doi: 10.1046/j.1529-8817.2005.00183.x

Polymorphic Alu Insertions and their Associations with MHC Class I Alleles and Haplotypes in the Northeastern Thais

D. S. Dunn[1], A. V. Romphruk[2,5], C. Leelayuwat[3,5], M. Bellgard[1] and J. K. Kulski[1,4,*]

[1] *Centre for Bioinformatics and Biological Computing, Murdoch University, Murdoch, 6150, Western Australia, Australia*
[2] *Blood Transfusion Center, Faculty of Medicine, Khon Kaen University, Khon Kaen, Thailand*
[3] *Department of Clinical Immunology, Faculty of Associated Medical Sciences, Khon Kaen University, Khon Kaen, Thailand*
[4] *Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara 259-1143, Japan*
[5] *The Centre for Research and Development of Medical Diagnostics Laboratories, Khon Kaen University, Khon Kaen, Thailand*

Summary

Polymorphic Alu insertions (POALINs) are known to contribute to the strong polymorphic nature of the Major Histocompatibility Complex (MHC). Previous population studies on MHC POALINs were limited to only Australian Caucasians and Japanese. Here, we report on the individual insertion frequency of the five POALINs within the MHC class I region, their HLA-A and -B associations, and the three and four locus alpha block POALIN haplotype frequencies in the Northeastern (NE) Thai population. Of the five POALINs, the lowest frequency was 0.018 for *AluyHF* and the highest frequency was 0.292 for *AluyHJ* and *AluyHG*. The strongest positive associations between the POALINs and HLA class I alleles was between *AluyMICB* and *HLA-B* 57*, *AluyHJ* and *HLA-A* 24* and *HLA-A* 01*, and *AluyHG* and *HLA-A* 02,* supporting previous findings in Caucasians and Japanese. Single POALIN haplotypes were found more frequently than multiple POALIN haplotypes. However, of the seven different POALIN haplotypes within the MHC alpha block, there were only two significant differences between the NE Thais, Caucasians and Japanese. This study confirms that the MHC POALINs are in linkage disequilibrium with *HLA-A* and *–B* alleles and that there are significant frequency differences for some of the POALINs when compared between NE Thai, Caucasians and Japanese.

Keywords: *Alu*, MHC, *HLA*, Haplotypes, Polymorphism

Introduction

The Major Histocompatibility Complex (MHC) in humans is a polymorphic genomic region of approximately 3.6 megabases on the short arm of chromosome 6 (6p21.3). The complex consists of ~224 closely linked genes and other genomic features (repeat elements) divided into 3 major sub-regions, Class II, Class III (Central MHC) and Class I, centromeric to telomeric (The MHC Sequencing Consortium, 1999). Strong linkage disequilibrium exists across the MHC, particularly among alleles of specific multilocus haplotypes and between particular genes (Begovich *et al*. 1992; Yunis *et al*. 2003). An increased list of polymorphisms identified in both the intragenic and intergenic regions of the MHC genomic region will permit rapid identification of changes (recombination) that can be localised to small segments (Carrington, 1999). This will most likely have an important impact on the definition and refinement of human haplotypes that are presently defined by a limited number of genetic markers,

* Corresponding author: Jerzy K Kulski, PhD, Centre for Bioinformatics and Biological Computing, Murdoch University, Murdoch, 6150, Western Australia, Australia. Phone: +61 8 93602492; Fax: +61 8 93607238 E-mail: jkulski@murdoch.edu.au

especially the HLA genes (*HLA-A, -B, -C, -DP, -DQ* and *–DR*). However, other markers have also been used in studying MHC haplotype variation, such as the polymorphic MHC-related genes (Romphruk *et al*. 2001), microsatellites (Li *et al*. 2004, Pascual *et al*. 2002), SNPs (Holloway *et al*. 1999; Walsh *et al*. 2003) and polymorphic Alu insertions (POALINs) (Dunn *et al*. 2002, 2003a; Kulski *et al*. 2001, 2002a).

A large number of polymorphic Alu insertions located on different chromosomes are widely used in population studies (Antunez-de-Mayolo *et al*. 2002; Carroll *et al*. 2001; Stoneking *et al*. 1997; Watkins *et al*. 2001). Alu sequences are the largest family of short interspersed nucleotide elements (SINEs) in humans and other primates, with more than a million copies per haploid genome (Lander *et al*. 2001; Mighell *et al*. 1997; Rowald & Herrera, 2000). They are derived ancestrally from the 7*SL* RNA gene and most are fixed in the human genome (Deininger & Batzer, 1991), with approximately 4500 polymorphic Alu insertion sites estimated within the human genome (Carroll *et al*. 2001; Rowald & Herrera, 2000; Roy-Engel *et al*. 2001). The POALINs are of interest as genomic markers because they are estimated to contribute to at least 0.4% - 1.0% of human genetic disorders, are biallelic (they are either present or absent within the genome at a particular site), may be selectively neutral and are generally free of homoplasy (Deninger & Batzer, 1999). The main advantage of using POALINs over SNPs in this context is that Alu insertions are single occurring events during human evolutionary history, whereas SNPs may result from parallel substitutions involving multiple evolutionary events. Thus, POALINs are useful molecular candidates for investigating the origins of human haplotypes, ethnic groups and disease associations (Deninger & Batzer, 1999).

A large number of small conserved haplotype blocks or linked genomic sequences were reported recently in a study of 201 SNPs covering nine classical HLA loci and two *TAP* genes (*TAP1* and *TAP2*) between the *HLA-A* and the *MLN* loci of the MHC genomic region (Walsh *et al*. 2003). However, the ~310-kb genomic region between the *HLA-J* and *HLA-F* genes, that we designate here as the alpha block, was not fully characterized. The alpha and beta blocks within the class I region of the MHC were previously defined as a linkage (haplotype) group of polymorphic genomic sequences that are restricted to blocks of sequences from the *HLA-J* to the *HLA-F* genes and from the *MICB* to *HLA-C* genes, respectively (Dawkins *et al*. 1999). The MHC alpha and beta blocks were also considered to be 'duplication blocks' that harbour clusters of coding and non-coding (pseudogenes) HLA class I and MIC genes (Kulski *et al*. 2002b). The diversity of HLA class I and MIC genes within the alpha and beta blocks, as well as the multi-locus (six locus *HLA-DQB1, HLA-DRB1, MICA, HLA-B, HLA-C* and *HLA-A)* haplotypes, have also been described in a large population sample of Northeastern (NE) Thai individuals (Romphruk *et al*. 2001). Although the beta block was represented in this analysis by the *HLA-B*, *HLA-C* and *MICA* loci, the alpha block was limited to only the *HLA-A* locus. We previously analysed and reported on *HLA-A* haplotypes composed of three POALINs within the MHC alpha block of Australian Caucasians and Japanese (Dunn *et al*. 2002). Therefore, a novel approach in MHC haplotype studies is to consider POALIN haplotypes in relation to the recognized *HLA-A* alleles as an alpha block haplotype of four loci. The population sample of NE Thais, together with their HLA class I genotypic data and POALINs, provides a unique opportunity to further investigate the alpha and beta blocks for new HLA-POALIN associations and haplotypes.

Here, we report on the individual insertion frequency of the five POALINs within the MHC class I region, their *HLA-A* and *-B* associations, and the three and four locus alpha block POALIN haplotype frequencies in the NE Thai population.

Materials and Methods

Genomic DNA and *HLA-A* and *–B* typing

Genomic DNA was obtained from 192 unrelated healthy NE Thai individuals (Romphruk *et al*. 2001). The *HLA-A* and *-B* alleles were typed by the polymerase chain reaction-sequence-specific primers (PCR-SSP) technique (Romphruk *et al*. paper in preparation), and some of the HLA class I alleles were differentiated at the sequence allelic level (e.g., *HLA-A* 0203, -A* 0205, -A* 0207*) whereas others, such as *HLA-A* 01, -A* 03, -A* 11, -B* 13, -B* 15 and –B* 57*, were designated as

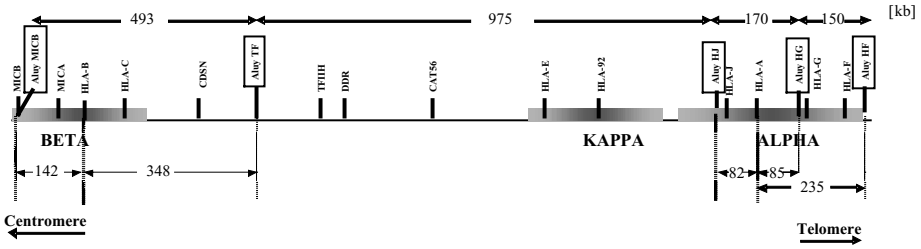

Figure 1 The location of the 5 POALINs (boxed) in relation to the *HLA*, *MIC*, *CDSN*, *TFIIH*, *DDR1* and *CAT56* loci within the 1800 kb class I region of the MHC. Three POALINs are located within the alpha block, one each within the beta block and the region between the beta and kappa blocks. The distances (kb) were calculated from the annotated GenBank files with accession numbers AP000503-AP000521.

alleles only at the broader group. A limitation of the PCR-SSP typing method was that many of the 'group' alleles could not be resolved into their specific sequence alleles. For example, the group *HLA-A* 11* allele was not resolved separately into its 17 sequence specific alleles, *HLA-A* 1101-A* 1117*. In this paper, we distinguish the specific alleles from the broader groups by using the standard nomenclature for specific sequence alleles such as the letter *A* or *B* for the *HLA-A* or *-B* gene, and an asterisk and four number code, as for *HLA-B* 4601*.

POALINs and PCR assay

The polymorphic Alu insertions (POALINs) used were located in the class I region from *MICB* to telomeric of *HLA-F* (Fig. 1). Primers were designed in the regions flanking the POALINs so that PCR products for the absence and the presence of the POALIN were distinguished from each other by their different sizes. Here, we define the absence of the Alu insertion at a POALIN locus as the *Aluy* 1* allele and the presence of the Alu insertion as the *Aluy* 2* allele. The PCR assays for the various POALINS were the same as those described in previous reports (Dunn *et al*. 2002, 2003a; Kulski *et al*. 2001, 2002). Briefly, the PCR solution (20 μl) contained 20 ng of template DNA, primers at 10 μM each, dNTPs at 2 mM each, 0.5 - 1.0 units of AmpliTaq polymerase (Applied Biosystems, Foster City, CA) and 3 mM $MgCl_2$ in 10 mM Tris-HCl buffer pH 8.3. PCR was performed using a Perkin Elmer 2700 Thermal cycler programmed for 35 cycles with a denaturation (96°C - 30 sec), annealing (68°C for *AluyMICB*, 62°C for other - 45 sec) and extension (72°C - 45 sec) step at each cycle. The reaction products were analysed by horizontal gel electrophoresis in 2% agarose using Tris-borate-EDTA running buffer. Fragments of different sizes were produced for either the presence or the absence of the POALIN, seen as a single fragment size in homozygous samples and as two fragments in the heterozygous samples (Fig. 2). Two DNA samples from the International Histocompatibility Workshop (IHW) panel were used as controls, one homozygous for the absence and the other homozygous for the presence of Alu insertions. Mixtures of the two controls were also included as heterozygous controls with each PCR run.

Calculation of Frequencies and Statistical Analysis

Allele frequencies (AF) were calculated using the formula: AF = sum of each individual allele/2N, where N equals the total number of individuals. Hardy-Weinberg (H-W) equilibrium analysis was performed for each of these POALINs. Heterozygosity (H) (Ott, 1992) was estimated as $1 - (p^2 + q^2)$, where p and q are the allele frequencies. Three loci haplotypes (three class I alpha block POALINs) and four loci haplotypes (*HLA-A* plus the three loci POALIN haplotype) were constructed for each individual sample. Haplotype frequencies ('observed') and H-W equilibrium were performed using the Arlequin computer program (Schneider *et al*. 2000). Significance of the differences between the haplotype frequencies for the NE Thai, Japanese and Australian Caucasians were determined by a contingency test (Fisher's exact test).

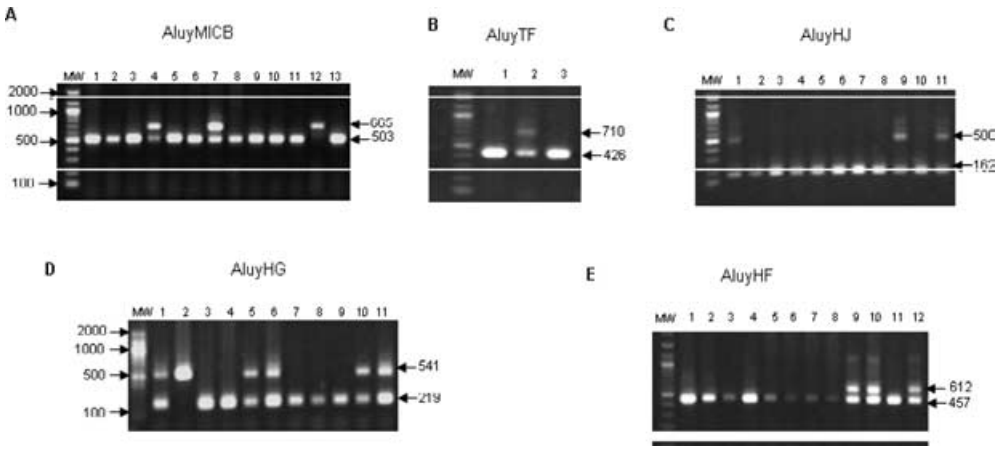

Figure 2 Gel photographic presentation of the MHC class I POALINs. The PCR products for the presence and/or absence of the respective POALINs are visually distinguishable. A marker (MW) control with known sizes (sizes shown on the left for A and D but not shown for B, C and E) was used for each gel (A-E) and the columns represent individual PCR products. The larger PCR product size for each POALIN represents the presence and the smaller size represents the absence of the POALIN. (A) Product 12 represents a homozygous *AluyMICB* individual, products 1-3, 5-6, 8-11 & 13 represents homozygous individuals without (absent) *AluyMICB*, and products 4 & 7 represent heterozygous individuals. (B) Product 2 represents the heterozygous presence of *AluyTF* and 1 & 3 represents the homozygous absence of *AluyTF*. (C) Products 1, 9 and 11 are heterozygous for *AluyHJ* and 2-8 & 10 are homozygous for the absence of *AluyHJ*. (D) Product 2 is homozygous for the presence of *AluyHG*, while 1, 5, 6, 10 and 11 are heterozygous, and the rest are homozygous for the absence of the POALIN. (E) Products 9-10 and 12 are heterozygous for *AluyHF*.

A contingency test was also performed to determine the significant difference between the 'expected' and 'observed' haplotype frequencies for single and multiple POALINs. 'Expected' haplotype frequencies were calculated as the products of the allele frequencies, while the frequency of the single POALIN-containing-haplotype was determined by summing all the haplotype frequencies that had only one of the POALINs. Similarly, the frequency of the multiple POALIN-containing-haplotype was determined by summing all the haplotype frequencies that contained more than one POALIN. The 2×2 contigency table was populated with proportions of the 192 individuals and constructed with the 'expected' versus 'observed' haplotype frequencies for both the single and multiple POALIN haplotypes.

HLA associations were determined by calculating the percentage of individuals sharing the same HLA allele and an Alu insertion. Linkage disequilibrium was represented as the delta measurement developed by Bengtsson & Thomson (1981) and defined as $(p_A - p_B)/(1 - p_B)$, where p_A is the frequency of HLA alleles in individuals with the POALIN and p_B is the frequency of HLA alleles in individuals without the POALIN. When a negative delta value was obtained, a rearrangement of the variables was applied, whereby the delta prime (delta') is defined as $(p_B - p_A)/(1 - p_A)$.

Results and Discussion

Fig. 1 shows the location of the five POALINs and the alpha, beta and kappa blocks within the ~1.8 Mb MHC class I region. Here we refer to the alpha, beta and kappa blocks as genomic regions within the MHC class I that harbour clusters of coding and non-coding (pseudogenes) HLA class I and MIC genes (Kulski *et al.* 2002). *AluyMICB* is located in the beta block within the first intron of the *MICB* gene, *AluyTF* is located in the region between the beta and kappa blocks close to the *TFIIH* and *CDSN* genes, and the remaining three elements are located within the alpha block: *AluyHJ* close to *HLA-J*, *AluyHG* close to *HLA-G* and *AluyHF* close to *HLA-F*.

The largest POALIN allele frequency of 0.292 ($p > 0.40$) in the NE Thais was detected for the *AluyHJ* and *AluyHG* loci (Table 1). The *AluyHG (AluyHG* 2)*

Table 1 Observed genotypes, allele frequencies, Hardy-Weinberg significance and heterozygosity for five MHC POALINs in a NE Thai population (n = 192)

MHC POALINs	Genotypes			*Allele Frequencies*			Heterozygosity	
	1,1	1,2	2,2	*Aluy* 1*	*Aluy* 2*	X^2	p	*H*
AluyMICB	150	39	3	0.883	0.117	0.064	0.80	0.207
AluyTF	161	29	2	0.914	0.086	0.286	0.59	0.152
AluyHJ	97	78	17	0.708	0.292	0.054	0.81	0.413
AluyHG	94	84	14	0.708	0.292	0.664	0.41	0.413
AluyHF	185	7	0	0.982	0.018	-	-	0.035

Genotypes: 1,1 = homozygote Alu absent._ 1,2 = heterozygote._ 2,2 = Homozygote Alu present.

Table 2 Alpha block POALIN haplotype identification (id), definition and frequencies in the NE Thai population and a comparison with the Australian Caucasian and Japanese populations

Alu Haplotype Id	Alu Haplotype			Expected from allele freq	Haplotype frequencies			Haplotype differences between Thai and	
	AluyHJ	AluyHG	AluyHF		Thai n = 192	Australian[c] n = 105	Japanese[c] n = 87	Australian (p)[b]	Japanese (p)[b]
A	1	1	1	0.492	0.421	0.355	0.364	0.267	0.368
B	1	1	2	0.009	0.016	0.104	0.038	<0.001	0.255
C	2	1	1	0.203	0.270	0.224	0.325	0.384	0.347
D	1	2	1	0.203	0.265	0.200	0.199	0.211	0.235
E	1	2	2	0.004	0.003	0.088	0.021	<0.0001	0.132
F	2	2	1	0.084	0.024	0.017	0.053	0.691	0.209
G	2	1	2	0.004	0.000	0.012	0.000	-	-

[a]Haplotype frequencies were determined by using Arlequin.
[b]Haplotype differences were calculated by the 2×2 contingency (haplotype numbers in gene pool).
[c]From Dunn *et al*. 2002.
Alu haplotype: 1 represents the absence of the insertion and 2 represents the presence of the insertion.

insertion frequency in the NE Thais was similar to the Australian Caucasians (0.301) and Japanese (0.270) whereas the *AluyHJ* frequency was lower in the Australians (0.252) and higher in the Japanese (0.376) (Dunn *et al*. 2002a) than in the NE Thais (0.292). The *AluyMICB* insertion (*AluyMICB* 2)* occurred at a frequency of 0.117, similar to the frequency of 0.118 previously reported in a sample size of 200 (Kulski *et al*. 2001). This frequency of 0.117 (p>0.80) in the NE Thais was lower than in the Australian Caucasians (0.157) (Kulski *et al*. 2002a) and Japanese (0.242) (Dunn *et al*. 2003b). Similarly, the frequency of the *AluyHF* 2* allele in the NE Thais (0.018) was lower than that observed in the Australian Caucasians (0.203) and Japanese (0.064) populations. All five POALINs were in H-W equilibrium. These observations revealed that the POALIN frequencies in the NE Thais were either between or below the POALIN frequencies of the Caucasians and the Japanese and, therefore, may be useful for population studies in a similar way as POALINs from other genomic regions (Antunea-de-Mayolo *et al*. 2002; Carroll *et al*. 2001; Stoneking *et al*. 1997; Watkins *et al*. 2001).

Haplotypes were constructed from the three POALINs '*AluyHJ-AluyHG-AluyHF'* that are located within a 320kb genomic region called the alpha block (Fig. 1). Table 2 lists the haplotype identification (A to G) and haplotype definition (POALIN allelic state), their expected frequencies (from allele frequencies), their observed frequencies (HF) and comparisons of the NE Thais with Australian Caucasian and Japanese haplotypes. Assuming no linkage disequilibrium for the three POALINs (loci), we would expect eight haplotypes. However, only six haplotypes were observed (A to F) in the NE Thais and Japanese, and seven haplotypes (A to G) in Australian Caucasians. The haplotype H, containing an Alu insertion (allele 2) at all three loci within the alpha block, was not observed. No significant difference was observed between the NE Thai and Japanese haplotypes, but a significant difference was observed between the NE Thai and Australian Caucasian

haplotypes B (<0.001) and E (<0.0001). This observation supports a closer relationship between the NE Thais and Japanese than with Caucasians.

In comparison to the three POALIN haplotypes, no significant differences ($p > 0.05$) were found between the NE Thais, Australian Caucasians and Japanese for the two locus POALIN haplotypes of the alpha block. Similarly, there was no significant difference ($p > 0.05$) for the *AluyMICB* and *AluyTF* haplotypes between the three populations (data not shown).

A large percentage (55.1%) of the NE Thai haplotypes detected had either one of the POALINs inserted, and 42.1% were found to have no insertions. Therefore, only 2.7% of the haplotypes had a POALIN insertion at more than one locus. Even though the individual haplotype frequencies appear to be similar, the difference is evident when the sum of the single POALIN haplotypes is considered relative to the multiple POALIN haplotypes. More single POALIN containing haplotypes than multiple POALIN containing haplotypes were observed when compared to the expected. This difference supports the idea that recombination events probably account for the small number of multiple POALIN haplotypes, and that recombinations may have occurred more frequently within the alpha block (Walsh *et al*. 2003) than previously considered (Dawkins *et al*. 1999). The probability of independent insertions of two or more elements at the same nucleotide site was estimated presumably to be close to zero (Batzer *et al*. 1990). Based on the estimation that with every new Alu insertion only 100-200 Alu elements within the human genome will be polymorphic after a million years (Baxter *et al*. 1990), we would expect to find only one polymorphic Alu element within the 1.8 Mb MHC class I region per haplotype. Because the likelihood of Alu insertions occurring at different loci within the same individual (haplotype) is extremely rare, haplotypes with multiple POALIN sites (two or more) have most probably arisen by recombination of haplotypes with single but different polymorphic Alu elements.

The number and percentage of *HLA-A* and *–B* alleles associated with POALINs are shown in Tables 3 and 4. The delta value of the linkage disequilibrium between the *HLA-A* or *–B* allele and a particular POALIN is also given. An association between the HLA alleles and the presence of a POALIN was

Table 3 The number and percentage of *HLA-A* alleles associated with four MHC POALINs in NEThais

HLA-A allele	Total HLA-A allele	No. with AluyHJ*2	% with AluyHJ*2	Delta	No. with AluyHG*2	% with AluyHG*2	Delta	No. with AluyHF*2	% with AluyHF*2	Delta	No. with AluyTF*2	% with AluyTF*2	Delta
01	6	5	83.3	0.80	0	0	-	0	0	-	2	33.3	0.50#
0207	54	21	38.9	0.36#	53	98.1	0.98	2	3.7	0.96#	8	14.8	0.83#
0203	35	11	31.4	0.54#	34	97.1	0.97	1	2.9	0.97#	7	20	0.75#
0205/0208	2	1	50	0.00	1	50	0.00	0	0	-	0	0	-
11	92	34	37	0.41#	44	47.8	0.08#	3	3.3	0.97#	20	21.7	0.72#
2601	3	1	33.3	0.50#	1	33.3	0.50#	2	66.7	0.50	1	33.3	0.50#
24	72	69	95.8	0.96	22	30.6	0.56#	4	5.6	0.94#	10	13.9	0.84#
30	4	3	75	0.67	1	25	0.67#	0	0	-	0	0	-
33	46	12	26.1	0.65#	18	39.1	0.36#	1	2.2	0.98#	3	6.5	0.93#
3101	4	2	50	0.00	0	0	-	0	0	-	1	25	0.67#
6801	1	0	0	-	0	0	-	0	0	-	0	0	-
7401	4	4	100	1.00	1	25	0.67#	0	0	-	1	25	0.67#
29	9	4	44.4	0.20#	4	44.4	0.20#	0	0	-	1	11.1	0.88#
3201	1	0	0	-	0	0	-	0	0	-	0	0	-
03	1	0	0	-	0	0	-	0	0	-	1	100	1.00
34/66	12	7	58.3	0.29	1	8.3	0.91#	1	8.3	0.91#	2	16.7	0.80#

Aluy*2 represents the presence of the insertion.
#delta values less than zero, therefore delta' calculated.

Table 4 The number and percentage of *HLA-B* alleles associated with *AluyMICB* and *AluyTF* in NE Thais

HLA-B allele	Total *HLA-B* alleles	No. with *AluyMICB* 2*	% with *AluyMICB* 2*	Delta	*No. with AluyTF* 2*	*% with AluyTF* 2*	Delta
57	8	7	87.5	0.86	2	25.0	0.67#
07(02-07)	16	2	13.3	0.85#	1	6.7	0.93#
13	34	6	18.2	0.78#	7	21.2	0.73#
27	25	8	32.0	0.53#	4	16.0	0.81#
35	12	2	16.7	0.80#	2	16.7	0.80#
4601	15	3	20.0	0.75#	5	33.3	0.50#
5401	2	1	50.0	0.00	0	0.0	-
55/56	11	5	45.5	0.17#	5	45.5	0.17#
4001	20	5	25.0	0.67#	1	5.0	0.95#
40(02,04-06)	8	3	37.5	0.40#	2	25.0	0.67#
4801	1	1	100.0	1.00	0	0.0	-
5801	33	8	24.2	0.68#	4	12.1	0.86#
38	9	3	33.3	0.50#	1	11.1	0.88#
39/6701	15	2	13.3	0.85#	3	20.0	0.75#
5101	2	0	0.0	-	0	0.0	-
18	26	9	34.6	0.47#	5	19.2	0.76#
15	46	15	32.6	0.52#	14	30.4	0.56#
51/5201	15	3	20.0	0.75#	4	26.7	0.64#
44	12	1	8.3	0.91#	1	8.3	0.91#
1521	3	0	0.0	-	0	0.0	-
3701	2	0	0.0	-	0	0.0	-

Aluy 2* represents the presence of the insertion.
#delta values less than 0, therefore delta' calculated.

considered to be insignificant if only one example of an association was observed in the population. Therefore, no positive associations were observed between the *AluyTF* POALIN (*AluyTF* 2*) and the HLA loci, even though large delta prime values (>0.50) were observed, which indicated that there is a negative allelic association with the loci in linkage disequilibrium. However, a strong percentage association was observed between *AluyHJ* 2* and *HLA-A* 24* (95.1%) and *HLA-A* 01* (85.7%), *AluyHG* 2* and *HLA-A* 0207* (98.3%) and *HLA-A* 0203* (97.2%). *HLA-A* 11* is present in the largest number of individuals, but does not show a particularly strong association with any of the POALINs. The strong frequency associations between some of the HLA alleles and POALINs in the NE Thais, such as between *HLA-A* 02* and *AluyHG*, *HLA-A* 24* and *AluyHJ*, *HLA-A* 26* and *AluyHF*, and *HLA-B* 57* and *AluyMICB*, were the same as in the Australian Caucasians and Japanese (Dunn *et al.* 2002, 2003a; Kulski *et al.* 2001, 2002a).

The haplotype analysis of the three POALIN loci and *HLA-A* locus within the alpha block (Fig. 1) of the NE Thais clearly supports the strong *HLA-A* and POALIN associations. Table 5 lists the haplotype identities, definition (allelic state) and frequencies of the four locus alpha block haplotypes. The list includes only the haplotypes that would yield at least one individual carrying the listed haplotype (a haplotype frequency greater than 0.0025); therefore, only thirty-two haplotypes were listed. The four loci (alpha block) haplotypes were named by employing an alpha-numeric code: a number that refers to the *HLA-A* allele, followed by the same letter that was used for the three loci POALIN haplotypes. Haplotypes deficient in Alu insertions were the most prevalent. The haplotype 24C with the *AluyHJ* insertion and the haplotypes 0207D and 0203D with the *AluyHG* insertion were relatively prevalent with frequencies of 0.196, 0.149 and 0.080, respectively. These five haplotypes, 11A, 33A, 24C, 0207D and 0203D, yielded an expected number of at least one homozygous individual in a total sample size of 192 individuals. The same *HLA-A* allele was observed in different haplotypes (table 5); *HLA-A* 11* was detected in five different haplotypes, *-A* 0203* in four, *-A* 24* and *–A* 0207* in three each, and *–A* 01*, *-A* 29*, *-A* 2601*, *-A* 3101* and *–A* 34/66* were each found in two different haplotypes.

Adding the *HLA-A* locus to the analysis of the POALIN haplotypes in NE Thais further supports the

Table 5 Alpha block haplotype identification (id), definition and frequencies in the NE Thais

	Haplotype				
Haplotype Id	AluyHJ	HLA-A	AluyHG	AluyHF	Frequency
1A	1	01	1	1	0.003
1C	2	01	1	1	0.015
3A	1	03	1	1	0.003
11A	1	11	1	1	0.208
11B	1	11	1	2	0.005
11D	1	11	2	1	0.031
11C	2	11	1	1	0.035
11F	2	11	2	1	0.006
24A	1	24	1	1	0.010
24C	2	24	1	1	0.196
24F	2	24	2	1	0.004
29A	1	29	1	1	0.014
29C	2	29	1	1	0.004
30A	1	30	1	1	0.009
33A	1	33	1	1	0.129
0203A	1	0203	1	1	0.003
0203D	1	0203	2	1	0.080
0203E	1	0203	2	2	0.003
0203F	2	0203	2	1	0.009
0205A	1	0205	1	1	0.005
0207A	1	0207	1	1	0.003
0207D	1	0207	2	1	0.149
0207F	2	0207	2	1	0.006
2601A	1	2601	1	1	0.005
2601B	1	2601	1	2	0.005
3101A	1	3101	1	1	0.003
3101C	2	3101	1	1	0.005
3201A	1	3201	1	1	0.003
6801C	2	6801	1	1	0.003
7401C	2	7401	1	1	0.008
34/66A	1	34/66	1	1	0.029
34/66B	1	34/66	1	2	0.003

Alu loci: 2 represents the presence of the insertion.

strong associations between the *HLA-A* alleles and some of the POALIN loci, using either percentage association or the delta value in linkage disequilibrium. For example, the *HLA-A* 0203* (founder) is strongly associated with *AluyHG* 2* in NE Thais (Tables 3 and 5) and in Australian Caucasians (Kulski *et al*. 2001), and therefore is likely to be the founder allele upon which the Alu insertion event originally occurred. Similarly, the strong correlation between *HLA-A* 24* and *AluyHJ* 2* suggests that *HLA-A* 24* was a founder allele in which the *AluyHJ* 2* insertion occurred. In comparison, the *HLA-A* 11* allele was found to be associated with one or other of the three different Alu elements as independent haplotypes. This suggests that the Alu insertions either occurred in different individuals with the same *HLA-A* allele or that there have been relatively high rates of recombination between individuals with the *HLA-A* 11* allele and different POALINs. In this case, the majority (founder) of *HLA-A* 11* alleles are without the Alu insertions, although we cannot dismiss the possibility of the existence of a strong association between a POALIN and a larger subgroup of *HLA-A* 11* alleles within a population group that we have not as yet studied. Nevertheless, it is evident from the present analysis that POALINs have considerable value as lineage and linkage markers for the study of human MHC genetics, diversity and evolution. Moreover, the five MHC POALINs used here to examine their relationship with *HLA-A* and *HLA-B* alleles, could be further extended to examine their associations with the alleles of other gene loci, such as *HLA-C*, *MICA* and *MICB* within the beta block and *HLA-G* within the alpha block.

Acknowledgements

CL was a TRF Research Scholar supported by Thailand Research Fund.

References

Antunez-de-Mayolo, G., Antunez-de-Mayolo, A., Antunez-de-Mayolo, P., Papiha, S. S., Hammer, M., Yunis, J. J., Yunis, E. J., Damodaran, C., Martinez de Pancorbo, M., Caeiro, J. L., Puzyrev, V. P. & Herrera, R. J. (2002) Phylogenetics of worldwide human populations as determined by polymorphic Alu insertions. *Electrophoresis* **23**, 3346–3356.

Batzer, M. A., Kilroy, G. E., Richard, P. E., Shaikh, T. H., Desselle, T. D., Hoppens, C. L. & Deininger, P. L. (1990) Structure and variability of recently inserted Alu family members. *Nucleic Acids Res* **18**, 6793–6798. (Erratum in *Nucleic Acids Res* 1991, **19**, 698–699).

Begovich, A. B., McClure, G. R., Suraj, V. C., Helmuth, R. C., Fildes, N., Bugawan, T. L., Erlich, H. A. & Klitz, W. (1992) Polymorphism, recombination, and linkage disequilibrium within the HLA class II region. *J Immunol* **148**, 249–258.

Bengtsson, B. O. & Thomson, G. (1981) Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* **18**, 356–363.

Carrington, M. (1999) Recombination within the human MHC. *Immunol Rev* **167**, 245–256.

Carroll, M. L., Roy-Engel, A. M., Nguyen, S. V., Salem, A. H., Vogel, E., Vincent, B., Myers, J., Ahmad, Z., Nguyen,

L., Sammarco, M., Watkins, W. S., Henke, J., Makalowski, W., Jorde, L. B., Deininger, P. L. & Batzer, M. A. (2001) Large-scale analysis of the Alu Ya5 and Yb8 subfamilies and their contribution to human genomic diversity. *J Mol Biol* **311**, 17–40.

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.

Deininger, P. L. & Batzer, M. A. (1999) Alu repeats and human disease. *Mol Genet Metab* **67**, 183–193.

Dunn, D. S., Inoko, H. & Kulski, J. K. (2003a) Identification and characterisation of a dimorphic Alu located between the TFIIH and CDSN genes within the Major Histocompatibility. *Electrophoresis* **24**, 2740–2748.

Dunn, D. S., Naruse, T., Inoko, H. & Kulski, J. K. (2002) The association between HLA-A alleles and young Alu dimorphism near HLA-J, -H and -F genes in workshop cell lines and Japanese and Australian populations. *J Mol Evol* **55**, 718–726.

Dunn, D. S., Ota, M., Inoko, H. & Kulski, J. K. (2003b) Association of MHC dimorphic Alu insertions with HLA class I and MIC genes in Japanese HLA-B48 haplotypes. *Tissue Antigens* **62**, 259–262.

Holloway, J. W., Beghe, B., Turner, S., Hinks, L. J., Day, I. N. & Howell, W. M. (1999) Comparison of three methods for single nucleotide polymorphism typing for DNA bank studies: sequence-specific oligonucleotide probe hybridisation, TaqMan liquid phase hybridisation, and microplate array diagonal gel electrophoresis (MADGE). *Hum Mutat* **14**, 340–347.

Kulski, J. K., Dunn, D. S., Hui, J., Martinez, P., Romphruk, A. V., Leelayuwat, C., Tay, G. K., Oka, A. & Inoko, H. (2002a) Alu polymorphism within the MICB gene and association with HLA-B alleles. *Immunogenetics* **53**, 975–979 (Erratum in *Immunogenetics* 2002, **54**, 365).

Kulski, J. K., Martinez, P., Longman-Jacobsen, N., Wang, W., Williamson, J., Dawkins, R. L., Shiina, T., Naruse, T. & Inoko, H. (2001) The association between HLA-A alleles and an Alu dimorphism near HLA-G. *J Mol Evol* **53**, 114–123.

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

Lander, E. S., Linton, L. M., Birren, B.,*et al.*, (2001) Initial sequencing and the analysis of the human genome. *Nature* **409**, 860–921.

Li, S., Kawata, H., Katsuyama, Y., Ota, M., Morishima, Y., Mano, S., Kulski, J. K., Naruse, T. & Inoko, H. (2004) Association of polymorphic MHC microsatellites with GVHD, survival, and leukemia relapse in unrelated hematopoietic stem cell transplant donor/recipient pairs matched at five HLA loci. *Tissue Antigens* **63**, 362–368.

Mighell, A. J., Markham, A. F. & Robinson, P. A. (1997) Minireview: Alu sequences. *FEBS Letters* **417**, 1–5.

Ott, J. (1992) Strategies for characterizing highly polymorphic markers in human gene mapping. *Am J Hum Genet* **51**, 283–290.

Pascual, M., Mataran, L., Jones, G., Shing, D., van der Slik, A. R., Giphart, M. J., Schreuder, G. M., de Vries, R. R., Breedveld, F. C., Roovers, E., Zanelli, E. & Martin, J. (2002) HLA haplotypes and susceptibility to rheumatoid arthritis. More than class II genes. *Scand J Rheumatol* **31**, 275–278.

Romphruk, A. V., Naruse, T. K., Romphruk, A., Kawata, H., Puapairoj, C., Kulski, J. K., Leelayuwat, C. & Inoko, H. (2001) Diversity of MICA (PERB11.1) and HLA haplotypes in Northeastern Thais. *Tissue Antigens* **58**, 83–89.

Rowold, D. J. & Herrera, R. J. (2000) Alu elements and the human genome. *Genetica* **108**, 57–72.

Roy-Engel, A. M., Carroll, M. L., Vogel, E., Garber, R. K., Nguyen, S. V., Salem, A. H., Batzer, M. A. & Deininger, P. L. (2001) Alu insertion polymorphisms for the study of human genomic diversity. *Genetics* **61**, 6640–6648.

Schneider, S., Roessli, D. & Excofier, L. (2000) Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept of Anthropology, University of Geneva.

Stoneking, M., Fontius, J. J., Clifford, S. L., Soodyall, H., Arcot, S. S., Saha, N., Jenkins, T., Tahir, M. A., Deininger, P. L. & Batzer, M. A. (1997) Alu insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* **7**, 1061–1071.

The, MHC sequencing consortium. (1999) Complete structure and gene maps of a human major histocompatibility complex (MHC). *Nature* **401**, 921–923.

Watkins, W. S., Ricker, C. E., Bamshad, M. J., Carroll, M. L., Nguyen, S. V., Batzer, M. A., Harpending, H. C., Rogers, A. R. & Jorde, L. B. (2001) Patterns of ancestral human diversity: an analysis of Alu-insertion and restriction-site polymorphisms. *Am J Hum Genet* **68**, 738–752.

Walsh, E. C., Mather, K. A., Schaffner, S. F., Farwell, L., Daly, M. J., Patterson, N., Cullen, M., Carrington, M., Bugawan, T. L., Erlich, H., Campbell, J., Barrett, J., Miller, K., Thomson, G., Lander, E. S. & Rioux, J. D. (2003) An integrated haplotype map of the human major histocompatibility complex. *Am J Hum Genet* **73**, 580–590.

Yunis, E. J., Larsen, C. E., Fernandez-Vina, M., Awdeh, Z. L., Romero, T., Hansen, J. A. & Alper, C. A. (2003) Inheritable variable sizes of DNA stretches in the human MHC: conserved extended haplotypes and their fragments or blocks. *Tissue Antigens* **62**, 1–20.

Received: 14 March 2004
Accepted: 17 November 2004

