J Mol Evol (2001) 53:114–123
DOI: 10.1007/s002390010199

JOURNAL OF MOLECULAR EVOLUTION

The Association Between HLA-A Alleles and an *Alu* Dimorphism Near HLA-G

Jerzy K. Kulski,[1,2,4] Patricia Martinez,[1] Natalie Longman-Jacobsen,[1] Wei Wang,[3] Joseph Williamson,[1] Roger L. Dawkins,[1] Takashi Shiina,[2] Taeko Naruse,[2] Hidetoshi Inoko[2]

[1] Centre for Molecular Immunology and Instrumentation, The University of Western Australia, Nedlands, 6009, Western Australia, Australia
[2] Department of Genetic Information, Division of Molecular Life Science, Tokai University School of Medicine, Bohseidai, Isehara, Kanagawa, 259–1193, Japan
[3] Centre for Human Genetics, Edith Cowan University, Joondalup, 6027, Western Australia, Australia
[4] Centre for Bioinformatics and Biological Computing, School of Information Technology, Murdoch University, Murdoch, 6150, Western Australia, Australia

Received: 7 December 2000 / Accepted: 28 February 2001

Abstract. The *Alu*Yb8 sequences are a subfamily of short interspersed *Alu* retroelements that have been amplified within the human genome during recent evolutionary time and are useful polymorphic markers for studies on the origin of human populations. We have identified a new member of the Yb8 subfamily, *Alu*yHG, located between the HLA-H and -G genes and 88-kb telomeric of the highly polymorphic HLA-A gene within the alpha block of the major histocompatibility complex (MHC). The *Alu*yHG element was characterised with a view to examining the association between *Alu*yHG and HLA-A polymorphism and reconstructing the history of the MHC alpha block. A specific primer pair was designed for a simple PCR assay to detect the absence or presence (dimorphism) of the *Alu*yHG element within the DNA samples prepared from a panel of 46 homozygous cell-lines containing complete or recombinant ancestral haplotypes (AH) of diverse ethnic origin and 92 Caucasoid and Asian subjects on which HLA-A typing was available. The *Alu*yHG insertion was most strongly associated with HLA-A2 and, to a lesser degree with HLA-A1, -A3, -A11, and A-19. The gene frequency of the *Alu*yHG insertion for 146 Caucasians and 94 Chinese-Han was 0.30 and 0.32 and there was no significant difference between the observed and expected frequencies. The results of the association studies and the phylogenetic analysis of HLA-A alleles suggest that the *Alu*yHG sequence was integrated within the progenitor of HLA-A2, but has been transferred by recombination to other human ancestral populations. In this regard, the dimorphic *Alu*yHG element is an important diagnostic marker for HLA association studies and could help in elucidating the evolution and functions of the MHC alpha block and polymorphism within and between ancestral haplotypes.

Key words: *Alu* — HLA-A alleles — Polymorphism — Haplotypes — Major histocompatibility complex (MHC)

Introduction

Alu sequences are a class of short interspersed retroelements that contribute to 10% of the human genomic content (Smit 1996). They have played a significant role in generating DNA sequence diversity during primate evolution and appear to have evolved from the 7SL RNA gene (Ullu and Tschudi 1984). Different *Alu* subfamilies have developed together with primate lineage history by retrotransposition via a RNA polymerase III-derived transcript through successive waves of amplification and fixation (Quentin 1988). Thus, at least three fundamental

Correspondence to: Jerzy K. Kulski; *email:* jkulski@murdoch.edu.au

categories of *Alu* sequences (*Alu*J, *Alu*S, and *Alu*Y) are recognized based upon commonly shared diagnostic mutations, genetic ages, and sequence differences (Jurka and Smith 1988; Batzer et al. 1996b; Kapitonov and Jurka 1996). In this regard, *Alu* subfamily sequences have been used successfully as molecular clocks to elucidate the genomic organization and evolutionary history of duplicated regions such as those consisting of the multicopy HLA class I gene family within the MHC (Kulski et al. 1999, 2000a, 2000b). Comparative genomic analysis between different HLA haplotypes has also revealed insertion/deletion events involving relatively modern *Alu* sequences (Gaudieri et al. 1997a, 1997b; Kulski et al. 1997, 1999, 2000b).

Members of the *Alu*Y subfamilies are potentially useful candidates for investigating the origins of human ancestral haplotypes, ethnic groups, and disease associations. They first emerged in primate history about 19 million years ago with about six *Alu*Y subfamilies now recognized in the human genome (Batzer et al. 1996b; Kapitonov and Jurka 1996; Schmid 1996). One of the young *Alu* subfamily members, *Alu*Yb8, appears to be human specific and transpositionally mobile (Zietkiewicz et al. 1994). It has been associated with different human genetic diseases (Miki, 1998; Deininger and Batzer, 1999), including acholinesteraemia (Muratani et al. 1991) and Huntington's disease (Goldberg et al. 1993). There are 500–1500 members of the *Alu*Yb8 subfamily distributed within the human genome where some have been fixed (Zietkiewicz et al. 1994; Batzer et al. 1995) and others are dimorphic and have been used in population studies to differentiate between ethnic groups and to trace the migration patterns of modern humans (Hammer 1994; Stoneking et al. 1997). However, little information is available on the evolutionary relationships between dimorphic *Alu*Yb8 subfamily members and the founder human haplotypes.

The alpha block within the MHC is one of at least six blocks that describe founder haplotypes within different population and ethnic groups, and disease associations (Dawkins et al. 1999). The classical class I HLA-A gene has evolved into one of the most polymorphic sequences within the alpha block whereas the closely-linked nonclassical class I genes HLA-F and -G have remained fairly conserved (Gaudieri et al. 2000). Consequently, the HLA-A gene is the preferred locus within the alpha block for routine HLA typing with the identification of at least 214 HLA-A alleles see IMGT/HLA statistics at www.ebi.ac.UK/imgt/hla/). During a comparative analysis of 319 kb of genomic sequence from the alpha block of the HLA class I region of the MHC (Kulski et al. 1999), we identified a new member of the *Alu*Yb8 subfamily, designated here as *Aluy*HG, located between the HLA-H and -G genes, approximately 88-kb telomeric of the HLA-A gene.

To elucidate the relationship between *Aluy*HG and the nearby HLA-A locus, we examined (1) the *Aluy*HG dimorphism (absent or present) in a panel of homozygous cell-lines with well-defined HLA haplotypes, and (2) the frequency of association between *Aluy*HG dimorphism and different HLA-A alleles in Caucasians and Asians.

Material and Methods

Cell-Lines, Blood Samples, and Preparation of Genomic DNA for PCR. The cell-lines used for the analysis of the *Aluy*HG insertion included Epstein-Barr virus-transformed human and chimpanzee B-cell-lines from the Fourth Asia-Oceania Histocompatibility Workshop (4AOH; Degli-Esposti et al. 1993) and the Tenth International Histocompatibility Workshop (10IHW; Prasad and Yang 1996), and ten cell-lines from a collection of homozygous Japanese typing cells (Naruse et al. 1998). The panel of cells shown in Table 1 covered mainly homozygous ancestral haplotypes with known disease associations. The AH of each cell-line is named according to the HLA-B allele and the order in which they were first described, X and Y are used when non-HLA-B alleles differ from the described AH (Degli-Esposti et al. 1993; Dawkins et al. 1999). Cells were cultured in RPMI 1640 with 10% fetal calf serum (37°C, 5% CO_2).

We also examined the DNA extracted from the peripheral blood of 191 Caucasians and Asians including 92 (73 Caucasians and 19 Asians) who had been typed for different HLA alleles. Forty-seven Chinese-Han, randomly selected from Liaoning Province in northeast China, and 52 Asians (Malays, Indians, and Malay-Chinese) from Malaysia were not typed for HLA alleles. Genomic DNA was prepared for PCR from peripheral blood and cell-lines by using the combination of proteinase K digestion and a salting-out method (Miller et al. 1988). HLA typing was performed by standard microlymphocytotoxicity assays.

PCR for Aluy*HG.* The sense and antisense PCR primers were *Alu*YF11 (5′-CAGGACAACCAGTAAAGAT GCTGG-3′) and *Alu*YR11 (5′-GCTTCAGTTAACATGCAAGTTTATGCC-3′), respectively (Pacific Oligos Pty. Ltd. Toowong, Australia). The *Alu*YF11 and *Alu*YR11 primer sequences were designed by reference to the MHC genomic sequence (Hampe et al. 1999) that has the *Aluy*HG sequence from nucleotide position 183727–184267 in GenBank accession number AF055066. The PCR was carried out in 50 μl containing 500 ng of template DNA, 0.5 pmol of each primer, 0.8 μmol dNTPs, 1.0 unit of AmpliTaq polymerase (Perkin Elmer), 3 mM $MgCl_2$ in 10 mM Tris-HCI buffer pH 8.3. Thirty or thirty-five cycles were performed in a Perkin Elmer 480 Thermal cycler with each cycle consisting of denaturation at 94 °C for 30 s annealing at 60 °C for 30 s and extension at 72 °C for 60 s. Following completion of PCR, 10 μl of the reaction products were analyzed by horizontal gel electrophoresis in 2% agarose using Tris-borate running buffer. The PCR products were visualized by staining with ethidium bromide and compared to molecular size markers (Gibco 1 Kb and 1 Kb DNA ladder from Biotechnologies) that were included with each run. The expected product sizes were 218 bp for the absence and 540 bp for the presence of the *Aluy*HG insertion. A positive (AH 62.1) and negative (AH 8.1) control sample were included in each PCR.

Sequence Confirmation of PCR Products as Aluy*HG.* PCR products were cloned into sequencing vector pCR2.1-TOPO (Invitrogen) and DNA sequencing was performed by the cycle sequencing method using AmpliTaq-DNA polymerase FS (PE Applied Biosystems, CA) and Bigdye terminators in a GeneAmp PCR system. The 373S or 377 Prism DNA sequencers. (PE Applied Biosytems, Foster City, CA) were used for automated fluorescence determination of sequence (Shiina et al. 1999).

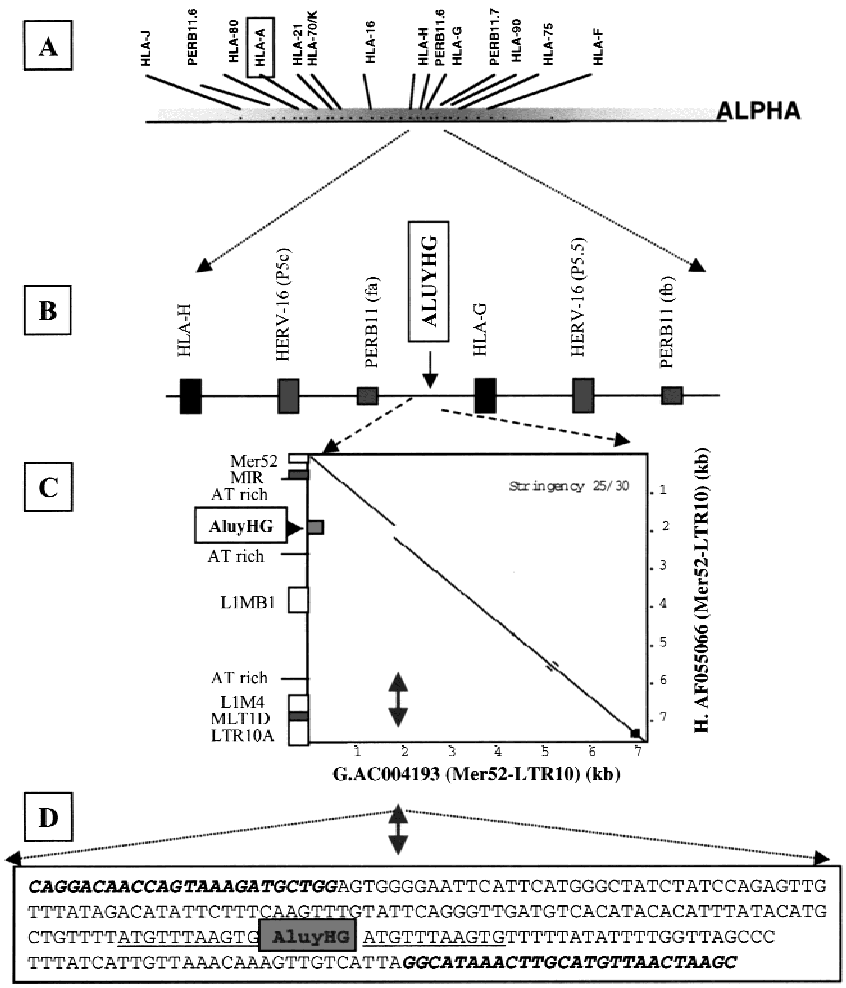

Fig. 1. Location of *Aluy*HG dimorphism in alpha block of MHC. (**A**) is a map of the location and distribution of the HLA class I coding and non-coding genes, including HLA-A (boxed) within the alpha block. (**B**) is a map of the location of the *Aluy*HG insertion site relative to the HLA-H and -G loci. (**C**) is a dot plot comparison of two genomic sequences showing the region of the *Aluy*HG insertion where one sequence (AC004193) lacks the *Aluy*HG and the other (AF055066) has the *Aluy*HG insertion. The gap in the diagonal line of the matrix shows the position of the *Aluy*HG insertion. The location of retroelements relative to the *Aluy*HG insertion site (shaded box inside the matrix of the Y-axis) that are shared by the two 7kb-sequences were identified by RepeatMasker and are labeled and indicated by the boxes on the left-hand side of the Y-axis. (**D**) shows the sequences in AC004192 and AF055066 used to design the 5′ and 3′ primers (blocked and in *italics*) for the *Aluy*HG PCR assay. The *Aluy*HG sequence (in reverse orientation) is represented by the labeled box that is flanked by the direct repeat sequence (*underlined*).

DNA Sequence Analysis, Construction of HLA-A Phylogenetic Tree, and Statistical Analysis. Genomic sequence for the analysis of *Alu*Yb8 elements in the alpha block were obtained from DDBJ/EMBI/GenBank accession numbers AC004193 and AF055066. Dot-plot matrix comparisons of sequences were performed using the program, Matrix, within Gene Jockey, Version 2 and other programs previously described (Kulski et al. 1999). Repeat elements, such as *Alu* subfamily members, L1, LTRs, and MERs were identified using the program RepeatMasker (http://ftp.genome.washington.edu/cgi-bin/RepeatMasker).

Aligned HLA-A cDNA sequences were obtained from the Anthony Nolan Bone Marrow Trust web page (http://www.anthonynolan.org.uk/HIG/seq/nuc/text/a.nt.txt). A total of 70 alleles for which complete information was available up to April 2000 were used in the initial analysis. Gogo (*Gorilla gorilla*), Patr (*Pan troglodytes*), Popy (*Pongo pygmaeus*), and Mamu (*Macaca mulatta*) cDNA sequences were obtained from GenBank (accession numbers X54376, X60257, X60259, X60258, AF168404, AF168401, AF168395, AF168394, AF168393, AF168392, U50084, U50085, U41832). These non-human primate sequences were aligned using the CLUSTAL W 1.8 program on the Baylor College of Medicine website (http://dot.imgen.bcm.tmc.edu.9331/multi-align/multi-align.html) and added to the alignment of human alleles. All invariant positions were removed to obtain a final alignment containing only positions that differed between alleles.

The phylogenetic tree was constructed using the Molecular Evolutionary Genetics Analysis (MEGA) package developed by Kumar et al. (1994), utilizing the Jukes-Cantor distance calculations and the neighbor-joining method. Bootstrap values from 500 iterations were also calculated. In order to simplify the resultant tree, some alleles were removed from groups for which there were multiple examples that clustered very closely together. The tree was then regenerated from the modified alignment containing 44 alleles, and bootstrap values recalculated. The graphic presentation of the phylogenetic tree was produced with TreeExplorer version 2.12 (Tamura 1997, unpublished).

Gene frequency, heterozygosity, and Hardy-Weinberg equilibrium tests were performed using the software programs provided by Genepop (Raymond and Rousset, 1995).

HLA-A Allele Nomenclature. HLA-A alleles have been grouped into at least six families A1/3/11, A9, A2/28, A10, A19, and A80 on the basis of serology and genetic relatedness (Kato et al. 1989; McKenzie et al. 1999). The HLA-A19 serological family includes HLA-A29, -A30, -A31, -A32, and -A33, although the HLA-A30 allele (HLA-A19^ group) is more closely related by nucleotide sequence to the A1/A3/A11 group. The reference to the HLA-A19 allele in this paper means we could not distinguish serologically between HLA-A29, -A30, -A31, -A32, or -A33.

Results

Identification of the Aluy*HG Insertion Located Between HLA-H and -G Duplicons*

Figure 1 shows a genomic map (**A** and **B**), dot plot (**C**), and nucleotide sequence (**D**) of the *Aluy*HG insertion in

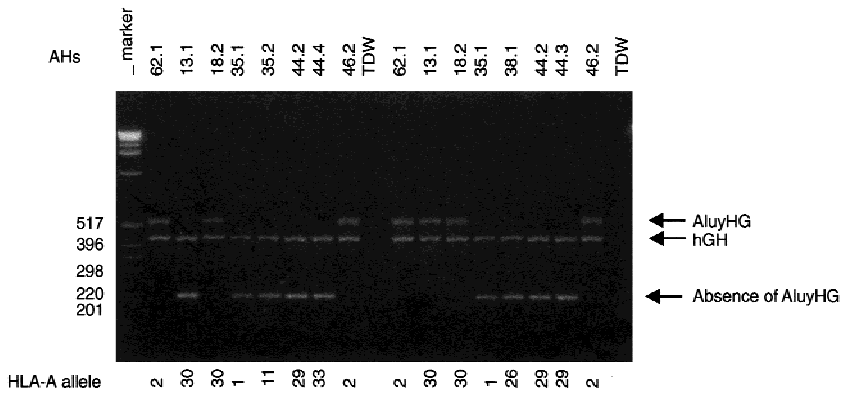

Fig. 2. Amplification of dimorphic *Aluy*HG element from the alpha block of different Ancestral Haplotypes (AHs). The presence and absence of *Aluy*HG are seen as 530 bp and 218 bp PCR products, respectively. The hGH 440 bp-band is an amplified growth hormone sequence used as an internal control for DNA integrity. Lanes 1–19 run horizontally across the gel with lane 1 labeled as the marker on the left-hand side of the figure. The numbers above each lane are the AHs. Those below each lane are the HLA-A alleles. Numbers on the left-hand side of the figure are molecular size markers (bp).

the BOLETH cell-line (GenBank accession number AF055066) compared to a sequence without the *Alu* element (GenBank accession number AC004193). The *Aluy*HG sequence has been inserted in the 3′ to 5′ (telomeric to centromeric) direction in close proximity to an AT-rich region between the remnants of a PERB11 (MIC) fragment and the 3′ end of the HLA-G gene (Fig. 1A, B) and it is flanked by the direct repeat sequence 5′-ATGTTTAAGTG-3′ (Fig. 1D). Nucleotide sequences flanking the *Aluy*HG insertion (Fig. 1C, D) were used to design the primers for the *Aluy*HG PCR assay and the expected sequence of the PCR products (218 bp) without the *Alu* insertion is shown in Fig 1D.

Aluy*HG PCR and Sequencing*

Figure 2 shows the gel electrophoresis results of the *Alu* and growth hormone (GH) PCR products amplified from the DNA of cell-lines with known human ancestral haplotypes (AHs). The *Alu* PCR produced the expected product sizes of 218 bp (absence of *Aluy*HG) and 530 bp (presence of *Aluy*HG) in human cell-lines (Fig. 2) but not in the macaque or any of the ten chimpanzee cell-lines (data not shown). By contrast, the GH sequence of 420 bp was amplified by PCR in all of the human, macaque, and chimpanzee cell-lines that were tested. The absence of *Aluy*HG amplification from the macaque and chimpanzee target DNA suggests that the *Alu* primers could not anneal to the target DNA because of mutations or deletions within the non human primate sequences. The *Aluy*HG element was detected in human cell-lines that were previously typed for HLA-A2 (AHs 18.3, 46.2, and 62.1) and -A30 alleles (AH 18.2). However, the *Aluy*HG element was absent from one of the AH 13.1 cell-lines that contains the HLA-A30 allele (Fig. 2, lane 3).

Fifteen PCR samples that were positive for *Aluy*HG were cloned and sequenced. Twelve of the fifteen *Aluy*HG sequences were associated with HLA-A2, whereas the others were associated with HLA-A3 (a Malaysian-Chinese subject with HLA-B38), homozygous HLA-A30 (Table 1, cell number 17), and the heterozygous HLA-A2 and -A19 alleles (Table 1, cell number 15). All of the sequenced PCR products were *Aluy*HG, therefore confirming the specificity of the PCR assay. The *Aluy*HG sequence has diverged from the *Alu*Yb8 consensus sequence (Jurka 1993) and 20 other members of the *Alu*Yb8 subfamily (data not shown) by 0.70–0.77%, with the rate of transversions approximately three times greater than transitions. In addition, the *Aluy*HG sequence lacks the first 11 nucleotides (5′-GGCCGGGCGCG-3′) usually found within the other *Alu*Yb8 subfamily members.

The Association Between the Aluy*HG Dimorphism and the HLA-A2 Allele*

Table 1 shows the *Aluy*HG results for 46 cell-lines that have been typed for different MHC alleles, including HLA-A, -B, -DR, and -DQ, and represent ancestral haplotypes (AH) (cell numbers 1–36) and Japanese cell-lines (cell numbers 37–46). All of the cell-lines were homozygous at the HLA-A locus with the exception of cell numbers 15, 22, 26, and 35. Of the 20 cell-lines that were homozygous for HLA-A2, 17 (85%) were homozygous and 2 (10%) were heterozygous for the presence of *Aluy*HG (cell numbers 13 and 14). One cell-line (cell number 15) heterozygous for HLA-A2/-A19 was homozygous for the presence of *Aluy*HG, whereas one cell-line (cell number 16, AH 46.1) homozygous for the HLA-A2 allele had no detectable *Alu* insertion. With the exception of HLA-A19 allelic group, the *Aluy*HG sequence was absent from all of the other non-HLA-A2 cell-lines. The *Aluy*HG was present as a homozygote in two homozygous HLA-A30 cell-lines (cell numbers 17 and 18) but absent from the other homozygous and heterozygous HLA-A19 allelic groups (cell numbers 32–36, 45 and 46). Therefore, on the basis of this, albeit, limited panel of homologous cell-lines, the *Aluy*HG insertion

Table 1. HLA and *Alu*yHG alleles in 46 different homozygous cell-lines

Cell No.	AH	Ethnicity	HLA-A	*Alu*yHG	HLA-B	HLA-DR	HLA-DQ	10WS	4AOH	JAPCL
1	13.1X	C/M	2	+	13	7	2	9093	100184C	
2	18.3	C	2	+	18	11	7	9039		
3	18.X	C	2	+	18.1	12	7	9038	100308M	
4	18.Z	C	2	+	18	14	5	9061		
5	46.1	M	2	+	46	9	3	9076	100059F	
6	46.2	M	2	+	46	8	6	9066	100063R	
7	46.2	M	2	+	46	8	1		100031E	
8	51.X	C	2	+	51	8	7	9070	100140Z	
9	57.1/X	C/M/N	2	+	57	7	9	9052	100084G	
10	62.1	C	2	+	75	4	8	9031	100072Q	
11	44.X	C	2	+	44	11	5	9036		
12			2	+	27	8	4	9067		
13	44.Y	C	2	+, −	44	14		9054		
14			2	+, −	27	16	5			
15	18.2/X	C	2, 19	+	18	3			100051Y	
16	46.1	M	2	−	46	9				
17	18.2	C	30	+	18	3	2	9085	100211C	
18	18.2	C	30	+	18	3	2		100112E	
19	8.1	C	1	−	8	3	2	9023	100210E	
20	35.1	C	1	−	35	11	7	9041	100019U	
21	57.1	C/M/N	1	−	17	7	3		100083J	
22	8.1/X	C	1, 9	−	8	3			100044V	
23	35.2	C	11	−	35	1	5	9006	100052W	
24	7.1	C	3	−	7	15	6	9082	100041B	
25			3	−	27	1	5			
26	18.2X/Y	C	3, 24	−	18	3	2	9018	100036U	
27	7.2	M	24	−	7	1	1		100042Z	
28	7.2	M	24	−	7	1	1		100043X	
29	35.4	C	24	−	35	11	7	9042	100018W	
30	18.1	C	25	−	18	15	6	9047	100064P	
31	38.1	C	26	−	38	4	8	9026	100009X	
32	44.2	C	29	−	44	7	2	9051	100057K	
33	44.3	M	29	−	44	7	2	9050	100058H	
34	13.1	C/M	30	−	13	7	2	9096	100202D	
35	42.1/X	N	30, 68	−	42	3	4	9021	100007B	
36	44.4	M	33	−	44	13	6	9053	100145P	
37			2	+	61	4	4			HY418
38			2	+	61	9	3			HY2566
39			2	+	61	9	3			HY595
40			2	+	61	9	3			HY3339
41			2	+	62	14	3			HY2176
42			3	−	24	15	6			HY435
43			24	−	52	15	6			HY628
44			24	−	52	15	6			HY1092
45			33	−	44	13	6			HY870
46			33	−	44	13	6			HY1417

C, caucasian; M, mongoloid; N, negroid.
+ is Alu presence and − is Alu absence.

is primarily associated with HLA-A2 and, to a lesser degree, with the HLA-A19 allelic group. The HLA-A2 subtypes in the cell-lines 1–12 were almost exclusively HLA-A*02011 and -A*0207, except that cell-line 5 was heterozygous for HLA-A*0206/0207. In the Japanese cell-lines, the HLA-A2 subtypes were A*0206 in cell 41, A*0210 in cell 37, and A*02011 in cells 38–40.

The association between *Alu*yHG dimorphism and HLA-A alleles was also determined for 92 Caucasians and Asians with 51 (55.4%) individuals having at least one A2 allele. Overall, 47 of 51 (92.2%) individuals with the A2 allele, versus 4 of 41 (9.8%) without the A2 allele, had the *Alu*yHG insertion. Of the four *Alu*-positive individuals without the A2 allele, three individuals had only one allele identified (one with HLA-A1 and two with HLA-A3) and one Japanese was heterozygous for HLA-A3 and -A11. In addition, two individuals who were heterozygous for HLA-A alleles (HLA-A2 and -A11) were homozygous for the presence of the *Alu*yHG insertion. The DNA from the blood of three Japanese showed either homozygosity or heterozygosity for *Alu*yHG. The *Alu*yHG heterozygosity was present in two

Table 2. Percentage frequency of association between *Alu*yHG$^+$ and HLA-A alleles

HLA-A Allele	*Alu*$^+$ Number	*Alu*$^-$ Number	Total Number	% Frequency	Category
A2	90	12	102	88.2	I
A1	1	40	41	2.4	II
A3	3	17	20	15	II
A11	3	18	21	14.3	II
A19	6	27	33	18.2	II
A9	0	35	35	0	III
A10	0	18	18	0	III
A68	0	1	1	0	III
Total	**103**	**168**	**271**	**38.0**	

This table combines the results obtained for the 73 Caucasians, 16 Asian-Malay, 3 Japanese, and the 43 cell-lines (excluding the 3 HLA-B27 cell-lines in Table 1) as described in text. Category I refers to alleles where the majority of a large number of samples have the *Alu*yHG (Alu+). Category II refers to alleles in which the majority of examples do not have the *Alu*yHG (Alu-). Category III refers to alleles in which all samples lack the *Alu*yHG sequence. HLA-A19 includes the HLA-A alleles -A29, -A30, -A31, -A32, -A33, and -A74; HLA-A9 includes HLA-23 and -24; and HLA-10 includes HLA-A25 and -A26.

Japanese with the HLA-A locus antigens, HLA-A2 and -A26, and HLA-A2 and -A24, respectively. One of the Japanese who was homozygote for the presence of *Alu*yHG was identified with the HLA-A3 and -A11 alleles.

The gene frequency of the *Alu*yHG insertion for 146 Caucasians and 94 Chinese-Han was 0.30 and 0.32, respectively. The *p* values for the Caucasians and Asians were 0.1534 and 0.999, respectively. Both *p* values are $P > 0.05$ showing no deviation from the Hardy-Weinberg equilibrium, and therefore confirming that this *Alu* element is distributed normally in the investigated populations.

Table 2 shows a summary of the overall frequency of association between *Alu*yHG presence and the HLA-A alleles after combining all of the study groups, including the cell-lines and the Caucasian, Asian, and Japanese individuals. The frequency of association between *Alu*yHG presence and HLA-A alleles was divided into three categories. Category I represents the high-frequency group where the majority of *Alu*yHG-positive samples (90 of 102, 88.2%) were associated with HLA-A2. Category II represents the low-frequency (13 of 115, 11.3%) groupings where the majority of HLA-A allelic groups lack the *Alu*yHG element. Category III represents the HLA allelic groups that were associated solely with the *Alu*yHG-negative genotype.

Phylogenetic Tree of Sequence Relationship Between Known HLA-A Alleles and Their Association with Alu*yHG*

To gain a better understanding of the origin of the *Alu*yHG insertion in association with the evolution of HLA-A alleles, we reconstructed a phylogenetic tree of different human, chimpanzee, and gorilla HLA-A alleles using a macaque sequence as an out-group. Figure 3 shows that the human HLA-A alleles within the tree have clustered separately into at least seven distinct groups (A groupings) with the *Patr* alleles clustering most closely with the human A3, A9, and A19ˆ group, and the *Gogo* alleles clustering with the A2, A10, A19*, and A28 group. The HLA-A groups were further subdivided into the three distinct categories of *Alu*yHG association frequency, where category I represents the high frequency of association between *Alu*yHG elements and HLA-A2, category II represents the low frequency of association between *Alu*yHG and the HLA-A3 and -A19 alleles, and category III is the HLA-A allelic groups that have no positive association with the presence of *Alu*yHG. On the phylogenetic tree, the HLA-A3 and -A19 groups that are associated with *Alu*yHG at a low frequency are well-separated from each other and from the HLA-A2 group that is associated at a high frequency with *Alu*yHG (Fig. 3). The distance measures in the phylogenetic tree suggest that the HLA-A3 and -A19 group may have originated before the HLA-A2 group.

A number of potential scenarios can be inferred about the evolution of the *Alu*yHG insertion from the phylogenetic tree shown in Fig. 3. The *Alu*yHG was first inserted into an individual belonging to the founding population of HLA-A2 alleles, and the association between the non-HLA-A2 alleles and the *Alu*yHG positive genotypes arose in some individuals after cross-over events between the chromosomes carrying the HLA-A2 and the non-HLA-A2 alleles. Alternatively, the *Alu*yHG was first associated with a non-HLA-A2 allelic group at a low frequency and then either the HLA-A2 allele evolved from the non-HLA-A2 allelic group carrying the *Alu*yHG insertion, or the *Alu*yHG was transferred by recombination to an individual belonging to the founding population of HLA-A2 alleles. The *Alu*yHG element may have been lost from some heterozygous individuals by cross-over or sequence exchange between the chromosomes carrying the HLA-A2 allele and the non-HLA-A2 allele. These different scenarios might be better resolved in the future by more extensive population studies, such as those undertaken within the International Histocompatibility Workshops, and by also examining the association of *Alu*yHG dimorphism with the HLA-G locus.

Discussion

This is the first study to examine the relationship between a particular member of the *Alu*Yb8 subfamily and HLA polymorphism within the MHC. We have used cell-lines that have been previously typed for polymorphic markers at different loci and define known ancestral

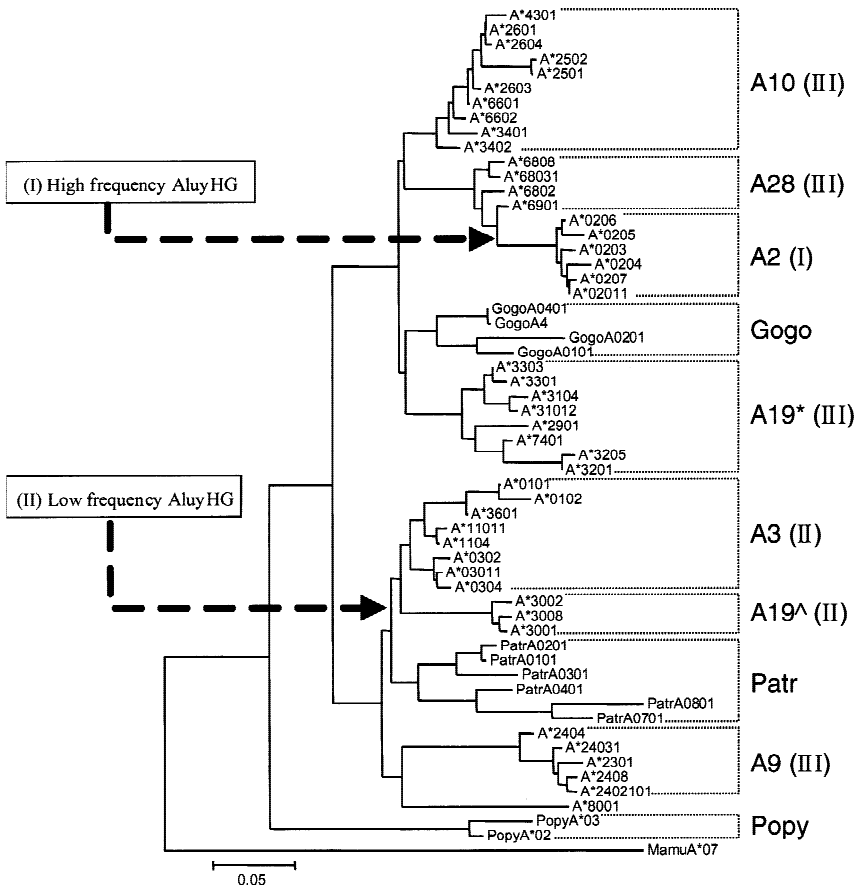

Fig. 3. Phylogenetic tree of HLA-A alleles and association with *Aluy*HG. The phylogenetic tree was constructed using the Jukes-Cantor distance calculations and the neighbor-joining method. Minimum evolution analysis of the complete cDNA sequences of HLA-A alleles (seven groupings labeled A2, A3, A9, A10, A19*, A19^, and A28) in comparison to the Gogo (*Gorilla gorilla*), Patr (*Pan troglodytes*), Popy (*Pongo pygmaeus*), and Mamu (*Macaca mulatta*) alleles. The high and low frequency association between *Aluy*HG insertion and HLA-A alleles are labeled and indicated by arrows. (I) represents the high-frequency association between *Aluy*HG insertion and HLA-A2. (II) represents the lower-frequency association between *Aluy*HG insertion and HLA-A groups A3, A19*, A19^. (III) has no positive association with *Aluy*HG. Branch lengths represent estimated percentage sequence divergence.

and ethnic haplotypes (Degli-Esposti et al. 1993; Prasad and Yang 1996; Naruse et al. 1998; Cattley et al 2000), and the DNA of Caucasians and Asians previously typed for HLA-A, -B, and -DR alleles. We found that the *Aluy*HG dimorphism in Caucasians was in strong linkage disequilibrium with the HLA-A locus within the alpha block, but not with the HLA-B locus within the beta block and not with the HLA-DR locus within the gamma block (data not shown). In Asian-Malays, the *Aluy*HG dimorphism was in strong linkage disequilibrium with both the HLA-A and -B loci (data not shown), reflecting the strong linkage disequilibrium between HLA-A and -B in this group of individuals. However, in this study, we have focused mainly on the relationship between the *Aluy*HG dimorphism and the HLA-A locus within the alpha block.

The degree of polymorphism within and around the HLA-A locus is extremely high when compared to the region of the *Aluy*HG insertion that is located between the HLA-H and -G genes (Gaudieri et al. 2000). It is believed that recombination or intergenomic exchange within the alpha block has been suppressed during the last 200,000 years of human migration (Dawkins et al. 1999). Therefore, if the *Aluy*HG element was inserted as a single event within the MHC at a time well after the generation of most AHs, then the *Aluy*HG should be associated with either one or a small number of closely interrelated HLA-A allelic groups. On the other hand, if the *Aluy*HG was inserted in a founding AH well before the generation of most of the other AHs, then the *Alu* element should be found associated at relatively high frequency with many different HLA-A allelic groups. Of the eight different HLA-A alleles that were typed in the present study, the *Aluy*HG integration was most frequently (at least 88%) associated with HLA-A2. Moreover, the HLA-A2-*Aluy*HG-positive phenotype appears to have flourished significantly over the HLA-A2-*Aluy*HG-negative phenotype in the different ethnic groups (Caucasian, Chinese, Malay, and Japanese). This suggests that the *Aluy*HG element was integrated early within the founding HLA-A2 allelic population and probably well after the generation of most of the other AH lineages.

The HLA-A2 allele is found in many different ethnic groups including Caucasians, Asians and Africans and probably has ancient origins. In this study, the detection of 51 of 92 (55.4%) individuals with at least one HLA-A2 allele is within the higher-than-expected range for Caucasians and Asians (Imanishi et al. 1992). Although the HLA-A2 antigen is at high frequency worldwide, there is significant allelic variation in different populations (Imanishi et al. 1992) with at least 46 HLA-A2 subtypes identified to date. For example, A*0201, which is within most of the Oxford population, is low in Sin-

gaporeans and Africans, whereas A*0202 is a common African subtype not found in Europeans and Singaporeans (Krausa et al. 1995). Because subtyping of HLA-A2 was not performed, we cannot conclude at this time which of the HLA-A2 subtypes are *Alu*yHG negative or positive.

It is reasonable to infer from the strong association between the HLA-A2 allele and the *Alu*yHG insertion that they have a common evolutionary history. However, the time when the HLA-A2 antigen first emerged still remains largely speculative (Browning and Krausa 1996). The phylogenetic analysis in the present study confirms previous findings (Lawlor et al. 1991; McAdam et al. 1995) that the descendants of the HLA-A2/A10/A19*/A28 lineage and the A3/A9/A19^ lineage were closely related to the gorilla and chimpanzee MHC class I genes, respectively. Although HLA-A2 is relatively frequent in most ethnic populations including Caucasians, Chinese, Japanese, and Africans (Imanishi et al. 1992), the branching order of HLA-A2 group in the phylogenetic tree suggests it has evolved fairly recently in comparison to the other HLA-A groups.

The absence of the *Alu*yHG element from some individuals with the HLA-A2 allele is intriguing. Either the individuals in this group have inherited the HLA-A2 from a founding population that predated the original *Alu*yHG insertion or the HLA-A2-*Alu*yHG-negative genotype has resulted from recombinations and genomic exchanges between the HLA-A2-*Alu*yHG-positive genotypes and the non-HLA-A2-*Alu*yHG-negative genotypes. Since the HLA-A2 allelic group appears to be distantly related to HLA-A3 and -A19 allelic groups in the phylogenetic analysis, genomic rearrangements or recombinations are a reasonable explanation for the low frequency of *Alu*yHG association with non-HLA-A2 alleles. By comparison, the generation of the HLA-A2-*Alu*yHG-negative haplotype and/or the non-HLA-A2-*Alu*yHG-negative haplotype by the excision of the *Alu* element exactly from the insertion site seems to be highly unlikely (see discussion by Batzer et al. 1996a). *Alu* deletions generally involve either the entire *Alu* element together with the flanking regions (Kulski et al. 1999) or a portion of the *Alu* sequence whereby a fragment or signature of the original insertion event is retained (Edwards and Gibbs 1992). On the basis of our confirmatory sequence analysis and observed PCR product sizes, the *Alu*yHG dimorphism is unlikely to have resulted from multiple deletion and integration events within the *Alu* insertion site, 5′-ATGTTTAAGTG-3′. The results obtained for the association between the *Alu*yHG dimorphism and non HLA-A2 alleles suggest that the history for the evolution of the alpha block is more complex than previously considered (Dawkins et al. 1999).

Because the *Alu*yHG element is present in a subset of the A3 and A19^ groups, it is potentially very informative in the analysis of phylogeny and population migration of these HLA allelic groups. HLA-A30 is the most frequent HLA-A19 antigen that is present in all major ethnic groups, and most frequently in populations on the African continent (Bodmer et al. 1997; Tanaka et al. 1997). While HLA-A30 is classified serologically within the HLA-A19 family, being especially difficult to distinguish from HLA-A31, both exonic sequence and phylogenetic analysis by us (see Fig. 3) and others (McKenzie et al. 1999) show that the A30 subtypes are more closely related to the HLA-A1/3/11 allelic group. The separation of the exonic sequences of the HLA-A30 subtypes from the HLA-A19 family is consistent with the phylogenetic trees generated from class I intronic sequences that reflect the diversity of broader groups rather than specific alleles (Blasczyk et al. 1997). In this regard, the presence of the *Alu*HG element in the cell-lines (see Table 1) with HLA-A*3002 (the 18.2 AH identified by the HLA-A*3002-Cw5-B18-DR3-DQ2 alleles) but not within HLA-A*3001 (the 13.1 AH identified by the HLA-A*3001-Cw6-B13-DR7-DQ2 alleles) confirms that the HLA-A30 group has evolved into at least two major subgroups that are seen as specific components of two different AHs. Both of these haplotypes are represented in many ethnic populations in varying frequencies. The 18.2 AH is most common in western Europe, including Sardinia and the Basque region, and several populations from the sub-Saharan continent (Bodmer et al. 1997), whereas the 13.1 AH is more common in populations of the Asian region, especially in central (Chinese-Han and Uygur) and northeastern China (Manchu), as well as in populations of the eastern European countries such as Slovinia, Hungary, and Cyprus (Bodmer et al. 1997; Tanaka et al. 1997). Therefore, HLA-A*3001 and A*3002 appear to have separated early in the evolution of class I alleles, and the 18.2 and 13.1 AHs with and without the *Alu*yHG insertion, respectively, are ancient haplotypes that probably originated in separate populations. Further work is required to determine whether the *Alu*HG insertion occurred during the time of the 18.2 AH formation or following the separation of A*3001 and A*3002 and/or during the divergence of the A3/A19^ groups.

The evolutionary time for *Alu*yHG insertion is difficult to estimate from the present results. The *Alu*yHG element is dimorphic, present in a broad spectrum of ethnic groups including Caucasians, Chinese, Japanese, Indians, and Malays, and strongly associated with HLA-A2 and to a lesser degree with some non-HLA-A2 alleles. Based on previous calculations of *Alu*Yb8 diversity from consensus (Kapitonov and Jurka 1996), the sequence diversity of 0.75% between *Alu*yHG and the consensus *Alu*Yb8 implies that the *Alu*yHG element is about 5 Myo and that it probably emerged about the time of the divergence between humans and chimpanzees. However, neither the insertion site nor the *Alu*yHG element could

be amplified by PCR in the ten different chimpanzee cell-lines, probably because sequence rearrangements and differences occur within this region of the non-human primates.

Nevertheless, it is evident from the present analysis that the association between *Alu*yHG dimorphism and the different HLA alleles and/or AHs provide highly informative markers to review human migration and the origins of contemporary populations. As the *Alu*yHG dimorphism (including determination of heterozygosity and homozygosity) can be readily obtained from a single PCR reaction, this marker should be investigated easily at the population level. The *Alu*yHG dimorphism could also be used as a marker in disease studies such as early onset Alzheimer's (Payami et al. 1997) and autoimmune thyroid diseases (Sudo et al. 1995) that have been associated with HLA-A2, or spontaneous abortions that have been associated with differential expression of the HLA-G gene (Le Bouteiller and Blaschitz 1999).

There are a few dozen members of the *Alu*Yb8 subfamily that have been cloned, sequenced, mapped, and investigated in population studies (Hammer 1994; Zietkiewicz et al. 1994; Batzer et al. 1995; Batzer et al. 1996a; Sherry et al. 1997). The *Alu*yHG sequence is the first member of the *Alu*Yb8 subfamily that we have identified within the alpha block of the MHC by comparative genomic analysis using the public domain programs RepeatMasker and CENSOR (Kulski et al. 1999). The *Alu*yHG sequence is a dimorphic marker, strongly associated with the HLA-A2 alleles within the alpha block of the MHC, that can be added to the other polymorphic *Alu* elements being used to study the population dynamics of humans.

Acknowledgments. This work was supported by a grant to JKK and RLD from the National Health and Medical Research Council and is publication No. MS0007 of the Centre for Molecular Immunology and Instrumentation, The University of Western Australia. NLJ was a recipient of a student fellowship from the Immunogenetics Research Foundation, and JKK was a recipient of a travelling fellowship from the Australian Academy of Science and the Japan Society for the Promotion of Science. We thank Dr. Maude Phipps for providing DNA from the Malaysian-Asian subjects, and Julie Wood of Genetype for providing us with the results of the A locus subgrouping of HLA-A*02 for the cell-lines.

References

Batzer MA, Arcot SS, Phinney JW, Alegria-Hartman M, Kass DH, Milligan SM, Kimpton C, Gill P, Hochmeister M, Ioannou PA, Herrera RJ, Boudreau DA, Scheer WD, Keats BJ, Deininger PL, Stoneking M (1996a) Genetic variation of recent *Alu* insertions in human populations. J Mol Evol 42:22–29

Batzer MA, Deininger PL, Hellmann-Blumberg U, Jurka J, Labuda D, Rubin CM, Schmid CW, Zietkiewicz E, Zuckerkandl E (1996b) Standardized nomenclature for *Alu* repeats. J Mol Evol 42:3–6

Batzer MA, Rubin CM, Hellmann-Blumberg U, Alegria-Hartman M, Leeflang EP, Stern JD, Bazan HA, Shaikh TH, Deininger PL, Schmid CW (1995) Dispersion and insertion polymorphism in two small subfamilies of recently amplified human *Alu* repeats. J Mol Biol 247:418–427

Blasczyk R, Kotsch K, Wehling J (1997) The nature of polymorphism of the HLA class I non-coding regions and their contribution to the diversification of HLA. Hereditas 127:7–9

Bodmer J, Cambon-Thomsen A, Hors J, Piazza A, Sanchez-Mazas A (1997) Anthropology Report. Introduction. In: D Charron (ed) Genetic diversity of HLA. Functional and medical implications EDK, Paris pp. 269–284

Browning M, Krausa P (1996) Genetic diversity of HLA-A2: evolutionary and functional significance. Immunol Today 17:165–170

Cattley SK, Williamson JF, Tay GT, Martinez OP, Gaudieri S, Dawkins RL (2000) Further characterisation of MHC haplotypes demonstrates conservation telomeric of HLA-A: update of the 4AOH and 10IHW cell panels. European Journal of Immunogenetics 27: 397–426

Dawkins R, Leelayuwat C, Gaudieri S, Tay G, Hui J, Cattley S, Martinez P, Kulski J (1999) Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunol Rev 167:275–304

Degli-Esposti MA, Griffiths MS, Daly LN, Witt CS, Simons M, Carcassi C, Albert ED, Giphart MJ, Dawkins RL (1993) Characterization of 4AOHW cell line panel including new data for the 10IHW panel. Hum Immunol 38:3–16

Deininger PL, Batzer MA (1999) *Alu* repeats and human disease. Mol Genet Metab 67:183–193

Edwards MC, Gibbs RA (1992) A human dimorphism resulting from loss of an *Alu.* Genomics 14:590–597

Gaudieri S, Dawkins RL, Habara K, Kulski JK, Gojobori T (2000) SNP profile within the human major histocompatibility complex reveals an extreme and interrupted level of nucleotide diversity. Genome Res 10:1579–1586

Gaudieri S, Giles KM, Kulski JK, Dawkins RL (1997a) Duplication and polymorphism in the MHC: *Alu* generated diversity and polymorphism within the PERB11 gene family. Hereditas 127:37–46

Gaudieri S, Kulski JK, Balmer L, Giles KM, Inoko H, Dawkins RL (1997b) Retroelements and segmental duplications in the generation of diversity within the MHC. DNA Sequence 8:137–141

Goldberg YP, Rommens JM, Andrew SE, Hutchinson GB, Lin B, Theilmann J, Graham R, Glaves ML, Starr E, McDonald H, et al. (1993) Identification of an *Alu* retrotransposition event in close proximity to a strong candidate gene for Huntington's disease. Nature 362:370–373

Hammer MF (1994) A recent insertion of an *Alu* element on the Y chromosome is a useful marker for human population studies. Mol Biol Evol 11:749–761

Hampe A, Coriton O, Andrieux N, Carn G, Lepourcelet M, Mottier S, Dreano S, Gatius MT, Hitte C, Soriano N, Galibert F (1999) A 356-Kb Sequence of the subtelomeric part of the MHC class I region. DNA Sequence 10:263–299

Imanishi T, Tatsuya A, Kimura A, Tokunaga k, Gojobori T (1992) Allele and haplotype frequencies for HLA and complement loci in various ethnic groups. In: (Tsuji L, Aizawa M, Sasazuki T, eds) HLA 1991, Vol. I Oxford University Press, Oxford, pp. 1065–1220

Jurka J, Smith T (1988) A fundamental division in the *Alu* family of repeated sequences. Proc Natl Acad Sci USA 85:4775–4778

Jurka J (1993) A new subfamily of recently retroposed human *Alu* genes. Nucl Acid Res 21:2252

Kapitonov V, Jurka J (1996) The age of *Alu* subfamilies. J Mol Evol 42:59–65

Kato K, Trapani JA, Allopenna J, Dupont B, Yang SY (1989) Molecular analysis of serologically defined HLA-Aw19 antigens: a genetically distinct family of HLA-A antigens comprising A29, A31, A32 and Aw33, but probably not A30. J Immunol 143:3371–3378

Krausa P, Brywka M, 3rd, Savage D, Hui KM, Bunce M, Ngai JL, Teo DL, Ong YW, Barouch D, Allsop CE, et al. (1995) Genetic polymorphism within HLA-A*02: significant allelic variation revealed in different populations. Tissue Antigens 45:223–231

Kulski JK, Gaudieri S, Bellgard M, Balmer L, Giles K, Inoko H, Dawkins RL (1997) The evolution of MHC diversity by segmental duplication and transposition of retroelements. J Mol Evol 45:599–609

Kulski JK, Gaudieri S, Dawkins RL (2000a) Using *Alu* J elements as molecular clocks to trace the evolutionary relationships between duplicated HLA Class I genomic segments. J Mol Evol 50:510–519

Kulski, JK, Gaudieri S, Dawkins RL (2000b) Transposable elements and the metamerismatic evolution of the HLA class I region. *In*: (Kasahara M, ed) The major histocompatibility complex: evolution, structure, and function. Springer-Verlag, Tokyo, pp. 158–177

Kulski JK, Gaudieri S, Martin A, Dawkins RL (1999) Coevolution of PERB11 (MIC) and HLA class I genes with HERV-16 and retroelements by extended genomic duplication. J Mol Evol 49:84–97

Kumar S, Tamura K, Nei M (1994) MEGA: molecular evolutionary genetics analysis software for microcomputers. Comput Appl Biosci 10:189–191

Lawlor DA, Warren E, Taylor P, Parham P (1991) Gorilla class I major histocompatibility complex alleles: comparison to human and chimpanzee class I. J Exp Med 174:1491–1509

Le Bouteiller P, Blaschitz A (1999) The functionality of HLA-G is emerging. Immunol Rev 167:233–244

McAdam SN, Boyson JE, Liu X, Garber TL, Hughes AL, Bontrop RE, Watkins DL (1995) Chimpanzee MHC class I A locus alleles are related to only one of the six families of human A locus alleles. J Immunol 154:6421–6429

McKenzie L, Pecon-Slattery J, Carrington M, O'Brien S (1999) Taxonomic hierarchy of HLA class I allele sequences. Genes and Immunity 1:120–129

Miki Y (1998) Retrotransposal integration of mobile genetic elements in human diseases. J Hum Genet 43:77–84

Miller SA, Dykes DD, Polesky HF (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nuc Acid Res 16:1215

Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K (1991) Inactivation of the cholinesterase gene by *Alu* insertion: possible mechanism for human gene transposition. Proc Natl Acad Sci USA 88:11315–11319

Naruse TK, Nose Y, Ando R, Araki N, Shigenari A, Ando A, Ishihara M, Kagiya M, Nabeya N, Isshiki G, Inoko H (1998) Extended HLA haplotypes in Japanese homozygous typing cells. Tissue Antigens 51:305–308

Payami H, Schellenberg GD, Zareparsi S, Kaye J, Sexton GJ, Head MA, Matsuyama SS, Jarvik LF, Miller B, McManus DQ, Bird TD, Katzman R, Heston L, Norman D, Small GW (1997) Evidence for association of HLA-A2 allele with onset age of Alzheimer's disease. Neurology 49:512–518

Prasad VK, Yang SY (1996) Allele assignment for HLA-A, -B, and -C genes to the Tenth International Histocompatibility Workshop cell-lines. Tissue Antigens 47:538–546

Quentin Y (1988) The *Alu* family developed through successive waves of fixation closely connected with primate lineage history. J Mol Evol 27:194–202

Raymond M, Rousset F (1995) Genepop (Version 1.2)—population genetics software for exact tests and ecumenicism. J Hered 86:248–249

Schmid CW (1996) *Alu*: structure, origin, evolution, significance and function of one-tenth of human DNA. Progr Nucl Acid Res Mol Biol 53:283–319

Sherry ST, Harpending HC, Batzer MA, Stoneking M (1997) *Alu* evolution in human populations: using the coalescent to estimate effective population size. Genetics 147:1977–1982

Shiina T, Tamiya G, Oka A, Takishima N, Inoko H (1999) Genome sequencing analysis of the 1.8 Mb entire human MHC class I region. Immunol Rev 167:193–199

Smit AF (1996) The origin of interspersed repeats in the human genome. Curr Opin Genet Dev 6:743–748

Stoneking M, Fontius JJ, Clifford SL, Soodyall H, Arcot SS, Saha N, Jenkins T, Tahir MA, Deininger PL, Batzer MA (1997) *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. Genome Res 7:1061–1071

Sudo T, Kamikawaji N, Kimura A, Date Y, Savoie CJ, Nakashima H, Furuichi E, Kuhara S, Sasazuki T (1995) Differences in MHC class I self peptide repertoires among HLA-A2 subtypes. J Immunol 155:4749–4756

Tanaka H, Tokonaga K, Inoko H, Tsuji K, Chimge NO, Jia GJ, Park MH, Han H, Han Y, Nose Y, Nagao N, Saitou S, Akaza T, Juji T (1997) Distribution of HLA-A, B and DRB1 alleles and haplotypes in Northeast Asia. In: (Charron D, ed) Genetic diversity of HLA. Functional and medical implications. EDK, Paris, pp. 285–291

Ullu E, Tschudi C (1984) *Alu* sequences are processed 7SL RNA genes. Nature 312:171–172

Zietkiewicz E, Richer C, Makalowski W, Jurka J, Labuda D (1994) A young *Alu* subfamily amplified independently in human and African great apes lineages. Nucl Acid Res 22:5608–5612