Tissue Antigens ISSN 0001-2815

BRIEF COMMUNICATION

HLA class I allele and haplotype diversity in Ugandans supports the presence of a major east African genetic cluster

G. H. Kijak[1], A. M. Walsh[1], R. N. Koehler[1], N. Moqueet[1], L. A. Eller[2], M. Eller[2], J. R. Currier[1], Z. Wang[1], F. Wabwire-Mangen[3], H. N. Kibuuka[3], N. L. Michael[4], M. L. Robb[1] & F. E. McCutchan[1]

1 Division of Retrovirology, US Military HIV Research Program/Henry M. Jackson Foundation, Rockville, MD, USA
2 Makerere University Walter Reed Project, Henry M. Jackson Foundation, Kampala, Uganda
3 Makerere University Walter Reed Project, Kampala, Uganda
4 Division of Retrovirology, US Military HIV Research Program/Walter Reed Army Institute of Research, Silver Spring, MD, USA

Key words
East Africa; genetic distance; haplotype frequency; HLA class I genes; Uganda

Correspondence
Gustavo Hernan Kijak, PharmD, PhD
Division of Retrovirology
US Military HIV Research Program/Henry M. Jackson Foundation
1600 East Gude Drive
Rockville
MD 20850
USA
Tel: +1 301 251 5046
Fax: +1 301 762 7460
e-mail: gkijak@hivresearch.org

Received 23 June 2008; revised 17 September 2008, 4 November 2008; accepted 19 November 2008

doi: 10.1111/j.1399-0039.2008.01192.x

Abstract

The objective of this study was to characterize the class I human leukocyte antigen (HLA) genetic composition of the Ugandan population to better define its relationship with other African groups. Samples from 175 individuals from Kampala (Uganda) were subjected to class I HLA-A, -B, and -C sequence-based typing. The high concordance between the major alleles and haplotypes found in the current and Kenyan populations and interpopulation genetic distance analysis strongly supported the presence of an East African cluster that contained the current Ugandan population along with Kenyan Luo and Nandi populations. The congruence of major alleles in different populations would permit consideration of East Africa as an integrated setting when designing and evaluating much needed malaria, tuberculosis, and AIDS vaccines.

Among world populations, Africans present the highest level of genetic diversity (1). This diversity is reflected in multiple markers (2–5), including the human leukocyte antigen (HLA) loci (6), located in the major histocompatibility complex. These highly polymorphic loci encode cell-surface molecules that present peptides sampled from the proteome and mediate self-antigen tolerance and cellular immune responses to pathogens. Class I HLA-A, -B, and -C loci are especially important for their crucial interaction with both T-cell receptors on cytotoxic T-lymphocytes and with killer immunoglobulin-like receptors on the surface of natural killer cells, thus forming a bridge between innate and adaptive immunity against viruses and intracellular parasites (7–9).

While only few genetic lineages at the HLA loci tend to be represented in populations outside of Africa, most of the existing allele families can be found in African populations (10). This is consistent with an evolutionary model proposing that modern humans originated in East Africa and migrated out of Africa to other regions of the world (11), supported by molecular analyses of HLA and other genetic markers (12) and archeological evidence (13).

The complex genetic makeup of the HLA loci in world populations also bears the marks of the history of each group (14), which include several waves of migration, different levels of admixture with other populations, changes in their effective population size and the strong selective pressure exerted by numerous pathogens in varied ecosystems (6). Additionally, the genetic diversity at HLA loci can condition the complexity of the antigens that would need to be included in preventive vaccines against epidemic diseases that affect

vast regions of the world. The application of high-resolution HLA typing has provided a great insight on the evolutionary processes that have shaped some global populations, and is used to understand immune responses mounted against pathogens. Nevertheless, and with few exceptions (10, 15-21), the class I HLA genetic diversity of sub-Sahara African populations remains incompletely described.

In East Africa, three populations from Kenya (10, 17) and one from Uganda (10) have been previously studied by high-resolution class I HLA genotyping. All of these groups were characterized by an extensive genetic diversity and they shared common alleles and haplotypic blocks in the studied loci. The Ugandan population, however, showed a much higher level of genetic complexity and contained alleles and haplotypes that are otherwise not found in other autochthonous African populations. Genetic distance analysis performed by Cao et al. (10) evidenced the Ugandan population as the most divergent group among the whole set of studied African groups, in fact more akin to the admixed African-Americans or Caucasians.

The objective of this study was to further characterize the genetic composition of the Ugandan population in order to better define its relationship with other African groups. Through a detailed comparison of class I HLA variation of a native population from Kampala (Uganda) with previously reported world populations, we established its close relationship with Kenyan populations, with which it forms a subcluster within the main African radiation.

A representative sample of 175 individuals from Kampala (latitude 00°20′N and longitude 32°30′E), the political and economic capital of Uganda, was studied. They corresponded to volunteers who were screened for participation in a phase I clinical trial to evaluate the safety and immunogenicity of a multiclade human immunodeficiency virus (HIV)-1 DNA Plasmid Vaccine at the Makerere University/Walter Reed Project (22). Sixty-six percent of the individuals were male, and the median age was 26 years (range: 18–40). All the individuals were black Africans from Kampala, but no other specific ethnic affiliation was recorded. All of the volunteers provided informed consent, and the study was reviewed and approved by the Ugandan National Council for Science and Technology, the US National Institutes of Health, and the US Army Medical Research and Material Command institutional review boards.

High-resolution typing of class I HLA-A, -B, and -C loci was performed by DNA sequence-based typing, by PCR amplification and dye-terminator nucleotide sequencing of exons 2, 3, and 4 (HLA-A, and -B) or exons 2 and 3 (HLA-C), with ambiguous types being resolved to four digits according to IMGT 2.16.0 (http://www.ebi.ac.uk/imgt/hla/ambig.html) (see Materials and methods, *Supporting Information* for details). Thirty-three HLA-A, 40 HLA-B, and 23 HLA-C alleles were identified (Table 1), including a new variant of HLA-C, temporarily named Cw*06New (GenBank accession number: EU707573). Cw*06New differs from Cw*0602 at amino acid 165, in exon 3, where Cw*0602 has a valine (encoded by GTG) and Cw*06new has a methionine (encoded by ATG). Note that no other reported class I molecule has a methionine at position 165; this position is otherwise conserved in all class I HLA molecules (23). The individual bearing the new allele carried the following class I HLA genotype: A*6601/A*7401:B*1303/B*5802:Cw*0602/Cw*06New.

Table 1 Allele frequencies for HLA-A, B, and -C in Kampala, Uganda (2n = 350)

HLA-A	Allele frequency	HLA-B	Allele frequency	HLA-C	Allele frequency
A*0101	0.0657	B*0702	0.0457	Cw*0210	0.0829
A*0102	0.0057	B*0705	0.0029	Cw*0302	0.0286
A*0103	0.0029	B*0801	0.0429	Cw*0304	0.0629
A*0109	0.0029	B*1302	0.0171	Cw*0401	0.1600
A*0123	0.0029	B*1303	0.0029	Cw*0404	0.0086
A*0201	0.1143	B*1401	0.0029	Cw*0407	0.0200
A*0202	0.0371	B*1402	0.0314	Cw*0501	0.0057
A*0205	0.0143	B*1503	0.0857	Cw*0602	0.1914
A*0214	0.0200	B*1510	0.0514	Cw*06New[a]	0.0029
A*0301	0.0429	B*1516	0.0029	Cw*0701	0.1114
A*2301	0.0914	B*1531	0.0029	Cw*0702	0.0371
A*2402	0.0086	B*1537	0.0029	Cw*0704	0.0171
A*2601	0.0086	B*1801	0.0171	Cw*0706	0.0029
A*2612	0.0057	B*1803	0.0114	Cw*0718	0.0514
A*2901	0.0143	B*3501	0.0114	Cw*0802	0.0400
A*2902	0.0514	B*3502	0.0029	Cw*0804	0.0086
A*3001	0.0629	B*3503	0.0029	Cw*1203	0.0057
A*3002	0.0829	B*3701	0.0057	Cw*1402	0.0029
A*3004	0.0114	B*3910	0.0029	Cw*1505	0.0114
A*3009	0.0057	B*3924	0.0029	Cw*1601	0.0486
A*3101	0.0086	B*4012	0.0086	Cw*1602	0.0029
A*3104	0.0086	B*4101	0.0114	Cw*1701	0.0686
A*3201	0.0086	B*4102	0.0029	Cw*1801	0.0286
A*3303	0.0143	B*4201	0.0629		
A*3402	0.0200	B*4202	0.0029		
A*3601	0.0286	B*4403	0.0057		
A*6601	0.0657	B*4415	0.0171		
A*6602	0.0057	B*4501	0.1086		
A*6603	0.0029	B*4703	0.0057		
A*6801	0.0200	B*4901	0.0543		
A*6802	0.0657	B*5001	0.0029		
A*7401	0.0943	B*5101	0.0086		
A*7403	0.0057	B*5301	0.1057		
		B*5702	0.0114		
		B*5703	0.0286		
		B*5801	0.0600		
		B*5802	0.1029		
		B*7301	0.0029		
		B*8101	0.0400		
		B*8202	0.0086		

HLA, human leukocyte antigen.

[a] New variant of HLA-C temporarily named Cw*06New (GenBank accession number: EU707573).

The overall distribution of genotypes and the observed heterozygosity at each of the HLA class I loci did not show

significant deviations from those expected under Hardy–Weinberg (Table S1, *Supporting Information*). For each of the loci, the observed homozygosity F statistic was lower than expected under neutrality, but the difference was only significant for the HLA-A locus ($P = 0.0228$), which implies that this locus is under balancing selection in the studied population. These results are in concordance with previous reports from sub-Saharan African populations (24), but the causal agents that are responsible for this phenomenon remain yet to be defined.

Table 1 shows the allelic frequencies for the HLA-A, -B, and -C loci. Throughout the studied HLA loci, there was an overall high concordance in the major alleles present in the current Ugandan population and the previously published Kenyan populations (10), and a lower agreement with the previously reported sample set from Uganda. In the HLA-A locus, the same set of alleles (namely A*0101, A*0201, A*0301, A*2301, A*2902, A*3001, A*3002, A*6601, A*6802, and A*7401) corresponded to the 10 most common variants in both the current Ugandan population and Kenyan Luo and were represented as 7 of the 10 most frequent alleles in Kenyan Nandi and the previously reported Ugandan sample set. In the latter, there was a significant over-representation of alleles A*0201, A*2402, and A*1101 compared with the other East African populations (Fisher's exact test $P < 0.01$). Of note, the previously published Ugandan sample set is the only population in sub-Saharan Africa where allele A*1101 has been reported at an allele frequency (AF) of 0.0429 (10); this allele is abundant in Asian populations and somewhat represented in most non-African populations (25). In the HLA-B locus, there was a concordance of the six most frequent alleles (i.e. B*1503, B*4201, B*4501, B*5301, B*5801, and B*5802) among the current Ugandan population and the Luo and Nandi Kenyan populations, but these alleles exhibited intermediate-to-low representation in the previously reported Ugandan sample set. The under-representation of B*4201 and B*5802 was statistically significant ($P < 0.0005$ and $P < 0.03$, respectively). However, alleles B*0702 and B*0801, that were, respectively, the second and third most frequent variants in the previously reported Ugandan sample set, and tend to predominate in Caucasian populations, exhibited only intermediate-to-low representation in the other East African populations. In the same vein, allele B*4402, which is abundant in Caucasian populations, has only been detected in sub-Saharan Africa in the previously reported Ugandan population (AF = 0.0342) (25) and was absent from the current Ugandan population. In the HLA-C locus, the allelic frequencies of the major variants showed a high degree of agreement between the current Ugandan population and the other East African populations, with the exception of Cw*1701, which was significantly underrepresented in the previously reported Ugandan sample set ($P < 0.02$).

The level of genetic complexity, measured at each locus as the number of alleles needed to provide a given coverage of the sampled alleles, varied among the East African populations (Figure S1, *Supporting Information*). The current Ugandan population showed levels of genetic complexity that virtually traced with those of the Luo and Nandi Kenyan populations. However, the previously reported Ugandan sample set exhibited a higher degree of complexity than all the other East African populations along the three loci, a difference that was statistically significant in the HLA-B (Kaplan–Meier survival logrank test, $P < 0.0001$) and -C loci ($P < 0.01$). For example, in the Luo, Nandi, and the current Ugandan populations, the 7 and 13 most frequent HLA-B alleles provided, respectively, a coverage of 50% and 75% of the sampled alleles, whereas 10 and 19 alleles were needed to provide a similar coverage in the previously reported Ugandan sample set.

Using the exact G-test of population differentiation (26), we were able to reject the null hypothesis that across all loci the alleles and genotypes reported in East African sample sets, including the current one, were drawn from the same distribution in all populations ($P < 0.000001$). In a locus-by-locus analysis, the null hypothesis was rejected for all populations pairs except from the HLA-A alleles and genotypes in the Kenyan Luo *vs* the current Ugandan population ($P = 0.15$).

Significant linkage disequilibrium (LD) was observed among all pairs of the studied HLA loci ($P < 0.0001$). The LD reached similar levels to those reported in other East African populations for HLA-A/-B ($D' = 0.68$ and Wn $= 0.45$), HLA-A/-C ($D' = 0.60$ and Wn $= 0.39$), and HLA-B/-C ($D' = 0.90$ and Wn $= 0.80$) (10). The strongest LD was observed among the HLA-B and -C loci. Overall, 77 different haplotypes were inferred, and the 31 haplotypes found in at least three individuals are presented in Table 2. All the HLA-B/-C haplotypes present at frequencies greater than 0.020 were found in at least one other East African population (10). The five most frequent haplotypes in the current population (i.e. B*5802:Cw*0602, B*5301:Cw*0401, B*1503:Cw*0210, B*4201:Cw*1701, and B*4501:Cw*0602) provided, in aggregate, a 39% population coverage, and, with the exception of B*4501:Cw*0602, coincided with the main haplotypes in the previously reported Luo and Nandi populations from Kenya, where they provided comparable population coverage (30% and 35%, respectively) (10). These haplotypes, with the exception of the widely distributed African haplotype B*4201:Cw*1701, were also represented in the previously reported Ugandan sample set, but they only covered 16% of the population (10). The association between alleles from the HLA-A and -B, and from the HLA-A and -C loci was weaker than the HLA-B and -C, and showed a high degree of atomization (Table S2, *Supporting Information*). It was previously noted that the HLA-A/-B/-C haplotypic variation in African populations is large (10). Very few

Table 2 Common HLA-B/-C, haplotypes in Kampala, Uganda (2n = 350)

HLA-B	HLA-C	Haplotypic frequency	D′
B*5802	Cw*0602	0.1029	1.0000
B*5301	Cw*0401	0.0943	0.9303
B*1503	Cw*0210	0.0743	0.8078
B*4501	Cw*0602	0.0600	0.4715
B*4201	Cw*1701	0.0600	0.6696
B*4901	Cw*0701	0.0543	0.5800
B*1510	Cw*0304	0.0486	0.5451
B*4501	Cw*1601	0.0429	0.4519
B*1402	Cw*0802	0.0314	0.3628
B*0702	Cw*0702	0.0286	0.3231
B*8101	Cw*0401	0.0257	0.2322
B*5801	Cw*0718	0.0226	0.2342
B*5801	Cw*0302	0.0200	0.2199
B*4415	Cw*0407	0.0171	0.2020
B*5703	Cw*0701	0.0140	0.1299
B*1302	Cw*0602	0.0114	0.0980
B*1803	Cw*0401	0.0114	0.1154
B*0702	Cw*0718	0.0086	0.0748
B*4101	Cw*0701	0.0086	0.0877
B*5301	Cw*0602	0.0086	−0.1403
B*0801	Cw*0702	0.0086	0.0839
B*4012	Cw*0404	0.0086	0.1022
B*1801	Cw*0704	0.0086	0.0995
B*5801	Cw*0401	0.0086	−0.0124
B*0801	Cw*0718	0.0086	0.0765
B*8202	Cw*0302	0.0086	0.1001
B*0801	Cw*0701	0.0086	0.0456
B*8101	Cw*1801	0.0086	0.0893
B*5702	Cw*1801	0.0086	0.0991
B*3501	Cw*0401	0.0086	0.0811
B*0801	Cw*0304	0.0086	0.0706

HLA, human leukocyte antigen.

haplotypes are extended throughout many populations, and only a few more are shared among neighboring groups. Consistent with this is the observation that the most common extended HLA-A/-B/-C haplotypes in the current Ugandan population were virtually absent from other East African settings, and just a few of the extended haplotypes were simultaneously represented as major blocks in the current Ugandan populations and other East African groups, namely A*0201:B*1503:Cw*0210, A*0202:B*5802:Cw*0602, A*3001:B*4201:Cw*1701, A*3601:B*5301:Cw*0401, A*6601:B*5802:Cw*0602, and A*7401:B*1503:Cw*0210. Of note, some of the haplotypes that were reported in the previously characterized Ugandan sample set at frequencies greater than 0.01 were also absent from other African groups, but were frequent in Caucasians (10, 25). These haplotypes were virtually not found in the current population.

Pairwise interpopulation genetic distances, which are a measure of the level of overlap of alleles between populations, were estimated between the current Ugandan population and all other published sample sets, using the definition proposed by Cavalli-Sforza and Bodmer (27, 28) (see Materials and methods, *Supporting Information*). Figure 1A shows the nonrooted dendrogram built using the matrix of genetic distances combining HLA-A, -B, and -C loci. A clear cluster that contained only sub-Saharan African populations, or admixed world populations that have a significant African component, was evident (see inset in Figure 1). Moreover, within the sub-Saharan African cluster, an East African subcluster was apparent, where the current Ugandan population was embedded. In contrast, the previously reported Ugandan sample set exhibited an outlier behavior, being located outside the main sub-Sahara African radiation, in a region of the tree mostly occupied by admixed world populations with a significant African component [i.e. African-Americans from the United States (29) and an admixed population from Brazil (30)]. The outlier location of the Ugandan sample persisted after repeating the analysis with the removal of each of the populations in the matrix, one at a time (data not shown). The presence of an East African subcluster, the inclusion of the current Ugandan population within this subcluster, and the outlier behavior of the previously reported Ugandan sample set were still evidenced when genetic distances were computed for each locus separately (Figure S2, *Supporting Information*).

When principal component analysis (PCA) based on HLA-A, -B, and -C loci variation was employed, the first and second principal components explained 78.2% of the variance in the sample set (Figure 1B). The PCA corroborated the separation between sub-Saharan African and non-African populations, and reiterated the main results of the genetic distance analysis: the inclusion of the current Ugandan population within the African cluster, and the significant exclusion of the previously reported Ugandan sample set from this group (see ellipse demarking the 99% confidence interval for the sub-Saharan African cluster).

In summary, the obtained results compellingly support the concept that Ugandans are integrated members of a major East African genetic cluster. The strong association between Ugandans and other East African populations is supported by (a) the high degree of overlap among the alleles present at all of the studied loci, (b) the concordance on the sets of alleles that constituted the major variants, (c) the similar levels of genetic complexity at each locus, evidenced in the cumulative frequency analyses, and (d) the reiteration of predominant HLA-B/-C haplotypic blocks in the Ugandan and Kenyan populations. Genetic distance analyses and PCA supported the conclusion that, at least in the light of the available information regarding Africa, the closest relatives to Ugandan populations are to be found among Kenyan groups.

Our results contrast with a previous report from Uganda (10), which led to the authors to argue that 'the Ugandans

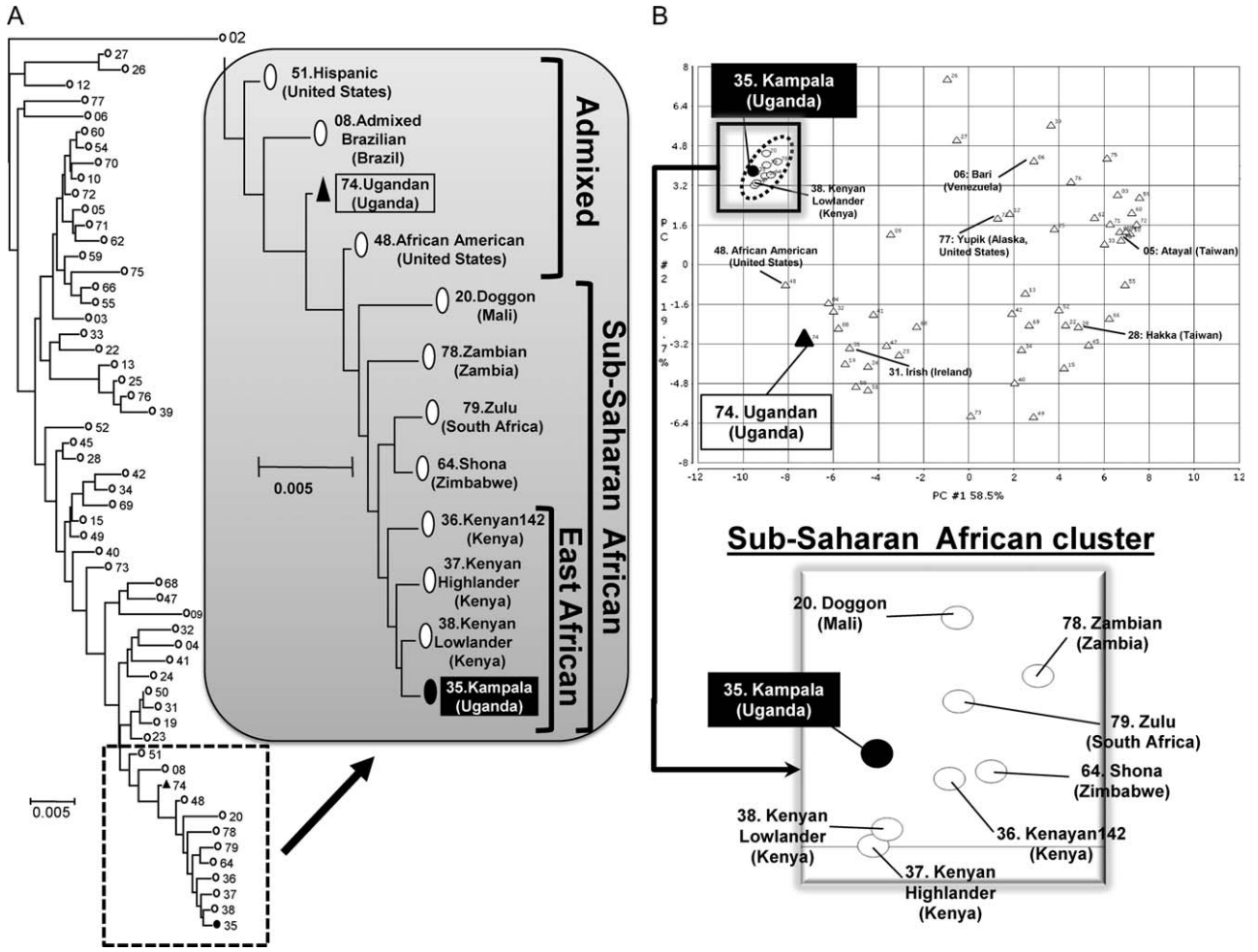

Figure 1 Location of the current Ugandan population in the context of global class I human leukocyte antigen (HLA) genetic diversity. (A) The nonrooted dendrogram was built using combined HLA-A, -B, and -C allele frequency-derived genetic distances, as described in *Materials and methods*. Open circles depict reference world populations retrieved from the dbMHC (25), a black circle represents the currently studied population [Kampala (Uganda)], and the black triangle shows the previously characterized Ugandan sample set [Ugandan (Uganda)] (10). The sections of the dendrograms containing the Ugandan populations, other sub-Saharan African, and admixed world populations sets are shown in detail. World populations are named as represented in the dbMHC (25). Other populations frequently referenced throughout the text include: Luo [Kenyan Lowlander(Kenya)] and Nandi [Kenyan Highlander(Kenya)] (10). (B) Principal component analysis based on genetic distances in HLA-A, -B, and -C loci. In interest of clarity, only the first and second principal components are shown, which account for 78.2% of the variance. The dotted ellipse demarks the 99% confidence interval for sub-Saharan African populations, excluding the previously reported Ugandan sample. The composition of this cluster is shown in detail in the inset, at the bottom of the figure. A black circle depicts the current Ugandan population, open circles show other sub-Saharan populations, open triangles show other world populations, and a black triangle depicts the previously reported Ugandan sample set (10). Populations are labeled as follows: 02: Amerindian (United States), 03: Ami97 (Taiwan), 04:Arab Druze (Israel), 05: Atayal (Taiwan), 06: Bari (Venezuela), 08: Brazilian Admixed (Brazil), 09: Bulgarian (Bulgaria), 10: Bunun (Taiwan), 12: Canoncito (New Mexico, United States), 13: Cape York (Australia), 15: Chinese (China), 20: Doggon (Mali), 22: Filipino (Philippines), 23: Finn90 (Finland), 24: Georgian (Georgia), 25: Groote Eylandt (Australia), 26: Guarani-Kaiowa (Brazil), 27: Guarani-Nandewa (Brazil), 28: Hakka (Taiwan), 31: Irish (Ireland), 32: Israeli Jews (Israel), 33: Ivatan (Philippines), 34: Javanese Indonesian (Singapore), 35: Kampala (Uganda) [*Current work*], 36: Kenyan142 (Kenya), 37: Kenyan Highlander [*Nandi*] (Kenya), 38: Kenyan Lowlander [*Luo*] (Kenya), 39: Kimberley (Australia), 40: Korean200 (South Korea), 41: Kurdish (Georgia), 42: Malay (Singapore), 45: Minnan (Taiwan), 47: New Delhi (India), 48: African American (United States), 49: Asian American (United States), 50: Caucasian (United States), 51: Hispanic (United States), 52: Okinawan (United States), 54: Paiwan51 (Taiwan), 55: Pazeh (Taiwan), 59: Puyuma49 (Taiwan), 60: Rukai (Taiwan), 62: Saisiat (Taiwan), 64: Shona (Zimbabwe), 66: Siraya (Taiwan), 68: Tamil (South Africa), 69: Thai (Singapore), 70: Thao (Taiwan), 71: Toroko (Taiwan), 72: Tsou (Taiwan), 73: Tuva (Tuva), 74: Ugandan [*previously published*] (Uganda), 75: Yami (Taiwan), 76: Yuendumu (Australia), 77: Yupik (Alaska, United States), 78: Zambian (Zambia), 79: Zulu (South Africa).

present a subset of the variability that is present in other sub-Saharan African populations'. The differences between the previous and the current study probably reflect different enrollment strategies in each of them. While the current study was only focused on Black Africans, the former might have included not only

indigenous East African individuals but also people from other descents living in Uganda, as suggested by Cao et al. (10). Their finding in Uganda of alleles that are abundant in Caucasian and Asian populations, but are virtually absent from other African groups, further supports this explanation. The population currently under study was drawn from Kampala, the political and economic capital of Uganda. This city is located in the Buganda region, with a predominance of Baganda peoples who speak Luganda (31), but is also the convergence center for the many different African ethnic components that are present throughout the country. Because of the demographic level at which the current population was defined, the information hereby reported may have limited application for anthropological studies. However, the current data on HLA allele and haplotype distributions are highly relevant for molecular epidemiological studies and public health interventions (12).

The fact that the same HLA alleles and haplotypes seem to be represented throughout Uganda and Kenya, although they may be differently balanced, provides the basis for evaluating the relative contributions of pathogen-driven selection (32), human migration (33), and changes in effective population sizes (34) in the current genetic makeup of populations residing in different areas of East Africa. Additionally, as most of the East African populations are heavily affected by highly variable pathogens (35), especially HIV (36), they provide a unique setting for exploration of the model that proposes that host genetic diversity is a major selecting force molding pathogen genetic variation (37). What is more, the congruence of major alleles in different populations would permit consideration of East Africa as an integrated setting when designing and evaluating much needed vaccines to prevent major epidemic diseases affecting the whole region, including malaria, tuberculosis, and AIDS (38, 39).

Acknowledgments

The authors would like to thank Dr Josephine Cox and Elana Dicht at the Cellular Immunology Laboratory, US Military HIV Research Program, Rockville, MD for their assistance in the expansion of the EBV-transformed B-cell lines. This work was supported through a cooperative agreement between the Henry M. Jackson Foundation for the Advancement of Military Medicine and the U.S. Department of Defense, by the National Institute for Allergy and Infectious Diseases, National Institutes of Health ('HIV Vaccine Research and Development—Project 2' Y1-AI-2642-11). The views and opinions expressed herein do not necessarily reflect those of the U.S. Army, the Department of Defense or the National Institutes of Health.

References

1. International HapMap Consortium. A haplotype map of the human genome. *Nature* 2005: **437**: 1299–320.
2. Cavalli-Sforza LL, Menozzi P, Piazza A. *The History and Geography of Human Genes. Princeton: Princeton University Press, 1994*.
3. Chen YS, Torroni A, Excoffier L, Santachiara-Benerecetti AS, Wallace DC. Analysis of mtDNA variation in African populations reveals the most ancient of all human continent-specific haplogroups. *Am J Hum Genet* 1995: **57**: 133–49.
4. Jorde LB, Watkins WS, Bamshad MJ et al. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. *Am J Hum Genet* 2000: **66**: 979–88.
5. Zietkiewicz E, Yotova V, Jarnik M et al. Nuclear DNA diversity in worldwide distributed human populations. *Gene* 1997: **205**: 161–71.
6. Prugnolle F, Manica A, Charpentier M, Guegan JF, Guernier V, Balloux F. Pathogen-driven selection and worldwide HLA class I diversity. *Curr Biol* 2005: **15**: 1022–7.
7. Rimmelzwaan GF, Fouchier RA, Osterhaus AD. Influenza virus-specific cytotoxic T lymphocytes: a correlate of protection and a basis for vaccine development. *Curr Opin Biotechnol* 2007: **18**: 529–36.
8. Stenger S. Cytolytic T cells in the immune response to mycobacterium tuberculosis. *Scand J Infect Dis* 2001: **33**: 483–7.
9. Parham P, MHC class I molecules and KIRs in human history, health and survival. *Nat Rev Immunol* 2005: **5**: 201–14.
10. Cao K, Moormann AM, Lyke KE et al. Differentiation between African populations is evidenced by the diversity of alleles and haplotypes of HLA class I loci. *Tissue Antigens* 2004: **63**: 293–325.
11. Templeton A. Out of Africa again and again. *Nature* 2002: **416**: 45–51.
12. Tishkoff SA, Williams SM. Genetic analysis of African populations: human evolution and complex disease. *Nat Rev Genet* 2002: **3**: 611–21.
13. Lahr MM, Foley RA. Towards a theory of modern human origins: geography, demography, and diversity in recent human evolution. *Am J Phys Anthropol* 1998: **107** (Suppl 27): 137–76.
14. Parham P, Ohta T. Population biology of antigen presentation by MHC class I molecules. *Science* 1996: **272**: 67–74.
15. Ellis JM, Mack SJ, Leke RF, Quakyi I, Johnson AH, Hurley CK. Diversity is demonstrated in class I HLA-A and HLA-B alleles in Cameroon, Africa: description of HLA-A*03012, *2612, *3006 and HLA-B*1403, *4016, *4703. *Tissue Antigens* 2000: **56**: 291–302.
16. Louie LG, Hartogensis WE, Jackman RP et al. Mycobacterium tuberculosis/HIV-1 coinfection and disease: role of human leukocyte antigen variation. *J Infect Dis* 2004: **189**: 1084–90.
17. Luo M, Embree J, Ramdahin S, et al. HLA-A and HLA-B in Kenya, Africa: allele frequencies and identification of HLA-B*1567 and HLA-B*4426. *Tissue Antigens* 2002: **59**: 370–80.

18. Middleton D, Williams F, Meenagh A et al. Analysis of the distribution of HLA-A alleles in populations from five continents. *Hum Immunol* 2000: **61**: 1048–52.
19. Sanchez-Mazas A, Steiner QG, Grundschober C, Tiercy JM. The molecular determination of HLA-Cw alleles in the Mandenka (West Africa) reveals a close genetic relationship between Africans and Europeans. *Tissue Antigens* 2000: **56**: 303–12.
20. Tang J, Naik E, Costello C et al. Characteristics of HLA class I and class II polymorphisms in Rwandan women. *Exp Clin Immunogenet* 2000: **17**: 185–98.
21. Williams F, Meenagh A, Darke C et al. Analysis of the distribution of HLA-B alleles in populations from five continents. *Hum Immunol* 2001: **62**: 645–50.
22. Eller MA, Eller LA, Opollo MS et al. Induction of HIV-specific functional immune responses by a multiclade HIV-1 DNA vaccine candidate in healthy Ugandans. *Vaccine* 2007: **25**: 7737–42.
23. Robinson J, Waller MJ, Parham P et al. IMGT/HLA and IMGT/MHC: sequence databases for the study of the major histocompatibility complex. *Nucleic Acids Res* 2003: **31**: 311–4.
24. Solberg OD, Mack SJ, Lancaster AK et al. Balancing selection and heterogeneity across the classical human leukocyte antigen loci: a meta-analytic review of 497 population studies. *Hum Immunol* 2008: **69**: 443–64.
25. Kitts A, Feolo M, Helmberg W. The major histocompatibility complex database, dbMHC. In: National Center for Biotechnology Information NIH, ed. *The NCBI Handbook. Bethesda: National Center for Biotechnology Information NIH, 2003, p.1-29*.
26. Raymond M, Rousset F. An exact test for population differentiation. *Evolution* 1995: **49**: 1283–6.
27. Cavalli-Sforza LL, Bodmer WF. *The Genetics of Human Populations. San Francisco: W.H. Freeman, 1971.*
28. Cavalli-Sforza LL, Edwards AW. Phylogenetic analysis. Models and estimation procedures. *Am J Hum Genet* 1967: **19**: 233–57.
29. Cao K, Hollenbach J, Shi X, Shi W, Chopek M, Fernandez-Vina MA. Analysis of the frequencies of HLA-A, B, and C alleles and haplotypes in the five major ethnic groups of the United States reveals high levels of diversity in these loci and contrasting distribution patterns in these populations. *Hum Immunol* 2001: **62**: 1009–30.
30. Louzada-Junior P, Deghaide NH, Araujo MB, Smith AG, Kraemer MHS, Donadi EA. Brazilian (admixed, African and European) from the Northeast region of State of São Paulo, Brazil. Anthropology/human genetic diversity population reports. In: Hansen A, ed. *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference. Seattle: IHWG Press, 2007, 647–8.*
31. Gordon R Jr. *Ethnologue: Languages of the World*, 5th edn. *Dallas: SIL International, 2005*. Online version: http://www.ethnologue.com/
32. Jeffery KJ, Bangham CR. Do infectious diseases drive MHC diversity? *Microbes Infect* 2000: **2**: 1335–41.
33. Spinola H, Brehm A, Williams F, Jesus J, Middleton D. Distribution of HLA alleles in Portugal and Cabo Verde. Relationships with the slave trade route. *Ann Hum Genet* 2002: **66**: 285–96.
34. Muirhead CA. Consequences of population structure on genes under balancing selection. *Evolution* 2001: **55**: 1532–41.
35. GIDEON database. *Global Infectious Diseases and Epidemiology Network*. Los Angeles. http://www.gideononline.com
36. Joint United Nations Programme on HIV/AIDS (UNAIDS) and World Health Organization (WHO). *AIDS epidemic update: December 2007. Geneva: WHO, 2007.*
37. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science* 2002: **296**: 1439–43.
38. The global fund to fight AIDS, tuberculosis and malaria. http://www.theglobalfund.org
39. Bill & Melinda Gates Foundation. http://www.gatesfoundation.org

Supporting information

Additional Supporting Information may be found in the online version of this article.

Materials and methods

Figure S1 Genetic complexity in class I human leukocyte antigen A, B, and C loci in East African populations. For each population and at each locus, alleles were sorted in decreasing order and cumulative frequencies were calculated. Depicted populations are the same ones as depicted in supplemental figure 1.

Figure S2 Subanalysis of the location of the current Ugandan population in the context of global class I human leukocyte antigen (HLA) genetic diversity. The nonrooted dendrograms were built using unilocus (a) HLA-A, (b)-B, and (c)-C allele frequency and (d) HLA-B:C haplotype-derived genetic distances, as described in *Materials and methods*. Open circles depict reference world populations retrieved from the dbMHC (1), a black circle represents the currently studied population [Kampala (Uganda)], and the black triangle shows the previously characterized Ugandan sample set [Ugandan (Uganda)] (2). The sections of the dendrograms containing the Ugandan populations, other sub-Saharan African, and admixed world populations sets are shown in detail. World populations are named as represented in the dbMHC (1). Other populations frequently referenced throughout the text include: Luo [Kenyan Lowlander(Kenya)] and Nandi [Kenyan Highlander(Kenya)] (2). Populations are labeled as follows: 01: American Samoa (American Samoa, United States), 02: Amerindian (United States), 03: Ami97 (Taiwan), 04:Arab Druze (Israel), 05: Atayal (Taiwan), 06: Bari (Venezuela), 07: Brazilian (Brazil), 08: Brazilian Admixed (Brazil), 09: Bulgarian (Bulgaria), 10: Bunun (Taiwan), 11: Buriat (Russia), 12: Canoncito (New Mexico, United States), 13: Cape York (Australia), 14: Chaouya (Morocco), 15:

Chinese (China), 16: Croatian (Croatia), 17: Cuban Mulatto (Cuba), 18: Cuban Caucasian (Cuba), 19: Czech (Czech Republic), 20: Doggon (Mali), 21: East Timorese (Indonesia), 22: Filipino (Philippines), 23: Finn90 (Finland), 24: Georgian (Georgia), 25: Groote Eylandt (Australia), 26: Guarani-Kaiowa (Brazil), 27: Guarani-Nandewa (Brazil), 28: Hakka (Taiwan), 29: Han-Chinese149 (Singapore), 30: Han-Chinese572 (China), 31: Irish (Ireland), 32: Israeli Jews (Israel), 33: Ivatan (Philippines), 34: Javanese Indonesian (Singapore), 35: Kampala (Uganda) [current study], 36: Kenyan142 (Kenya), 37: Kenyan Highlander [Nandi] (Kenya), 38: Kenyan Lowlander [Luo] (Kenya), 39: Kimberley (Australia), 40: Korean200 (South Korea), 41: Kurdish (Georgia), 42: Malay (Singapore), 43: Mandenka (Senegal), 44: Metalsa (Morocco), 45: Minnan (Taiwan), 46: Moluccan (Indonesia), 47: New Delhi (India), 48: African American (United States), 49: Asian American (United States), 50: Caucasian (United States), 51: Hispanic (United States), 52: Okinawan (United States), 53: Omani (Oman), 54: Paiwan51 (Taiwan), 55: Pazeh (Taiwan), 56: Pima99 (Arizona), 57: PNG Highlander (Papua New Guinea), 58: PNG Lowlander95 (Papua New Guinea), 59: Puyuma49 (Taiwan), 60: Rukai (Taiwan), 61: Ryukuan (Japan), 62: Saisiat (Taiwan), 63: Seri (Mexico), 64: Shona (Zimbabwe), 65: Singapore-Chinese (Singapore), 66: Siraya (Taiwan), 67: South Indian (India), 68: Tamil (South Africa), 69: Thai (Singapore), 70: Thao (Taiwan), 71: Toroko (Taiwan), 72: Tsou (Taiwan), 73: Tuva (Tuva), 74: Ugandan [previously published] (Uganda), 75: Yami (Taiwan), 76: Yuendumu (Australia), 77: Yupik (Alaska, United States), 78: Zambian (Zambia), 79: Zulu (South Africa). At each locus, only populations that had data available were included in the analysis.

Table S1 Comparison between the observed heterozygosity and homozygosity, and their expected values under Hardy–Weinberg equilibrium

Table S2 Common HLA-A/-B, HLA-A/-C, and HLA-A/-B/-C haplotypes in Kampala, Uganda (2n = 350)

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.