MHC genotyping with massively parallel pyrosequencing

Roger W. Wiseman, Julie A. Karl, Benjamin N. Bimber, Claire E. O'Leary, Simon M. Lank, Jennifer J. Tuscher, Ann M. Detmer, Pascal Bouffard, Natalya Levenkova, Cynthia L. Turcotte, Edward Szekeres Jr., Chris Wright, Timothy Harkins & David H. O'Connor

Supplementary Item & Number	Title
Supplementary Figure 1	Novel MHC class I sequences detected in cynomolgus, rhesus, and pig-tailed macaques.
Supplementary Figure 2	MHC class I genotypes of cynomolgus macaques.
Supplementary Figure 3	MHC class I genotypes of rhesus macaques.
Supplementary Figure 4	MHC class I genotypes of pig-tailed macaques.
Supplementary Figure 5	Detailed analysis of sequence artifacts.
Supplementary Figure 6	Additional shared MHC class I transcript abundance profiles.
Supplementary Figure 7	Inferred *Mamu-A* and *Mamu-B* class I haplotypes.
Supplementary Note	

Supplementary Figure 1.

Cynomolgus macaques

Working Allele Name	GenBank Accession Number	Representative Animal(s)
*Mf-A*nov001*	GQ153320	AG102, AG108
*Mf-B*nov001*	GQ153321	CY0157, CR355
*Mf-B*nov002*	GQ153322	CY0157, CR355
*Mf-B*nov003*	GQ153323	CY0157, CR355
*Mf-B*nov004*	GQ153324	CY0161, AG108
*Mf-B*nov005*	GQ153325	CY0161, AG108
*Mf-B*nov006*	GQ153326	CY0165
*Mf-B*nov007*	GQ153327	CY0165
*Mf-B*nov008*	GQ153328	CY0165
*Mf-B*nov010*	GQ153329	CR354, CR340
*Mf-B*nov011*	GQ153330	CR352, AG102
*Mf-B*nov012*	GQ153331	CR352, AG102
*Mf-B*nov014*	GQ153332	CR352, AG102
*Mf-B*nov015*	GQ153333	CR355, CR340
*Mf-B*nov017*	GQ153334	CR355, CR340
*Mf-B*nov018*	GQ153335	CR354, CR340
*Mf-B*nov019*	GQ153336	CR355, CR340
*Mf-B*nov020*	GQ153337	AG102, AG109
*Mf-B*nov021*	GQ153338	AG102, AG109
*Mf-B*nov022*	GQ153339	AG102, AG109
*Mf-E*nov001*	GQ153340	CY0161, CY0165
*Mf-E*nov002*	GQ153341	CR352, AG102
*Mf-E*nov003*	GQ153342	CR355, CR340
*Mf-E*nov004*	GQ153343	CR352, CR340
*Mf-E*nov005*	GQ153344	CY0157, CR355
*Mf-E*nov006*	GQ153345	CY0161, CY0165
*Mf-E*nov007*	GQ153346	AG102, AG109
*Mf-E*nov008*	GQ153347	CR355, CR340
*Mf-E*nov009*	GQ153348	CR352, CR403
*Mf-E*nov010*	GQ153349	CY0161, CY0165
*Mf-E*nov011*	GQ153350	CR352, CR403

Rhesus macaques

Working Allele Name	GenBank Accession Number	Representative Animal(s)
*Mm-A*nov001*	GQ153351	LP10, CR12
*Mm-A*nov002*	GQ153352	NM15
*Mm-A*nov003*	GQ153353	CR11, NM42
*Mm-A*nov004*	GQ153354	CR45, NM31
*Mm-A*nov006*	GQ153355	LP05, LP25
*Mm-A*nov007*	GQ153356	LP11, LP13
*Mm-A*nov008*	GQ153357	LP11, LP13
*Mm-A*nov009*	GQ153358	LP12, LP15
*Mm-A*nov010*	GQ153359	LP05
*Mm-A*nov011*	GQ153360	LP11
*Mm-A*nov017*	GQ153361	CR062
*Mm-A*nov018*	GQ153362	CR069, NM44
*Mm-A*nov019*	GQ153363	CR12, NM36
*Mm-B*nov001*	GQ153364	LP03, LP15
*Mm-B*nov002*	GQ153365	
*Mm-B*nov003*	GQ153366	CR45, NM06
*Mm-B*nov004*	GQ153367	CR059, NM15
*Mm-B*nov005*	GQ153368	CR05, NM15
*Mm-B*nov006*	GQ153369	LP10, NM15
*Mm-B*nov007*	GQ153370	LP10, NM15
*Mm-B*nov008*	GQ153371	LP03, LP15
*Mm-B*nov009*	GQ153372	CR04, CR055
*Mm-B*nov010*	GQ153373	LP33, NM43
*Mm-B*nov011*	GQ153374	CR45, CR50
*Mm-B*nov012*	GQ153375	LP21, CR48
*Mm-B*nov013*	GQ153376	LP01, PJ02
*Mm-B*nov014*	GQ153377	CR50, NM38
*Mm-B*nov015*	GQ153378	CR51, NM38
*Mm-B*nov016*	GQ153379	CR11, CR12
*Mm-B*nov017*	GQ153380	LP03, LP10
*Mm-B*nov018*	GQ153381	CR51, NM39
*Mm-B*nov019*	GQ153382	LP03, CR45
*Mm-B*nov020*	GQ153383	LP24, CR45
*Mm-B*nov021*	GQ153384	PJ02, CR08
*Mm-B*nov022*	GQ153385	CR08, NM43
*Mm-B*nov023*	GQ153386	LP05, CR10
*Mm-B*nov024*	GQ153387	CR11
*Mm-B*nov025*	GQ153388	CR11
*Mm-B*nov026*	GQ153389	CR48, NM42
*Mm-B*nov027*	GQ153390	CR068, CR48
*Mm-B*nov028*	GQ153391	CR48, NM35
*Mm-B*nov029*	GQ153392	CR48, NM42
*Mm-B*nov030*	GQ153393	CR062, CR51
*Mm-B*nov031*	GQ153394	LP21, LP25
*Mm-B*nov032*	GQ153395	CR069, NM06
*Mm-B*nov033*	GQ153396	CR059, NM15
*Mm-B*nov034*	GQ153397	NM18
*Mm-B*nov035*	GQ153398	PJ05, NM18
*Mm-B*nov036*	GQ153399	CR065, CR07
*Mm-B*nov037*	GQ153400	PJ02, PJ09
*Mm-B*nov038*	GQ153401	LP06, LP34
*Mm-B*nov039*	GQ153402	LP06, LP34
*Mm-B*nov040*	GQ153403	LP12, PJ01
*Mm-B*nov041*	GQ153404	LP21, PJ04
*Mm-B*nov042*	GQ153405	PJ06, PJ11
*Mm-B*nov045*	GQ153408	PJ04
*Mm-B*nov047*	GQ153410	NM32
*Mm-B*nov048*	GQ153411	NM32
*Mm-B*nov050*	GQ153412	NM39
*Mm-B*nov051*	GQ153413	NM41
*Mm-B*nov052*	GQ153414	NM39
*Mm-B*nov053*	GQ153415	NM41
*Mm-B*nov054*	GQ153416	NM42
*Mm-B*nov055*	GQ153417	NM43
*Mm-B*nov056*	GQ153418	CR03
*Mm-B*nov057*	GQ153419	CR055
*Mm-B*nov058*	GQ153420	CR06, CR08
*Mm-B*nov059*	GQ153421	CR068
*Mm-B*nov060*	GQ153422	CR068
*Mm-B*nov061*	GQ153423	NM36, NM44
*Mm-B*nov062*	GQ153424	CR12, NM35
*Mm-B*nov063*	GQ153425	CR060, CR067
*Mm-B*nov064*	GQ153426	LP30, CR062
*Mm-B*nov065*	GQ153427	NM32
*Mm-B*nov066*	GQ153428	NM36, NM44
*Mm-B*nov067*	GQ153429	NM36, NM44
*Mm-B*nov068*	GQ153430	NM36, NM44
*Mm-B*nov069*	GQ153431	NM36, NM44
*Mm-B*nov070*	GQ153432	NM39
*Mm-B*nov071*	GQ153433	NM39
*Mm-B*nov072*	GQ153434	CR06, NM41
*Mm-B*nov073*	GQ153435	CR06, NM41
*Mm-B*nov074*	GQ153436	CR05, NM35
*Mm-B*nov075*	GQ153437	NM39, NM41
*Mm-I*nov001*	GQ153448	CR48
*Mm-I*nov002*	GQ153449	NM31
*Mm-I*nov003*	GQ153450	LP33, NM43
*Mm-I*nov004*	GQ153451	LP05, CR068
*Mm-E*nov001*	GQ153438	CR45, NM35
*Mm-E*nov002*	GQ153439	CR48, NM35
*Mm-E*nov003*	GQ153440	LP10, PJ07
*Mm-E*nov004*	GQ153441	PJ11, CR10
*Mm-E*nov005*	GQ153442	CR08, NM37
*Mm-E*nov006*	GQ153443	CR48, NM42
*Mm-E*nov007*	GQ153444	LP11, LP15
*Mm-E*nov008*	GQ153445	NM31
*Mm-E*nov009*	GQ153446	NM41
*Mm-E*nov010*	GQ153447	NM41

Pig-tailed macaques

Working Allele Name	GenBank Accession Number	Representative Animal(s)
*Mn-A*nov001*	GQ153452	PT020
*Mn-A*nov002*	GQ153453	PT009, PT023
*Mn-A*nov003*	GQ153454	PT020
*Mn-A*nov004*	GQ153455	PT020
*Mn-A*nov005*	GQ153456	PT023
*Mn-A*nov006*	GQ153457	PT025
*Mn-A*nov007*	GQ153458	PT021
*Mn-A*nov008*	GQ153459	PT021
*Mn-A*nov009*	GQ153460	PT022
*Mn-A*nov010*	GQ153461	PT025
*Mn-A*nov011*	GQ153462	PT021, PT027
*Mn-A*nov012*	FJ875218	PT025, PT027
*Mn-B*nov001*	FJ875219	PT009, PT020
*Mn-B*nov002*	FJ875220	PT009, PT020
*Mn-B*nov003*	FJ875221	PT009, PT020
*Mn-B*nov004*	GQ153463	PT009, PT019
*Mn-B*nov005*	FJ875222	PT009
*Mn-B*nov006*	GQ153464	PT009
*Mn-B*nov007*	GQ153465	PT009, PT026
*Mn-B*nov008*	FJ875223	PT009, PT026
*Mn-B*nov009*	GQ153466	PT009, PT026
*Mn-B*nov010*	GQ153467	PT009, PT019
*Mn-B*nov011*	GQ153468	PT009, PT010
*Mn-B*nov012*	GQ153469	PT009, PT026
*Mn-B*nov013*	GQ153470	PT009, PT026
*Mn-B*nov014*	FJ875224	PT020, PT025
*Mn-B*nov015*	GQ153471	PT020, PT025
*Mn-B*nov016*	GQ153472	PT020, PT025
*Mn-B*nov017*	FJ875225	PT010, PT019
*Mn-B*nov018*	GQ153473	PT020, PT027
*Mn-B*nov019*	FJ875226	PT019, PT020
*Mn-B*nov020*	GQ153474	PT020, PT023
*Mn-B*nov021*	GQ153475	PT020, PT026
*Mn-B*nov022*	GQ153476	PT020, PT023
*Mn-B*nov023*	GQ153477	PT020, PT025
*Mn-B*nov024*	FJ875227	PT023
*Mn-B*nov025*	FJ875228	PT023
*Mn-B*nov026*	FJ875229	PT019, PT022
*Mn-B*nov027*	GQ153478	PT023
*Mn-B*nov028*	GQ153479	PT023
*Mn-B*nov029*	GQ153480	PT022, PT023
*Mn-B*nov030*	FJ875230	PT010, PT022
*Mn-B*nov031*	FJ875231	PT021, PT026
*Mn-B*nov033*	GQ153482	PT023
*Mn-B*nov034*	GQ153483	PT010
*Mn-B*nov035*	GQ153484	PT019
*Mn-B*nov036*	FJ875232	PT019
*Mn-B*nov037*	FJ875233	PT019
*Mn-B*nov038*	GQ153485	PT019
*Mn-B*nov039*	GQ153486	PT019
*Mn-B*nov040*	GQ153487	PT021
*Mn-B*nov041*	GQ153488	PT021
*Mn-B*nov042*	GQ153489	PT021
*Mn-B*nov043*	GQ153490	PT021
*Mn-B*nov044*	GQ153491	PT022
*Mn-B*nov045*	GQ153492	PT022
*Mn-B*nov047*	GQ153494	PT024
*Mn-B*nov048*	FJ875234	PT024
*Mn-B*nov049*	GQ153495	PT024
*Mn-B*nov050*	FJ875235	PT026
*Mn-B*nov052*	FJ875237	PT026
*Mn-B*nov053*	FJ875238	PT026
*Mn-B*nov054*	FJ875239	PT026
*Mn-B*nov055*	GQ153496	PT026
*Mn-B*nov056*	FJ875240	PT028
*Mn-B*nov057*	GQ153497	PT028
*Mn-B*nov058*	GQ153498	PT028
*Mn-B*nov059*	FJ875241	PT010, PT022
*Mn-B*nov060*	GQ153499	PT010, PT022
*Mn-B*nov061*	FJ875242	PT010, PT019
*Mn-B*nov063*	FJ875244	PT010
*Mn-B*nov064*	FJ875245	PT010, PT027
*Mn-B*nov065*	GQ153500	PT010
*Mn-B*nov066*	FJ875246	PT019
*Mn-B*nov067*	GQ153501	PT019
*Mn-B*nov068*	GQ153502	PT019
*Mn-B*nov069*	GQ153503	PT021, PT022
*Mn-B*nov070*	GQ153504	PT021, PT022
*Mn-B*nov071*	FJ875247	PT021, PT022
*Mn-B*nov073*	GQ153505	PT021, PT024
*Mn-B*nov074*	GQ153506	PT021, PT024
*Mn-B*nov075*	GQ153507	PT021, PT028
*Mn-B*nov076*	GQ153508	PT021, PT024
*Mn-B*nov077*	GQ153509	PT024
*Mn-B*nov078*	GQ153510	PT024, PT027
*Mn-B*nov079*	GQ153511	PT025, PT027
*Mn-B*nov080*	GQ153512	PT025, PT027
*Mn-B*nov081*	FJ875249	PT025, PT027
*Mn-B*nov083*	FJ875250	PT026
*Mn-B*nov084*	GQ153513	PT028
*Mn-B*nov085*	GQ153514	PT019
*Mn-B*nov086*	GQ153515	PT025
*Mn-B*nov087*	GQ153516	PT025
*Mn-B*nov088*	GQ153517	PT026
*Mn-B*nov089*	GQ153518	PT024, PT026
*Mn-B*nov090*	FJ875251	PT026, PT028
*Mn-I*nov002*	GQ153527	PT021
*Mn-E*nov001*	GQ153519	PT009, PT010
*Mn-E*nov003*	GQ153520	PT023, PT026
*Mn-E*nov004*	GQ153521	PT019
*Mn-E*nov005*	GQ153522	PT022
*Mn-E*nov006*	GQ153523	PT024
*Mn-E*nov007*	GQ153524	PT027
*Mn-E*nov008*	GQ153525	PT010, PT019
*Mn-E*nov009*	GQ153526	PT021, PT028

Supplementary Figure 1. Novel cynomolgus, rhesus and pig-tailed macaque MHC class I sequences.

GeneBank accession numbers as well as representative animals in which we observed these sequences are listed.

Supplementary Figure 2.

Allele/Animal ID	CY0157	CY0161	CY0165	CR354	CR352	CR403	CR355	CR340	AG108	AG102	AG109	AG107	112003	Total # Reads in Contigs
# Total Reads/Animal	11754	7689	7955	7582	10183	7441	10803	6392	10169	9733	10374	11928	21.3	Ave Alleles/Animal
# Alleles / Animal	14	13	13	12	28	28	24	21	23	27	27	25		
Mafa Haplotype 1	M1	M2	M3	M4	M1	M1	M1	M4	M1	M4/M6	M2	M2		
Mafa Haplotype 2	Homo	Homo	Homo	Homo	M6	M6	M7	M7	M2	Fi	Fi	M4/M6		
***Mafa-A* Sequences**													# Obs.	
*Mafa-A1*03101*				23.7				16.5		8.2		11.1	4	
*Mafa-A1*03201*					12.8	12.6							2	
*Mafa-A1*04701*					7.3	7.5							2	
*Mafa-A1*06001/02/03*							20.0	17.4					2	
*Mafa-A1*06301/02*	20.9	23.0	27.7		9.5	8.2	10.1		21.2		7.8	14.1	9	
*Mafa-A2*05g*	0.3	0.5	0.7	1.0	0.2	0.2	0.1	0.4	0.6	0.4	0.2	0.4	12	
*Mafa-A2*2402*										9.4	10.3		2	
*Mafa-A4*0101*	4.9	6.1	8.7		4.7	6.4	4.0		5.6		2.9	3.4	9	
*Mafa-A5*3001*				2.7			1.0	2.4		0.7		0.4	5	
*Mf-A*nov001*										4.1	6.5		2	
***Mafa-B* Sequences**														
*Mafa-B*02702*	12.8				7.5	6.2	8.3		5.5				5	
*Mafa-B*03602*										10.6	10.5		2	
*Mafa-B*0430101*	3.2				1.8	1.2	1.9		1.1				5	
*Mafa-B*0440101*	27.9				12.3	13.9	14.3		14.0				5	
*Mafa-B*04501*			2.4										1	
*Mafa-B*0460101*	7.2				3.3	3.2	3.7		4.0				5	
*Mafa-B*0470101*				18.5				17.6					2	
*Mafa-B*0480101*		12.5							5.7		3.7	5.5	4	
*Mafa-B*0490101*					7.8	8.6				8.3		9.6	4	
*Mafa-B*05101*			27.3										1	
*Mafa-B*05401*	3.7				2.2	1.9	1.3		1.9				5	
*Mafa-B*0550101/02*	1.4				0.6	0.5	0.7		0.4	0.2			7	
*Mafa-B*05504*					0.5	0.4				0.2		0.2	4	
*Mafa-B*0600101*		18.1							6.7		6.7	7.5	4	
*Mafa-B*0620101*				0.2				0.5					2	
*Mafa-B*0630101*		21.4							10.8		11.4	13.8	4	
*Mafa-B*0640101*					3.2	3.1				3.9		4.2	4	
*Mafa-B*0650101*					3.6	3.6				2.1		2.9	4	
*Mafa-B*0660101*		0.3							0.1		0.1	0.3	4	
*Mafa-B*07701/02*					3.9	3.3				6.8	2.6	5.6	5	
*Mafa-B*08101*										7.3	6.4		2	
*Mafa-B*08201*										4.5	7.5		2	
*Mafa-B*08401*										0.8	0.4		2	
*Mafa-B*0860101/02*										3.6	1.7		2	
*Mafa-B*09001*							0.8	0.8					2	
*Mafa-B*09301*		2.4	6.2						0.4		0.9	1.1	5	
*Mafa-B*09401/02*				15.8				3.1					2	
*Mafa-B*09502*							4.5	6.7					2	
*Mf-B*nov001*	2.2				1.5	1.6	1.2		0.2				5	
*Mf-B*nov002*	0.9				0.4	1.0	0.7		0.2				5	
*Mf-B*nov003*	0.5			1.1	0.4	0.4	0.4	1.0					6	
*Mf-B*nov004*		0.7							0.2		0.2	0.3	4	
*Mf-B*nov005*		0.5							0.1			0.04	3	
*Mf-B*nov006*			0.6										1	
*Mf-B*nov007*			0.5										1	
*Mf-B*nov008*			0.3										1	
*Mf-B*nov010*				0.2				0.1					2	
*Mf-B*nov011*					2.1	1.8				1.7		2.2	4	
*Mf-B*nov012*					0.4	0.6				0.4		0.4	4	
*Mf-B*nov014*					0.1	0.2				0.2			3	
*Mf-B*nov015*							10.5	13.3					2	
*Mf-B*nov017*							0.3	0.3					2	
*Mf-B*nov018*				0.5				0.2					2	
*Mf-B*nov019*							0.1	0.2					2	
*Mf-B*nov020*										6.9	5.6		2	
*Mf-B*nov021*										0.9	0.4		2	
*Mf-B*nov022*										0.4	0.2		2	
***Mafa-I* Sequences**														
*Mafa-I*g*			1.5		0.7	1.0	0.5	0.6		2.4	1.7	0.4	8	
*Mafa-I*110101*				9.8				3.7					2	
***Mafa-E* Sequences**														
*Mafa-E*01/02*	13.2			24.2	6.5	5.6	8.1	8.1	10.4	8.4		7.8	9	
*Mf-E*nov001*		13.5	22.7						9.6		6.5	7.4	5	
*Mf-E*nov002*					6.0	5.9				6.2	4.9		4	
*Mf-E*nov003*							6.5	5.8					2	
*Mf-E*nov004*				2.1				0.7		0.6		0.9	4	
*Mf-E*nov005*	0.8				0.3	0.3	0.6		0.7				5	
*Mf-E*nov006*		0.6	1.1						0.2		0.3	0.2	5	
*Mf-E*nov007*										0.6	0.3		2	
*Mf-E*nov008*							0.5	0.4					2	
*Mf-E*nov009*					0.3	0.3							2	
*Mf-E*nov010*		0.3	0.2						0.1		0.1	0.2	5	
*Mf-E*nov011*					0.1	0.3							2	

>5% Total Reads

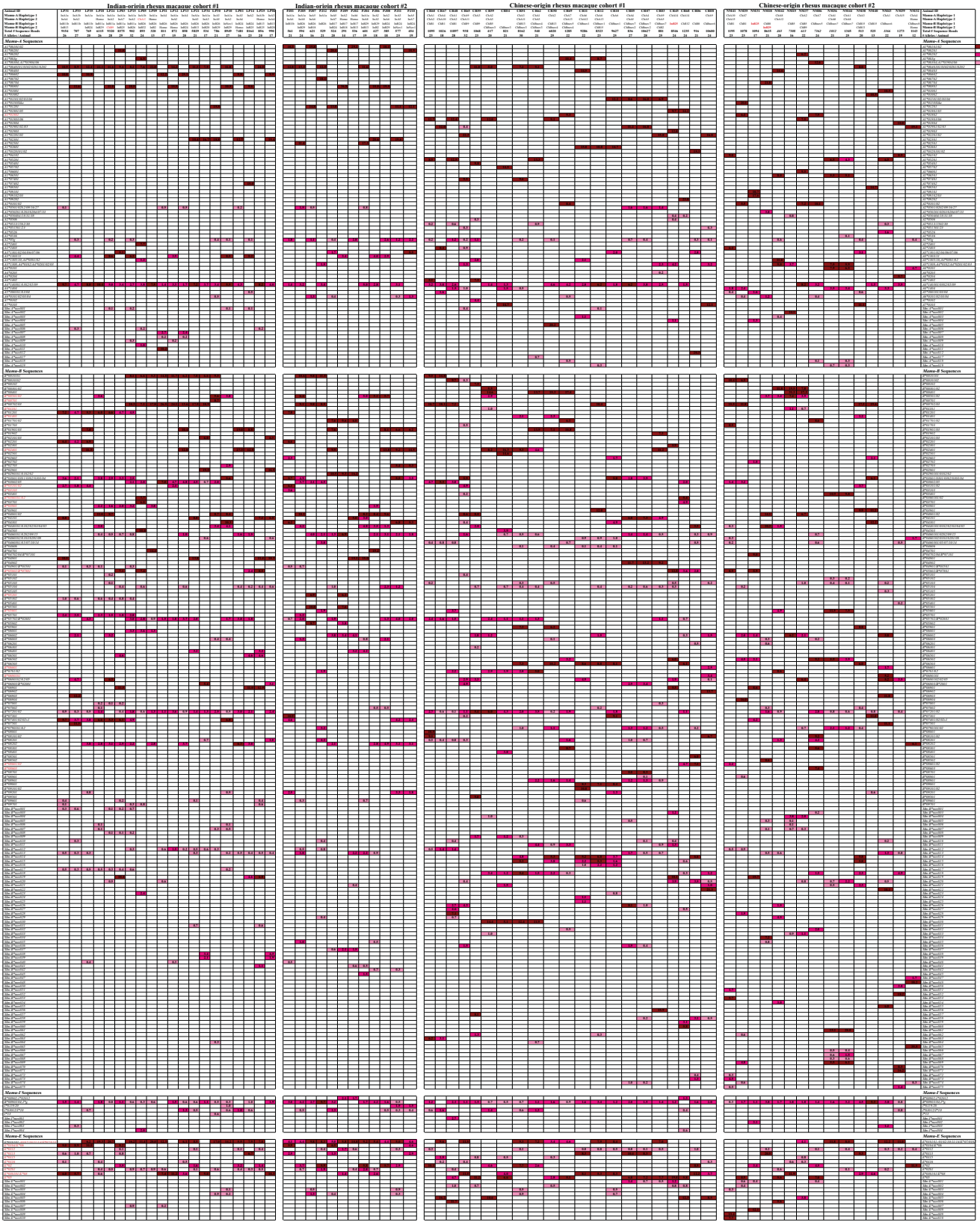
1-5% Total Reads

<1% Total Reads

Supplementary Figure 2. MHC class I genotypes of cynomolgus macaques.

Transcript abundance levels are presented as a percentage of the total MHC class I sequences reads evaluated for each animal.

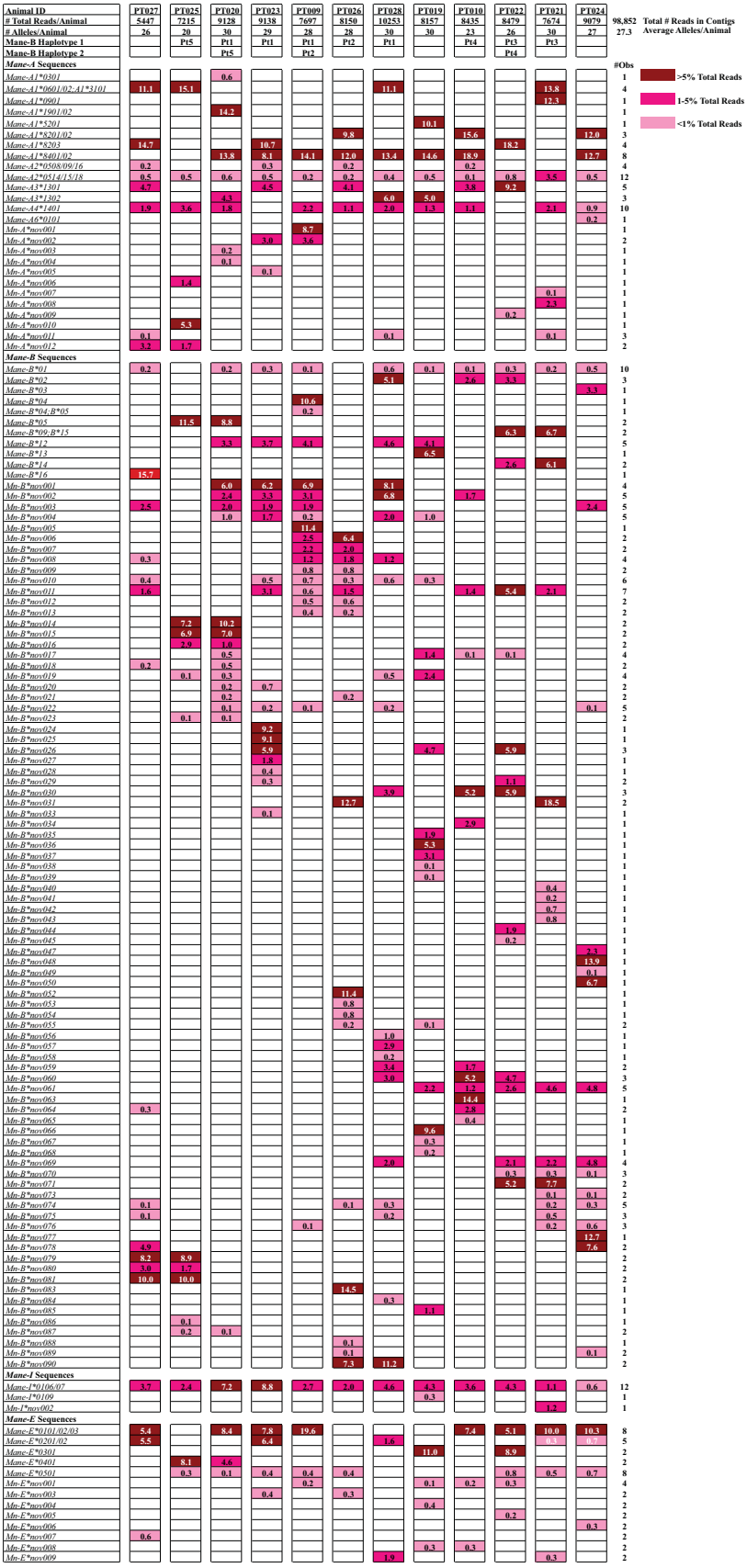
Supplementary Figure 3.



Supplementary Figure 3. MHC class I genotypes of rhesus macaques.

Transcript abundance levels are presented as a percentage of the total MHC class I sequences reads evaluated for each anmals. Class I sequence names with primer mismatches are highlighted in red in the left column.

Supplementary Figure 4.



Supplementary Figure 4. MHC class I genotypes of pig-tailed macaques.

Transcript abundance levels are presented as a percentage of the total MHC class I sequences reads evaluated for each animal.

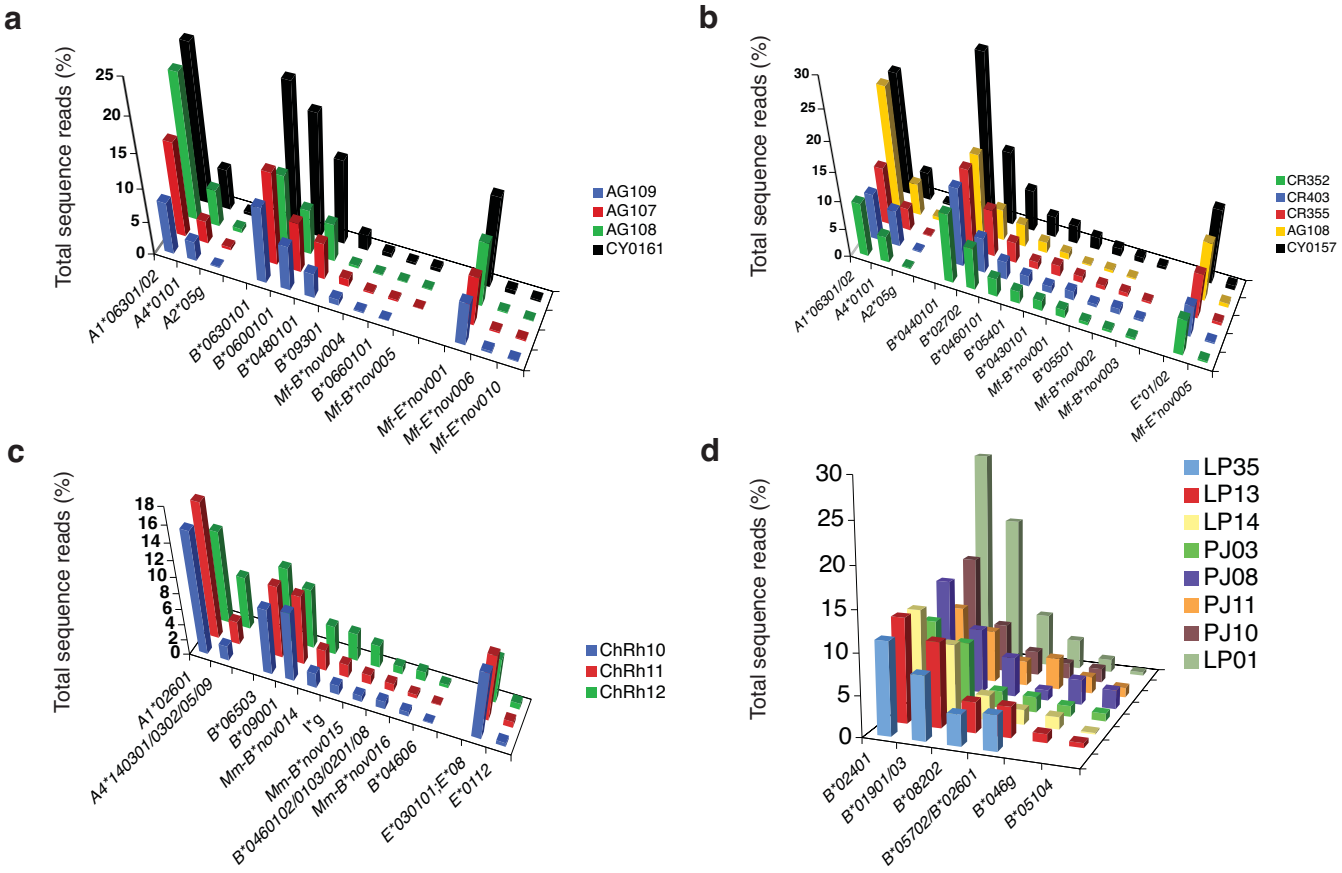
Supplementary Figure 5.

Animal ID	MID Tag	Total Reads	Failed Assembly	Assembled	Known or Novel	Insertions/ Deletions	Single Base Substitutions	Multiple Base Substitutions
	Overall Totals	**484,985**	**73,114**	**411,871**	**406,274**	**2,914**	**565**	**2,118**
	Pilot Study Totals	**444,142**	**68,173**	**375,969**	**371,104**	**2,703**	**278**	**1,884**
CY0157	MID1	12,564	591	11,973	11,740	202	7	24
CY0161	MID2	8,839	979	7,860	7,682	97	14	67
CY0165	MID3	10,307	2,070	8,237	7,961	136	5	135
CR354	MID4	11,898	3,981	7,917	7,570	260	5	82
CR403	MID5	8,265	601	7,664	7,446	62	5	151
CR352	MID6	11,952	1,617	10,335	10,193	92	0	50
CR355	MID7	12,249	1,366	10,883	10,779	91	6	7
CR340	MID8	9,041	2,518	6,523	6,371	152	0	0
AG108	MID9	14,176	3,958	10,218	10,130	57	10	21
AG102	MID10	11,697	1,368	10,329	9,691	72	6	560
AG109	MID11	13,345	2,891	10,454	10,337	106	5	6
AG107	MID12	13,326	1,231	12,095	11,903	153	0	39
LP01	MID1	8,481	892	7,589	7,486	96	7	0
LP03	MID2	9,300	828	8,472	8,403	54	0	15
LP10	MID5	10,381	1,434	8,947	8,871	27	41	8
LP14	MID6	8,729	496	8,233	8,164	58	5	6
LP18	MID7	7,993	1,849	6,144	6,115	29	0	0
LP21	MID8	7,385	1,410	5,975	5,922	27	26	0
LP24	MID9	10,431	1,096	9,335	9,325	10	0	0
LP31	MID10	10,553	1,340	9,213	9,174	31	0	8
ChRh07	MID1	12,485	1,858	10,627	10,463	124	0	40
ChRh08	MID2	10,622	1,690	8,932	8,856	69	0	7
ChRh10	MID3	10,160	875	9,285	9,237	18	0	30
ChRh11	MID4	10,631	1,794	8,837	8,764	29	0	44
ChRh12	MID5	8,468	689	7,779	7,773	6	0	0
ChRh45	MID6	11,177	3,370	7,807	7,743	44	8	12
ChRh48	MID7	12,169	2,106	10,063	9,997	48	6	12
ChRh50	MID8	5,812	731	5,081	5,044	31	0	6
ChRh51	MID9	8,147	1,149	6,998	6,954	29	6	9
NM06	MID10	8,706	1,308	7,398	7,362	36	0	0
NM15	MID11	9,785	2,203	7,582	7,508	74	0	0
NM18	MID12	10,288	2,004	8,284	8,227	49	0	8
PT020	MID1	9,944	775	9,169	9,122	6	14	27
PT021	MID2	8,565	836	7,729	7,599	29	6	95
PT022	MID3	9,777	1,245	8,532	8,443	31	5	53
PT023	MID4	10,165	972	9,193	9,088	45	9	51
PT024	MID5	10,176	1,060	9,116	9,012	52	17	35
PT025	MID6	8,893	1,663	7,230	7,201	14	0	15
PT026	MID7	8,792	562	8,230	8,115	30	5	80
PT027	MID8	6,101	624	5,477	5,407	33	7	30
PT028	MID9	11,708	1,720	9,988	9,870	25	5	88
PT009	MID10	10,374	2,639	7,735	7,662	35	0	38
PT010	MID11	10,026	1,732	8,294	8,254	17	23	0
PT019	MID12	10,259	2,052	8,207	8,140	17	25	25
	Follow-up Study Totals	**40,843**	**4,941**	**35,902**	**35,170**	**211**	**287**	**234**
ChRh03	MID1	889	72	817	802	9	4	2
ChRh04	MID2	865	71	794	784	4	4	2
ChRh05	MID3	1,165	386	779	775	2	0	2
ChRh06	MID4	909	141	768	755	7	4	2
ChRh055	MID5	1,064	85	979	965	6	8	0
ChRh059	MID6	450	91	359	354	3	2	0
ChRh060	MID7	1,132	96	1,036	1,030	0	4	2
ChRh062	MID8	583	84	499	488	9	2	0
ChRh065	MID9	1,028	98	930	907	13	6	4
ChRh067	MID10	1,263	319	944	925	7	8	4
ChRh068	MID11	1,038	80	958	939	2	5	12
ChRh069	MID12	1,400	134	1,266	1,256	4	4	2
NM31	MID1	969	108	861	820	11	10	20
NM32	MID2	905	65	840	824	2	8	6
NM35	MID3	1,098	125	973	918	15	17	23
NM36	MID4	1,118	94	1,024	1,012	10	0	2
NM37	MID5	685	60	625	609	2	8	6
NM38	MID6	582	98	484	471	9	4	0
NM39	MID7	1,031	87	944	931	2	6	5
NM40	MID8	586	57	529	525	0	2	2
NM41	MID9	1,185	138	1,047	990	6	19	32
NM42	MID10	759	274	485	465	0	13	7
NM43	MID11	1,341	138	1,203	1,176	12	6	9
NM44	MID12	1,231	112	1,119	1,080	6	28	5
LP33	MID1	813	89	724	707	9	4	4
LP35	MID2	900	109	791	769	6	12	4
LP06	MID3	1,159	157	1,002	990	4	4	4
LP15	MID4	1,071	143	928	898	2	16	12
LP05	MID5	1,075	170	905	886	3	14	2
LP09	MID6	413	93	320	320	0	0	0
LP30	MID7	903	88	815	787	0	7	21
LP34	MID8	604	87	517	513	0	4	0
LP11	MID9	910	87	823	813	6	2	2
LP12	MID10	1,125	231	894	872	7	7	8
LP13	MID11	992	123	869	858	3	6	2
LP25	MID12	975	98	877	856	11	6	4
PJ01	MID1	589	20	569	563	0	0	6
PJ02	MID2	576	30	546	538	0	8	0
PJ03	MID3	571	39	532	524	4	2	2
PJ04	MID4	646	39	607	601	2	4	0
PJ05	MID5	627	28	599	594	2	3	0
PJ06	MID6	467	36	431	427	0	4	0
PJ07	MID7	658	26	632	621	7	2	2
PJ08	MID8	423	34	389	385	0	0	4
PJ09	MID9	331	36	295	291	0	2	2
PJ10	MID10	547	91	456	452	2	2	0
PJ11	MID11	619	34	585	575	0	4	6
PJ12	MID12	573	40	533	529	2	2	0

Supplementary Figure 5. Detailed analysis of sequence artifacts.

The total number of high quality reads obtained for each animal are given along with the number of reads that assembled (contigs of 5 or more for the pilot study and 2 or more for the follow-up study) vs. the number that failed to assemble. Sequence reads that passed our assembly filter were evaluated further by BLASTN analysis and broken down into known or novel MHC class I sequences and various types of sequence artifacts.

Supplementary Figure 6.



Supplementary Figure 6. Additional shared MHC class I transcript abundance profiles.

a. Transcript profiles for four Mauritian cynomolgus macaques that share the M2 haplotype[7]. Transcript levels are approximately twice as great for CY0161 since this animal is homozygous for the M2 haplotype while the remaining animals are heterozygous (see **Fig 3a** for the alternate haplotypes of these animals). Mf-B*nov005 was detected as 42, 15 and five sequences reads in CY0161, AG108 and AG107, respectively but fell below our detection limit of five sequence reads for AG109. **b.** Transcript profiles for four Mauritian cynomolgus macaques that share the most frequent M1 haplotype[7]. Transcript levels are approximately twice as great for CY0157 since this animal is homozygous for the M1 haplotype while the remaining animals are M1 heterozygotes (**Supplementary Fig. 2** online). This figure only includes class I sequences associated with the M1 haplotype. **c.** Transcript profiles for three Chinese rhesus macaques that share the novel ChBnov1 haplotype (**Supplementary Fig. 7** online). Once again, only sequences from the shared haplotype are illustrated for these heterozygous animals. **d.** Transcript profiles for eight Indian rhesus macaques that share the common InB24 haplotype[5, 6] (**Supplementary Fig. 7** online). LP01 is homozygous for this Mamu-B haplotype and displays transcript levels approximately twice those of the remaining heterozygous animals. For simplicity, only Mamu-B sequences associated with the InB24 haplotype are presented.

Supplementary Figure 7.

Haplotype	# Obs	Mamu-A sequences
Indian rhesus chromosomes (n = 64)		
InA1a	17	*A1*0040101/0102/0201/0202, A4*140301/0302/05/09*
InA1b	4	*A1*0040101/0102/0201/0202, A4*140301/0302/05/09, Mm-A*nov006*
InA2	9	*A1*00801, A3*1303/11, {A2*05g}, {A4*140301/0302/05/09}*
InA3	7	*A1*02301, A4*140301/0302/05/09*
InA4	6	*A1*00602, A4*140301/0302/05/09, {A2*0501/0202/09/16/27}*
InA5	5	*A1*01201, A2*05g, A6*0101/02/03/04*
InA6	5	*A1*00101/02, A2*05g*
InA7	3	*A1*00201, A3*1301/02/04/06/07/08*
InA8	2	*A1*02501, A2*0501/0202/09/16/27*
Chinese rhesus chromosomes (n = 72)		
ChA1	6	*A1*05201, {A2*0513/1502/30}*
ChA2	5	*A1*01803/06, A4*140301/0302/05/09, {A2*05g}*
ChA3	5	*A1*0040101/0102/0201/0202, {A2*05g}*
ChA4	5	*A1*01901/02/05, A4*140301/0302/05/09, {A2*05g}*
ChA5	3	*A1*01101/02/03/04, A2*0501/0202/09/16/27, A4*140301/0302/05/09*
ChA6	3	*A1*02601, A4*140301/0302/05/09*
ChA7	3	*A1*01802, A2*0101/02, A6*0101/02/03/04, {Mm-A*nov018}*
ChA8	3	*A1*06501, A3*1308/A4*0102/A4*0201/02/03, {A2*05g}*
ChA9	3	*A1*02201/02, A3*1308/A4*0102/A4*0201/02/03, {A2*051501/23}*
ChA10	2	*A2*2403, {A5*300101/03/04}*
ChA11	2	*A1*07401, A2*05g, A4*1404*
ChA12	2	*A1*00403, A3*1305/10/A4*0301/02, Mm-A*nov003*
ChA13	2	*A1*003g, A4*140301/0302/05/09*
ChA14	2	*A1*01801/05, A2*050404/18/31/38*
ChA15	2	*A1*04102, A4*140301/0302/04/05/09*

Haplotype	# Obs	Mamu-B sequences
Indian rhesus chromosomes (n = 64)		
InB26	14	*B*0010101, B*00702/03, {B*03002/05}, {B*02601/B*05702}, {B*07201/02}, {Mm-B*nov012}*
InB24	9	*B*01901/03, B*02401, B*05702/B*02601, B*08202, {B*0460101/0202/09/15}, {B*05104}*
InB29	8	*B*04001/02, B*0440101/0102/02/03/04/05, B*00501/02, {B*06003}*
InB13	6	*B*04101, B*04801, Mm-B*nov013, {B*06401}, {B*09601}*
InB25	5	*B*06901, B*06501/02, B*05002/B*07801, Mm-B*nov019, {B*05104}*
InB11a	4	*B*01201, B*03801, B*07401/02/02v1, B*08202, B*05701, B*03001/0301/0302/0303/04, B*0460101/0202/09/15, B*05302, Mm-B*nov017, {B*07001}, {Mm-B*nov001}, {Mm-B*nov008}*
InB11b	4	*B*01201, B*02201, B*07401/02/02v1, B*03101/02, B*05701/02/B*02601, {B*03001/0301/0302/0303/04}, {B*04901/B*03501}, {Mm-B*nov013}, {Mm-B*nov017}*
InB17	3	*B*01701, B*0290102, B*06002, Mm-B*nov037*
InB2	2	*B*03801, B*04701, B*08202*
InB18	2	*B*05501, B*05802, B*05201, B*0460101/0202/09/15*
InB20	2	*B*02101/03, B*02801, B*02001/B*06803, B*0460102/0103/0201/08, B*07201/02, B*08201, Mm-B*nov038, Mm-B*nov039*
InB32	1	*B*03601, B*03701, B*04503*
InBnov1	3	*B*02703, B*03001/0301/0302/0303/04, B*04301, B*07401/02/02v1, B*09201*
Chinese rhesus chromosomes (n = 72)		
ChB1	5	*B*0010101/02, B*00702/03, B*03002/05, Mm-B*nov012, {B*0460301/05/07/10/14}*
ChB5	4	*B*00301/01, B*00401, B*00501/02, B*01101, B*06002, Mm-B*nov004, Mm-B*nov033*
ChB9	3	*B*04001/02, B*0680102/02/05, {B*06002}*
ChB13	3	*B*01401, B*05901, B*06503, B*0790101/02, Mm-B*nov014, Mm-B*nov015, Mm-B*nov018, {B*05103}, {B*04606}*
ChB15	3	*B*00702/03, B*03901*
ChB8	2	*B*0440101/0102/02/03/04/05, B*085001/02, B*06201*
ChB2	1	*B*01902, B*02401, B*05702/B*02601, {B*0460101/0202/09/15}, {B*05104}*
ChBnov1	3	*B*06503, B*09001, Mm-B*nov014, Mm-B*nov015, Mm-B*nov016, B*0460102/0103/0201/08, B*04606*
ChBnov2	2	*B*03002/05, B*06803/B*02001, B*08201, B*08701, Mm-B*nov026, Mm-B*nov036, Mm-B*nov074*
ChBnov3	4	*B*02401, B*05702/B*02601, B*6701/02, B*07201/02, Mm-B*nov018, Mm-B*nov030, {B*0460101/0202/09/15}, {B*05104}*
ChBnov4	3	*B*00401, B*01902, B*08901, B*07201/02, Mm-B*nov011*
ChBnov5	2	*B*01701/02, B*08301, B*06502, B*08201, B*0460301/05/07/10/14, B*07201/02, Mm-B*nov018, Mm-B*nov032*
ChBnov6	2	*Mm-B*nov022, B*06902/03, B*0680101, B*06601, B*06002, B*05103, Mm-B*nov020*
ChBnov7	3	*B*04101, B*04802, B*0460101/0202/09/15, B*05104, B*0790101/02, B*08901, Mm-B*nov013*
ChBnov8	2	*Mm-B*nov065, Mm-B*nov048, B*08202, B*0460102/0103/0201/08, Mm-B*nov047, Mm-B*nov074*

Supplementary Figure 7. Inferred Mamu-A and Mamu-B class I haplotypes.

Mamu-A haplotypes are inferred based on allele sharing between multiple unrelated individuals. Mamu-B haplotypes are based on the framework sequences reported by Otting and coworkers[5,6]. Brackets { } indicate class I sequences that were only detected in at least two thirds of the observed haplotypes in our cohort.

Supplementary Note

Macaque samples.

We obtained fresh, whole blood samples from eight feral Mauritian cynomolgus macaques housed at the Wisconsin National Primate Research Center (Madison, WI) or Charles River BRF, Inc. (Houston, TX). Additionally, we obtained four blood samples from a cynomolgus macaque breeding group from Alpha Genesis, Inc. (Yeemasse, SC). Thirty-two whole blood or frozen peripheral blood mononuclear cell samples from Indian-origin rhesus macaques were provided by investigators at the Oregon National Primate Research Center and the New England Primate Research Center, respectively. Likewise, 36 Chinese-origin rhesus macaque samples were provided by NIAID (Bethesda, MD) and Battelle (Columbus, OH). Finally, we acquired 12 pig-tailed macaque frozen peripheral blood mononuclear cell samples from the University of Pennsylvania (Philadelphia, PA) and Johns Hopkins University (Baltimore, MD). All animals were cared for according to the regulations and guidelines of the Institutional Care and Use Committees at their respective institutions.

Primary cDNA-PCR and pooling strategy.

We isolated total cellular RNAs with a MagNAPure LC (Roche Applied Sciences) using a protocol that included a DNAse treatment to remove contaminating genomic DNA. We converted RNA samples to cDNA using a Superscript™III First-Strand Synthesis System (Invitrogen). We generated primary cDNA-PCR amplicons spanning 190 bp of exon two of macaque class I sequences with high-fidelity Phusion™ polymerase (New England Biolabs). The PCR primers contained one of 12 distinct 10 bp MID tags along with adaptor sequences necessary for emulsion PCR. We used the following primer sequences; each of the 10 bp Multiplex Identifier (MID) tags has been highlighted in bold:

Sense-strand PCR primers

	GS-FLX-A adaptor	MIDx Tag	Class I-specific
GSA-MID1-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**ACGAGTGCGT**	GCTACGTGGACGACACG-3′
GSA-MID2-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**ACGCTCGACA**	GCTACGTGGACGACACG-3′
GSA-MID3-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**AGACGCACTC**	GCTACGTGGACGACACG-3′
GSA-MID4-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**AGCACTGTAG**	GCTACGTGGACGACACG-3′
GSA-MID5-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**ATCAGACACG**	GCTACGTGGACGACACG-3′
GSA-MID6-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**ATATCGCGAG**	GCTACGTGGACGACACG-3′
GSA-MID7-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**CGTGTCTCTA**	GCTACGTGGACGACACG-3′
GSA-MID8-SBT190F	5′-GCCTCCCTCGCGCCATCAG	**CTCGCGTGTC**	GCTACGTGGACGACACG-3′

GSA-MID9-SBT190F 5′-GCCTCCCTCGCGCCATCAG**TAGTATCAGA**GCTACGTGGACGACACG-3′
GSA-MID10-SBT190F 5′-GCCTCCCTCGCGCCATCAG**TCTCTATGCG**GCTACGTGGACGACACG-3′
GSA-MID11-SBT190F 5′-GCCTCCCTCGCGCCATCAG**TGATACGTCT**GCTACGTGGACGACACG-3′
GSA-MID12-SBT190F 5′-GCCTCCCTCGCGCCATCAG**TACTGAGCTA**GCTACGTGGACGACACG-3′

Antisense-strand PCR primers

	GS-FLX-B adaptor	MIDx Tag	Class I-specific
GSB-MID1-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**ACGAGTGCGT**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID2-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**ACGCTCGACA**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID3-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**AGACGCACTC**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID4-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**AGCACTGTAG**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID5-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**ATCAGACACG**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID6-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**ATATCGCGAG**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID7-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**CGTGTCTCTA**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID8-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**CTCGCGTGTC**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID9-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**TAGTATCAGA**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID10-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**TCTCTATGCG**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID11-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**TGATACGTCT**	TCGCTCTGGTTGTAGTAGC-3′
GSB-MID12-SBT190R	5′-GCCTTGCCAGCCCGCTCAG	**TACTGAGCTA**	TCGCTCTGGTTGTAGTAGC-3′

We used the following PCR program with a MJ Research Tetrad Thermocycler (Bio-Rad Laboratories) to generate the primary cDNA-PCR amplicons: initial denaturation at 98 °C for 3 min; amplification over 23 cycles of 98 °C for 5 s, 57 °C for 1 s, 72 °C for 20 s, and a final extension of 72 °C for 5 min. Aliquots of each reaction were monitored by agarose gel electrophoresis and subjected to an additional 3–6 cycles of PCR if necessary to generate sufficient product. After agarose gel electrophoresis and MinElute PCR purification (Qiagen), we quantified primary amplicons with a Qubit Fluorometer (Invitrogen) and normalized them to equimolecular concentrations. We pooled amplicons from groups of 12 animals and confirmed their purity with a 2100 BioAnalyzer (Agilent Technologies). Four additional Indian rhesus samples were lost in our pilot study due to a pipeting error.

Emulsion PCR and pyrosequencing.

We performed the emulsion PCR, bead recovery, and pyrosequencing steps following the manufacturer′s GS FLX protocols[1,2] at the 454 Sequencing Center (Branford, CT) and the University of Illinois at Urbana-Champaign High-Throughput Sequencing Center for the pilot and follow-up studies, respectively. For our pilot study, we physically subdivided a single 70x75 PicoTiterPlate into four regions to allow simultaneous pyrosequencing of pooled amplicons with twelve distinct MID tags from cynomolgus, pig-tailed, Indian-origin and Chinese-origin rhesus macaques. To examine the effect of reducing sequencing depth on MHC class I sequence profiles, we sequenced four pools containing amplicons from 48 rhesus macaques in 1/16 regions

of a 70x75 PicoTiterPlate for our follow-up study. From this instrument run using only 1/4 of a PicoTiterPlate, we acquired 40,879 high quality sequence reads containing a total of 8,713,026 high quality bases. Pyrosequence images were captured by a high resolution CCD camera and processed to generate flowgrams for each well as described previously[1,2]. We acquired 490,243 high quality sequence reads containing a total of 102,177,657 high quality bases from the pilot run. We obtained an average high quality read length of 208.4 bases, allowing detection of the MID tags at both the 5′ and 3′ ends of most amplicons (each 190 bp amplicon contained the same 10 bp MID on each end).

Data analysis.

Improvements in GS FLX pyrosequencing technology have substantially decreased error rates compared to the previous GS 20 platform[3]. In order to insure the accuracy of genotypes and to evaluate the error rate of MHC genotyping by GS FLX pyrosequencing, we employed a multi-step analysis procedure combining both automated and manual analyses to cull sequence artifacts associated with cDNA-PCR and GS FLX pyrosequencing. We initially assembled unfiltered high quality reads into contigs of identical sequences (five or more sequences for the pilot study, two or more sequences for the follow-up study) using SeqMan Pro Version 8.0.2 (DNASTAR). This automated filter step removed most obvious artifacts, allowing for minimal manual downstream editing. We then established a first pass genotype of each animal by aligning all sequence contigs against a custom in-house database of all known macaque MHC class I sequences. The majority of contigs comprised sequence reads that were identical to previously described class I sequences (**Supplementary Fig. 5** online), and we used this complement of known sequences per animal to evaluate the remaining unknown sequence contigs detected in each animal. We designated the remaining unknown contigs as putative novel class I alleles if they were: A) not closely related to other known class I sequences expressed by the animal and either B) detected as >10 reads or C) detected in multiple animals. In a detailed analysis of all high quality sequence reads for the three animals illustrated in **Fig. 2**, we found that 93.4% were consistent with known MHC class I sequences or novel class I sequences confirmed by inter- and intra-animal comparison (**Table 1**). Unknown contigs comprising a limited number of individual sequence reads which differed from known sequences present in the animal by a single nucleotide were excluded from further analysis. Most often these minor sequence contigs contained single base insertions or deletions in

homopolymeric nucleotide runs created by errors in the GS FLX pyrosequencing process. For the animals shown in **Fig. 2**, more than half of these minor sequence contigs were insertions or deletions in short homopolymer tracts that have been well documented as errors associated with the GS FLX pyrosequencing process[1–4]. A subset of these minor contigs was consistent with single nucleotide misincorporations during reverse transcription or the primary cDNA-PCR. One quarter of the putative artifactual reads found in the **Fig. 2** animals were likely due to nucleotide misincorporations during reverse transcription or the primary PCR amplification with Phusion™ High-Fidelity DNA polymerase; these likely sequence artifacts contained single base substitutions relative to another class I sequence expressed by the same animal. A third, minor class of artifacts contained multiple base substitutions relative to high abundance class I sequences expressed by an animal. Given the high degree of sequence polymorphism within the 190 bp region of our amplicon between MHC class I sequences expressed by an individual animal, it is unlikely that such unknown sequences represent novel class I alleles. These sequences most likely resulted from template switching during reverse transcription or the primary cDNA-PCR (resulting in chimera formation), although we cannot fully exclude minor contamination of the RNA samples used for cDNA synthesis with genomic DNA (each macaque MHC haplotype is expected to contain multiple class I pseudogenes). Additional artifacts are those sequences that failed to reach at least 180 bp in length, adding some uncertainty to their identity during the assembly process. 1.1% of the total sequence reads for the three animals shown in **Fig. 2** constituted reads that were <180 bp long. As expected, we identified more putative novel sequences in Chinese-origin rhesus and pig-tailed macaques with less well-characterized MHC class I genetics than in the well-characterized Mauritian-origin cynomolgus and Indian-origin rhesus macaques (**Supplementary Fig. 1** online).

Amplification primer mismatches with MHC class I sequences.

We also manually edited a small subset of class I sequences that we missed by BLASTN analysis as perfect hits due to single base mismatches introduced by the PCR primers (*e.g. Mamu-B*03801* differs from the 5′ primer by a single nucleotide). The MHC class I sequences of rhesus macaques shown below have single base pair mismatches with the amplification primers and may therefore be underrepresented to varying degrees in the transcript abundance data. These

class I sequences are highlighted in red in **Supplementary Fig. 3** online. Interestingly, the proof-reading capacity of Phusion™ High-Fidelity DNA polymerase is able to correct the 3′-terminal mismatched primer nucleotides shown below with high efficiency.

SBT190F class I primer	G	C	T	A	C	G	T	G	G	A	C	G	A	C	A	C	G
*Mamu-A1*00410* (EU403493)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
*Mamu-A1*01909* (EU403495)	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.
*Mamu-A3*1309* (EF112556)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	G	.	.
*Mamu-B*01101* (U41838)	.	.	.	.	.	.	.	.	.	.	.	.	T	.	.	.	.
*Mamu-B*03801* (AJ556889)	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
*Mamu-B*06601* (AJ844597)	.	.	.	.	.	.	.	.	.	T	.	.	.	.	.	.	.
*Mamu-B*0680101* (AJ844599)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
*Mamu-B*08601* (EF580166)	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*08602* (AM902582)	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*08603* (AM902583)	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*09801* (FJ544408)	.	T	.	.	.	.	.	.	T	.	.	.	.	.	.	.	.
*Mamu-E*010102* (EU934770)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0104* (EU305662)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0110* (EU109710)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0111* (EU109711)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0112* (EU305658)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0113* (EU305659)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0114* (EU305660)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0115* (EU305661)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0116* (EU934771)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0117* (FJ544410)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0201* (EU109709)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*02* (U02979)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*030101* (EF611171)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*03* (U02980)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*04* (U02981)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*05* (U41837)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*06* (L41822)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*070102* (EF611170)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*07* (L41823)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*08* (L41824)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C
*Mamu-E*0901* (EF611169)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	C

SBT190R class I primer	T	C	G	C	T	C	T	G	G	T	T	G	T	A	G	T	A	G	C
*Mamu-A1*01802* (AM295899	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-A1*02801*like(AB444925).	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-A1*05106* (EU403529)	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-A2*0532* (EF112544)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A	.
*Mamu-B*00201* (U41833)	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*00501* (U41827)	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*00502* (EU589240)	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*00701* (U41829)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	T	.
*Mamu-B*01401* (AM902540)	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*02102* (AM902536)	.	.	.	.	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*03101* (EF602319)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
*Mamu-B*03102* (EU589242)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
*Mamu-B*03103* (EU934765)	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	A
*Mamu-B*0360101* (AJ556886)	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*0360102* (EU682518)	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*03602* (AJ556887)	C	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*05002* (AJ620415)	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*05201* (EF457755)	.	G	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*05601* (AM902562)	.	.	.	.	.	.	.	.	A	.	.	.	.	.	.	.	.	.	.
*Mamu-B*06101* (AJ556906)	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.
*Mamu-B*06102* (AM902564)	.	T	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.

```
Mamu-B*06102like(AB430442) . T . . . . . . . . . . . . . . . . . .
Mamu-B*07801 (EF611164)    . T . . . . . . . . . . . . . . . . . .
Mamu-B*09301 (AM902585)    . . . . . . . . . . . . . . . . . . T .
Mamu-B*09801 (FJ544408)    . . . . . . G . . . . . . . . . . . . .
Mamu-E*0112 (EU305658)     A . . . . . . . . . . . . . . . . . . .
Mamu-E*0113 (EU305659)     A . . . . . . . . . . . . . . . . . . .
Mamu-E*0115 (EU305661)     A . . . . . . . . . . . . . . . . . . .
Mamu-I*010603 (EU305656)   . . A . . . . . . . . . . . . . . . . .
```

References

1. Thomas, R.K. *et al*. Sensitive mutation detection in heterogeneous cancer specimens by massively parallel picoliter reactor sequencing. *Nat. Med.* **12**, 852–855 (2006).

2. Wheeler, D.A. *et al.* Complete genome sequence of an individual by massively parallel DNA sequencing. *Nature* **452**, 872–876 (2008).

3. Quinlan, A.R. Stewart, D.A., Stromberg, M.P. & Marth, G.T. Pyrobayes: An improved base caller for SNP Discovery in pyrosequences. *Nat. Methods* **5**, 179–181 (2008).

4. Campbell, P.J. *et al.* Subconal phylogenetic structures in cancer revealed by ultra-deep sequencing. *Proc. Natl. Acad. Sci. USA* **105**, 13081–13086 (2008).

5. Otting, N. *et al*. Unparalleled complexity of the MHC class I region in rhesus macaques. *Proc. Natl. Acad. Sci. USA* **102**, 1626–1631 (2005).

6. Otting, N. *et al*., A snapshot of the *Mamu-B* genes and their allelic repertoire in rhesus macaques of Chinese origin. *Immunogenetics* **60**, 507–514 (2008).

7. Wiseman, R.W. *et al*. Siman immunodeficiency virus SIVmac239 infection of major histocompatibility complex-identical cynomolgus macaques from Mauritius. *J. Virol*. **81**, 349–361 (2007).