Open Journal of Genetics, 2014, 4, 99-124
Published Online April 2014 in SciRes. http://www.scirp.org/journal/ojgen
http://dx.doi.org/10.4236/ojgen.2014.42013

Scientific Research

Diversity and Frequencies of HLA Class I and Class II Genes of an East African Population

Trevor A. Peterson[1], Thomas Bielawny[1], Philip Lacap[1], Rae-Anne Hardie[1], Christina Daniuk[1], Lillian Mendoza[1], Subotheni Thavaneswaran[2], Tony Kariri[3], Joshua Kimani[2,3], Charles Wachihi[3], Maboku Kimani[3], Terry Blake Ball[1,2], Francis A. Plummer[1,2], Ma Luo[1,2*]

[1]National Microbiology Laboratory, Winnipeg, Canada
[2]Department of Medical Microbiology, University of Manitoba, Winnipeg, Canada
[3]Department of Medical Microbiology, University of Nairobi, Nairobi, Kenya
Email: [*]Ma.Luo@phac-aspc.gc.ca

Received 11 January 2014; revised 10 February 2014; accepted 9 March 2014

Abstract

Human Leukocyte Antigens (HLAs) play an important role in host immune responses to infectious pathogens, and influence organ transplantation, cancer and autoimmune diseases. In this study we conducted a high resolution, sequence-based genotyping of HLA class I and class II genes of more than 2000 women from Kenya, eastern Tanzania and southern Uganda around Lake Victoria and analyzed their allele, phenotype and haplotype frequencies. A considerable genetic diversity was observed at both class I and II loci. A total of 79 HLA-A, 113 HLA-B, 53 HLA-C, 25 HLA-DPA1, 60 HLA-DPB1, 15 HLA-DQA1, 44 HLA-DQB1 and 38 HLA-DRB1 alleles have been identified. The most common class I alleles were A * 02:01:01 (10.90%), B * 58:02 (8.79%), and C * 06:02:01 (16.98%). The most common class II alleles were DPA1*01:03:01 (40.60%), DPB1 * 01:01:01 (23.45%), DQA1 * 01:02:01 (31.03%), DQB1 * 03:01:01 (21.79%), DRB1 * 11:01:02 (11.65%), DRB3 * 02:02:01 (31.65%), DRB4 * 01:01:01 (10.50%), and DRB5 * 01:01:01 (10.50%). Higher than expected homozygosity was observed at HLA-B (P = 0.022), DQA1 (P = 0.004), DQB1 (P = 0.023), and DRB1 (P = 0.0006) loci. The allele frequency distribution of this population is very similar to the ones observed in other sub-Saharan populations with the exception of lower frequencies of A * 23 (5.55% versus 11.21%) and DQA1 * 03 (4.79% versus 11.72%), and higher frequencies of DPB1 * 30 (2.26% versus 0.37%) and DRB1 * 11 (21.51% versus 15.89%). The knowledge of the diversity and allele/phenotype frequencies of the HLA alleles of this east African population, can contribute to the understanding of how host genetic factors influence disease susceptibility and effective anti-retroviral treatment of HIV infections and future vaccine trials.

[*]Corresponding author.

How to cite this paper: Peterson, T.A., *et al.* (2014) Diversity and Frequencies of HLA Class I and Class II Genes of an East African Population. *Open Journal of Genetics*, **4**, 99-124. http://dx.doi.org/10.4236/ojgen.2014.42013

Keywords

HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRB1, HLA Frequencies, HLA Class I, HLA Class II, East African

1. Introduction

The Human Leukocyte Antigen (HLA) genes are the most polymorphic genes within the human genome and play an important role in initiating the immune response to invading pathogens. Associations between specific HLA alleles and increased/decreased risk of autoimmunity diseases [1] [2] as well as resistance and susceptibility to pathogenic infections have been reported in many studies [3]. The information related to the HLA allele frequencies and their compositions in a given population can contribute not only to the anthropological and transplantation studies, but also to the understanding of disease epidemiology. In turn, it can also provide valuable information for designing effective vaccine and clinical trials.

As a founder population under the extensive selective pressure exerted by many infectious diseases, East African populations are expected to have a very diverse HLA class I and II allele composition. A previous analysis of the genetic makeup of African populations and their genetic diversity has shown extensive variations in multiple genetic markers such as microsatellites, insertions/deletions and single nucleotide polymorphisms [4]. Comparison of HLA composition of Kenya and Cameroon populations showed various similarities and differences for specific frequencies of HLA-A and -B alleles [5] and similar genetic distance in HLA-C has been observed between the Caucasian and African population and between the African and Asian populations [6].

In this study, we conducted high resolution sequence-based typing of the HLA class I and II genes [7] and comprehensive analysis of allele frequencies of more than 2000 women from Kenya, eastern Tanzania and southern Uganda around lake Victoria. A considerable genetic diversity was observed at both class I and II loci. The class I and class II HLA allele frequencies of the study population were compared with that of several sub-Saharan African and North African, Western European Caucasoid and Asian populations [8]. The knowledge of the diversity of the HLA alleles of different populations, their similarities and differences can contribute to the understanding of how host genetic factors influence disease susceptibility and effectiveness of anti-retroviral treatment of HIV infections, as well as better preparation for future vaccine trials.

2. Materials and Methods

2.1. Study Population

The study population is consisted of more than 2000 women enrolled before 2008 in the Pumwani sex worker cohort in Nairobi, Kenya. HIV status is not the requirement for enrolment. In this study, 2161 individuals were typed for HLA-A, 2179 individuals for HLA-B, 2152 individuals for HLA-C, 2160 individuals for DPA1, 2215 individuals for DPB1, 2098 individuals for DQA1, 2070 individuals for DQB1, and 1090 individuals for DRB1, DRB3, DRB4 and DRB5. Due to insufficient DNA quantity for some samples, not all individuals were typed for each locus. The Ethics Committee of the University of Manitoba and the Ethics and Research Committee of Kenyatta National Hospital has approved this study and informed consent was obtained from all women enrolled in the study.

2.2. DNA Preparation

The DNA used for HLA typing was isolated from patients of the cohort using QIAmp DNA Mini Kit and QIAgen EZ1 Blood Robot (QIAgen Inc, Mississauga, Ontario, Canada) and was quantified by standard UV spectrophotometric analysis.

2.3. PCR Amplification and Sequencing Primers

Exons 2 and 3 and intron 2 were amplified for HLA-A, HLA-B, and HLA-C. Both gene-specific and allele specific primers were used to sequence the PCR products. Exon 2 of DQA1, DQB1, DPA1 and DPB1 was amplified and

sequenced. DRB typing was accomplished using a two-step sequence based genotyping method described previously [9]. Allele-specific primers were then developed to resolve ambiguous allele combinations. PCR, sequencing PCR and allele specific primers are listed in **Table 1**.

2.4. PCR Reactions

The 50 µl PCR reaction mixture consisted of 60mM Tris-HCl (pH 9.0), 15mM $(NH4)_2SO_4$, 1.5 mM $MgCl_2$, 0.1% gelatin, 100 mM each dNTP, 25 pmol of each primer, 1.25 Unit of Taq DNA polymerase (Invitrogen Life Technologies, Burlington, ON, Canada) and 100 - 200 ng DNA. The cycle parameters used in the PTC-100 programmable Thermal Controller (MJ Research, Inc., Waltham, MA USA) were 35 cycles of 1 min denaturation at 96˚C, 1 min at a specific annealing temperature (**Table 1**), 2 min extension at 72˚C, followed by a final 10 minutes extension at 72˚C. Five µl PCR product was checked by 1.0% agarose gel electrophoresis for the correctly sized PCR products. The remaining PCR products were purified with the "High Pure PCR Product Purification Kit" (Roche Molecular Biochemicals, Laval QC, Canada) or with Amicon Microcon-PCR Centrifugal filter device (Millipore, Bedford, MA USA) for sequencing PCR. Sequencing PCR programs consisted of 80 cycles of 1 minute denaturation at 96˚C, followed by 1 min at the primer's specific annealing temperature, then 2 minutes of extension at 60˚C, followed by a final extension at 10 minutes. Allele specific primers were used to resolve allele ambiguities (**Table 2**). For the allele specific primers, the annealing step was done at the melting temperature of the primer for 15 seconds to allow for highly specific binding to the specific allele.

2.5. Sequencing and HLA Typing

ABI PRISM BigDye Terminator Cycle Sequencing Ready Reaction Kits (Applied Biosystems, Foster City, CA, USA) were used for all sequencing. The amplified PCR products from this were purified and then analyzed using ABI PRISM 310 GENETIC ANALYZER (Applied Biosystems). Class I and II alleles were typed using software Codon Express™, a computer program developed based on a Taxonomy Based Sequencing Analysis (TBSA) [7]. The HLA databases were downloaded from IMTG/HLA Database (http://www.ebi.ac.uk/imgt/hla/).

2.6. Statistical Analysis

PyPop 0.7.0 was used to calculate allele frequencies, Hardy-Weinberg equilibrium, deviations from expected genotype and homozygosity frequencies, and for the Ewens-Watterson homozygosity test of neutrality. SPSS 13.0 was used to calculate phenotype frequencies and combined counts of homozygosity for members typed at all loci. Linkage disequilibrium analysis was done by using an online tool developed for LD analysis (http://www.hiv.lanl.gov/). Allele frequencies from different populations were obtained from an online collection of past studies on HLA allele frequencies (http://allelefrequencies.net/).

3. Results

3.1. HLA Class I Allele Diversity and Frequencies

Extensive diversity has been observed at each HLA class I locus in this East African population. Similar to other populations the greater allele diversity is observed in HLA-B, followed by HLA-A, and then HLA-C. A total of 79 HLA-A alleles, 113 HLA-B alleles, and 53 HLA-C alleles were identified in this population excluding many potential new alleles to be identified.

Of the 79 HLA-A alleles identified, 9 with frequencies above 5% (**Table 3**). These alleles were A * 02:01:01 (10.90%), A * 68:02:01 (9.93%), A * 74:01:01 (7.47%), A * 30:02:01 (7.43%), A * 30:01:01 (6.55%), A * 01:01:01 (6.36%), A * 02:02 (5.76%), A * 03:01:01 (5.76%), and A * 23:01:01 (5.48%). Six of the 113 HLA-B alleles identified in this population are with frequencies above 5% (**Table 3**). These alleles were B * 58:02 (8.79%), B * 15:03:01 (8.47%), B * 53:01:01 (7.85%), B * 45:01:01 (6.61%), B * 42:01:01 (6.47%), and B *49:01:01 (5.26%). Seven out of the 53 HLA-C alleles identified are with frequencies above 5% (**Table 3**). These alleles were C * 06:02:01 (16.98%), C * 07:01:01 (14.29%), C * 04:01:01 (13.01%), C * 17:01:01 (9.36%), C * 02:10 (6.67%), C * 16:01:01 (5.95%), and C * 18:01:01 (5.81%). A total of 25 HLA-A alleles, 39 HLA-B al- leles and 17 HLA-C alleles were only identified in one individual in the population.

Table 1. List of HLA Class I and II PCR primers and Sequencing PCR primers.

Primer Name	Specificity	Primer Sequence (5' to 3')	Location	Annealing temp (˚C)	PCR Product length	Purpose
APCRF	5' HLA-A	GAAACGGCCTCTGTGGGGAGAAGCAA	Intron 1	58.7	984	PCR
APCRR	3'HLA-A	TGTTGGTCCCAATTGTCTCCCCTC	Intron 3	58.7	984	PCR
ASEQ5F	Exon 3 of HLA-A	GGTTTCATTTTCAGTTTAGGCCA	Intron 2	51.6	-	Sequencing PCR
ASEQ3R	Exon 2 of HLA-A	ATCTCGGACCCGGAGACTGTG	Intron 2	58.9	-	Sequencing PCR
BPCRF	5' HLA-B	GGGAGGAGCGAGGGGACCGCAG	Intron 1	64.0	942	PCR
BPCRR	3'HLA-B	GGAGGCCATCCCCGGCGACCTAT	Intron 3	64.0	942	PCR
BSEQ5F	Exon 3 of HLA-B	GGGGACGGGGCTGA	Intron 2	53.8	-	Sequencing PCR
BSEQ3RN	Exon 2 of HLA-B	GGATGGGGAGTCGTGACCTG	Intron 2	59.0	-	Sequencing PCR
CPCRF	5' HLA-C	AGCGAGGTGCCCGCCCGGCGA	Intron 1	64.0	946	PCR
CPCRR	3'HLA-C	ATCTCCCGGGATGGCCTCCCACG	Intron 3	64.0	946	PCR
CSEQ5F	Exon 3 of HLA-C	GGGGACGGGGCTGAC	Intron 2	57.8	-	Sequencing PCR
CSEQ3R	Exon 2 of HLA-C	GCCGTCCGTGGGGGATG	Intron 2	59.2	-	Sequencing PCR
DPAPCRF	5' HLA-DPA1	ACATTTTGTCGTGTTTTTCTCT	Exon 2	47.8	334	PCR
DPAPCRR	3' HLA-DPA1	CTCTCATCCCTTCCAGTTG	Exon 2	47.8	334	PCR
DPASEQF	5' of Exon 2 of HLA-DPA1	AATGTTAGCCAGCCCGCC	Exon 2	55.1	-	Sequencing PCR
DPASEQR	3' of Exon 2 of HLA-DPA1	GCCTGAGTGTGGTTGGAACG	Exon 2	57.9	-	Sequencing PCR
DPBPCRF	5' HLA-DPB1	GAGAGTGGCGCCTCCGCTCAT	Exon 2	60.8	326	PCR
DPBPCRR	3' HLA-DPB1	GCCGGCCCAAAGCCCTCACTC	Exon 2	60.8	326	PCR
DPBSEQF	5' of Exon 2 of HLA-DPB1	CCTCCCCGCAGAGAATTAC	Exon 2	54.7	-	Sequencing PCR
DPBSEQR	3' of Exon 2 of HLA-DPB1	GAGGTGAGTGAGGGCTTTG	Exon 2	54.7	-	Sequencing PCR
DQAPCRF	5' HLA-DQA1	ATCTTCACTCATCAGCTGACCA	Exon 2	55.0	726	PCR
DQAPCRR	3' HLA-DQA1	GCTGACCCAGTGTCACGGGAG	Exon 2	55.0	726	PCR
DQASEQ3	5' of Exon 2 of HLA-DQA1	GCCTCTTGTGGTGTAAACTTG	Exon 2	53.0	-	Sequencing PCR
DQASEQ2	3' of Exon 2 of HLA-DQA1	CATTGGTAGCAGCAGTAG	Exon 2	50.0	-	Sequencing PCR
DQBPCRF	5' HLA-DQB1	TCCCCGCAGAGGATTTCGTG	Exon 2	57.0	293	PCR
DQBPCRR	3' HLA-DQB1	GGCGACGACGCTCACCTC	Exon 2	57.0	293	PCR
DQBSEQF	5' of Exon 2 of HLA-DQB1	GCAGAGGATTTCGTGTTCCAG	Exon 2	55.0	-	Sequencing PCR
DQBSEQR	3' of Exon 2 of HLA- DQB1	GTAGTTGTGTCTGCA	Exon 2	43.1	-	Sequencing PCR
DRBPCRF	5' HLA-DRB1	GTTCGTGTCCCCACAGCACGTTTC	Exon 2	55.0	293	PCR
DRBPCRR	3' HLA-DRB1	CATGCTCACCTCGCCGCTGCAC	Exon 2	55.0	293	PCR
DRBSEQ4	3' of Exon 2 of HLA-DRB1	CACTGTGAAGCTCTC	Exon 2	59.0	-	Sequencing PCR

Continued

DRBSEQ1	5' of Exon 2 of HLA-DRB1	TCGACAGCGACGTGGGGGAG	Exon 2	55.0	-	Sequencing PCR
DRB031234PCR	DRB1 loci for DRB1 * 03 * 11 * 12 * 13 * 14	CGTTTCTTGGAGTACTCTAC	Exon 2	50.8	279*	High Res PCR
DRB04PCR	DRB1 loci for DRB1 * 04	CCTGGACAGATACTTCTATC	Exon 2	51.8	216*	High Res PCR
DRB08PCR	DRB1 loci for DRB1 * 08	TTCTTGGAGTACTCTACGG	Exon 2	51.4	275*	High Res PCR
DRB1516PCR	DRB1 loci for DRB1 * 15 * 16	TGTGGCAGCCTAAGAGG	Exon 2	53.0	281*	High Res PCR
DRB3PCR	DRB3 loci	ATTACTGCAGACACAACTAC	Exon 2	48.7	243*	High Res PCR
DRB4PCR	DRB4 loci	GAGCGAGTGTGGAACCTGATC	Exon 2	56.9	189*	High Res PCR
DRB5PCR	DRB5 loci	CACGTTTCTTGCAGCAGGA	Exon 2	52.6	270*	High Res PCR

*The primers were combined with DRBPCRR primer in PCR reaction to produce the PCR products.

3.2. HLA Class II Allele Diversity and Frequencies

As expected, the HLA class II loci is less diverse than the class I loci in this population. A total of 25 DPA1 alleles, 60 DPB1 alleles, 15 DQA1 alleles, 44 DQB1 alleles and 40 DRB1 alleles were identified excluding potential new alleles to be confirmed.

Class II alleles with frequencies above 5% in this population were 4 DPA1 alleles [DPA1 * 01:03:01 (40.60%), DPA1 * 02:01:01 (18.40%), DPA1 * 03:01 (17.89%), DPA1 * 02:02:02 (15.83%)], 6 DPB1 alleles [DPB1 * 01:01:01 (23.45%), DPB1 * 04:02:01 (17.56%), DPB1 * 02:01:02 (14.47%), DPB1 * 04:01:01 (9.59%), DPB1 * 03:01:01 (7.86%), DPB1 * 18:01 (5.35%)], 5 DQA1 alleles [DQA1 * 01:02:01 (31.03%), DQA1 * 05:01:01 (25.79%), DQA1 * 01:01:01 (15.59%), DQA1 * 04:01:01 (9.10%), DQA1 * 02:01 (7.22%)], 6 DQB1 alleles [DQB1 * 03:01:01 (21.79%), DQB1 * 06:02:01 (17.46%), DQB1 * 02:01:01 (17.10%), DQB1 * 05:01:01 (16.09%), DQB1 * 06:04:01 (6.91%), DQB1 * 04:02 (6.33%)], and 8 DRB1 alleles [DRB1 * 11:01:02 (11.65%), DRB1 * 13:02:01 (10.78%), DRB1 * 15:03:01 (10.23%), DRB1 * 07:01:01 (7.57%), DRB1 * 11:02 (7.06%), DRB1 * 03:01:01 (6.70%), DRB1 * 03:02:01 (6.42%), and DRB1 * 13:01:01 (5.05%)] (**Table 4**). For other functional DRB loci, the DRB3 phenotype was most common in the population at 84.77%, the DRB4 phenotype was observed in 22.57% population, and the frequency of the DRB5 phenotype was 21.28%. Only 5 unique DRB3 alleles were identified in this population (**Table 4**) with DRB3 * 02:02:01 (54.76%), DRB3 * 03:01:01 (25.79%) and DRB3 * 01:01:02 (18.57%) as the most abundant alleles. Only one allele was identified at DRB4 (DRB4 * 01:01:01) and DRB5 (DRB5*01:01:01). Four HLA-DPA1 alleles, 11 DPB1 alleles, 1 DQA1 allele, 5 DQB1 alleles, and six DRB1 alleles were only found in one individual in the population.

3.3. Homozygosity at Class I and Class II Loci

As expected the frequency of homozygosity was lower in HLA class I genes than that in class II genes. The frequencies of homozygosity for class I loci are 5.78%, 6.71% and 9.11% for HLA-B, -A, and -C respectively. Whereas, the frequencies of homozygosity for class II loci are 27.50%, 13.23%, 23.31%, 16.68%, and 8.50% for DPA1, DPB1, DQA1, DQB1 and DRB1, respectively.

Among the 2103 individuals who were fully typed at the class I loci, 297 individuals were homozygous at one class I locus (14.12%), 56 individuals (2.66%) were homozygous at two class I loci, and 16 individuals (0.76%) were homozygous at all three class I loci. Among the 867 individuals who were fully typed at the class II loci, 249 were homozygous at one class II locus, 137 (15.80%) were homozygous at two class II loci, 61 (7.04%) were homozygous at three class II loci, 12 (1.38%) were homozygous at four class II loci, and 8 (0.92%) were homozygous at all the class II loci.

We analyzed homozygosity at all class I and class II loci (excluding DRB3, DRB4, and DRB5) for 834 individuals who have been fully typed for HLA-A, -B, -C, -DRB1, -DPA1, -DPB1, -DQA1 and -DQB1. Of these

Table 2. Allele specific primers for determining multiple results.

Primer Name	Specificity	Primer Sequence (5' to 3')	Specific Binding Location	Melting Temp (˚C)
A24SEQN	Exon 2 of HLA-A	CCCCGCTTCATCGCA	Codon 24	56.3
A82SEQR	Exon 2 of HLA-A	TCTGGTTGTAGTAGCCGC	Codon 82	57.3
A97SEQ	Exon 3 of HLA-A	TTCTCACACCGTCCAGAG	Codon 97	57.3
A99SEQ	Exon 3 of HLA-A	CACACCATCCAGATGATGTA	Codon 99	55.8
A114SEQN	Exon 3 of HLA-A	CTCCGCGGGTACCA	Codon 114	55.9
A116SEQ	Exon 3 of HLA-A	GCGGGTACCACCAGT	Codon 116	59.2
A70SSP1	Exon 2 of HLA-A	CTCTCGGTCAGTCTGTGAG	Codon 70	59.7
A70SSP2	Exon 2 of HLA-A	CTCTCGGTCAGTCTGTGAC	Codon 70	59.7
A77SSP1	Exon 2 of HLA-A	GCGCAGGGTCCCCAGGTC	Codon 77	66.4
A77SSP2	Exon 2 of HLA-A	GCGCAGGGTCCCCAGGCT	Codon 77	66.4
B23SEQN	Exon 2 of HLA-B	AGCCCCGCTTCATC	Codon 23	52.9
B24SEQN	Exon 2 of HLA-B	AGCCCCGCTTCATCG	Codon 24	56.3
B24SEQ2	Exon 2 of HLA-B	AGCCCCGCTTCATCTCA	Codon 24	57
B45SEQN	Exon 2 of HLA-B	CGCCAGTCCGAGGAT	Codon 45	56.3
B83SEQN	Exon 2 of HLA-B	TCTGGTTGTAGTAGCC	Codon 83	51.5
B94SEQ	Exon 3 of HLA-B	AGCGAGGCCGGGTCTCACAT	Codon 94	64
B94SEQNS	Exon 3 of HLA-B	AGGCCGGGTCTCACAT	Codon 94	56.7
B97SEQ	Exon 3 of HLA-B	GGTCTCACATCATCCAGA	Codon 97	55.0
B99SEQ	Exon 3 of HLA-B	ACCCTCAGAGGATGTAC	Codon 99	54.6
B99SEQ2N	Exon 3 of HLA-B	TCACATCCAGAGGATGTAT	Codon 99	53.2
B116SEQN	Exon 3 of HLA-B	CGGGCATGACCAGTA	Codon 116	53.6
B116SEQ2N	Exon 3 of HLA-B	GCGGGCATAACCAGTA	Codon 116	54.1
B156SEQN	Exon 3 of HLA-B	TCCAGGTAGGCTCTCC	Codon 156	56.7
B163SEQRN	Exon 3 of HLA-B	CCACTCCACGCACTC	Codon 163	56.3
B45SEQN	Exon 2 of HLA-B	GCCGCGAGTCCGAGGAT	Codon 45	61.8
C35SEQ	Exon 2 of HLA-C	ACGACACGCAGTTCGTGCG	Codon 35	61.9
C43SEQ	Exon 2 of HLA-C	GACGCCGCGAGTCCG	Codon 43	61.8
C129SEQR	Exon 3 of HLA-C	GGTCCAGGAGCGCAGG	Codon 129	61.8
C24SEQ	Exon 2 of HLA-C	GAGCCCCGCTTCATCT	Codon 24	56.7
DPA15SEQ	Exon 2 of HLA-DPA1	TATGCCGCGTTTGTACAGACC	Codon 15	60.0
DPA15SEQRV	Exon 2 of HLA-DPA1	TATGCCATGTTTGTACAGACC	Codon 15	56.1
DPB11SEQ	Exon 2 of HLA-DPB1	AGAATTACCTTTTCCAGG	Codon 11	50.5
DPB36SEQ	Exon 2 of HLA-DPB1	ACCGGGAGGAGTTCGC	Codon 36	59.2
DPB36SEQRV	Exon 2 of HLA-DPB1	AACCGGGAGGAGTTCGT	Codon 36	57.0
DPB76SEQR	Exon 2 of HLA-DPB1	CTCGTAGTTGTGTCTGCAT	Codon 76	55.4
DQB9SEQ	Exon 2 of HLA-DQB1	CGCAGAGGATTTCGTGTA	Codon 9	55.0
DQB0201F	Exon 2 of HLA-DQB1	GTGCGTCTTGTGAGCAGAAG	Codon 30	59.9
DQB0301F	Exon 2 of HLA-DQB1	GGACGGAGCGCGTGCGTT	Codon 26	64.1
DRB58SEQN	Exon 2 of HLA-DRB1	TGGGGCGGCCTGATGA	Codon 58	59.2
DRB86SEQ	Exon 2 of HLA-DRB1	ACTGTGAAGCTCTCAC	Codon 86	51.5

Table 3. Frequency of HLA class I (HLA-A, -B, and -C) phenotypes and genotypes in the Pumwani sex worker cohort.

HLA-A	Phenotype Count (n = 2161)	Allele Count (2n = 4322)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-A	Phenotype Count (n = 2161)	Allele Count (2n = 4322)	Phenotype Frequency (%)	Allele Frequency (%)
01:01:01	261	275	12.08	6.36	26:09	1	1	0.05	0.02
01:01:02	3	3	0.14	0.07	26:12	18	18	0.83	0.42
01:02	19	19	0.88	0.44	26:18	1	1	0.05	0.02
01:03	36	36	1.67	0.83	26:30	5	5	0.23	0.12
01:09	23	23	1.06	0.53	29:01:01	65	66	3.01	1.53
01:14	3	3	0.14	0.07	29:02:01	167	174	7.73	4.03
02[a]	1	1	0.05	0.02	29:03	1	1	0.05	0.02
02:01:01	443	471	20.50	10.90	29:04	5	5	0.23	0.12
02:01:03	1	1	0.05	0.02	29:06	1	1	0.05	0.02
02:02	243	249	11.24	5.76	29:10	1	1	0.05	0.02
02:04	11	11	0.51	0.25	29:11	6	6	0.28	0.14
02:05:01	104	105	4.81	2.43	29:15	2	2	0.09	0.05
02:06:01	1	1	0.05	0.02	29:28	1	1	0.05	0.02
02:07:01	4	4	0.19	0.09	30:01:01	274	283	12.68	6.55
02:09	1	1	0.05	0.02	30:02:01	310	321	14.35	7.43
02:14	55	57	2.55	1.32	30:04	74	75	3.42	1.74
02:16:01	2	2	0.09	0.05	30:07	1	1	0.05	0.02
02:20:01	1	1	0.05	0.02	30:09	19	19	0.88	0.44
02:22:01	1	1	0.05	0.02	30:10	21	21	0.97	0.49
02:25	1	1	0.05	0.02	31:01:02	15	15	0.69	0.35
02:40	8	8	0.37	0.19	31:01:06	1	1	0.05	0.02
02:58	1	1	0.05	0.02	31:04	46	46	2.13	1.06
02:85	4	4	0.19	0.09	32:01:01	55	57	2.55	1.32
02:102	2	2	0.09	0.05	32:03	2	2	0.09	0.05
03[a]	1	1	0.05	0.02	33:01:01	27	28	1.25	0.65
03:01:01	243	249	11.24	5.76	33:03:01	50	50	2.31	1.16
03:02:01	2	2	0.09	0.05	34:01:01	1	1	0.05	0.02
03:07	3	3	0.14	0.07	34:02:01	95	97	4.40	2.24
03:08	1	1	0.05	0.02	34:03	6	6	0.28	0.14
03:75	1	1	0.05	0.02	36:01	120	122	5.55	2.82
11:01:01	2	2	0.09	0.05	66:01:01	144	147	6.66	3.40
23:01:01	233	237	10.78	5.48	66:02	5	5	0.23	0.12
23:04	1	1	0.05	0.02	66:03	6	6	0.28	0.14
23:05	2	2	0.09	0.05	68:01:01	55	55	2.55	1.27
24:01:01	5	5	0.23	0.12	68:01:05	1	1	0.05	0.02
24:02:01	92	92	4.26	2.13	68:02:01	406	429	18.79	9.93
24:02:02	1	1	0.05	0.02	68:31	1	1	0.05	0.02
24:05:01	1	1	0.05	0.02	74:01:01	301	323	13.93	7.47
24:13	1	1	0.05	0.02	74:03	12	12	0.56	0.28
24:35	1	1	0.05	0.02	80:01:01	11	11	0.51	0.25
26:01:01	12	12	0.56	0.28	unidentified[b]	7	13	0.32	0.30

Continued

HLA-B	Phenotype Count (n = 2179)	Allele Count (2n = 4358)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-B	Phenotype Count (n = 2179)	Allele Count (2n = 4358)	Phenotype Frequency (%)	Allele Frequency (%)
07[a]	1	1	0.05	0.02	39:01:01	2	2	0.09	0.05
07:02:01	183	185	8.40	4.25	39:06:02	1	1	0.05	0.02
07:05:01	24	24	1.10	0.55	39:10	66	67	3.03	1.54
07:09	3	3	0.14	0.07	39:16	1	1	0.05	0.02
07:12	1	1	0.05	0.02	39:24	7	7	0.32	0.16
07:26	3	3	0.14	0.07	40:12	9	10	0.41	0.23
07:33	1	1	0.05	0.02	40:16	10	10	0.46	0.23
08:01:01	111	111	5.09	2.55	41:01	46	50	2.11	1.15
08:02	1	1	0.05	0.02	41:02:01	13	13	0.60	0.30
08:07	1	1	0.05	0.02	42:01:01	276	282	12.67	6.47
08:10	1	1	0.05	0.02	42:02	10	10	0.46	0.23
08:12	1	1	0.05	0.02	42:05:02	1	1	0.05	0.02
08:23	1	1	0.05	0.02	44:02:01	2	2	0.09	0.05
13:01:01	3	3	0.14	0.07	44:03:01	103	105	4.73	2.41
13:02:01	83	87	3.81	2.00	44:03:02	21	21	0.96	0.48
13:03	14	14	0.64	0.32	44:15	42	42	1.93	0.96
14:01:01	28	28	1.28	0.64	44:18	3	3	0.14	0.07
14:02:01	65	66	2.98	1.51	44:28:01	2	2	0.09	0.05
14:03	3	3	0.14	0.07	44:38	1	1	0.05	0.02
14:05	4	4	0.18	0.09	45:01:01	274	288	12.57	6.61
14:06:01	9	9	0.41	0.21	45:02	3	3	0.14	0.07
14:06:02	15	15	0.69	0.34	45:03	2	2	0.09	0.05
15[a]	1	1	0.05	0.02	45:04	8	9	0.37	0.21
15:02:01	3	3	0.14	0.07	45:06	1	1	0.05	0.02
15:03:01	353	369	16.20	8.47	47:01:01	22	22	1.01	0.50
15:10:01	203	211	9.32	4.84	47:03	38	39	1.74	0.89
15:16:01	13	13	0.60	0.30	47:04	1	1	0.05	0.02
15:17:01	43	43	1.97	0.99	47:05	6	6	0.28	0.14
15:18:01	1	1	0.05	0.02	48:01:01	1	1	0.05	0.02
15:29	1	1	0.05	0.02	48:05	9	9	0.41	0.21
15:31	6	6	0.28	0.14	48:06	1	1	0.05	0.02
15:37	1	1	0.05	0.02	49:01:01	220	229	10.10	5.25
15:55	3	3	0.14	0.07	49:02	2	2	0.09	0.05
15:61	2	2	0.09	0.05	50:01:01	5	5	0.23	0.11
15:64	1	1	0.05	0.02	50:02	1	1	0.05	0.02
15:67	22	22	1.01	0.50	51:01:01	78	81	3.58	1.86
15:69	1	1	0.05	0.02	51:01:02	1	1	0.05	0.02
15:90	1	1	0.05	0.02	51:01:04	1	1	0.05	0.02
15:91	1	1	0.05	0.02	51:02:02	1	1	0.05	0.02
15:123:01	1	1	0.05	0.02	51:04	1	1	0.05	0.02
18:01:01	139	141	6.38	3.24	51:13:02	1	1	0.05	0.02

Continued

18:03	36	36	1.65	0.83	52:01:01	1	1	0.05	0.02
18:05	1	1	0.05	0.02	53:01:01	323	342	14.82	7.85
18:07	3	3	0.14	0.07	53:08:01	3	3	0.14	0.07
18:11	2	2	0.09	0.05	53:08:02	1	1	0.05	0.02
18:18	1	1	0.05	0.02	56:01:01	5	5	0.23	0.11
27:03	23	23	1.06	0.53	57:01:01	10	10	0.46	0.23
27:05:02	6	6	0.28	0.14	57:02:01	75	77	3.44	1.77
27:26	4	4	0.18	0.09	57:03:01	160	163	7.34	3.74
3[a]	1	1	0.05	0.02	57:07	1	1	0.05	0.02
35:01:01	93	95	4.27	2.18	58[a]	10	10	0.46	0.23
35:02:01	32	32	1.47	0.73	58:01:01	197	204	9.04	4.68
35:03:01	1	1	0.05	0.02	58:02	369	383	16.93	8.79
35:08:01	2	2	0.09	0.05	58:06	2	2	0.09	0.05
35:10	1	1	0.05	0.02	73:01	9	9	0.41	0.21
35:25	1	1	0.05	0.02	81:01:01	188	192	8.63	4.41
35:34	1	1	0.05	0.02	82:02	2	2	0.09	0.05
35:37	1	1	0.05	0.02	unidentified[b]	6	10	0.28	0.23
37:01:01	12	12	0.55	0.28					

HLA-C	Phenotype Count (n = 2152)	Allele Count (2n = 4304)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-C	Phenotype Count (n = 2152)	Allele Count (2n = 4304)	Phenotype Frequency (%)	Allele Frequency (%)
01:02:01	6	6	0.28	0.14	07:08	3	3	0.14	0.07
02:02:02	61	61	2.83	1.42	07:10	1	1	0.05	0.02
02:02:03	2	2	0.09	0.05	07:12	1	1	0.05	0.02
02:02:05	5	5	0.23	0.12	07:14	3	3	0.14	0.07
02:05	1	1	0.05	0.02	07:16	1	1	0.05	0.02
02:10	277	287	12.87	6.67	08:01:01	1	1	0.05	0.02
03:02:01	47	48	2.18	1.12	08:02:01	153	157	7.11	3.63
03:03:01	3	3	0.14	0.07	08:04	57	57	2.65	1.32
03:04:01	38	38	1.77	0.88	08:07	2	2	0.09	0.05
03:04:02	188	195	8.74	4.53	12:02:01	1	1	0.05	0.02
03:10	1	1	0.05	0.02	12:03:01	61	61	2.83	1.42
04:01:01	523	560	24.30	13.01	12:03:04	1	1	0.05	0.02
04:04:01	19	21	0.88	0.49	14:02:01	17	17	0.79	0.40
04:05	8	8	0.37	0.19	14:02:03	14	14	0.65	0.33
04:07	51	51	2.37	1.19	14:03	18	18	0.84	0.42
04:08	1	1	0.05	0.02	14:05	1	1	0.05	0.02
04:11	1	1	0.05	0.02	14:08	1	1	0.05	0.02
05:01:01	16	16	0.74	0.37	15:01:01	2	2	0.09	0.05
06:02:01	675	731	31.37	16.98	15:05:01	69	70	3.21	1.63
06:03	1	1	0.05	0.02	16:01:01	249	256	11.57	5.95
06:06	1	1	0.05	0.02	16:02:01	23	23	1.07	0.53
06:08	1	1	0.05	0.02	16:04:01	10	10	0.46	0.23
07:01:01	567	615	26.35	14.29	16:21	1	1	0.05	0.02

Continued

07:01:05	3	3	0.14	0.07	17:01:01	391	403	18.17	9.36
07:02:01	180	186	8.36	4.32	17:01:04	1	1	0.05	0.02
07:04:01	94	94	4.37	2.18	18:01:01	245	250	11.38	5.81
07:05	9	9	0.42	0.21	unidentified[b]	2	2	0.09	0.05

a: Incomplete low-res typing results, b: Cannot determine allele due to SNPs or SNP combinations in sequence that do not exist in database.

fully typed individuals, 246 (29.50%) were homozygous for one loci, 148 (17.56%) were homozygous at two loci, 75 (8.99%) were homozygous at three loci, 20 (2.40%) at four loci, 8 (0.96%) were homozygous at five loci, 1 (0.12%) at 6 loci, 4 (0.48%) at 7 loci and 2 (0.24%) at all loci. The frequency of individuals with one or more homozygous loci was 60.43% in this fully typed subgroup.

The observed homozygosity in HLA-B is higher than expected (P = 0.022), while there is no significant difference between expected and observed homozygosity in HLA-A (P = 0.063) and HLA-C (P = 0.729). At the class II loci, higher than expected homozygosity was observed in HLA-DQA1 (P =0.004), HLA-DQB1 (P = 0.023), and HLA–DRB1 (P = 0.006). The Ewens-Watterson homozygosity test of neutrality was tested on each locus however there were no significant results at any loci. All the class I and class II Allele frequency distributions of the population are visualized in **Figure 1** from highest to lowest frequency.

3.4. HLA Genotypes, Haplotypes and Linkage Disequilibrium

Genotype frequencies were analyzed at each locus using Pypop 0.7.0. The following genotypes were more frequent than expected A * 74:01:01-A * 36:01:01 (P = 0.003), A * 30:01:01-24:02:01 (P = 0.003), A * 24:02:01-02: 02 (P = 0.010), A * 23:01:01-A * 03:01:01 (P = 0.012), A * 74:01-A * 74:01 (P = 0.007), B * 45:01:01-B * 44:03:01 (P = 0.037), B * 57:03:01-B * 07:02:01 (P = 0.003), C * 03:04:02-C * 02:10 (P = 0.008), C * 04:01:01-C * 03:04:01 (P = 0.007), C * 17:01:01-C * 04:07 (P = 0.020), C * 18:01-C * 07:04:01 (P = 0.020), C * 18:01-C * 16:01:01 (P = 0.012), DPA1 * 03:01-DPA1 * 01:07 (P = 0.001), DPA1 * 01:03:01-DPA1 * 01:03:01 (P = 0.004), DPA1 * 02:02:02-DPA1 * 02:02:02 (P = 0.027), DPB1 * 03:01:01-DPB1 * 30:01 (P = 0.003), DPB1 * 04:01:01-DPB1 * 01:01:01 (P = 0.046), DQA1 * 01:01:01-DQA1 * 01:01:01 (P = 0.008), DQA1 * 03:01:01-DQA1 * 03:01:01 (P = 0.006), DQB1 * 05:01:01-DQB1 * 05:01:01 (P = 0.036), DQB1 * 06:03:01-DQB1 * 03:01:01 (P = 0.034), DQB1 * 06:04:01-DQB1 * 02:01:01 (P = 0.045), and DRB1 * 07:01:01-DRB1 * 15:03:01 (P = 0.004) (**Table 5(a)**). The following genotypes were less frequent than expected A * 74:01:01-A * 30:01:01 (P = 0.027), C * 16:01:01-C * 06:02:01 (P = 0.030), DPA1 * 02:02:02-DPA1 * 01:03:01 (P = 0.024), DPB1 * 04:02:01-DPB1 * 03:01:01 (P = 0.035) and DQB1 * 03:01:01-DQB1 * 02:01:01 (P = 0.048) (**Table 5(b)**)

3.5. Linkage Disequilibrium Analysis of HLA Class I and II Alleles Identified Specific Class I and Class II Haplotypes

The most abundant two-locus haplotypes were class II haplotypes [DPA1 * 01:03:01-DQA1 * 01:02:01 (16.26%), DQA1 * 05:01:01-DQB1 * 03:01:01 (16.05%), DQA1 * 01:02:01-DQB1 * 06:02:01 (14.88%) and DPA1 * 02:02:02-DPB1 * 01:01:01 (14.43%)]. The most frequent class I two-locus haplotypes were B * 58:02-C * 06:02:01 (8.04%), B * 42:01:01-C * 17:01:01 (6.37%) and B * 15:03:01-C * 02:10 (6.37%). The most common two-locus class I and II haplotypes were C * 07:01:01-DPA1 * 01:03:01 (8.02%), C * 06:02:01-DPA1 * 01:03:01 (6.64%), C * 07:01:01-DQA1 * 01:02:01 (5.93%), C * 04:01:01-DQA1 * 01:02:01 (5.89%), and C * 04:01:01-DPA1 * 01:03:01 (5.89%) (**Table 6**).

Linkage disequilibrium analysis of HLA class I and II alleles identified specific class I and class II haplotypes (**Table 7** and **Table 8**). This data was obtained using the online tool previously described in the Methods section.

There were 13 unique class I (A-B-C) haplotypes with frequencies above 1% in the population (**Table 9**) that were identified using Pypop 0.7.0. The most prevalent class I haplotypes were 30:01:01-42:01:01-17:01:01 (3.12%), 02:02-58:02-06:02:01 (2.09%), 36:01-53:01:01-04:01:01 (2.09%) 66:01:01-58:02-06:02:01 (2.04%), 68:02:01-15:10:01-03:04:02 (1.89%), 02:01:01-15:03:01-02:10 (1.74%), 30:02:01-45:01:01-16:01:01 (1.64%), 74:01:01-15:03:01-02:10 (1.57%), 01:01:01-81:01:01-18:01:01 (1.49%), 68:02:01-07:02:01-07:02:01 (1.42%), 74:01:01-58:02-06:02:01 (1.41%), 02:01:01-45:01:01-16:01:01 (1.27%), and 74:01:01-49:01:01-07:01:01

Table 4. Frequency of HLA class II (-DPA1, -DPB1, -DQA1, -DQB1, -DRB1, -DRB3, -DRB4, -DRB5) phenotypes and genotypes in the Pumwani sex worker cohort.

HLA-DPA1	Phenotype Count (n = 2160)	Allele Count (2n = 4320)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-DPA1	Phenotype Count (n = 2160)	Allele Count (2n = 4320)	Phenotype Frequency (%)	Allele Frequency (%)
01:03:01	1368	1754	63.33	40.60	02:01:05	7	7	0.32	0.16
01:03:02	12	12	0.56	0.28	02:01:06	1	1	0.05	0.02
01:03:03	16	16	0.74	0.37	02:01:07	5	5	0.23	0.12
01:04	36	36	1.67	0.83	02:02:01	1	1	0.05	0.02
01:05	40	42	1.85	0.97	02:02:02	615	684	28.47	15.83
01:06:01	3	3	0.14	0.07	02:02:03	7	7	0.32	0.16
01:06:02	4	4	0.19	0.09	02:03	11	11	0.51	0.25
01:07	13	13	0.60	0.30	02:04	1	1	0.05	0.02
01:09	3	3	0.14	0.07	03:01	704	773	32.59	17.89
02:01:01	728	795	33.70	18.40	03:02	30	30	1.39	0.69
02:01:02	23	23	1.06	0.53	03:03	1	1	0.05	0.02
02:01:03	1	1	0.05	0.02	04:01	82	83	3.80	1.92
02:01:04	10	10	0.46	0.23	unidentified[b]	3	4	0.14	0.09
HLA-DPB1	Phenotype Count (n = 2215)	Allele Count (2n = 4430)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-DPB1	Phenotype Count (n = 2215)	Allele Count (2n = 4430)	Phenotype Frequency (%)	Allele Frequency (%)
01:01:01	922	1039	41.63	23.45	33:01	3	3	0.14	0.07
01:01:02	46	46	2.08	1.04	34:01	30	30	1.35	0.68
02:01:02	598	641	27.00	14.47	35:01	1	1	0.05	0.02
02:01:04	2	2	0.09	0.05	39:01	19	19	0.86	0.43
02:01:06	2	2	0.09	0.05	40:01	18	18	0.81	0.41
02:02	1	1	0.05	0.02	46:01	3	3	0.14	0.07
03:01:01	333	348	15.03	7.86	47:01	9	9	0.41	0.20
04:01:01	405	425	18.28	9.59	48:01	2	2	0.09	0.05
04:02:01	700	778	31.60	17.56	49:01	8	8	0.36	0.18
04:03	3	3	0.14	0.07	50:01	1	1	0.05	0.02
05:01:01	6	6	0.27	0.14	51:01	4	4	0.18	0.09
06:01	1	1	0.05	0.02	55:01	87	87	3.93	1.96
08:01	2	2	0.09	0.05	57:01	1	1	0.05	0.02
09:01	17	17	0.77	0.38	61:01N	2	2	0.09	0.05
10:01	2	2	0.09	0.05	65:01	3	3	0.14	0.07
11:01:01	80	83	3.61	1.87	66:01	3	3	0.14	0.07
11:01:02	1	1	0.05	0.02	68:01	2	2	0.09	0.05
13:01:01	119	121	5.37	2.73	79:01	2	2	0.09	0.05
14:01	7	7	0.32	0.16	81:01	2	2	0.09	0.05
15:01	32	32	1.44	0.72	86:01	2	2	0.09	0.05
16:01	7	7	0.32	0.16	87:01	2	2	0.09	0.05
17:01:01	197	202	8.89	4.56	89:01	4	4	0.18	0.09
18:01	228	237	10.29	5.35	90:01	5	5	0.23	0.11
19:01:01	34	34	1.53	0.77	92:01	2	2	0.09	0.05
21:01	1	1	0.05	0.02	95:01	1	1	0.05	0.02

Continued

23:01:01	32	32	1.44	0.72	98:01	7	7	0.32	0.16
24:01	1	1	0.05	0.02	107:01	1	1	0.05	0.02
25:01	8	8	0.36	0.18	116:01	1	1	0.05	0.02
26:01:02	10	10	0.45	0.23	122:01	2	2	0.09	0.05
29:01	2	2	0.09	0.05	unidentified[b]	6	12	0.27	0.27
30:01	99	100	4.47	2.26					
HLA-DQA1	Phenotype Count (n = 2098)	Allele Count (2n = 4196)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-DQA1	Phenotype Count (n = 2098)	Allele Count (2n = 4196)	Phenotype Frequency (%)	Allele Frequency (%)
01:01:01	587	654	27.98	15.59	04:03N	11	11	0.52	0.26
01:02:01	1082	1302	51.57	31.03	05:01:01	929	1082	44.28	25.79
01:03:01	165	171	7.86	4.08	05:02	14	14	0.67	0.33
01:06	7	7	0.33	0.17	05:04	48	49	2.29	1.17
02:01	292	303	13.92	7.22	05:05:01	1	1	0.05	0.02
03:01:01	190	201	9.06	4.79	06:01:01	15	15	0.71	0.36
04:01:01	362	382	17.25	9.10	06:02	4	4	0.19	0.10
HLA-DQB1	Phenotype Count (n = 2070)	Allele Count (2n = 4140)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-DQB1	Phenotype Count (n = 2070)	Allele Count (2n = 4140)	Phenotype Frequency (%)	Allele Frequency (%)
02:01:01	643	708	31.06	17.10	05:02:01	8	8	0.39	0.19
02:03	12	12	0.58	0.29	05:03:01	22	23	1.06	0.56
02:05	1	1	0.05	0.02	05:04	2	2	0.10	0.05
03:01:01	793	902	38.31	21.79	06:01:01	3	3	0.14	0.07
03:01:02	21	21	1.01	0.51	06:02:01	658	723	31.79	17.46
03:02:01	46	46	2.22	1.11	06:03:01	90	91	4.35	2.20
03:02:02	1	1	0.05	0.02	06:04:01	272	286	13.14	6.91
03:03:02	28	28	1.35	0.68	06:04:02	3	3	0.14	0.07
03:03:03	6	6	0.29	0.14	06:05:01	96	99	4.64	2.39
03:04	23	23	1.11	0.56	06:08:01	10	10	0.48	0.24
03:05:01	4	4	0.19	0.10	06:09	90	94	4.35	2.27
03:05:02	2	2	0.10	0.05	06:11:01	2	2	0.10	0.05
03:05:03	1	1	0.05	0.02	06:11:02	15	15	0.72	0.36
03:08	2	2	0.10	0.05	06:12	5	5	0.24	0.12
03:09	1	1	0.05	0.02	06:15	3	3	0.14	0.07
03:10	11	11	0.53	0.27	06:16	3	3	0.14	0.07
03:11	5	5	0.24	0.12	06:18	3	3	0.14	0.07
03:13	4	4	0.19	0.10	06:19	3	3	0.14	0.07
04:01:01	43	43	2.08	1.04	06:24	1	1	0.05	0.02
04:02	253	262	12.22	6.33	06:31	1	1	0.05	0.02
04:04	5	5	0.24	0.12	06:33	1	1	0.05	0.02
05:01:01	600	666	28.99	16.09	unidentified[b]	1	1	0.05	0.02
05:01:02	6	6	0.29	0.14					

Continued

HLA-DRB1	Phenotype Count (n = 1090)	Allele Count (2n = 2180)	Phenotype Frequency (%)	Allele Frequency (%)	HLA-DRB1	Phenotype Count (n = 1090)	Allele Count (2n = 2180)	Phenotype Frequency (%)	Allele Frequency (%)
01[a]	14	14	1.28	0.64	10:01:01	56	58	5.14	2.66
01:01:01	34	35	3.12	1.61	11:01:01	32	32	2.94	1.47
01:02:01	103	108	9.45	4.95	11:01:02	236	254	21.65	11.65
03/11/13/14[a]	56	70	5.23	3.21	11:02	145	154	13.30	7.06
03:01:01	143	150	13.12	6.88	11:04:01	21	23	1.93	1.06
03:02:01	133	140	12.20	6.42	11:04:02	7	7	0.64	0.32
03:05:01	3	3	0.28	0.14	12[a]	11	11	1.01	0.50
03:07	4	4	0.37	0.18	12:01:01	46	49	4.22	2.25
03:08	1	1	0.09	0.05	12:01:02	1	1	0.09	0.05
04[a]	11	11	1.01	0.50	12:02:01	1	1	0.09	0.05
04:01:01	12	12	1.10	0.55	13:01:01	108	110	9.91	5.05
04:04:01	7	7	0.64	0.32	13:01:02	2	2	0.18	0.09
04:05:01	25	26	2.29	1.19	13:02:01	218	235	20.00	10.78
04:08:01	3	3	0.28	0.14	13:03:01	47	48	4.31	2.20
04:10:01	5	5	0.46	0.23	13:03:02	3	3	0.28	0.14
07:01:01	164	165	15.05	7.57	13:37	1	1	0.09	0.05
08[a]	46	50	4.22	2.29	14:01:01	13	14	1.19	0.64
08:02:01	3	3	0.28	0.14	14:01:02	1	1	0.09	0.05
08:02:02	2	2	0.18	0.09	15/16[a]	17	19	1.56	0.87
08:04:01	89	92	8.17	4.22	15:01:01	2	2	0.18	0.09
08:04:03	1	1	0.09	0.05	15:03:01	209	223	19.17	10.23
08:06	1	1	0.09	0.05	16:02:01	4	4	0.37	0.18
09:01:02	24	25	2.20	1.15					
DRB3	(n = 881)[c]	(2n = 1260)[c]			DRB4	(n = 222)[c]	(2n = 229)[c]		
01:01:02G	225	234	24.35	18.57	01:01:01G	222	229	100	100
02:01:01G	2	2	0.22	0.16	untyped[a]	24	25	-	-
02:02:01G	571	690	61.80	54.76	DRB5	(n = 214)	(2n = 229)		
02:10	9	9	0.97	0.71	01:01:01	214	229	100	100
03:01:01G	297	325	32.14	25.79	untyped[a]	18	19	-	-
untyped[a]	43	59	-	-					

[a]High res not typed; [b]Cannot determine allele due to SNPs or SNP combinations in sequence that do not exist in database; [c]Sample sizes only included high-res typed individuals at that loci.

(1.17%). Nine unique class II haplotypes (DPA1-DPB1-DQA1-DQB1-DRB1) were found in more than 1% of the population (**Table 10**). These common class II haplotypes were 02:02:02-01:01:01-04:01:01G-04:02-03:02:01 (2.71%), 03:01-04:02:01-05:01:01-03:01:01-11:02 (2.42%), 01:03:01-02:01:02-01:02:01-06:02:01-15:03:01 (2.25%), 02:02:02-01:01:01-05:01:01-03:01:01-11:01:02 (2.02%), 01:03:01-04:01:01-01:02:01-06:02:01-15:03:01 (1.56%), 03:01-04:02:01-01:02:01-06:02:01G-11:01:02 (1.56%), 02:01:01-01:01:01-02:01-02:01:01-07:01:01 (1.21%), 03:01-04:02:01-02:01-02:01:01-07:01:01 (1.21%), and 03:01-04:02:01-01:02:01-06:02:01-15:03:01 (1.10%).

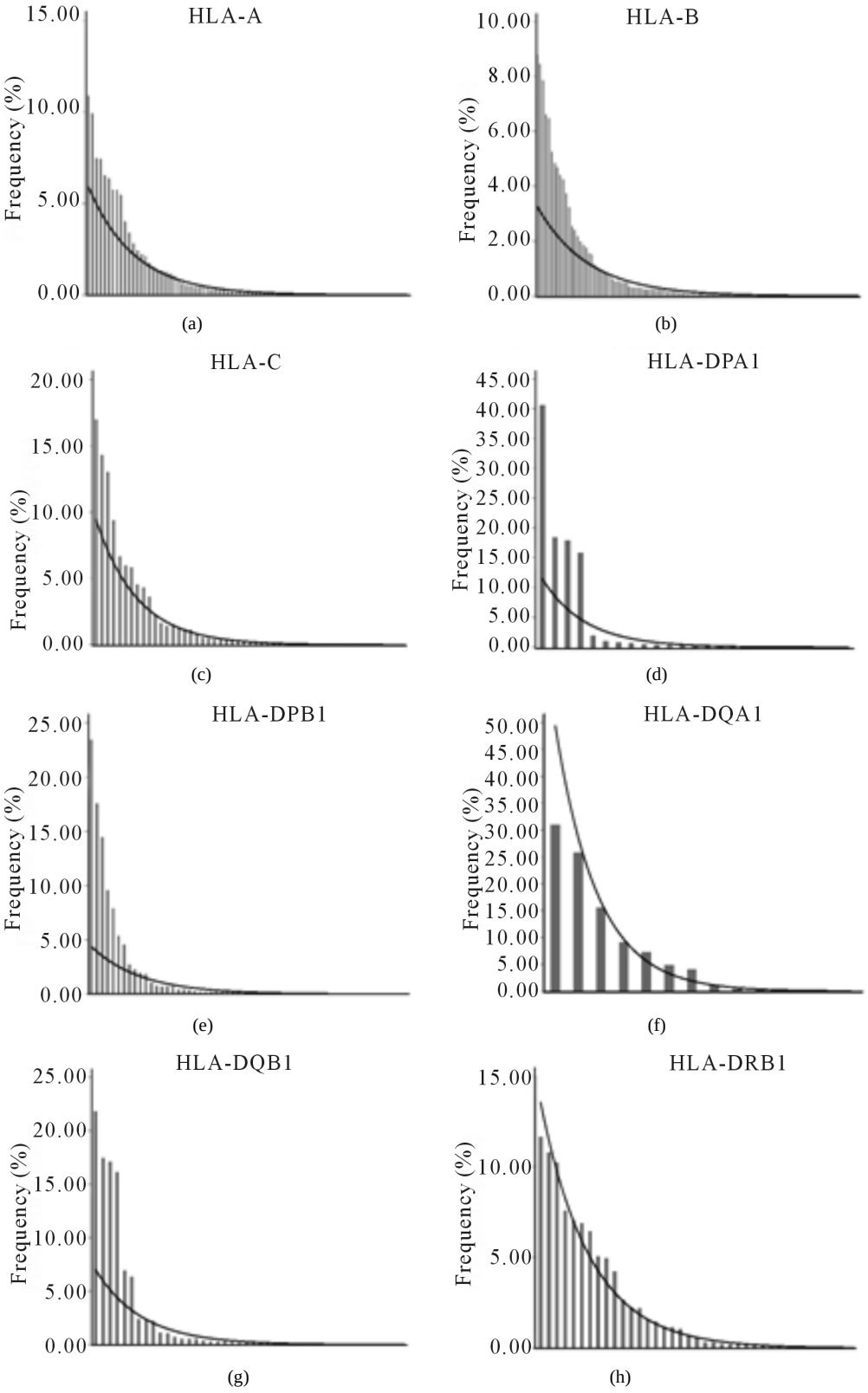

Figure 1. Distribution of HLA alleles arranged by highest to lowest frequency in the Pumwani sex worker cohort. (a) HLA-A. (b) HLA-B. (c) HLA-C. (d) HLA-DPA1. (e) HLA-DPB1. (f) HLA-DQA1. (g) HLA-DQB1. (h) HLA-DRB1.

Table 5. HLA Genotypes that deviated significantly from Hardy-Weinberg expectations in the Pumwani sex worker cohort.

(a)

Higher than expected counts		
HLA Loci	Genotype	P-value
A	74:01:01-36:01:01	0.003
A	30:01:01-24:02:01	0.003
A	24:02:01-02:02	0.010
A	23:01:01-03:01:01	0.012
A	74:01-74:01	0.007
B	45:01:01-44:03:01	0.037
B	57:03:01-07:02:01	0.003
C	03:04:02-02:10	0.008
C	04:01:01-03:04:01	0.007
C	17:01:01-04:07	0.020
C	18:01-07:04:01	0.020
C	18:01-16:01:01	0.012
DPA1	03:01-01:07	0.001
DPA1	01:03:01-01:03:01	0.004
DPA1	02:02:02-02:02:02	0.027
DPB1	03:01:01-30:01	0.003
DPB1	04:01:01-01:01:01	0.046
DQA1	01:01:01-01:01:01	0.008
DQA1	03:01:01-03:01:01	0.006
DQB1	05:01:01-05:01:01	0.036
DQB1	06:03:01-03:01:01	0.034
DQB1	06:04:01-02:01:01	0.045
DRB1	07:01:01-15:03:01	0.004

(b)

Lower than expected counts		
A	74:01:01-30:01:01	0.027
C	16:01:01-06:02:01	0.030
DPA1	02:02:02-01:03:01	0.024
DPB1	04:02:01-03:01:01	0.035
DQB1	03:01:01-02:01:01	0.048

3.6. Comparison of HLA Allele Frequencies of Pumwani Cohort with other World Populations

We compared the HLA class I and class II allele frequencies of this population with other populations in the world including Sub-Saharan Africa, western Africa, Europe and Asia [8] (**Table 11** and **Table 12**). As expected the HLA class I and class II allele frequency distribution of Pumwani cohort is very similar to that of other Sub-Saharan Africa and North Africa populations with the exception of lower frequencies of A * 23 (5.55% versus 11.21%) and DQA1 * 03 (4.79% versus 11.72%), and higher frequencies of DPB1 * 30 (2.26% versus 0.37%) and DRB1 * 11 (21.51% versus 15.89%). The frequencies of specific allele groups in Pumwani cohort are much higher than Asian and Caucasian populations, such as A * 29, A * 30, A * 36, A * 66, A * 68, A * 74, B*42, B * 46, B * 58, B * 81, C * 02, C * 06, C * 17, C * 18, DPA1 * 03, DPB1 * 01, DPB1 * 17, DPB1 * 18,

Table 6. Common HLA Two-locus Haplotypes in the Pumwani sex worker cohort.

Loci	Haplotype	Haplotype Frequency (%)	Count	Loci	Haplotype	Haplotype Frequency (%)	Count
A-B	30:01:01-42:01:01	3.02	129	C-DPA1	04:01:01-01:03:01	5.89	242
A-B	02:02-58:02	2.14	91	C-DPB1	06:02:01-04:02:01	4.53	191
A-B	36:01-53:01:01	2.12	91	C-DPB1	17:01:01-01:01:01	4.36	184
A-C	30:01:01-17:01:01	3.55	151	C-DPB1	04:01:01-01:01:01	2.91	123
A-C	02:02-06:02:01	2.75	117	C-DQA1	07:01:01-01:02:01	5.93	242
A-C	66:01:01-06:02:01	2.25	96	C-DQA1	04:01:01-01:02:01	5.89	240
A-DPA1	02:01:01-01:03:01	4.46	185	C-DQA1	06:02:01-05:01:01	4.84	198
A-DPA1	74:01:01-01:03:01	3.61	149	C-DQB1	04:01:01-06:02:01	4.54	182
A-DPA1	03:01:01-01:03:01	3.30	137	C-DQB1	06:02:01-03:01:01	4.43	177
A-DPB1	02:01:01-01:01:01	2.87	122	C-DQB1	02:10-03:01:01	3.57	143
A-DPB1	68:02:01-01:01:01	2.73	116	C-DRB1	17:01:01-03:02:01	4.83	95
A-DPB1	30:01:01-01:01:01	2.42	102	C-DRB1	06:02:01-11:02	3.82	76
A-DQA1	68:02:01-01:02:01	3.69	151	C-DRB1	07:01:01-13:02:01	3.44	68
A-DQA1	30:02:01-01:02:01	3.26	134	DPA1-DPB1	02:02:02-01:01:01	14.43	616
A-DQA1	68:02:01-05:01:01	3.18	130	DPA1-DPB1	01:03:01-02:0102	13.26	566
A-DQB1	02:01:01-03:01:01	3.72	149	DPA1-DPB1	03:01-04:02:01	12.47	532
A-DQB1	02:02-03:01:01	2.78	111	DPA1-DQA1	01:03:01-01:02:01	16.26	660
A-DQB1	74:01:01-03:01:01	2.53	101	DPA1-DQA1	01:03:01-05:01:01	9.49	385
A-DRB1	30:01:01-03:02:01	2.34	46	DPA1-DQA1	01:03:01-01:01:01	6.29	255
A-DRB1	02:01:01-11:01:02	2.04	40	DPA1-DQB1	01:03:01-06:02:01	8.64	350
A-DRB1	68:02:01-03:01:01	1.31	26	DPA1-DQB1	01:03:01-03:01:01	7.64	309
B-C	58:02-06:02:01	8.04	341	DPA1-DQB1	01:03:01-02:01:01	7.21	292
B-C	42:01:01-17:01:01	6.37	270	DPA1-DRB1	01:03:01-15:03:01	5.61	112
B-C	15:03:01-02:10	6.37	270	DPA1-DRB1	01:03:01-13:02:01	5.00	100
B-DPA1	53:01:01-01:03:01	3.25	135	DPA1-DRB1	02:02:02-03:02:01	4.00	80
B-DPA1	49:01:01-01:03:01	3.25	135	DPB1-DQA1	01:01:01-05:01:01	6.96	289
B-DPA1	15:03:01-01:03:01	3.10	128	DPB1-DQA1	02:0102-01:02:01	5.84	242
B-DPB1	42:01:01-01:01:01	3.89	165	DPB1-DQA1	04:02:01-05:01:01	5.34	222
B-DPB1	58:02-04:02:01G	3.21	136	DPB1-DQB1	04:02:01-03:01:01	5.94	243
B-DPB1	15:03:01-01:01:01	3.04	129	DPB1-DQB1	01:01:01-02:01:01	4.97	203
B-DQA1	15:03:01-05:01:01	4.00	165	DPB1-DQB1	01:01:01-03:01:01	4.77	195
B-DQA1	53:01:01-01:02:01	3.90	160	DQA1-DQB1	05:01:01-03:01:01	16.05	638
B-DQA1	42:01:01-04:01:01	3.85	158	DQA1-DQB1	01:02:01-06:02:01	14.88	591
B-DQB1	15:03:01-03:01:01	3.86	155	DQA1-DQB1	01:01:01-05:01:01	12.59	500
B-DQB1	42:01:01-04:02:01	3.38	136	DQA1-DRB1	01:02:01-13:02:01	9.97	187
B-DQB1	58:02-03:01:01	3.31	133	DQA1-DRB1	01:02:01-15:03:01	9.18	172
B-DRB1	42:01:01-03:02:01	4.68	92	DQA1-DRB1	02:01-07:01:01	5.97	112
B-DRB1	15:03:01-11:01:02	3.49	69	DQB1-DRB1	06:02:01-15:03:01	8.97	175
B-DRB1	58:02-11:02	3.09	61	DQB1-DRB1	02:01:01-07:01:01	6.75	132
C-DPA1	07:01:01-01:03:01	8.02	330	DQB1-DRB1	02:01:01-03:01:01	6.14	120
C-DPA1	06:02:01-01:03:01	6.64	273				

Table 7. Linkage Disequilibrium between HLA class I alleles in the Pumwani sex worker cohort.

HLA-A	HLA-B	HLA-C	p-value	HLA-A	HLA-B	HLA-C	p-value
30:01	42:01	17:01	9.63E−60	74:01	35:01	07:05	4.21E−12
30:01	42:01	06:02	1.56E−59	29:01	07:05	15:05	1.05E−11
36:01	53:01	04:01	1.63E−54	74:01	15:03	02:10	3.58E−11
36:01	53:01	17:01	1.63E−54	74:01	49:01	02:10	3.58E−11
66:01	58:02	06:02	2.82E−36	30:02	18:01	17:01	4.28E−11
66:01	58:02	07:02	2.82E−36	30:02	18:01	18:01	4.28E−11
01:01	15:67	18:01	1.22E−28	30:02	18:01	05:01	4.28E−11
01:01	37:01	18:01	1.22E−28	29:02	42:01	17:01	5.19E−11
01:01	81:01	18:01	1.22E−28	24:02	18:01	05:01	8.40E−11
32:01	81:01	08:04	7.80E−28	24:02	18:01	07:04	8.40E−11
68:02	07:02	03:04	4.62E−26	24:02	35:02	04:01	9.53E−11
68:02	15:10	03:04	4.62E−26	30:02	57:02	18:01	1.54E−10
01:01	81:01	08:04	1.69E−25	30:02	57:03	18:01	1.54E−10
01:09	44:15	04:07	6.68E−25	01:02	27:03	03:02	1.38E−09
68:02	15:10	07:02	2.63E−22	31:04	47:03	07:01	1.61E−09
68:02	15:10	07:01	2.63E−22	30:04	45:01	06:02	2.03E−09
02:02	58:02	06:02	1.61E−20	30:04	45:01	07:01	2.03E−09
02:02	58:02	07:02	1.61E−20	30:04	45:01	16:01	2.03E−09
02:14	18:03	16:04	2.10E−19	02:01	15:03	16:02	2.42E−09
02:14	18:03	04:01	2.10E−19	02:01	51:01	16:02	2.42E−09
30:02	18:01	07:04	1.52E−18	29:02	42:01	06:02	3.00E−09
30:02	45:01	07:04	1.52E−18	01:02	27:03	02:02	1.52E−08
30:02	57:02	07:04	1.52E−18	02:01	51:01	02:10	1.54E−08
30:02	57:03	07:04	1.52E−18	02:01	51:01	06:02	1.54E−08
30:02	18:01	16:01	3.15E−18	31:04	58:01	07:01	1.67E−08
30:02	45:01	16:01	3.15E−18	31:04	58:01	03:02	1.67E−08
30:02	57:02	16:01	3.15E−18	02:01	15:03	06:02	1.03E−07
30:02	57:03	16:01	3.15E−18	74:01	15:03	06:02	1.63E−07
32:01	81:01	18:01	5.29E−18	74:01	15:03	16:01	1.63E−07
34:02	44:03	04:01	1.98E−16	03:01	49:01	07:01	2.86E−07
34:02	44:03	14:03	1.98E−16	03:01	49:01	16:01	2.86E−07
30:02	45:01	17:01	9.11E−16	30:02	57:02	17:01	3.41E−07
30:02	45:01	18:01	9.11E−16	30:02	57:03	17:01	3.41E−07
30:02	45:01	06:02	9.11E−16	74:01	49:01	07:01	6.71E−07
30:02	45:01	07:01	9.11E−16	74:01	49:01	16:01	6.71E−07
02:01	15:03	16:01	5.83E−14	02:01	15:03	02:10	7.57E−07
02:01	51:01	16:01	5.83E−14	01:01	15:67	14:02	1.36E−06
33:01	41:02	17:01	1.90E−13	33:03	53:01	04:01	1.69E−06
30:10	13:02	06:02	7.47E−13	33:03	53:01	17:01	1.69E−06
68:02	07:02	07:02	2.19E−12	30:02	57:03	07:01	2.55E−06
74:01	35:01	02:10	4.21E−12	30:01	42:01	07:01	3.98E−06
74:01	35:01	04:01	4.21E−12				

Table 8. Linkage Disequilibrium between HLA Class II alleles in the Pumwani sex worker cohort.

HLA1	HLA2	HLA3	p-value	HLA1	HLA2	HLA3	p-value
DQA1 * 01:01	DQB1 * 05:01	DRB1 * 01:02	8.81E−235	DPA1 * 01:03	DPB1 * 30:01	DQA1 * 01:02	1.22E−14
DQA1 * 01:01	DQB1 * 05:01	DRB1 * 10:01	8.81E−235	DPA1 * 01:03	DQA1 * 01:02	DQB1 * 02:01	1.22E−14
DQA1 * 01:01	DQB1 * 05:01	DRB1 * 12:01	8.81E−235	DPA1 * 01:03	DQA1 * 01:02	DQB1 * 03:01	1.22E−14
DPA1 * 02:02	DPB1 * 01:01	DQA1 * 01:02	4.78E−208	DPA1 * 01:03	DQA1 * 01:02	DQB1 * 04:02	1.22E−14
DPA1 * 02:02	DPB1 * 01:01	DQA1 * 04:01	4.78E−208	DPA1 * 01:03	DQA1 * 01:02	DQB1 * 06:02	1.22E−14
DPA1 * 02:02	DPB1 * 01:01	DQB1 * 04:02	4.78E−208	DPA1 * 01:03	DQA1 * 01:02	DQB1 * 06:04	1.22E−14
DPA1 * 02:02	DPB1 * 01:01	DRB1 * 03:02	4.78E−208	DPA1 * 01:03	DQA1 * 01:02	DQB1 * 06:05	1.22E−14
DPA1 * 03:01	DPB1 * 04:02	DQB1 * 03:01	6.80E−178	DPA1 * 01:03	DQA1 * 01:02	DQB1 * 06:09	1.22E−14
DPA1 * 03:01	DPB1 * 04:02	DRB1 * 08:04	6.80E−178	DPA1 * 01:03	DQA1 * 01:02	DRB1 * 01:02	1.22E−14
DPA1 * 03:01	DPB1 * 04:02	DRB1 * 11:02	6.80E−178	DPA1 * 01:03	DQA1 * 01:02	DRB1 * 03:02	1.22E−14
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 03:01	1.12E−136	DPA1 * 01:03	DQA1 * 01:02	DRB1 * 07:01	1.22E−14
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 07:01	1.12E−136	DPA1 * 01:03	DQA1 * 01:02	DRB1 * 11:02	1.22E−14
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 08:04	1.12E−136	DPA1 * 01:03	DQA1 * 01:02	DRB1 * 13:02	1.22E−14
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 11:01	1.12E−136	DPA1 * 01:03	DQA1 * 01:02	DRB1 * 15:03	1.22E−14
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 11:02	1.12E−136	DPA1 * 03:01	DPB1 * 02:01	DRB1 * 11:02	1.10E−13
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 11:04	1.12E−136	DPA1 * 03:01	DPB1 * 40:01	DRB1 * 11:02	1.10E−13
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 13:01	1.12E−136	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 11:02	1.10E−13
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 13:02	1.12E−136	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 01:02	1.24E−13
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 13:03	1.12E−136	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 03:01	1.24E−13
DQA1 * 05:01	DQB1 * 03:01	DRB1 * 15:03	1.12E−136	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 07:01	1.24E−13
DQA1 * 01:02	DQB1 * 06:02	DRB1 * 01:02	1.40E−122	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 10:01	1.24E−13
DQA1 * 01:02	DQB1 * 06:02	DRB1 * 03:02	1.40E−122	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 12:01	1.24E−13
DQA1 * 01:02	DQB1 * 06:02	DRB1 * 07:01	1.40E−122	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 13:01	1.24E−13
DQA1 * 01:02	DQB1 * 06:02	DRB1 * 11:02	1.40E−122	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 13:02	1.24E−13
DQA1 * 01:02	DQB1 * 06:02	DRB1 * 13:02	1.40E−122	DQA1 * 05:01	DQB1 * 05:01	DRB1 * 15:03	1.24E−13
DQA1 * 01:02	DQB1 * 06:02	DRB1 * 15:03	1.40E−122	DQA1 * 05:01	DQB1 * 06:02	DRB1 * 03:01	1.27E−13
DQA1 * 04:01	DQB1 * 04:02	DRB1 * 03:02	4.97E−116	DQA1 * 05:01	DQB1 * 06:02	DRB1 * 07:01	1.27E−13
DQA1 * 04:01	DQB1 * 04:02	DRB1 * 08:04	4.97E−116	DQA1 * 05:01	DQB1 * 06:02	DRB1 * 13:01	1.27E−13
DQA1 * 02:01	DQB1 * 02:01	DRB1 * 03:01	2.93E−108	DQA1 * 05:01	DQB1 * 06:02	DRB1 * 13:02	1.27E−13
DQA1 * 02:01	DQB1 * 02:01	DRB1 * 07:01	2.93E−108	DQA1 * 05:01	DQB1 * 06:02	DRB1 * 15:03	1.27E−13
DQA1 * 02:01	DQB1 * 02:01	DRB1 * 09:01	2.93E−108	DPA1 * 01:05	DPB1 * 04:01	DQB1 * 03:03	2.00E−13
DQA1 * 02:01	DQB1 * 02:01	DRB1 * 11:01	2.93E−108	DQA1 * 04:01	DQB1 * 04:01	DRB1 * 08:04	2.14E−13
DPA1 * 01:03	DPB1 * 02:01	DQA1 * 01:02	8.22E−80	DQA1 * 04:01	DQB1 * 05:01	DRB1 * 08:04	2.14E−13
DPA1 * 01:03	DPB1 * 02:01	DQA1 * 04:01	8.22E−80	DQA1 * 05:01	DQB1 * 02:01	DRB1 * 03:01	2.53E−13
DPA1 * 01:03	DPB1 * 02:01	DRB1 * 03:02	8.22E−80	DQA1 * 05:01	DQB1 * 02:01	DRB1 * 07:01	2.53E−13
DPA1 * 02:01	DPB1 * 17:01	DQA1 * 01:03	8.47E−72	DQA1 * 05:01	DQB1 * 02:01	DRB1 * 09:01	2.53E−13
DPA1 * 02:01	DPB1 * 17:01	DQA1 * 05:02	8.47E−72	DQA1 * 05:01	DQB1 * 02:01	DRB1 * 11:01	2.53E−13
DPA1 * 02:01	DPB1 * 17:01	DQB1 * 06:03	8.47E−72	DQA1 * 05:01	DQB1 * 02:01	DRB1 * 13:01	2.53E−13
DQA1 * 04:01	DQB1 * 04:01	DRB1 * 03:02	3.83E−56	DQA1 * 05:01	DQB1 * 02:01	DRB1 * 13:02	2.53E−13
DQA1 * 04:01	DQB1 * 05:01	DRB1 * 03:02	3.83E−56	DQA1 * 05:01	DQB1 * 02:01	DRB1 * 15:03	2.53E−13
DQA1 * 01:02	DQB1 * 06:04	DRB1 * 01:02	1.16E−45	DQA1 * 01:02	DQB1 * 02:01	DRB1 * 01:02	9.39E−13
DQA1 * 01:02	DQB1 * 06:04	DRB1 * 03:02	1.16E−45	DQA1 * 01:02	DQB1 * 02:01	DRB1 * 03:01	9.39E−13

Continued

DQA1 * 01:02	DQB1 * 06:04	DRB1 * 07:01	1.16E−45	DQA1 * 01:02	DQB1 * 02:01	DRB1 * 03:02	9.39E−13
DQA1 * 01:02	DQB1 * 06:04	DRB1 * 11:02	1.16E−45	DQA1 * 01:02	DQB1 * 02:01	DRB1 * 07:01	9.39E−13
DQA1 * 01:02	DQB1 * 06:04	DRB1 * 13:02	1.16E−45	DQA1 * 01:02	DQB1 * 02:01	DRB1 * 09:01	9.39E−13
DQA1 * 01:02	DQB1 * 06:04	DRB1 * 15:03	1.16E−45	DQA1 * 01:02	DQB1 * 02:01	DRB1 * 11:01	9.39E−13
DQA1 * 01:03	DQB1 * 06:03	DRB1 * 13:01	4.07E−44	DQA1 * 01:02	DQB1 * 02:01	DRB1 * 11:02	9.39E−13
DQA1 * 01:03	DQB1 * 06:08	DRB1 * 13:01	4.07E−44	DQA1 * 01:01	DQB1 * 02:01	DRB1 * 12:01	1.33E−12
DPA1 * 01:03	DPB1 * 04:01	DQA1 * 01:02	8.62E−39	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 12:01	1.33E−12
DPA1 * 01:03	DPB1 * 04:01	DQA1 * 04:01	8.62E−39	DQA1 * 01:01	DQB1 * 05:03	DRB1 * 12:01	1.33E−12
DPA1 * 01:03	DPB1 * 04:01	DRB1 * 03:02	8.62E−39	DQA1 * 01:01	DQB1 * 06:02	DRB1 * 12:01	1.33E−12
DPA1 * 04:01	DPB1 * 04:02	DQB1 * 03:01	6.33E−37	DPB1 * 30:01	DQA1 * 01:02	DQB1 * 02:01	8.06E−12
DPA1 * 04:01	DPB1 * 04:02	DRB1 * 11:02	6.33E−37	DPB1 * 30:01	DQA1 * 01:02	DQB1 * 03:01	8.06E−12
DQA1 * 03:01	DQB1 * 03:02	DRB1 * 04:05	1.04E−34	DPB1 * 30:01	DQA1 * 01:02	DQB1 * 04:02	8.06E−12
DQA1 * 03:01	DQB1 * 03:02	DRB1 * 04:10	1.04E−34	DPB1 * 30:01	DQA1 * 01:02	DQB1 * 06:02	8.06E−12
DQA1 * 03:01	DQB1 * 03:02	DRB1 * 09:01	1.04E−34	DPB1 * 30:01	DQA1 * 01:02	DQB1 * 06:05	8.06E−12
DQA1 * 01:01	DQB1 * 02:01	DRB1 * 01:02	1.57E−34	DPB1 * 30:01	DQA1 * 01:02	DQB1 * 06:09	8.06E−12
DQA1 * 01:01	DQB1 * 03:01	DRB1 * 01:02	1.57E−34	DPB1 * 30:01	DQA1 * 01:02	DRB1 * 01:02	8.06E−12
DQA1 * 01:01	DQB1 * 05:03	DRB1 * 01:02	1.57E−34	DPB1 * 30:01	DQA1 * 01:02	DRB1 * 03:02	8.06E−12
DQA1 * 01:01	DQB1 * 06:02	DRB1 * 01:02	1.57E−34	DPB1 * 30:01	DQA1 * 01:02	DRB1 * 07:01	8.06E−12
DPA1 * 01:03	DPB1 * 01:01	DQA1 * 01:02	9.07E−33	DPB1 * 30:01	DQA1 * 01:02	DRB1 * 11:02	8.06E−12
DPA1 * 01:03	DPB1 * 01:01	DQA1 * 04:01	9.07E−33	DPB1 * 30:01	DQA1 * 01:02	DRB1 * 13:02	8.06E−12
DPA1 * 01:03	DPB1 * 01:01	DQB1 * 04:02	9.07E−33	DPB1 * 30:01	DQA1 * 01:02	DRB1 * 15:03	8.06E−12
DPA1 * 01:03	DPB1 * 01:01	DRB1 * 03:02	9.07E−33	DQA1 * 01:02	DQB1 * 04:02	DRB1 * 07:01	1.21E−11
DPA1 * 01:03	DPB1 * 18:01	DQA1 * 01:02	3.23E−31	DPA1 * 02:01	DPB1 * 02:01	DQA1 * 01:03	2.62E−11
DPA1 * 01:03	DPB1 * 18:01	DQA1 * 04:01	3.23E−31	DPA1 * 02:01	DPB1 * 02:01	DQA1 * 05:02	2.62E−11
DPA1 * 01:03	DPB1 * 18:01	DQB1 * 05:01	3.23E−31	DPA1 * 02:01	DPB1 * 02:01	DQB1 * 06:03	2.62E−11
DPA1 * 01:03	DPB1 * 18:01	DRB1 * 03:02	3.23E−31	DPB1 * 01:01	DQA1 * 01:02	DQB1 * 04:02	3.13E−11
DPA1 * 01:03	DPB1 * 18:01	DRB1 * 12:01	3.23E−31	DPA1 * 02:01	DPB1 * 04:02	DQA1 * 01:03	7.86E−11
DPA1 * 03:01	DPB1 * 55:01	DQB1 * 03:01	3.40E−28	DPA1 * 02:01	DPB1 * 04:02	DQA1 * 05:02	7.86E−11
DPA1 * 03:01	DPB1 * 55:01	DRB1 * 08:04	3.40E−28	DPA1 * 02:01	DPB1 * 04:02	DQB1 * 03:01	7.86E−11
DPA1 * 03:01	DPB1 * 55:01	DRB1 * 11:02	3.40E−28	DPA1 * 02:01	DPB1 * 04:02	DQB1 * 06:03	7.86E−11
DPA1 * 01:03	DPB1 * 03:01	DQA1 * 01:02	1.28E−27	DPA1 * 02:01	DPB1 * 04:02	DRB1 * 11:02	7.86E−11
DPA1 * 01:03	DPB1 * 03:01	DQA1 * 04:01	1.28E−27	DQA1 * 01:01	DQB1 * 05:03	DRB1 * 14:01	1.28E−10
DPA1 * 01:03	DPB1 * 03:01	DRB1 * 03:02	1.28E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 07:01	1.29E−10
DPB1 * 30:01	DQB1 * 06:04	DRB1 * 13:02	2.40E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 08:04	1.29E−10
DPB1 * 30:01	DQA1 * 01:02	DQB1 * 06:04	2.40E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 11:01	1.29E−10
DQA1 * 01:02	DQB1 * 02:01	DRB1 * 13:02	3.05E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 11:02	1.29E−10
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 13:02	3.05E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 11:04	1.29E−10
DQA1 * 01:02	DQB1 * 04:02	DRB1 * 13:02	3.05E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 13:02	1.29E−10
DQA1 * 01:02	DQB1 * 06:05	DRB1 * 13:02	3.05E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 13:03	1.29E−10
DQA1 * 01:02	DQB1 * 06:09	DRB1 * 13:02	3.05E−27	DQA1 * 01:01	DQB1 * 03:01	DRB1 * 15:03	1.29E−10
DPA1 * 02:01	DPB1 * 13:01	DQA1 * 01:03	9.85E−27	DPA1 * 01:03	DQA1 * 04:01	DQB1 * 04:01	4.61E−10
DPA1 * 02:01	DPB1 * 13:01	DQA1 * 05:02	9.85E−27	DPA1 * 01:03	DPB1 * 30:01	DQA1 * 04:01	4.61E−10
DPA1 * 02:01	DPB1 * 13:01	DQB1 * 06:03	9.85E−27	DPA1 * 01:03	DQA1 * 04:01	DQB1 * 04:02	4.61E−10

Continued

DPA1 * 02:01	DPB1 * 11:01	DQA1 * 01:03	3.60E−26	DPA1 * 01:03	DQA1 * 04:01	DQB1 * 05:01	4.61E−10
DPA1 * 02:01	DPB1 * 11:01	DQA1 * 05:02	3.60E−26	DPA1 * 01:03	DQA1 * 04:01	DRB1 * 03:02	4.61E−10
DPA1 * 02:01	DPB1 * 11:01	DQB1 * 06:03	3.60E−26	DPA1 * 01:03	DQA1 * 04:01	DRB1 * 08:04	4.61E−10
DPA1 * 02:02	DQB1 * 04:02	DRB1 * 03:02	7.36E−24	DPA1 * 02:01	DPB1 * 04:01	DQA1 * 01:03	6.26E−10
DPA1 * 02:02	DQB1 * 402	DRB1 * 3020	7.36E−24	DPA1 * 02:01	DPB1 * 04:01	DQA1 * 05:02	6.26E−10
DPA1 * 02:02	DPB1 * 02:01	DQB1 * 04:02	7.36E−24	DPA1 * 02:01	DPB1 * 04:01	DQB1 * 06:03	6.26E−10
DPA1 * 02:02	DPB1 * 03:01	DQB1 * 04:02	7.36E−24	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 11:02	1.07E−09
DPA1 * 02:02	DPB1 * 04:02	DQB1 * 04:02	7.36E−24	DQA1 * 01:02	DQB1 * 04:02	DRB1 * 11:02	2.91E−09
DPA1 * 02:02	DPB1 * 30:01	DQB1 * 04:02	7.36E−24	DQA1 * 01:01	DQB1 * 02:01	DRB1 * 03:01	3.88E−09
DPA1 * 02:02	DQA1 * 01:02	DQB1 * 04:02	7.36E−24	DQA1 * 01:01	DQB1 * 02:01	DRB1 * 07:01	3.88E−09
DPA1 * 02:02	DQA1 * 04:01	DQB1 * 04:02	7.36E−24	DQA1 * 01:01	DQB1 * 02:01	DRB1 * 09:01	3.88E−09
DQA1 * 01:02	DQB1 * 02:01	DRB1 * 15:03	2.90E−23	DQA1 * 01:01	DQB1 * 02:01	DRB1 * 11:01	3.88E−09
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 15:03	2.90E−23	DPA1 * 02:01	DPB1 * 19:01	DQA1 * 01:03	4.17E−09
DQA1 * 01:02	DQB1 * 04:02	DRB1 * 15:03	2.90E−23	DPA1 * 02:01	DPB1 * 19:01	DQA1 * 05:02	4.17E−09
DQA1 * 01:02	DQB1 * 06:05	DRB1 * 15:03	2.90E−23	DPA1 * 02:01	DPB1 * 19:01	DQB1 * 06:03	4.17E−09
DQA1 * 01:02	DQB1 * 06:09	DRB1 * 15:03	2.90E−23	DPA1 * 02:02	DPB1 * 30:01	DQA1 * 01:02	1.98E−08
DPA1 * 02:01	DPB1 * 01:01	DQA1 * 01:02	8.36E−22	DPA1 * 02:02	DPB1 * 30:01	DQB1 * 06:04	1.98E−08
DPA1 * 02:01	DPB1 * 01:01	DQA1 * 01:03	8.36E−22	DPA1 * 02:02	DPB1 * 30:01	DRB1 * 13:02	1.98E−08
DPA1 * 02:01	DPB1 * 01:01	DQA1 * 04:01	8.36E−22	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 07:01	3.61E−08
DPA1 * 02:01	DPB1 * 01:01	DQA1 * 05:02	8.36E−22	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 08:04	3.61E−08
DPA1 * 02:01	DPB1 * 01:01	DQB1 * 04:02	8.36E−22	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 11:01	3.61E−08
DPA1 * 02:01	DPB1 * 01:01	DQB1 * 06:03	8.36E−22	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 11:04	3.61E−08
DPA1 * 02:01	DPB1 * 01:01	DRB1 * 03:02	8.36E−22	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 13:02	3.61E−08
DPA1 * 02:02	DPB1 * 02:01	DRB1 * 03:02	2.51E−20	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 13:03	3.61E−08
DPA1 * 02:02	DPB1 * 03:01	DRB1 * 03:02	2.51E−20	DPB1 * 04:02	DQB1 * 03:01	DRB1 * 15:03	3.61E−08
DPA1 * 02:02	DPB1 * 04:02	DRB1 * 03:02	2.51E−20	DQA1 * 01:01	DQB1 * 06:02	DRB1 * 11:02	4.41E−08
DPA1 * 02:02	DPB1 * 30:01	DRB1 * 03:02	2.51E−20	DQA1 * 01:01	DQB1 * 06:02	DRB1 * 15:03	4.41E−08
DPA1 * 02:02	DQA1 * 01:02	DRB1 * 03:02	2.51E−20	DQA1 * 04:03	DQB1 * 04:02	DRB1 * 03:02	8.81E−08
DPA1 * 02:02	DQA1 * 04:01	DRB1 * 03:02	2.51E−20	DPB1 * 18:01	DQB1 * 05:01	DRB1 * 12:01	1.34E−07
DPA1 * 01:03	DPB1 * 04:02	DQA1 * 01:02	5.61E−20	DQA1 * 01:02	DQB1 * 04:02	DRB1 * 03:02	1.35E−07
DPA1 * 01:03	DPB1 * 04:02	DQA1 * 04:01	5.61E−20	DPA1 * 03:01	DPB1 * 02:01	DRB1 * 08:04	1.41E−07
DPA1 * 01:03	DPB1 * 04:02	DQB1 * 03:01	5.61E−20	DPA1 * 03:01	DPB1 * 40:01	DRB1 * 08:04	1.41E−07
DPA1 * 01:03	DPB1 * 04:02	DRB1 * 03:02	5.61E−20	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 08:04	1.41E−07
DPA1 * 01:03	DPB1 * 04:02	DRB1 * 11:02	5.61E−20	DPA1 * 02:01	DQA1 * 05:02	DQB1 * 06:03	1.87E−07
DQA1 * 01:02	DQB1 * 06:09	DRB1 * 01:02	1.13E−19	DPA1 * 02:02	DPB1 * 04:02	DQA1 * 01:02	2.39E−07
DQA1 * 01:02	DQB1 * 06:09	DRB1 * 03:02	1.13E−19	DPA1 * 02:02	DPB1 * 04:02	DQB1 * 03:01	2.39E−07
DQA1 * 01:02	DQB1 * 06:09	DRB1 * 07:01	1.13E−19	DPA1 * 02:02	DPB1 * 04:02	DRB1 * 11:02	2.39E−07
DQA1 * 01:02	DQB1 * 06:09	DRB1 * 11:02	1.13E−19	DQA1 * 01:02	DQB1 * 04:02	DRB1 * 01:02	2.60E−07
DPA1 * 02:02	DQA1 * 04:01	DQB1 * 04:01	2.14E−19	DPA1 * 01:03	DPB1 * 30:01	DQB1 * 06:04	2.77E−07
DPA1 * 02:02	DPB1 * 02:01	DQA1 * 04:01	2.14E−19	DPA1 * 01:03	DPB1 * 30:01	DRB1 * 03:02	2.77E−07
DPA1 * 02:02	DPB1 * 03:01	DQA1 * 04:01	2.14E−19	DPA1 * 01:03	DPB1 * 30:01	DRB1 * 13:02	2.77E−07
DPA1 * 02:02	DPB1 * 04:02	DQA1 * 04:01	2.14E−19	DPA1 * 03:01	DPB1 * 40:01	DQB1 * 03:01	3.75E−07
DPA1 * 02:02	DPB1 * 30:01	DQA1 * 04:01	2.14E−19	DPB1 * 18:01	DQB1 * 05:01	DRB1 * 01:02	4.37E−07

Continued

DPA1 * 02:02	DQA1 * 04:01	DQB1 * 05:01	2.14E−19	DPB1 * 18:01	DQB1 * 05:01	DRB1 * 10:01	4.37E−07
DPA1 * 02:02	DQA1 * 04:01	DRB1 * 08:04	2.14E−19	DPA1 * 02:01	DQB1 * 06:03	DRB1 * 13:01	4.96E−07
DQA1 * 03:01	DQB1 * 03:03	DRB1 * 04:05	2.73E−18	DPA1 * 02:01	DQA1 * 01:03	DQB1 * 06:03	4.96E−07
DPB1 * 19:01	DQA1 * 01:03	DQB1 * 06:03	3.89E−18	DPA1 * 02:01	DQA1 * 01:03	DQB1 * 06:08	5.23E−07
DPB1 * 19:01	DQA1 * 01:03	DQB1 * 06:08	3.89E−18	DPA1 * 02:01	DQA1 * 01:03	DRB1 * 13:01	5.23E−07
DPB1 * 19:01	DQA1 * 01:03	DRB1 * 13:01	3.89E−18	DPA1 * 03:01	DPB1 * 02:01	DQB1 * 03:01	5.52E−07
DPA1 * 03:01	DPB1 * 01:01	DQA1 * 01:02	6.39E−18	DPB1 * 01:01	DQA1 * 01:02	DQB1 * 02:01	7.21E−07
DPA1 * 03:01	DPB1 * 01:01	DQA1 * 04:01	6.39E−18	DPB1 * 01:01	DQA1 * 01:02	DQB1 * 03:01	7.21E−07
DPA1 * 03:01	DPB1 * 01:01	DQB1 * 03:01	6.39E−18	DPB1 * 01:01	DQA1 * 01:02	DQB1 * 06:02	7.21E−07
DPA1 * 03:01	DPB1 * 01:01	DQB1 * 04:02	6.39E−18	DPB1 * 01:01	DQA1 * 01:02	DQB1 * 06:04	7.21E−07
DPA1 * 03:01	DPB1 * 01:01	DRB1 * 03:02	6.39E−18	DPB1 * 01:01	DQA1 * 01:02	DQB1 * 06:05	7.21E−07
DPA1 * 03:01	DPB1 * 01:01	DRB1 * 08:04	6.39E−18	DPB1 * 01:01	DQA1 * 01:02	DQB1 * 06:09	7.21E−07
DPA1 * 03:01	DPB1 * 01:01	DRB1 * 11:02	6.39E−18	DPB1 * 01:01	DQA1 * 01:02	DRB1 * 01:02	7.21E−07
DPB1 * 01:01	DQA1 * 04:01	DQB1 * 04:01	8.21E−18	DPB1 * 01:01	DQA1 * 01:02	DRB1 * 07:01	7.21E−07
DPB1 * 01:01	DQA1 * 04:01	DQB1 * 04:02	8.21E−18	DPB1 * 01:01	DQA1 * 01:02	DRB1 * 11:02	7.21E−07
DPB1 * 01:01	DQA1 * 04:01	DQB1 * 05:01	8.21E−18	DPB1 * 01:01	DQA1 * 01:02	DRB1 * 13:02	7.21E−07
DPB1 * 01:01	DQA1 * 04:01	DRB1 * 03:02	8.21E−18	DPB1 * 01:01	DQA1 * 01:02	DRB1 * 15:03	7.21E−07
DPB1 * 01:01	DQA1 * 04:01	DRB1 * 08:04	8.21E−18	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 07:01	8.45E−07
DQA1 * 01:02	DQB1 * 06:05	DRB1 * 01:02	2.47E−17	DPA1 * 03:01	DQB1 * 3010	DRB1 * 1102	8.45E−07
DQA1 * 01:02	DQB1 * 06:05	DRB1 * 03:02	2.47E−17	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 11:01	8.45E−07
DQA1 * 01:02	DQB1 * 06:05	DRB1 * 07:01	2.47E−17	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 11:04	8.45E−07
DQA1 * 01:02	DQB1 * 06:05	DRB1 * 11:02	2.47E−17	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 13:02	8.45E−07
DQA1 * 01:01	DQB1 * 02:01	DRB1 * 10:01	3.96E−17	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 13:03	8.45E−07
DQA1 * 01:01	DQB1 * 03:01	DRB1 * 10:01	3.96E−17	DPA1 * 03:01	DQB1 * 03:01	DRB1 * 15:03	8.45E−07
DQA1 * 01:01	DQB1 * 05:03	DRB1 * 10:01	3.96E−17	DPA1 * 03:01	DQB1 * 3010	DRB1 * 8040	8.45E−07
DQA1 * 01:01	DQB1 * 06:02	DRB1 * 10:01	3.96E−17	DQA1 * 04:01	DQB1 * 05:01	DRB1 * 01:02	9.13E−07
DPB1 * 01:01	DQB1 * 04:02	DRB1 * 03:02	7.87E−17	DQA1 * 04:01	DQB1 * 05:01	DRB1 * 10:01	9.13E−07
DPB1 * 01:01	DQA1 * 01:02	DRB1 * 03:02	7.87E−17	DQA1 * 04:01	DQB1 * 05:01	DRB1 * 12:01	9.13E−07
DQA1 * 05:01	DQB1 * 02:01	DRB1 * 11:02	6.61E−16	DPA1 * 02:02	DPB1 * 03:01	DQA1 * 01:02	9.72E−07
DQA1 * 05:01	DQB1 * 05:01	DRB1 * 11:02	6.61E−16	DPA1 * 02:02	DQA1 * 01:02	DQB1 * 02:01	9.72E−07
DQA1 * 05:01	DQB1 * 06:02	DRB1 * 11:02	6.61E−16	DPA1 * 02:02	DQA1 * 01:02	DQB1 * 03:01	9.72E−07
DQA1 * 03:01	DQB1 * 03:03	DRB1 * 04:10	1.97E−15	DPA1 * 02:02	DQA1 * 01:02	DQB1 * 06:02	9.72E−07
DQA1 * 03:01	DQB1 * 03:03	DRB1 * 09:01	1.97E−15	DPA1 * 02:02	DQA1 * 01:02	DQB1 * 06:04	9.72E−07
DPA1 * 02:02	DPB1 * 02:01	DQA1 * 01:02	2.35E−15	DPA1 * 02:02	DQA1 * 01:02	DQB1 * 06:05	9.72E−07
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 01:02	9.91E−15	DPA1 * 02:02	DQA1 * 01:02	DQB1 * 06:09	9.72E−07
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 03:02	9.91E−15	DPA1 * 02:02	DQA1 * 01:02	DRB1 * 01:02	9.72E−07
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 07:01	9.91E−15	DPA1 * 02:02	DQA1 * 01:02	DRB1 * 07:01	9.72E−07
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 08:04	9.91E−15	DPA1 * 02:02	DQA1 * 01:02	DRB1 * 11:02	9.72E−07
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 11:01	9.91E−15	DPA1 * 02:02	DQA1 * 01:02	DRB1 * 13:02	9.72E−07
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 11:02	9.91E−15	DPA1 * 02:02	DQA1 * 01:02	DRB1 * 15:03	9.72E−07
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 11:04	9.91E−15	DPA1 * 01:05	DQB1 * 03:03	DRB1 * 04:10	1.02E−06
DQA1 * 01:02	DQB1 * 03:01	DRB1 * 13:03	9.91E−15				

Table 9. Frequencies of the most abundant HLA class I haplotypes in the Pumwani sex worker cohort.

HLA-A	HLA-B	HLA-C	Haplotype Frequency (%)	Count (2n = 4206)
30:01:01	42:01:01	17:01:01	3.12	131
36:01	53:01:01	04:01:01	2.09	88
02:02	58:02	06:02:01	2.09	88
66:01:01	58:02	06:02:01	2.04	8
68:02:01	15:10:01	03:04:02	1.89	80
02:01:01	15:03:01	02:10	1.74	73
30:02:01	45:01:01	16:01:01	1.64	69
74:01:01	15:03:01	02:10	1.57	66
01:01:01	81:01:01	18:01:01	1.49	63
68:02:01	07:02:01	07:02:01	1.42	60
74:01:01	58:02	06:02:01	1.41	59
02:01:01	45:01:01	16:01:01	1.27	53
74:01:01	49:01:01	07:01:01	1.17	49

Table 10. Frequencies of the most abundant HLA class II haplotypes in the Pumwani sex worker cohort.

HLA-DPA1	HLA-DPB1	HLA-DQA1	HLA-DQB1	HLA-DRB1	Haplotype Frequency (%)	Count (2n = 1734)
02:02:02	01:01:01	04:01:01	04:02	03:02:01	2.71	47
03:01	04:02:01	05:01:01	03:01:01	11:02	2.42	42
01:03:01	02:01:02	01:02:01	06:02:01	15:03:01	2.25	39
02:02:02	01:01:01	05:01:01	03:01:01	11:01:02	2.02	35
01:03:01	04:01:01	01:02:01	06:02:01	15:03:01	1.56	27
03:01	04:02:01	01:02:01	06:02:01	11:01:02	1.56	27
03:01	04:02:01	02:01	02:01:01	07:01:01	1.21	21
02:01:01	01:01:01	02:01	02:01:01	07:01:01	1.21	21
03:01	04:02:01	01:02:01	06:02:01	15:03:01	1.10	19

DPB1 * 30, and DQA1 * 04. Whereas, the frequencies of B * 51, C * 01, C * 12, DQA1 * 03, DRB1 * 04 and DRB1 * 07 are much lower in Pumwani cohort than Asian and Caucasian populations. B * 38, B * 46, B * 54 and B * 55 were not detected in the more than 2000 women genotyped.

4. Discussion

Analysis of HLA class I and II allele distributions in different populations is an important part of vaccine, anthropological and disease association studies [1]-[3] [5]-[7] [9]-[12]. While many studies have been conducted in a diversity of world populations, few studies have analyzed both class I and class II alleles at the high resolution and few have studied population size as large as ours in this study. The identification of 245 HLA class I alleles and 184 class II alleles in this population, further demonstrates the genetic diversity in this East African population. The HLA class I and class II allele frequency distribution of the East African population will no doubt play an important role in shaping pathogen diversity and influencing pathogen evolution through host-pathogen interactions.

In this study population, homozygosity at the HLA-B -DQA1 and -DPB1 loci was significantly higher than expected. HLA association studies have shown that homozygosity for certain alleles/supertypes was associated with both better and worse clinical outcomes [10]. Frequencies of specific alleles such as A * 74:01, DPA1 * 01:03:01, DPA1 * 02:02:02, DQA1 * 01:01:01, DQA1 * 03:01:01, and DQB1 * 05:01:01 were higher than expected. The significant deviations from expected frequencies may suggest a potential benefit conferred to individuals homozygous for these alleles against common pathogens. For example, A * 74:01 is associated with

Table 11. Comparison of HLA class I (-A, -B, and -C) allele distributions between Pumwani cohort, and other world populations.

HLA-A	Pumwani Cohort (2n = 4322)	Sub-Saharan Africa (2n = 6098)	North Africa (2n = 2356)	Asia (2n = 224,854)	Western Europe Caucasoids (2n = 615,598)	HLA-A	Pumwani Cohort (2n = 4322)	Sub-Saharan Africa (2n = 6098)	North Africa (2n = 2356)	Asia (2n = 224854)	Western Europe Caucasoids (2n = 615,598)
01	8.31	6.42	10.55	3.23	13.42	31	1.43	1.21	2.02	3.19	2.55
02	21.31	19.51	22.04	29.31	26.84	32	1.37	1.66	4.48	1.15	4.45
03	5.95	5.09	7.32	3.07	11.89	33	1.81	4.67	4.20	10.10	0.12
11	0.05	0.66	3.61	20.42	6.07	34	2.41	3.16	1.55	0.34	-
23	**5.55**	**11.31**	6.18	0.35	2.73	36	2.82	2.09	0.57	<0.01	0.03
24	2.34	2.18	7.08	17.87	10.73	66	3.66	3.69	1.40	0.08	0.28
26	0.85	2.04	2.19	3.14	4.02	68	11.24	9.32	7.02	0.91	4.10
29	5.95	5.82	4.07	0.85	4.13	74	7.75	5.60	0.54	0.15	0.07
30	16.66	14.21	11.04	5.26	3.81						
HLA-B	Pumwani Cohort (2n = 4358)	Sub-Saharan Africa (2n = 6580)	North Africa (2n = 2356)	Asia (2n = 225,578)	Western Europe Caucasoids (2n = 613,024)	HLA-B	Pumwani Cohort (2n = 4358)	Sub-Saharan Africa (2n = 6580)	North Africa (2n = 2356)	Asia (2n = 225,578)	Western Europe Caucasoids (2n = 613,024)
07	5.00	5.53	5.21	3.43	7.98	45	6.95	6.09	3.87	0.17	0.74
08	2.66	4.04	5.04	1.19	8.00	46	-	-	-	9.04	0.02
13	2.39	1.60	1.61	9.75	2.60	48	0.25	0.11	0.39	2.08	0.04
14	2.87	3.53	3.39	0.44	4.39	49	5.30	3.32	4.25	0.21	3.01
15	15.63	13.43	6.82	12.93	5.64	50	0.14	0.49	7.14	0.47	1.92
18	4.22	3.95	4.03	1.54	7.32	51	1.97	3.19	6.95	6.77	8.91
27	0.76	1.11	2.17	2.94	2.73	52	0.02	0.91	2.15	2.83	1.09
35	3.10	6.54	6.71	5.31	12.87	53	7.94	9.31	3.03	0.07	0.83
38	-	0.09	2.81	2.90	2.70	54	-	0.02	<0.01	2.93	0.02
39	1.79	0.87	2.50	2.40	2.22	55	-	0.08	1.38	2.37	2.10
40	0.46	1.22	2.55	13.61	3.89	57	5.76	3.92	2.90	1.32	4.16
41	1.45	1.86	5.00	0.27	1.12	58	13.74	11.74	3.55	6.38	0.63
42	6.72	5.75	1.51	0.02	0.08	81	4.41	2.17	0.08	0.07	0.01
44	4.04	5.03	8.55	4.49	12.20						
HLA-C	Pumwani Cohort (2n = 4304)	Sub-Saharan Africa (2n = 4904)	North Africa (2n = 1372)	Asia (2n = 16,282)	Western Europe Caucasoids (2n = 135,416)	HLA-C	Pumwani Cohort (2n = 4304)	Sub-Saharan Africa (2n = 4904)	North Africa (2n = 1372)	Asia (2n = 16282)	Western Europe Caucasoids (2n = 135,416)
01	0.14	0.65	0.53	8.32	3.33	08	5.02	5.19	3.53	5.03	3.85
02	8.27	8.51	5.89	2.53	4.01	12	1.46	1.87	7.79	6.76	3.65
03	6.62	8.03	2.64	13.17	12.97	14	1.18	1.76	1.00	3.06	1.09
04	14.92	16.68	13.71	10.94	8.87	15	1.67	1.50	4.18	3.41	2.28
05	0.37	1.34	3.28	1.56	10.75	16	6.74	8.92	6.91	0.87	4.60
06	17.05	14.09	13.84	6.51	8.89	17	9.39	8.20	4.78	1.10	0.64
07	21.28	20.98	20.59	16.50	34.18	18	5.81	2.91	0.48	0.03	0.07

-Alleles with frequencies below 2.00% were not included.

slower disease progression to AIDS in HIV infected individuals [13], DPA1 * 01:03:01 is associated with slower seroconversion in this population [14].

The frequencies of multiple class I (A-B-C) alleles in this cohort were higher than 1% and majority of them were unique to Sub-Saharan populations including Kenyan Luo and Kenyan Nandi populations. Haplotype A * 30:01-B * 42:01-C * 17:01 has also been identified in a sub-Saharan population [15] and a North African population from Morocco [16]. The existence of similar haplotypes between different ethnic populations suggests an ancestral linkage.

Comparing allele frequencies between different world populations showed similarities among African

Table 12. Comparison of HLA class II (-DPA1, -DPB1, -DQA1, -DQB1, and -DRB1) allele distributions between Pumwani Cohort, and other world populations.

HLA-DPA1	Pumwani Cohort (2n = 4320)	Sub-Saharan Africa (2n = 1938)	Gambia (2n = 292)	Asia (2n = 2700)	Western Europe Caucasoids (2n = 1816)	HLA-DPA1	Pumwani Cohort (2n = 4320)	Sub-Saharan Africa (2n = 1938)	Gambia (2n = 292)	Asia (2n = 2700)	Western Europe Caucasoids (2n = 1816)
01	43.59	29.37	21.50	60.91	82.11	03	18.61	17.58	8.20	0.03	0.16
02	35.79	51.67	70.10	37.71	17.36	04	1.92	1.48	-	0.15	0.06
HLA-DPB1	Pumwani Cohort (2n = 4430)	Sub-Saharan Africa (2n = 3098)	North Africa (2n = 794)	Asia (2n = 16106)	Western Europe Caucasoids (2n = 1816)	HLA-DPB1	Pumwani Cohort (2n = 4430)	Sub-Saharan Africa (2n = 3098)	North Africa (2n = 794)	Asia (2n = 16354)	Western Europe Caucasoids (2n = 13610)
01	24.49	28.96	6.11	0.85	3.41	11	1.90	1.27	2.00	0.06	1.57
02	14.58	11.34	26.72	22.71	15.79	13	2.73	3.97	1.73	6.22	1.42
03	7.86	5.17	10.26	4.46	17.08	14	0.16	0.36	1.28	2.55	1.35
04	27.22	23.75	34.85	18.64	47.80	15	0.72	0.69	2.51	0.15	0.51
05	0.14	-	0.25	32.23	1.85	17	4.56	7.72	8.31	1.01	1.18
06	0.02	0.13	0.49	0.25	1.23	18	5.35	5.61	0.28	<0.01	0.02
09	0.38	0.03	0.87	3.74	1.16	30	2.26	0.37	1.51	<0.01	<0.01
HLA-DQA1	Pumwani Cohort (2n = 4196)	Sub-Saharan Africa (2n = 2806)	North Africa (2n = 996)	Asia (2n = 18574)	Western Europe Caucasoids (2n = 1816)	HLA-DQA1	Pumwani Cohort (2n = 4196)	Sub-Saharan Africa (2n = 2806)	North Africa (2n = 996)	Asia (2n = 18230)	Western Europe Caucasoids (2n = 12914)
01	50.86	52.36	30.62	41.07	39.85	04	9.37	7.02	2.67	3.49	3.11
02	7.22	7.91	16.02	10.20	14.44	05	27.31	20.71	28.38	17.69	26.73
03	4.79	11.72	15.44	26.01	15.79	06	0.45	0.04	0.42	3.23	0.69
HLA-DQB1	Pumwani Cohort (2n = 4140)	Sub-Saharan Africa (2n = 5070)	North Africa (2n = 2874)	Asia (2n = 35480)	Western Europe Caucasoids (2n = 112432)	HLA-DQB1	Pumwani Cohort (2n = 4140)	Sub-Saharan Africa (2n = 5070)	North Africa (2n = 2874)	Asia (2n = 33678)	Western Europe Caucasoids (2n = 1816)
02	17.42	17.79	29.88	14.04	24.29	05	17.03	19.42	14.55	21.92	13.73
03	25.53	21.65	30.02	37.35	31.27	06	32.51	35.57	20.12	21.87	20.79
04	7.49	5.47	3.36	6.33	2.08						
HLA-DRB1	Pumwani Cohort (2n = 2180)	Sub-Saharan Africa (2n = 5000)	North Africa (2n = 3484)	Asia (2n = 260724)	Western Europe Caucasoids (2n = 666392)	HLA-DRB1	Pumwani Cohort (2n = 2180)	Sub-Saharan Africa (2n = 5000)	North Africa (2n = 3484)	Asia (2n = 260724)	Western Europe Caucasoids (2n = 666392)
01	7.20	6.78	5.82	2.94	10.11	11	21.51	15.89	12.01	6.54	16.31
03	13.44	11.85	15.30	4.35	11.31	12	2.84	3.75	0.90	12.37	0.78
04	2.94	3.76	13.50	13.04	11.78	13	18.07	22.03	14.91	6.20	13.89
07	7.57	7.77	14.85	9.14	13.42	14	0.69	1.31	1.36	6.63	1.41
08	6.83	4.37	4.31	7.10	3.17	15	10.32	14.93	10.52	14.35	12.08
09	1.15	1.88	0.69	13.12	0.86	16	0.18	0.85	1.52	2.70	0.95
10	2.66	3.98	3.56	1.60	1.37						

Alleles with frequencies below 2.00% were not included.

populations, and significant difference between African population and other world populations, such as Asian and European populations. While most allele groups were found in every population, there were a few that were unique to specific regions such as B * 46 and B * 54 which was frequent in Asian populations but was rare in western Caucasoid Europeans and African populations, while A * 74 was highly prevalent in sub-Saharan Africa but was rare in other populations. As expected, similar frequencies were observed between this east African population and other sub-Saharan populations (http://www.allelefrequencies.net/). There were a few differences, such as A * 23, DQA1 * 03, DPB1 * 30 and DRB1 * 11.

Genetic diversity at the HLA loci enables this east African population to deal with the great number of infectious pathogens [17] at the population level. The large number of alleles and the new alleles identified in this population [5] [18]-[25] could reflect the founder population under the intensive balanced selective pressure by infectious pathogens. The insignificant deviation from Ewens-Watterson homozygosity test of neutrality could be due to the large number of alleles identified in this population. Furthermore, a recent paper showed that negative

frequency-dependent selection can limit the utility of the EW test in detecting selection acting on the HLA genes [26].

Acknowledgements

We would like to thank Chris Czarnecki, Janis Huntington, Kulvinder Gill, Binhua Liang, Erin Knight, Kimberly, Summer Pham, Cara Weessies, Felicia Foster for their help in HLA typing. We would also like to thank Gary Van Domselaar for the HLA allele format conversion program.

References

[1] Azurdia, R.M., Luzzi, G.A., Byren, I., *et al.* (1999) Lichen Sclerosus in Adult Men: A Study of HLA Associations and Susceptibility to Autoimmune Disease. *British Journal of Dermatology*, **140**, 79-83. http://dx.doi.org/10.1046/j.1365-2133.1999.02611.x

[2] Gladman, D.D., Anhorn, K.A., Schachter, R.K. and Mervart, H. (1986) HLA Antigens in Psoriatic Arthritis. *The Journal of Rheumatology*, **13**, 586-592.

[3] Rohowsky-Kochan, C., Skurnick, J., Molinaro, D. and Louria, D. (1998) HLA Antigens Associated with Susceptibility/Resistance to HIV-1 Infection. *Human Immunology*, **59**, 802-815. http://dx.doi.org/10.1016/S0198-8859(98)00086-X

[4] Tishkoff, S.A., Reed, F.A., Friedlaender, F.R., *et al.* (2009) The Genetic Structure and History of Africans and African Americans. *Science*, **324**, 1035-1044. http://dx.doi.org/10.1126/science.1172257

[5] Luo, M., Embree, J., Ramdahin, S., *et al.* (2002) HLA-A and HLA-B in Kenya, Africa: Allele Frequencies and Identification of HLA-B*1567 and HLA-B*4426. *Tissue Antigens*, **59**, 370-380. http://dx.doi.org/10.1034/j.1399-0039.2002.590503.x

[6] Sanchez-Mazas, A., Steiner, Q.G., Grundschober, C. and Tiercy, J.M. (2000) The Molecular Determination of HLA-Cw Alleles in the Mandenka (West Africa) Reveals a Close Genetic Relationship between Africans and Europeans. *Tissue Antigens*, **56**, 303-312. http://dx.doi.org/10.1034/j.1399-0039.2000.560402.x

[7] Luo, M., Blanchard, J., Pan, Y., Brunham, K. and Brunham, R.C. (1999) High-Resolution Sequence Typing of HLA-DQA1 and -DQB1 Exon 2 DNA with Taxonomy-Based Sequence Analysis (TBSA) Allele Assignment. *Tissue Antigens*, **54**, 69-82. http://dx.doi.org/10.1034/j.1399-0039.1999.540108.x

[8] Middleton, D., Menchaca, L., Rood, H. and Komerofsky, R. (2003) New Allele Frequency Database: http://www.allelefrequencies.net. *Tissue Antigens*, **61**, 403-407. http://dx.doi.org/10.1034/j.1399-0039.2003.00062.x

[9] Luo, M., Blanchard, J., Brunham, K., *et al.* (2001) Two-Step High Resolution Sequence-Based HLA-DRB Typing of Exon 2 DNA with Taxonomy-Based Sequence Analysis Allele Assignment. *Human Immunology*, **62**, 1294-310. http://dx.doi.org/10.1016/S0198-8859(01)00339-1

[10] Flores-Villanueva, P.O., Yunis, E.J., Delgado, J.C., *et al.* (2001) Control of HIV-1 Viremia and Protection from AIDS Are Associated with HLA-Bw4 Homozygosity. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 5140-5145. http://dx.doi.org/10.1073/pnas.071548198

[11] Hill, A.V., Allsopp, C.E., Kwiatkowski, D., *et al.* (1991) Common West African HLA Antigens Are Associated with Protection from Severe Malaria. *Nature*, **352**, 595-600. http://dx.doi.org/10.1038/352595a0

[12] Kijak, G.H., Walsh, A.M., Koehler, R.N., *et al.* (2009) HLA Class I Allele and Haplotype Diversity in Ugandans Supports the Presence of a Major East African Genetic Cluster. *Tissue Antigens*, **73**, 262-269. http://dx.doi.org/10.1111/j.1399-0039.2008.01192.x

[13] Peterson, T.A., Kimani, J., Wachihi, C., *et al.* (2013) HLA Class I Associations with Rates of HIV-1 Seroconversion and Disease Progression in the Pumwani Sex Worker Cohort. *Tissue Antigens*, **81**, 93-107. http://dx.doi.org/10.1111/tan.12051

[14] Hardie, R.A., Knight, E., Bruneau, B., *et al.* (2008) A Common Human Leucocyte Antigen-DP Genotype Is Associated with Resistance to HIV-1 Infection in Kenyan Sex Workers. *AIDS*, **22**, 2038-2042. http://dx.doi.org/10.1097/QAD.0b013e328311d1a0

[15] Cao, K., Moormann, A.M., Lyke, K.E., *et al.* (2004) Differentiation between African Populations Is Evidenced by the Diversity of Alleles and Haplotypes of HLA Class I Loci. *Tissue Antigens*, **63**, 293-325. http://dx.doi.org/10.1111/j.0001-2815.2004.00192.x

[16] Canossi, A., Piancatelli, D., Aureli, A., *et al.* (2010) Correlation between Genetic HLA Class I and II Polymorphisms and Anthropological Aspects in the Chaouya Population from Morocco (Arabic Speaking). *Tissue Antigens*, **76**, 177-193. http://dx.doi.org/10.1111/j.1399-0039.2010.01498.x

[17] Dunn, R.R., Davies, T.J., Harris, N.C. and Gavin, M.C. (2010) Global Drivers of Human Pathogen Richness and Prevalence. *Proceedings of the Royal Society*: *Biological Sciences*, **277**, 2587-2595. http://dx.doi.org/10.1098/rspb.2010.0340

[18] Isabey, E.P., Pham, H.M., Peterson, T.A., Brunham, R.C., Luo, M. and Plummer, F.A. (2011) Identification of a Novel HLA-A Allele, A*29:28, in an East African Population. *Tissue Antigens*, **78**, 451-452. http://dx.doi.org/10.1111/j.1399-0039.2011.01741.x

[19] Luo, M., Blanchard, J., Maclean, I. and Brunham, R. (1999) Identification of a Novel HLA-DQA1 Allele (DQA1*0106) by Sequence-Based DQA1 Typing. *Tissue Antigens*, **53**, 595-596. http://dx.doi.org/10.1034/j.1399-0039.1999.530613.x

[20] Luo, M., Blanchard, J., Maclean, I. and Brunham, R. (1999) Identification of a Novel DQB1 Allele DBQ1*0616. *Tissue Antigens*, **53**, 381-382. http://dx.doi.org/10.1034/j.1399-0039.1999.530410.x

[21] Luo, M., Blanchard, J., Maclean, I. and Brunham, R. (2001) Identification of a Novel DRB1 Allele, DRB1*1112, by Sequence-Based DRB Typing. *Tissue Antigens*, **57**, 85-86. http://dx.doi.org/10.1034/j.1399-0039.2001.057001085.x

[22] Luo, M., Cohen, C.R., Narayansingh, M.J., *et al.* (2004) Identification of a Novel HLA-DQA1 Null Allele, DQA1*0403N, from an East African Woman. *Tissue Antigens*, **63**, 609-611. http://dx.doi.org/10.1111/j.0001-2815.2004.00231.x

[23] Luo, M., Mao, X. and Plummer, F.A. (2005) Identification of Four Novel HLA-B Alleles, B*1590, B*1591, B*2726, and B*4705, from an East African Population by High-Resolution Sequence-Based Typing. *Tissue Antigens*, **65**, 187-191. http://dx.doi.org/10.1111/j.1399-0039.2005.00338.x

[24] Peterson, T.A., Luo, M., Mao, X., Brunham, R.C. and Plummer, F.A. (2008) Identification of a Novel DPA1 Allele, DPA1*010602, in an East African Population. *Human Immunology*, **69**, 885-886. http://dx.doi.org/10.1016/j.humimm.2008.09.001

[25] Turk, W.J., Mao, X.J., Bielawny, T., Brunham, R., Luo, M. and Plummer, F.A. (2008) Identification of Novel Human Leukocyte Antigen G Alleles in an East African Population by High-Resolution Sequence-Based Typing. *Human Immunology*, **69**, 426-429. http://dx.doi.org/10.1016/j.humimm.2008.05.007

[26] Ejsmond, M.J., Babik, W. and Radwan, J. (2010) MHC Allele Frequency Distributions under Parasite-Driven Selection: A Simulation Model. *BMC Evolutionary Biology*, **10**, 332. http://dx.doi.org/10.1186/1471-2148-10-332