J Mol Evol (1997) 45:17–23

JOURNAL OF MOLECULAR EVOLUTION

The Major Histocompatability Complex (MHC) Contains Conserved Polymorphic Genomic Sequences That Are Shuffled by Recombination to Form Ethnic-Specific Haplotypes

Silvana Gaudieri, Chanvit Leelayuwat, Guan K. Tay, David C. Townend, Roger L. Dawkins

Centre for Molecular Immunology and Instrumentation, GPO Box 507, Subiaco, Western Australia 6008

Received: 20 January 1997 / Accepted: 11 March 1997

Abstract. The major histocompatibility complex (MHC) consists of polymorphic frozen blocks (PFBs) that are linked to form megabase haplotypes. These blocks consist of polymorphic sequences and define regions where recombination appears to be inhibited. We have been able to show, using a highly polymorphic sequence centromeric of *HLA-B* (within the beta block), that PFBs are conserved and contain specific insertions/deletions and substitutions that are the same for individuals with the same MHC haplotype but that differ between at least most different haplotypes. A sequence comparison between ethnic-specific haplotypes shows that these sequences have remained stable and predate the formation of these haplotypes. To determine whether the same conserved block has been involved in the generation of multiple haplotypes, we compared the block typing profiles of different ethnic specific haplotypes. Block typing profiles have previously been shown to be identical in individuals with the same MHC haplotype but, generally, to differ between different haplotypes. It was found that some PFBs are common to more than one haplotype, implying a common ancestry. Subsequently, haplotypes have been generated by the shuffling and exchange of these PFBs. The regions between these PFBs appear to permit the recombination sites and therefore could be expected to exhibit either low polymorphism or a localized "hotspot."

Correspondence to: R.L. Dawkins; *e-mail:* dawkins@erythema.dci.uwa.edu.au

Publication number 9329 of the Centre for Molecular Immunology and Instrumentation, and the University of Western Australia, Perth Western Australia

Key words: Polymorphism — Recombination — Ancestral haplotypes — Major histocompatibility complex — *Homo sapiens*

Introduction

The major histocompatibility complex (MHC) has attributes that lend themselves to genomic and evolutionary analysis (Klein et al. 1991; Dawkins et al. 1991). For example some but not all HLA loci (class I and class II) are extremely polymorphic. Some of the non-HLA genes (e.g., *C4*) are also very polymorphic (Tokunaga et al. 1992). The MHC region of several megabases contains at least 100 genes, many of which are clustered within a segment of tens of kilobases and seem to have been subject to segmental duplication and deletion throughout their evolutionary history (Tokunaga et al. 1992; Zhang et al. 1993; Leelayuwat et al. 1995). Thus the number of segments varies within and between species (Tokunaga et al. 1992; Zhang et al. 1993). As a result of examining polymorphic loci, whether HLA or non-HLA, it appears that at least some alleles have been largely retained through sequential speciation (Klein et al. 1991) and the same may be true of genomic segments (Kay et al. 1988).

In this study we examine the evolution of the MHC region at the genomic level using polymorphic sequences from different ethnic-specific haplotypes. When different human ethnic groups are compared, it is apparent that HLA and non-HLA alleles may be shared, but the combinations of alleles at multiple MHC loci (haplotypes) can be classified depending upon whether they occur in

Table 1. List of cell lines used in the sequence, block profile analysis, and the accession numbers obtained from GenBank for the sequences shown in Fig. 1[a]

Haplotype	4AOH Number	Cell line	HLA designation	Sequence accession number
7.1	100040D	Q6/3975	A3; B7; C4A3; C4B1; DR15; DQ6	U28053
7.1		R6/12366		U46852
7.2	100042Z	R7/4708	A24; B7; C4A3+3; C4B1; DR1; DQ5	U26802
7.2	100043X	R7/12583		U26803
46.1		R9/52519	A2; B46; C4A4; C4B2; DR9; DQ9	U26804
46.1		R9/52523		U26805
46.1	100059F	R6/12361		U26806
46.2	100063R	R6/12351	A2; B46; C4A4; C4B2; DR8; DQ6	U26808
62.1	100072Q	R6/12316	A2; B62; C4A3; C4B3; DR4; DQ8	U28263
62.1	100073N	R6/12317		U28264
62.2	100074K	RO/26468	B62; C4A4; C4B2; DR4; DQ8	U46848
18.2	100051Y	R5/5054	A30; B18; C4A3; C4BQ0; DR3; DQ2	U28258
18.2	100036U	R6/12303		U28259
18.2	100211C	R6/12370		U28260
18.2		R1/3849		U46849
8.1		R2/8189	A1; B8; C4AQ0; C4B1; DR3; DQ2	U46850
8.1	100210E	R6/12308		U46851
8.1	100208R	R6/12373		U28256
8.1	100044V	R5/1518		U28257
57.1	100084G	R2/23091	A1; B57; C4A6; C4B1; DR7; DQ9	U46853

[a]Designation of the haplotypes is based on the integer representing the B allele and the decimal the order of discovery

one or more ethnic groups. For example, haplotype 57.1 appears to have a wide distribution whereas others such as 7.1 and 46.1 are ethnic specific and appear to be more recent (Kay et al. 1988; Tokunaga et al. 1992). For this reason, we propose that common MHC haplotypes are ancestral in that the entire genomic sequence of a haplotype has been maintained since the formation of that particular ethnic group or population (Degli-Esposti et al. 1992). In fact, it is now clear that the major ethnic-specific haplotypes carry specific genomic sequences (Abraham et al. 1991) in addition to carrying specific alleles at polymorphic loci. These haplotypes carry haplospecific sequence motifs, substitutions, deletions/insertions, and a specific gene content in regions where gene copy number varies.

These conserved polymorphic regions have been termed "polymorphic frozen blocks" (PFBs) (Klein et al. 1991; Dawkins et al. 1991). There are at least four blocks currently described: the alpha block, containing approximately 300 kb, including *HLA-A* and *HLA-G;* the beta block, spanning 250–300 kb, including *HLA-B* and *HLA-C;* the gamma block, covering 60–180 kb, including *C2, Bf, RP, C4, Cyp-21;* and the delta block, covering 60–300 kb, including *HLA-DR* and *HLA-DQ* (Marshall et al. 1993). Observed recombination events have occurred at the boundaries of these PFBs. Recombination appears to be inhibited by the level of polymorphism of the sequences within these blocks. This constraint on genetic exchange has also been observed in bacteria (Matic et al. 1995). By examining ethnic-specific haplotypes it can be shown that they have been formed by recombination between, and exchange of, these conserved PFBs. For example, by comparing 7.1 (Caucasoid) and 7.2 (Japanese) we can show that the beta block is shared but that other blocks differ. Similarly, 46.1 (Chinese) and 46.2 (Japanese) share the beta block but differ elsewhere. Thus the blocks define remote common ancestry and the mix of blocks (haplotypes) characterizes populations and ethnic groups.

Materials and Strategy

The aim of this study was to compare the conservation of different haplotypes. Therefore we needed to determine the degree of conservation within PFBs prior to considering whether these had been recombined to create new haplotypes. It was crucial to distinguish between the conservation of polymorphic sequences and the formation of haplotypes by shuffling of the conserved genomic sequences. Specifically, we aimed to determine whether the specific haplotypes shared some blocks but not others.

Primers were designed to amplify the CL region (Leelayuwat et al. 1992) centromeric of *HLA-B.* Since the sequences are polymorphic and duplicated, we have developed an assay, termed "block matching," for comparing the sequence of different haplotypes. The profiles observed with these primers were haplotypic (i.e., different individuals with the same haplotype exhibit the same pattern) (Abraham et al. 1992). To determine the conservation of some of these highly polymorphic regions we compared actual sequences. The sequence analyzed contains a direct duplication of a haplospecific geometric element which comprises complex patterns including dinucleotide repeats and is characterized by varying lengths of insertions and deletions in the element of different haplotypes (Abraham et al. 1992). The *TNF* microsatellites which are between the beta and gamma blocks, and which contain less complex sequences than these elements, have been shown to be relatively unstable and were not used in this study (Abraham et al. 1993; Udalova et al. 1993). The list of cell lines used and the corresponding sequence accession numbers are listed in Table 1.

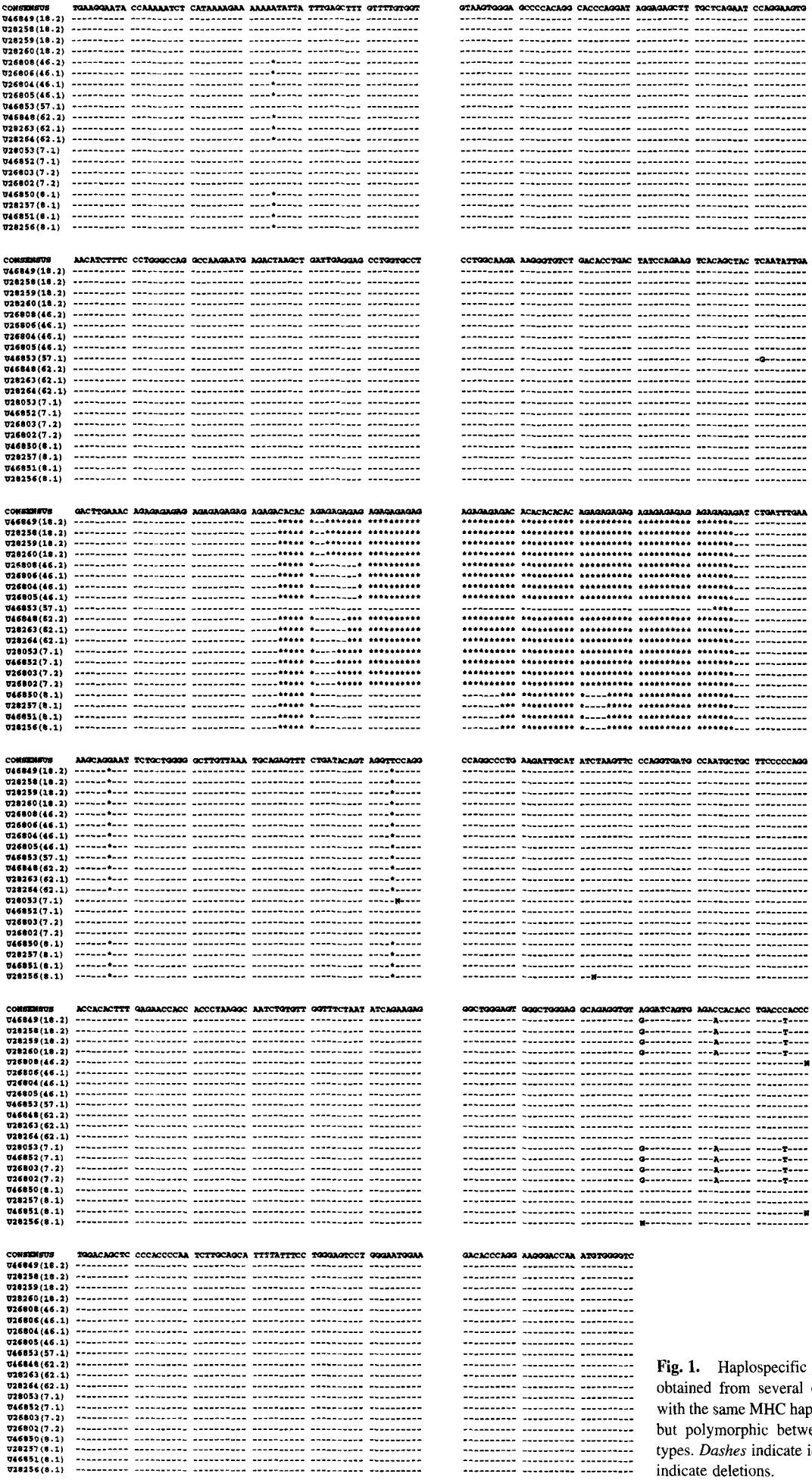

Fig. 1. Haplospecific geometric sequences obtained from several different individuals with the same MHC haplotype are conserved but polymorphic between different haplotypes. *Dashes* indicate identity and *asterisks* indicate deletions.

Methods

Sequencing. The method and primers have been described previously (Leelayuwat et al. 1993). Briefly, a 2.2-kb product in the 6.5-kb *Bam*HI fragment in the *CL* region was amplified as a template to differentiate the two copies of this sequence, using primers CMIIT950023 and CMIIT950025 (previously 65287B and 655782A respectively) in a 50-μl standard PCR for 1 min at 95°C, 30 s at 55°C, and 1 min at 72°C, with an autoextension of 2 s per cycle for 35 cycles. The resulting PCR product was amplified in a nested PCR amplification using CMIIT950026 (previously 65531B) and CMIIT950025 under the same conditions. The nested PCR product was purified by Centricon column centrifugation according to the manufacturer's recommendations (Centricon-100 microconcentrator; Amicon, W.R. Grace and Co., USA), and sequenced with the primer CMIIT950026 using a fluorescent-labeled dideoxy termination reaction on an automated 373A DNA sequencer (Applied Biosystems, Inc., USA). After blind manual editing, the accuracy of the sequence was confirmed by comparison with sequences for those haplotypes obtained by oligo-walking on pBCKS+ clones subcloned from lambda (Leelayuwat et al. 1994). To obtain the 3′ end and reverse strand of the repetitive sequence, a 750-bp fragment was amplified using the primers CMIIT950026 and CMIIT950027 (previously named 65617) in a 50-μl standard PCR for 2 min at 93°C for 1 cycle and then 1 min at 93°C, 30 s at 63°C, and then 1 min at 72°C with an autoextension of 2 s per cycle for 35 cycles. The resultant PCR fragment was extracted from the agarose gel and purified by agarase digestion (Agarase; Boehringer Mannheim, Germany) and sequenced using the primer CMIIT950027 as above.

Block Matching. Method and primers were developed to discriminate between haplotypes and have been used in bone marrow transplant matching and family genotyping (see Tay et al. 1995). Briefly, 200 ng of genomic DNA was amplified using the primers CMIIT950001 and CMIIT950002 (previously CLHGE3 and CLHGE4, respectively) in a 20-μl standard PCR for 2 min at 95°C for 1 cycle and then 10 s at 95°C, 10 s at 55°C, and 20 s at 72°C for 35 cycles. Five microliters of the resultant PCR product was run on a polyacrylamide gel (5% Long Ranger [Hydrolink] gel, AT Biochem, Malvern) and electrophoresed on a GS2000 Gelscanner (Corbett Research, Sydney, Australia). The densitometric scan from the resultant output was stored on a computer disk.

Extensive sequence differences can be marked by the beta block profiles. Accordingly, the same strategy was employed in a second region of the MHC, the delta block (Tay et al. 1995).

Results

Conservation of Polymorphic Sequences

Sequences obtained from the same haplotypes using different cell lines show that the geometric elements are conserved (Fig. 1). That is, the actual sequences at and between the typing primers show that different individuals with the same MHC haplotype are identical. The sequences differ between different MHC haplotypes by insertions/deletions and substitutions (Fig. 1).

There are several ways of presenting the polymorphic sequence data. For example 57.1 can be written as $(GA)_{12}$ $(CA)_3$ $(GA)_{14}$ $(CA)_6$ $(GA)_3$ $(GA)_9$ (Fig. 2). All examples share $(GA)_{12}$, again emphasizing that this complex sequence is different from simple microsatellites in which there would be expected to be differences of at least one dinucleotide of the $(GA)_{12}$. With the exception of 57.1, all examples lack $(CA)_3$ and $(CA)_6$. The differences between the remaining examples are a function of the number of GA dinucleotides, but the number appears to be fixed within each group (e.g., 46.1 and 46.2, 62.1 and 62.2, 7.1 and 7.2). The other sequence polymorphisms shown (Fig. 1) also appear to be fixed within haplotypes.

57.1	(GA)12	(CA)3	(GA)14	(CA)6	(GA)12
8.1	(GA)12	-	(GA)14	-	(GA)2
46.1	(GA)12	-	(GA)5	-	-
62.1	(GA)12	-	(GA)4	-	-
7.1	(GA)12	-	(GA)3	-	-
18.2	(GA)12	-	(GA)2	-	-

Fig. 2. Dinucleotide repeat structure in different haplotypes. Size differences between haplotypes are due to large insertions/deletions of dinucleotide repeats. *Dash* indicates no repeat present.

Conserved Polymorphic Sequences Predate Ethnic-Specific MHC Haplotypes

Given such remarkable conservation of sequence, we wished to compare the sequences from two haplotypes which are specific to Caucasoids (7.1) and Japanese (7.2). From earlier studies, we were aware that these two haplotypes appeared to be the same in the region around *HLA-B*. Both have *Cw7* and *B7* and a particular combination of restriction fragment length polymorphism patterns (Wu et al. 1992). The two haplotypes differ elsewhere in the MHC. Figure 1 shows sequences from both haplotypes (from several different individuals) are identical in insertions/deletions and substitutions. Therefore, these conserved polymorphic sequences predate the formation of the two ethnic-specific MHC haplotypes observed today.

The MHC haplotypes 46.1 and 46.2 are predominantly Chinese and Japanese, respectively. Figure 1 shows both haplotype sequences are identical in insertions/deletions and substitutions. As described above, these conserved polymorphic sequences predate the formation of the ethnic-specific haplotypes. In addition, haplotypes 7.1 and 7.2 are clearly different from the sequence observed from haplotypes 46.1 and 46.2 (Fig. 1).

Formation of Common MHC Haplotypes by Recombination of Polymorphic Frozen Blocks

Figure 3 illustrates the superimposed beta block profiles of haplotypes 7.1 and 7.2. As expected, the beta block profiles are identical between the two haplotypes. These profiles reflect the sequences shown in Fig. 1. Therefore, the beta block region predates the separation of these two ethnic-specific haplotypes.

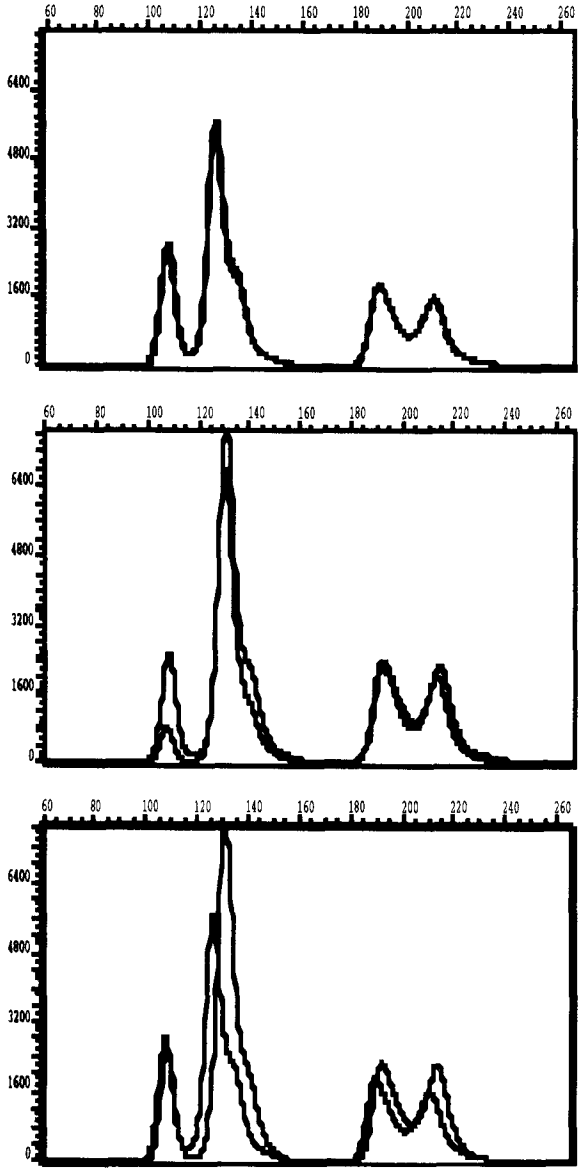

Fig. 3. The superimposed beta block profiles of the ethnic-specific haplotypes. The superimposed beta block profiles of 7.1 and 7.2 are indistinguishable as are the profiles of 46.1 and 46.2. However, a composite profile of the two groups of haplotypes 46.1, 46.2, and 7.1, 7.2 are distinctly different from one another.

Figure 3 also shows the superimposed beta block profiles of the two haplotypes 46.1 and 46.2. The profiles are identical and reflect the sequences shown in Fig. 1. Again, the beta block predates the separation of these two ethnic-specific haplotypes. Figure 3 also shows the superimposed beta block profiles of 7.1 and 46.1. The profiles are different and reflect the small sequence differences observed in Fig. 1.

This strategy was used in another region of the MHC, the delta block. Figure 4 shows the beta block and the delta block profiles of the 7.1 and 7.2 haplotypes. It is obvious that the delta block profiles are very different. Therefore, it would follow that there are considerable differences between the two haplotypes in this region.

The beta block and delta block profiles of haplotypes 46.1 and 46.2 are also shown in Fig. 4. Again, the delta block profiles are very different between the two haplotypes, and it could be assumed that large differences would also be observed between the two haplotypes in this region. By contrast, the beta block predates the separation of these ethnic-specific haplotypes.

Another example of shared beta blocks is provided by the two Caucasian haplotypes 62.1 and 62.2 which share the same sequence as shown in Fig. 1. They also share identical beta block profiles (data not shown).

Discussion

Polymorphic Frozen Blocks

For the purpose of this analysis, we have been able to demonstrate, using several different individuals with the

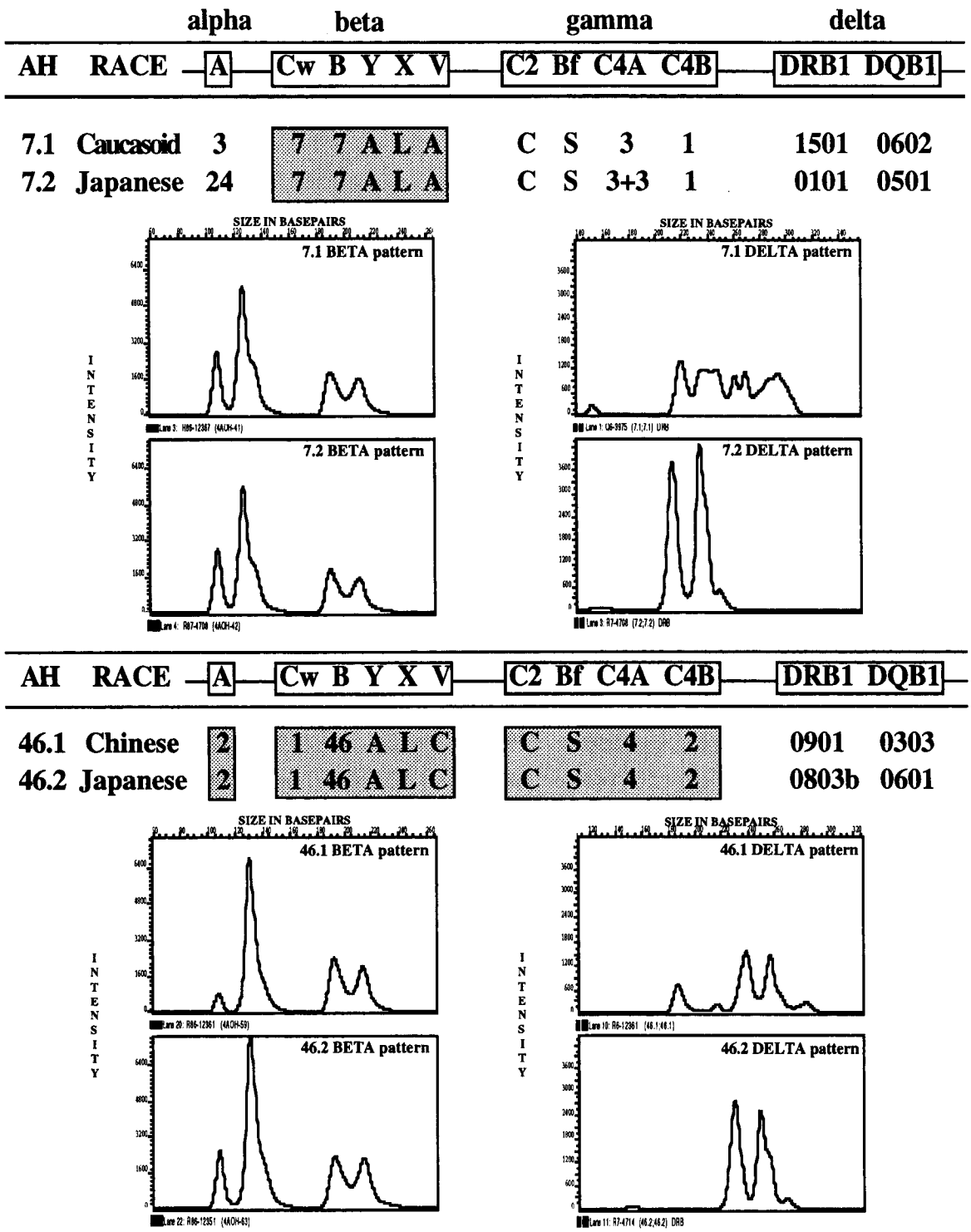

Fig. 4. The beta block and delta block profiles of the haplotypes 7.1, 7.2, 46.1, and 46.2. The beta block profiles for the ethnic specific haplotypes 7.1 and 7.2 are identical as they are for 46.1 and 46.2. The delta block profiles are very different between the ethnic-specific haplotypes. The *Y*, *X*, and *V* designations in the beta block are probes which have been used in earlier studies (Wu et al. 1992).

same haplotype, that polymorphic sequences centromeric of *HLA-B* are remarkably conserved for common MHC haplotypes. These conserved polymorphic sequences predate the formation of certain ethnic specific haplotypes and must therefore have been stable for periods in excess of 10^4 years (Cavalli-Sforza et al. 1988). With the exception of five sequencing ambiguities (two of which relate to CCC at positions 416 and 600 in Fig. 1) there is no evidence of a single intrahaplotype mutation within 690 bp. Given the magnitude of the sequence differences between haplotypes such as 57.1, 8.1, 7.1, 18.2, and 62.1, all of which are common in Caucasoids, there would appear to be three major possibilities.

Firstly, these haplospecific sequence differences must have accumulated over either a long period or at a rapid rate *before* the appearance of this ethnic group.

Secondly, the differences seen are not simply due to single base mutation or even dinucleotide slippage but to other events such as multibase insertion and deletion. Much of the sequence shown in Fig. 1 could be explained in terms of indels affecting one, six, 12, 18, or 24 bases.

Thirdly, polymorphic sequences have become "frozen" since separation of ethnic groups if not since speciation.

In fact, we favor a combination of these three phenomena and propose that haplospecific sequence differences largely predate separation of ethnic groups, reflect selection, and become "frozen" when a certain level of polymorphism is achieved. This suppression affects not only classical recombination and single base mutations but also mini indels of the type shown here.

Formation of Haplotypes

In spite of the conservation of sequence within polymorphic frozen blocks it appears that ethnic-specific haplotypes have been formed and/or selected relatively recently.

Using block profiles that reflect the actual sequences between the different haplotypes, we have been able to show that, in these examples, the beta block is the common block and the ethnic-specific haplotypes observed today have arisen via recombination between this and other polymorphic frozen blocks. No doubt other examples could have been chosen to demonstrate common alpha, gamma, or delta blocks (Zhang et al. 1990; Degli-Esposti et al. 1992). The shuffling of these blocks involves not simply the allele in the region but also the highly polymorphic sequences within the entire block. Further analysis is needed to determine the boundaries of these "frozen" blocks and to define the extent of cold or hot "spots" of recombination. It would be expected that sequences within a region of recombination would exhibit either a very low level of polymorphism or, if very localized, a specific "hotspot."

Acknowledgments. We would like to acknowledge Catena Causareno, Denise Witt, Linda Smith, and Renae Anderson for their technical assistance. This work has been supported by the National Health and Medical Research Council and the Immunogenetics Research Foundation grants. S. Gaudieri is the holder of an APRA (I) scholarship supported by AMRAD.

References

Abraham LJ, Du DC, Zahedi K, Dawkins RL, Whitehead AS (1991) Haplotypic polymorphisms of the *TNFB* gene. Immunogenetics 33: 50–53

Abraham LJ, Leelayuwat C, Grimsley G, Degli-Esposti MA, Mann A, Zhang WJ, Christiansen FT, Dawkins RL (1992) Sequence differences between *HLA-B* and *TNF* distinguish different MHC ancestral haplotypes. Tissue Antigens 39:117–121

Abraham LJ, Marley J, Nedospasov SA, Cambon-Thomsen A, Dawkins RL, Giphart M (1993) Microsatellite, FRCP, SSO typing of the *TNF* region; Comparisons on the 4AOH cell panel. Hum Immunol 38:17–23

Cavalli-Sforza LL, Piazza A, Menozzi P, Mountain J (1988) Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc Natl Acad Sci USA 85:6002–6006

Dawkins RL, Degli-Esposti MA, Abraham LJ, Zhang WJ, Christiansen FT (1991) Conservation versus polymorphism of the MHC in relation to transplantation, immune responses and autoimmune disease. In: Klein J, Klein D (eds) Molecular evolution of the major histocompatibility complex. Springer-Verlag, Heidelberg, pp 391–402

Degli-Esposti MA, Leaver AL, Christiansen FT, Witt CS, Abraham LJ, Dawkins RL (1992) Ancestral haplotypes: conserved population MHC haplotypes. Hum Immunol 34:242–252

Kay PH, Dawkins RL, Williamson J, Tokunaga K, Christiansen FT (1988) Coexistence of MHC chromosal segment marked by *HLA, B17, BfS, C4, A6, B1, DR7,* and *DQW9* in different ethnic groups. Hum Immunol 23:27–35

Klein J, O'Huigin C, Kasahara M, Vincek V, Klein D, Figueroa F (1991) Frozen haplotypes in Mhc evolution. In: Klein J, Klein D (eds) Molecular evolution of the major histocompatibility complex. Springer-Verlag, Berlin, pp 261–286

Leelayuwat C, Abraham LJ, Tabarias H, Christiansen FT, Dawkins RL (1992) Genomic organisation of a polymorphic duplicated region centromeric of HLA-B. Immunogenetics 36:208–212

Leelayuwat C, Zhang WJ, Townend DC, Gaudieri SG, Dawkins RL (1993) Differences in the central MHC between man and chimpanzee: implications for autoimmunity and acquired immune deficiency development. Hum Immunol 38:30–41

Leelayuwat C, Townend DC, Degli-Esposti MA, Abraham LJ, Dawkins RL (1994) A new polymorphic and multicopy MHC gene family related to nonmammalian class I. Immunogenetics 40:339–351

Leelayuwat C, Pinelli M, Dawkins RL (1995) Clustering of diverse replicated sequences in the major histocompatibility complex (MHC): evidence for en bloc duplication. J Immunol 155:692–698

Marshall B, Leelayuwat C, Degli-Esposti MA, Pinelli M, Abraham LJ, Dawkins RL (1993) New MHC genes. Hum Immunol 38:24–29

Matic I, Rayssiguier C, Radman M (1995) Interspecies gene exchange in bacteria: the role of SOS and mismatch repair systems in evolution of species. Cell 80:507–515

Tay GK, Witt CS, Christiansen FT, Charron D, Baker D, Hermann R, Smith LK, Diepeveen D, Mallal S, McCluskey J, Lester S, Loiseau P, Teisserence H, Chapman J, Tait B, Dawkins RL (1995) Matching for MHC haplotypes results in improved survival following unrelated bone marrow transplantation. Bone Marrow Transplant 15:381–385

Tokunaga K, Tanaka H, Mauff G et al. (1992) Complement studies of the Eleventh International Histocompatibility Workshop. In: Tsuji K, Aizawa M, Sasazuki T (eds) HLA 1991, vol 1. Oxford University Press, New York, pp 947–954

Udalova IA, Nedospasov SA, Webb GC, Chaplin DD, Turetskaya RL (1993) Highly informative typing of the human *TNF* locus using six adjacent polymorphic markers. Genomics 16:180–186

Wu X, Zhang WJ, Witt CS, Abraham LJ, Christiansen FT, Dawkins RL (1992) Haplospecific polymorphism between *HLA B* and tumor necrosis factor. Hum Immunol 33:89–97

Zhang WJ, Degli-Esposti MA, Cobain TJ, Cameron PU, Christiansen FT, Dawkins RL (1990) Difference in gene copy number carried by different MHC ancestral haplotypes: quantitation after physical separation of haplotypes by Pulsed Field Gel Electrophoresis. J Exp Med 171:2101–2114

Zhang WJ, Christiansen FT, Wu X, Abraham LJ, Giphart MJ, Dawkins RL (1993) Organization and evolution of *C4* and *CYP21* genes in primates: importance of genomic segments. Immunogenetics 37: 170–176