times from the 5′ and 3′ ends and analyzed for the presence of SNPs at the specific genomic loci, *HG* or *AK*, and nucleotide sequence variation between the *HG* and *AK* genomic loci.

The SNPs at 11 nucleotide positions for the *3′ MER9.HG* sequences that were linked to one of six common *HLA-A* alleles (*HLA-A3*, *-A11*, *-A24*, *-A30*, *-A31*, and *-A33*) are shown in Table 2. The *3′ MER9.HG* sequence linked to *HLA-A3* was compared against the *3′ MER9.HG* sequences linked to the other five *HLA-A* alleles. The *3′ MER9.HG/HLA-A3* haplotype sequence was differentiated from the other *3′ MER9.HG/HLA-A* haplotypes at nine different *3′ MER9.HG* nucleotide positions, but at only nucleotide position 201 when compared to the *HLA-A31* haplotype. The *3′ MER9.HG* sequences associated with *HLA-A31* and *-A33* were identical. The *3′ MER9.HG/HLA-A24* and *3′ MER9.HG/HLA -A30* haplotypes were differentiated from each other by only one *3′ MER9.HG* nucleotide (position 493), even though up to six different *3′ MER9.HG* nucleotide positions differentiated them from the *3′ MER9.HG/HLA-A3* haplotype. In addition, the two nucleotides at positions −64 and −60 upstream of the *3′ MER9.HG* sequence and located within the internal *HERVK9* sequence differentiated the *3′ MER9.HG/HLA-30* haplotype from the other haplotypes. In comparison to the one to six nucleotide differences for the *3′ MER9.HG/HLA-A* sequence haplotypes, the number of nucleotide differences between the *3′ MER9.HG/HLA-A* haplotypes and the chimpanzee or gorilla *3′ MER9* or *5′ MER9* sequences were substantially greater at 12 to 35 nucleotide differences. The *3′ MER9* sequence of the orangutan (*Popy*) had an 87 bp deletion compared to its *5′ MER9* sequence (accession number AB453920). There was no sequence difference for the multiple *DNA* samples with the same *3′ MER9.HG/HLA-A* haplotypes as exemplified by the *3′ MER9.HG DNA* sequences of the seven *HLA-A3* and five *HLA-A24* homozygous *DNA* samples (Table 2).

The SNPs at nine loci for the *sMER9.HG* sequences that were linked to one of eight different *HLA-A* alleles (*HLA-A1*, *-A2*, *-A10*, *-A11*, *-A26*, *-A29*, *-A32*, and *-A34*) are presented in Table 3. The *sMER9.HG* sequence linked to *HLA-A1* was compared against the *sMER9.HG* sequences that were linked to the other seven *HLA-A* alleles. There was at least one *MER9* haplospecific nucleotide marker for five of the eight *sMER9.HG/HLA-A* haplotypes. An identical s*MER9* sequence was obtained for the three haplotype pairs, *sMER9.HG/A2* and *sMER9.HG/A11*, *sMER9.HG/A10* and *sMER9.HG/A34*, and between *sMER9.HG/A26* and s*MER9.HG/A29*. The greatest sequence difference of six nucleotides was obtained between the *sMER9* sequences of the *sMER9.HG/A1* and *sMER9.HG/A34* haplotypes. In comparison to the small number of differences (up to six) for the *sMER9.HG/HLA-A*, the number of nucleotide differences between the *sMER9.HG/HLA-A* haplotypes and the human, chimpanzee, or gorilla *3′ MER9* or *5′ MER9* sequences was much larger at 24 to 33 nucleotide differences. There was no sequence difference for multiple *DNA* samples with the same *sMER9.HG/HLA-A* haplotype as exemplified by the *sMER9.HG DNA* sequences of the 12 *HLA-A1* and 19 *HLA-A2* homozygous *DNA* samples (Table 3).

The differences at 31 nucleotide positions for the *sMER9.AK* sequences that were linked to one of seven different HLA-A alleles (*HLA-A1*, *-A2*, *-A3*, *-A26*, *-A29*, *-A32*, and *-A34*) and the chimpanzee *PatrA* allele are shown in Table 4. The *sMER9.AK* sequence linked to *HLA-A2* was compared against the *sMER9.AK* sequences linked to the other *HLA-A* and *PatrA* alleles. There is at least one haplotypic *MER9* nucleotide marker and up to four for five of the seven *sMER9.AK/HLA-A* haplotypes. Overall, there were up to 29 SNP sites that differentiated between the seven *sMER9.AK/HLA-A* haplotypes. The number of nucleotide differences in a comparison between the *sMER9.AK/HLA-A2* haplotype and other human *sMER9.AK/HLA-A* haplotypes was generally higher (2.7–3.1%) than the differences between the *sMER9.AK/HLA-A2* haplotype and the chimpanzee (*Patr*) *sMER9.AK/PatrA* haplotype (2.2%).

In order to assess the degree of *MER9-LTR.AK* sequence variation associated with the same *HLA-A* alleles, we sequenced the *MER9.AK* solo elements for 11 *HLA-A1*, 14 *HLA-A2*, seven *HLA-A3*, and four *HLA-A26 DNA* samples extracted from the homozygous and heterozygous *HLA-A* typed cell lines (Table 1). Little or no *MER9* sequence variation was detected within the same *MER9-LTR.AK/HLA-A* haplotypes, such as *MER9-LTR.AK/HLA-A1*, *MER9-LTR.AK/HLA-A3* alleles, except for the *MER9-LTR.AK/HLA-A2* haplotype at nucleotide position 94 (C/A) of the *sMER9.AK* sequences (Table 4).

MER9-LTR phylogenetic tree

Figure 2 shows a phylogenetic tree constructed by the maximum parsimony method of 36 *MER9-LTR* sequences including the solo and flanking *MER9.HG* sequences and the *sMER9.AK* sequences associated with different *MER9/HLA-A* haplotypes. The *5′ MER.HC.HosaB44C16* sequence that is located telomeric of the *HLA-C* gene as part of a *HERVK9* sequence (Kulski et al. 1999) and present in the MANN cell line (*HLA-A29/HLA-B44/HLA-C16*) was used as the out-group sequence for reconstructing the tree. The rhesus macaque (Old World monkey) solo *MER9* sequences that are duplicated at the *BALSL1* to *BALSL6* loci within the *Mhc alpha* block (Kulski et al. 2004) clustered together with the *5′ MER9.HLA-C* sequence as part of the outgroup for the *AK* and *HG MER9* sequences. The human and non-human hominid *AK* and *HG MER9* sequences divided into

Table 2 Nucleotide differences at 11 basepair positions for *3′ MER9.HG* sequence alignments (510/512 bp) associated with homozygous *HLA-A* alleles

Haplotype	−64	−60	Basepair position in sequence alignment									% nt difference to *3′ MER9.HG/A3*	No. nt differences to *3′ MER9.HG/A3*	Gaps no. nt	Sample no. this study	Sample no. total	Accession no. ref. sequence
			101	175	200	201	227	384	385	395	493						
3′ MER9.HG/A3	A	G	G	T	A	A	G	C	G	C	T	0	0	0	7	7	AB443936
3′ MER9.HG/A11	A	G	G	T	A	A	G	T	G	A	T	0.4	2	0	1	1	AB443932
3′ MER9.HG/A24	A	G	A	C	G	A	A	C	A	C	T	1	5	0	5	5	AB443933
3′ MER9.HG/A30	G	A	A	C	G	A	A	C	A	C	C	1.2	6	0	1	1	AB443934
3′ MER9.HG/A31	A	G	G	T	A	G	G	C	G	C	T	0.2	1	0	4	4	AB443935
3′ MER9.HG/A33	A	G	G	T	A	G	G	C	G	C	T	0.2	1	0	4	4	AB447379
3′ MER9.HG Patr	G	G	G	T	A	A	A	G	G	C	T	2.9	15	1		1	AC192848
3′ MER9.HG Gogo	G	G	G	T	A	A	A	G	G	C	T	2.3	12	1		1	CU104658
3′ MER9.HG. Popy	A	A	A	T	A	A	G	-	-	-	T	7.7	33 of 426 nt	87	1	1	AB453920
5′ MER9.HG/A3			G	T	A	A	G	G	G	C	T	6.8	35	1		1	AL645929
5′ MER9.HG Patr			G	T	A	A	G	G	G	C	T	6	31	2		1	AC192848
5′ MER9.HG Gogo			G	T	A	A	G	G	G	C	T	6	31	1		1	CU104658
5′ MER9.HG. Popy			G	T	A	A	G	G	G	C	C	5.7	29	0	1	1	AB453920
5′ MER9.HLA-C			G	T	A	G	G	A	A	C	C	15	77	1		1	CR847781

A3, *A11*, *A24*, *A30*, *A31*, *A33* are the *HLA-A* alleles linked to the *MER9* sequences at the *HG* locus.

Patr, *Gogo* and *Popy* are chimpanzee, gorilla and orangutan *MER9* sequences, respectively, at the orthologous *HG* locus.

5′MER9.HLA-C is the *MER9* sequence in close proximity to the *HLA-C* locus.

Basepair positions -64 and -60 are within the three prime-end of the *HERVK9* internal sequence.

The four samples for *3′MER9.HG/A31* includes the heterozygous cell-lines 54 and 55 (Table 1).

The deleted nucleotide sequences (gaps) are not included in the base difference count.

Table 3 Nucleotide differences at nine basepair positions for solo *MER9.HG* sequences (512 bp) associated with homozygous *HLA-A* alleles

Haplotype	Basepair position in multialignment of *MER9* sequences									% difference to *MER9.HG/A1*	No. nt differences to *MER9.HG/A1*	Gaps No. nt	Sample no. this study	Sample no. total	Accession no. Ref. Sequence
	120	123	136	157	211	227	274	295	388						
sMER9.HG/A1	A	T	G	C	C	A	G	T	C	0	0	0	11	12	AB447380
sMER9.HG/A2	A	T	G	C	C	G	A	T	C	0.4	2	0	16	19	AB447382
sMER9.HG/A10	G	C	C	C	A	G	G	A	C	1.3	6	0	1	1	AB447383
sMER9.HG/A11	A	T	G	C	C	G	A	T	C	0.4	2	0	2	2	AB447384
sMER9.HG/A26	G	T	G	C	C	G	G	T	C	0.4	2	0	4	4	AB447381
sMER9.HG/A29	G	T	G	C	C	G	G	T	C	0.4	2	0	2	2	AB447385
sMER9.HG/A32	G	T	G	T	C	G	G	T	A	0.7	4	0	0	1	BX284699
sMER9.HG/A34	G	C	C	C	A	G	G	A	C	1.3	6	0	2	2	AB443937
sMER9.AK/A1	G	T	G	C	T	G	G	T	C	9.4	51	12	11	12	AB447373
sMER9.AK/A2	G	T	G	C	T	G	G	T	C	9.4	55	10	16	19	AB447374
sMER9.AK/Patr	G	T	G	C	C	G	G	T	C	8.9	52	10		1	AC192848
5′ MER9.HG/A3	G	T	G	C	T	G	G	T	C	4	24	0		1	AL645929
3′ MER9.HG/A3	A	T	G	C	C	G	G	T	C	6.6	31	0	7	7	AB443936
5′ MER9.HG/Patr	G	T	G	C	C	G	G	T	C	4.3	25	0		1	AC192848
3′ MER9.HG/Patr	A	T	G	C	C	A	G	T	C	6.9	33	2		1	AC192848
5′ MER9.HG/Gogo	G	T	G	C	C	G	G	T	C	4.3	24	1		1	CU104658
3′ MER9.HG/Gogo	G	T	G	C	C	A	G	T	C	6.2	31	2		1	CU104658

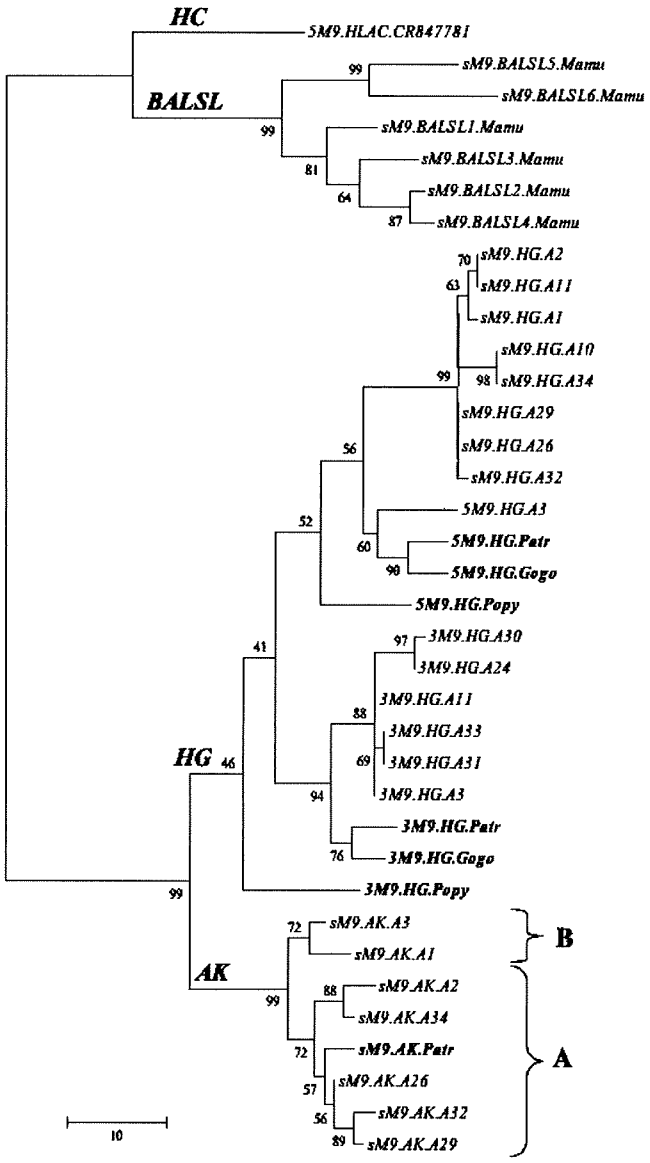
A1, *A2*, *A10*, *A11*, *A26*, *A29*, *A32*, and *A34* are the *HLA-A* alleles linked to the *sMER9* sequences at the *HG* or *AK* locus. 5′ and 3′ *MER9* human sequences (*HG/A3*) and non-human primates (*Patr* and *Gogo*) are included for comparison. Increased sample numbers in the total column include sequences found in GenBank. The deleted nucleotide sequences (gaps) are not included in the base difference count. Sample numbers in this study include heterozygous cell lines (Table 1)

Table 4 Nucleotide differences at 31 positions for solo *MER9.AK* sequences (503 bp) linked to the listed HLA-A alleles *A1*, *A2*, *A3*, *A26*, *A29*, *A32*, and *A34*

Nucleotide position	*sMER9.AK* nucleotides associated with *HLA-A* alleles							
	A1	*A2*	*A3*	*A26*	*A29*	*A32*	*A34*	*Patr*
42	C	T	C	C	C	C	T	C
45	G	C	G	G	G	G	G	G
50	G	G	G	G	G	T	G	G
59	A	A	A	A	A	A	A	G
64	T	C	T	C	C	C	C	C
93	G	C	G	G	G	G	C	G
94	C	C/A	C	C	C	C	C	C
96	C	T	C	T	T	T	T	T
114	G	G	G	G	T	G	G	G
127	C	T	C	T	C	C	T	C
128	A	G	G	G	A	A	G	G
144	A	A	G	A	A	A	A	A
153	C	C	C	C	C	T	C	C
155	C	T	C	C	C	C	C	C
175	T	T	T	T	T	T	T	C
251	A	G	A	A	A	A	G	A
273	C	T	C	T	T	T	T	T
282	C	C	C	A	A	A	C	A
322	T	C	T	C	C	C	C	C
356	G	G	G	G	A	A	G	G
357	T	C	T	T	T	T	C	T
373	T	C	T	C	C	C	C	C
381	Gap	T	Gap	T	T	T	T	T
382	Gap	C	Gap	C	C	C	C	C
428	A	A	A	A	G	G	A	A
462	G	A	A	A	A	A	A	A
464	T	C	C	C	C	C	T	C
465	G	G	G	A	A	A	G	G
476	G	A	A	A	A	A	A	A
487	C	T	C	C	C	C	C	T
494	A	G	G	G	G	G	G	G
No. haplospecific bases	3	4	1	0	1	2	0	
No. base diff. v *sMER9.AK/A2*	18	0	14	9	15	16	4	12
% difference v *sMER9.AK/A2* wrt *sMER9.AK/A2*	3.1	0	2.7	1.8	2.9	3.1	0.8	2.2
No. gaps	3	0	2	0	0	0	0	0
No. samples this study (Table 1)	11	14	7	4	2	0	2	0
No. samples total	11	15	8	4	2	1	2	1
Accession no. (sequence reference)	AB447373	AB447374 AB447375	AB447376	AB447377	CR382333	BX005091	AB447378	AC192848

A1, *A2*, *A3*, *A26*, *A29*, *A32*, and *A34* are the *HLA-A* alleles linked to the *sMER9* sequences at the *AK* locus. *Patr* is a *sMER9* chimpanzee sequence at the *AK* locus. Nucleotide deletions (gaps) are not included in the base difference (diff.) count. The number of total samples sequenced includes the sequences found in GenBank at NCBI

Sample numbers in this analysis include the heterozygous cell lines (Table 1)



◀ **Fig. 2** Phylogenetic tree of 5′, 3′, and solo *MER9 DNA* nucleotide sequences from the *AK* or *HG* loci that were associated with particular *HLA-A* alleles, chimpanzee *PatrA*, gorilla *GogoA*, and orangutan *PopyA*. The tree was reconstructed by the maximum parsimony method. The labeled sequences in the tree indicate the *MER9* type (5′, 3′, or solo), the *MER9* locus (*AK*, *HG*, or *HLA-C*), and the *HLA-A* allele (A#) linked to the human *MER9* sequence or identified in the chimpanzee (*Patr*), gorilla (*Gogo*), or orangutan (*Popy*) *Mhc*. The *Gogo* (*Gorilla gorilla*), *Patr* (*Pan troglodytes*), and *Popy* (*Pongo pygmaeus*) *MER9* sequences are emphasized by blocked letters. *Mamu* (*Macaca mulatta*) is rhesus macaque. The two main ancestral *HLA-A* allelic clusters are represented by *A* and *B*, where in previous branching studies the *HLA-A2*, *-A26*, *-A29*, *-A32*, and *-A34* grouped together, separately from *HLA-A1* and *-A2* allelic group (McKenzie et al. 1999; Kulski et al. 2001). The *HC* lineage is the *MER9* sequence near the *HLA-C* gene and it is used as an outgroup. The *BALSL* lineage is the genomic region of retroelements within the *Mamu Mhc* and near to the *MamuA* gene, and it is present in rhesus macaque, but not the human, as previously described by Kulski et al. (2004). *HG* and *AK* are the lineages represented by the *MER9* sequences within the *HG* and *AK* loci of the human and ape *Mhc* region. The bootstrap values are shown at the nodes as a percentage. The *horizontal bar* that is labeled *10* means 10 nt substitutions for the indicated branch length

two main clusters, one cluster with all the sequences from the *AK* locus and the other cluster with all the sequences from the *HG* locus. The mean (±SD) percentage distance of the *MER9* sequence at the *HLA-C* locus was 18.2% (±0.70%) from the *HG* locus and 16.9% (± 0.38%) from the *AK* locus, which was significant ($p<0.05$) in the unpaired *t* test. The 5′ and 3′ *MER9* sequences clustered into separate groups, apparently diverging from each other and from the 5′ and 3′ *MER9* ancestral sequences that were probably identical when the *HERVK9* retrovirus was inserted within the *HG* locus. Generally, the divergence of the 5′ and 3′ *MER9* sequences was more pronounced than between the 5′ *MER9* or the 3′ *MER9* group.

The *sMER9.HG* sequences from different *HLA-A* haplotypes grouped together separately to the *5′* and *3′ MER9.HG* sequences, but more closely to the *5′ MER9.HG* sequences than to the *3′ MER9.HG* sequences. The *sMER9.HG* sequences diverged from each other according to their *HLA-A* haplotype differences. No sequence differences were found between the *MER9.HG* from different cell lines with the same *HLA-A/MER9* haplotypes. The *sMER9.AK* sequences separated into two groups, A and B, similar to the previously reported phylogenetic grouping of the *HLA-A* alleles where *HLA-A2*, *-A26*, *-A29*, *-A32*, and *-A34* grouped together (group A) separately to *HLA-A1* and *-A2* (group B; McKenzie et al. 1999; Kulski et al. 2001), suggesting that the *HLA-A* alleles are linked more strongly with the *MER9.AK* than the *MER9.HG* allelic markers. The *MER9 AK* sequences were absent from the gorilla (accession number CU104664) and the orangutan (unpublished data) *Mhc* genomic sequences and therefore are not part of this analysis.

Discussion

The polymorphic *MER9-LTR* sequences at the *AK* and *HG* loci are potentially informative *HLA-A* haplotypic and stratification markers for population and disease studies because of their association with *HLA-A* alleles (Kulski et al. 2008). The *5′*, *3,′* and *sMER9 DNA* nucleotides at the *HG* or *AK* loci have from nine to 29 SNP sites for the assessment of *MER9* polymorphisms, haplotypes, and associations with *HLA-A* alleles. The difference in the number of SNP diagnostic positions associated with particular *HLA-A* alleles for the *MER9-LTR* sequences at the *AK* and *HG* loci appears to be related to the different genomic distances between the *MER9* and *HLA-A* loci and/or to the relative age of the *MER9* loci associated with the *HLA-A* alleles. The solo *MER9.HG* and *3′ MER9.HG* sequences had at least nine SNP loci with one or more

associated with different common *HLA-A* alleles. However, some of the *sMER9-HG/HLA-A* haplotypes, such as the *sMER9.HG/A26* and the *sMER9.HG/A29*, which had identical sequences for the *sMER9-LTR*, may have been formed by an ancestral crossing over of chromosomes with the recombination breakpoint located between the *HLA-A* and *MER9.HG* loci. In this regard, the nucleotide sequences at the *HLA-A*, *MER9.AK*, and *MER9.HG* loci might be used as haplotype lineage markers or as crossing-over markers to identify and assess the frequency of crossing over events within this genomic region and possibly for stratification and classification of particular *HLA-A* alleles into different *HLA-A/MER9* sequence haplotypes in population and disease studies.

No *MER9* sequence variation was detected in samples with the same *HLA-A* allelic samples except for a single nucleotide variation (G/C) that was identified for the *MER9.AK/HLA-A2* haplotypes. Thus, the association between *MER9-LTR* nucleotide sequences and common *HLA-A* alleles appears to be conserved, although substantially more analyses of the same and different *MER9/HLA-A* haplotypes will need to be performed to estimate the degree of association stability. In this study, we concentrated on the relationships of the *MER9.HG* or *MER9.AK* SNPs with the *HLA-A* alleles separately, but combining the SNPs results at the three loci may substantially increase the power of future haplotype analyses within this genomic region. In addition, the number and variety of SNPs were not investigated within the *5′ MER9* sequence at the *HG* locus. The present analysis of the *MER9 DNA* nucleotide sequence variations nevertheless demonstrates a potentially useful rate of nucleotide mutations within the *MER9-LTR* and highlights distinct *MER9* nucleotide differences between a number of different *MER9/HLA-A* haplotypes that warrant further studies at the level of population diversity and disease associations.

The number of nucleotide differences between *sMER.AK/HLA-A2* and the *sMER.AK*s of the other *HLA-A* haplotypes was higher than between *sMER.AK/HLA-A2* and the chimpanzee *sMER9.AK/PatrA*. This difference suggests that the *sMER9.AK/HLA-A2* is either the oldest or youngest *MER9* allele compared to the other *sMER9.AK/HLA-A* alleles. Because the sequence difference was smaller between *sMER9.AK/HLA-A2* and *sMER9.AK/PatrA* (2.2%) than between *sMER9.AK/HLA-A2* and the other human *sMER9.AK* alleles (2.7–3.1%), it could be surmised that the *sMER9.AK/HLA-A2* sequence is the oldest of the *sMER9.AK/HLA-A* alleles in this study. Alternatively, the *sMER9.AK* nucleotide sequence in linkage with *HLA-A2* might have mutated at a faster rate than the other *sMER9.AK* alleles due to some unspecified form of selection pressure. More comparative analyses between human and chimpanzee *MER9.AK* alleles may help to resolve these observed differences.

The usefulness of the *MER9-LTR* sequences at the *AK* and *HG* loci as evolutionary and haplotype markers stems from their structural polymorphism as well as their nucleotide variations. The *HERVK9* and associated *MER9-LTR* structural polymorphisms (absence or presence of the endogenous retrovirus) have at least four possible and distinct ancestral states, the initial "empty" state that has no inserted sequence, the "occupied" ancestral state with the proviral insertion, the "modified" ancestral state with the remnants of a solitary *MER9-LTR* marker remaining at the location of the proviral deletion, and the "lost" ancestral state due to a genomic deletion or rearrangement of the insertion locus. While the first and second states are once only events for a specific genomic location, the third and fourth states might have happened on a number of occasions involving different *HLA-A* haplotypes and individuals within a population. The results of the phylogenetic analysis of the *MER9-LTR* sequences suggest that the generation of the solo *MER9* sequence by the deletion of its *HERVK9* sequence at the *HG* locus was probably a single event and that the different s*MER9-LTR/HLA-A* haplotypes were then generated by sequence mutation and other mechanisms such as unequal crossing over within the genomic region between the *HLA-A* and *HERVK9.HG* loci (Kulski et al. 2008). The small number of nucleotide differences (up to 1.3%) detected between the *sMER9.HG* sequences from different HLA-A haplotypes implies that the *sMER9.HG/HLA-A* haplotypes identified in this study first emerged about 3 or 4 Myrs after the separation of the human and chimpanzee lineage, 6–7 Mya (Goodman et al. 1998; Steiper and Young 2006). The *sMER9.HG* sequences are different to the *5′* and *3′ MER9.HG* sequences on average by 5.5%, which suggests that the *sMER9.HG* first appeared as a *HERVK9* deletion 15.1 Mya, only a few Myrs after its integration into the ancestral *HG* locus 18.3 Mya (Kulski et al. 2008). The estimate of the *HERVK9* integration date at 18.3 Mya is consistent with the physical data available for the presence of the *HG* locus in hominids (orangutan, gorilla, chimpanzee, and human) and its absence in the Old World monkeys such as rhesus macaque (Kulski et al. 2004).

The *HERVK9/MER9.HG* and *MER9.AK* loci appear to have originated from a multigenic duplication event about the same time or after the *HERVK9/MER9* was first inserted into the progenitor *DNA* segment (Kulski et al. 2005). The direct evidence for the generation of the *MER9-LTR HG* and *AK* loci by duplication is that the same flanking sequences exist at the *AK* and *HG* insertion loci (Fig. 1c) rather than different flanking sequences as would be expected if the *HERVK9* integrations at the *HG* and the *AK* loci were separate events. Comparatively, however, the sequence differences (0.4–1.3%) between the *sMER9* at the *HG* locus were at least three times less than the sequence differences (up to 4.2%) between the *sMER9* at

the *AK* locus. In addition, the divergence rates appear to be slower between the *5′* and *3′ MER9* sequences at the *HG* locus than between the *solo MER9* sequences at the *HG* and *AK* loci possibly due to selection forces or a hitchhiking effect of the *HLA-A* locus (Shiina et al. 2006). The *HLA-A* region encompassing the *AK* locus was previously found to generate a much greater SNP diversity than the *HG* locus in genomic comparative studies between different *HLA* haplotypes (Gaudieri et al. 2000; Stewart et al. 2004; Horton et al. 2008) and between the human and chimpanzee *MHC* class I genomic sequences (Anzai et al. 2003).

The absence of the *MER9.AK* locus in the gorilla and the orangutan is possibly due to a single deletion event in an ancestral species or to a multiple deletion event after the emergence of these species, similar to the deletion of *MER9.AK* locus in humans with the *HLA-A24* allele (Table 1). In spite of the likelihood of different mutation rates and deletion events at the *HG* and *AK* loci, the *HERVK9* insertion and duplication, whether considered as two independent events or as a single combined event, appear to have occurred at a time well after the separation of the human and the OWM lineages 28 Mya (Goodman et al. 1998; Steiper and Young 2006). Whereas some *HERVK9* sequences at other genomic positions were integrated into the ancestral human genome about 35 Mya and before the emergence of the OWM (Mayer and Meese 2005), the *HERVK9.HG* insertion is present in the *Mhc* of the chimpanzee (Kulski et al. 2005), the gorilla (Sanger Institute, NCBI accession number CU104658), and the orangutan (accession number AB453920), but not in the orthologous locations of the rhesus macaque (Kulski et al. 2004), which is a member of the OWM.

In conclusion, the sequencing and analysis of the full-length *MER9* from clones of PCR-amplified heterozygous *DNA* or directly from PCR-amplified homozygote *DNA* is a relatively rapid, easy, and inexpensive laboratory method. The *MER9 DNA* sequence mutations are informative genetic markers that might be used in fine mapping *MHC-A* genomic haplotypes for population and evolutionary studies in humans as well as the chimpanzee, gorilla, and orangutan. The analysis of *MER9* SNP and structural variation in combination with *HLA-A* gene haplotypes in the *MHC* region is potentially useful for the identification and analysis of new haplotypes generated by crossover events.

References

Anzai T, Shiina T, Kimura N, Yanagiya K, Kohara S, Shigenari A, Yamagata T, Kulski JK, Naruse TK, Fujimori Y et al (2003) Comparative sequencing of human and chimpanzee MHC class I regions unveils insertions/deletions as the major path to genomic divergence. Proc Natl Acad Sci USA 100:7708–7713. doi:10.1073/pnas.1230533100

Batzer MA, Deininger PL (2002) Alu repeats and human genomic diversity. Nat Rev Genet 3:370–379. doi:10.1038/nrg798

Belancio VP, Hedges DJ, Deininger P (2008) Mammalian non-LTR retrotransposons: for better or worse, in sickness and in health. Genome Res 18:343–358. doi:10.1101/gr.5558208

Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE (2004) Natural genetic variation caused by transposable elements in humans. Genetics 168:933–951. doi:10.1534/genetics.104.031757

Dunn DS, Inoko H, Kulski JK (2006) The association between non-melanoma skin cancer and a young dimorphic Alu element within the major histocompatibility complex class I genomic region. Tissue Antigens 68:127–134. doi:10.1111/j.1399-0039.2006.00631.x

Dunn DS, Choy MK, Phipps ME, Kulski JK (2007) The distribution of major histocompatibility complex class I polymorphic Alu insertions and their associations with HLA alleles in a Chinese population from Malaysia. Tissue Antigens 70:136–143. doi:10.1111/j.1399-0039.2007.00868.x

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. doi:10.1101/gr.127200

Goodman M, Porter CA, Czelusniak J, Page SL, Schneider H, Shoshani J, Gunnell G, Groves CP (1998) Toward a phylogenetic classification of Primates based on DNA evidence complemented by fossil evidence. Mol Phylogenet Evol 9:585–598. doi:10.1006/mpev.1998.0495

Hampe A, Coriton O, Andrieux N, Carn G, Lepourcelet M, Mottier S, Dréano 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 Seq 10:263–299

Horton R, Gibson R, Coggill P, Miretti M, Allcock RJ, Almeida J, Forbes S, Gilbert JG, Halls K, Harrow JL, Hart E et al (2008) Variation analysis and gene annotation of eight MHC haplotypes: the MHC Haplotype Project. Immunogenetics 60:1–18. doi:10.1007/s00251-007-0262-2

Kapitonov VV, Pavlicek A, Jurka J (2004) Anthology of human repetitive DNA. In: Meyers RA (ed) Encyclopedia of Molecular Cell Biology and Molecular Medicine, vol. 1. Wiley-VCH Verlag GmbH and Co, KGaA Weinheim, pp 251–305

Kulski JK, Dunn DS (2005) Polymorphic Alu insertions within the Major Histocompatibility Complex class I genomic region. A brief review. Cytogenet Genome Res 110:193–202. doi:10.1159/000084952

Kulski JK, Gaudieri S, Inoko H, Dawkins RL (1999) Comparison between two human endogenous retrovirus (HERV)-rich regions within the major histocompatibility complex. J Mol Evol 48:675–683. doi:10.1007/PL00006511

Kulski JK, Martinez P, Longman-Jacobsen N, Wang W, Williamson J, Dawkins RL, 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. doi:10.1007/s002390010251

Kulski JK, Anzai T, Shiina T, Inoko H (2004) Rhesus macaque class I duplicon structures, organization, and evolution within the alpha block of the major histocompatibility complex. Mol Biol Evol 21:2079–2091. doi:10.1093/molbev/msh216

Kulski JK, Anzai T, Inoko H (2005) ERVK9, transposons and the evolution of MHC class I duplicons within the alpha-block of the human and chimpanzee. Cytogenet Genome Res 110:181–192. doi:10.1159/000084951

Kulski JK, Shigenari A, Shiina T, Ota M, Hosomichi K, James I, Inoko H (2008) Human endogenous retrovirus (HERVK9) structural polymorphism with haplotypic HLA-A allelic associations. Genetics 180:445–457. doi:10.1534/genetics.108.090340

Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921. doi:10.1038/35057062

Mager DL, Medstrand P (2003) Retroviral repeat sequences. In: Cooper DN (ed) Nature encyclopedia of the human genome, vol 5. Macmillan, London, pp 57–63

Mayer J, Meese E (2005) Human endogenous retroviruses in the primate lineage and their influence on host genomes. Cytogenet Genome Res 110:448–456. doi:10.1159/000084977

McKenzie L, Pecon-Slattery J, Carrington M, O'Brien S (1999) Taxonomic hierarchy of HLA class I allele sequences. Genes Immun 1:120–129. doi:10.1038/sj.gene.6363648

Medstrand P, Blomberg J (1993) Characterization of novel reverse transcriptase encoding human endogenous retroviral sequences similar to type A and type B retroviruses: differential transcription in normal human tissues. J Virol 67:6778–6787

Seifarth W, Frank O, Zeilfelder U, Spiess B, Greenwood AD, Hehlmann R, Leib-Mosch C (2005) Comprehensive analysis of human endogenous retrovirus transcriptional activity in human tissues with a retrovirus-specific microarray. J Virol 79:341–352. doi:10.1128/JVI.79.1.341-352.2005

Shiina T, Ota M, Shimizu S, Katsuyama Y, Hashimoto N, Takasu M, Anzai T, Kulski JK, Kikkawa E, Naruse T et al (2006) Rapid evolution of major histocompatibility complex class I genes in primates generates new disease alleles in humans via hitchhiking diversity. Genetics 173:1555–1570. doi:10.1534/genetics.106.057034

Steiper ME, Young NM (2006) Primate molecular divergence dates. Mol Phylogenet Evol 41:384–394. doi:10.1016/j.ympev.2006.05.021

Stewart CA, Horton R, Allcock RJ et al (2004) Complete MHC haplotype sequencing for common disease gene mapping. Genome Res 14:1176–1187. doi:10.1101/gr.2188104

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599. doi:10.1093/molbev/msm092

Terreros MC, Martinez L, Herrera RJ (2005) Polymorphic Alu insertions and genetic diversity among African populations. Hum Biol 77:675–704. doi:10.1353/hub.2006.0009

Watanabe Y, Tokunaga K, Geraghty DE, Tadokoro K, Juji T (1997) Large-scale comparative mapping of the MHC class I region of predominant haplotypes in Japanese. Immunogenetics 46:135–141. doi:10.1007/s002510050252

TISSUE ANTIGENS
IMMUNE RESPONSE GENETICS

Tissue Antigens ISSN 0001-2815

Polymorphic major histocompatibility complex class II *Alu* insertions at five loci and their association with *HLA-DRB1* and *-DQB1* in Japanese and Caucasians

J. K. Kulski[1,2], A. Shigenari[2], T. Shiina[2] & H. Inoko[2]

1 Centre for Forensic Science, The University of Western Australia, Nedlands, WA, Australia
2 Division of Molecular Life Science, Department of Genetic Information, School of Medicine, Tokai University, Isehara, Kanagawa, Japan

Key words
Alu; *human leukocyte antigen* class II alleles; dimorphism; haplotypes; major histocompatibility complex; polymorphism

Correspondence
Jerzy K. Kulski, PhD
Centre for Forensic Science
The University of Western Australia
Mailbag M420
35 Stirling Highway
Crawley
WA 6009
Australia
Tel: +61 8 6488 7286
Fax: +61 8 6488 7285
e-mail: jkulski@me.com

Received 31 October 2009; revised 29 December 2009, 13 January 2010; accepted 24 January 2010

doi: 10.1111/j.1399-0039.2010.01465.x

Abstract

We investigated polymorphic *Alu* insertion (*POALIN*) frequencies at five loci in the major histocompatibility complex (*MHC*) class II genomic region to determine their allele and haplotype frequencies and associations with the human leukocyte antigen (*HLA*)-*DRB1* and *-DQB1* genes for 100 Japanese, 174 Australian Caucasians and 67 *HLA* reference cell lines obtained from different ethnic groups. The *POALIN*s varied in frequency between 11% and 57% with significant differences between the Japanese and Caucasians at three loci. One *POALIN* locus deviated significantly from Hardy–Weinberg equilibrium (HWE) and four *POALIN* loci were in significant linkage disequilibrium and had a high percentage association with a variety of *HLA-DRB1* or *-DQB1* two-digit alleles. Inferred haplotype analysis among two-locus, five-locus and seven-locus haplotype structures showed maximum differences between the Japanese and Caucasians with the seven-locus haplotypes. The most common multilocus haplotype in Caucasians was *DRB1**1501/*DQB1**0602/*AluDQ1*/*AluDRB1*/*AluORF10*/*AluDPB2* (6.7%), whereas the second most common allele *HLA-DRB1**15 (17.5%) in Japanese was associated with three or four *Alu* insertions. The *HLA* class II *POALIN*s also differentiated within and between *HLA-DRB1* super-haplotypes *DR1, DR8, DR51, DR52* and *DR53*. This is the first comparative population study of multilocus *POALIN*s in the *HLA* class II region, which shows that *POALINs* whether investigated alone or together with the *HLA* class II alleles are informative genetic markers for the identification of allele and haplotype lineages and variations within the same and/or different populations.

Introduction

Polymorphic *Alu* insertions (*POALINs*) are either absent or present at their integration site and they are informative population markers that may carry signature alleles or haplotypes for different populations (1–3). They are potentially useful lineage and evolutionary markers because they have an inherited identity by descent arising from a known initial ancestral state (no *Alu* insertion), whereby their presence and/or absence define the ancestral lineages within a population (4). One million members of the *Alu* retrotransposon family contribute to ~10% of the human genomic content with most firmly fixed within the genome and only about 0.06% of the elements still actively mobile and structurally dimorphic in that they are either absent or present at the insertion site (1, 5). The presence or absence of some of the duplicated *Alu* J, S and Y family members within the major histocompatibility complex (*MHC*) genomic region have been used as evolutionary molecular signatures and clocks to infer the ancestral duplication history of the human leukocyte antigen (*HLA*) class I and class II gene copies (6–9). Most *POALIN*s belong to the *Alu*Y subfamilies, particularly the *Alu* Ya5 and Yb8 subgroups, (1, 10, 11). The *POALINs* are excellent genetic markers for population diversity, forensic and evolutionary studies because their allelic frequency distribution can vary markedly between geographically distinct human populations (3, 4, 12, 13). Although *POALIN* loci are distributed widely across the whole genome, some are clustered within close vicinity to each other to allow for a more formal haplotypic analysis of their combined allelic relationships, linkage disequilibrium (LD) and their

percentage associations with neighbouring gene alleles. For example, allele and haplotype frequencies and *HLA* class I gene associations for five *POALIN* loci within the 1.8-Mb *MHC* class I region of chromosome 6*p*21 have been investigated in Australian Caucasians, Japanese, North-eastern Thai, South and Central Saharan Africans (2), Chinese Han (14) and other ethnic Chinese populations (15, 16), and case–control cohorts with non-melanoma skin cancers (17). Although comparative DNA sequence analysis of the entire *MHC* genomic region between two different homozygous *HLA* haplotypes had previously showed the existence of *POALIN* within the *MHC* class II region (18), none had yet been investigated or characterized in human population studies.

In this study, we genotyped the *POALIN* by polymerase chain reaction (PCR) at five loci in the *MHC* class II genomic region and determined their allele and haplotype frequencies and percentage associations with the pre-typed alleles of the *HLA-DRB1* and *-DQB1* genes in 100 Japanese, 174 Australian Caucasians and 67 *HLA* reference cell lines obtained from different ethnic groups. The Japanese and Caucasians were genetically different on the basis of the inferred *HLA* class II gene and *POALIN* haplotype patterns and frequencies and the standard statistical *Fst* and analyses of molecular variance (*AMOVA*) differentiation computations. The *POALIN* patterns also differentiated among some members of the ancestral *HLA-DRB1* super-haplotypes, *DR1, DR8, DR51, DR52* and *DR53*.

Materials and methods

DNA samples and *HLA* class II genotypes

Reference DNA samples were obtained as previously described (19, 20) from 100 Japanese (Department of Legal Medicine, Shinshu University School of Medicine, Matsumoto, Nagano, Japan), 174 Australian–Caucasian (Department of Clinical Immunology and Biochemical Genetics, Royal Perth Hospital, Perth, Australia) from the seaside town of Busselton in Western Australia (http://www.busseltonhealthstudy.com/) and an *HLA* reference set of 67 B-lymphoblastoid cell lines (European Collection of Cell Cultures) of different ethnic origins that were genotyped and/or serotyped at least for *HLA* alleles at the *HLA-DRB1* and *-DQB1*. The Caucasian DNA samples were also genotyped for *DRB1, DRB3, DRB4* and *DRB5* class II gene loci by DNA sequencing. Ethics approval for the use of the human DNA samples in this study was obtained from the Tokai University Ethics Committee as ethics approval no. 07I-38.

Nomenclature of HLA alleles

The *HLA* alleles are reported and analysed here statistically as two-digit alleles as well as four-digit alleles. The first two digits of an allele such as *01* in *DRB1*0102* (Table S1, *Supporting Information*) represent the ancestral group or type of highly related alleles, which often corresponds to the serological antigen carried by an allotype. The third and fourth digits such as the *02* in *DRB1*0102* describe the subtype that has been assigned in the order of the determined DNA sequences and represent differences in one or more nucleotide substitutions that have changed the amino acid sequence of the encoded protein (21).

PCR assays for the detection of the absence and presence of *POALIN*s

Table 1 shows the PCR primer nucleotide sequences, amplicon product sizes, the annealing temperatures and cycle times used for the amplification of the five *POALIN*s. Each PCR assay was performed in 10 µl aliquots using 2 pmol of each primer (200 nmol/l), 1 ng of genomic DNA, 0.25 U of TaKaRa LA *Taq* polymerase, 0.08 µl of dNTP mixture (2.5 mM each) and 5 µl of 2 × GC reaction buffer 1 with 5 mM $MgCl_2$ purchased from TaKaRa, Shiga, Japan. The PCR was performed in eight strips of 0.2-ml thin-walled PCR tubes (QSP) using a GeneAmp PCR System 9700 Thermal cycler (Applied Biosystems Inc., Foster City, CA) programmed for an initial denaturation step at 96°C for 5 min and then 35 cycles with each cycle consisting of a denaturation (96°C for 30 s), and a single annealing and extension step using the temperatures (60–63°C) and times (3–5 min) for each PCR assay that is shown in Table 1. Small aliquots (2–5 µl) of the reaction products were stained with ethidium bromide and the sizes compared with molecular size markers by horizontal gel electrophoresis in 2% agarose using Tris-borate-EDTA running buffer.

Figure S1 (*Supporting Information*) shows an example of the electrophorograms of the *POALIN*-PCR products stained with ethidium bromide for all five PCR assays using 21 cell line DNA samples (lanes 22–42), molecular markers and reference DNA from the cell lines COX and PGF (Table S1, *Supporting Information*). The *POALIN*-PCR assays allowed the simultaneous detection of the presence and absence of the *Alu* insertions in heterozygous samples as seen in lane 32 for the *AluDQA2* and *AluORF10* PCR assays. The homozygous or heterozygous *Alu* insertion was seen as one or two amplified bands of predicted sizes, respectively, for all of the *Alu* PCRs except for the PCR genotyping of *AluDPB2*, which produced single bands for the homozygotes, but more than two bands for the heterozygotes. Control samples (without DNA template) were also run for most assays to ensure there was no amplification of contaminating DNA. Reference control DNA from the COX and PGF cell lines were used to verify the identified polymorphisms.

Assignment of *HLA-DRB1* alleles and *POALIN* to five *DRB1* supertypes

The 12 different Caucasian *HLA-DRB1* two-digit alleles that represent groups of highly related alleles were correlated with the absence or the presence of the *POALIN* and the

Table 1 Primer sequences, amplicon product sizes and PCR conditions for the DNA amplification of five MHC class II POALIN loci

Alu name	Laboratory code number	Primer name	Primer type	Primer sequence	Primer length	Fragment size (bp) Presence	Fragment size (bp) Absence	*Alu* subfamily type	PCR conditions (35 cycles) Denaturation and annealing Temperature (°C)	PCR conditions (35 cycles) Denaturation and annealing Time (min)
AluDPB2	13a	13aS	Sense	5′-AGACTAAGGAGTGTGGATTTCTC-3′	23	754	422	*AluYb8*	60	3
		13aAS	Antisense	3′-ACTTCTATCCTCCTCTTTCCTC-5′	22					
AluDQA2	12c	12cS	Sense	5′-CTGAAATCTTAATGTGGTTGG-3′	21	423	100	*AluYa5*	60	5
		12cAS	Antisense	3′-GAGTAGAATAAGGAGAATTGC-5′	21					
AluDQA1	10b	10Sb	Sense	5′-AACTTTACI[C/T]CATCTCTACCTCTC-3′	23	1082	769	*AluY*	60	5
		10ASb	Antisense	3′-TGTTTCTCATACTTGACTGTGG-5′	22					
AluDQA1	10c	10cFP1	Sense	5′-TGCCATGTAGCCTGGTCTA-3′	19	855	536	*AluY*	63	5
		10cRP1	Antisense	3′-CTGTCTTCATCTAGAGGTGT-5′	20					
AluDRB1	9a	9aS	Sense	5′-CACTAGTCAGTTCATCCTCTTGT-3′	23	763	474/424	*AluY*	60	5
		9aAS	Antisense	3′-TATGTCTGTTG[G/T]AGATCTTGTG-5′	22					
AluORF10	7c	Sab.10	Sense	5′-AGGGATGAATAGCTTCCTGT-3′	20	459	136	*AluYb8*	63	5
		ASb.10	Antisense	3′-GAATTGTCTTTGGATGGTGAG-5′	21					

PCR, polymerase chain reaction; MHC, major histocompatibility complex; POALIN, polymorphic *Alu* insertion.

HLA-DRB3, *-DRB4* and *-DRB5* alleles and arranged into the five *HLA-DRB1* haplotypes or supertypes *DR1, DR8, DR51, DR52* and *DR53*, as previously designated (21–23).

Statistical analyses

Gene and allele frequencies, heterozygosity, HWE and LD tests in a pairwise analysis of the association between the *HLA* class II gene two-digit alleles and the *Alu* alleles were performed using the GENEPOP software programmes (24) online at http://genepop.curtin.edu.au/, the De Finetti programme online at http://ihg.gsf.de/cgi-bin/hw/hwa1.pl (25) or the GENODIVE software (26) downloaded from the URL http://www.bentleydrummer.nl/software/software/GenoDive.html. The LD values D' and r^2 were calculated in a multiallele pairwise analysis of the association between the *HLA* class II gene alleles and the *Alu* alleles using the MIDAS software (27) downloaded from the URL http:// www.oege.org/software/midas/index.shtml. A 2 × 2 contingency two-sided test with 2 df and Fisher's exact test in the De Finetti programme detected the significant difference for *POALIN* frequencies between Japanese and Caucasians. Haplotype frequencies, pairwise LD and HWE of *POALIN*s were also calculated using the ARLEQUIN computer programme v3.1 (28) downloaded from the URL http://cmpg.unibe.ch/software/arlequin3. The haplotype frequencies were calculated for genotypic data with unknown gametic phase using expectation-maximization (EM) algorithm and parameter ARLEQUIN settings of 10–100 for the number of starting values for EM, 100–500 for the initial conditions for bootstraps, 100–5000 for the number of iterations and 10–1000 for the number of bootstrap replicates. The percentage association between an *Alu* insertion and *HLA* allele was calculated as the percentage of the total *HLA-DRB1* or *HLA-DQB1* allele frequency that was associated with the presence of the *Alu* insertion at an inferred *HLA* class II gene/*POALIN* haplotype using the haplotype frequency data generated by the ARLEQUIN software. The primary P-values of <0.05 obtained by Fisher's exact test or Chi square estimates were adjusted using the Bonferroni correction (29) for multiple testing when necessary by multiplying the P-value by the number of independent comparisons performed. Unpaired t-tests between the means were performed with GRAPHPAD software online at http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=SD. Inter-population methods included the use of hierarchical *AMOVA* (26) and *Fst* (30) to evaluate the amount of population genetic structure, using the algorithms provided by the GENEPOP, ARLEQUIN and GENODIVE programmes.

Results and discussion

Location of *POALIN*s within the *HLA* class II region and PCR primer design

Figure 1 shows the names and map positions, respectively, of the five *POALIN*s and *HLA* class II genes distributed

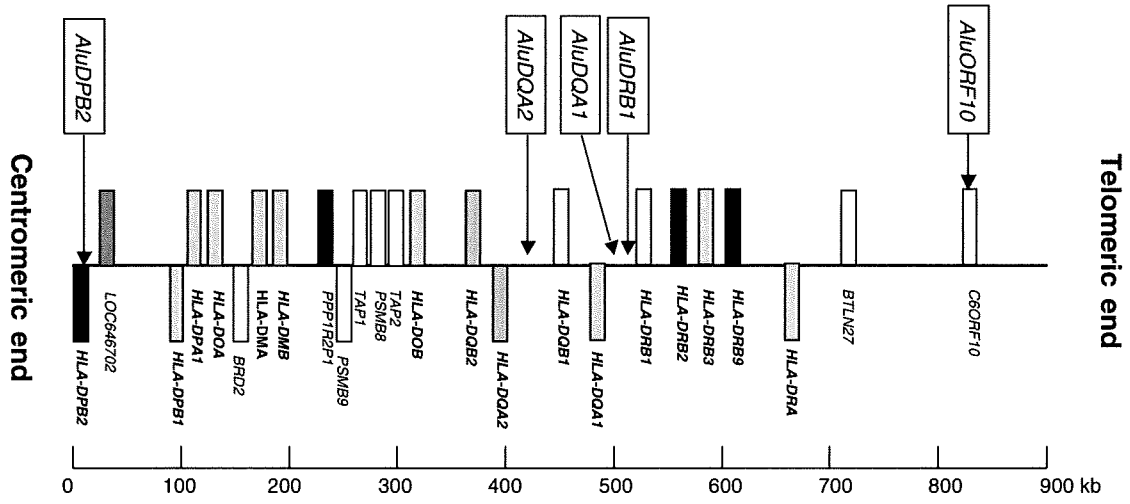


Figure 1 Map of the location of the five polymorphic *Alu* insertion (*POALIN*) within the major histocompatibility complex (*MHC*) class II region.

across 850 kb and located between the *HLA-DPB2* and *C6ORF10* genes within or bordering the *MHC* class II genomic region. The locations of the *POALIN*s for PCR primer design were determined from the accession numbers previously reported (18): *AluDPB2* in intron 2 of the *HLA-DPB2* gene (nt position 10685–10998 in AL645940); *AluDQA2*, 1.2 kb telomeric of the 5′-end of *HLA-DQA2* (nt position 16927–17233 in Al713890); *AluDQA1*, 11 kb telomeric of the 5′-end of *HLA-DQA1* (nt position 37834–38134 in AL662789); *AluDRB1*, 33 kb telomeric of the 5′-end of *HLA-DQA1* and 14 kb centromeric of 5′-end of *HLA-DRB1* (nt position 15466–15734 in AL662789); and *AluORF10*, in intron 8 of *C6ORF10* at the telomeric edge of the *HLA* class II genomic region and ~300 kb from *HLA-DRB1* (nt position 21377–21684 in AL662789).

The *POALIN*-PCR results in the cell line analysis (Table S1, *Supporting Information*) conformed to the *HLA* genomic sequence information available for the *POALIN* DNA structures in the *HLA* reference cell lines COX, SSTO, PGF, DBB and MANN (31). On the basis of sequence analysis of selected PCR products and available GenBank sequences (data not shown), the single nucleotide polymorphism (SNP) (*C*/*A*) within the 9aAS antisense primer of the *AluDRB1* PCR assay (Table 1) has the '*C*' nucleotide linked with the *Alu* insertion and the '*A*' nucleotide linked with the absence of the *Alu* insertion. The SNP (*C*/*A*) within the 9aAS antisense primer of the *AluDRB1* PCR assay (Table 1) was amplified successfully in the homozygous cell lines (SSTO, PGF, DBB and MANN) and heterozygous DNA samples where the '*C*' nucleotide was linked to the *Alu* insertion and the '*A*' nucleotide was linked to the absence of the *Alu* insertion.

HLA and *POALIN* extended haplotypes in homozygous and heterozygous reference cell lines

The five *Alu* PCR assays (Table 1) were tested using the DNA samples from the 67 reference cell lines shown in Table S1, *Supporting Information*. The 50 homozygous *HLA-DRB1* alleles permitted a precise analysis of the *HLA* and *Alu* allelic relationships and haplotypes, and a summary of the *Alu* insertion/*HLA* class II haplotype diversity results is shown in Table 2. All seven cell lines with the *HLA-A*03/-B*07/-DRB1*1501* haplotype, which is represented by the PGF reference cell line that was previously fully sequenced for the *HLA* genomic region (18), had the two *Alu* insertions, *AluDRB1* and *AluDQA1*, six cells had the homozygous or heterozygous *AluDPB2* insertion, five had the *AluORF10* insertion and none had the *AluDQA2* insertion. Three other cell lines with the haplotypes *A*3001/B*1302/DRB1*0701* (cell number 6), *A*3301/B*1402/DRB1*0102* (cell number 8) and *A*3301/B*1402/DRB1*0701* (cell number 12) had *Alu* heterozygous or homozygous insertions at four *POALIN* loci. There were only six different *HLA* haplotypes in the *DRB1* homozygous cell lines that had no *Alu* insertion at the five *POALIN* loci, designated here as the *POALIN* null haplotype. The reference cell lines with the *POALIN* null haplotypes are numbered 17 (BM16), 18 (EJ32B), 21 (WT9), 22 (KOSE), 26 (TAB089), 37 (L0081785) and 47 (QBL) in Table S1, *Supporting Information*. Interestingly, four of the six *POALIN* null haplotypes carried the *HLA-B*1801* allele or the combination of *HLA-B*1801* and *HLA-DRB1*0301* alleles.

POALIN allele and haplotype frequencies at five Alu loci

Table S2, *Supporting Information*, shows the *HLA-DRB1* and *HLA-DQB1* four-digit genotypes and POALIN two-digit

Table 2 A summary of the *Alu* insertion-*HLA* class II haplotype combinations in the *HLA-DRB1/HLA-DQB1* homozygous cell lines presented in Table S1, *Supporting Information*

Haplotype ID	*HLA* class II alleles and/or haplotypes	*Alu* insertions within the haplotype region (a), (b) or (c)	Number of *Alu* insertion loci	Number of homozygous cell lines
(a) All five Alu loci				
1	*DRB1*0301/DQB1*0501*	None	0	1 of 1
2	*DRB1*0301/DQB1*0201*	None	0	2 of 9
3	*DRB1*0803/DQB1*0103*	None	0	1 of 1
4	*DRB1*1201/DQB1*0301*	None	0	1 of 1
5	*DRB1*1401/DQB1*0503*	None	0	1 of 2
6	*DRB1*0301/DQB1*0201*	*AluORF10* or *AluDQA2*	1	2 of 9
7	*DRB1*09/DQB1*03*	*AluDQA1*	1	3 of 3
8	*DRB1*0401/DQB1*0301*	*AluDQA1*	1	3 of 4
9	*DRB1*0701/DQB1*0201*	*AluDQA1*	1	2 of 6
10	*DRB1*0301/DQB1*0201*	*AluDPB2/AluDQA2*	2	4 of 9
11	*DRB1*0101/DQB1*0501*	*AluDPB2/AluDRB1*	2	4 of 5
12	*DRB1*0401/DQB1*0301*	*AluDPB2/AluDQA1*	2	1 of 4
13	*DRB1*0701/DQB1*0201*	*AluDPB2/AluDQA2/AluDRB1/AluORF10*	4	2 of 3
14	*DRB1*1501/DQB1*0602*	*AluDPB2/AluDQA1/AluDRB1/AluORF10*	4	5 of 5
(b) AluDPB2 only				
15	*DPB1*0401*	None	0	4 of 13
16	*DPB1*0201/*0202*	None or heterozygous insertion	0	8 of 8
17	*DPB1*0301*	*AluDPB2*	1	2 of 2
18	*DPB1*0101*	*AluDPB2*	1	3 of 3
19	*DPB1*0401*	*AluDPB2*	1	9 of 13
20	*DPB1*0402*	*AluDPB2*	1	4 of 4
21	*DPB1*0501*	*AluDPB2*	1	4 of 4
(c) AluDPB2/AluDQA2 only				
22	*DQB1*0201*	None	0	5 of 13
23	*DQB1*0301*	None	0	3 of 5
24	*DQB1*0302*	None	0	2 of 4
25	*DQB1*05*	*AluDPB2*	1	5 of 6
26	*DQB1*0201*	*AluDPB2* or *AluDQA2*	1	3 of 13
27	*DQB1*0303*	*AluDPB2*	1	2 of 3
28	*DQB1*0201*	*AluDPB2/AluDQA2*	2	6 of 13
29	*DQB1*0501*	*AluDPB2/AluDQA2*	2	6 of 6

HLA, human leukocyte antigen.

genotypes in 100 Japanese and 174 Caucasians. Table 3 shows the allele and haplotype frequencies of the five Alu loci and population differences at $P < 0.05$. The most frequent *POALIN* allele in the Japanese (0.568) and the Caucasians (0.497) was the *AluDQA1* insertion and the least frequent *POALIN* in the Japanese (0.1) and Caucasians (0.21) was the *AluDQA2* insertion. The allele frequency differences between the Japanese and Caucasians were significant ($P < 0.05$) for the *AluDQA2*, *AluDQA1* and *AluORF10* insertions, but not for the *AluDRB1* and *AluDPB2* insertions. The interpopulation genetic differentiation of the Japanese and the Caucasians was also examined using the *F* statistical measures *Fst* and *AMOVA* of the *HLA-DRB1*, *HLA-DQB1* and *POALIN* genotypes and alleles, and a significant variation was detected between the Japanese and Caucasian populations for the *Phi'_st* values at four of the seven loci with a *P*-value of 0.001 and an overall 7.1% variation between the two populations (Table S3, *Supporting Information*). The pairwise *Fst* values ranged between −0.0014 and 0.1519 with an overall value of 0.0365 for all seven loci and moderate genetic differentiation in the range of 0.05–0.15.

Maximum-likelihood (ML) haplotype frequencies were inferred using the EM algorithm in ARLEQUIN for the *POALIN* 5-locus haplotypes (Table 3). Twenty-six *POALIN* haplotypes were inferred in total with 16 haplotypes in the Japanese and 23 haplotypes in the Caucasians. Thirteen of the haplotypes were present in both Japanese and Caucasians, three were unique to the Japanese at a combined frequency of 3.6% and ten were unique to the Caucasians at a combined frequency of 18.2%. The three most frequent haplotypes in the Japanese were the *AluDQA1/AluDPB2* insertions at 19.8%, the *Alu* null haplotype (no *Alu* insertions) at 17.6% and the single *AluDQA1* insertion at 12.6%. In Caucasians, the three most frequent haplotypes were the single *AluDQA1* insertion at 15.3%, the *Alu* null haplotype at 14% and the *AluDQA1/AluDPB2* insertions at 9.5%. Eleven of the 13 *Alu* haplotypes shared between the Japanese and

Table 3 Allele and haplotype frequencies of the five *Alu* loci and population differences (*P*)

	Japanese			Caucasian					
Marker	Frequency	SD	Number	Frequency	SD	Number	*P*	5-*POALIN* haplotypeID#	*AluDRB1/AluDQA1* haplotypeID#
AluDPB2	0.500	0.039	100	0.454	0.026	174	0.299		
AluDQA2	0.010	0.007	100	0.210	0.021	174	5.74E−11		
AluDQA1	0.568	0.043	96	0.497	0.024	174	0.031		
AluDRB1	0.225	0.029	100	0.256	0.024	174	0.299		
AluORF10	0.145	0.024	100	0.236	0.024	174	0.01104		
Haplotypes									
1 1 1 1 1	0.176	0.034	200	0.140	0.025	348	<0.0001	5hap1	2hap1
1 1 1 1 2	0.091	0.029	200	0.091	0.021	348	0.9396	5hap2	2hap1
1 1 1 2 2	0.061	0.020	200	0.015	0.009	348	<0.0001	5hap3	2hap1
1 1 2 1 1	0.126	0.031	200	0.153	0.025	348	<0.0001	5hap4	2hap2
1 1 2 1 2	0.198	0.034	200	0.095	0.021	348	<0.0001	5hap5	2hap2
1 1 2 2 1	0.051	0.020	200	0.023	0.012	348	<0.0001	5hap6	2hap2
1 1 2 2 2	0.028	0.018	200	0.036	0.015	348	<0.0001	5hap7	2hap2
1 2 1 1 1	0.010	0.009	200	0.080	0.021	348	<0.0001	5hap8	2hap1
1 1 ? 1 2	0.025	0.014	200					?	
1 2 1 1 2	0.010	0.007	200	0.060	0.017	348	<0.0001	5hap9	2hap1
2 1 1 1 1	0.021	0.017	200	0.040	0.016	348	<0.0001	5hap10	2hap3
2 1 1 1 2	0.052	0.021	200	0.053	0.017	348	0.6246	5hap11	2hap3
2 1 1 2 1	0.006	0.006	200					5hap12	2hap3
2 1 2 1 1	0.106	0.026	200	0.021	0.012	348	<0.0001	5hap13	2hap4
2 1 2 1 2	0.036	0.019	200	0.011	0.010	348	<0.0001	5hap14	2hap4
2 1 ? 1 1	0.005	0.006	200					?	
1 2 1 2 1				0.004	0.005	348		5hap15	2hap1
1 2 2 1 1				0.013	0.009	348		5hap16	2hap2
1 2 2 2 1				0.034	0.012	348		5hap17	2hap2
2 1 1 2 2				0.004	0.005	348		5hap18	2hap3
2 1 2 2 1				0.033	0.014	348		5hap19	2hap4
2 1 2 2 2				0.075	0.019	348		5hap20	2hap4
2 2 1 1 2				0.005	0.006	348		5hap21	2hap3
2 2 1 2 1				0.005	0.005	348		5hap22	2hap3
2 2 1 2 2				0.007	0.007	348		5hap23	2hap3
2 2 2 1 2				0.003	0.003	348		5hap24	2hap4

POALIN, polymorphic *Alu* insertion.

The differences (*P*) between the Japanese and Caucasians were analysed using the unpaired *t*-test performed with GRAPHPAD Software using the mean, SD and N data entry format online at http://www.graphpad.com/quickcalcs/ttest1.cfm?Format=SD. The haplotypes are the presence (2) or absence (1) of an *Alu* insertion at the loci *AluDRB1, AluDQA2, AluDQA1, AluORF10, AluDPB2* from the left to the right-hand side of the haplotype configuration column. The symbol '?' indicates missing genotype at the *AluDQA1* locus in Japanese. Number of bootstraps for generating SDs was 1000 with initial conditions for bootstrap of 10. The ARLEQUIN analysis was performed using the default settings and allowed level of missing data at 0.05.

the Caucasians were significantly ($P < 0.0001$) different in frequency between the two populations. Only haplotypes 2 (*AluDPB2* insertion only) and 12 (*AluDRB1* and *AluDPB2* insertions) were not significantly ($P > 0.05$) different in frequency between the two populations.

HWE tests

The genotype distributions were consistent with HWE at all loci except for *AluDQA1* ($P < 0.01$) in the Japanese and Caucasians (Table 4) possibly due to PCR genotyping errors because of nucleotide sequence mutations and indels at the PCR primers sites. In addition, the PCR amplification of the *AluDQA1* locus failed for 4 of the 100 Japanese DNA samples, suggesting that there may have been sequence mutations (SNPs or indels) at the primer sites of these samples that prevented efficient PCR. There was a significant ($P < 0.001$) heterozygote deficit for the *AluDQA1* genotypes in the Japanese and a significant ($P < 0.01$) level of heterozygote excess in the Caucasians even after Bonferroni's correction for multiple testing.

The statistical significance ($P = 0.047$) of the *AluDPB2* deviation from HWE and the heterozygote deficit ($P = 0.0315$) in Japanese was lost ($P > 0.05$) after Bonferroni's correction for multiple testing. The inbreeding coefficient *Fis* was generally low for each of the *Alu* polymorphisms except for *AluDQA1* in Japanese and Caucasians, which reflected the significant level of variation between the expected and the observed number of heterozygotes in each population.

Table 4 Hardy–Weinberg tests

Locus	*P*	
	Japanese	Caucasian
AluDPB2	0.047	0.879
AluDQA2	1.000	0.358
HLA-DQB1	0.934	0.869
AluDQA1	2.181E–06	0.010
AluDRB1	0.774	0.842
HLA-DRB1	0.400	0.927
AluORF10	0.686	0.204

LD and two-locus paired allelic associations

The exact *P*-value for the genotypic LD of two-locus genotypic counts for all possible pairs of loci in each population was estimated by a Markov chain method used in the GENEPOP computer programme. The LD between different pairs of loci was significant ($P < 0.05$) for *HLA-DRB1* and *HLA-DQB1* and most paired relationships between the *HLA-DRB1* or *HLA-DQB1* loci and the *POALIN* loci with a few exceptions, which were between *HLA-DRB1* or *HLA-DQB1* and *AluDBP2* in Caucasians and Japanese, and between *HLA-DQB1* and *AluDQA2* in Japanese.

A more detailed LD analysis was implemented using the computer software MIDAS to compute the LD values D' and r^2 and the significance levels based on the Yates' corrected and uncorrected Chi square statistics. The *AluDQA1* was in Hardy–Weinberg disequilibrium ($P < 0.01$) in Caucasians and Japanese and, therefore, not included in this analysis. Overall, 518 *POALIN/POALIN*, *POALIN/HLA-DRB1* and POALIN/*DQB1* multiallelic pairs were analysed (285 Caucasian and 233 Japanese pairs) with 27 Yates' corrected and 29 uncorrected significant pairs ($P < 0.05$) for the D' values in Caucasians and 12 Yates corrected and 19 uncorrected significant pairs ($P < 0.05$) for the D' values in Japanese (Table 5). Four pairs of *POALIN* loci were in significant ($P < 0.05$) LD in Caucasians, but no *POALIN* loci were in LD in Japanese.

Four of ten possible *POALIN* pairs were in significant LD in Caucasians, but not the Japanese, presumably because of their different evolutionary histories and the relatively different times that they were inserted into the founder individuals. The *POALIN AluDRB1* was in significant ($P < 0.05$), but relatively weak to moderate LD ($D'$of 0.29–0.67) with *AluDPB2*, *AluDQA2* and *AluORF10*. The *POALIN*s, *AluDPB2* and *AluORF10*, which are located at opposite ends of the *HLA* class II region (Figure 1), were also in significant LD ($P < 0.05$) but with a low D' measure of 0.31. The absence of strong LD between many *POALIN* loci paired combination suggests that they were inserted into the genomes of the ancestral individuals with unrelated *HLA-DRA1* and *HLA-DRB1* alleles.

The calculation of the percentage *HLA-DRB1* or *-DQB1* allele frequency that is associated with a *POALIN* provides a better quantitative measure of the linkage between an *Alu* insertion and an *HLA* class I or class II gene allele. The frequency and the percentage association of the *POALIN*s with the *HLA-DRB1* and *HLA-DQB1* alleles are shown in Tables S4 and S5, *Supporting Information*, respectively. The strongest *POALIN-DRB1* associations (>70% of *HLA-DRB1* allelic frequencies) were between *AluDQA2* and *DRB1*03* or *DRB1*10* in Japanese and Caucasians; *AluDQA1* and *DRB1 *04, *09, *15* and **16* in Japanese and Caucasians; *AluDQA1* and *DRB1 *07* and *DRB1 *08* in Caucasians; *AluDRB1* and *DRB1 *01, *15, *16* in Japanese and Caucasians; *AluORF10* and *DRB1 *15* in Caucasians; and *AluORF10* and *DRB1 *16* in Japanese (Table 5 and Table S4, *Supporting Information*). The other *POALIN* and *DRB1* percentage associations ranged from 0% to 70% (Table S4, *Supporting Information*).

Five-locus *POALIN* on *HLA-DRB1-DQB1* haplotypes (four-digit alleles)

The largest number of haplotypes (145 haplotypes) and differences between Japanese and Caucasians (72 *vs* 85 haplotypes, respectively) was obtained in the seven-locus ML haplotype analysis for the five-locus *POALIN* on the *HLA-DRB1-DQB1* haplotypes (four-digit alleles) using the EM algorithm in ARLEQUIN. This number of inferred haplotypes in each population represented 72% of the Japanese population and almost 50% of the Caucasian population. The three most common seven-locus haplotypes in the Japanese were *DRB1*0901/DQB1*0303/AluDQA1/AluDPB2* (9.2%), *DRB1*1502/DQB1*0603/AluDRB1/AluDQA1* (7.3%) and *DRB1*0101/DQB1*0501/AluDRB1/AluDPB2* (6%), and in Caucasians they were *DRB1*1501/DQB1*0602/AluDQA1/AluDRB1/AluORF10/AluDPB2* (6.7%), *DRB1*0101/DQB1*0501/AluDRB1/AluDPB2* (6.1%) and *DRB1*0301/DQB1*0201/AluDQA2* (5.3%).

A relatively small number of inferred haplotypes were found to have no *Alu* insertions, designated here as the *POALIN* null haplotype. In the reference cell lines, four of the six *POALIN* null haplotypes carried the *HLA-B*18* allele or the combination of the *HLA-B*1801* and *HLA-DRB1*0301* alleles as a haplotype. The other two *POALIN* null haplotypes in cell lines carried either the *HLA*-B**3503*, *-DRB1*1302/1401* alleles or the *HLA*-B**4601*, *-DRB1*0803* alleles. The inferred *POALIN* null haplotype was found in 17.6% of Japanese and 14% of Caucasians on different *DQB1/DRB1* haplotypes. The Japanese and Caucasians shared the *POALIN* null haplotype with four different *HLA-DQB1* and *HLA-DRB1* allelic combinations, *DQB1*03/DRB1*11*, *DQB1*03/DRB1*12*, *DQB1*06/DRB1*13* and *DQB1*05/DRB1*13*. The *POALIN* null haplotype was also found in a small proportion of *HLA-DRB1*0301* and *-DRB1*0803* allelic populations. The frequency of the five-locus *POALIN* null haplotype in the *MHC* class II region of Japanese and Caucasians (present study) is

Table 5 LD estimations among *Alu* insertion, *HLA-DRB1* and *HLA-DQB1* loci for different allele (two-digit) paired combinations in Caucasians and Japanese

Haplotype (Loc1AlleleName_Loc2AlleleName)	LD (D')	LD (r^2)	χ^2 (P-value) (1 df)	Yates χ^2 corrected (P-value) (1 df)	Percentage association
Japanese					
*AluDPB2*2_HLA-DQB1*06*	−0.43	0.07	<0.01	<0.05	28
*AluDPB*2_HLA-DRB1*15*	−0.48	0.05	<0.05	ns	27
*AluDQA2*2_HLA-DQB1*05*	1.00	0.06	<0.05	ns	7
*AluDQA2*2_HLA-DRB1*10*	1.00	0.50	<1E−10	ns	100
*AluDQA2*2_HLA-DRB1*13*	0.47	0.07	<0.01	ns	0
*HLA-DQB1*03_AluDRB1*2*	−0.88	0.16	<0.0001	<0.001	1
*HLA-DQB1*04_AluDRB1*2*	−1.00	0.06	<0.05	<0.05	3
*HLA-DQB1*05_AluDRB1*2*	0.28	0.05	<0.05	ns	40
*HLA-DQB1*06_AluDRB1*2*	0.47	0.17	<0.0001	<0.0001	57
*HLA-DQB1*03_AluORF10*2*	−1.00	0.12	<0.001	<0.01	2
*HLA-DQB1*04_AluORF10*2*	0.78	0.53	<1E−10	<1E−10	73
*HLA-DQB1*06_AluORF10*2*	−1.00	0.06	<0.05	<0.05	0
*AluDRB1*2_HLA-DRB1*01*	0.81	0.20	<1E−5	<1E−4	75
*AluDRB1*2_HLA-DRB1*04*	−1.00	0.09	<0.01	<0.01	2
*AluDRB1*2_HLA-DRB1*09*	−1.00	0.07	<0.01	<0.05	0
*AluDRB1*2_HLA-DRB1*15*	0.76	0.42	<1E−10	<1E−10	77
*HLA-DRB1*04_AluORF10*2*	0.95	0.47	<1E−5	<1E−5	53
*HLA-DRB1*09_AluORF10*2*	−1.00	0.04	<0.05	ns	0
*HLA-DRB1*16_AluORF10*2*	1.00	0.06	<0.05	ns	100
Caucasian					
*AluDPB2*2_HLA-DQB1*05*	0.38	0.03	<0.05	<0.05	58
*AluDPB2*2_AluDRB1*2*	0.29	0.03	<0.05	<0.05	0
*AluDPB2*2_HLA-DRB1*01*	0.44	0.03	<0.05	<0.05	59
*AluDPB2*2_HLA-DRB1*04*	−0.36	0.03	<0.05	ns	29
*AluDPB2*2_AluORF10*2*	0.31	0.04	<0.05	<0.05	0
*AluDQA2*2_HLA-DQB1*02*	0.41	0.15	<0.000001	<0.000001	11
*AluDQA2*2_HLA-DQB1*03*	−0.91	0.13	<0.000001	<0.000001	39
*AluDQA2*2_AluDRB1*2*	−0.67	0.04	<0.01	<0.05	0
*AluDQA2*2_HLA-DRB1*03*	0.67	0.25	<1E−10	<1E−10	76
*AluDQA2*2_HLA-DRB1*04*	−0.70	0.03	<0.05	<0.05	2
*AluDQA2*2_HLA-DRB1*11*	−1.00	0.02	<0.05	ns	0
*AluDQA2*2_HLA-DRB1*15*	−1.00	0.03	<0.05	<0.05	0
*HLA-DQB1*02_AluDRB1*2*	−1.00	0.10	<0.001	<0.001	1
*HLA-DQB1*03_AluDRB1*2*	−0.84	0.14	<1E−6	<1E−5	2
*HLA-DQB1*05_AluDRB1*2*	0.78	0.33	<1E−10	<1E−10	82
*HLA-DQB1*06_AluDRB1*2*	0.29	0.07	<0.001	<0.01	51
*HLA-DQB1*03_AluORF10*2*	−0.65	0.08	<0.001	<0.001	7
*AluDRB1*2_HLA-DRB1*01*	0.96	0.37	<1E−10	<1E−10	100
*AluDRB1*2_HLA-DRB1*03*	−1.00	0.05	<0.01	<0.01	0
*AluDRB1*2_HLA-DRB1*04*	−0.64	0.03	<0.05	<0.05	5
*AluDRB1*2_HLA-DRB1*07*	−1.00	0.06	<0.01	<0.01	0
*AluDRB1*2_HLA-DRB1*11*	−1.00	0.03	<0.05	<0.05	0
*AluDRB1*2_HLA-DRB1*13*	−1.00	0.05	<0.01	<0.05	0
*AluDRB1*2_HLA-DRB1*15*	1.00	0.38	<1E−15	<1E−10	100
*AluDRB1*2_AluORF10*2*	0.35	0.11	<1E−5	<1E−5	0
*HLA-DRB1*03_AluORF10*2*	−1.00	0.04	<0.01	<0.05	2
*HLA-DRB1*04_AluORF10*2*	−0.79	0.04	<0.01	<0.05	4
*HLA-DRB1*07_AluORF10*2*	0.51	0.14	<1E−6	<1E−5	62
*HLA-DRB1*11_AluORF10*2*	−1.00	0.03	<0.05	ns	0
*HLA-DRB1*15_AluORF10*2*	0.85	0.31	<1E−10	<1E−10	89

LD, linkage disequilibrium; HLA, human leukocyte antigen.

The % association is the percentage of the total *HLA-DRB1* or *HLA-DQB1* allele frequency that is associated with an Alu insertion at an inferred *HLA* class II gene/POALIN haplotype. ns, not significant. *Aluname*2* is the insertion allele. *AluDQA1* was in HWD in Caucasians and Japanese and *AluDPB2* was in HWD in Japanese and, therefore, were not included in this table.

similar to the frequency of the five-locus *POALIN* haplotype in the *MHC* class I region of Japanese and Caucasians (2).

We did not find a five-locus-*Alu* insertion haplotype in the Japanese, Caucasian or in the DNA samples of the reference cell lines that we investigated. The Caucasians and Japanese had *Alu* insertions ranging from none and up to four loci per individual. We found that the four-*Alu*-insertion haplotype was mostly linked to the relatively frequent (10.3%) Caucasian ancestral haplotype *7.1 AH*, which has the *HLA-DRB1*1501* and *-DQB1*0602* alleles that confer protection against insulin-dependent diabetes mellitus (32, 33). In comparison to the Caucasian *7.1 AH*, the *HLA-DRB1*15* extended haplotype in Japanese only had two or three *Alu* insertions and none had the *AluORF10* insertion. The relatively common *HLA* ancestral haplotype *8.1 AH* with *HLA-DRB1*0301* and *-DQB1*0201* (32, 33) had homozygous *Alu* insertions at the *AluDPB2* locus and heterozygous *Alu* insertions at *AluDQA2* and *AluORF10*. One of the *8.1 AH*, represented by the cell line named COX, however, only had a homozygous *Alu* insertion at *AluDPB2* (18). This result shows that even highly conserved extended haplotypes such as the *8.1 AH* are not identical for *Alu* insertions at all loci.

HLA class II *POALINs* and the five *DR* supertypes

Table 6 shows the *DR* supertype assignment of 12 different Caucasian *HLA-DRB1* two-digit alleles that represent groups of highly related alleles correlated to four *POALIN* types. We correlated the genotyped *HLA-DRB3, -DRB4 and -DRB5* genes with the Caucasian *HLA-DRB1* two-digit alleles and arranged them into the five *HLA-DR* supertypes *DR1, DR8, DR51, DR52* and *DR53*, as previously described (21–23). In Caucasians, there were strong correlations between *DR1* and *AluDRB1*, *DR8* and *AluDQA1*, *DR51* and *AluDQA1*, *AluDRB1* and *AluORF10*, and between *DR53* and *AluDQA1*. In the case of the *DR52* supertype, the *HLA-DRB1*11*, *-DRB1*12, -DRB1*13* and *-DRB1*14* alleles had no or few *POALIN* associations, whereas the *HLA-DRB1*03* allele had a strong association (76%) with *AluDQA2*. In addition, *HLA-DRB1*08*, which belongs to the *DR8* supertype, appears to have had no associated *POALIN*s in the Japanese, but was associated moderately to strongly with *AluDQA2* and *AluDQA1*, respectively, in Caucasians.

Evolution of the class II *POALIN*s and *HLA-DQB1, -DQA1* and *-DRB1* haplotypes

From the age estimates of *Alu* subfamilies, the oldest of the young *Alu* subfamily members is *Alu*Y at 30 MYA, whereas the youngest is *Alu*Yb8 at about 3.3 MYA (11). Almost 10% of *Alu*Yb8 are identical in sequence and therefore of recent origin (10). The *POALIN*s in our study are all members of the young *Alu* subfamily, with *AluDQA1* and *AluDRB1* belonging to the *Alu*Y subgroup and *AluDQA2*, *AluDPB2* and *AluORF10* belonging to the youngest *Alu*Ya5 or *Alu*Yb8 subgroup (Table 1). *AluDQA1* appears to be the oldest of the five *POALIN*s on the basis of having the highest *POALIN* frequency in Japanese and Caucasians (Table 3) and its association with most of the *DRB1* supertypes (Table 6). The other *Alu*Y subfamily member, *AluDRB1*, has half the frequency (0.22–0.26) of *AluDQA1* (0.50–0.57) and is strongly haplotypic for *HLA-DRB1*01*, and *-DRB1*15/*16* and the supertypes *DR1* and *DR51* (Table 6), respectively, in both the Japanese and Caucasians. These differences suggest that the *AluDRB1* insertion probably originated in an ancestral *HLA-DRB1* allele as a progenitor of the *DR1* and *DR51* supertypes. *AluDQA2* is a member of the *Alu*Ya5 subgroup that has an estimated origin of 2–3 MYA (11). The frequency of this *POALIN* is 0.21 in Caucasians and 0.01 in Japanese (Table 3), and it appears to be emerging selectively in Caucasians because of its strong association or linkage with the *HLA-DRB1*03* allele, which is absent in Japanese (Table S4, *Supporting Information*). The two *POALINs*, *AluDPB2* and *AluORF10*, which are members of the youngest *Alu* subgroup *Alu*Yb8, are located at either end of the extremities of the *HLA* class II region (Figure 1) and are in moderate ($P < 0.05$) to high ($P < 1E-10$) LD with some of the *HLA-DRB1* or *HLA-DQB1* alleles (Table 5). *AluDPB2* has a frequency of 0.45 and 0.5 in the Japanese and Caucasians, respectively, with low- to high-level percentage associations with many different *HLA-DRB1* alleles and haplotypes (Table S4, *Supporting Information*). In comparison, the *AluORF10* insertion frequency (0.15–0.24) is relatively low in the Japanese and Caucasians, and its association with *HLA-DRB1* alleles is different between the two populations (Table S4, *Supporting Information*). In Japanese, *AluORF10* was associated moderately with *HLA-DRB1*04*, strongly with *-DRB1*16* and not at all with *-DRB1*07* or *DRB1*15*, whereas in Caucasians *AluORF10* was associated moderately with *HLA-DRB1*07*, strongly with *-DRB1*15* and not at all with *-DRB1*04* or *-DRB1*15*. Therefore, the *AluORF10* insertion within intron 8 of the *C6ORF10* gene at the telomeric edge of the *HLA* class II genomic region can be differentiated between the Japanese and Caucasians when it is associated with *HLA-DRB1* alleles, indicating that this difference between the populations in the *C6ORF10* gene (NCBI geneID: 10665), where the *AluORF10* is inserted, deserves further attention, particularly as the genetics and function of *C6ORF10* is not yet known and little studied.

Recently, 20 human haplotypes of various combinations of the *HLA-DQB1, -DQA1* and *-DRB1* coding regions and intergenic regions as well as orthologous chimpanzee and gorilla ~100-kbp genomic regions were sequenced to determine their divergence, phylogeny and evidence for recent recombinations (34). The exceptionally high divergences between some pairs of haplotypes (up to 9.3%) suggested a long history of independent haplotype evolution, at least 40 Myr for the most dissimilar haplotypes. The evolution

Table 6 *POALIN* association with *HLA-DR* supertypes in Caucasians

			Percentage association of *HLA-DRB1* allele frequency			
DR supertypes	*DRB* gene and *DRB1* allele	*DRB* frequency	*AluDQA2*	*AluDQA1*	*AluDRB1*	*AluORF10*
DR1	*DRB1*01*	0.121	16	3	100	12
	*DRB1*10*	0.006	98	0	50	0
DR8	*DRB1*08*	0.020	54	100	0	0
DR51	*DRB1*15*	0.115	0	100	100	89
	*DRB1*16*	0.014	20	100	100	0
	DRB5	0.129				
DR52	*DRB1*03*	0.126	76	0	0	2
	*DRB1*11*	0.083	0	3	0	0
	*DRB1*13*	0.118	0	0	0	0
	*DRB1*14*	0.017	8	0	0	0
	DRB3	0.345				
DR53	*DRB1*04*	0.187	2	92	5	4
	*DRB1*07*	0.144	17	100	0	62
	*DRB1*09*	0.023	29	100	0	25
	DRB4	0.354				

POALIN, polymorphic *Alu* insertion; HLA, human leukocyte antigen.
The percentage association between an *Alu* insertion and *HLA* allele was calculated as the percentage of the total *HLA-DRB1* allele frequency that is associated with an Alu insertion at an inferred *HLA* class II gene/*POALIN* haplotype using the haplotype frequency data generated by the ARLEQUIN software.

of independent deeply diverged haplotypes over tens of millions of years apparently reflects either infrequent recombination or selection against recombinant haplotypes. Raymond et al. (34) classified the 20 human haplotypes into five groups A–E based on genomic similarity/diversity and phylogenetic analysis of selected genomic regions and the haplotypic combinations of the *HLA-DQB1, -DQA1* and *-DRB1* alleles. The deeply divergent C-group haplotype (*DQB1*030101/DQA1*0505/DRB1*1101*) and E-group haplotype (*DQB1*0602/DQA1*0102/DRB1*1503*) were found in many world-wide populations with frequencies ranging from a few to more than 20%, whereas a reciprocal recombinant with the *DQB1*0602/DQA1*0101/DRB1*1101* alleles was found only in 9% of South Africans and at a few per cent in African Americans (35).

In order to compare the five haplotype groups of Raymond et al. (34) with potential *Alu* haplotype pairs within these haplotypes, we examined the 20 haplotype genomic sequences (AY663393–AY663415) for the presence and absence of *AluDRB1* and *AluDQA1* (Table 7) using the PCR paired primer sequences (Table 1) in a BLAST search of GenBank. The other three polymorphic *POALIN* loci were outside the genomic sequences of these haplotypes and therefore were not examined. Table 7 shows the absence and presence of the *AluDRB1* and *AluDQA1* insertions in each of the haplotype groups and their association with the various combinations of the *HLA-DQB1, -DQA1* and *-DRB1* alleles and the haplotype groups A–E derived by Raymond et al. (34). Essentially, the haplotype groups A–C had no *Alu* insertions (*Alu* haplotype 2hap1), haplotype group D had the *AluDQA1* insertion but not the *AluDRB1* (*Alu* haplotype 2hap2) and haplotype group E had both the *AluDQA1* and *AluDRB1* insertions (*Alu* haplotype 2hap4). The *Alu* haplotype 2hap3, which contained the *AluDRB1* but not the *AluDQA1* insertions, was associated with the recent recombination haplotypes representing a mixture of the haplotype groups A/E or C/D. Interestingly, the gorilla and chimpanzee haplotype sequences carried the *Alu* haplotype 2hap2 (*AluDQA1* insertion but not the *AluDRB1* insertion). In this regard, the *AluDQA1* insertion in the gorilla (AY663402) and chimpanzee (AY663401) sequences confirms that it is at least 6–9 MYA, and therefore likely to be the oldest of the five *POALIN* markers investigated in this study. In addition, Raymond et al. (34) had difficulty in classifying their NA03715_2 haplotype genomic sequence *DQB1*0601/DQA1*0103/DRB1*1502* into a haplotype group on the basis of phylogenetic analysis of selected genomic regions. We found that this haplotype carried the *Alu* 2hap4 and that the *HLA-DQB1/-DQA1/-DRB1* alleles were representative of the group E haplotype. Based on the phylogenetic analysis of the haplotypes by Raymond et al. (34) and our *Alu* analysis (Table 7), it appears that many of the haplotypes without the *Alu* insertions are probably much older than those with one or both of the *AluDRB1* and *AluDQA1* insertions and that the haplotype group D (*AluDQA1* insertion but not the *AluDRB1* insertion) stems from the haplotype groups B and C, which have no *Alu* insertions. Thus, the advantage of *POALIN*s over microsatellites and SNP as historical lineage and/or evolutionary markers is that their ancestral state is known (absence of insertion) and the inherited insertion state is identical by descent. The *Alu* insertions are independent, stable events with different timelines for their insertions that

Table 7 The *AluDRB1* and *AluDQA1* haplotypes and their association with various combinations of the *HLA-DQB1, -DQA1* and *-DRB1* alleles and the haplotype groups A–E derived by Raymond et al.(34)

GenBank accession number	*AluDRB1* allele 9a primer set	*AluDQA1* allele 10b and 10c primer set	*AluDRB1/AluDQA1* Haplotype ID#	*AluDRB1/AluDQA1* Haplotype	*HLA-DQB1**	*HLA-DQA1**	*HLA-DRB1**	NA haplotype # or cell line [s] ID	Haplotype group
AY663415	Absent	Absent	2hap1	11	*DQB1*0603*	*DQA1*0103*	*DRB1*130101*	NA03715_1	A
AY663413	Absent	Absent	2hap1	11	*DQB1*0609*	*DQA1*010201*	*DRB1*130201*	NA14663_2	A
AY663408	Absent	Absent	2hap1	11	*DQB1*050301*	*DQA1*010401*	*DRB1*140501*	NA04535_2	A
AY663405	Absent	Absent	2hap1	11	*DQB1*050301*	*DQA1*010401*	*DRB1*1410*	NA10540_1	A
AY663396	Absent	Absent	2hap1	11	*DQB1*050101*	*DQA1*0105*	*DRB1*120101*	NA14660_1	A
AY663407	Absent	Absent	2hap1	11	*DQB1*020101*	*DQA1*050101*	*DRB1*030101*	NA10923_1	B
AY663399	Absent	Absent	2hap1	11	*DQB1*020101*	*DQA1*05010*	*DRB1*030101*	NA01018_2	B
AY663397	Absent	Absent	2hap1	11	*DQB1*030101*	*DQA1*0505*	*DRB1*110102*	NA01960_2	C
AY663394	Absent	Absent	2hap1	11	*DQB1*030101*	*DQA1*0505*	*DRB1*110401*	NA14661_1	C
AY663412	Absent	Absent	2hap1	11	*DQB1*030101*	*DQA1*0505*	*DRB1*110101*	NA00576-1	C
AL662842	Absent	Absent	2hap1	11	*DQB1*0201*		*DRB1*0301*	COX [s]	
Z84489	Absent	Absent	2hap1	11		*DQA1*05011*	*DRB1*03011*	RPCI-1 [s]	
AY663401	Absent	Insertion	2hap2	12				NS03646	Chimpanzee
AY663402	Absent	Insertion	2hap2	12				NG05251	Gorilla
AY663393	Absent	Insertion	2hap2	12	*DQB1*030101*	*DQA1*0303*		NA14661_2	C/D
AY663410	Absent	Insertion	2hap2	12	*DQB1*030302*	*DQA1*0302*	*DRB1*090102*	NA00576_2	D
AY663404	Absent	Insertion	2hap2	12	*DQB1*030302*	*DQA1*0302*		NA10540_2	D
AY663398	Absent	Insertion	2hap2	12	*DQB1*030201*	*DQA1*03010*		NA01960_1	D
BX248406	Absent	Insertion	2hap2	12	*DQB1*0305*		*DRB1*0403*	SSTO [s]	
AL137064	Absent	Insertion	2hap2	12			*DRB1*04011*	RPCI-5 [s]	
CR753309	Absent	Insertion	2hap2	12	*DQB1*030302*		*DRB1*0701*	DBB [s]	
CR753835	Absent	Insertion	2hap2	12	*DQB1*0202*		*DRB1*07011*	MANN [s]	
AY663400	Insertion	Absent	2hap3	21	*DQB1*050101*	*DQA1*010102*	*DRB1*010201*	NA01018_1	A/E
AY663409	Insertion	Absent	2hap3	21	*DQB1*030101*	*DQA1*0505*		NA04535_1	C/E
AY663414	Insertion	Insertion	2hap4	22	*DQB1*060101*	*DQA1*0103*	*DRB1*150201*	NA03715_2	A?
AY663411	Insertion	Insertion	2hap4	22	*DQB1*0602*	*DQA1*010201*	*DRB1*1503*	NA14663_1	E
AY663406	Insertion	Insertion	2hap4	22	*DQB1*0602*	*DQA1*010201*	*DRB1*150101*	NA10923_2	E
AY663395	Insertion	Insertion	2hap4	22	*DQB1*0602*	*DQA1*010201*	*DRB1*1503*	NA14660_2	E
AL662789	Insertion	Insertion	2hap4	22	*DQB1*0602*		*DRB1*150101*	PGF [s]	

HLA, human leukocyte antigen.

NA haplotype# and haplotype group are taken from Raymond et al. (34). Cell line [s] ID are the cell-line names (Table S1, *Supporting Information*) used at the Sanger Institute to sequence the HLA haplotypes by Horton et al. (31).

occur only once at each locus as there is no known mechanism for their excision from a specific locus (36).

Conclusion

The allele frequencies at three of the *MHC* class II *POALIN* loci were significantly different between the two populations and there were significant differences in the frequency of nearly all the *POALIN* haplotypes between the populations. The seven-locus haplotype of the five *POALIN*s on *DRB1/DQB1* either at the *DRB1/DQB1* two-digit or four-digit allelic level of analysis also showed considerable differences at the structural and frequency levels. A number of the common haplotypes shared between Japanese and Caucasians, such as *DRB1*1501/DQB1*0602*, *DRB1*09/DQB1*03/AluDRB1/AluDPB2* or *DRB1*0101/DQB1*0505/AluDRB1/AluDPB2*, although significantly different in frequency, may have been already established as ancestral haplotypes before the separation of Japanese and Caucasians from their common ancestor. On the other hand, the private haplotypes that are unique to either the Japanese or the Caucasians were probably formed after the separation of the two populations from their common ancestor. In this regard, there were 12 haplotypes shared between the Japanese and Caucasians and 59 private haplotypes for the Japanese and 74 for the Caucasians. It also seems that the *HLA-DRB1*08* allele represented by the supertype *DR8* has variable *AluDQA1* insertions in Japanese, but a complete number of *AluDQA1* insertions in Caucasians. These differences further highlight the potential application of the *MHC* class II *POALIN* markers for studying population differences and stratification (37). The *MHC* class II *POALIN*s seem well suited as ancestral signature markers and together with the *HLA-DRB1* and *HLA-DQB1* alleles may help to identify common ancestral haplotypes as well as their divergence from common haplotypes because of crossing over and other recombination events.

In conclusion, this study confirms the existence of *Alu* structural polymorphisms within at least five different locations of the *MHC* class II genomic region and provides the first results for class II *POALIN* allelic and haplotypic frequency and structural information in two different populations, the Japanese and the Caucasians. We used newly developed *POALIN*-PCR assays to show that they are informative genetic and haplotype markers for population, evolutionary, forensic and disease studies either with or without the *HLA-DRB1* and/or *-DQB1* allelic data and that they will complement the other five *POALINs* that we previously characterized and described in the *MHC* class I region (2). However, the *AluDPB2* and *AluDQA1* PCR genotyping methods may require refinement, and more work is required to determine the *MHC* class II *POALIN* genotypes in other human populations and examine their usefulness in transplantation and disease studies. The five *MHC* class II *POALIN* are nevertheless easily detected by these simple and cost effective PCR methods and are well suited to laboratories that are supported only at the low-end of the economic scale.

Acknowledgments

We thank Paula M. Moolhuijzen for her help with the initial genomic analysis for the PCR primer sets, Professor M. Ota for the Japanese HLA-typed DNA samples and Dr Campbell Witt for the Australian Caucasian HLA-typed DNA samples

References

1. Bennett EA, Coleman LE, Tsui C, Pittard WS, Devine SE. Natural genetic variation caused by transposable elements in humans. *Genetics* 2004: **168**: 933–51.
2. Kulski JK, Dunn DS. Polymorphic *Alu* insertions within the Major Histocompatibility Complex class I genomic region: a brief review. *Cytogenet Genome Res* 2005: **110**: 193–202.
3. Ray DA, Walker JA, Batzer MA. Mobile element-based forensic genomics. *Mutat Res* 2007: **616**: 24–33.
4. Antunez-de-Mayolo G, Antunez-de-Mayolo A, Antunez-de-Mayolo P et al. Phylogenetics of worldwide human populations as determined by polymorphic *Alu* insertions. *Electrophoresis* 2002: **23**: 3346–56.
5. Lander ES, Linton LM, Birren B et al. and the International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* 2001: **409**: 860–921.
6. Mnukova-Fadjelova M, Satta Y, O'hUigin C, Mayer WE, Figueroa F, Klein J. *Alu* elements of the primate major histocompatibility complex. *Mamm Genome* 1994: **5**: 405–15.
7. Svensson A, Setterblad C, Sigurdardottir N, Rask S, Andersson L, Andersson G. Evolutionary relationship between human major histocompatibility complex *HLA*-DR haplotypes. *Immunogenetics* 1996: **43**: 304–14.
8. Kulski JK, Gaudieri S, Martin A, Dawkins RL. Coevolution of *PERB11* (*MIC*) and *HLA* class I genes with HERV-16 and retroelements by extended genomic duplication. *J Mol Evol* 1999: **49**: 84–97.
9. Kulski JK, Gaudieri S, Dawkins RL. Using *alu* J elements as molecular clocks to trace the evolutionary relationships between duplicated *HLA* class I genomic segments. *J Mol Evol* 2000: **50**: 510–9.
10. Gibbons R, Dugaiczyk LJ, Girke T, Dulstermars B, Zielinski R, Dugaiczyk A. Distinguishing humans from great apes with *Alu*yB8 repeats. *J Mol Biol* 2004: **339**: 721–9.
11. Xing J, Hedges DJ, Han K, Wang H, Cordaux R, Batzer MA. *Alu* element mutation spectra: molecular clocks and the effect of DNA methylation. *J Mol Biol* 2004: **344**: 675–82.
12. de Pancorbo MM, Lopez-Martinez M, Martinez-Bouzas C et al. The Basques according to polymorphic *Alu* insertions. *Hum Genet* 2001: **109**: 224–33.
13. Stoneking M, Fontius JJ, Clifford SL et al. *Alu* insertion polymorphisms and human evolution: evidence for a larger population size in Africa. *Genome Res* 1997: **7**: 1061–71.
14. Dunn DS, Choy MK, Phipps ME, Kulski JK. The distribution of major histocompatibility complex class I polymorphic *Alu*