Published in final edited form as:

J Med Primatol. 2018 December ; 47(6): 379–387. doi:10.1111/jmp.12353.

Determination of major histocompatibility class I and class II genetic composition of the Caribbean Primate Center specific pathogen-free rhesus macaque (*Macaca mulatta*) colony based on massively parallel sequencing

SREETHARAN KANTHASWAMY^{1,2,*}, ROBERT F. OLDT², JILLIAN NG¹, DAVID GLENN SMITH¹, MELWEEN I. MARTÍNEZ³, and CARLOS A. SARIOL^{3,4,5}

¹California National Primate Research Center, University of California, Davis, CA, USA

²School of Mathematics and Natural Sciences, Arizona State University (ASU) at the West Campus, Glendale, AZ, USA

³Caribbean Primate Research Center, University of Puerto Rico-Medical Sciences Campus, San Juan, Puerto Rico

⁴Department of Microbiology and Medical Zoology, University of Puerto Rico-Medical Sciences Campus, San Juan, Puerto Rico

⁵Department of Medicine, University of Puerto Rico-Medical Sciences Campus, San Juan, Puerto Rico

Abstract

Background—Knowledge of MHC composition and distribution in rhesus macaque colonies is critical for management strategies that maximize the utility of this model for biomedical research.

Methods—Variation within the *Mamu-A* and *Mamu-B* (class I) and *DRB*, *DQA/B* and *DPA/B* (class II) regions of 379 animals from the Caribbean Primate Research Center's specific pathogen free (SPF) colony was examined using massively parallel sequencing.

Results—Analyses of the seven MHC loci revealed a background of Indian origin with high levels of variation despite past genetic bottlenecks. All loci exhibited mutual linkage disequilibria while conforming to Hardy–Weinberg expectations suggesting the achievement of mutation–selection balance.

Conclusion—The CPRC's SPF colony is a significant resource for research on AIDS and other infectious agents. Characterizing colony-wide MHC variability facilitates the breeding and selection of animals bearing desired haplotypes and increases the investigator's ability to understand the immune responses mounted by these animals.

*Corresponding Author: Sree Kanthaswamy, School of Mathematics and Natural Sciences, ASU at the West Campus, 4701 W Thunderbird Rd., Glendale, AZ 85306-4908, Phone: (530) 219-2017, sree.kanthaswamy@asu.edu.
DR. SREE KANTHASWAMY (Orcid ID: 0000-0003-4503-6886)

Keywords

MHC genotyping; Next Generation Sequence (NGS); genetic management; Mamu haplotypes; colony genetic structure

Introduction

As genes within the major histocompatibility complex (MHC) class I and class II regions facilitate prediction of the immune response, the use of MHC-defined macaques as research subjects improves the reproducibility and validity of experimental results and is thus a vital component of many NIH funded biomedical research programs. Figure 1 illustrates the various human MHC class I and class II regions including the genes coding for the human leukocyte antigens (HLA)-*A*, *-B*, *-E*, and *-F* orthologs exhibited by macaques [1–3]. As Figure 1 depicts, the macaque MHC class I genomic region has many more class I genes than the HLA-class I genomic region as a result of gene duplication events. For instance, rhesus macaques exhibit 4 fold more potentially functional class I proteins than humans [4]. While the macaque MHC class I region is far more complicated in its organization than the HLA-class I region of humans, the class II genes are more conserved between macaques and humans [5] (although many linked macaque MHC class I and II genes have yet to be adequately described [6]). The advent of massively parallel sequencing (MPS) technologies has aided the comprehensive examination and characterization of MHC organization of colony bred macaques [7]. However, there is no efficient protocol for breeding colony managers to compile this information and utilize it to maximize the long-term efficiency of MHC-defined colony production.

As part of its MHC screening program to provide larger cohorts of animals with enhanced MHC characterization to investigators, the Caribbean Primate Research Center (CPRC) has been testing its specific pathogen-free (SPF) rhesus macaques (*Macaca mulatta*) for the Mamu-A001, A002, A011, B001, B003, B008 and B017 haplotypes by genotyping MHC-linked short tandem repeat (STR) loci and/or performing sequence specific primer-polymerase chain reaction (SSP-PCR) amplification [8, 9]. Since some MHC haplotypes, groups of alleles of MHC genes that are co-inherited, serve a critical simian immunodeficiency virus (SIV) epitope presentation function in macaques that is associated with an adaptive cellular immune response in animals infected with SIV [10–12], individuals exhibiting these haplotypes are considered highly desirable in AIDS research. For instance, approximately 20% of *Mamu-B*17*-positive and 50% of *Mamu-B*008*-positive rhesus macaques can control replication of SIVmac239 spontaneously [13, 14]. Individuals with the *Mamu-A*01* allele, which, like the *Mamu B*01* allele, occurs in 20–30% of Indian rhesus macaques [15], can control viral replication better and survive longer than those without the allele [16, 17], and those with the class I alleles *Mamu-A*01*, *Mamu-B*17*, and *Mamu-B*29* can control the replication of a highly pathogenic SIV clone [18]. Several class II haplotypes, including Mamu-DRB1 0306 and Mamu-DRB1 1003, have also been shown to be over-represented in virus-infected macaque elite controllers [21], and ongoing efforts in the search for effective therapeutics and vaccines against HIV/AIDS have also identified several CD8⁺ T epitopes that play a crucial role in responses against SIV [18, 19, 20]. As

more haplotypes that influence the acute-phase cytotoxic T lymphocyte responses (CTR) and long-term clinical outcomes of SIV/HIV infection are identified [7], requests by investigators for information on these additional MHC haplotypes will also increase.

The CPRC is now relying on high throughput MHC sequencing to increase efficiency in characterizing additional class I and class II haplotypes. The introduction of macaque MHC genotyping via deep sequencing techniques [22, 23] has enhanced sample throughput while lowering per sample costs, which allows increased efficiency in MHC typing by generating complete haplotype libraries from significant numbers of colony animals. This study reports MHC data from the CPRC SPF colony that was recently generated using the deep sequencing method first described in Wiseman et al. [23]. The composition and distribution of MHC class I and class II haplotypes from the CPRC SPF colony described here will directly benefit the development of the center's SPF colony for AIDS-related research as well as initiatives on infectious or autoimmune diseases in which the immune response of the MHC class I or II haplotypes play a crucial role.

Methods and Materials

MHC genotyping of a study cohort of 379 CPRC SPF rhesus macaques was performed by the Genetics Services Unit at the Wisconsin National Primate Research Center (WNPRC) on a fee for service basis. Genomic DNA (gDNA) extracted from frozen blood samples was amplified with a Fluidigm Access Array using a multiplex panel of PCR primers flanking the hypervariable (i.e., most polymorphic [7]) peptide binding domains encoded by exon 2 of class I (*Mamu-A* and *-B*) and class II (*Mamu-DRB*, *-DQA*, *-DQB*, *-DPA* and *-DPB*) loci (Figure 1). Over 20 million sequence reads were generated with an average of 53,000 reads per animal. The amplicons were cleaned-up, pooled and then sequenced on an Illumina MiSeq instrument [R. Wiseman, pers. comm. August 2017; September 2017] [24]. The Genetics Services Unit's custom database of rhesus macaque sequences of major class I and class II alleles was used to manually map the sequence reads, each of which was tagged with a unique barcode that referenced the animal from which it was derived, to these reference sequences and assign them to the allele they most likely represent [R. Wiseman, pers. comm. August 2017; September 2017] [24]. These 'major' alleles that are defined by known haplotypes represent the vast majority of sequence reads and are assumed to initiate most of the immune responses [24].

Genotyping based on MiSeq sequence reads is limited because the sequence reads do not span entire open reading frames to allow an exact identification of alleles but rather rely on low-resolution definitions of lineages as families of closely related sequences that differ by only one or a few nucleotide differences [R. Wiseman, pers. comm. November 2017]. A consequence of the short amplicons is that the class II sequence data that have only been defined by MiSeq sequencing, which is the case for some *DPA/DPB* and *DQA/DQB* analyses, cannot be directly inferred from the detailed class II allele calls reported by Otting et al. [25] who relied on full length sequences of these regions in their study (see Supplementary table II). These haplotypes are given preliminary names such as Mamu-DPA02g1 (Table 4).

As systems of nomenclature for abbreviated haplotypes that define MHC class I and II alleles vary and the data for this study were generated by the WNPRC Genetics Services Unit, their nomenclature system was employed in the present study [23–30]. Three of the seven MHC class I and II regions examined in this study, *Mamu-A*, *-B* and *-DRB*, contain alleles of more than a single locus (see Fig. 1). A combination of alleles of these loci comprises a haplotype of which less inclusive sub-haplotypes that define minor alleles are subsets, and these haplotypes are indicated by a short name/abbreviation for simplification. For example, the haplotype “A008” of the *Mamu-A* region is defined by its major allele, *A1*008* of the *A1* locus, and minor alleles of the *A2* and *A3* loci. Supplementary tables I and II list the abbreviated haplotypes that were identified in each animal by matching sequence reads from the class I and class II regions to previously described class I and class II alleles.

Haplotype diversity, or the number of haplotypes per locus (H_a), observed (H_O) and expected (H_E) heterozygosity, and F_{IS} were calculated using the program Arlequin version 3 [31]. Deviations from Hardy–Weinberg Equilibrium (HWE) for each of the seven MHC loci were tested using the exact HWE test implemented in GENEPOP 3.2 [32] with the Markov chain method [33] (dememorization: 1000; batches: 500; iterations per batch: 1000). Linkage disequilibrium (LD) among the seven loci, which can be expressed in terms of the correlation coefficient (LD_r) [34], was tested using the LINKDOS program described by Garnier-Gere and Dillmann [35].

Results

Due to the sensitivity of the MPS technique employed, several thousand (53,000, on average) class I and class II sequence reads were identified in each animal allowing the definition of haplotypes as rare as those that occurred at a frequency of 1%. The *Mamu-A* haplotypes are common in Indian origin rhesus macaques with A004, which exhibited a frequency of 57.9% in the cohort studied, being the most predominant haplotype. The *Mamu-B* haplotypes in this colony are also common Indian origin rhesus haplotypes with B012b being the most prevalent at 36.3%. The *DRB* haplotypes DR03a and DR04a, respectively, were detected in over 30% of the chromosomes surveyed, and a previously undescribed *DRB* haplotype, DR-U, was also discovered. Within the *DQA/DQB* region, the most common *DQA* haplotypes were 01g2 (34.4%) and 26g1 (31.3%) while the most prevalent *DQB* haplotypes were 06_01 (34.8%) and 18g1 (31.2%). A majority of *DPA* haplotypes were 02g1, which occurred among 67.5% of the chromosomes assayed. Among the *DPB* haplotypes, 15g1 was most frequently detected in the colony (56.9%).

Table 2 provides estimates of genetic diversity parameters for the *Mamu-A*, *Mamu-B*, *DRB*, *DQA/DQB* and *DPA/DPB* loci/regions, all of which were relatively high. The number of haplotypes per locus (or per region, in the case of *Mamu-A*, *-B* and *-DRB* [G. Doxiadis, pers. comm. January 2018]) fell between 5 (*DPA*) and 15 (*Mamu-B*) with an average of 10.14. High heterozygosity was exhibited for all seven of the surveyed MHC regions. H_O estimates ranged from approximately 51% (*DPA*) to 80% (*Mamu-B*) with an average of 68.6% while H_E estimates ranged from 51% (*DPA*) to 82% (*Mamu-B*) with an average of 68.6%. All loci/regions with the exception of *Mamu-B* ($F_{IS} = 0.019$) exhibited insignificant

levels of inbreeding ($FIS = -0.023$ to 0.005). While class I and class II haplotype frequencies at all seven loci/regions fit HWE expectations (Table 2), indicating mutation-selection balance; estimates of highly significant LD ($p < 0.0001$) suggest that the haplotypes within each of these different genomic regions have a strong tendency to co-segregate due to their physical linkage. The wide range of correlation coefficients (LD_r) from 0.099 (*Mamu-A-DPB*) to 0.465 (*DQA-DQB*) (Table 3a), however, reveals variation in recombination intensity within the MHC region. For instance, recombination within the chromosomal region between *Mamu-A* and *Mamu-B* is relatively more common than that between either of these loci and the *DRB* region while allelic recombination with the *DQA-DQB* and *DPA-DPB* gene pairs occurred at a much lower rate (Table 3b). Figure 1 illustrates the map of the rhesus macaque MHC region based on these LD estimates.

Discussion

The distributions and frequencies of MHC class I and class II haplotypes can be examined collectively and comprehensively (Tables 1 and 2) in a large cohort of rhesus macaques. Concordant with previous studies of smaller samples of rhesus macaques by Karl et al.'s [22], Wiseman et al.'s [7, 23] and Otting et al.'s [25], the *Mamu-A*, *Mamu-B*, *DRB*, *DQA/DQB* and *DPA/DPB* loci that were analyzed in the present study were all polymorphic with moderate (over 50%) to high (over 80%) levels of heterozygosity. However, our colony-wide assessment of MHC composition revealed that a limited number of *Mamu-A*, *B*, and *DRB* haplotypes accounted for more than half of the MHC variability in the CPRC's SPF colony. As indicated by the *Mamu*-class I and class II haplotype frequencies, the MHC composition of this colony, particularly that for the *Mamu-A* and *B* haplotypes, is comparable to that of other predominantly Indian origin colonies of the US National Primate Research Centers (NPRC) [23]. For instance, haplotype A002a is common in Indian rhesus macaques and was observed in many individuals in this study [23, 28, 29].

Consistent with the study by Doxiadis et al. [36], highly significant linkage disequilibria ($p < 0.0001$) among all seven loci/regions were observed in this study confirming that the MHC region surveyed tends to behave as a single linkage group. While recombination is low within this region, varying recombination rates across the region have contributed to the rhesus macaque MHC haplotypic structure and diversity. Based on the analyses of the *Mamu-A*, *Mamu-B* and *DRB* regions of a sample of Indian rhesus macaques, Doxiadis et al. [36] argued that MHC haplotype diversity is maintained by recombination events involving a relatively few segments within these loci. However, the overall low rates of recombination and the long persistence times of MHC haplotypes within populations [25, 37] make MHC loci reliable indicators of ancestry and/or geographic origins. For instance, haplotypes Mamu-A 002a and 023 and the Mamu-B-012a, 012b, 017a, 017d, 017f, 024a, 029, 043a, 47a, and 93 are derived haplotypes that are only found in Indian animals while haplotype A002b, which appears to be Chinese in origin, was not observed in this colony [23, 28, 29].

Although rare MHC haplotypes that appear at or below frequencies of 1% in various rhesus macaque breeding colonies have been attributed to admixture between animals from different regional populations [23], none of the rare MHC haplotypes in the present study have been documented to be derived from locations outside the Indian subcontinent. The

attribution of the MHC haplotype structure of the CPRC SPF colony solely to its Indian founders is consistent with Kanthaswamy et al.'s [38, 39] studies based on autosomal short tandem repeat (STR) loci which support the sole Indian provenance of these founders. Ancestry assignments based on STRs have demonstrated that the CPRC SPF colony's degree of Indian ancestry is among the highest of those that have been calculated within primate breeding facilities [38, 39], further confirming the CPRC as a high-value resource for AIDS research [40, 41]. Despite the pure Indian ancestry of the study animals, haplotypes such as Mamu-A 001, 004, 008 and 224a and Mamu-B 001a, 002, 015b, 048 and 069a that were identified in these animals that are shared with Chinese rhesus macaques [23, 28, 29] are regarded as ancestral haplotypes probably due to incomplete lineage sorting of MHC [42].

The extended MHC haplotypes of most CPRC SPF animals (Table 4) were easily inferred mainly due to the extent of LD within this region and the limited number of founders in the source population of Cayo Santiago from which animals were translocated to the CPRC to establish this SPF colony [38, 43]. In addition to being a closed colony, the Cayo Santiago island population, the sole source of the CPRC SPF colony, has experienced extreme founder effects and genetic bottlenecks since its establishment 80 years ago [38, 39, 43, 44]. This is probably why several MHC homozygous individuals with five of the most common Indian-origin haplotypes were identified in this study including 16 individuals that were homozygous for the extended haplotype A004/B012b/DR04a/DQA01g2/DQB06_01/DPA02g1/DPB15g1, which is a haplotype reported to be very common among US primate centers that are predominantly composed of Indian rhesus macaques [R. Wiseman, pers. comm. August 2017; September 2017]. The seven *Mamu-A* haplotypes observed in this study represent fewer than half of the common *Mamu-A* haplotypes in Indian-origin rhesus macaques reported in Wiseman et al. [7], reflecting the relatively lower genetic diversity of the CPRC animals. The presence in the CPRC SPF colony of only 18 Mamu-B008 (2.4%) and 4 Mamu-B017 (0.5%) haplotypes (see Table 1 and Supplementary table 1), which have been identified as being critically important for SIV/HIV studies because of their association with exceptional control of SIVmac239 replication during the chronic phase of viral infection, also underscores the relative isolation of this colony compared to other breeding colonies in the US [13].

MHC typing of the CPRC SPF colony is critical for selecting animals for specific experimental studies and the management of this colony. While the results reveal that this relatively genetically homogeneous colony would be a good source for MHC haplotype-matched groups of animals for investigators, the array of MHC haplotypes exhibited within it still allows a significant likelihood for MHC incompatibility among study animals. In either case, information on the composition and distribution of MHC class I and class II haplotypes within the CPRC SPF colony will enhance the investigator's ability to predict and understand immune responses mounted by the majority of these SPF animals and prevent risks of incompatibility.

As knowledge of the MHC genetic composition of colony animals is also fundamental for making sound colony management decisions, lacking this data will make it impossible to anticipate and meet investigator requirements for particular types of MHC haplotypes. For

instance, in the past, selection regimes at the CPRC were designed to increase the percentage of *Mamu-A*001*-positive animals in the colony. While such schemes for selective breeding can increase the proportion of animals in a colony that is homozygous for *Mamu-A*001*, they are ineffective for increasing *Mamu-A*001* allele frequencies unless the reproductive success of *Mamu-A*001*-positive breeders can be increased above that of *Mamu-A*001*-negative breeders. Nevertheless, more recently, HIV/AIDS researchers have more often routinely requested *Mamu-A001* negative animals instead. In order to avoid any dramatic fluctuation in colony-wide frequency of MHC haplotypes or unintentional elimination of valuable haplotypes from the colony, as one haplotype is favored over another by researchers, colony managers at the CPRC are now choosing male breeders with common MHC haplotypes and female breeders with rare MHC haplotypes. In the short term, a breeding scheme that prioritizes the maintenance of rare alleles or haplotypes over heterozygosity could result in substantially depleting diversity at the MHC as well as other loci in the genome [45]. To ameliorate any potential long-term losses in MHC variability, novel MHC variation can be introduced into the colony by translocating individuals from Cayo Santiago and/or other genetically and demographically healthy NPRC colonies.

Beyond the characterization of class I and class II haplotypes, information about the larger-scale genetic structure of the macaque MHC region and diversity of MHC genes other than those within class I and class II regions remains somewhat restricted. Advances in the application of cost-effective, high throughput and deep sequencing technologies have triggered comprehensive MHC testing of large macaque cohorts [7]. Continued improvements in the detection methods for MHC typing, including enhanced next-generation techniques, will facilitate a better understanding of the immune response, not only to SIV infection but infection by many other pathogens as well, and expand the usefulness of rhesus macaques in biomedical research. However, with the availability of more MHC information, colony managers must be cautioned that breeding and animal selection schemes designed strictly around this single linkage group will risk losing variability at other potentially important loci [46–48]. Overall genetic diversity can be conserved by employing colony management protocols that rely on strategies to monitor genetic diversity across a panel of unlinked autosomal loci that is distributed throughout the genome such as STRs or SNPs [38, 39].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the dedicated staff at Cayo Santiago and Sabana Seca Field Station for their continued dedication and care in support of the CPRC's animal population. We thank the WNPRC's Genetics Services Unit for performing the MPS analysis on a fee for service basis. We would like to thank WNPRC's Dr Roger Wiseman for his valuable insight into the MHC analysis and data. We are very grateful to Dr Gaby Doxiadis of the Biomedical Primate Research Centre in Rijswijk, The Netherlands, for kindly reviewing this manuscript, and providing critical comments to improve it.

Animals involved in this study were managed in compliance with Institutional Animal Care and Use Committee (IACUC) regulations and with the National Institutes of Health (NIH) guidelines prescribing the humane care and use of laboratory animals. This research complies with the protocols approved by the IACUC (protocol numbers:

7890112 and 7890113) at the University of Puerto Rico and with the American Society of Primatologists (ASP) Principles for the Ethical Treatment of Nonhuman Primates. All research reported in this manuscript was performed in accordance with the legal requirements of Puerto Rico. This research was supported by NIH-ORIP grant number: 5U42OD021458-16.

References

1. Boyson JE, Iwanaga KK, Golos TG, Watkins DI. Identification of the rhesus monkey HLA-G ortholog. Mamu-G is a pseudogene. *J Immunol.* 1996; 157:5428–5437. [PubMed: 8955191]
2. Slierendregt BL, Otting N, van Besouw N, Jonker M, Bontrop RE. Expansion and contraction of rhesus macaque DRB regions by duplication and deletion. *J Immunol.* 1994; 152:2298–2307. [PubMed: 8133043]
3. Watanabe N, McAdam SN, Boyson JE, Piekarczyk MS, Yasutomi Y, Watkins DI, Letvin NL. A simian immunodeficiency virus envelope V3 cytotoxic T-lymphocyte epitope in rhesus monkeys and its restricting major histocompatibility complex class I molecule Mamu-A*02. *J Virol.* 1994; 68:6690–6696. [PubMed: 7521921]
4. 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]
5. Shiina T, Suzuki S, Kulski JK. MHC Genotyping in Human and Nonhuman Species by PCR based Next-Generation Sequencing, Next Generation Sequencing - Advances, Applications and Challenges. Kulski Jerzy, Dr, editorInTech; 2016. Available from: <https://www.intechopen.com/books/next-generation-sequencing-advances-applications-and-challenges/mhc-genotyping-in-human-and-nonhuman-species-by-pcrbased-next-generation-sequencing>
6. Shen S, Pyo CW, Vu Q, Wang R, Geraghty DE. The essential detail: the genetics and genomics of the primate immune response. *ILAR J.* 2013; 54:181–195. [PubMed: 24174441]
7. Wiseman RW, Karl JA, Bohn PS, Nimityongskul FA, Starrett GJ, O'Connor DH. Haplessly hoping: macaque major histocompatibility complex made easy. *ILAR J.* 2013; 54:196–210. [PubMed: 24174442]
8. Penedo MC, Bontrop RE, Heijmans CM, Otting N, Noort R, Rouweler AJ, de Groot N, de Groot NG, Ward T, Doxiadis GG. Microsatellite typing of the rhesus macaque MHC region. *Immunogenetics.* 2005; 57:198–209. [PubMed: 15900491]
9. Knapp LA, Lehmann E, Piekarczyk MS, Urvater JA, Watkins DI. A high frequency of Mamu-A*01 in the rhesus macaque detected by polymerase chain reaction with sequence-specific primers and direct sequencing. *Tissue Antigens.* 1997; 50:657–661. [PubMed: 9458122]
10. Janeway CA, Jr, Travers P, Walport M, Shlomchik MJ. *Immunobiology: The Immune System in Health and Disease.* 5. New York: Garland Science; 2001.
11. Carrington M, O'Brien SJ. The influence of HLA genotype on AIDS. *Annu Rev Med.* 2003; 54:535–551. [PubMed: 12525683]
12. Bonhomme M, Doxiadis GG, Heijmans CM, Vervoort V, Otting N, Bontrop RE, Crouau-Roy B. Genomic plasticity of the immune-related Mhc class I B region in macaque species. *BMC Genomics.* 2008; 9:514. [PubMed: 18973694]
13. Yant LJ, Friedrich TC, Johnson RC, May GE, Maness NJ, Enz AM, Lifson JD, O'Connor DH, Carrington M, Watkins DI. The high-frequency major histocompatibility complex class I allele Mamu-B*17 is associated with control of simian immunodeficiency virus SIVmac239 replication. *J Virol.* 2006; 80:5074–5077. [PubMed: 16641299]
14. Loffredo JT, Maxwell J, Qi Y, Glidden CE, Borchardt GJ, Soma T, Bean AT, Beal DR, Wilson NA, Rehauer WM, Lifson JD, Carrington M, Watkins DI. Mamu-B*08-positive macaques control simian immunodeficiency virus replication. *J Virol.* 2007; 81:8827–8832. [PubMed: 17537848]
15. Kanthaswamy S, Kou A, Satkoski J, Penedo MC, Ward T, Ng J, Gill L, Lerche NW, Erickson BJ, Smith DG. Genetic characterization of specific pathogen-free rhesus macaque (*Macaca mulatta*) populations at the California National Primate Research Center (CNPRC). *Am J Primatol.* 2010; 72:587–599. [PubMed: 20162538]
16. Mothe BR, Weinfurter J, Wang C, Rehauer W, Wilson N, Allen TM, Allison DB, Watkins DI. Expression of the major histocompatibility complex class I molecule Mamu-A*01 is associated

- with control of simian immunodeficiency virus SIVmac239 replication. *J Virol.* 2003; 77:2736–2740. [PubMed: 12552014]
17. Muhl T, Krawczak M, Ten Haaft P, Hunsmann G, Sauermann U. MHC class I alleles influence set-point viral load and survival time in simian immunodeficiency virus-infected rhesus monkeys. *J Immunol.* 2002; 169:3438–3446. [PubMed: 12218167]
 18. O'Connor DH, Mothe BR, Weinfurter JT, Fuenger S, Rehauer WM, Jing P, Rudersdorf RR, Liebl ME, Krebs K, Vasquez J, Dodds E, Loffredo J, Martin S, McDermott AB, Allen TM, Wang C, Doxiadis GG, Montefiori DC, Hughes A, Burton DR, Allison DB, Wolinsky SM, Bontrop R, Picker LJ, Watkins DI. Major histocompatibility complex class I alleles associated with slow simian immunodeficiency virus disease progression bind epitopes recognized by dominant acute-phase cytotoxic-T-lymphocyte responses. *J Virol.* 2003; 77:9029–9040. [PubMed: 12885919]
 19. Valentine LE, Watkins DI. Relevance of studying T cell responses in SIV-infected rhesus macaques. *Trends Microbiol.* 2008; 16:605–611. [PubMed: 18964016]
 20. 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]
 21. Giraldo-Vela JP, Rudersdorf R, Chung C, Qi Y, Wallace LT, Bimber B, Borchardt GJ, Fisk DL, Glidden CE, Loffredo JT, Piaskowski SM, Furlott JR, Morales-Martinez JP, Wilson NA, Rehauer WM, Lifson JD, Carrington M, Watkins DI. The major histocompatibility complex class II alleles Mamu-DRB1*1003 and -DRB1*0306 are enriched in a cohort of simian immunodeficiency virus-infected rhesus macaque elite controllers. *J Virol.* 2008; 82:859–870. [PubMed: 17989178]
 22. Karl JA, Wiseman RW, O'Connor DH. Cost-effective sequence-based nonhuman primate MHC class I genotyping from RNA. *Methods.* 2009; 49:11–17. [PubMed: 19442739]
 23. Wiseman RW, Karl JA, Bimber BN, O'Leary CE, Lank SM, Tuscher JJ, Detmer AM, Bouffard P, Levenkova N, Turcotte CL, Szekeres E Jr, Wright C, Harkins T, O'Connor DH. Major histocompatibility complex genotyping with massively parallel pyrosequencing. *Nat Med.* 2009; 15:1322–1326. [PubMed: 19820716]
 24. O'Connor DH, Karl JA. MHC Genotyping via Deep Sequencing. UW-Madison: WNPRC; 2016. https://www.primate.wisc.edu/wp-content/uploads/2016/09/GS_WhitePaper-MHC-16.09.23.pdf [accessed April 2018]
 25. Otting N, van der Wiel MK, de Groot N, de Vos-Rouweler AJ, de Groot NG, Doxiadis GG, Wiseman RW, O'Connor DH, Bontrop RE. The orthologs of HLA-DQ and -DP genes display abundant levels of variability in macaque species. *Immunogenetics.* 2017; 69:87–99. [PubMed: 27771735]
 26. 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]
 27. Otting N, Heijmans CM, van der Wiel M, de Groot NG, Doxiadis GG, Bontrop RE. A snapshot of the Mamu-B genes and their allelic repertoire in rhesus macaques of Chinese origin. *Immunogenetics.* 2008; 60:507–514. [PubMed: 18618105]
 28. Doxiadis GG, de Groot N, Otting N, Blokhuis JH, Bontrop RE. Genomic plasticity of the MHC class I A region in rhesus macaques: extensive haplotype diversity at the population level as revealed by microsatellites. *Immunogenetics.* 2011; 63:73–83. [PubMed: 20949353]
 29. Budde ML, Wiseman RW, Karl JA, Hanczaruk B, Simen BB, O'Connor DH. Characterization of Mauritian cynomolgus macaque major histocompatibility complex class I haplotypes by high-resolution pyrosequencing. *Immunogenetics.* 2010; 62:773–780. [PubMed: 20882385]
 30. Karl JA, Graham ME, Wiseman RW, Heimbruch KE, Gieger SM, Doxiadis GG, Bontrop RE, O'Connor DH. Major histocompatibility complex haplotyping and long-amplicon allele discovery in cynomolgus macaques from Chinese breeding facilities. *Immunogenetics.* 2017; 69:211–229. [PubMed: 28078358]
 31. Excoffier L, Laval G, Schneider S. Arlequin (version 3. 0): an integrated software package for population genetics data analysis. *Evol Bioinform Online.* 2007; 1:47–50. [PubMed: 19325852]

32. Raymond M, Rousset F. GENEPOP (Version 1. 2): Population Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity*. 1995; 86:248–249.
33. Guo SW, Thompson EA. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*. 1992; 48:361–372. [PubMed: 1637966]
34. Weir BS. Inferences about linkage disequilibrium. *Biometrics*. 1979; 35:235–254. [PubMed: 497335]
35. Garnier-Gere P, Dillmann C. A computer program for testing pairwise linkage disequilibria in subdivided populations. *J Hered*. 1992; 83:239.
36. Doxiadis GG, de Groot N, Otting N, de Vos-Rouweler AJ, Bolijn MJ, Heijmans CM, de Groot NG, van der Wiel MK, Remarque EJ, Vangenot C, Nunes JM, Sanchez-Mazas A, Bontrop RE. Haplotype diversity generated by ancient recombination-like events in the MHC of Indian rhesus macaques. *Immunogenetics*. 2013; 65:569–584. [PubMed: 23715823]
37. Watkins DI, McAdam SN, Liu X, Strang CR, Milford EL, Levine CG, Garber TL, Dogon AL, Lord CI, Ghim SH, et al. New recombinant HLA-B alleles in a tribe of South American Amerindians indicate rapid evolution of MHC class I loci. *Nature*. 1992; 357:329–333. [PubMed: 1589035]
38. Kanthaswamy S, Ng J, Hernandez-Pacheco R, Ruiz-Lambides A, Maldonado E, Martinez MI, Sariol CA. The Population Genetic Composition of Conventional and SPF Colonies of Rhesus Macaques (*Macaca mulatta*) at the Caribbean Primate Research Center. *J Am Assoc Lab Anim Sci*. 2016; 55:147–151. [PubMed: 27025804]
39. Kanthaswamy S, Oldt RF, Ng J, Ruiz-Lambides AV, Maldonado E, Martinez MI, Sariol CA. Population Genetic Structure of the Cayo Santiago Colony of Rhesus Macaques (*Macaca mulatta*). *J Am Assoc Lab Anim Sci*. 2017; 56:396–401. [PubMed: 28724489]
40. Ling B, Veazey RS, Luckay A, Penedo C, Xu K, Lifson JD, Marx PA. SIV(mac) pathogenesis in rhesus macaques of Chinese and Indian origin compared with primary HIV infections in humans. *AIDS*. 2002; 16:1489–1496. [PubMed: 12131186]
41. Trichel AM, Rajakumar PA, Murphey-Corb M. Species-specific variation in SIV disease progression between Chinese and Indian subspecies of rhesus macaque. *J Med Primatol*. 2002; 31:171–178. [PubMed: 12390539]
42. Rogers J, Gibbs RA. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat Rev Genet*. 2014; 15:347–359. [PubMed: 24709753]
43. Kessler MJ, Rawlins RG. A 75-year pictorial history of the Cayo Santiago rhesus monkey colony. *Am J Primatol*. 2016; 78:6–43. [PubMed: 25764995]
44. Buettner-Janusch J, Sockol M. Genetic studies on free-ranging macaques of Cay Santiago. II. 6-phosphogluconate dehydrogenase and NADH-methemoglobin reductase (NADH-diaphorase). *Am J Phys Anthropol*. 1977; 47:375–379. [PubMed: 412423]
45. Hill AV, Allsopp CE, Kwiatkowski D, Anstey NM, Twumasi P, Rowe PA, Bennett S, Brewster D, McMichael AJ, Greenwood BM. Common west African HLA antigens are associated with protection from severe malaria. *Nature*. 1991; 352:595–600. [PubMed: 1865923]
46. Gilpin M, Wills C. MHC and Captive Breeding: A Rebuttal. *Conservation Biology*. 1991; 5:554–555.
47. Hughes AL. MHC Polymorphism and the Design of Captive Breeding Programs. *Conservation Biology*. 1991; 5:249–251.
48. Miller PS, Hedrick PW. MHC Polymorphism and the Design of Captive Breeding Programs: Simple Solutions Are Not the Answer. *Conservation Biology*. 1991; 5:556–558.
49. Meyer D, Thomson G. How selection shapes variation of the human major histocompatibility complex: a review. *Ann Hum Genet*. 2001; 65:1–26. [PubMed: 11415519]

Figure 1.

Map of the macaque MHC regions (above), based on LDr estimates generated by the present study, compared to the orthologous human MHC region, also called the HLA (human leukocyte antigen) complex (below) adapted from [49] and [5]. An A001 haplotype is the summary of all A alleles as shown in Supplementary table 1. The same is true for the B and DRB haplotypes (Supplementary table I).

Table 1

MHC class I and class II haplotype frequencies. See supplementary tables I and II for *Mamu-A*, *Mamu-B* and *DRB* haplotype-specific alleles. For all haplotypes, please refer to <https://www.ebi.ac.uk/ipd/mhc/>

<i>Mamu-A</i>													
A001	A002a	A224a	A004	A057	A008	A023	A031						
0.131	0.055	0.009	0.579	0.001	0.222	0.001	0.001						
<i>Mamu-B</i>													
B001a	B002	B008	B012a	B012b	B015b	B017a	B024a	B028	B028'	B043a	B047a	B048	B055
0.095	0.017	0.024	0.05	0.363	0.082	0.005	0.004	0.128	0.011	0.012	0.123	0.044	0.001
													0.04
													0.043
<i>DRB</i>													
DR01a	DR01d	DR02	DR02'	DR03a	DR03f	DR04a	DR06	DR08	DR09a	DR14b	DR-U		
0.015	0.001	0.009	0.008	0.313	0.074	0.347	0.111	0.003	0.112	0.001	0.007		
<i>DQA</i>													
01_05_02	01_06	01g1	01g2	02g1	23_01	23_02	23_03	24g1	26g1	26g2			
0.001	0.007	0.067	0.344	0.001	0.113	0.013	0.001	0.004	0.313	0.135			
<i>DQB</i>													
06_01	06_06	06_09	06g2	18_02	18_10	18g1	18g2	18g3					
0.348	0.007	0.001	0.068	0.112	0.004	0.312	0.013	0.134					
<i>DPA</i>													
02_08	02g1	02g2	04g1	06g	07_01								
0.016	0.675	0.127	0.009	0.131	0.042								
<i>DPB</i>													
01g1	03g	04g	06_02	06_04	07_01	08_01	15g1	19g					
0.132	0.009	0.001	0.094	0.015	0.124	0.015	0.569	0.042					

Table 2

Genetic diversity estimates across all seven MHC loci/regions. N: sample size, Ha: haplotype number per locus; HO: observed heterozygosity, HE: expected heterozygosity, FIS: inbreeding coefficient and sd: standard deviation. P values denote significant departure from HWE ($p > 0.01$).

Locus/region	N	Ha	HO	HE	FIS	p value	sd
<i>Mamu-A</i>	379	7	0.610	0.596	-0.023	0.799	0.00027
<i>Mamu-B</i>	379	15	0.799	0.815	0.019	0.611	0.00033
<i>DRB</i>	379	11	0.752	0.752	0.000	0.318	0.00032
<i>DQA</i>	379	10	0.749	0.749	-0.001	0.262	0.00045
<i>DQB</i>	376	8	0.745	0.747	0.003	0.272	0.00041
<i>DPA</i>	379	5	0.507	0.509	0.005	0.365	0.00046
<i>DPB</i>	379	8	0.633	0.635	0.001	0.409	0.00046
<i>Average</i>	378.57	9.14	0.69	0.69	0.00	0.43	0.00

Table 3a

Analysis of Linkage Disequilibrium between pairs of loci (where r is correlation coefficient, df is degrees of freedom and p is 0.001). Analysis was performed using the LinkDos program [35]

Locus 1	Locus 2	LD _r	X ²	df
<i>Mamu-A</i>	<i>Mamu-B</i>	0.142	1237.67	105
<i>Mamu-A</i>	<i>DRB</i>	0.155	1091.56	77
<i>Mamu-A</i>	<i>DQA</i>	0.158	865.67	70
<i>Mamu-A</i>	<i>DQB</i>	0.110	512.2	56
<i>Mamu-A</i>	<i>DPA</i>	0.108	385.67	35
<i>Mamu-A</i>	<i>DPB</i>	0.099	533.41	56
<i>Mamu-B</i>	<i>DRB</i>	0.148	2080.12	165
<i>Mamu-B</i>	<i>DQA</i>	0.151	1795.01	150
<i>Mamu-B</i>	<i>DQB</i>	0.137	1378.72	120
<i>Mamu-B</i>	<i>DPA</i>	0.109	799.93	75
<i>Mamu-B</i>	<i>DPB</i>	0.101	1110.98	120
<i>DRB</i>	<i>DQA</i>	0.297	3032.04	110
<i>DRB</i>	<i>DQB</i>	0.349	2677.88	88
<i>DRB</i>	<i>DPA</i>	0.155	1102.5	55
<i>DRB</i>	<i>DPB</i>	0.190	1847.95	88
<i>DQA</i>	<i>DQB</i>	0.465	2964.3	80
<i>DQA</i>	<i>DPA</i>	0.142	860.72	50
<i>DQA</i>	<i>DPB</i>	0.193	1669.33	80
<i>DQB</i>	<i>DPA</i>	0.166	850.22	40
<i>DQB</i>	<i>DPB</i>	0.224	1614.54	64
<i>DPA</i>	<i>DPB</i>	0.461	1907.07	40

Table 3b

Pairwise LDr estimates among the seven loci analyzed.

	Mamu-B	DRB	DQA	DQB	DPA	DPB
<i>Mamu-A</i>	0.142	0.155	0.158	0.110	0.108	0.099
<i>Mamu-B</i>		0.148	0.151	0.137	0.109	0.101
<i>DRB</i>			0.297	0.349	0.155	0.19
<i>DQA</i>				0.465	0.142	0.193
<i>DQB</i>					0.166	0.224
<i>DPA</i>						0.461

Table 4

The 23 common extended MHC founder haplotypes. The column on the extreme right presents the frequency of that particular founder haplotype (R. Wiseman, pers. comm. August 2017; September 2017).

Founder Haplotype #	Mamu-A Haplotype	Mamu-B Haplotype	Mamu-DRB Haplotype	Mamu-DQA Haplotype	Mamu-DQB Haplotype	Mamu-DPA Haplotype	Mamu-DPB Haplotype	Founder haplotype frequency
1	A004	B012b	DR04a	01g2	06_01	02g1	15g1	0.465
2	A004	B012b	DR03a	26g1	18g1	02g1	15g1	0.014
3	A002a	B012a	DR03f	01g1	06g2	07_01	19g	0.085
4	A004	B012a	DR04a	01g2	06_01	02_08	06_04	0.014
5	A008	B028	DR09a	26g2	18g3	06g	01g1	0.014
6	A004	B028'	DR03a	26g1	18g1	02g1	15g1	0.113
7	A008	B047a	DR06	23_01	18_02	02g1	15g1	0.014
8	A001	B047a	DR03a	26g1	18g1	02g2	07_01	0.056
9	A001	B047a	DR09a	26g2	18g3	06g	01g1	0.014
10	A001	B047a	DR04a	23_01	06_01	02g1	15g1	0.014
11	A001	B015b	DR03a	26g1	18g1	02g1	06_02	0.014
12	A004	B001a	DR03a	26g1	18g1	02g2	07_01	0.028
13	A004	B001a	DR04a	01g2	06_01	02g1	15g1	0.014
14	A224a	B001a	DR01a	26g2	18g3	07_01	19g	0.014
15	A004	B048	DR03a	26g1	18g1	02g1	06_02	0.014
16	A004	B048	DR03f	01g1	06g2	06g	01g1	0.014
17	A004	B069a	DR03a	26g1	18g1	02g2	07_01	0.014
18	A002a	B069a	DR04a	01g2	06_01	02g1	15g1	0.014
19	A001	B002	DR03f	01g1	06g2	02g1	15g1	0.014
20	A004	B008	DR06	23_01	18_02	02g1	15g1	0.014
21	A004	B017a	DR01a	26g2	18g3	02g2	07_01	0.014
22	A008	B043a	DR03a	26g1	18g1	02g2	07_01	0.014
23	A002a	B024a	DR02	24g1	18_10	06g	01g1	0.014