Available online at www.sciencedirect.com

ScienceDirect

Genomics 89 (2007) 402–412

GENOMICS

www.elsevier.com/locate/ygeno

A BAC-based contig map of the cynomolgus macaque (*Macaca fascicularis*) major histocompatibility complex genomic region

Atsushi Watanabe [a], Takashi Shiina [a], Sayoko Shimizu [a], Kazuyoshi Hosomichi [a], Kazuyo Yanagiya [a], Yuki F. Kita [a], Tetsuaki Kimura [a], Eiichi Soeda [b], Ryuzo Torii [c], Kazumasa Ogasawara [d], Jerzy K. Kulski [a,e], Hidetoshi Inoko [a,*]

[a] *Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1143, Japan*

[b] *Advanced GenoTechs Co., Nishioui, Tsukuba, Ibaragi 1710-234, Japan*

[c] *Department of Pathology, Shiga University of Medical Science, Seta, Tsukinowa-cho, Ohtsu, Shiga 520-2192, Japan*

[d] *Research Center for Animal Life Science, Shiga University of Medical Science, Seta, Tsukinowa-cho, Ohtsu, Shiga 520-2192, Japan*

[e] *Centre for Comparative Genomics, School of Information Technology, Murdoch University, Murdoch, WA 6150, Australia*

Received 7 July 2006; accepted 1 November 2006
Available online 14 December 2006

Abstract

The construction of a cynomolgus macaque (*Macaca fascicularis, Mafa*) BAC library for genomic comparison between rhesus and cynomolgus macaques is necessary to promote the cynomolgus macaque as one of the important experimental animals for future medical and biological research. In this paper, we constructed a cynomolgus macaque BAC library and a map of the MHC (*Mafa*) genomic region for comparison of the genomic organization and nucleotide similarities between the human, the chimpanzee, and the rhesus macaque. The BAC library consists of 221,184 clones with an average insert size of 83 kb, providing a sixfold coverage of the haploid genome. A total of 114 BAC clones and 54 PCR primer sets were used to construct a 4.3-Mb contig of the MHC region. Diversity analysis of genomic sequence from selected subregions of the MHC revealed that the cynomolgus sequence varied compared to rhesus macaque, human, and chimpanzee sequences by 0.48, 4.15, and 4.10%, respectively. From these findings, we conclude that the BAC library and *Mafa* genomic map are useful tools for genome analysis and will have important applications for comparative genomics and identifying regions of consequence in medical research.

Keywords: Cynomolgus; BAC library; MHC; Contig; Comparative genomics

The macaques (genus *Macaca*) are Old World monkeys that diverged from the human lineage about 29.4 to 34.5 million years ago [1]. They are important experimental animals for the study of human diseases, vaccines, and biology because of their genetic and behavioral similarities to humans and the nonhuman hominoids, such as chimpanzee and gorilla, which are in danger of extinction (http://www.iucn.org/). One of the macaque species, cynomolgus macaque (*Macaca fascicularis, Mafa*), alias crab-eating macaque or long-tailed macaque, has been well established since the early 20th century as an experimental animal for biomedical research, such as transplantation studies [2,3] and vaccine development against measles [4,5], HIV, or various other infectious diseases [6–10]. Although 65% of the approximately 13,000 nonhuman primates used for research at the National Institutes of Health in 1999 (http://www.ncrr.nih.gov/compmed/rhesusworkshopreport.pdf) were rhesus macaques (*Macaca mulatta, Mamu*), cynomolgus macaques have a greater potential as a surrogate animal model [11,12] because they are smaller and easier to manage, are more economical to maintain, and breed seasonally in captivity. The comparison between the genomes of the rhesus and the cynomolgus macaque is of structural, functional, and evolutionary interest because the divergence time between the two species is estimated to be only about 1.8–2.0 million years [13].

The organizational and functional comparisons between the major histocompatibility complex (MHC) genomic regions of the human and nonhuman primates has revealed the primate

* Corresponding author. Fax: +81 463 94 8884.
E-mail address: hinoko@is.icc.u-tokai.ac.jp (H. Inoko).

0888-7543/$ - see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.ygeno.2006.11.002

MHC to have an unusually high level of complexity with respect to sequence diversity and conservation and variability in the rates and types of gene duplications and indel activity [14]. The human MHC (also known as HLA) genomic region is a relatively small (3.6 Mb) portion of the whole genome, but it provides many challenges for the identification and analysis of loci predisposed to complex disorders [15]. The HLA is defined by a set of highly polymorphic (>1600 alleles) antigen-presenting HLA class I and II genes, embedded within well over 239 loci, collectively associated with more than 100 pathologies [16–18]. Remarkably, recent genome-wide scans have shown that for a majority of autoimmune diseases the MHC remains the first and foremost genetic component in pathogenesis [19]. Recently, a bacterial artificial chromosome (BAC) library (CHORI250 from http://bacpac.chori.org/) of the rhesus macaque genome was constructed and released as a tool for comparative genomics between humans and primates, and the 5.1-Mb genomic sequence of the *Mamu* MHC region was determined [20,21]. However, there is now a need to also have a cynomolgus macaque BAC library available for genomic sequencing and comparison between the rhesus and the cynomolgus macaque genomic sequences.

In this paper, we report on the construction of a cynomolgus macaque BAC library, the preparation of a BAC contig to map the *Mafa* MHC region, and the comparison of the MHC genomic features and nucleotide similarities between human, chimpanzee, and rhesus macaque.

Results and discussion

Construction of a cynomolgus macaque BAC library

The cynomolgus macaque BAC library was constructed in collaboration with the commercial company Advanced GenoTechs Co. (http://www.geno-gtac.co.jp/) using standard techniques. A total of 221,184 individual BAC clones were picked and arrayed into 576 different 384-well microtiter plates. We used a two-step PCR system as described in a previous report [22] to screen the BAC library for genes known to reside in the MHC region. All of the 160 BACs selected randomly had insert DNA and the average insert size was estimated by pulsed-field gel electrophoresis to be 82.6 kb, ranging from 30 to 185 kb (Fig. 1). On this basis, the library is expected to cover 1.8×10^{10} bp (sixfold) of the genome if the genome size of the cynomolgus macaque is 3.0×10^{9} bp as it is in the human genome.

Construction of a BAC contig of the Mafa MHC region

A total of 114 BAC clones were obtained by PCR screening of the library using 54 PCR primer sets to locate the corresponding HLA and *Mamu* regions between the *MHC-F* and the *KIFC1* gene (Tables 1 and 2). By fingerprinting, sequencing of PCR products, and comparing the BAC end sequences for all of the 114 BAC clones with genomic sequences corresponding to those of HLA and *Mamu*, 56 BACs were classified to have the *Mafa* class I region, 23 the *Mafa* class III region, and 35 the *Mafa* class II and extended class II regions

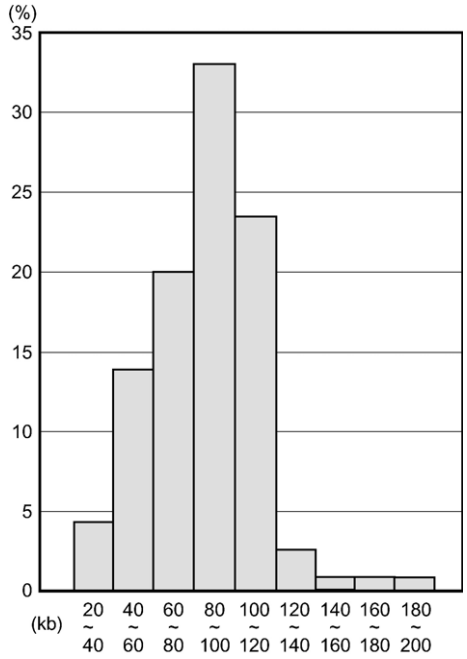


Fig. 1. Insert size distribution and percentage frequency of the randomly selected 160 BAC clones.

(Fig. 2). Although, at this stage, we could not construct a BAC contig to cover completely the *Mafa* class I gene-rich segments, such as *Mafa-A/-AG/-G/-F* and *Mafa-B* (see below), we successfully mapped 2.8 Mb of a non-*Mafa* class I gene-rich segment onto nine contigs (C1–C10) (Fig. 2). The segment lengths, genes, and associated contig were as follows: 180-kb segment between *C6orf12* and *TRIM26* (C1), 120-kb segment between STS_2170 and *RPP21* (C2), 230-kb segment between *TRIM39* and *ABCF1* (C3), 280-kb segment between *DHX16* and *DDR1* (C4), 250-kb segment between *DPCR1* and *POU5F1* (C5), 410-kb segment between *MIC* and *HSPA1L* (C6), 410-kb segment between *NEU1* and *NOTCH4* (C7), 130-kb segment around *Mafa-DRA* (C8), 420-kb segment between *Mafa-DQB* and *Mafa-DOA* (C9), and 390-kb segment between *Mafa-DPA1* and *KIFC1* (C10). Genomic size differences between orthologous BACs of *Mafa, Mamu,* and HLA genomic sequences were identified by correlating the results of pulsed-field gel electrophoresis (sizing) with gene-specific PCR screening and BAC-end sequence mapping of BACs. Namely, one *C6orf12*-positive BAC (1223K17) in C1 was 65 kb shorter than the *Mamu* ortholog, and two *NEU1*- and *BAT8*-positive BACs (2263O16 and 2016D1) in C7 were 30 kb shorter than the *Mamu* ortholog. This difference in size between HLA, *Mamu,* and *Mafa* for C7 seems due to the deletion of the *SLC44A4* gene in the *Mafa* genome. The other genomic features, such as contig length and gene content, in each contig were almost the same as in the HLA and *Mamu* orthologs [16–18,23]. On the other hand, there were eight gaps (Gaps 1–8), a 20-kb segment between *TRIM26* and STS_2170, 5 kb between *RPP21* and *TRIM39*, 40 kb around *PPP1R10*, 40 kb between *DDR1* and *DPCR1*, 30 kb around *HSPA1A*, 50 kb between the class III and the class II boundary, and 5 kb between *Mafa-DOA* and *Mafa-DPA1*, compared to the *HLA* region (Fig. 2). Gaps 3, 4, and 5 possibly

Table 1
Primer sets for construction of a BAC contig

Primer name	Forward primer	Reverse primer	Product size (bp)
Mamu-F	GCTCAGATCCTCCAAAGGCA	TTCGTGTTGCACATGGCAC	248
Mamu-A	TGGAGTGGCTCCGCAGATA	GCAGACCCTCATGCTGCAC	902
Mamu-AG	TGGAGTGGCTCGGCAGATA	CCCTCATGTTGCACATGACAG	897
Mamu-G	CAAGTGTGAGGCGGCCAAA	GCCCCTGGTGCTGCACATA	972
C6orf12	CACCTTGCTTTACTAGCAGAACACC	CAGTTGAACTCAGCAATTTATGTGTCC	781
TRIM31	TTAGTGGAATTGGGTCGAAGGAGT	TAATTTCCCTTGCTGATATGAGGGGT	403
TRIM10	TCAGTTCTCCTACAAATGGCAGAA	AAAGAAGGGAATGACCTTCCTAGTG	409
TRIM26	GACGAAGATGAGGAATCGTTGGG	GAGTCCTGGAATTCCAAAGAAGTGA	447
STS2170[a]	TGACCTCTACCTACAGATGCAAAC	GCTTTTACAGTTCCACATATGAGTGA	532
RPP21	GAGAAGTACAGGGAAACCAGG	TGTCTGCCAGTCCCCAACC	202
TRIM39	CTTTCCAAGTTCCCTTGGGAACAC	TTGACCCAGGATAGTCTTGCTTCT	464
1249Sp6[a]	ACGTGGGAACACTTTTGATCCAG	AGTTTGCTGTTTCAGCTTCGTGTAA	298
Mamu-E	GTGGAGTGGCTCCGCAGATAC	GTAGCCCCTCATGCTGCACAT	923
ABCF1	ACCACCTGGACCTCAACGC	CCCGCCAACTCACTCCTTG	1255
DHX16	CGTTGTCAGCATGGACGACC	GGAGGACTGGGCTCAACTCTC	804
IER3	TCACCTCCTAAACTTACGACCCAC	TACGCTTTGGACCTCAGCACTTT	574
STS1600[a]	CCCATCCGTATTAGGGTACTCCA	GAGATTTGTACCATTAGCTCCCCT	605
DDR1	CACCTCTAATAGAGGCAGTGAGA	GCTTAGAGAAGTTCTAGCTACAAGAC	515
DPCR1	TGCAGGGAATGAATCATTCTCCC	TACATCAGCCAATGTGATCACACC	571
STS1400[a]	CCTCCAGTTCATGTTCTAGGCTC	ATCTTCACATCCCAAACCTTGGTC	618
CDSN	TGGGCTACTTCACCAAAGATAACC	AAGGATGATTTTGCCACTGGATTG	374
POU5F1	GGCTGCCTCTCACTCGGTTC	CCAAGCTCCTGAAGCAGAAGA	799
Mamu-B	AGAGCAGCGGAGAGCCTACCT	GCTGCACATGGCACGTGTATC	903
Mamu-MIC3	CTTGGCGAATTCTGGTGGC	GCCGTCAGGAGAACAGGTACC	813
Mamu-MIC2	TGTGTTCCATGTAGCAGGCGA	CGGGCAGACTGCCTGAAGA	901
BAT1	TTTCCAGCCATTGCCATCC	CAAAGCGATCCTGCACATCAT	510
LTB	CAGGACGGCCTCTATTACCTCTAC	AATCAATTTCCAAACAGTCTCCTACA	453
BAT2	CCCAGAAAACCAGAGCTGCTAC	GACCTGATGACACCTGAGAATCC	848
LY6G5B	TTGTATGACGAAGACAGCACATGG	GGTGGTTTTGACAAGGTCTGAGC	303
LY6G6C	ATGTGGGTTTTCTCCAATCTGC	AGACGATTCTGTAAAGCCAGGG	302
C6orf27	CAGACCCAGCTGCTGGTAGAAGTG	GAAGCTGGGGTTGACGAAAGTCCT	492
HSPA1L	CCATCATCTATGGTCAGAATTGACAC	ATCCTGTTGTACAAGCAGATATGAAAC	406
NEU1	CAGTCCTAAAGGAAGTGATTTCCC	GCAGGAAAATGATTTCAATCCTGATG	923
BAT8	GGGAGAGATGGGATAATCAACCATA	AAGGGCTTGCTTTGAGGTAGAAAG	303
DOM3Z	CAAGACCTTTCCTACCATGAAGATGT	CCTCCCTCTAAAGCATTACTATTTGG	524
CYP21A2	TCCTCAAGATGCAGCCTTTCCAAGT	GCCAGCTCAGAATTAAGCCTCAATCC	551
TNXB	TAGCCCGTGTCGCAAATGCATTC	CTGCCCAGATGACTGCAATGATCAG	624
PRRT1	CTATAAGCACCGCCCCAGCT	CCAGGAACTTCCCAAATGTCC	305
NOTCH4	CCCAGGCTACGATGGACAGA	CCTCCATGAAAGCAGGGTTG	560
C6orf10	TCATTTGCTTAGGTCTGATCAATCTG	TGTTGCATCATCTGGATTTCTTTG	168
Mamu-DRA	AGAACACGTGATCATCCAGGC	TGGAGCGCTTTGTCATGATTT	224
Mamu-DRB	TGTCCCCCCAGCACGTTTC	TCACCTCGCCGCTGCACTGT	285
Mamu-DQB	GCATGTGCTACTTCACCAACG	GTAGTTGTGTCTGCACAC	209
Mamu-DOB	GTGATTCAGGCAAAGGCTGAC	CGCCCAGCCTGTAGTTGTG	238
TAP1	TACCGCCTTCGTTGTCAGTTATG	GAAGTAGAAAAGCCTCCTGTTAGAGA	495
Mamu-DMB	GGCCACCATCTGTGGAAGT	GCCCCAATGTGCTCTACCAC	259
Mamu-DMA	TCGCTGAAGTGTTCACGCTG	GCGGTCAATTTCGTGAGTCAC	248
BRD2	AAGCGGAAAGCAGATACTACCAC	GGTGCTTAATGATGTCATGGTAGTC	340
Mamu-DOA	ACCACATGGGCTCCTACGG	GCTCCACCAGGATGTCCAGAT	219
Mamu-DPA1	CCCCTGAGGTGACCGTGTT	TCTGCTGAGGGCACAAAGGT	217
RXRB	TCAGCAGGAGTAGGAGCCATCTTT	GCTCTAGACACTTAAGGCCAATGGA	831
RGL2	ATGGCAGTGTCTATAAGAGCATT	TTTCTTAAGGACACGACTGATGA	461
DAXX	GCTAGAGATTGAAGCTTTGCCCC	GCTGGGCAGGGTACATATCTTTTT	831
KIFC1	ATGATTTTTCCTTTGACCGGGTATT	TCCCGGACAGTCTCATTGTAGATCT	304

[a] From rhesus macaque genomic sequence.

contain the three expressed genes, *PPP1R10, VARS2L*, and *HSP1A1A*, respectively, based on the gene annotation of the *HLA* region [23]. Of the 111 expressed non-MHC genes in the HLA region, 108 genes (97%) are probably included in the contig on the basis of genomic structural similarities with the HLA and *Mamu* regions. From these findings, the BAC library covers most of the genes known to be in the *Mamu* MHC genomic region and it will therefore be a useful genomic tool for biomedical research and for sequence comparisons with the rhesus macaque MHC genomic region, although the insert size for the *Mafu* BAC library is shorter than the commercial *Mamu* BAC library (average 163 kb; CHORI250).

Table 2
Characteristics of BACs derived from PCR screening

Primer name	Positive BAC clone
Mamu-F	1232A11 (110)
Mamu-A	1183E17 (105), 2043L2 (105), 2084N19 (110), 2091F22 (120), 2174J9 (95)
Mamu-AG	1007P2 (80), 1184P5 (95), 1190L16 (85), 2091F22 (120), 2271F21 (105)
Mamu-G	2074J4 (80), 2091F22 (120), 2271F21 (105)
C6orf12	1113J2 (70), 1223K17 (40), 1282I3 (45)
TRIM31	1027J1 (65), 1113J2 (70), 2234F2 (125)
TRIM10	2234F2 (125)
TRIM26	2234F2 (125)
STS2170 [a]	1227J12 (115)
RPP21	1220K18 (185), 1227J12 (115)
TRIM39	1090H13 (85)
1249Sp6 [a]	1057H19 (100), 1090H13 (85), 1249M2 (95)
Mamu-E	1102K17 (80), 1187B1 (80), 1224O3 (70), 1249M2 (95)
ABCF1	1028K22 (50)
DHX16	1043I22 (95), 1096O13 (45)
IER3	1043I22 (95), 1281C8 (140)
STS1600 [a]	1191O10 (100), 1281C8 (140), 2073A10 (80), 2273B17 (105)
DDR1	2073A10 (80), 2273B17 (105)
DPCR1	1238J17 (60)
STS1400 [a]	1195K23 (110)
CDSN	2070L9 (95), 2267K22 (115)
POU5F1	1046N11 (65), 2070L9 (95)
Mamu-B	1018I12 (55), 1058D19 (30), 1219O4 (100), 1220N4 (85), 1223K1 (85), 1270N21 (85), 1288B13 (60), 2023G22 (60), 2062O24 (80), 2080I6 (85), 2093K16 (70), 2099E9 (45), 2116N18 (95), 2142J6 (75), 2180L22 (110), 2245K9 (100), 2277I22 (120)
Mamu-MIC3	1219O4 (100), 2116N18 (95), 2127H20 (100)
Mamu-MIC2	1103P19 (55), 1204D4 (90)
BAT1	1103P19 (55), 1252D11 (90), 2254D13 (110)
LTB	1040H16 (45), 1065J3 (55), 1252D11 (90), 2254D13 (110)
BAT2	1151G6 (30), 1201B3 (100), 1263J23 (85), 1273H6 (100)
LY6G5B	1151G6 (30), 1201B3 (100), 1263J23 (85), 1273H6 (100)
LY6G6C	1047P10 (70), 2073C16 (75)
C6orf27	1047P10 (70), 2073C16 (75)
HSPA1L	1047P10 (70), 2073C16 (75)
NEU1	2016D1 (60), 2263O16 (85)
BAT8	2016D1 (60), 2054B12 (165), 2232G2 (85), 2263O16 (85)
DOM3Z	1187L18 (115), 1208L6 (90), 2054B12 (165), 2232G2 (85)
CYP21A2	1030L11 (95), 1187L18 (115), 1208L6 (90), 1245H15 (105), 1277M12 (110), 2054B12 (165)
TNXB	1030L11 (95), 1185J20 (55), 1245H15 (105), 1277M12 (110)
PRRT1	2029L21 (80)
NOTCH4	1217P4 (75), 1226L20 (65)
C6orf10	2106P19 (80), 2147G1 (90)
Mamu-DRA	1192B9 (45), 2147G1 (90), 2259G20 (75)
Mamu-DRB	1096O7 (50), 1111F2 (95), 1276M5 (30), 1282J22 (85), 2099L11 (90), 2157D19 (100), 2264K20 (95)
Mamu-DQB	1110L24 (75), 1174M20 (100), 2268J11 (130)
Mamu-DOB	1177H3 (110), 1186E15 (100)
TAP1	1177H3 (110), 1045H7 (80)
Mamu-DMB	1051O21 (35), 1094F2 (50), 1114B15 (40), 1264E5 (65), 2040I21 (80)
Mamu-DMA	1051O21 (35), 2040I21 (80), 2099L24 (75)
BRD2	2040I21 (80), 2099L24 (75)
Mamu-DOA	1040G6 (80), 2099L24 (75)
Mamu-DPA1	1005F22 (45), 1267B12 (50), 2127M21 (60)
RXRB	1153N6 (80), 1196C1 (105), 2224E7 (100)
RGL2	2093M3 (70)
DAXX	1055N6 (35), 2093M3 (70), 2127M8 (60)
KIFC1	1015B20 (70), 2117I24 (75)

Parentheses indicate estimated insert size based on pulsed-field gel electrophoresis.
[a] From rhesus macaque genomic sequence.

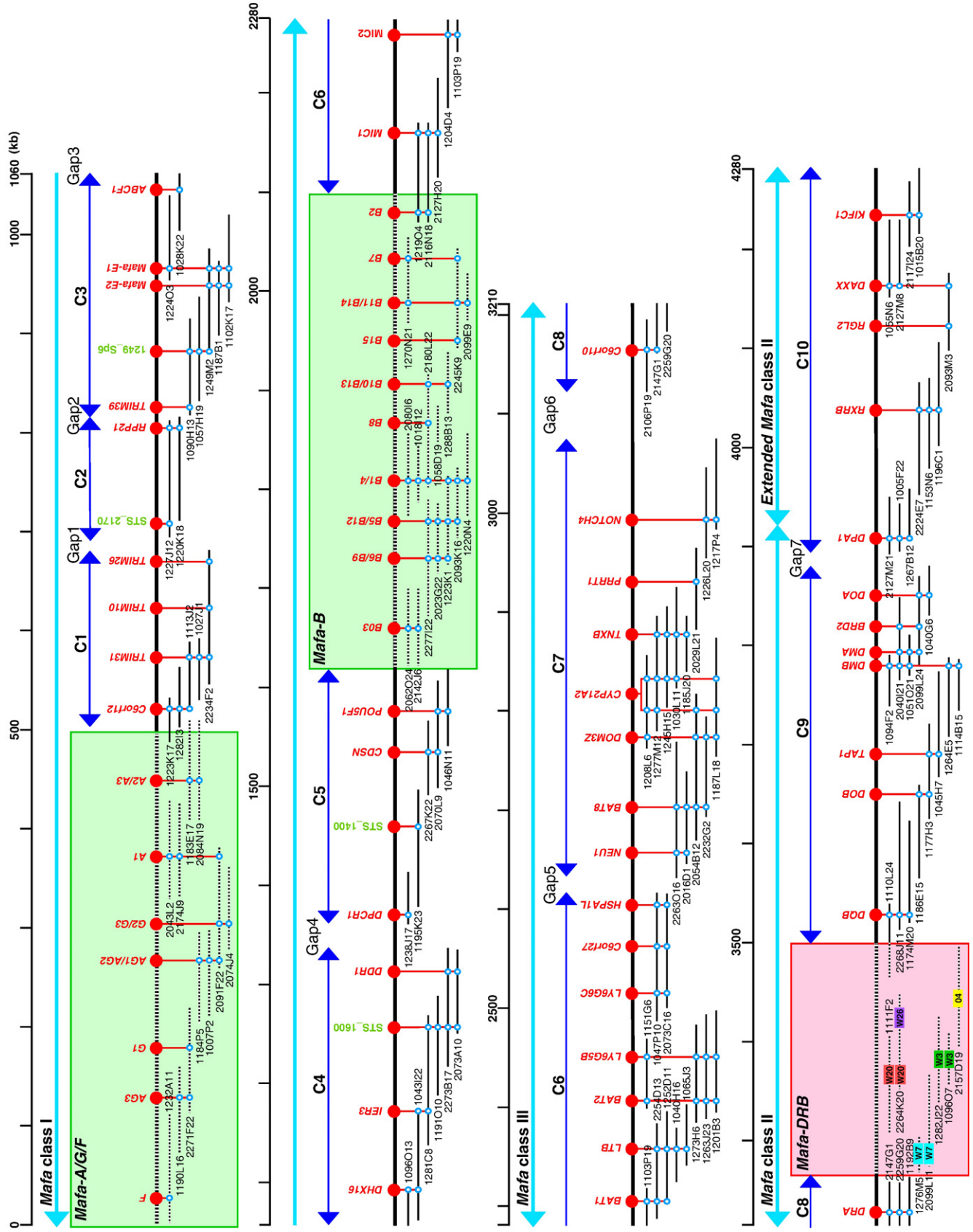
Characteristics of the Mafa class II region

Seven *Mafa-DRB*-positive BACs were identified (Fig. 2 and Table 2) with the *Mafa-DRB* sequences found in the three BAC sets, 1096O7 and 1282J22, 2099L11 and 1276M5, and 2157D19, perfectly matched to *Mamu-DRB*W0307* (GenBank Accession No. Z26173), *Mamu-DRB*W0703* (U57940), and *Mamu-DRB1*0415* (AY487264), respectively. In addition, BAC 1111F2 had 97% nucleotide identity with *Mamu-DRB*W2003* (AF163283) and BAC 2264K20 had 97 and 100% identity with *Mamu-DRB*W2003* and *-DRB*W2603* (AF163286), respectively. From these findings, one of the *Mafa-DRB* haplotypes is thought to correspond to a previously reported *Mamu-DRB* haplotype, *DRB*W3–DRB*W7–DRB1*04* [24], whereas the other may be a diverged *Mafa*-specific haplotype, *DRB*20–DRB*26*, which has not been previously observed as a *Mamu-DRB* haplotype [25].

Eight other MHC class II loci (*Mafa-DRA, -DQB, -DOB, -DMB, -DMA, -DOA, -DPA1*, and *-DPB1*) were covered by more than one BAC (Fig. 2 and Table 2). Their gene order was the same as that found within the HLA and *Mamu* orthologous regions [20,23]. Of the nucleotide sequences of exon 2 on *Mafa-DRA, -DQB*, and *-DPB1* loci, two (1192B9 and 2147G1) and one (2259G20) perfectly matched to *Mamu-DRA*01027* (AJ586880) and *-DRA*1041* (AJ586882), respectively [26]; two (1110L24 and 2268J11) and one (1174M20) perfectly matched to *Mafa-DQB*0614* (AJ308062) and *-DQB*1503* (AJ308063), respectively [27]; and one (2127M21) and two (1005F22 and 1267B12) perfectly matched to *Mafa-DPB1*03* (AB235858) and *-DPB1*28* (AB235883), respectively (Fig. 2 and Table 2) [28].

Characteristics of the Mafa class I region

A total of 11 and 17 positive BACs were connected to the *Mafa-A/-AG/-G/-F* and *Mafa-B* regions of the MHC, respectively, and partial sequences of their PCR products were determined for phylogenetic analysis (Figs. 2 and 3 and Supplementary Fig. 1). As a result, 3, 3, 3, 1, and 15 distinct sequences (AB263221 to AB263241 and AB263246 to AB263249) were identified at 2, 2, 2, 1, and 10 loci, respectively, that corresponded to the *Mafa-A, -AG, -G, -F*, and *-B* genes, respectively (Figs. 3A–3E, Supplementary Fig. 1, and Table 3). The nucleotide similarities ranged from 95 to 99% compared with the *Mamu* orthologous sequences (Table 3). From our previous analyses of the *Mamu* class I region, we identified the presence of 2 *Mamu-A*, 5 *-AG*, 6 *-G*, 1 *–F*, and 19 *-B* as a single haplotypic sequence (AB128049) [21]. However, the cynomolgus


Mafa class I
Mafa-A/G/F
Mafa-B
Mafa class III
Extended Mafa class II
Mafa class II
Mafa-DRB
Gap1
Gap2
Gap3
Gap4
Gap5
Gap6
Gap7
C1
C2
C3
C4
C5
C6
C7
C8
C9
C10
0
500
1000
1060 (kb)
1500
2000
2280
2500
3000
3210
3500
4000
4280
AB/CF1
Mafa-E1
Mafa-E2
1249_Sp6
TRIM39
RPP21
STS_2170
TRIM26
TRIM10
TRIM31
C6orf12
A2/A3
A1
G2/G3
AG1/AG2
G1
AG3
F
MIC2
MIC1
B2
B7
B11/B14
B15
B10/B13
B8
B1/4
B5/B12
B6/B9
B03
POU5F1
CDSN
STS_1400
DPCR1
DDR1
STS_1600
IER3
DHX16
C6orf10
NOTCH4
PRRT1
TNXB
CYP21A2
DOM3Z
BAT8
NEU1
HSPA1L
C6orf27
LY6G6C
LY6G5B
BAT2
LTB
BAT1
DRA
DQB
TAP1
DOB
DMA
DMB
BRD2
DOA
DPA1
RXRB
RGL2
DAXX
KIFC1

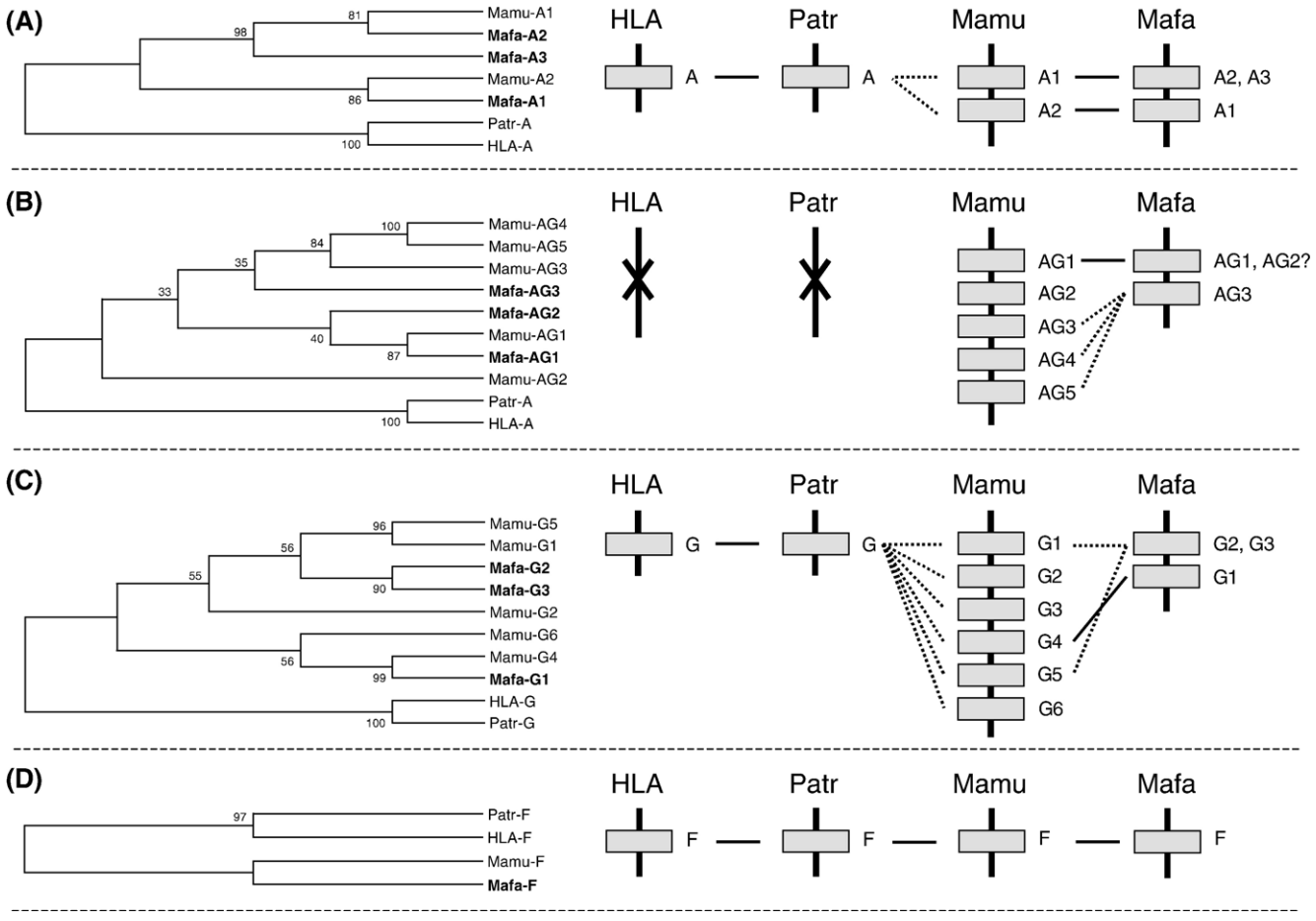


Fig. 3. Phylogenetic analysis and gene organization of the primate MHC class I segments. (A–F) The phylogenetic trees and the inferred organization of the *Hosa, Patr, Mamu*, and *Mafa* genes corresponding to the gene families of *MHC-A* (A), *MHC-AG* (B), *MHC-G* (C), *MHC-F* (D), *MHC-B* (E-1), *MIC* (E-2), and *MHC-E* (F). Gray, white, and black labeled boxes indicate the MHC class I genes, MIC genes, and non-MHC genes, respectively. Solid and dotted lines indicate orthologous correlations and possible orthologous correlations, respectively, suggested by the phylogenetic trees. Gene organization of *Hosa, Patr*, and *Mamu* MHC segments reflect actual gene order [21,31,32], but gene order of the *Mafa* MHC segments is unknown.

macaque that was used to construct the present BAC library is heterozygous for the MHC alleles (see Materials and methods for further details). Therefore, by comparing the *Mamu* class I locus numbers per diploid (2*n*) with the *Mafa* locus numbers, *Mafa-A* and *Mafa-F* were found to have the same number as the *Mamu* orthologs, whereas the other loci appear to have fewer loci than the corresponding *Mamu* orthologs (Table 3). The data, nevertheless, confirm that the *Mafa-A/-AG/-G/-F* and *Mafa-B* are also multiply duplicated as in the *Mamu*, but the degree of gene duplication is less in *Mafa* than in *Mamu*.

Four *Mafa-E*-positive BACs were identified and are shown in Fig. 2 and Table 2. Although the genomic sequences of HLA and *Mamu* contain one conserved *MHC-E* locus per haploid, it was reported previously that the rhesus macaque may have up to two highly polymorphic *Mamu-E* loci [29]. Two (1249M2 and 1187B1) of three *Mafa-E*-positive BACs revealed the existence of two *Mafa-E* loci. The loci in 1249M2 were tentatively named *Mafa-E*01* and **02* and those in 1187B1 were named *Mafa-E*03* and **04* (AB263242 to AB263245) (Supplementary Fig. 1). Therefore, *Mafa-E*01/*04* and *Mafa-E*02/*03* are thought to be allelic correlations on the basis of phylogenetic analysis (Fig. 3F).

Five MHC class I-like gene (*MIC*)-positive BACs were identified at the genomic boundary of the class I and class III regions (Fig. 2, Table 2). Three BACs (1219O4, 2116N18, and 2127H20) and two BACs (1204D4 and 1103P19) (AB263250, AB263251) have 99% nucleotide identities with *Mamu-MIC1*01* and *Mamu-MIC2*02* sequences (AJ242438 and AJ242440), respectively [30]. *Mafa-MIC1*01* and *Mafa-MIC2*02* are thought to be *Mamu*-orthologous loci on the basis of phylogenetic analysis (Fig. 3E).

We have so far determined the complete genomic sequences and the comparative organization of the human, chimpanzee, and rhesus macaque MHC class I regions [21,31,32]. From the

Fig. 2. The BAC contig map of the 4.3-Mb *Mafa* MHC region. The blue horizontal double arrows labeled C1 to C10 represent the genomic regions of the overlapping BAC clones. Red vertical lines and circles indicate the positions of the gene-specific PCR markers. The shaded green and pink boxes highlight the regions of the *Mafa* class I and class II gene duplications, respectively. Dotted horizontal lines show the estimated BAC locations.

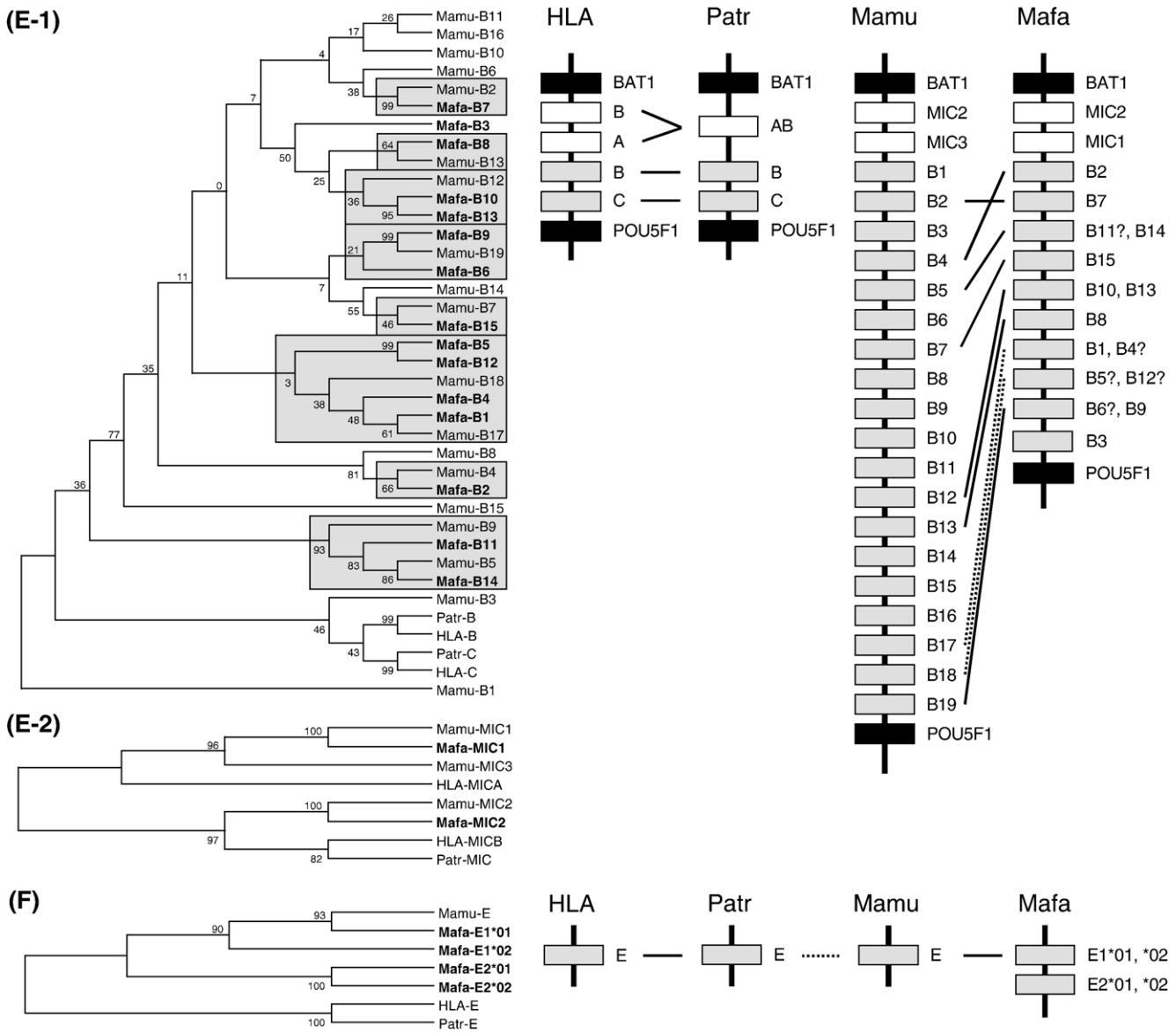


Fig. 3 (*continued*).

comparative data, the genomic organization of the non-MHC gene-rich segments located between *MHC-C* and *MHC-A* are well conserved among the three species, but the organization of the MHC class I gene-rich segments, such as the *Mafa-A/-AG/-G/-F* and *Mafa-B* regions, are generally species-specific structures exhibiting considerable variability between species. In this connection, BAC end sequencing of the BAC clones containing the *Mafa-A/-AG/-G/-F* and *Mafa-B* regions showed <96% nucleotide identity with the *Mamu* sequences (data not shown). Therefore, the basic structure of the MHC regions is largely conserved among various mammalian species, but the *Mafa-A/-AG/-G/-F* and *Mafa-B* regions are often remodeled as a response to environmental pathogens by the "birth and death" of MHC and MHC-linked genes [31–34].

Table 3
Comparison of *Mafa* and *Mamu* class I locus numbers

Mafa				*Mamu*
Class I family	Seq. No. from PCR	Locus No. from phylogenetic tree	Nt similarity with *Mamu* seq.	Locus No. from genomic segment
MHC-A/-AG/-G/-F				
A	3	2	97–99%	2
AG	3	2	96–99%	5
G	3	2	96–99%	6
F	1	1	98%	1
MHC-B	15	10	95–99%	19
MHC-E	4	2	97–99%	1

Diversity analysis of MHC genomic sequences of cynomolgus and other primates

To investigate the nucleotide diversity among human (*Hosa*), chimpanzee (*Patr*), *Mamu*, and *Mafa*, we determined the nucleotide sequences for 34 *Mafa* non-MHC genes to

Table 4
Genomic diversity among four primate species on 34 non-MHC gene segments

Gene symbol	Determined nt. length in *Mafa* (bp)	Accession No.	*Mafa* vs *Mamu*			*Mafa* vs *Hosa*			*Mafa* vs *Patr*			*Hosa* vs *Hosa*			*Hosa* vs *Patr*		
			Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity
C6orf12	696	AB263258	688	696	1.15%	656	696	5.75%	656	696	5.75%	703	704	0.14%	702	704	0.28%
TRIM31	352	AB263284	339	352	3.69%	337	352	4.26%	336	351	4.27%	352	355	0.85%	348	355	1.97%
TRIM10	341	AB263282	336	341	1.47%	326	341	4.40%	324	341	4.99%	340	341	0.29%	338	341	0.88%
TRIM26	399	AB263283	396	399	0.75%	385	399	3.51%	387	398	2.76%	398	398	0.00%	394	398	1.01%
RPP21	153	AB263278	153	153	0.00%	145	149	2.68%	145	149	2.68%	154	155	0.65%	153	155	1.29%
TRIM39	416	AB263285	414	416	0.48%	392	416	5.77%	391	411	4.87%	415	415	0.00%	412	415	0.72%
ABCF1	1127	AB263252	1124	1127	0.27%	1081	1127	4.08%	1082	1125	3.82%	1134	1135	0.09%	1128	1136	0.70%
DHX16	669	AB263265	667	669	0.30%	636	669	4.93%	636	669	4.93%	681	681	0.00%	677	681	0.59%
IER3	488	AB263270	488	488	0.00%	468	488	4.10%	469	488	3.89%	504	504	0.00%	485	488	0.61%
DDR1	465	AB263264	465	465	0.00%	449	465	3.44%	450	465	3.23%	464	465	0.22%	460	465	1.08%
DPCR1	516	AB263267	514	516	0.39%	488	516	5.43%	487	516	5.62%	515	516	0.19%	509	516	1.36%
CDSN	331	AB263261	329	331	0.60%	320	331	3.32%	319	331	3.63%	331	331	0.00%	330	331	0.30%
POU5F1	764	AB263276	759	764	0.65%	708	740	4.32%	711	741	4.05%	732	733	0.14%	726	735	1.22%
BAT1	431	AB263253	430	431	0.23%	416	431	3.48%	415	431	3.71%	433	433	0.00%	432	433	0.23%
LTB	383	AB263272	382	383	0.26%	370	383	3.39%	369	383	3.66%	383	383	0.00%	382	383	0.26%
BAT2	771	AB263254	768	771	0.39%	736	767	4.04%	738	768	3.91%	774	775	0.13%	772	775	0.39%
LY6G5B	462	AB263268	462	462	0.00%	445	462	3.68%	442	462	4.33%	467	467	0.00%	464	467	0.64%
LY6G6C	574	AB263273	570	574	0.70%	520	556	6.47%	NA	NA	NA	576	576	0.00%	NA	NA	NA
C6orf27	442	AB263259	440	442	0.45%	418	442	5.43%	413	441	6.35%	440	441	0.23%	432	441	2.04%
HSPA1L	323	AB263269	323	323	0.00%	315	323	2.48%	317	323	1.86%	323	323	0.00%	320	323	0.93%
NEU1	813	AB263274	812	813	0.12%	781	813	3.94%	782	813	3.81%	814	815	0.12%	811	815	0.49%
BAT8	256	AB263255	256	256	0.00%	243	256	5.08%	238	251	5.18%	254	254	0.00%	247	251	1.59%
DOM3Z	568	AB263266	567	568	0.18%	540	568	4.93%	540	567	4.76%	572	572	0.00%	566	572	1.05%
CYP21A2	488	AB263262	486	488	0.41%	460	488	5.74%	459	488	5.94%	490	491	0.20%	481	491	2.04%
TNXB	446	AB263281	441	446	1.12%	432	446	3.14%	432	446	3.14%	445	446	0.22%	440	446	1.35%
PRRT1	486	AB263260	484	486	0.41%	461	483	4.55%	NA	NA	NA	486	486	0.00%	NA	NA	NA
NOTCH4	510	AB263275	509	510	0.20%	490	510	3.92%	490	510	3.92%	508	510	0.39%	507	510	0.59%
C6orf10	125	AB263257	125	125	0.00%	124	125	0.80%	125	125	0.00%	123	125	1.60%	124	125	0.80%
TAP1	563	AB263280	550	560	1.79%	529	564	6.21%	522	556	6.12%	564	564	0.00%	545	557	2.15%
BRD2	271	AB263256	271	271	0.00%	269	271	0.74%	NA	NA	NA	271	271	0.00%	NA	NA	NA
RXRB	755	AB263279	754	755	0.13%	742	755	1.72%	735	755	2.65%	756	756	0.00%	749	756	0.93%
RGL2	489	AB263277	489	489	0.00%	467	489	4.50%	466	487	4.31%	490	490	0.00%	487	490	0.61%
DAXX	406	AB263263	406	406	0.00%	397	406	2.22%	397	406	2.22%	406	406	0.00%	406	406	0.00%
KIFC1	243	AB263271	243	243	0.00%	241	243	0.82%	242	243	0.41%	243	243	0.00%	242	243	0.41%
Total or average	16,522		16,440	16,519	0.48%	15,787	16,470	4.15%	14,515	15,136	4.10%	16,541	16,560	0.11%	15,069	15,204	0.89%

NA, not analyzed.

Table 5
Genomic diversity among four primate species on coding region of 25 non-MHC gene segments

Gene symbol	CDR nt. length in *Mafa* (bp)	*Mafa* vs *Mamu*			*Mafa* vs *Hosa*			*Mafa* vs *Patr*			*Hosa* vs *Hosa*			*Hosa* vs *Patr*		
		Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity	Matched length	Aln. length	Nt. diversity
TRIM10	341	336	341	1.47%	326	341	4.40%	324	341	4.99%	340	341	0.29%	338	341	0.88%
TRIM26	361	359	361	0.55%	352	361	2.49%	352	361	2.49%	361	361	0.00%	359	361	0.55%
ABCF1	444	443	444	0.23%	436	444	1.80%	437	444	1.58%	444	444	0.00%	442	444	0.45%
DHX16	81	81	81	0.00%	81	81	0.00%	81	81	0.00%	81	81	0.00%	81	81	0.00%
IER3	18	18	18	0.00%	18	18	0.00%	18	18	0.00%	18	18	0.00%	18	18	0.00%
DPCR1	465	463	465	0.43%	440	465	5.38%	440	465	5.38%	465	465	0.00%	459	465	1.29%
CDSN	331	329	331	0.60%	320	331	3.32%	319	331	3.63%	331	331	0.00%	330	331	0.30%
POU5F1	224	224	224	0.00%	222	224	0.89%	224	224	0.00%	224	224	0.00%	224	224	0.00%
BAT1	206	206	206	0.00%	205	206	0.49%	205	206	0.49%	206	206	0.00%	206	206	0.00%
LTB	327	326	327	0.31%	316	327	3.36%	315	327	3.67%	327	327	0.00%	326	327	0.31%
BAT2	237	237	237	0.00%	234	237	1.27%	234	237	1.27%	237	237	0.00%	237	237	0.00%
LY6G5B	160	159	160	0.62%	155	160	3.13%	153	160	4.38%	160	160	0.00%	158	160	1.25%
LY6G6C	172	171	172	0.58%	168	172	2.33%	NA	NA	NA	172	172	0.00%	NA	NA	NA
C6orf27	363	362	363	0.28%	347	363	4.41%	342	363	5.79%	362	363	0.28%	358	363	1.38%
HSPA1L	323	323	323	0.00%	315	323	2.48%	317	323	1.86%	323	323	0.00%	320	323	0.93%
NEU1	450	450	450	0.00%	436	450	3.11%	432	450	4.00%	450	450	0.00%	450	450	0.00%
DOM3Z	345	345	345	0.00%	331	345	4.06%	332	345	3.77%	345	345	0.00%	342	345	0.87%
TNXB	446	441	446	1.12%	432	446	3.14%	432	446	3.14%	445	446	0.22%	440	446	1.35%
NOTCH4	256	256	256	0.00%	251	256	1.95%	251	256	1.95%	255	256	0.39%	255	256	0.39%
TAP1	284	280	284	1.41%	271	284	4.58%	268	284	5.63%	284	284	0.00%	275	284	3.17%
BRD2	271	271	271	0.00%	269	271	0.74%	NA	NA	NA	271	271	0.00%	NA	NA	NA
RXRB	233	233	233	0.00%	233	233	0.00%	231	233	0.86%	223	223	0.00%	221	223	0.90%
RGL2	16	16	16	0.00%	16	16	0.00%	16	16	0.00%	16	16	0.00%	16	16	0.00%
DAXX	246	246	246	0.00%	244	246	0.81%	244	246	0.81%	246	246	0.00%	246	246	0.00%
KIFC1	243	243	243	0.00%	241	243	0.82%	242	243	0.41%	243	243	0.00%	242	243	0.41%
Total or average	6843	6818	6843	0.37%	6659	6843	2.69%	6209	6400	2.98%	6829	6833	0.06%	6343	6390	0.74%

NA, not analyzed.

obtain a total of 16,522 bp of sequence (AB263252 to AB263285) (Table 4) using the PCR primer sets in Table 1. The average diversities of the regions between the aligned sequences of *Mafa* and *Mamu* (16,519 bp), *Mafa* and *Hosa* (16,470 bp), and *Mafa* and *Patr* (15,136 bp) were 0.48, 4.15, and 4.10%, respectively (Table 4). In comparison, the nucleotide diversity between the two aligned (16,560 bp) *Hosa* haplotypes, PGF and COX, was 0.11% and that between the aligned *Hosa* and *Patr* sequences (15,204 bp) was 0.89% (Table 4). The difference between the human haplotypes was four times less than the diversity between the two macaque species, suggesting that the human haplotypes diverged from each about 500,000 years ago, that is, if the rate of divergence was almost the same between them as between the rhesus and cynomolgus, which split 1.8–2.0 Mya [13]. The average diversities for the coding regions aligned between *Mafa* and *Mamu* (6843 bp), *Mafa* and *Hosa* (6843 bp), and *Mafa* and *Patr* (6400 bp) were 0.37, 2.69, and 2.98%, respectively (Table 5). Diversity between *Mafa* and *Hosa* was slightly higher than the recently reported diversity of 2.21% [35]. The diversities between *Mafa* and *Hosa* were over 4% for the coding regions of *TRIM10, DPCR1, C6orf27, DOM3Z*, and *TAP1*. In addition, the diversities between *Mafa* and *Mamu* were over 1% for the coding regions of *TRIM10, TNXB*, and *TAP1*. These genes contributed to raising the diversity level, suggesting that there may be functional differences for them among primates (Table 5).

Conclusion

To undertake comparative genomic analysis of the cynomolgus macaque MHC region, we constructed the *Mafa* BAC library, developed a PCR screening system to construct a BAC contig of the MHC region, and performed a diversity sequence analysis between human and some other nonhuman primates. We have demonstrated here that the library is a useful tool for cynomolgus macaque genome analysis and that we need not only the rhesus macaque genomic sequence data, but also the cynomolgus macaque genomic sequence data for the development of animal models for medical research and for future comparative genomic studies within and between primate species.

Materials and methods

Sample

We used the genomic DNA of a MHC heterozygous cynomolgus macaque donated by Shiga University of Medical Science for the construction of the *Mafa* BAC library. This individual has *Mafa-DPB1*08/*28, Mafa-DQB1*05/*12, Mafa-DRA*02/*15*, and different PCR product sizes for three microsatellite markers (MS_MIC, MS_B2, and MS_A1) that are located within the class I region ([28], Shiina et al., unpublished). Genomic DNA was prepared from spleen tissue.

Construction of a BAC library and screening

A BAC library for *Mafa* was constructed from a female kidney according to the modified method of Osoegawa [36]. The kidney was crushed in a mortar in the presence of liquid nitrogen to prepare agarose plugs. The genomic DNA was extracted and purified with protease K in 0.4 M EDTA and 1% *N*-lauryl sarcosine at 50°C and digested partially with restriction enzyme, *Mbo*I. DNA of more than 100 kb was fractionated by pulsed-field gel electrophoresis and ligated into the *Bam*HI cloning site of pBACe3.6 vector. The recombinant DNA was transfected into *Escherichia coli* DH10B competent cells (Invitrogen) and plated on LB plates containing 5% sucrose and chloramphenicol (20 μg/ml). The colonies grown on the plates were picked up into 384-well plates with 80 μl LB containing 10% glycerol and Cm with Flexys Colony and incubated overnight at 37°C. Two-dimensional PCR screening was achieved using DNA pools and screening plates [22] (Advanced GenoTechs Co., http://www.geno-gtac.co.jp/).

Isolation and assembly of contiguous BACs

A total of 221,184 individual BAC clones taken from the library were individually screened with 54 locus-specific PCR markers, including 34 conserved genes located in the *HLA* region between *KIFC1* and *C6orf12*, 16 rhesus macaque MHC genes, 3 rhesus macaque intergenic regions, and 1 sequence-tagged site made from the BAC end sequence (Table 1). The PCR markers were designed to yield 168- to 1255-bp amplicons after 30 cycles of amplification using the following conditions: 30 s at 96°C, 30 s at 60°C, and 30 s at 72°C using Ampli*Taq* Gold DNA polymerase (Applied Biosystems). BACs were assembled by PCR-based mapping with 54 locus-specific primer sets, fingerprinting by *Eco*RI digestion, BAC-end sequencing, and PCR-based sequencing of portions of all markers (Fig. 1). DNA sequencing was performed by the cycle sequencing method using Ampli*Taq* DNA polymerase FS and fluorescently labeled BigDye terminators in a GeneAmp PCR system (Applied Biosystems). An ABI 3100 DNA analyzer was used for automated fluorescence sequencing (Applied Biosystems).

Phylogenetic analysis of primate MHC class I and MIC sequences

Multiple sequence alignments were created using the ClustalW sequence alignment program at DDBJ (http://www.ddbj.nig.ac.jp/). The phylogenetic tree was constructed using the neighbor-joining method on intron 3 for *MHC-A, -AG, -G, -B, -E*, and *MIC* and exon 4 for *MHC-F* using the Molecular Evolution Genetics Analysis (MEGA3.1) software [37]. We used the following primate MHC class I and *MIC* sequences (DNA accession numbers) for phylogenetic analysis: *HLA-A*0312* (AY310505), *Patr-A* (BA000041), and *Mamu-A1* and *-A2* (AB128833 and AB128834) for *MHC-A; HLA-A*0312* (AY310505), *Patr-A* (BA000041), and *Mamu-AGs* (AB128835–AB128839) for *MHC-AG; HLA-G*010101* (AY645774), *Patr-G* (AB210189), and *Mamu-Gs* (AB128049) for *MHC-G; HLA-F* (NM_018950), *Patr-F* (BA000041), and *Mamu-F* (AB128841) for *MHC-F; HLA-B*0702* (D83956), *HLA-Cw*0701* (D83957), *Patr-B* (AB054536), *Patr-C* (BA000041), and *Mamu-Bs* (AB128842–AB128860) for *MHC-B; HLA-E*0101* (AB210186), *Patr-E* (AB210186), and *Mamu-E* (AB128840) for *MHC-E*; and *Hosa-MICA* (D84394), *Hosa-MICB* (AB088107), *Patr-MIC* (AB210174), *Mamu-MIC1* (AJ242438), and *Mamu-MIC2* and *Mamu-MIC3* (AB128049) for the *MIC* genes.

Diversity analysis with other primates

To investigate the evolutionary differences among primates, the partial sequences derived from 54 PCR sequences were compared with human DNA sequences ([23], http://www.sanger.ac.uk/HGP/Chr6/MHC/consensus.fasta), HLA haplotypic sequences obtained from the cell lines COX and PGF ([38], http://www.sanger.ac.uk/HGP/Chr6/MHC/sequence.shtml), chimpanzee DNA (AB054536) [32], and rhesus macaque DNA (AB128049, AC148659–AC148717) [20,21]. The sequence identities were calculated by GENETYX-WIN version 5.1 (SDC).

Acknowledgment

This work was supported by a KAKENHI (Grant-in-Aid for Scientific Research) on Priority Area “Comparative Genomics” from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ygeno.2006.11.002.

References

[1] M.E. Steiper, N.M. Young, T.Y. Sukarna, Genomic data support the hominoid slowdown and an Early Oligocene estimate for the hominoid–cercopithecoid divergence, Proc. Natl. Acad. Sci. USA 101 (2004) 17021–17026.

[2] T. Kawai, et al., Thrombophilia associated with anti-CD154 monoclonal antibody treatment and its prophylaxis in nonhuman primates, Transplantation 77 (2004) 460–462.

[3] H.J. Schuurman, et al., Incidence of hyperacute rejection in pig-to-primate transplantation using organs from hDAF-transgenic donors, Transplantation 73 (2002) 1146–1151.

[4] P. Putkonen, et al., Experimental infection of cynomolgus monkeys (*Macaca fascicularis*) with HIV-2, J. Acquired Immune Defic. Syndr. 2 (1989) 366–373.

[5] H. Feldmann, S. Jones, H.D. Klenk, H.J. Schnittler, Ebola virus: from discovery to vaccine, Nat. Rev. Immunol. 3 (2003) 677–685.

[6] B. Zuber, et al., An in vivo model for HIV resistance development, AIDS Res. Hum. Retroviruses 17 (2001) 631–635.

[7] P. Putkonen, et al., Experimental infection of cynomolgus monkeys (Macaca fascicularis) with simian immunodeficiency virus (SIVsm), J. Acquired Immune Defic. Syndr. 2 (1989) 366–373.

[8] M.G. O'Sullivan, et al., Experimental infection of cynomolgus monkeys with simian parvovirus, J. Virol. 71 (1997) 4517–4521.

[9] Y.K. Leong, A. Awang, Experimental group A rotaviral infection in cynomolgus monkeys raised on formula diet, Microbiol. Immunol. 34 (1990) 153–162.

[10] D.N. Lees, A. Baskerville, L.M. Cropper, D.W. Brown, Herpesvirus simiae (B virus) antibody response and virus shedding in experimental primary infection of cynomolgus monkeys, Lab. Anim. Sci. 41 (1991) 360–364.

[11] J.V. Lawler, et al., Cynomolgus macaque as an animal model for severe acute respiratory syndrome, PLoS Med. 3 (2006) 677–686.

[12] W.H. Chen, et al., Analysis of 10,000 ESTs from lymphocytes of the cynomolgus monkey to improve our understanding of its immune system, BMC Genomics 7 (2006) 82.

[13] K. Hayasaka, K. Fujii, S. Horai, Molecular phylogeny of macaques: implications of nucleotide sequences from an 896-base pair region of mitochondrial DNA, Mol. Biol. Evol. 13 (1996) 1044–1053.

[14] J.K. Kulski, T. Shiina, T. Anzai, S. Kohara, H. Inoko, Comparative genomic analysis of the MHC: the evolution of class I duplication blocks, diversity and complexity from shark to man, Immunol. Rev. 190 (2002) 95–122.

[15] J.N. Hirschhorn, M.J. Daly, Genome-wide association studies for common diseases and complex traits, Nat. Rev. Genet. 6 (2005) 95–108.

[16] J.A. Hansen, B. Dupont, HLA 2004: immunobiology of the human MHC, Proceedings of the 13th International Histocompatibility Workshop and Conference, IHWG Press, Seattle, 2004.

[17] T. Shiina, H. Inoko, J.K. Kulski, An update of the HLA genomic region, locus information and disease associations, Tissue Antigens 64 (2004) 631–649.

[18] R. Horton, et al., Gene map of the extended human MHC, Nat. Rev. Genet. 5 (2004) 889–899.

[19] S. Onengut-Gumuscu, P. Concannon, Mapping genes for autoimmunity in humans: type 1 diabetes as a model, Immunol. Rev. 190 (2002) 182–194.

[20] R. Daza-Vamenta, G. Glusman, L. Rowen, B. Guthrie, D.E. Geraghty, Genetic divergence of the rhesus macaque major histocompatibility complex, Genome Res. 14 (2004) 1501–1515.

[21] T. Shiina, et al., Rapid evolution of MHC class I genes in primates generates new disease alleles in man via hitchhiking diversity, Genetics 173 (2006) 1555–1570.

[22] C.G. Kim, A. Fujiyama, N. Saitou, Construction of a gorilla fosmid library and its PCR screening system, Genomics 82 (2003) 571–574.

[23] The MHC Sequencing Consortium, Complete sequence and gene map of a human major histocompatibility complex, Nature 401 (1999) 921–923.

[24] G.G. Doxiadis, N. Otting, N.G. de Groot, R.E. Bontrop, Differential evolutionary MHC class II strategies in humans and rhesus macaques: relevance for biomedical studies, Immunol. Rev. 183 (2001) 76–85.

[25] N. Leuchte, et al., *MhcDRB*—Sequences from cynomolgus macaques (*Macaca fascicularis*) of different origin, Tissue Antigens 63 (2004) 529–537.

[26] N. de Groot, et al., Genetic makeup of the DR region in rhesus macaques: gene content, transcripts, and pseudogenes, J. Immunol. 172 (2004) 6152–6157.

[27] N. Otting, N.G. de Groot, G.G. Doxiadis, R.E. Bontrop, Extensive MHC-DQB variation in humans and non-human primate species, Immunogenetics 54 (2002) 230–239.

[28] K. Sano, et al., Novel cynomolgus macaque *MHC-DPB1* polymorphisms in three South-East Asian populations, Tissue Antigens 67 (2006) 297–306.

[29] M. Alvarez, et al., 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 (1997) 160–167.

[30] J.W. Seo, R. Bontrop, L. Walter, E. Gunther, Major histocompatibility complex-linked MIC genes in rhesus macaques and other primates, Immunogenetics 50 (1999) 358–362.

[31] T. Shiina, et al., Molecular dynamics of MHC genesis unraveled by sequence analysis of the 1,796,938-bp HLA class I region, Proc. Natl. Acad. Sci. USA 96 (1999) 13282–13287.

[32] T. Anzai, et al., Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence, Proc. Natl. Acad. Sci. USA 100 (2003) 7708–7713.

[33] J.K. Kulski, T. Anzai, T. Shiina, H. Inoko, Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex, Mol. Biol. Evol. 21 (2004) 2079–2091.

[34] H. Piontkivska, M. Nei, Birth-and-death evolution in primate MHC class I genes: divergence time estimates, Mol. Biol. Evol. 20 (2003) 601–609.

[35] C.L. Magness, et al., Analysis of the Macaca mulatta transcriptome and the sequence divergence between Macaca and human, Genome Biol. 6 (2005) R60.1–R60.16.

[36] K. Osoegawa, et al., An improved approach for construction of bacterial artificial chromosome libraries, Genomics 52 (1998) 1–8.

[37] S. Kumar, K. Tamura, M. Nei, MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment, Brief. Bioinformat. 5 (2004) 150–163.

[38] C.A. Stewart, et al., Complete MHC haplotype sequencing for common disease gene mapping, Genome Res. 14 (2004) 1176–1187.