Published in final edited form as:

Hum Immunol. 2018 December ; 79(12): 839–847. doi:10.1016/j.humimm.2018.09.004.

Allelic resolution NGS HLA typing of Class I and Class II loci and haplotypes in Cape Town, South Africa

Yvonne R. Thorstenson^{a,*}, Lisa E. Creary^b, Huang Huang^c, Virginie Rozot^d, Tracy T. Nguyen^a, Farbod Babrzadeh^a, Sandeep Kancharla^a, Marilyn Fukushima^a, Raquel Kuehn^a, Chunlin Wang^a, Ming Li^a, Sujatha Krishnakumar^a, Michael Mindrinos^a, Marcelo A. Fernandez Viña^b, Thomas J. Scriba^d, Mark M. Davis^e

^aImmucor, Sirona Genomics, Mountain View, CA, USA

^bDepartment of Pathology, Stanford University School of Medicine, Stanford, CA, USA

^cInstitute for Immunity, Transplantation and Infection, Stanford University, Stanford, CA, USA

^dSouth African Tuberculosis Vaccine Initiative, Institute of Infectious Disease and Molecular Medicine, Division of Immunology, Department of Pathology, University of Cape Town, Cape Town, South Africa

^eStanford University and Howard Hughes Medical Institute, Stanford, CA

Abstract

The development of next-generation sequencing (NGS) methods for HLA genotyping has already had an impact on the scope and precision of HLA research. In this study, allelic resolution HLA typing was obtained for 402 individuals from Cape Town, South Africa. The data were produced by high-throughput NGS sequencing as part of a study of T-cell responses to *Mycobacterium tuberculosis* in collaboration with the University of Cape Town and Stanford University. All samples were genotyped for 11 HLA loci, namely HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1, -DRB3, -DRB4, and -DRB5. NGS HLA typing of samples from Cape Town inhabitants revealed a unique cohort, including unusual haplotypes, and 22 novel alleles not previously reported in the IPD-IMGT/HLA Database. Eight novel alleles were in Class I loci and 14 were in Class II. There were 62 different alleles of HLA-A, 72 of HLA-B, and 47 of HLA-C. Alleles A*23:17, A*43:01, A*29:11, A*68:27:01, A*01:23, B*14:01:01, B*15:10:01, B*39:10:01, B*45:07, B*82:02:01 and C*08:04:01 were notably more frequent in Cape Town compared to other populations reported in the literature. Class II loci had 21 different alleles of DPA1, 46 of DPB1, 27 of DQA1, 26 of DQB1, 41 of DRB1, 5 of DRB3, 4 of DRB4 and 6 of DRB5. The Cape Town cohort exhibited high degrees of HLA diversity and relatively high heterozygosity at most loci. Genetic distances between Cape Town and five other sub-Saharan African populations were also calculated and compared to European Americans.

*Corresponding author at: Immucor Inc, 1916A Old Middlefield Way, Mountain View, CA 94043, USA. yvonne@sironagenomics.com (Y.R. Thorstenson).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.humimm.2018.09.004>.

Keywords

HLA; NGS; Allele frequency; Haplotype frequency; Population; Linkage disequilibrium; Genetic distance

1. Introduction

Africans represent the most genetically diverse group of populations in the world, [1] but have not been as well studied with respect to their human leukocyte antigen (HLA) genotypes compared to the inhabitants of European countries. Thus it was of interest to apply NGS high-throughput HLA sequencing technology to characterize the HLA allele diversity in a cohort of 402 healthy individuals from the Cape Town region of South Africa, recruited as part of a study of T-cell responses to *Mycobacterium tuberculosis* [2]. Genetic admixture is expected in these subjects based on the 2011 Census data from the City of Cape Town indicating a population of 38.6% Black African, 42.4% Cape Mixed Ancestry (CMA or Coloured), 1.4% Indian or Asian, 15.7% White, and 1.9% Other (www.statssa.gov.za).

HLA typing methods have evolved over time to reduce typing ambiguities and improve haplotype predictions [3]. HLA matching of patients and donors at high or allelic resolution has been shown to have a positive impact on graft outcome, reducing morbidity and complications in bone marrow transplantation [4,5]. The current study used a new method, MIA FORA NGS, to produce allelic resolution, whole gene HLA typing in a high throughput format. HLA genotypes that were called by MIA FORA NGS software were validated in a clinical study performed at three different sites. Eleven HLA loci were tested in 206 subjects (3692 alleles) that were previously typed by Sanger-based typing (SBT) or single-strand oligonucleotide probing (SSO) methods. A comparison of MIA FORA NGS with SBT and SSO showed that the overall concordance was 99.74% (Krishnakumar, Wang, Li, Mindrinos, Saldanha and Dwivedi, personal communication.)

2. Methods

Venous blood was collected for PBMC isolation, QuantiFERON® TB Gold In-tube (Qiagen) (QFT) and a tuberculin skin test (TST) were administered. Cryopreserved PBMC from 402 healthy individuals were collected and shipped to Stanford University from Cape Town, South Africa. Genomic DNA was extracted as described previously [6]. The majority of subjects were of Cape Mixed Ancestry (CMA). As shown in Table 1, the breakdown of ethnicity was 62% CMA, 31% Black African, and 7% Caucasian.

This study was performed in accordance with approvals from the Human Research Ethics Committee of the University of Cape Town. All participants were adolescents, who provided written informed assent. Written informed consent was also provided by a parent or legal guardian.

To add context to the HLA typing results in this study, we compared them with three previously published datasets of African or African American subjects that were genotyped to high resolution. The US National Bone Marrow Donor Program registry (NMDP)

includes subjects from four broad ethnic classifications as defined by the US census, namely European American (EUR), Asian Pacific Islander (API), Hispanic (HIS), and African American (AFA.) The NMDP dataset included more than 16,000 African American individuals that were genotyped to the four-digit level of resolution for HLA-A, -B, -C, -DRB1, and -DQB1 [3,7]. The Cao et al. dataset included 852 unrelated individuals from five different African populations. The samples were typed for HLA-A, -B, and -C using sequence-specific oligonucleotide probe hybridization (SSOP) for exons 2 and 3 [8]. The Paximadis et al. dataset included 200 Black African and 102 Caucasian individuals that worked at the South African Electricity Supply Commission (ESKOM.) The samples were typed for HLA-A, -B, -C, and -sDRB1 by direct sequencing (SBT) of exons 2, 3, and 4 [9].

2.1. Sequencing and genotyping methods

Allelic resolution HLA typing was performed using MIA FORA NGS kit and analysis software (Immucor, Inc.). All samples were genotyped for 11 HLA loci, namely HLA-A, -B, -C, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1, -DRB3, -DRB4, and -DRB5. The coverage for HLA-A, -B, and -C included all exons and introns, at least 200 base pairs of the 5' UTR and 100–1100 base pairs of the 3' UTR; coverage for -DPA1, and -DQA1, included all exons and introns, at least 45 base pairs of the 3' UTR and 25–190 base pairs of the 3' UTR. Coverage for -DRB1 included all exons, introns 2–6, at least 440 base pairs of the 5' UTR, 12 base pairs of the 3' UTR, 275 base pairs of intron 1 adjacent to exon 1, and 210 base pairs of intron 1 adjacent to exon 2; coverage for -DRB3/4/5 included exons 2–6, introns 2–5, and 260 base pairs of intron 1 adjacent to exon 2; coverage for -DQB1 included exons 1–5 and introns 1–4; coverage for -DPB1 included exons 2–4 and introns 2–3. The gene coverage is diagrammed in Fig. 1.

The extensive coverage across HLA loci minimized the possibility for ambiguous genotypes. First, intron coverage permitted direct phasing of polymorphisms whenever they were spaced within the insert size of a library fragment. Second, the complete coding sequence was obtained for all loci, except DPB1 and DRB3/4/5, allowing for nonambiguous typing to the third field of resolution. Genotypes were called to the closest fourth field allele according to the MIA FORA NGS v3.0 software. No attempt was made to record novel intron variation, which occurred in 11.8% of the allele calls for Class I and 66.4% for Class II loci. The percentages include copy number variation in introns that cannot be resolved by sequencing. Alleles with novel intron variation were consolidated to the closest fourth field name. The reported genotypes are unambiguous to the third field in every case, except DPB1 where pairwise ambiguities were possible [10]. Specifically, the pair DPB1*03:01:01/DPB1*105:01 was observed five times and was ambiguous with the pair DPB1*124:01/DPB1*463:01; the pair DPB1*03:01:01/DPB1*04:01:01:01 was observed three times and was ambiguous with DPB1*124:01/DPB1*350:01. The output of the MIA FORA software included all ambiguous pairs.

Sample and sequencing library preparation were performed using a MIA FORA NGS LT24 library preparation kit (Immucor, Inc.). Each HLA gene was amplified from genomic DNA using long-range PCR. Following amplification, PCR products were quantified and balanced by gene locus, fragmented, end-repaired and ligated to index adaptors containing unique

barcodes. In contrast to other NGS methods, all loci for a sample were combined before barcoding. Then the barcoded samples were consolidated into a single sequencing library, size selected, and amplified 12 rounds to incorporate the P5 and P7 adaptor sequences needed for binding to the Illumina flow cell. NGS sequencing libraries were prepared in sets of 24 or 96 samples using a semi-automated protocol on a Biomek liquid handling workstation. Up to 384 samples were sequenced using the paired-end protocol on Illumina NextSeq and up to 24 samples on Illumina MiSeq.

HLA allele candidates were computed and the final HLA typing was called and reviewed using MIA FORA NGS FLEX version 3.0 analysis software. Raw NGS sequence reads were used as input to call genotypes for all 11 HLA loci with high allele resolution. Three orthogonal algorithms, including independent mapping and de novo assembly strategies, were employed to calculate a probability score and rank the genotype candidates, and to generate consensus sequences for individual alleles. Specifically, paired-end reads were competitively mapped to HLA loci based on known reference sequences in the IPD-IMGT/HLA Database version 3.24.0 [11]. A mismatch filter eliminated alignments with mismatches or gaps and a paired-end filter increased specificity by requiring both ends to map to a single HLA reference. Then minimum coverage was computed for each allele candidate and allele pairs were computed as described in Wang et al. [10]. Novel alleles were identified by variation in the coding sequence only.

Further specificity for allele calling was obtained by quantifying the continuity of the tiling pattern using an empirically derived metric called “central read.” When reads were mapped onto a correct reference sequence, they formed a continuous tiling pattern over the entire sequenced region. When reads were mapped onto a similar, but incorrect reference sequence, they formed a staggered tiling pattern at mismatched positions. To quantify this difference between the two alignment patterns, the minimum value of “central read” was calculated for any given point; this value was then used as part of the probability score for each allele candidate. Central reads were defined as mapped reads for which the ratio between the length of the left arm and that of the right arm related to a particular position is between 0.5 and 2.0 [10].

2.2. Statistical analyses

Allele frequencies were determined by direct counting. Allele frequencies at the G group level were tested for deviations of Hardy-Weinberg equilibrium (HWE) proportions at each locus using the exact test of Guo and Thompson [12], and by chi-square testing when expected counts were equal or greater than 5. Chi-square tests were applied to investigate overall most frequent genotypes, the rare binned genotypes (pooled set of genotypes that individually have expected counts < 5), all homozygotes, all heterozygotes, and most frequent heterozygotes for specific alleles. PyPop version 0.7.0 (<http://www.pypop.org>) was used to perform most of the population analyses, including allele frequencies, deviations from Hardy Weinberg equilibrium (HWE) [12] at each locus, linkage disequilibrium (LD) between pairs of loci, and 4 locus haplotypes [13].

Haplotype frequencies for 2, 3, and 4 loci were estimated using an iterative Expectation-Maximization (EM) algorithm [14,15]. Linkage disequilibrium (LD) between pairs of alleles

at different loci and overall ‘global’ LD between pairs of loci were calculated [16]. Overall LD was computed using two formulae normalized to range from 0 to 1; Hedricks D' statistic (weighs the contribution to LD of specific allele pairs by the product of their allele frequencies) [17] and Cramer’s V statistic also described as W_n [18] calculated a chi-squared statistic for deviations between observed and expected haplotype frequencies. The significance of the overall pair-wise LD was determined using the loglikelihood permutation test. Association between allele pairs was expressed as normalized LD values D'_{ij} which is the coefficient of LD (D) divided by the theoretical maximum of D (when $D = 0$) or the minimum value of D (if $D < 0$) for the particular alleles at a given locus.

Frequencies of extended haplotypes (encompassing more than four loci) were estimated using the BIGDAWG (Bridging Immuno Genomic Data Analysis Workflow Gaps) [19] v1.8 package implemented in R (R foundation for statistical computing, <http://www.r-project.org>).

Genetic distances between the Cape Town cohort and populations previously studied from sub-Saharan African, Mali, Uganda, Zimbabwe, Zulu, [8] African Americans and European Americans [7] were estimated using the formula described by Cavalli-Sforza and Bodmer [20,21]. Calculations were performed using MS Excel. Genetic distance comparisons were made using class I allele frequencies that were reduced to G groups to match the data available from previously tested populations.

3. Results

NGS HLA typing of Cape Town subjects revealed a mixed population, including novel alleles and haplotypes not previously reported in the IPD-IMGT/HLA Database [11]. A total of 357 unique alleles were identified in the Cape Town cohort. Eight novel alleles were identified in class I, and fourteen were identified in class II loci; 11 alleles were detected in two or more individuals (3 in Class I and 8 in Class II loci) and 11 were found in a single individual (5 in Class I and 6 in Class II loci). Sanger sequencing was used to confirm the 11 novel alleles found at least twice. Descriptions of the novel alleles are listed in Table 2. Four novel alleles were found in HLA-A, two in HLA-B, and two in HLA-C. DPA1 and DPB1 included five novel alleles each, DQA1 had three, and DQB1 had one. Relative to the nearest known alleles in the IMGT/HLA database, there were 24 codon changes in 22 novel alleles; 7 were synonymous (no amino acid change) and 17 were non-synonymous. Eight of the novel alleles could be most simply explained by gene conversion (short segmental exchange) because the specific nucleotide substitutions or sequence motifs in the novel cases were found in known alleles.

Most of the alleles found in the Cape Town cohort (74% of Class I and 65% of Class II) were listed in the common and well-documented (CWD) alleles catalog [22,23]. The non-CWD alleles included 22 Class I, and 18 Class II alleles that differed from CWD alleles in the fourth field of the allele name.

The allele diversity of G groups in the Cape Town subjects was comparable to that of HLA-A, -B, and -C reported by Cao et al. [8]. The higher resolution typing in this study allowed

alleles within G groups to be distinguished from one another due to NGS typing of introns and exons beyond those that code for the antigen recognition domain. There were 62 unique NGS alleles identified in HLA-A compared to 49 G groups; 72 of HLA-B versus 64 G groups; 47 of HLA-C versus 32 G groups; 21 versus 13 for DPA1; 46 versus 35 for DPB1; 27 versus 10 for DQA1; 26 versus 15 for DQB1; and 41 versus 40 for DRB1. All of the alleles previously reported in all five African populations by Cao et al [8] and the most frequent alleles in South African Black Africans [9] were also found in Cape Town.

Guo and Thompson test *p*-values for the four-field and G group alleles in three population groups are listed in Table 3. The distributions of HLA-A, -DPA1, -DPB1, -DQA1, -DQB1, -DRB1, and -DRB4 genotypes were in HWE. However there was significant deviation from HWE for HLA-B, HLA-C, HLA-DRB3, and HLA-DRB5 in at least one group. The HWE deviations may be due to several causes such as population substructure or admixture as 63% of the cohort is known to have mixed ancestry. Other reasons for large HWE deviations may be due to nonrandom sample selection, possible relatedness of the subjects, and genotyping ambiguity or error.

3.1. Comparison of populations in Cape Town

Table 4 shows the genetic distances between seven populations estimated by the distribution of HLA-A, -B, and -C. Overall, the Cape Town cohort was most similar to African Americans and Sub-Saharan Africans. Interestingly, the genetic distance between Cape Town and European Americans was smaller compared to South Asian populations, possibly reflecting a greater contribution of European alleles.

Fig. 2 illustrates the geographic distribution of the HLA-C alleles observed in this study compared to Asian Americans (API) and European Americans (EUR) in the NMDP registry [3,7]. The alleles were ordered by descending frequency in Cape Town (thick, medium blue bars) compared to API (medium thick, light blue bars) and EUR (thin, black bars.) Nine HLA-C alleles (C*15:02:01:01, C*14:02:01, C*08:01:01, C*03:02:02:01, C*04:03:01, C*01:02:01, C*12:02:02, C*16:02:01, and C*14:03) were predominantly found in API, defined as being at least two-fold more frequent in API than any other category in the NMDP registry. In addition, five alleles (C*02:02:05, C*05:01:01:02, C*07:01:01:01, C*08:02:01:01, and C*12:03:01:01) were predominantly found in EUR. The presence of API and EUR alleles is evidence of admixture.

3.2. Allele frequencies in Class I loci

There were 62 unique HLA-A alleles found in the subjects from Cape Town, listed in Supplementary Table 1. Ten alleles of HLA-A that were observed in at least 1% of the subjects from every group in the NMDP registry [7] were also observed in the Cape Town subjects. Similarly, all seven alleles that were observed in every African population sampled in Cao et al. [8] were also observed in the Cape Town subjects. The minimum number of HLA-A alleles with a cumulative frequency of at least 0.5 was 10. Two alleles, A*23:17 and A*43:01, were notably more frequent in Cape Town compared to other populations. A*23:17 had a frequency of 0.053 compared to 0.00021 in African Americans [7]. The allele A*43:01 had a frequency of 0.03980 and was previously reported in South African

Black Africans (frequency 0.0352), [9] African Americans (frequency 0.00021), [7] Zambians (frequency 0.0116) and Zulu (frequency 0.0300.) [8]. Six alleles that were observed in at least two subjects in Cape Town were not found in the Cao et al study [8]. Three of them (A*29:11, A*68:27:01 and A*01:23) were very rare in the NMDP registry (< 0.01%) [3] and were observed in individual cell lines with ethnic origin reported as Black African (IPD-IMGT/HLA Database [11].) A*68:27:01 was observed at a frequency of 0.0050 in South African Black Africans [9]. The other three alleles that were not found in the Cao et al. study [8] (A*24:07:01, A*34:01:01, and A*02:03:01) were also not observed in South African Black Africans [9]; instead they were found at the highest global frequency in Asian populations (www.allelefrequencies.net) [24] and were at least 20-fold more common in the broad API group compared to AFA or EUR [3,7]. Two alleles in this study (A*29:02:01:02, and A*66:02) were greater than four-fold more frequent in South African Black Africans compared to CMA or Caucasians.

There were 72 unique HLA-B alleles found in the subjects from Cape Town, listed in Supplementary Table 2. Seven alleles that were greater than 1% in all groups in the NMDP registry [7] were also observed in the Cape Town subjects. In addition, the ten alleles that were observed in every African population in Cao et al. [8] were also observed in Cape Town. The minimum number of HLA-B alleles with a cumulative frequency of at least 0.5 was 9. Two alleles (B*45:07 and B*82:02:01) were not found in Cao et al. [8] and were very rare in the NMDP registry [7]. The allele B*45:07 was observed at a frequency of 0.0051 in South African Black Africans [9] and in a single cell line with South African origin (IMGT/HLA Database [11].) Three additional alleles (B*14:01:01, B*15:10:01, and B*39:10:01) were found at a two-fold higher frequency in Cape Town than in Cao et al. [8] but were comparable to the frequency observed in South African Black Africans [9]. Three alleles in this study (B*42:01:01, B*45:07, and B*81:01) were greater than four-fold more frequent in South African Black Africans compared to CMA or Caucasians.

There were 47 unique HLA-C alleles found in the subjects from Cape Town, listed in Supplementary Table 3. Seven alleles that were greater than 1% in all groups in the NMDP registry [7] were also observed in the Cape Town subjects. An additional four alleles were observed in every African population sampled in Cao et al. [8] and all were also observed in the Cape Town subjects. The minimum number of HLA-C alleles with a cumulative frequency of at least 0.5 was 6. The allele C*08:04:01 was at a higher frequency in Cape Town than previously reported in three African populations [8] but was comparable to South African Black Africans [9]. The allele C*02:10:01:01 was not reported in any of the African populations but was reported in African and Hispanic Americans in the NMDP registry [7]. However, the former study [8] did not test exon 4, which differentiates C*02:10 from C*02:02. The allele C*02:10 was named previously C*02:02:04; when the sequence of this allele was extended it was found there was one amino acid difference with the C*02:02 allele in the alpha-3 domain. Six alleles that were observed at least twice in Cape Town were not found in any subject in the Cao et al study [8]. Four of the six were found in a lower proportion of subjects in the NMDP registry, but one (C*18:02) was high in African Americans and another one (C*04:03:01) was high in Asian Americans [7].

3.3. Allele frequencies in Class II loci

The observed frequencies of the Class II alleles are listed in Supplementary Tables 4–9. There were 21 unique DPA1 alleles, 46 alleles of DPB1, 27 alleles of DQA1, 26 alleles of DQB1, 41 alleles of DRB1, 5 DRB3, 4 DRB4, and 6 DRB5 alleles. In general, the allele frequencies were consistent with a mixed population that includes individuals of African, European and South Asian descent.

DPA1 typing has not been extensively characterized in different populations because it is less diverse than other loci (see www.allelefrequencies.net and Middleton et al.) [24]. The alleles are listed in Supplementary Table 4. The minimum number of DPA1 alleles with a cumulative frequency of at least 0.5 was 3. The novel allele DPA1*04:02 described in this study was only observed in South African Black African and CMA subjects.

DPB1 had more than twice as many alleles as in DPA1. The alleles are listed in Supplementary Table 5. The minimum number of DPB1 alleles with a cumulative frequency of at least 0.5 was 3. One of the novel alleles in this study (DPB1*654:01) was homozygous in one subject.

DQA1 and DQB1 had a similar number of alleles, 27 and 26 respectively. The alleles are listed in Supplementary Tables 6 and 7. The minimum number of DQA1 alleles with a cumulative frequency of at least 0.5 was 4, and the minimum number for DQB1 was 5.

DRB1 had 41 different alleles, including one low frequency allele (DRB1*11:17) that was homozygous in a subject where three other Class II loci were also homozygous. The alleles are listed in Supplementary Table 8. The minimum number of DRB1 alleles with a cumulative frequency of at least 0.5 was 6. The five most frequent alleles in Cape Town were consistent with that observed in South African Black Africans [9]. Five alleles that were observed in at least two subjects in Cape Town (DRB1*08:01:03, DRB1*08:12, DRB1*11:14:02, DRB1*12:02:02, and DRB1*15:02:01:01) were not observed in the South African Black cohort [9]. The last two alleles had the highest frequency in the Asian Pacific Islander (API) group in the NMDP database [7]. Alleles of DRB3, DRB4, and DRB5 are listed in Supplementary Table 9.

3.4. Haplotype frequencies

Haplotype frequencies were estimated for seven selected combinations of two, three, and four loci and the results are shown in Supplementary Tables 10–16. As expected, the number of haplotypes found in three or more chromosomes correlated with allele diversity and map distance. For example, there were 92 haplotypes of B–C ($D' = 0.91$) and 150 of A–B ($D' = 0.75$); similarly, there were 44 haplotypes of DQA1–DQB1 ($D' = 0.91$) and 112 of DQB1–DPA1 ($D' = 0.43$).

Supplementary Table 10 lists 16 different haplotypes of A–B found in 1% or more of the chromosome counts, representing 22.8% of all chromosomes. All sixteen were previously observed in either Cao et al. or the NMDP registry [3,7,8]. Of the 11 A–B haplotypes observed in three or more African groups, [8] the Cape Town cohort had all except one (A*02:01–B*15:03).

Supplementary Table 11 lists 28 haplotypes of B–C that were found at a frequency of 1% or greater, representing 60% of all chromosomes. All of the B–C haplotypes with frequencies greater than 2% were reported previously; [3,8,9] one possible exception was B*15:03:01:02–C*02:10:01:01 because C*02:10 was not differentiated from C*02:02 in the previous studies. One haplotype (B*53:01–C*04:01) previously found in a South African Black African cohort [9] was not among the estimated haplotypes in Cape Town.

Nine A–B–C haplotypes that had estimated frequencies of 1% or greater in the Cape Town cohort and are listed in Supplementary Table 12. All except A*23:01:01–B*15:10:01–C*16:01:01:01 were found in previous studies [8,9]. The haplotype A*03:01:01:01–B*07:02:01–C*07:02:01:01 was only found in the Caucasian cohort [9].

Supplementary Table 13 lists 18 different haplotypes of DPA1–DPB1 found at a frequency of 1% or greater, representing 77% of all chromosomes. One frequent haplotype, DPA1*04:02–DPB1*105:01, included a novel allele described in this manuscript. Interestingly, DPA1*04:02 was strongly linked to DPB1*105:01, whereas the most closely related allele, DPA1*04:01 was linked to two different alleles of DPB1, namely DPB1*296:01 (N = 5), and DPB1*13:01:01 (N = 3).

Supplementary Table 14 shows 26 most frequent haplotypes of DQA1–DQB1–DRB1 observed in 1% or more of the Cape Town subjects, representing 78% of all chromosomes. All of the haplotypes corresponded with two locus, G group DQB1–DRB1 haplotypes that were observed in African Americans (AFA) in the NMDP registry [7]. One haplotype, DQB1*03:01:01:03–DRB1*12:02:01, was found at a high frequency (greater than 7%) in Asian Americans (API)[7].

The frequency of A–B–DRB1 haplotypes in Cape Town observed in 0.5% or more of the Cape Town subjects are listed in Supplementary Table 15. They were compared to those represented in the NMDP registry [7]. The most frequent haplotype in the AFA population (A*30:01:01–B*42:01:01–DRB1*03:02:01) was also the most frequent in Cape Town. However, two of the most frequent AFA haplotypes (top 10) were not found in Cape Town (A*33:03–B*53:01–DRB1*08:04 and A*34:02–B*44:03–DRB1*15:03.) In contrast, two relatively common (> 0.5%) A–B–DRB1 haplotypes in Cape Town (A*43:01–B*15:10:01–RB1*04:01:01 and A*43:01–B*44:03:01:01–DRB1*04:01:01) were not found in any of the subjects in the NMDP registry [7]. Two additional haplotypes (A*02:01:01:01–B*45:01:01–DRB1*13:01:01 and A*23:01:01–B*15:10:01–DRB1*11:01:02) were rare in US individuals (0–0.12%) but were in the top five most frequent haplotypes in Cape Town.

The haplotype organization of the region encompassing DRB1–DRB3/4/5 has been shown to be correlated with the allele present on the DRB1 locus [25]. As shown in Supplementary Table 16, the expected linkages were observed in most cases. However, DRB5 was absent in some individuals carrying DRB1*15 alleles. The DRB1*15:03:01:01–DRBX*NNNN haplotype was previously reported in AFA [26,27] while the DRB1*15:01:01:01–DRBX*NNNN haplotype was most frequent in EUR and API [3]. As expected, only DRB1*15:02:01:01 was observed with DRB5*01:02 and only DRB1*16:01:01:01 was observed with DRB5*02:02 [28]. The unusual haplotype DRB1*10:01:01–DRB5*01:01:01

was observed in two subjects; possibly representing a rare haplotype that has persisted in the population.

The most frequent extended haplotypes of A–B–C–DQA1–DQB1–DRB1–DRB3/4/5 are listed in Table 5. There were 46 haplotypes estimated to be present in three or more subjects, representing 26.4% of the chromosomes in the study. All of the top five extended haplotypes were composed of alleles that were present at 1% or greater in the Cape Town cohort. The most common extended haplotype in Cape Town was identical to the most frequent A–B–C–DQB1–DRB1–DRB3 haplotype observed in African Americans (AFA), Black Africans and Caribbean Africans [3]. Extended haplotypes that included DPA1 and DPB1 were not presented because of low linkage disequilibrium.

4. Discussion

Allelic resolution NGS typing of extended HLA gene products is a powerful method to characterize HLA allele and haplotype diversity in population studies. In this study, sequencing resolution allowed detection of previously unreported variants, including coding sequence changes. In addition, whole gene coverage identified polymorphic sites in both coding and noncoding regions and enabled unambiguous phasing by defining the physical linkage between exons.

The subjects typed in this study were from a relatively small cohort from Cape Town, South Africa. As expected, the HLA genotypes included many that are common across the world and across Africa, and contributions of Asian and European alleles that reflect the demographic history of Cape Town. The majority of the Cape Town cohort was South African Cape Mixed Ancestry (CMA), a population that has been shown to have a complex admixture of ancestral populations, including African, European, Indian and South-East Asian [1,29–31]. As a result of the mixed ancestry of the Cape Town population, the frequency data provided in the Supplementary Tables should be used with caution. They may not be appropriate for use as reference data in future population studies or for comparisons, because they present a heterogeneous cohort that does not necessarily represent a true population.

In contrast to the lower resolution HLA typing of many previous studies [7–9], the alleles reported in this study were based on nonambiguous, allelic resolution typing. One exception is DPB1 where some allele combinations could not be unambiguously phased because of the low SNP diversity across intron 1. The phasing ambiguities could be reduced by expanding the number of genomic sequences in the IMGT/HLA database and by using sequencing methods, such as PacBio or Nanopore, that span longer distances. Illustrating the improved resolution of NGS typing in HLA-A, five sets of alleles were distinguished at the fourth field based on intron variation, and four sets were distinguished at the third field based on synonymous base substitutions. Four alleles of A*30:02 were distinguished, namely A*30:02:01:01, A*30:02:01:02, A*30:02:01:03 and A*30:02:02; and three alleles of A*68:01 were distinguished, namely A*68:01:01:02, A*68:01:02:01, and A*68:01:02:02. (See Supplementary Table 1). Other alleles, such as A*23:01:01 and A*23:17 were not distinguished previously because they are in the same G group.

Additional, larger scale NGS studies will be needed to capture a more complete picture of the world's HLA diversity, especially in communities with complex immigration patterns and the resulting admixture. Fortunately, new methods for allelic resolution HLA typing, such as the semi-automated NGS approach presented here, are rapidly becoming available.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We are grateful to the reviewers of this manuscript for their helpful suggestions. This work was funded by Sirona Genomics, the Bill and Melinda Gates Foundation [grant numbers OPP1066265 and 116468], the National Institutes of Health [2U19AI057229] and the Howard Hughes Medical Institute.

References

- [1]. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, et al., The genetic structure and history of Africans and African Americans, *Science* 324 (5930) (2009) 1035–1044. [PubMed: 19407144]
- [2]. Mahomed H, Hawkrigde T, Verver S, Abrahams D, Geiter L, Hatherill M, Ehrlich R, Hanekom WA, Hussey GD, The tuberculin skin test versus QuantiFERON TB Gold(R) in predicting tuberculosis disease in an adolescent cohort study in South Africa, *PLoS One* 6 (3) (2011) e17984. [PubMed: 21479236]
- [3]. Gragert L, Madbouly A, Freeman J, Maiers M, Six-locus high resolution HLA haplotype frequencies derived from mixed-resolution DNA typing for the entire US donor registry, *Hum. Immunol* 74 (10) (2013) 1313–1320. [PubMed: 23806270]
- [4]. Lee SJ, Klein J, Haagenson M, Baxter-Lowe LA, Confer DL, Eapen M, Fernandez-Vina M, Flomenberg N, Horowitz M, Hurley CK, et al., High-resolution donor-recipient HLA matching contributes to the success of unrelated donor marrow transplantation, *Blood* 110 (13) (2007) 4576–4583. [PubMed: 17785583]
- [5]. Speiser DE, Tiercy JM, Rufer N, Grundschober C, Gratwohl A, Chapuis B, Helg C, Loliger CC, Siren MK, Roosnek E, et al., High resolution HLA matching associated with decreased mortality after unrelated bone marrow transplantation, *Blood* 87 (10) (1996) 4455–4462. [PubMed: 8639808]
- [6]. McKinney DM, Southwood S, Hinz D, Oseroff C, Arlehamn CS, Schulten V, Taplitz R, Broide D, Hanekom WA, Scriba TJ, et al., A strategy to determine HLA class II restriction broadly covering the DR, DP, and DQ allelic variants most commonly expressed in the general population, *Immunogenetics* 65 (5) (2013) 357–370. [PubMed: 23392739]
- [7]. Maiers M, Gragert L, Klitz W, High-resolution HLA alleles and haplotypes in the United States population, *Hum. Immunol* 68 (9) (2007) 779–788. [PubMed: 17869653]
- [8]. Cao K, Moormann AM, Lyke KE, Masaberg C, Sumba OP, Doumbo OK, Koech D, Lancaster A, Nelson M, Meyer D, et al., Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci, *Tissue Antigens* 63 (4) (2004) 293–325. [PubMed: 15009803]
- [9]. Paximadis M, Mathebula TY, Gentle NL, Vardas E, Colvin M, Gray CM, Tiemessen CT, Puren A, Human leukocyte antigen class I (A, B, C) and II (DRB1) diversity in the black and Caucasian South African population, *Hum. Immunol* 73 (1) (2012) 80–92. [PubMed: 22074999]
- [10]. Wang C, Krishnakumar S, Wilhelmy J, Babrzadeh F, Stepanyan L, Su LF, Levinson D, Fernandez-Vina MA, Davis RW, Davis MM, et al., High-throughput, high-fidelity HLA genotyping with deep sequencing, *Proc. Natl. Acad. Sci. U.S.A* 109 (22) (2012) 8676–8681. [PubMed: 22589303]

- [11]. Robinson J, Soormally AR, Hayhurst JD, Marsh SG, The IPD-IMGT/HLA Database – New developments in reporting HLA variation, *Hum. Immunol* 77 (3) (2016) 233–237. [PubMed: 26826444]
- [12]. Guo SW, Thompson EA, Performing the exact test of Hardy-Weinberg proportion for multiple alleles, *Biometrics* 48 (2) (1992) 361–372. [PubMed: 1637966]
- [13]. Lancaster AK, Single RM, Solberg OD, Nelson MP, Thomson G, PyPop up-date—a software pipeline for large-scale multilocus population genomics, *Tissue Antigens* 69 (Suppl 1) (2007) 192–197. [PubMed: 17445199]
- [14]. Excoffier L, Slatkin M, Maximum-likelihood estimation of molecular haplotype frequencies in a diploid population, *Mol. Biol. Evol* 12 (5) (1995) 921–927. [PubMed: 7476138]
- [15]. Dempster A, Laird N, Rubin D, Maximum likelihood from incomplete data via the EM algorithm, *J. R. Stat. Soc* 39 (1977) 1–38.
- [16]. Lewontin RC, The interaction of selection and linkage. I. general considerations; heterotic models, *Genet.* 49 (1) (1964) 49–67.
- [17]. Hedrick PW, Gametic disequilibrium measures: proceed with caution, *Genetics* 117 (2) (1987) 331–341. [PubMed: 3666445]
- [18]. Cramer H, *Mathematical Models of Statistics*, Asia Publishing House, New York, NY, 1946.
- [19]. Pappas DJ, Marin W, Hollenbach JA, Mack SJ, Bridging immunogenomic data analysis workflow gaps (BIGDAWG): an integrated case-control analysis pipeline, *Hum Immunol* 77 (3) (2016) 283–287. [PubMed: 26708359]
- [20]. Cavalli-Sforza L, Bodmer W, *The Genetics of Human Populations*, W.H. Freeman, San Francisco, CA, USA, 1971.
- [21]. Pickbourne P, Piazza A, Bodmer W: Population analysis. In: *Histocompatibility testing*. Copenhagen, Denmark: Munksgaard; 1977: 259–278.
- [22]. Cano P, Klitz W, Mack SJ, Maier M, Marsh SG, Noreen H, Reed EF, Senitzer D, Setterholm M, Smith A, et al., Common and well-documented HLA alleles: report of the Ad-Hoc committee of the american society for histocompatibility and immunogenetics, *Hum. Immunol* 68 (5) (2007) 392–417. [PubMed: 17462507]
- [23]. Mack SJ, Cano P, Hollenbach JA, He J, Hurley CK, Middleton D, Moraes ME, Pereira SE, Kempenich JH, Reed EF, et al., Common and well-documented HLA alleles: 2012 update to the CWD catalogue, *Tissue Antigens* 81 (4) (2013) 194–203. [PubMed: 23510415]
- [24]. Middleton D, Menchaca L, Rood H, Komarovsky R, New allele frequency database, *Tissue Antigens* 61 (5) (2003) 403–407. [PubMed: 12753660]
- [25]. Andersson G, Evolution of the human HLA-DR region, *Front. Biosci* 3 (1998) d739–745. [PubMed: 9675159]
- [26]. Caillier SJ, Briggs F, Cree BA, Baranzini SE, Fernandez-Vina M, Ramsay PP, Khan O, Royal W 3rd, Hauser SL, Barcellos LF, et al., Uncoupling the roles of HLA-DRB1 and HLA-DRB5 genes in multiple sclerosis, *J. Immunol* 181 (8) (2008) 5473–5480. [PubMed: 18832704]
- [27]. Robbins F, Hurley CK, Tang T, Yao H, Lin YS, Wade J, Goeken N, Hartzman RJ, Diversity associated with the second expressed HLA-DRB locus in the human population, *Immunogenetics* 46 (2) (1997) 104–110. [PubMed: 9162096]
- [28]. Moraes ME, Fernandez-Vina M, Stastny P: DNA typing for class II HLA antigens with allele-specific or group-specific amplification. IV. Typing for alleles of the HLA-DR2 group, *Hum. Immunol* 31 (2) (1991) 139–144. [PubMed: 2066273]
- [29]. Chimusa ER, Daya M, Moller M, Ramesar R, Henn BM, van Helden PD, Mulder NJ, Hoal EG, Determining ancestry proportions in complex admixture scenarios in South Africa using a novel proxy ancestry selection method, *PLoS One* 8(9) (2013) e73971. [PubMed: 24066090]
- [30]. Patterson N, Petersen DC, van der Ross RE, Sudoyo H, Glashoff RH, Marzuki S, Reich D, Hayes VM, Genetic structure of a unique admixed population: implications for medical research, *Hum. Mol. Genet* 19 (3) (2010) 411–419. [PubMed: 19892779]
- [31]. Quintana-Murci L, Harmant C, Quach H, Balanovsky O, Zaporozhchenko V, Bormans C, van Helden PD, Hoal EG, Behar DM, Strong maternal Khoisan contribution to the South African coloured population: a case of gender-biased admixture, *Am. J. Hum. Genet* 86 (4) (2010) 611–620. [PubMed: 20346436]

Fig. 1.

Coverage of HLA loci using MIA FORA NGS HLA typing kit. Human genomic DNA was amplified by long-range PCR, targeting 11 HLA genes in nine all-in-one master mixes. Location of HLA region on human chromosome 6 is indicated at the top of the figure; each following line represents a different targeted HLA locus. Purple boxes at each end represent untranslated regions; light blue boxes represent exons in Class I loci; red boxes represent exons in Class II loci; red arrows represent positions of primers used for long-range PCR. HLA-DRB1, 3, 4, 5 loci were amplified together in two separate reactions; the boxed region on the left represents the exon 1 amplified region.

Fig. 2. Geographic distribution of HLA-C alleles observed in this study compared to Asian Americans (API) and European Americans (EUR) in the NMDP registry. Alleles were ordered by descending frequency in Cape Town (thick, medium blue bars) compared to API (medium thick, light blue bars) and EUR (thin, black bars.)

Table 1

Demographics of cape town cohort (n = 402). Ethnicity was self-reported as locally defined.

		Number	Percentage
Gender	Female	223	55.5%
	Male	179	44.5%
Age	12, 13, 14	136	33.8%
	15, 16	181	45.0%
	17, 18	85	21.1%
Ethnicity	Black African	125	31.1%
	CMA ^a	250	62.2%
	White	27	6.7%
QuantiFERON Status	Negative	203	50.5%
	Positive	199	49.5%
Previous TB Diagnosis	No	355	88.3%
	Yes	46	11.4%
	Unknown	1	0.2%

^aCape mixed ancestry.

Table 2

Novel HLA alleles observed in cape town cohort.

Temporary Name ^a	Number Alleles	Nucleotide Substitution ^b	Exon	Codon and Amino Acid Change ^c	Genbank Accession	IMGT Name
A*23:17×1	1	TTC > CTC	exon 3	109 Phe > Leu (nonsynonymous)	MF371301	TBD
A*26:01:01:01×1	1	TAC > CAC	exon 2	85 Tyr > His (nonsynonymous)	MF371310	TBD
A*32:01:01×1	19	TGG > AGG	exon 1	(-2) Trp > Arg (nonsynonymous)	KY575119	TBD
A*32:01:01×2	3	AGG > AGA	exon 5	307 Arg > Arg (synonymous)	MG470795	TBD
B*07:02:01×1	1	GTC > ATC	exon 4	247 Val > Met (nonsynonymous)	MF371302	TBD
B*08:01:01×1	1	AAC > GAG	exon 2	63 Asn > Glu (nonsynonymous)	MF405183	TBD
C*17:01:01:02×1	3	CCC > CTC	exon 3	193 Pro > Leu (nonsynonymous)	KY859857	C*17:37
C*17:01:01:02×2	1	GCT > ACT	exon 5	311 Ala > Thr (nonsynonymous)	MF371303	TBD
DPA1*02:01:01:02×1	12	CGT > TGT	exon 2	76 Arg > Cyt (nonsynonymous)	MF285076	DPA1*02:09
DPA1*02:02:01×1 ^d	14	GTC > ATC	exon 3	91 Val > Met (nonsynonymous)	KY807148	DPA1*02:07:01:03
		CCA > CCG	exon 3	127 Pro > Pro (synonymous)		
		GTA > GTG	exon 3	154 Val > Val (synonymous)		
DPA1*03:01:01×1	7	GTC > GTC	exon 4	204 Val > Val (synonymous)	KY807146	DPA1*03:01:02
DPA1*03:01:01×2	1	ACG > ATG	exon 3	120 Thr > Met (nonsynonymous)	MF371304	TBD
DPA1*04:01×1	14	GCG > ACG	exon 4	187 Ala > Thr (nonsynonymous)	KY807145	DPA1*04:02
DPB1*03:01:01×1	1	CGG > CAG	exon 4	194 Arg > Gln (nonsynonymous)	MF371307	TBD
DPB1*11:01:01×1	7	CGG > CAG	exon 4	189 Arg > Gln (nonsynonymous)	KY859858	DPB1*654:01
DPB1*55:01×1	1	AAC > AAT	exon 3	144 Asn > Asn (synonymous)	MF371308	TBD
DPB1*106:01×1	2	AGG > GGG	exon 2	75 Arg > Gly (nonsynonymous)	KY859859	DPB1*558:01
DPB1*377:01×1	1	GCG > GTG	exon 2	36 Ala > Val (nonsynonymous)	MF371309	TBD
DQA1*01:02:01:01×1	1	GAT > GAA	exon 3	145 Asp > Glu (nonsynonymous)	MF371311	TBD
DQA1*03:03:01:01×1	1	ATC > ATT	exon 2	66 Ile > Ile (synonymous)	MF371305	TBD
DQA1*04:01:01×1	9	CTC > CAC	exon 2	51 Leu > His (nonsynonymous)	KY859860	TBD
DQB1*05:01:01:01×1	3	TCC > TCT	exon 1	(-1 0) Ser > Ser (synonymous)	KY859861	DQB1*05:01:24

^aTemporary name based on most similar allele.

^bNucleotide of the previously reported allele is listed first, differences are underlined.

^cAmino acid encoded by the previously reported allele is shown first.

DPA1*02:01×1 has three variants in exon 3.
p

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3

Guo and Thompson Test for Hardy Weinberg Equilibrium in Three Cape Town Populations. Calculations of overall p-value were performed for both four field and G group genotypes.

	Black African (2n = 250)		Cape Mixed Ancestry (2n = 500)		White (2n = 54)	
	Four Field	G Group	Four Field	G Group	Four Field	G Group
A	0.3527	0.5719	0.0540	0.0312 *	0.6479	0.8279
B	0.0029 **	0.0016 **	0.1808	0.2914	0.0182 *	0.0182 *
C	0.0053 **	0.0540	0.1342	0.0293 *	0.4192	0.2256
DPA1	0.2216	0.4120	0.1161	0.8981	0.9462	1.0000
DPB1	0.0548	0.0609	0.1967	0.0685	0.8933	0.9671
DQA1	0.5224	0.2172	0.4131	0.4740	0.4100	0.3436
DQB1	0.3415	0.5831	0.1721	0.3104	0.7939	0.4550
DRB1	0.4713	0.3900	0.1756	0.1805	0.8211	0.6688
DRB3	0.3288	0.3159	0.0324 *	0.6454	0.0419 *	0.1126
DRB4	0.5436	1.0000	0.1038	0.3968	1.0000	1.0000
DRB5	0.6172	0.6172	0.2772	0.2520	0.0289 *	0.0291 *

* P < 0.05, statistically significant;

** P < 0.01, highly statistically significant.

Table 4

Genetic distances between seven populations estimated by the distribution of HLA-A, HLA-B, and HLA-C alleles.

Population	EuAm ^a	AfAm ^a	Cape Town ^b	SSAfrica ^c	Zulu ^c	Tamil ^c	NDehli ^c
<i>HLA-A</i>							
EuAm	0						
AfAm	0.491	0					
Cape Town	0.469	0.332	0				
SSAfrica	0.515	0.214	0.340	0			
Zulu	0.580	0.302	0.352	0.279	0		
Tamil	0.573	0.722	0.663	0.771	0.792	0	
NDehli	0.483	0.669	0.610	0.711	0.752	0.382	0
<i>HLA-B</i>							
EuAm	0						
AfAm	0.565	0					
Cape Town	0.562	0.381	0				
SSAfrica	0.633	0.250	0.401	0			
Zulu	0.712	0.381	0.367	0.343	0		
Tamil	0.702	0.777	0.761	0.791	0.853	0	
NDehli	0.562	0.693	0.686	0.725	0.791	0.445	0
<i>HLA-C</i>							
EuAm	0						
AfAm	0.412	0					
Cape Town	0.399	0.200	0				
SSAfrica	0.367	0.278	0.296	0			
Zulu	0.524	0.453	0.424	0.327	0		
Tamil	0.507	0.615	0.553	0.562	0.620	0	
NDehli	0.546	0.579	0.579	0.574	0.649	0.445	0

^aData from our unpublished study: EuAm, European American; AfAm, African American.

^bData from this study.

^cData from the MHC database <http://www.ncbi.nlm.nih/projects/gv/mhc/ihwg.cgi> SSAfrica, Sub-Saharan Africa includes ethnic populations from Kenya (three groups), Dogon Mali, Uganda, Zambia and Zimbabwe. Tamil Indian population collected in Durban, South Africa; NDehli; New Delhi, India.

Table 5

Extended Haplotypes of A-B-C-DQA1-DQB1-DRB1-DRB3-DRB4-DRB5.

	Black (2n = 250)			Cape Mixed (2n = 500)			White (2n = 54)		
	Number	Frequency		Number	Frequency		Number	Frequency	TOTAL
A*01:01:01-01-B*08:01:01-C*07:02:01-01-DQA1*05:01:01-02-DQB1*02:01:01-DRB1*03:01:01-01-DRB3*01:01:02:01	0	0		6	0.012		2	0.038	8
A*02:01:01-01-B*07:02:01-C*07:02:01-03-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:01:01-01-DRB5*01:01:01:01	0	0		4	0.008		1	0.019	5
A*02:01:01-01-B*40:01:02-C*03:04:01:01-DQA1*01:02:01:04-DQB1*06:04:01-DRB1*13:02:01-01-DRB3*03:01:01:01	0	0		3	0.006		1	0.019	4
A*02:01:01-01-B*44:02:01:01-C*05:01:01:02-DQA1*03:03:01:01-DQB1*03:01:01-01-DRB1*04:01:01-01-DRB4*01:03:01:01	0	0		3	0.006		0	0	3
A*02:01:01-01-B*45:01:01-C*16:01:01:01-DQA1*01:03:01:02-DQB1*06:03:01-DRB1*13:01:01-01-01-DRB3*01:01:02:01	4	0.016		7	0.014		0	0	11
A*02:01:01-01-B*45:07-C*16:01:01:01-DQA1*01:03:01:02-DQB1*06:03:01-DRB1*13:01:01-01-01-DRB3*01:01:02:01	2	0.008		2	0.004		0	0	4
A*02:02:01-01-B*57:03:01-C*07:02:01-02-DQA1*01:01:02-DQB1*05:01:01-01-DRB1*01:02:01-01-DRB3*NNNN	3	0.012		1	0.002		0	0	4
A*02:05:01-01-B*14:01:01-C*08:04:01-DQA1*01:03:01:02-DQB1*06:03:01-DRB1*13:01:01-01-01-DRB3*01:01:02:01	2	0.008		1	0.002		0	0	3
A*02:05:01-01-B*58:01:01:01-C*07:18-DQA1*05:05:01:01-DQB1*03:19:01-DRB1*11:02:01-01-DRB3*03:01:01:01	3	0.012		1	0.002		0	0	4
A*03:01:01-01-B*07:02:01-C*07:02:01-03-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:01:01-01-01-DRB5*01:01:01:01	0	0		1	0.002		4	0.077	5
A*03:01:01-05-B*08:01:01-C*07:02:01-01-DQA1*04:01:01×1-DQB1*04:02:01-DRB1*03:02:01-01-DRB3*01:01:02:01	2	0.008		1	0.002		0	0	3
A*03:01:01-05-B*35:02:01-C*04:01:01-06-DQA1*05:05:01:01-DQB1*03:01:01-02-DRB1*11:04:01-01-DRB3*02:02:01:02	2	0.008		1	0.002		0	0	3
A*03:01:01-05-B*47:01:01-03-C*06:02:01:01-DQA1*05:01:01-01-DQB1*02:01:01-DRB1*03:01:01-01-01-DRB3*02:02:01:01	0	0		3	0.006		0	0	3
A*23:01:01-B*15:10:01-C*16:01:01-01-DQA1*05:02-DQB1*03:19:01-DRB1*11:01:02-01-DRB3*02:02:01:02	2	0.008		5	0.01		0	0	7
A*23:17-B*44:03:01:01-C*04:01:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:03:01:01-01-DRB5*01:01:01:01	1	0.004		2	0.004		0	0	3
A*23:17-B*58:02:01-C*06:02:01:01-DQA1*03:03:01:01-DQB1*03:02:01-DRB1*04:01:01-01-DRB4*01:01:01:01	0	0		4	0.008		0	0	4
A*24:02:01-01-B*07:02:01-C*07:02:01:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:03:01:01-01-DRB5*01:01:01:01	4	0.016		1	0.002		0	0	5

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Black (2n = 250)		Cape Mixed (2n = 500)		White (2n = 54)		TOTAL
	Number	Frequency	Number	Frequency	Number	Frequency	
A*24:02:01:01-B*57:01:01-C*06:02:01:01-DQA1*02:01:01:01-DRB1*07:01:01:01-DRB4*01:03:01:02N	0	0	4	0.008	0	0	4
A*24:07:01-B*35:05:01-C*04:01:01:01-DQA1*06:01:01:01-DQB1*03:01:01:03-DRB1*2:02:01-DRB3*03:01:03	0	0	3	0.006	0	0	3
A*29:01:01:01-B*18:01:01:02-C*07:04:01:01-DQA1*01:02:01:04-DQB1*06:09:01-DRB1*13:02:01-DRB3*03:01:01	2	0.008	4	0.008	0	0	6
A*29:02:01:01-B*44:03:02-C*07:06-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*11:01:02-DRB3*NNNN	1	0.004	2	0.004	0	0	3
A*29:02:01:01-B*44:03:02-C*07:06-DQA1*05:05:01:01-DQB1*03:19:01-DRB1*11:01:02-DRB3*02:02:01:02	2	0.008	1	0.002	0	0	3
A*29:02:01:02-B*15:03:01:02-C*02:10:01:01-DQA1*01:02:01:04-DQB1*06:09:01-DRB1*13:02:01-DRB3*03:01:01	3	0.012	0	0	0	0	3
A*29:02:01:02-B*42:01:01-C*17:01:01:02-DQA1*04:01:01-DQB1*04:02:01-DRB1*03:02:01-DRB3*01:01:02:01	3	0.012	1	0.002	0	0	4
A*30:01:01-B*07:02:01-C*07:02:01:01-DQA1*01:01:02-DQB1*05:01:01:01-DRB1*01:02:01-DRB3*NNNN	2	0.008	1	0.002	0	0	3
A*30:01:01-B*18:01:01:02-C*07:04:01:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*11:01:02-DRB3*02:02:01:02	1	0.004	2	0.004	0	0	3
A*30:01:01-B*42:01:01-C*17:01:01:02-DQA1*04:01:01-DQB1*04:02:01-DRB1*03:02:01-DRB3*01:01:02:01	15	0.06	5	0.01	0	0	20
A*30:01:01-B*42:01:01-C*17:01:01:02-DQA1*04:01:01-DQB1*04:02:01-DRB1*03:02:01-DRB3*01:01:02:01	5	0.02	1	0.002	0	0	6
A*30:01:01-B*42:01:02-C*17:01:01:02-DQA1*01:05:01-DQB1*05:01:01:02-DRB1*12:01:01:03-DRB3*01:01:02:01	2	0.008	1	0.002	0	0	3
A*30:01:02-B*81:01-C*04:01:01:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:03:01:01-DRB5*01:01:01	4	0.016	0	0	0	0	4
A*30:02:01:02-B*45:01:01-C*16:01:01:01-DQA1*01:01:02-DQB1*05:01:01:01-DRB1*01:02:01-DRB3*NNNN	1	0.004	2	0.004	0	0	3
A*30:02:01:02-B*58:02:01-C*06:02:01:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:03:01:01-DRB5*01:01:01	2	0.008	1	0.002	0	0	3
A*30:02:01:03-B*08:01:01-C*07:01:01:01-DQA1*05:01:01:02-DQB1*02:01:01-DRB1*03:01:01:01-DRB3*02:02:01:02	2	0.008	2	0.004	0	0	4
A*30:04:01-B*39:10:01-C*15:05:02-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:03:01:01-DRB3*NNNN	1	0.004	3	0.006	0	0	4
A*32:01:01x1-B*44:03:01:01-C*02:10:01:01-DQA1*03:03:01:01-DQB1*03:02:01-DRB1*04:01:01-DRB4*01:01:01	0	0	3	0.006	0	0	3
A*33:03:01-B*44:03:02-C*07:06-DQA1*02:01:01:01-DQB1*02:02:01:01-DRB1*07:01:01:01-DRB4*01:03:01:01	0	0	4	0.008	0	0	4

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

	Black (2n = 250)		Cape Mixed (2n = 500)		White (2n = 54)		TOTAL
	Number	Frequency	Number	Frequency	Number	Frequency	
A*34:01:01-B*40:02:01-C*15:02:01:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*15:01:01:01-DRB5*01:01:01	0	0	3	0.006	0	0	3
A*34:02:01-B*08:01:01-C*07:01:01:01-DQA1*04:01:01-DQB1*04:02:01-DRB1*03:02:01-DRB3*01:01:02:01	3	0.012	0	0	0	0	3
A*36:01-B*53:01:01-C*04:01:01:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*11:01:02-DRB3*02:02:01:02	1	0.004	2	0.004	0	0	3
A*43:01-B*15:10:01-C*04:01:01:01-DQA1*03:03:01:01-DQB1*03:02:01-DRB1*04:01:01-DRB4*01:01:01:01	1	0.004	4	0.008	0	0	5
A*43:01-B*44:03:01:01-C*02:10:01:01-DQA1*03:03:01:01-DQB1*03:02:01-DRB1*04:01:01-DRB4*01:01:01:01	0	0	4	0.008	0	0	4
A*66:02-B*42:01:01-C*17:01:01:02-DQA1*04:01:01-DQB1*04:02:01-DRB1*03:02:01-DRB3*01:01:02:01	4	0.016	1	0.002	0	0	5
A*68:01:01:02-B*58:02:01-C*06:02:01:01-DQA1*03:03:01:01-DQB1*02:02:01:02-DRB1*07:01:01:01-DRB4*01:01:01:01	4	0.016	1	0.002	0	0	5
A*68:02:01:01-B*15:10:01-C*03:04:02-DQA1*05:01:01:01-DQB1*02:01:01-DRB1*03:01:01:01-DRB3*02:02:01:01	7	0.028	4	0.008	0	0	11
A*68:02:01:01-B*15:10:01-C*08:04:01-DQA1*01:02:01:01-DQB1*06:02:01-DRB1*11:01:02-DRB3*03:01:01	3	0.012	1	0.002	0	0	4
A*74:01:01-B*15:03:01:02-C*02:10:01:01-DQA1*01:02:01:04-DQB1*06:09:01-DRB1*13:02:01-DRB3*03:01:01	0	0	4	0.008	0	0	4