We also found nine clones with the nucleotide sequences showing high homology with those of *Mamu-I* alleles. Phylogenetic analysis showed that these clones were derived from nine *Mafa-I* alleles. It was reported that novel MHC class I *I* locus in rhesus monkeys, *Mamu-I*, could be amplified with *B* locus-specific primers, and that the *I* locus was recently evolved from a classical MHC class I *B* locus by duplication (Urvater et al. 2000b).

The haplotypes of rhesus MHC class I composed of at least one *A* locus and at least two *B* loci (Boyson et al. 1996b). In cynomolgus monkeys, we previously reported that the *A* locus had been duplicated, because one to four *Mafa-A* alleles were found in an animal (Uda et al. 2004). The presence of up to six *Mamu-B* alleles in a rhesus monkey (Urvater et al. 2000a) indicates that rhesus monkeys have three class I *B* loci. In this study, we also showed that two to four *Mafa-B* alleles were present in each individual, strongly suggesting that cynomolgus monkeys have multiple MHC class I *B* loci. Regarding the *I* locus, it seemed possible that at least one locus was present in each animal, although some individual appeared not to have the locus. The apparent lack of the *I* locus in some individual was probably due to low efficiency of amplification of the *I* locus because of the presence of the multiple *B* loci.

Information on MHC class I molecule is particularly important in better understanding of pathogenesis of various infectious diseases including HIV infection. So far the nucleotide sequence data are available for the alleles of

*Mafa-A* (Uda et al. 2004), *Mafa-E* (Alvarez et al. 1997; Boyson et al. 1995), *Mafa-G* (Arnaiz-Villena et al. 1997; Castro et al. 1996), *Mafa-I* (Urvater et al. 2000b), *Mafa-DRB* (Gaur et al. 1997; Kriener et al. 2000; Leuchte et al. 2004), *Mafa-DQA* (Kenter et al. 1992), and *Mafa-DQB* (Otting et al. 2002) in cynomolgus monkeys. The identification of *Mafa-B* alleles would, therefore, greatly help understand the pathogenesis of various pathogens that naturally or experimentally infect cynomolgus monkeys.

**Acknowledgements** We thank K. Ono, A. Hiyaoka, M. Maeshima, and other staff in The Corporation of Production and Breeding of Primate for animal care and blood collection. This study was supported by the Program of Fundamental Studies in Health Science of the Organization for Pharmaceutical Safety and Research of Japan, and by the Health Science Research Grants from the Ministry of Health and Welfare of Japan.

## References

- Alvarez M, Martinez-Laso J, Varela P, Diaz-Campos N, Gomez-Casado E, Vargas-Alarcon G, Garcia-Torre C, Arnaiz-Villena A (1997) High polymorphism of *Mhc-E* locus in non-human primates: alleles with identical exon 2 and 3 are found in two different species. *Tissue Antigens* 49:160–167
- Arnaiz-Villena A, Martinez-Laso J, Alvarez M, Castro MJ, Varela P, Gomez-Casado E, Suarez B, Recio MJ, Vargas-Alarcon G, Morales P (1997) Primate *Mhc-E* and *-G* alleles. *Immunogenetics* 46:251–266
- 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, Iwanaga KK, Golos TG, Watkins DI (1996a) Identification of the rhesus monkey HLA-G ortholog. *Mamu-G* is a pseudogene. *J Immunol* 157:5428–5437
- Boyson JE, Shufflebotham C, Cadavid LF, Urvater JA, Knapp LA, Hughes AL, Watkins DI (1996b) The MHC class I genes of the rhesus monkey. Different evolutionary histories of MHC class I and II genes in primates. *J Immunol* 156:4656–4665
- Castro MJ, Morales P, Fernandez-Soria V, Suarez B, Recio MJ, Alvarez M, Martin-Villa M, Arnaiz-Villena A (1996) Allelic diversity at the primate *Mhc-G* locus: exon 3 bears stop codons in all Cercopitheciinae sequences. *Immunogenetics* 43:327–336
- Evans DT, Jing P, Allen TM, O'Connor DH, Horton H, Venham JE, Piekarczyk M, Dzuris J, Dykhuizen M, Mitchen J, Rudersdorf RA, Pauza CD, Sette A, Bontrop RE, DeMars R, Watkins DI (2000) Definition of five new simian immunodeficiency virus cytotoxic T-lymphocyte epitopes and their restricting major histocompatibility complex class I molecules: evidence for an influence on disease progression. *J Virol* 74:7400–7410
- Gaur LK, Nepom GT, Snyder KE, Anderson J, Pandarpurkar M, Yadock W, Heise ER (1997) *MHC-DRB* allelic sequences incorporate distinct intragenic trans-specific segments. *Tissue Antigens* 49:342–355
- Kenter M, Otting N, Anholts J, Leunissen J, Jonker M, Bontrop RE (1992) Evolutionary relationships among the primate *Mhc-DQA1* and *DQA2* alleles. *Immunogenetics* 36:71–78
- Kriener K, O'huigin C, Tichy H, Klein J (2000) Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics* 51:169–178
- 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* 56:142–154



- Leuchte N, Berry N, Kohler B, Almond N, LeGrand R, Thorstensson R, Titti F, Sauermann U (2004) *MhcDRB*-sequences from cynomolgus macaques (*Macaca fascicularis*) of different origin. *Tissue Antigens* 63:529–537
- Otting N, Bontrop RE (1993) Characterization of the rhesus macaque (*Macaca mulatta*) equivalent of *HLA-F*. *Immunogenetics* 38:141–145
- Otting N, de Groot NG, Doxiadis GG, Bontrop RE (2002) Extensive *Mhc-DQB* variation in humans and non-human primate species. *Immunogenetics* 54: 230–239
- Prilliman K, Lawlor D, Ellexson M, McElwee N, Confer D, Cooper DK, Kennedy RC, Hildebrand W (1996) Characterization of baboon class I major histocompatibility molecules. Implications for baboon-to-human xenotransplantation. *Transplantation* 61:989–996
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- Sidebottom DA, Kennedy R, Hildebrand WH (2001) Class I MHC expression in the yellow baboon. *J Immunol* 166:3983–3993
- Uda A, Tanabayashi K, Yamada YK, Akari H, Lee YJ, Mukai R, Terao K, Yamada A (2004) Detection of 14 alleles derived from the MHC class I *A* locus in cynomolgus monkeys. *Immunogenetics* 56:155–163
- Urvater JA, McAdam SN, Loehrke JH, Allen TM, Moran JL, Rowell TJ, Rojo S, Lopez de Castro JA, Taurog JD, Watkins DI (2000a) A high incidence of *Shigella*-induced arthritis in a primate species: major histocompatibility complex class I molecules associated with resistance and susceptibility, and their relationship to *HLA-B27*. *Immunogenetics* 51:314–325
- Urvater JA, Otting N, Loehrke JH, Rudersdorf R, Slukvin II, Piekarczyk MS, Golos TG, Hughes AL, Bontrop RE, Watkins DI (2000b) Mamu-I: a novel primate MHC class I B-related locus with unusually low variability. *J Immunol* 164:1386–1398