



Different Evolutionary Histories in Two Subgenomic Regions of the Major Histocompatibility Complex

Silvana Gaudieri, Jerzy K. Kulski, Roger L. Dawkins, et al.

Genome Res. 1999 9: 541-549

Access the most recent version at doi:[10.1101/gr.9.6.541](https://doi.org/10.1101/gr.9.6.541)

References This article cites 29 articles, 5 of which can be accessed free at:
<http://genome.cshlp.org/content/9/6/541.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Different Evolutionary Histories in Two Subgenomic Regions of the Major Histocompatibility Complex

Silvana Gaudieri,^{1,3} Jerzy K. Kulski,² Roger L. Dawkins,² and Takashi Gojobori¹

¹Centre for Information Biology, National Institute of Genetics, Mishima, Shizuoka, 411-8540 Japan; ²Centre for Molecular Immunology and Instrumentation, University of Western Australia, Subiaco, Western Australia, 6008 Australia

Two subgenomic regions within the major histocompatibility complex, the alpha and beta blocks, contain members of the multicopy gene families HLA class I, human endogenous retroviral sequence (HERV-16; previously known as P5 and PERB3), hemochromatosis candidate genes (HCG) (II, IV, VIII, IX), 3.8-1, and MIC (PERB1). In this study we show that the two blocks consist of imperfect duplicated segments, which contain linked members of the different gene families. The duplication and truncation sites of the segments are associated with retroelements. The retroelement sites appear to generate the imperfect duplications, insertions/deletions, and rearrangements, most likely via homologous recombination. Although the two blocks share several characteristics, they differ in the number and orientation of the duplicated segments. On the 62.1 haplotype, the alpha block consists of at least 10 duplicated segments that predominantly contain pseudogenes and gene fragments of the HLA class I and MIC (PERB1) gene families. In contrast, the beta block has two major duplications containing the genes *HLA-B* and *HLA-C*, and *MICA* (PERB1.1) and *MICB* (PERB1.2). Given the common origin between the blocks, we reconstructed the duplication history of the segments to understand the processes involved in producing the different organization in the two blocks. We then found that the beta block contains four distinct duplications from two separate events, whereas the alpha block is characterized by multisegment duplications. We will discuss these results in relation to the genetic content of the two blocks.

The major histocompatibility complex (MHC) has been described as a composite of conserved polymorphic blocks (Marshall et al. 1993). These blocks are ~200–300 kb in length, and their combination is observed in a population as MHC haplotypes (Degli-Esposti et al. 1992). The central and telomeric region of the human MHC spans ~2 Mb on the short arm of chromosome six and contains at least two blocks, the alpha and beta blocks (Marshall et al. 1993). The blocks contain members of the multicopy gene families HLA (human leukocyte antigen) class I (Geraghty et al. 1992); MIC (MHC class I-related chain) [alternatively termed PERB11 (Perth beta block transcript 11)] (Bahram et al. 1994; Leelayuwat et al. 1994); P5 (Vernet et al. 1993); PERB3 (Marshall et al. 1993); 3.8-1 (Pichon et al. 1997); hemochromatosis candidate genes (HCGs) II, IV, VIII, and IX (Pichon et al. 1996b); and 1A3 (Totaro et al. 1995).

At least two of the multicopy gene families, HLA class I and MIC (PERB11), have an immunological role. The HLA class I family includes genes that encode cell surface glycoproteins that bind peptides and present them to cytotoxic T cells (Parham 1988). The function of MIC (PERB11) may involve $\gamma\delta$ T cells and the protection of epithelial surfaces (Groh et al. 1998). Of the other multicopy gene families, P5 and PERB3 have

been shown to form a genomic structure related to a human endogenous retroviral sequence-16 (HERV-16) sequence (Kulski and Dawkins 1999), HCGIX has similarity to transcription factors (Pichon et al. 1996c), and 1A3 has no known transcripts.

The organization of the multicopy gene families in the alpha and beta blocks reflects a common origin and duplicative history (Leelayuwat et al. 1995; Pichon et al. 1996a). Within the beta block, we have recently shown that there are two large segmental duplications containing the loci *HLA-B* and *HLA-C*, and *MICA* (PERB1.1) and *MICB* (PERB1.2) (Gaudieri et al. 1997; Kulski et al. 1997). Based on L1 and *Alu* dating, the two duplications occurred as separate events resulting in four distinct segments (Kulski et al. 1997; Yamazaki et al. 1999). These duplications contain segments that share ~30 kb of sequence, including coding and intergenic sequences. The major differences between the duplications are large insertions/deletions (indels) (>100 bp) containing retroelements [*Alu*, L1, HERV, MaLR (mammalian apparent LTR retrotransposon)], and processed pseudogenes and other repetitive sequences [mammalian wide interspersed repeat (MIR)/medium reiterated frequency repeat (MER)] (Kulski et al. 1997). The duplication and insertion sites of sequences within the duplications of the beta block are closely associated with retroelement sequences (Gaudieri et al. 1997; Kulski et al. 1997).

³Corresponding author.
E-MAIL sgaudier@lab.nig.ac.jp; FAX 81-559-81-6848.

In contrast to the beta block, the alpha block contains a greater number of members of the multicopy gene families, particularly more pseudogenes and gene fragments of the HLA class I and MIC (PERB11) families (Geraghty et al. 1992; Leelayuwat et al. 1994). Within the alpha block, tandem gene duplications, gene conversion, and “block” duplications have been proposed to explain the complex arrangement of HLA class I loci resulting in the separation of phylogenetically related loci by more distantly related loci (Geraghty et al. 1992; Hughes 1995).

The purpose of this study was (1) to determine whether segmental duplications, as in the beta block, are involved in the organization of the multicopy gene families in the alpha block; (2) to determine the role of retroelements in the processes of duplication, truncation, and indels; and (3) to reconstruct the history of the duplications to understand the differing number of multicopy gene family members and complexity between the two blocks. To achieve these purposes we used recently released sequences spanning the alpha and beta blocks to compare the structural organization of the two blocks. We show that both blocks contain duplicated segments composed of linked multicopy gene families. The duplication and indel sites of the segments are associated with retroelements and other fragile sequences within the region and generate the dispersion and diversity of the multicopy gene families. However, the blocks differ in segment number, complexity, and orientation. The beta block contains four distinct segments from two separate duplication events, whereas the alpha block contains multisegment duplications. The multisegment duplications contain at least two tandem segments, one each from the two major segment types in the alpha block. We will discuss the implications of the different duplication modes in relation to the genetic contents of each block.

RESULTS

Different Members of Multicopy Gene Families Are Linked Together as a Single Unit Within Duplicated Segments in Both Alpha and Beta Blocks

Figure 1 shows two dot plots of duplicated genomic segments within the alpha and beta blocks and the locations of the multicopy gene families HLA class I, 1AD3, P5 + PERB3 (HERV-16), 3.8-1, MIC (PERB11), and HCG (II, IV, VIII, and IX). The arrangement of the multicopy gene families and the lines within the dot plots indicate they are linked and form duplicated segments in both blocks, designated a–n. In this regard, segmental groups a–n depicted as homologous sequences by dot plots are defined by the genes and retroelements contained within the segment, as shown in Figure 2.

Comparison of the beta block against itself reveals two distinct duplications that are well separated and likely to be the result of single segment duplications. Each duplication set shares ~30 kb of sequence. The two sets of duplications are in the same orientation but share limited sequence (Fig. 1). The segments k and l together contain HCGII, HLA class I, 1AD3, HERV-16, 3.8-1, MIC (PERB11), HCGIX, whereas segments m and n contain HCGII, HLA class I, and HCGIV.

Comparison of the alpha block against itself reveals a complex array of segment duplications (Fig. 1). All duplications are in the same orientation. There are no clear distinctions between the segments indicating little intervening sequence. The length of the segments varies between ~10 and 52 kb. The alpha block segments a–j share HLA class I, HCGIV, 1AD3, and HERV-16. The duplications are imperfect and are interrupted with HCGII, HCGVIII, HCGIX, 3.8-1, and MIC (PERB11). The large insertion corresponding to the HLA-A31 haplotype contains a similar organization to the segments within the alpha block of the 62.1 haplotype (Fig. 1A). Although the segments within both blocks contain a similar order of the multicopy gene families, there are clear differences between the blocks that will be described in a later section.

Duplication, Truncation, and Indel Sites Associated with HERV-16 and Other Retroelement Sequences Within Duplicated Segments

The endpoints on several segments in both blocks, including the large indels corresponding to the HLA-A31 and HLA-A9 haplotypes, coincide with HERV-16 and other retroelement sequences (Figs. 1 and 2) (Kulski et al. 1999). Indel sites within the beta block have been shown to be associated with retroelement sequences (Gaudieri et al. 1997). Similarly, deletion sites in at least two segments in the alpha block are also associated with HERV-16 (Fig. 2). The insertion of retroelements THE1C, L1PA, MER, and L1ME3A is shared by several segments in the alpha block (Fig. 2). This insertion has occurred between an MLT and LTR16B (5' end of HERV-16), and in segment i some of these retroelements form another internal duplication within the initial insertion (Fig. 2).

Other sites of truncations and rearrangements occur within the HLA class I and MIC (PERB11) family. Within the HLA class I family, the following loci contain only part of the full-length gene: HLA-80 and HLA-90 exon 3-8, HLA-16 exon 4-8, HLA-75 exon 1-3, HLA-17 exon 6-8, HLA-21 exon 3, and HLA-X intron C (Fig. 2) (Geraghty et al. 1992). These observations indicate a potential hot spot(s) between exons 2 and 4.

Figure 3 shows the MIC (PERB11) gene region in 6 of 14 segments from the alpha and beta blocks. The members differ from each other by large indels corresponding to L1 and *Alu* sequences. MICD (PERB11.4) is

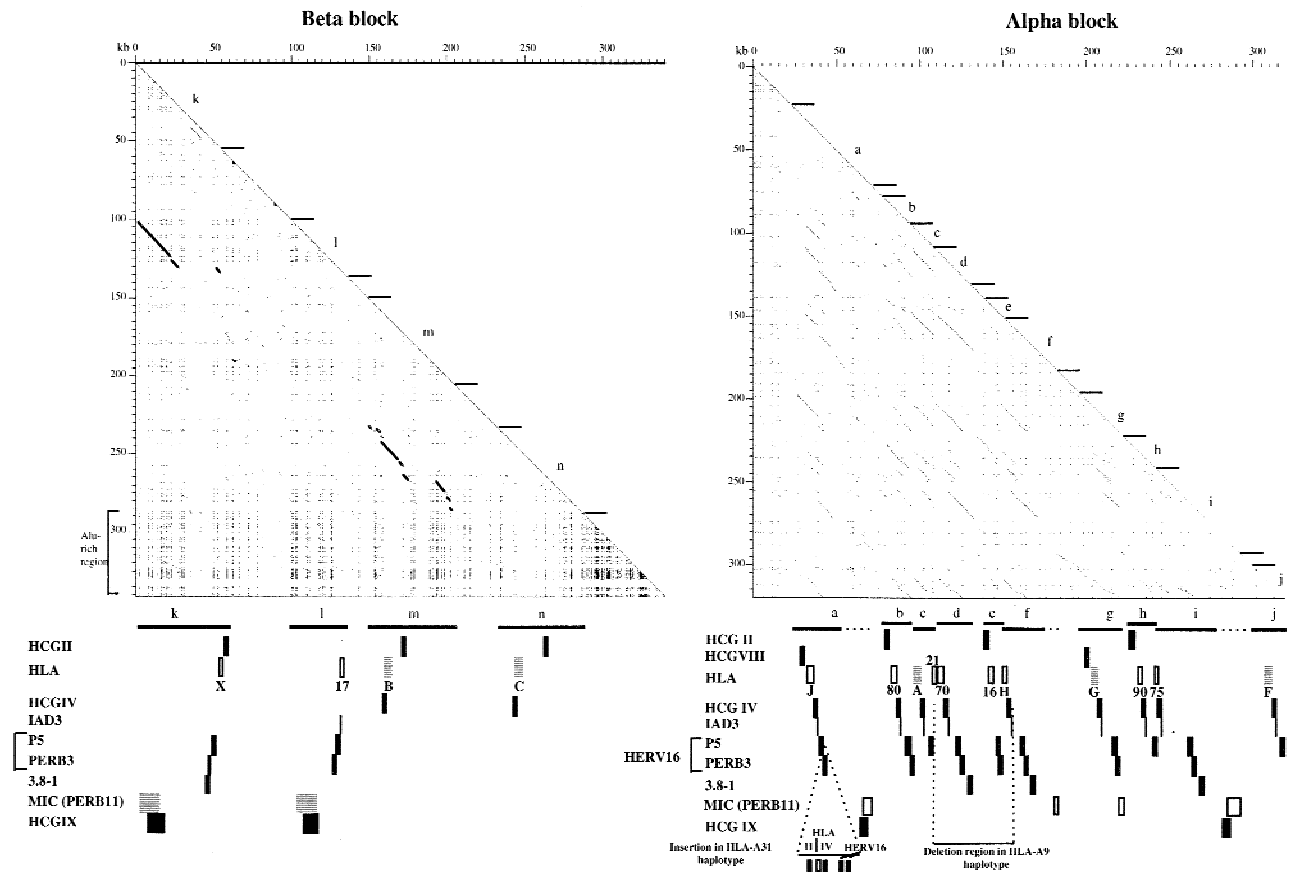


Figure 1 Segmental duplication of linked multicopy gene families. Dot-plot analysis of the intracomparison of the alpha (*right*) and beta (*left*) blocks. Diagonal lines within the dot plots indicate regions of duplication, with interruptions corresponding to indels. The position of the multicopy gene families HLA class I, HCG (II, IV, VIII, and IX), IAD3, PS + PERB3 (HERV-16), 3.8-1, and MIC (PERB11) are shown below the *x*-axis. Based on the duplication end points in the dot plot and the ordering of the multicopy gene families, duplicated segments have been designated a–n. In the alpha block, there is minimal distance between duplicated segments indicating that the duplication of the MIC (PERB11) members may have involved, though less likely, a neighboring segment. For the well-characterized HLA class I and MIC (PERB11) members, transcribed genes are indicated by shaded boxes, and pseudogenes and gene fragments are shown by open boxes. The HCG, IAD3, and HERV-16 families are shown by black boxes. The endpoints of the human haplotype indels are indicated by horizontal dashed lines. Segment j is truncated at the 3' end following an *Alu* Ya5/8 insert coinciding with the end of the alpha block sequence (accession no. AF055066). An *Alu*-rich region in the beta block is indicated.

truncated at the 3' end in intron 5, which directly correlates with the start of MIC.b (PERB11.b). MIC.a (PERB11.a) commences further downstream of the MICD (PERB11.4) truncation in intron 5. In addition, the *L1* sequence, L1MA4A, in MIC.a (PERB11.a) corresponds to a similar L1MA sequence within intron 1 of *MICA* (PERB11.1) and *MICB* (PERB11.2) (L1MA3). These L1 sequences are not present in MICD (PERB11.4) or MICE (PERB11.5). However, L1MB3 sequence is present in *MICA* (PERB11.1), MICD (PERB11.4), and MICE (PERB11.5), possibly in *MICB* (PERB11.2), but not in MIC.a (PERB11.a) and MIC.b (PERB11.b). The organization of MIC.a (PERB11.a) and MIC.b (PERB11.b) indicates introns 1 and 5 are fragile sites susceptible to change.

The association of retroelement sequences with duplication, truncation, and indel sites is a common feature of both blocks. However, significant differences

between the blocks exist, which will be described in the next section.

Different Orientation and Complexity of Segmental Duplications in the Alpha and Beta Blocks

Comparison of the alpha and beta blocks shows the following differences:

1. There is opposite orientation of duplicated segments corresponding to segments containing both MIC (PERB11) and HLA class I members. The sequence surrounding *HLA-B* and *HLA-C* in the beta block does not form large segment duplications with the class I region of the alpha block; only the HLA class I gene or pseudogene and limited surrounding sequence is retained (~5–10 kb) (Fig. 4).
2. The alpha block, on the 62.1 haplotype, contains at least 10 segments compared with the four segments within the beta block.

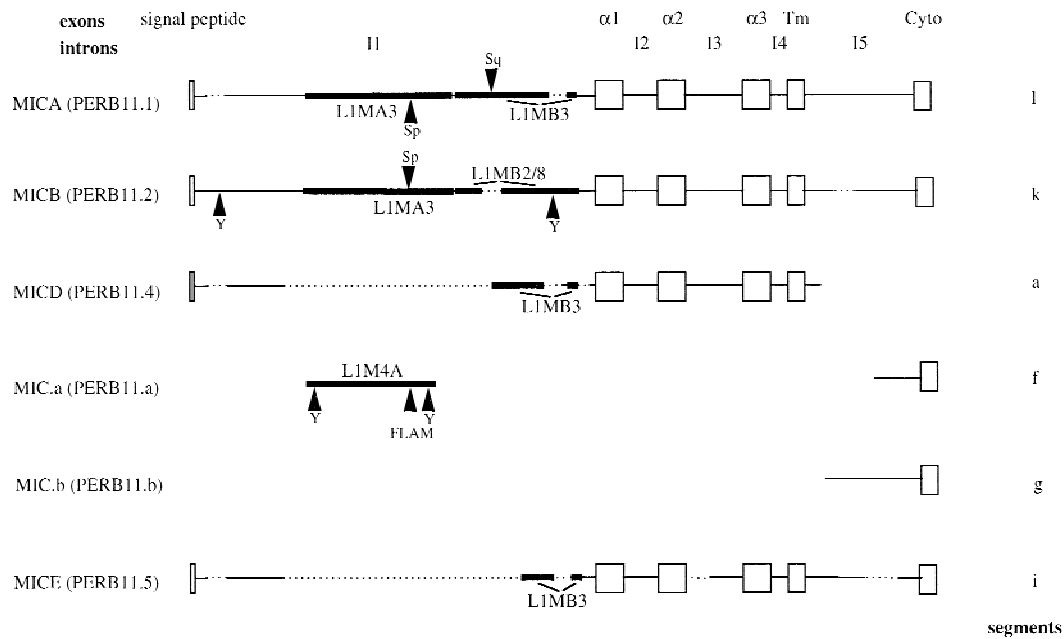


Figure 3 Indels/rearrangements within MIC (PERB11) members. A schematic diagram of the MIC (PERB11) gene family. The intron and exon structure is depicted on *MICA* (*PERB11.1*). Exons are shown as open boxes and labeled as signal peptide, $\alpha 1$, $\alpha 2$, $\alpha 3$, Tm (transmembrane region), and cyto (cytoplasmic region). The introns are shown as lines and labeled in order as I1–I5. The positions of gaps in MIC (PERB11) members are indicated by a dashed line. L1 sequences are depicted by solid bars. The positions of Alu sequences are shown by arrowheads. The signal peptide of MICD (PERB11.4) is shown by a grey box, as the level of identity with other members is low in this region and the corresponding open reading frame does not encode hydrophobic amino acids as found in the signal peptide of other members. *MICA* (PERB11.1) and *MICB* (PERB11.2) are full-length genes transcribed in the beta block. MICD (PERB11.4) and MICE (PERB11.5) are pseudogenes, and MIC.a (PERB11.a) and MIC.b (PERB11.b) are fragments containing the 3' end of the gene.

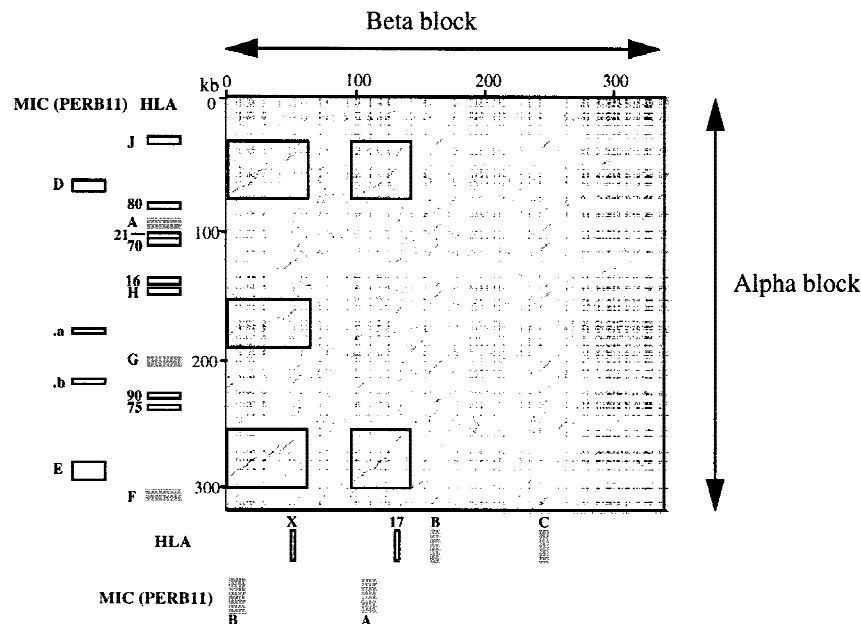


Figure 4 Different orientation and complexity of segments in the alpha and beta blocks. Intercomparison of the alpha and beta blocks. The orientation of the duplications shown by the small diagonal lines indicates that the segments from the two blocks are in opposite orientation to each other. Large segment duplications corresponding to segments containing MIC (PERB11) and HLA class I members are shown by boxes. For the well-characterized HLA class I and MIC (PERB11) members, transcribed genes are indicated by grey boxes, and pseudogenes and gene fragments are shown by open boxes along the x - and y -axes.

duplicated segments (Fig. 2) share a common phylogenetic organization (Fig. 5), supporting the duplication of linked multicopy gene families.

Figure 5 shows that segments b, h, and e (containing HLA-80, HLA-90, and HLA-16, respectively), separate from the other segments within the alpha block, with high bootstrap values and form a separate lineage in agreement with previous results for HLA-80, HLA-90, and HLA-16 (Hughes 1995). Nucleotide comparison of paralogous L1 sequences between the alpha block segments shows segments b and h share 90.3% identity but only 76.3% with the other alpha block segments (Table 2). In addition, the segments b, e, and h share a distinct set of retroelements at the 5' end corresponding to the region containing HCGII, which separates them from the other alpha block segments, which lack both HCGII and the retroelements

Table 1. Different Properties of the Alpha and Beta Blocks

Block	Pseudogene/ functional ^a	No. of segments	Retroelements (%) ^b	Paralogous retroelements (%)	Pattern of duplication	Human indel polymorphisms
Alpha	4	at least 10	39.8	57.7	single, bisegmental and transpositions	several ^c
Beta	0.5	4	49.7	28.5	single	?

^aFor HLA and MIC (PERB11) members.^bFrom RepeatMasker output.^cVenditti and Chorney (1992); Watanabe et al. (1997).

(Fig. 2). There are two main groups of segments in the alpha block that must have diverged early in the evolution of the segments.

The segments a, d, g, and i, which harbor HLA-J, HLA-70, *HLA-G*, and HLA-75, respectively, do not have distinct clustering patterns in the different phylogenetic trees (Fig. 5A,B,D,E) indicating similar levels of identity, supported by the comparison of paralogous L1 sequences (88%–91%) (Table 2). Similarly, the segments b, h, and e share 90.3% identity between paralogous L1 sequences. The position of the two major groups of segments in the alpha block and equivalent levels of identity between paralogous L1 sequences suggest at least two instances of bisegmental duplications involving a member from each group (i.e., a bisegmental duplication of segments h and i, and segments e and f) (Fig. 2). In addition, the sequence content between segments f and g is similar to that between segments i and j, suggesting bi- or trisegmental duplications within this block (Fig. 2). A more detailed analysis of the evolution of the segments in the alpha and beta blocks based on retroelement dating will be presented elsewhere (Kulski et al. 1999). The complex history of the segments in the alpha block differs from the single segment duplications present in the beta block.

Furthermore, the segments f (containing HLA-H) and c (containing *HLA-A*) form a cluster (Fig. 5A,B,D,E) with high reliability. The level of identity between paralogous L1 sequences in segments f and c is ~94.1%, greater than any other comparison, supporting a recent common ancestor (Table 2). Previous phylogenetic results of *HLA-A* and HLA-H showed they had diverged recently (Hughes 1995). The segments c and f can be further separated from the other alpha block segments by the insertion of a MER9 and *Alu Y* insert (Fig. 2). To account for the large distance between the two segments, a likely translocation event has occurred subsequent to duplication. The alpha block segment history includes single and multisegment duplications as well as translocations. In contrast, individual segment duplications have been described in the beta block (Gaudieri et al. 1997; Kulski et al. 1997).

DISCUSSION

Common Features of the Alpha and Beta Blocks

The alpha and beta blocks are characterized by the duplication of linked multicopy gene families. We have shown that both blocks contain imperfect segmental duplications resulting in the dispersion and diversity of the families and associated retroelements. An important consequence of the duplications is the increased possibility of recombination between paralogous retroelements and other repetitive sequences resulting in genomic changes (Mazzarella and Schlessinger 1997).

Ancestral Segment Contained Linked Members of the HLA Class I, HERV-16, and MIC (PERB11) Gene Families

The composition of the segments in the alpha and beta blocks suggests that the ancestral HLA class I gene was linked to other multicopy gene families through retroelements such as HERV-16. The ancestral segment of the alpha and beta blocks contained linked members of the HLA class I, HCGIV, HERV-16, 3.8-1, and MIC (PERB11)/HCGIX gene families and the associated retroelements L2, MLT1E, MLT, MLT1F, and LTR16B (Figs. 1 and 2). Although the functions of HCGIV, HERV-16, and 3.8-1 are unknown, the HLA class I and MIC (PERB11) genes are involved in antigen presentation to T cells and their linkage may be of functional importance.

The distinct phylogenetic cluster of the segments b, h, and e away from the other segments within both blocks (Fig. 5E) suggests that the ancestral segment of b, h, and e was the result of an imperfect duplication at HERV-16, or the ancestral segment did not extend beyond HERV-16. In the HLA class I exon 4–7 tree (Fig. 5E), segments m and n containing *HLA-B* and *HLA-C*, respectively, clustered between the two major groups of alpha block segments, suggesting that the ancestral segment of m and n diverged from the alpha block segments after the separation of the two main groups of segments. Therefore, the ancestral segment of m and n was imperfect, resulting in limited common sequence with other segments.

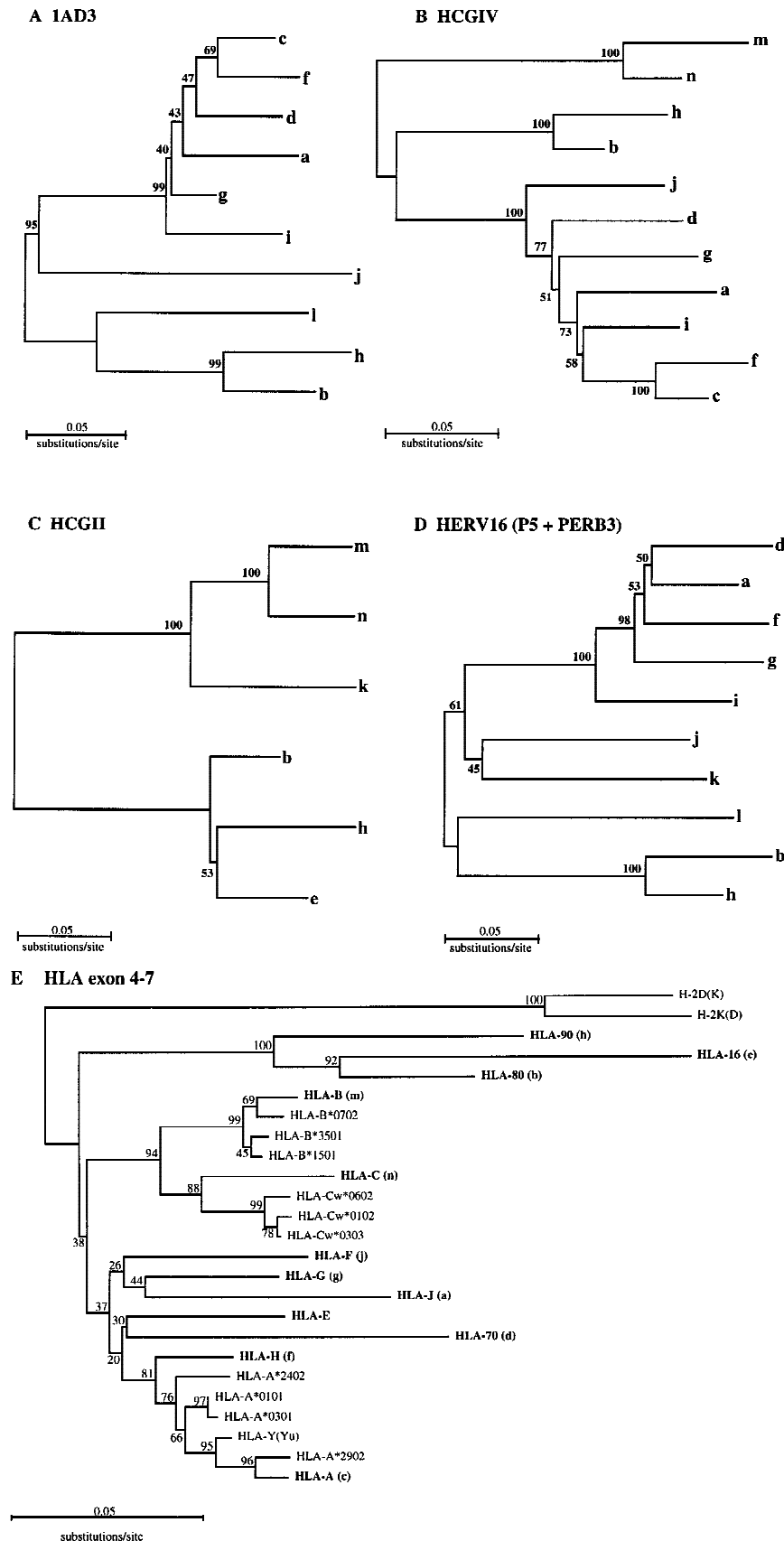


Figure 5 The alpha block segments can be separated into two main groups. The phylogenetic relationships of segments a–n were examined by analyzing the multicopy gene families of 1AD3 (A), HCGIV (B), HCGII (C), HERV-16 (D), and HLA class I (E). A scale bar for the number of substitutions per site is included for each tree and the bootstrap percentage from 1000 replications is given for each node. All trees except the HLA class I tree are unrooted. In the HLA class I tree (E), mouse H2 sequences were used as an outgroup and segments corresponding to HLA class I loci are indicated by brackets.

Homologous Recombination Between Retroelements Is an Important Process Within the MHC

The segment duplications, truncations, indels, and internal duplications coincide with regions associated with retroelements, particularly HERV-16. These results indicate that homologous recombination between retroelements is an important mechanism of such changes.

There is increasing evidence that retroelements (including HERVs) contribute to the recombination and duplications observed in the MHC (Kulski et al. 1997; Andersson et al. 1998) and in other gene clusters in the genome (Erickson et al. 1992; Schwartz et al. 1998). Furthermore, instability of operon structures in bacteria has been shown to be associated with transposon insertion sequences (Itoh et al. 1999).

Differences Between Alpha and Beta Blocks

Although the two blocks in this study shared a common origin, the duplicative history is different in their overall genomic organization. The blocks differ in the number of duplicated segments and in the proportion of pseudogenes and gene fragments. The alpha block contains a greater proportion of paralogous retroelements (Table 1) resulting in a greater level of identity between the sequences, and therefore, an in-

Table 2. Nucleotide Identity (%) Between Paralogous L1 Sequences Within the Alpha and Beta Block Segments

Segments	No. of paralogous L1 ^a											
	a	b	c	d	e	f	g	h	i	j	k	l
a		N.A.	90.3	91.1	N.A.	86.9	90.7	N.A.	88	87.4	76.8	76.5
b	0		76.3	76.4	89.8	75.8	75.9	90.6 ^b	77.3	N.A.	N.A.	N.A.
c	1	2		89.2	N.A.	94.1 ^c	89.1	77	90.1	85.8	N.A.	N.A.
d	2	1	3		N.A.	89.6	89.8	76.1	89.2	91.4	N.A.	N.A.
e	0	1	0	0		N.A.	N.A.	90.6	N.A.	N.A.	N.A.	N.A.
f	4	2	3	4	0		88.6	75.4	88.6	84	80.2	N.A.
g	3	1	3	3	0	3		76.3	89.9	87.4	N.A.	N.A.
h	0	3	1	1	1	1	1		78.9	N.A.	N.A.	N.A.
i	2	1	3	4	0	4	3	1		86.4 ^d	87.5	81.6
j	2	0	1	1	0	1	2	0	1		N.A.	N.A.
k	2	0	0	0	0	1	0	0	1	0		93.7 ^e
l	1	0	0	0	0	0	0	0	1	0		

(N.A.) Not applicable.

^aOnly paralogous L1 sequences >200 bp are included.^bStatistically different from other segment b comparisons. $P < 0.05$ using *t*-test.^cStatistically different from other segment f comparisons. $P < 0.05$ using *t*-test.^dPercentage in black box represents internal duplication of segment i.^eObtained from Yamazaki et al. (1999).

crease in the possibility of exchange and indel events within the block. The process of imperfect duplications and the subsequent indels and rearrangements results in a large number of nonfunctional sequences. The alpha block appears to have accumulated nonfunctional sequences. The functional *HLA-A*, *HLA-G*, and *HLA-F* sequences in the alpha block have not been disrupted, however, the importance of these three loci has been questioned (Klein et al. 1998).

The differences between the two blocks are a result of the mode of segment duplications, which in turn have been generated by retroelements. The blocks differ either through an intrinsic property of the genomic sequence within the blocks, or more likely, through chance events whereby duplicated segments have provided a greater opportunity for homologous recombination between retroelements, with the result of additional gene copies. The formation of noncoding sequences of the HLA class I and MIC (PERB11) gene families within the alpha block (particularly early in the duplicative history through the formation of the ancestral segment of b, h, and e) may have allowed the propagation of the segment duplications in the alpha block, resulting in the present complex organization. The arrangement of segments in the alpha block appears to have occurred by several duplications of more than one segment, whereas two separate duplication events have occurred in the beta block. Therefore, the complex segment history in the alpha block is different from the history of the beta block segments that have arisen via single segment duplications.

METHODS

DNA Sequences

A 319-kb sequence within the alpha block spanning 32 kb centromeric of *HLA-J* to 6 kb telomeric of *HLA-F* was obtained from the DDBJ/EMBL/GenBank accession no. AF055066. A 337-kb sequence within the beta block spanning *MICB* (*PERB11.2*) to 87 kb telomeric of *HLA-C* was obtained from DDBJ/EMBL/GenBank accession nos. AB000878, AB000879, AB000880, and D84394. AF055066, AB000878, AB000879, and AB000880 are from the cell line BOLETH and contain the HLA alleles *HLA-A2*, *HLA-B62*, *HLA-Cw10*, and *HLA-DR4* (62.1 haplotype). D84394 is from the heterozygous cell line CGM1 with the HLA alleles *HLA-A3,29*; *HLA-B8,14*; *HLA-Cw-;*; and *HLA-DR3,7*.

The *HLA-A*, *HLA-B*, and *HLA-C* alleles used for the phylogenetic tree in Figure 5E were obtained from the World Wide Web (http://square.umin.ac.jp/JSHI/hla_data/data.html). The sequences for *HLA-Y* (YU) (accession no. AB012685), *HLA-E* (accession no. M20022), *H-2D* (K) (accession no. U47327), and *H-2K* (D) (accession no. U47329) were obtained from the DDBJ/EMBL/GenBank database.

Location of Multicopy Gene Families, Retroelements, and Duplications

The locations of the HLA class I, HERV-16 (P5 + PERB3), MIC (PERB11), 1A3, 3.8-1, and HCGII, HCGIV, HCGVIII, and HCGIX family members on the sequences were determined using the program BLASTN (National Center for Biotechnology Information, Bethesda, MD) and the annotation information of the database entry. The position and type of retroelement within each sequence was determined using the RepeatMasker program (<http://ftp.genome.washington.edu/cgi-bin/RepeatMasker>). The inter- and intracomparison of the sequences was performed using the dot matrix program Dotter (Sonnhammer and Durbin 1996).

Phylogenetic Analysis

The members of the HLA class I, HERV-16 (P5 + PERB3), 1AD3, HCGII, and HCGIV families were each aligned using the CLUSTALW program (GCG, Madison, WI). The resultant alignments were used in the construction of the phylogenetic trees by the neighbor-joining method (Saitou and Nei 1987) based on Kimura's two-parameter model (Kimura 1981). To assess the reliability of the clusters, 1000 bootstrap replications were performed (Felsenstein 1985).

ACKNOWLEDGMENTS

We thank Dr. Naoko Takezaki, Professor Yoshio Tateno, Dr. Kazuho Ikeo, Dr. Annalise Martin, Sonia Cattle, and two anonymous reviewers for helpful comments regarding the manuscript. S.G. is supported by a Japanese Society for the Promotion of Science fellowship (97119). J.K.K. and R.L.D. are supported by the National Health and Medical Research Council.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Andersson, G., A. Svensson, N. Setterblad, and L. Rask. 1998. Retroelements in the human MHC class II region. *Trends Genet.* **14**: 109–114.
- Bahram, S., M. Bresnahan, D.E. Geraghty, and T. Spies. 1994. A second lineage of mammalian major histocompatibility complex class I genes. *Proc. Natl. Acad. Sci.* **91**: 6259–6263.
- Degli-Esposti, M.A., A.L. Leaver, F.T. Christiansen, C.S. Witt, L.J. Abraham, and R.L. Dawkins. 1992. Ancestral haplotypes: Conserved population MHC haplotypes. *Hum. Immunol.* **34**: 242–252.
- Erickson, L.M., H.S. Kim, and N. Maeda. 1992. Junctions between genes in the haptoglobin gene cluster of primates. *Genomics* **14**: 948–958.
- Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* **39**: 783–791.
- Gaudieri, S., K. Giles, J.K. Kulski, and R.L. Dawkins. 1997. Duplication and polymorphism in the MHC: Alu generated diversity and polymorphism within the PERB11 gene family. *Hereditas* **127**: 37–46.
- Geraghty, D.E., B.H. Koller, J.A. Hansen, and H.T. Orr. 1992. The HLA class I gene family includes at least six genes and twelve pseudogenes and gene fragments. *J. Immunol.* **149**: 1934–1946.
- Groh, V., A. Steinle, S. Bauer, and T. Spies. 1998. Recognition of stress-induced MHC molecules by intestinal epithelial $\gamma\delta$ T cells. *Science* **279**: 1737–1740.
- Hughes, A.L. 1995. Origin and evolution of HLA class I pseudogenes. *Mol. Biol. Evol.* **12**: 247–258.
- Itoh, T., K. Takemoto, H. Mori, and T. Gojobori. 1999. Evolutionary instability of operon structures disclosed by sequence comparisons of complete microbial genomes. *Mol. Biol. Evol.* (in press).
- Kimura, M. 1981. Estimation of evolutionary distances between homologous nucleotide sequences. *Proc. Natl. Acad. Sci.* **78**: 454–458.
- Klein, J., A. Sato, and C. O'Uigin. 1998. Evolution by gene duplication in the major histocompatibility complex. *Cytogenet. Cell Genet.* **80**: 123–127.
- Kulski, J.K. and R.L. Dawkins. 1999. The P5 multicopy gene family in the MHC is related in sequence to human endogenous retroviruses HERV-L and HERV-16. *Immunogenetics* **49**: 404–412.
- Kulski, J.K., S. Gaudieri, M. Bellgard, L. Balmer, K. Giles, H. Inoko, and R.L. Dawkins. 1997. The evolution of MHC diversity by segmental duplication and transposition of retroelements. *J. Mol. Evol.* **45**: 599–609.
- Kulski, J.K., S. Gaudieri, A. Martin, and R.L. Dawkins. 1999. Coevolution of PERBII (MIC) and HLA Class I Genes with HERV-16 and retroelements by extended genomic duplication. *J. Mol. Evol.* (in press).
- Leelayuwat, C., D.C. Townend, M.A. Degli-Esposti, L.J. Abraham, and R.L. Dawkins. 1994. A new polymorphic and multicopy MHC gene family related to nonmammalian class I. *Immunogenetics* **40**: 339–351.
- Leelayuwat, C., M. Pinelli, and R.L. Dawkins. 1995. Clustering of diverse replicated sequences in the MHC. Evidence for en bloc duplication. *J. Immunol.* **155**: 692–698.
- Marshall, B., C. Leelayuwat, M.A. Degli-Esposti, M. Pinelli, L.J. Abraham, and R.L. Dawkins. 1993. New major histocompatibility complex genes. *Hum. Immunol.* **38**: 24–29.
- Mazzarella, R. and D. Schlessinger. 1997. Duplication and distribution of repetitive elements and non-unique regions in the human genome. *Gene* **205**: 29–38.
- Parham, P. 1988. Function and polymorphism of human leukocyte antigen-A,B,C molecules. *Am. J. Med.* **85**: 2–5.
- Pichon, L., G. Carn, P. Bouric, T. Giffon, B. Chauvel, M. Lepourcelet, J. Mosser, J. Legall, and V. David. 1996a. Structural analysis of the HLA/HLA-F subregion: Precise localization of two new multigene families closely associated with the HLA Class I sequences. *Genomics* **32**: 236–244.
- Pichon, L., T. Giffon, B. Chauvel, G. Carn, P. Bouric, A. El Kahloun, J. Legall, and V. David. 1996b. Physical map of the HLA-A/HLA-F subregion and identification of two new coding sequences. *Immunogenetics* **43**: 175–181.
- Pichon, L., A. Hampe, T. Giffon, G. Carn, J.Y. Legall, and V. David. 1996c. A new non-HLA multigene family associated with the PERB11 family within the MHC class I region. *Immunogenetics* **44**: 259–267.
- Pichon, L., C. Venditti, J. Harris, A. Elshof, M. Pinelli, and M. Chorney. 1997. Studies of the 6.7 family of dispersed genomic fragments within the MHC class I region. *Exp. Clin. Immunogenet.* **14**: 131–140.
- Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.
- Schwartz, A., D.C. Chan, L.G. Brown, R. Alagappan, D. Pettay, C. Disteche, B. McGillivray, A. de la Chapelle, and D.C. Page. 1998. Reconstructing hominid Y evolution: X-homologous block, created by X-Y transposition, was disrupted by Yp inversion through LINE-LINE recombination. *Hum. Mol. Genet.* **7**: 1–11.
- Sonnhammer, E.L.L. and R. Durbin. 1996. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GC1–GC10.
- Totaro, A., A. Grifa, A. Roetto, C. Lunardi, L. D'Agruma, L. Sbaiz, L. Zelante, G. De Sandre, C. Camaschella, and P. Gasparini. 1995. New polymorphisms and markers in the HLA class I region: Relevance to hereditary hemochromatosis (HFE). *Hum. Genet.* **95**: 429–434.
- Venditti, C.P. and M.J. Chorney. 1992. Class I gene contraction within the HLA-A subregion of the human MHC. *Genomics* **14**: 1003–1009.
- Vernet, C., M.T. Ribouchon, G. Chimini, A.M. Jouanolle, I. Sidibe, and P. Pontarotti. 1993. A novel coding sequence belonging to a new multicopy gene family mapping within the human MHC class I region. *Immunogenetics* **38**: 47–53.
- Watanabe, Y., K. Tokunaga, D.E. Geraghty, K. Tadokoro, and T. Juji. 1997. Large-scale comparative mapping of the MHC class I region of predominant haplotypes in Japanese. *Immunogenetics* **46**: 135–141.
- Yamazaki, M., Y. Tateno, and H. Inoko. 1999. Genomic organization around the centromeric end of the HLA Class I region: Large-scale sequence analysis. *J. Mol. Evol.* **48**: 317–327.

Received December 15, 1998; accepted in revised form May 10, 1999.