Supporting Online Material for

The Shaping of Modern Human Immune Systems by Multiregional Admixture with Archaic Humans

Laurent Abi-Rached, Matthew J Jobin, Subhash Kulkarni, Alasdair McWhinnie,
Klara Dalva, Loren Gragert, Farbod Babrzadeh, Baback Gharizadeh, Ma Luo,
Francis A Plummer, Joshua Kimani, Mary Carrington, Derek Middleton,
Raja Rajalingam, Meral Beksac, Steven GE Marsh, Martin Maiers, Lisbeth A Guethlein,
Sofia Tavoularis, Ann-Margaret Little, Richard E Green, Paul J Norman and
Peter Parham*

*To whom correspondence should be addressed. E-mail: peropa@stanford.edu

This PDF file includes

Materials and Methods
Figs. S1 to S26
References

TABLE OF CONTENTS (1)

Materials and Methods

TABLE OF CONTENTS (2)

Supplementary Figures

Supporting Online Material

Materials and Methods

Part I. *HLA-B*73*

I.1 Characterization of the *HLA-B*73:01* haplotype and of a chimpanzee allele of the *B*73* lineage. An Epstein-Barr virus-transformed B-cell line derived from a *HLA-B*73:01* donor was used as source of genomic DNA for the construction of a cosmid library, using the SuperCos 1 Cosmid Vector Kit (Stratagene, Santa Clara, CA) per manufacturer's instructions. The library was screened with a ^{32}P-labeled *B*73:01* cDNA probe, and hybridizing clones analyzed by end-sequencing and targeted amplification of *B*73:01* using primers B*73-Forward (5'-AACCGTCCTCCTGCTGCTC-3') and B*73-Reverse (5'-GCGACCACAGCTACGGTGA-3'). One cosmid of ~38kb, which contained the complete *B*73:01* gene, flanked by ~27kb of DNA at the 5' end and ~8kb at the 3' end, was selected for shotgun sequencing (TOPO Shotgun Subcloning Kit, Invitrogen, Carlsbad, CA).

Genomic DNA from a panel of 95 chimpanzees was typed by PCR specific for alleles of the *B*73* lineage (*MHC-BII* lineage). PCR primers were TypeF (5'-GCTTAGATGAACATGAAAAGATAG-3') and TypeR (5'-TAGTTGATTGACATGTGGATTA-3').

I.2 Characterization of extended *MHC-B* gene sequences in hominoids. To compare the region of the *HLA-B*73:01* haplotype that is directly derived from the second lineage of *MHC-B* alleles (*MHC-BII*), a region of ~8.5kb around the *MHC-B* gene was characterized in several common chimpanzees and gorillas, and one bonobo. Primers 8.5F (5'-GTCAGGACAAGAGGAACAGAGAAAC-3') and 8.5R (5'-GTTCAGCACCAAGATCACTAGAACC-3') were used in chimpanzees and 8.5GorF (5'-GTCAGGAWAAGAGGAACAGAGAAAC-3') and 8.5R in gorillas. PCR products were isolated by gel electrophoresis, cloned into the pcr2.1-TOPO vector (Invitrogen) and clones from three independent amplifications were sequenced directly with specific overlapping primers.

In addition to these sequences, we used BAC clones to characterize a ~38kb gorilla haplotype carrying *Gogo-B*06:01*, an ~8.5kb orangutan region carrying *Popy-B*03:02*, and the *Popy-C*02:02* gene. BAC clones carrying these genes were obtained by screening the NCBI trace archive database (http://www.ncbi.nlm.nih.gov/Traces/) for *MHC-B* gene-containing BAC clones. Clones 498J12 and 308J7 were isolated from the CHORI-277 gorilla library, together with clones 253H24 and 97B06 from the CHORI-276 orangutan library. Clones were shotgun-sequenced using a Genome Sequencer FLX (Roche/ 454 Life Sciences); samples were prepared and sequenced using the GS Titanium reagents per manufacturer's instructions. The ~6,900 reads we generated were assembled using MIRA (*32*) and the assembly edited with the GAP4 program of the STADEN package (*33*). The assembly was then supplemented by 63 sequence reads generated at the Wellcome Trust Sanger Institute as part of the gorilla reference genome sequencing project. These data can be obtained from the NCBI Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/). The assembly was further supplemented by three sequence reads generated by Washington University School of Medicine; these data were produced by the Genome Institute at Washington University School of Medicine in St. Louis and can be obtained from the NCBI Trace Archive (http://www.ncbi.nlm.nih.gov/Traces/). These 66 reads were obtained by BLAST search (*34*), using as probes the genomic segments with low coverage in our assembly. Those sequence reads that displayed a well-supported mismatch (i.e. a difference at a position with good quality in both directions) with sequence reads in our assembly, likely represented the other *MHC-B* haplotype and were not used. Finishing was performed by PCR amplifications on the two BAC clones and Sanger sequencing of the PCR products. The same approach was used for the orangutan analysis, but sequence reads were from the Washington University Genome Sequencing Center and Baylor College of Medicine (orangutan genome project (*35*)). The names of the animals used for the characterization of these sequences are given in figure S2.

I.3 Sequencing and accession numbers. Sequencing of the *B*73* cosmid clone, the 8.5kb extended chimpanzee and gorilla *MHC-B* genes, as well as finishing for the gorilla and orangutan genomic segments characterized from BAC clones, used an ABI377 DNA sequencer (Applied Biosystems, Carlsbad, CA), a CEQ2000XL (Beckman), and McLab (South San Francisco, CA). Sequences were assembled with the STADEN package (*33*), and finished to high-quality standard: each base pair of the final assembly was minimally covered by three sequences representing two different templates and had a quality score >40-80 (error rate $<1/10^4-1/10^8$). Completed sequences were deposited in GenBank under accession numbers JF974053-70.

I.4 Sequence analysis of the *HLA-B*73:01* haplotype. Haplotype and extended gene sequences were aligned with MAFFT (*36*) and corrected manually. Insertions of >100bp and unique to one sequence, to *MHC-C* or to macaque *MHC-B* sequences were excluded. The pattern of historical recombination in the region was investigated with the program RDP (*37*) and confirmed by Neighbor-Joining (NJ) analyses with MEGA4 (*38*) and alignment inspection. Following this analysis, the alignment was divided into 14 segments, which were each analyzed by three phylogenetic methods: maximum-likelihood (ML), NJ and parsimony. NJ analysis was performed with MEGA4 (*38*) using the Tamura-Nei method with 500 replicates. PAUP*4.0b10 (*39*) and the tree bisection-reconnection branch swapping algorithm were used for parsimony analyses with 500 replicates and heuristic search. ML analyses were performed with RAXML7 (*40*) under the GTR+CAT model with 500 replicates (rapid bootstrapping).

I.5 *MHC-BI-BII* divergence time analysis. Mean and 95% credibility interval values for the divergence time between the *MHC-BI* and *MHC-BII* lineages were estimated using a Bayesian approach with the program MCMCTREE (v4.4c) from the PAML package (*41*). Analysis was conducted with an independent-rates molecular clock, the HKY85+Γ model of nucleotide substitution, on a ~4.4kb segment located 5’ of the *MHC-B* gene (segment #5 of figure S3), and used a ML tree topology (fig. S3). Several fossil calibration points were used for the following speciation events: human-chimpanzee, 6.5-10 million years ago (MYA) (*42*); gorilla-orangutan, >10MYA (*43, 44*); human-gorilla, >10MYA (*45*); human-Old World monkey 23-33.9MYA (*42*).

Priors for the transition/transversion rate ratio (κ) and the shape parameter for gamma rates (α) were set as follows: means were estimated using the BASEML program of the PAML package (*41*) and SD were set so that prior SD would be about two times the posterior SD. Priors for the overall rate parameter (μ) and for the parameter σ2, which specifies how variable the rates are across branches, were set as follows: prior means were set according to the posterior means and prior SD were set to be about four times the posterior SD. Step lengths used in the MCMC algorithm were set so that acceptance proportions were all in the interval 0.2-0.4. Three independent runs were conducted and in each case produced similar results.

I.6 Distribution of *B*73* and *C*15:05*. Allele frequencies for *B*73* and *C*15:05* were obtained from the Allele Frequency Net Database (*12*). Distribution maps were generated using the GMT software package (*46*) and a previously developed script (*47*). Only studies of more than 40 individuals were included: either anthropology studies or studies documenting the donors as unrelated. Data from bone marrow registries were excluded (unless frequencies were established on a group of randomly-selected donors), as well as data from admixed populations (for example African-Americans) or recent migrant populations for which the geographical origin was not well defined (for example 'USA Asian').

I.7 Estimation of the extent of linkage between *B*73:01* and *C*15:05*. *HLA-C*15:05* was discovered as a novel *HLA-C* allele on a haplotype with *HLA-B*73:01* (*13*). To determine how frequently *C*15:05* and *B*73* are on the same haplotype, we obtained data from anthropology studies and bone marrow registries (fig. S7). The former were directly obtained from the authors, while the latter were obtained with permission from the Bone Marrow Donors Worldwide (BMDW) and the National Marrow Donor Program (NMDP) organizations.

HLA-B and *-C* genotyping was performed to confirm the *B*73* type of an African American *B*73* donor using LABType® SSO reagents (One Lambda Inc, Canoga Park, CA) and detection by a Luminex 100 instrument (Luminex corp, Austin TX). High-resolution *HLA-C* typing was performed for five *B*73* donors from the Anthony Nolan Trust using LABType® reagents. Eight unrelated *B*73* donors from Ankara (Turkey) were genotyped for *HLA-B* and *-C* using Lifecodes typing kits (Tepnel Lifecodes, Stamford, CT). Six donors from Kenya and Tanzania were sequence-based genotyped for *HLA-B* and *-C*, as previously described (*48*). A further 120 *B*73* donors from the Canadian OneMatch Stem Cell and Marrow Network were retyped at high resolution for *HLA-B* and *-C* using the AlleleSEQR HLA sequence-based typing (SBT) method (Atria Genetics, South San Francisco, CA), with additional tests using LABType® SSO reagents.

Because bone marrow registry data include low-resolution *HLA-C* typing, for the LD analysis (fig. S7) the *C*15* alleles with no defined subtype were allocated to the *C*15:05* or *C*15:not05* groups based on the *C*15:05*/*C*15:not05* ratio, on a region-by-region basis. This approach allowed us to have an 'Estimated *C*15:05*' category for each geographical region.

HLA-C-B haplotype structures were determined for 46 *B*73*$^+$ individuals lacking *C*15* using an implementation of the Expectation Maximization algorithm that accepts allelic ambiguities as inputs (*49*). Due to the small sample size, only *C-B* haplotypes containing *B*73:01* with count >2 are provided.

I.8 Haplotype data for population simulations. The frequencies of *B*73* and *C*15:05* used in simulations were obtained as indicated above (section I.6), except that data for a few additional African populations were obtained from dbMHC (*50*) (lacking *B*73* and *C*15:05*, these populations were not included in the distribution maps); data are summarized in figure S9. Because anthropology studies commonly investigate a single locus (e.g. *HLA-B* or *HLA-C*), allele frequencies for both *B*73* and *C*15:05* are not always available from the same set of individuals. Thus, to prepare haplotype data for population simulation we used three steps to set the linkage between the *B* and *C* alleles according to the linkage data estimated from anthropology studies that investigated both loci and data from bone marrow donors (fig. S7). First, the total number of haplotypes and the number of *B*73* haplotypes in the simulation were set to match the data observed in populations with *HLA-B* genotyping in west Asia or in Africa (fig. S9A). Second, the number of *C*15:05* haplotypes in these *HLA-B* genotyped populations was calculated using the frequencies observed in *HLA-C* genotyped populations (fig. S9B), on a region-by-region basis. Finally, the extent of linkage between *B*73* and *C*15:05* was set according to linkage data summarized in figure S7 (on a region-by-region basis), but because of the heterogeneity in typing data and the possibility that some bone marrow donors represent related donors, we investigated nine models of linkage between *B*73* and *C*15:05* (fig. S10).

I.9 Population simulations. We compared two alternative models to explain how the *B*73-C*15:05* haplotypes were formed and propagated in western Asia and in Africa: a model of archaic admixture (Model 'a') and a model of African ancestry (Model 'b') (Fig. 1). Comparison of the models was conducted with the rejection-based approximate Bayesian inference and hypothesis testing software REJECTOR (*16*). In a first phase, simulated data were generated for each model using a coalescent simulation (*51*) based on the population demographic events, parameters (see Fig. 1 and section I.10) and haplotype data (fig. S10) supplied for each model. In a second phase, the simulated data were compared to the observed data and either 'accepted' or 'rejected'; this process was repeated several million times.

To compare simulated and observed data, two summary statistics were used: 'Derived Fraction', which is the proportion of derived alleles present in the extant populations, and 'LD', the linkage disequilibrium measured as r^2 (*52*). For each statistic, when the simulated values matched the observed values (with some tolerance) in both western Asia and Africa, the iteration was considered a ‘single accept’; if an iteration produced ‘single accepts’ for both summary statistics, it was treated as a ‘double accept’. Because of the generally low frequencies of *B*73* and *C*15:05*, the tolerance for the ‘Derived Fraction’ statistic was set at 99% (i.e. a range of 0.01-1.99% is tolerated for an observed allele frequency of 1%) and at 50% for the ‘LD’ statistic. A test was also conducted with these two tolerance values halved (fig. S11B).

One model was considered to be significantly favored by the simulations when it gained >19 times the number of double accepts achieved by the alternative model (>95:5 ratio, 5% significance level).

I.10 Parameters for population simulation. For both models a and b, population sizes were considered under exponential growth with effective population sizes and growth rates derived from Pilkington et al. (*53*); recombination was set a 0.44cM/Mb (*31*). The Out-of-Africa migration was set at 3,375 generations ago (~67.5kya), without affecting the size of the African population, and a back-migration into Africa was set so that 10% of African ancestry traced back to western Asia, 500 generations ago (~10kya) (*21*). In both models, *C*15:05* was formed in western Asia 3,250 generations ago (~65kya), so that the two models would only differ regarding where and when *B*73* emerged: in western Asia 2,500 generations ago (~50kya) for model ‘a’ and in Africa 5,250 generations ago (~105kya) for model ‘b’.

I.11 Note: explanation of *HLA* nomenclature. *HLA* alleles are distinguished by a rational nomenclature (*54*) in which the name of the gene is followed by up to four sets of digits separated by colons. The digits before the first colon distinguish groups of alleles that roughly correspond to the serological types first used to define HLA class I polymorphisms. The next set of digits distinguishes subtypes that differ in their amino-acid sequences. Subsequent sets of digits distinguish alleles that differ by synonymous (third set) or intronic (fourth set) differences.

I.12 Note: study approval. This study was approved by the Stanford University administrative panels on human subjects in medical research and laboratory animal care.

Part II. Analysis of the *HLA class I* content of the Denisova and Neandertal genomes

II.1 Isolation and re-mapping of the Neandertal and Denisovan *HLA class I* sequence reads. To investigate Denisovan (*4*) and Neandertal (*3*) *HLA class I* content, we downloaded from the UCSC Genome Browser website (http://genome.ucsc.edu) all sequence reads previously assembled to the human (hg18) and chimpanzee (panTro2) genomes. Using SAMTOOLS 0.1.12a (*55*), we extracted all reads within the coordinates chr6:28,810,000-33,500,000 of hg18 (127,562 reads) and chr6:30,978,159-34,115,302 of panTro2 (185,092 reads), corresponding to an interval including *TRIM27*, *KIFC1* and all classical *HLA*/*MHC* loci.

Re-mapping was conducted using a set of *HLA-A,-B,-C,-E,-F,-G* reference sequences (one sequence for each subtype), as well as sequences for the pseudogenes *HLA-H,-J,-K,-*

L,-P,-S,-T,-U,-V,-W,-X,-Y (*10*). Assembly of *HLA class I* from the Denisovan genome and the three Neandertal genomes was performed using MIRA (*32*). The contigs were investigated and edited using GAP4 of the STADEN package (*33*); only a single copy of the reads mapped both in the hg18 and panTro2 *MHC* regions was kept.

The Neandertal sequences presented a special problem for alignment because of errors characteristic of ancient DNA (*3, 56*), namely a high rate of C to T transitions at the 5' end of reads and an equally high rate of G to A transitions at the 3' end. Identifying relevant reads by sequence similarity searching was hampered by these ancient DNA associated base misincorporations. The Denisova sequence data suffered much less from this problem, because an enzymatic pre-treatment of the DNA (*57*) largely corrected the errors. To generate a more comprehensive list of relevant Neandertal reads, we re-mapped all the primary Neandertal data against the reference panel of *HLA class I* sequences using a custom mapper, *mia*, (available from the authors upon request) with the following command-line parameters: -k 12 –N 500 –S 100. This program incorporates the special position-specific error profile of these data into its scoring scheme for identifying and aligning sequences. Because it performs a full Smith-Waterman alignment this program is not hampered by insertion-deletion differences that can erode the sensitivity of heuristic aligners. With this approach 164 additional Neandertal sequences were identified from the original mapping to human and chimpanzee genomes.

II.2 Characterization of the Denisovan and Neandertal *HLA-A/-B/-C* content. Locus-specific screen: following re-mapping, we screened the assemblies to ensure all reads associated with *HLA-A,-B,* or *-C* were specific to that locus: each sequence read was compared to the reference *HLA class I* dataset (*10*) and any read shared with another locus was discarded. For reads having substitutions absent from the reference dataset, comparison to a larger dataset including all coding sequences (*10*) was conducted. If this approach failed to identify a sequence with the same SNP, BLAST was used (*34*) to examine *MHC class I* in other species. SNPs not represented in the known human and non-human primate *MHC class I* sequences (more than 6,800 sequences (*10, 58*)) were attributed to misincorporation errors made during PCR amplification of ancient DNA and ignored.

Allele-specific screen: *HLA-A,-B,* and *-C* allele content was determined for each archaic individual by comparing all locus-specific sequences to the reference panel of *HLA class I* sequences. For Neandertal *HLA-A/B/C* and Denisovan *HLA-A/C,* this analysis allowed the sequences to be split in three groups: sequences from the first allele, sequences from the second allele, and sequences that could come from either allele (which were discarded). For Denisovan *HLA-B,* at least one allele was novel, and to associate several reads to one allele or the other we used a parsimony approach to minimize differences with modern sequences (fig. S12).

The allele-specific sequences obtained for each gene and archaic human genome were compared to genomic and cDNA sequences from the corresponding gene of modern humans through pairwise comparisons, using MEGA4 (*38*). This analysis allowed a first assessment of Neandertal and Denisovan *HLA-A/B/C* content. This information was used to design a second and more specific re-mapping to ensure capture of all the relevant sequence reads. For this reanalysis the initial set of reference *HLA-A/B/C* genomic sequences, which included one sequence for each of the allele groups, was increased to include all the available genomic sequences for the allele groups that we had determined were present in the Denisovan/Neandertal genomes. Following this remapping we used the same locus and allele-specific determination to refine the allele string.

II.3 Distribution and diversity of Denisovan/Neandertal alleles and haplotypes. Allele frequency data and distribution were investigated as described above (Section I.6). For the study of haplotype distributions, data were also obtained from dbMHC (*50*), and from Melanesian populations (*59*). *HLA* genotype phasing for the Melanesian populations was performed using PHASE 2.1 (*60*, *61*).

Analysis of the diversity of *HLA* haplotypes carrying alleles of Denisovan/Neandertal origin was conducted using the same data sources as described above (*12, 50*). Because low-frequency haplotypes are not always reported in these databases, to investigate haplotype structures in Africa (where haplotypes carrying Denisovan *HLA-C* alleles are rare) we used data from a panel of ~2,400 African-American donors, for whom the haplotypes are well defined (*62*).

II.4 Divergence time analyses for *HLA-A*11, C*12:02* and *C*15.* Estimation of the emergence times of *A*11*, *C*12:02* and *C*15* was performed as described for the *B*73* lineage (section I.5); specifics such as the datasets used or differences with the analysis described in section I.5 are given below for each allele. The phylogenetic tree topologies used for the divergence time analyses and presented in figure S15 were obtained using the methods described in section I.4.

For *HLA-A*11*, a dataset of ~2.2kb representing the genomic segment between intron 3 and the 3'UTR of *HLA-A* was used, because this segment has not been a target for recombination (*63*). Divergence time analysis was conducted as indicated in section I.5, except that prior SD for the overall rate parameter (μ) and for the parameter $\sigma 2$ were set to be about two times the posterior SD.

For *C*15*, we used a dataset of ~3kb representing the complete *HLA-C* gene, except for two recombinant segments in exon 3 (*64*). For *C*12:02*, we used a dataset of ~2.6kb representing a partial *HLA-C* gene sequence, with exon 2 and part of exon 3 excluded due to the presence of two gene conversion events. To avoid recombinants, a minimum number of sequences was used in these analyses. Divergence time analyses were conducted as described in section I.5, except that prior SD for the overall rate parameter (μ) was set to be about twice the posterior SD. The mean and SD priors for the parameter $\sigma 2$ were also set to one divided by the root age (*65*), and the SD for the shape parameter for gamma rates (α) was set to the mean value. In both analyses, the root of the tree corresponds to the separation between humans and orangutans; in addition to a minimum bound of 10MYA (*43, 44*), a maximum bound of 18MYA was used to reflect the probable separation of orangutans from human/African apes after the African continent became combined with Eurasia (*66*).

II.5 Frequency of *HLA-A*02, A*11* and *A*24* in Asia. To assess the overall frequencies of *HLA-A*02, A*11* and *A*24* in Asia, we obtained *HLA* frequency data for 93 populations from 20 Asian countries representing >91% of the population of Asia. Frequencies in each country were defined as the median value for all considered populations. Using these median frequencies and the population size of each country (obtained from the Population Reference Bureau website, http://www.prb.or), we estimated the number of *A*02*, *A*11* and *A*24* alleles in each country and, by integrating over all Asia, obtained the average frequency. *HLA* frequencies were obtained as indicated in section I.6, with one exception to avoid sampling bias. For Taiwan, data from aborigine populations was not included, as they represent only ~2% of the population of the country but have been extensively sampled and studied.

II.6 Note: Long and ancient haplotypes shared between archaic and modern non-Africans are the products of archaic admixture. Over time recombination breaks haplotypic associations, so that the longer two populations have been separated the shorter their shared haplotypes become. Consistent with this thesis, Green et al. (*3*) found in a genome-wide scan across 50kb windows only 13 regions for which there is a much deeper coalescence time separating non-African haplotypes than African haplotypes. The presence of these 13 long haplotypes could be explained by archaic admixture in non-Africans. For 10 of these regions, Neandertal sequences matched the deep clade unique to non-Africans, suggesting that when this haplotype configuration is observed, it is often the result of Neandertal admixture (*3*). Thus, in the genomes of non-Africans, deeply divergent haplotypes of more than >50kb deriving from African ancestry (as opposed to admixture) are seen to be very uncommon (*3*).

The Denisovan and Neandertal *HLA* haplotypes we characterized in this study follow the same pattern but have more dramatic features. For example, the haplotype formed by *A*11* and either *C*12:02* or *C*15* is shared by Denisovans and modern Asian and Oceanian populations. This haplotype is more than eight times the size of the longest shared haplotype observed by Green et al in the genome-wide study (~1.3Mb vs ~160kb, respectively) (*3*). This *A*11-C*12:02* or *A*11-C*15* haplotype does not exist in Africa but carries allele groups whose formation predates the separation between archaic and modern humans. Similarly and more specifically, studies of isolated Amerindian populations show that *HLA class I* alleles are rapidly diversified by recombination (*17, 18*) so that long-term (>250ky) conservation of allele sequences or haplotype structures is not an expected feature of *HLA class I* evolution, despite strong balancing selection (*67*).

Part III. *HLA class I* recombination rates and LD decay in four reference populations

III.1 Estimation of *HLA class I* recombination rates. Recombination rates were estimated in four populations (YRI, Yoruba (Nigeria); CEU, Europeans from Utah; CHB, Han Chinese; JPT, Japanese) using high-resolution phased SNP data (*31*). Three genomic intervals were analyzed for each population, *HLA-A-B*, *A-C* and *C-B.* For the *HLA-C-B* interval all SNP between the two loci were used (89-101 SNP; ~1SNP/kb), while for the *HLA-A-C* and *A-B* intervals, one of every five SNP was used (299-321 SNPS; ~1SNP/4.5kb). For each run of PHASE 2.1 (*68, 69*), we used the general model for recombination rate (-MR0), and sampling was performed for 500 iterations (thinning interval of 1) after a burn-in phase of 500 iterations; three independent runs were conducted for each population and interval. The run producing the best goodness-of-fit measure was retained (fig. S26). The population recombination rate (ρ) for each interval was then converted to cM/Mb using an effective population size of 7,500 for the African population and 3,100 for the non-African populations (*70*).

III.2 Characterization of haplotypes with rapid LD decay. High-resolution genotyping has identified SNP markers that correlate with individual HLA alleles (*31*). We defined haplotypes with rapid LD decay as those where all the HLA-identifying SNPs (r^2>0.2) occurred within 500kb of that *HLA* locus. This analysis was performed separately for each locus and population.

III.3 Correlation between haplotypes with rapid LD decay and archaic alleles. The correlation between archaic alleles and alleles defining haplotypes with rapid LD decay was studied using a binomial distribution: considering $\Omega=(0,1,2,\ldots,n)$, $\forall k\in\Omega$, $p=(X=k)= {}_nC_k * p^k * q^{n-k}$.

This indicates for example that under a random distribution it is unlikely to have more than 28 of the 100 archaic alleles defining haplotypes with rapid LD decay, as the latter only represents 49 of the 231 haplotypes (α=0.05); so the fact that 33 of the 100 archaic alleles define haplotypes with rapid LD decay indicates a biased distribution (α =0.0042).

Part IV. *KIR3DL1/S1* in the Denisova genome

To investigate the *KIR3DL1/S1* content of the Denisova genome (*4*), we used the approach detailed in section II.1 for the *HLA class I* genes: all reads that had been mapped to the *KIR* locus of human (chr19:59,920,000-60,075,000) and chimpanzee (chr19:60,452,452-60,582,180) reference genomes were extracted and remapped to the variant *KIR* genomic sequences (*71*). *KIR3DL1* and *KIR3DS1* are alleles of the *KIR3DL1/S1* gene (*24, 58*). There were 405 reads that mapped exclusively to *3DL1/S1*, 66 of these reads identified *3DS1* (*3DS1*013* closest known allele) and 76 identified *3DL1* (*3DL1*005* closest known allele). Demonstrating two alleles were present, 41 nucleotide positions were heterozygous for *3DL1* and *3DS1* and these occurred throughout the 15kb genomic sequence of *3DL1/S1*.

Allele frequencies of *3DS1*013* were obtained from 117 populations (*12*). Distribution maps were generated using the GMT software package (*46*), as indicated in section I.6.

Supplemental References

32. B. Chevreux, T. Wetter, S. Suhai, *Computer Science and Biology: Proceedings of the German Conference on Bioinformaticss (GCB)* **99**, 45 (1999).
33. R. Staden, K. F. Beal, J. K. Bonfield, *Methods Mol. Biol.* **132**, 115 (2000).
34. S. F. Altschul *et al.*, *Nucleic Acids Research* **25**, 3389 (1997).
35. D. P. Locke *et al.*, *Nature* **469**, 529 (2011).
36. K. Katoh, K. Misawa, K. Kuma, T. Miyata, *Nucleic Acids Res.* **30**, 3059 (2002).
37. D. P. Martin, C. Williamson, D. Posada, *Bioinformatics* **21**, 260 (2005).
38. S. Kumar, K. Tamura, M. Nei, *Brief. Bioinform.* **5**, 150 (2004).
39. D. L. Swofford. (Sinauer, Sunderland, Massachusetts, 2001).
40. A. Stamatakis, *Bioinformatics* **22**, 2688 (2006).
41. Z. Yang, *Comput Appl Biosci* **13**, 555 (1997).
42. M. J. Benton, P. C. Donoghue, *Mol. Biol. Evol.* **24**, 26 (2007).
43. Y. Chaimanee *et al.*, *Nature* **427**, 439 (2004).
44. Y. Chaimanee *et al.*, *Nature* **422**, 61 (2003).
45. G. Suwa, R. T. Kono, S. Katoh, B. Asfaw, Y. Beyene, *Nature* **448**, 921 (2007).
46. P. Wessel, W. H. F. Smith, *Eos, Transactions, American Geophysical Union* **79**, 579 (1998).
47. O. D. Solberg *et al.*, *Hum. Immunol.* **69**, 443 (2008).
48. M. Luo *et al.*, *Tissue Antigens* **59**, 370 (2002).
49. C. Kollman *et al.*, *Hum Immunol* **68**, 950 (2007).
50. D. Meyer *et al.*, in *Immunobiology of the Human MHC: Proceedings of the 13th International Histocompatibility Workshop and Conference, Volume I*, J. A. Hansen, Ed. (IHWG Press, Seattle, WA, 2007), pp. 653-704.
51. G. Laval, L. Excoffier, *Bioinformatics* **20**, 2485 (2004).
52. W. G. Hill, A. Robertson, *TAG Theoretical and Applied Genetics* **38**, 226 (1968).
53. M. M. Pilkington *et al.*, *Mol Biol Evol* **25**, 517 (2008).
54. S. G. Marsh *et al.*, *Tissue Antigens* **75**, 291 (2010).
55. H. Li *et al.*, *Bioinformatics* **25**, 2078 (2009).
56. A. W. Briggs *et al.*, *Proc Natl Acad Sci U S A* **104**, 14616 (2007).
57. A. W. Briggs *et al.*, *Nucleic Acids Res* **38**, e87 (2010).
58. J. Robinson, K. Mistry, H. McWilliam, R. Lopez, S. G. Marsh, *Nucleic Acids Res* **38**, D863 (2010).
59. P. Main, R. Attenborough, G. Chelvanayagam, K. Bhatia, X. Gao, *Hum Biol* **73**, 365 (2001).
60. M. Stephens, P. Donnelly, *Am J Hum Genet* **73**, 1162 (2003).
61. M. Stephens, N. J. Smith, P. Donnelly, *Am J Hum Genet* **68**, 978 (2001).
62. M. Maiers, L. Gragert, W. Klitz, *Hum Immunol* **68**, 779 (2007).
63. M. Gleimer *et al.*, *J Immunol* **186**, 1575 (2011).
64. L. Sanz *et al.*, *Tissue Antigens* **47**, 329 (1996).
65. J. Inoue, P. C. Donoghue, Z. Yang, *Syst Biol* **59**, 74 (2010).
66. P. J. Waddell, D. Penny, in *Handbook of Human Symbolic Evolution*, A. Lock, C. R. Peters, Eds. (Oxford University Press, Oxford, 1996), pp. 53-73.

67. K. Gendzekhadze *et al.*, *Proc Natl Acad Sci U S A* **106**, 18692 (2009).
68. N. Li, M. Stephens, *Genetics* **165**, 2213 (2003).
69. D. C. Crawford *et al.*, *Nat Genet* **36**, 700 (2004).
70. A. Tenesa *et al.*, *Genome Res* **17**, 520 (2007).
71. C. W. Pyo *et al.*, *PLoS One* **5**, e15115 (2010).
72. T. Shiina *et al.*, *Genetics* **173**, 1555 (2006).
73. N. Lee, D. R. Goodlett, A. Ishitani, H. Marquardt, D. E. Geraghty, *J Immunol* **160**, 4951 (1998).
74. P. Parham *et al.*, *J Med Primatol* **39**, 194 (2010).
75. L. Abi-Rached, A. K. Moesta, R. Rajalingam, L. A. Guethlein, P. Parham, *PLoS Genet* **6**, e1001192 (2010).
76. L. Abi-Rached *et al.*, *J Immunol* **184**, 1379 (2010).
77. E. J. Adams, G. Thomson, P. Parham, *Immunogenetics* **49**, 865 (1999).
78. A. Mohyuddin, S. Q. Mehdi, *Tissue Antigens* **66**, 691 (2005).
79. M. Zetlaoui, *Ainsi vont les enfants de Zarathoustra : Parsis de l'Inde et Zartushtis d'Iran* (Imago 2003).
80. R. N. Koehler *et al.*, *PLoS One* **5**, e10751 (2010).
81. J. Tang *et al.*, *Exp Clin Immunogenet* **17**, 185 (2000).
82. J. N. Torimiro *et al.*, *Tissue Antigens* **67**, 30 (2006).
83. S. Farjadian *et al.*, *Tissue Antigens* **64**, 581 (2004).
84. D. Middleton *et al.*, *Hum Immunol* **61**, 1048 (2000).
85. S. A. Tishkoff *et al.*, *Science* **324**, 1035 (2009).
86. B. M. Henn *et al.*, *Proc Natl Acad Sci U S A* **108**, 5154 (2011).
87. J. Bruges Armas *et al.*, *Tissue Antigens* **62**, 233 (2003).

Supporting Figures

Figure S1
Simplified map of the *HLA class I* region

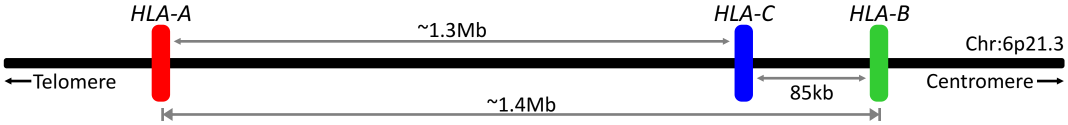


Fig. S1. Simplified map of the *HLA class I* region. Schematic representation of the location and distances between the *HLA-A*, *HLA-B* and *HLA-C* loci. *HLA* or *human leukocyte antigen* is the name given to the *Major Histocompatibility Complex* (*MHC*) in humans. *HLA* alleles are distinguished by a rational nomenclature (*54*) in which the name of the gene is followed by up to four sets of digits separated by colons. The digits before the first colon distinguish groups of alleles that roughly correspond to the serological types first used to define HLA class I polymorphisms. The next set of digits distinguishes subtypes that differ in their amino-acid sequences. Subsequent sets of digits distinguish alleles that differ by synonymous (third set) or intronic (fourth set) differences.

Figure S2
Characterization of extended *MHC-B* gene sequences in hominoids

A

	Individual		*MHC-B* allele
	Species	Name	
Extended gene sequence (8.5kb)	Common chimpanzee	Harry	*05:01
			*09:01
		Miss Eve	*18:01
		Phineas	*24:01
		Renee	*36:01
			*38:01
		Tank	*03:01
		Cheeta	*17:01
			*04:02
	Bonobo	Matata	*07:01
	Gorilla	Machi	*04:01
		Beta	*07:01
		Oko	*02:01
		Shamba	*03:01
	Orangutan	Susie#	*03:02
Haplotype sequence (38kb)	Gorilla	Kamilah#	*06:01

B

Chimpanzee	Allele of the *B*73* lineage	Source
Cheeta	Patr-B*17:01	Sequencing
Leslie	Patr-B*17:01	Sequencing
Lindsey	Patr-B*17:01	Filiation
Fiona	Patr-B*17:01	Filiation
Ericka	Patr-B*17:01	Sequencing

Fig. S2. Characterization of extended *MHC-B* gene sequences in hominoids. (**A**) Chimpanzees, gorillas and orangutans from whom *MHC-B* extended gene or haplotype sequences were characterized. These sequences were named according to the *MHC-B* allele sequences they contain (IPD-MHC Database (*58*)). #, sequence determined from BAC DNA. (**B**) *Patr-B*17:01* is a chimpanzee allele of the *B*73* lineage (*MHC-BII*). Five chimpanzees from a panel of 95 tested were positive for *B*73* lineage-specific PCR amplification. Sequencing or pedigree typing analysis showed all five individuals carried *Patr-B*17:01*. *Patr-B*17:01* from Cheeta and Leslie was sequenced in this study. Lindsey and Fiona inherited *Patr-B*17:01* from their mother, Leslie, and the sequence of *Patr-B*17:01* from Ericka was determined previously (*72*).

Figure S3
Phylogenetic analysis of the segments forming the *HLA-B*73:01* haplotype

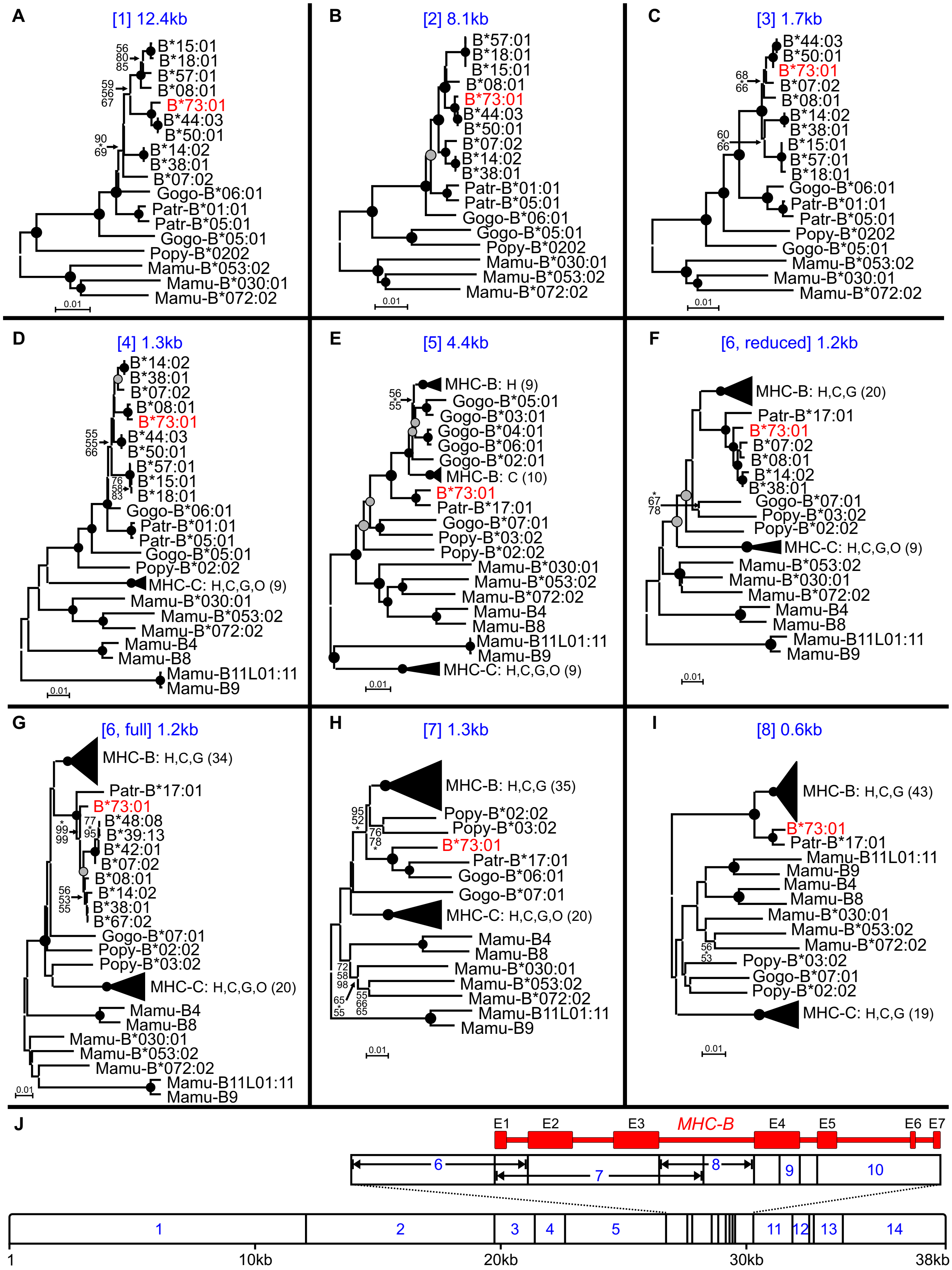


Figure S3
Phylogenetic analysis of the segments forming the *HLA-B*73:01* haplotype

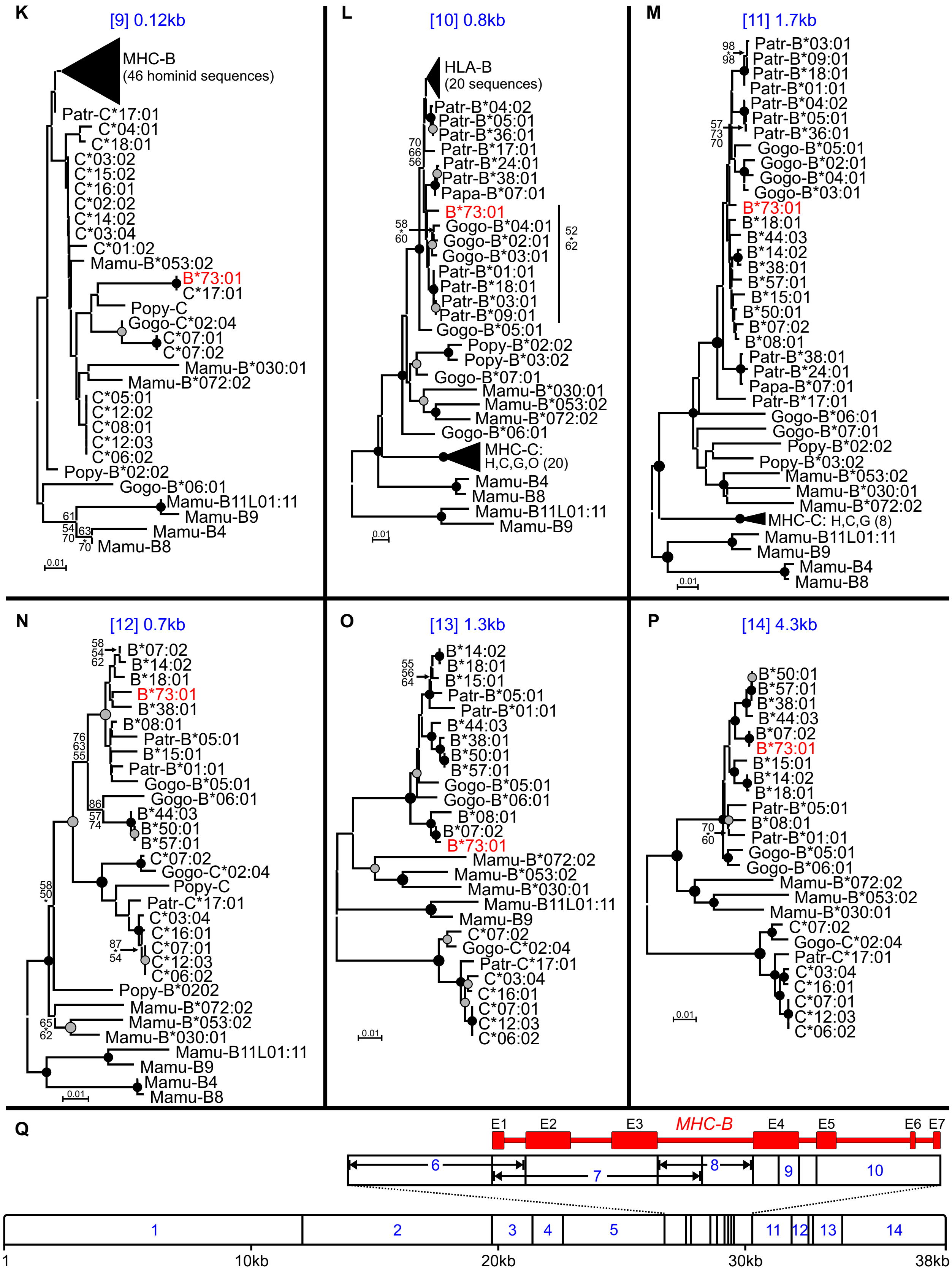
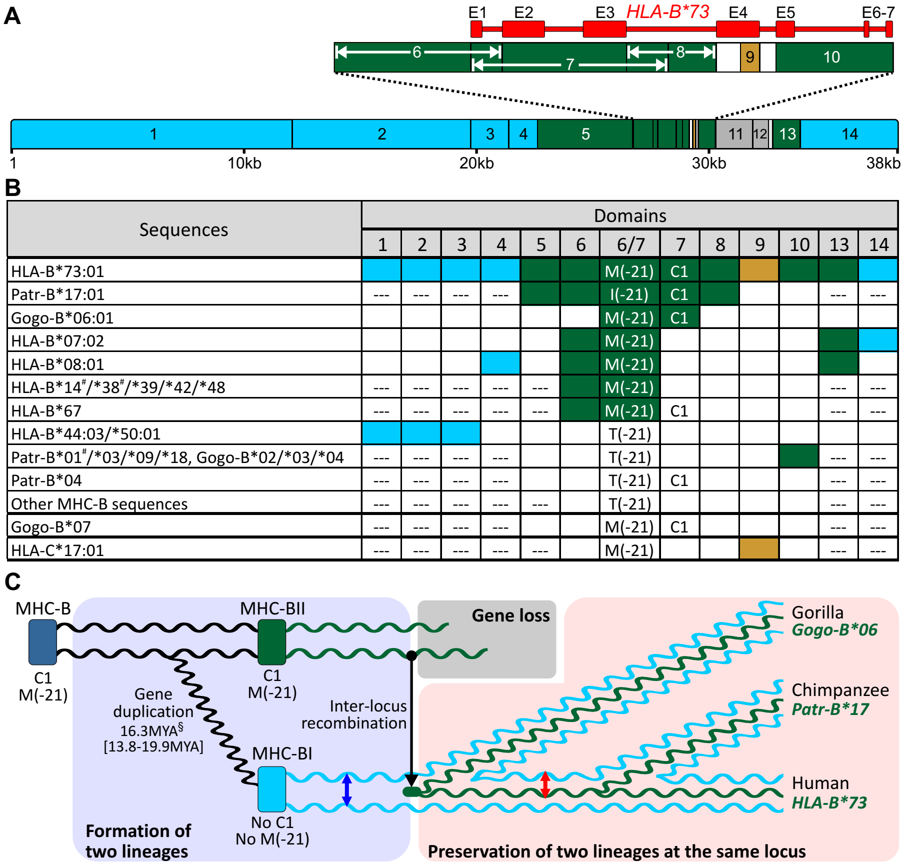


Fig. S3. Phylogenetic analysis of the segments forming the *HLA-B*73:01* haplotype. (**A**-**Q**) Shown are phylogenetic reconstructions performed on each of 14 genomic segments (**A**-**I**, **K**-**P**). In blue at the top of each panel is shown the segment number (corresponds to the segments shown in panels (**J**, **Q**)) and its size. Analysis was performed using Maximum-Likelihood (ML), Neighbor-Joining (NJ) and parsimony with NJ tree topologies used for the display (midpoint rooting). Circles at nodes indicate strong (black) or moderate (grey) phylogenetic support, where bootstrap proportions from the three methods are ≥80% or ≥60%, respectively. Otherwise, bootstrap values are indicated when at least two methods gave ≥50% support (from top to bottom: NJ, parsimony, ML). To simplify the display some groups of *MHC-B* sequences were collapsed as indicated by black triangles: the number of sequences is given in parentheses together with the species composing the group (H, human; C, chimpanzees; G, gorilla; O, orangutan). Patr, *Pan troglodytes*; Papa, *Pan paniscus*; Gogo, *Gorilla gorilla*; Popy, *Pongo pygmaeus*; Mamu, *Macaca mulatta*.

Figure S4

*HLA-B*73* is the only remnant in modern humans of a deeply divergent allelic lineage



Sequences	Domains												
	1	2	3	4	5	6	6/7	7	8	9	10	13	14
HLA-B*73:01	🟦	🟦	🟦	🟦	🟩	🟩	🟩 M(-21)	🟩 C1	🟩	🟫	🟩	🟩	🟦
Patr-B*17:01	---	---	---	---	🟩	🟩	🟩 I(-21)	🟩 C1	🟩			---	---
Gogo-B*06:01							🟩 M(-21)	🟩 C1					
HLA-B*07:02						🟩	🟩 M(-21)					🟩	🟦
HLA-B*08:01				🟦		🟩	🟩 M(-21)					🟩	
HLA-B*14#/*38#/*39/*42/*48	---	---	---	---	---	🟩	🟩 M(-21)					---	---
HLA-B*67	---	---	---	---	---	🟩	🟩 M(-21)	C1				---	---
HLA-B*44:03/*50:01	🟦	🟦	🟦				T(-21)						
Patr-B*01#/*03/*09/*18, Gogo-B*02/*03/*04	---	---	---	---			T(-21)				🟩	---	---
Patr-B*04	---	---	---	---			T(-21)	C1				---	---
Other MHC-B sequences	---	---	---	---	---		T(-21)					---	---
Gogo-B*07	---	---	---	---			M(-21)	C1				---	---
HLA-C*17:01	---	---	---	---	---		M(-21)			🟫		---	---

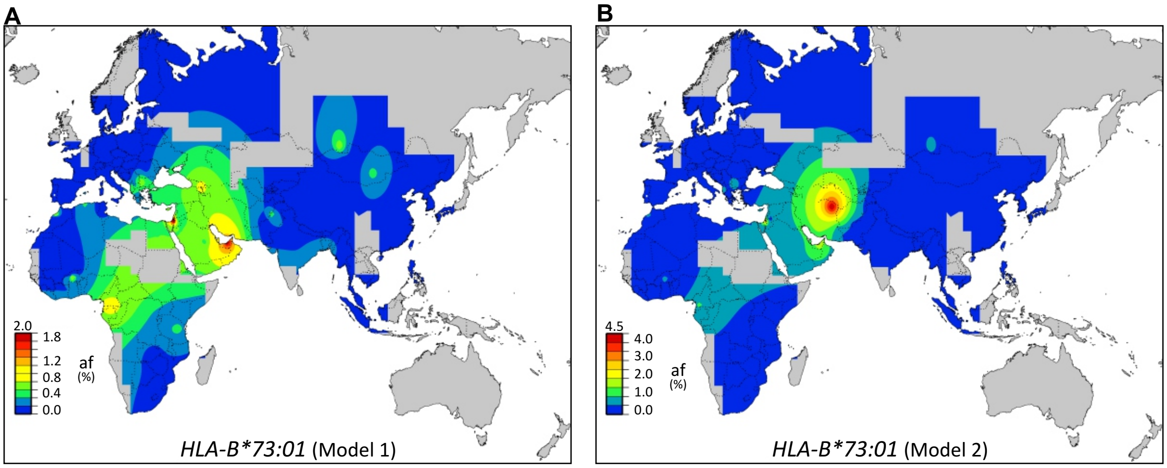
Fig. S4. *HLA-B*73* is the only remnant in modern humans of a deeply divergent allelic lineage. (**A**) Phylogenetic analyses (fig. S3) show the *B*73* haplotype contains segments most closely related to chimpanzee and gorilla *MHC-B* alleles (green) and flanking segments highly related to other *HLA-B* (blue) (brown segment is related to *HLA-C*). (**B**) Sequences with segments equivalent to those of *B*73:01*. Some human alleles acquired *B*73:01* segments by recombination. 'M(-21)' and 'C1' are functional characteristics in the leader peptide and α1 domain, respectively. Presence of methionine at position -21 ensures the HLA-B leader peptide is processed and bound to HLA-E, a ligand for CD94:NKG2 natural killer (NK) cell receptors (73). M(-21) is associated with MHC-BII but not with the main lineage of MHC-B sequences (MHC-BI). Additionally, MHC-BII molecules have the C1 epitope (formed by valine 76 and asparagine 80 in the α1 domain) which engages the KIR2DL2/3 NK cell receptors (74); in contrast a minority of MHC-BI molecules have C1 (75). ---, no sequence. #, complete sequence. (**C**) Model to explain how the sequence and functional divergence of *B*73* emerged and was maintained, despite the high level of recombination at *MHC-B*. *B*73*'s divergent core was formed >16 million years ago (MYA), suggesting this divergence is linked to the *MHC-B* gene duplications that occurred in early hominoids (76, 77). Accordingly, following duplication and divergence of *MHC-BI* and *BII*, one allele of *BII* recombined to the *BI* locus giving rise to the ancestor of *B*73* and its gorilla and chimpanzee equivalents. Sequence divergence likely reduced intra-locus recombination of *BI-BII* lineages (red arrow) comparing to the high levels of *BI-BI* recombination (blue arrow), thus contributing to the long-term structural conservation of the *MHC-BII* allele at the *MHC-BI* locus. §, mean and 95% credibility interval.

Figure S5
Diversity of the *HLA-B* allele groups

HLA-B allele group	Number of specific HLA allotypes	Notes
B*07	124	
B*08	69	
B*13	46	
B*14	21	
B*15	221	
B*18	59	
B*27	79	
B*35	160	
B*37	28	
B*38	27	
B*39	63	
B*40	162	
B*41	18	
B*42	16	
B*44	127	
B*45	13	
B*46	27	
B*47	8	
B*48	25	
B*49	16	
B*50	13	
B*51	109	
B*52	23	
B*53	24	
B*54	24	
B*55	49	
B*56	31	
B*57	43	
B*58	33	
B*59	5	
B*67	3	
B*73	2	Second allotype: single observation
B*78	7	
B*81	5	
B*82	3	
B*83	1	Uncommon recombinant

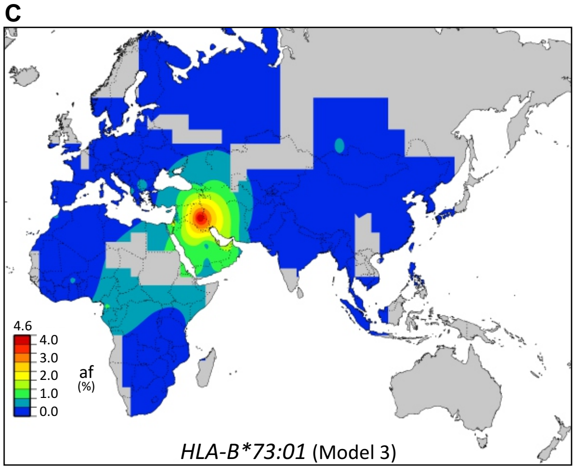
Fig. S5. Diversity of the *HLA-B* allele groups. The number of specific allotypes is given for each of the 36 *HLA-B* allele groups (from the IMGT-HLA Database, release 3.3.0 (*10*)). *B*73:02* was identified in one person and is possibly resulting from mutation in that individual.

Figure S6
Models for the distribution of *HLA-B*73*



*HLA-B*73:01* (Model 1)

*HLA-B*73:01* (Model 2)



*HLA-B*73:01* (Model 3)

D

Population	Freq (%)	n
Parsi	4.90	91
Israel Jews	2.30	146
United Arab Emirates	2.00	373
Cameroon Beti	1.10	174
Georgia Tibilisi	0.90	109
Bulgaria	0.90	55
Morocco Casablanca	0.90	100
Burkina Faso Mossi	0.90	53
Russia Tuva	0.80	169
Oman	0.80	118
Albania	0.60	160
Cameroon Bamileke	0.60	77
Mongolia Khalkha	0.50	200
Pakistan Pathan	0.50	100
Iran Baloch	0.50	100
Tanzania Dodoma Kongwa	0.50	212
Israel Arab Druze	0.50	101
Algeria	0.50	106
Saudi Arabia	0.40	383
Turkey	0.40	142
Bangladesh Dhaka Bangalee	0.30	141
Jordan Amman	0.30	146
Uganda Kampala	0.30	175
Kenya Nandi	0.20	240
Kenya Luo	0.20	265
Zimbabwe Harare Shona	0.20	230
Rwanda	0.20	280
Greece	0.20	246
Macedonia	0.20	216

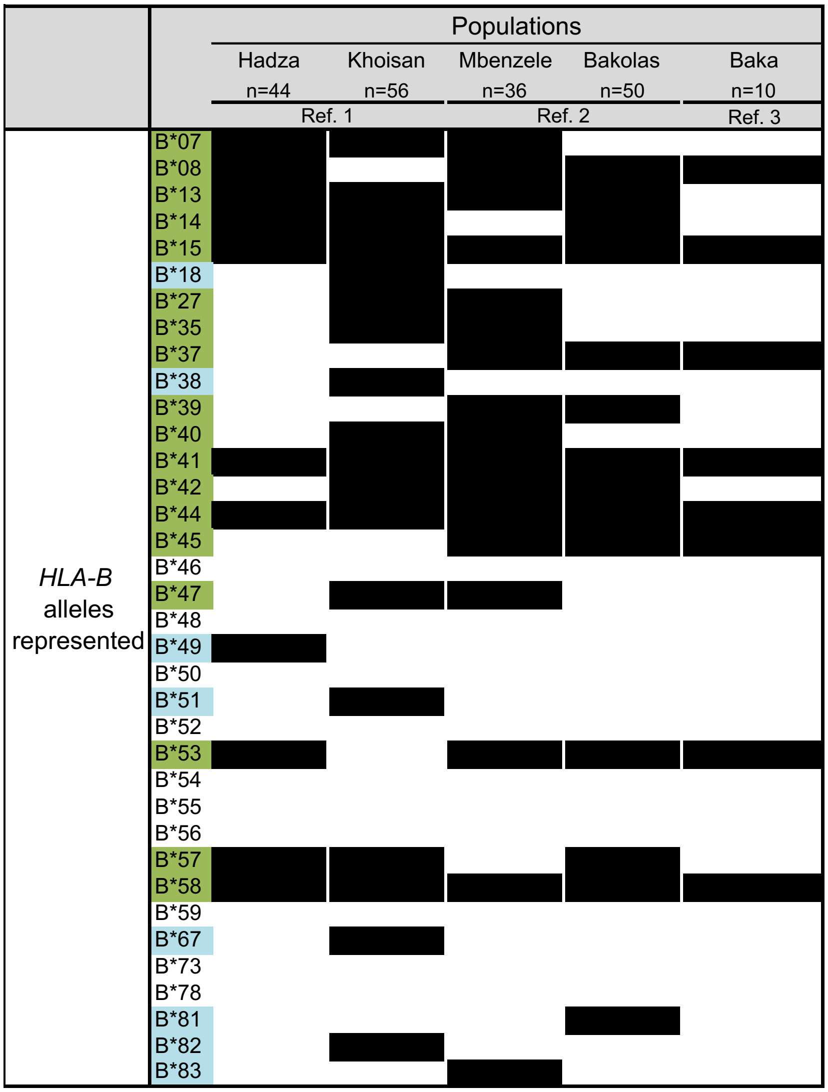
Fig. S6. Models for the distribution of *HLA-B*73*. (**A-D**) Shown is the distribution of *B*73* (**A-C**) and the *B*73:01* allele frequency of all populations with *B*73* (**D**). Because of uncertainty in the ancient geographical location of Parsis, the population with the highest *B*73* frequency (12, 78) (**D**), we considered three models for the distribution of *B*73* (**A-C**). While some Parsis now live in Pakistan, they represent a 'recent' migrant population from Persia (79) and are not included in model 1. Models 2 and 3 are based on two alternative hypotheses of Parsi origin. Model 2 uses the historical Greater Khorasan region as the point of origin of their migration (city of Nishapur, 58.8E 36.2N), as this region became a refuge for Zoroastrians following the fall of the Sasanian empire, and subsequent migration to India originated from this region (79). The Khorasan region was only a refuge for Zoroastrians who used to live in other regions of Persia however, so in Model 3 the Sasanian capital Ctesiphon (present-day Iraq, 44.62E 33.15N) is the point of origin of the migration.

Figure S7
HLA-C characteristics of individuals with _HLA-B*73_

Origin of the data (individuals with _HLA-B*73_)		Ethnicity	_HLA-C_ characteristics of individuals with _B*73_					
			n	_C*15_ subtype				Not _C*15_
				_:05_	Allele string		not _:05_	
					includes _:05_	excludes _:05_		
Bone Marrow Registries	USA-NMDP	Europe	2,158	1,271	853	6	4	24
		Western Asia	64	28	35			1
		Western Asia**	6	1	5			
		N/S/E Asia	39	12	27			
		Africa	15	7	8			
		Africa*	49	12	36		1	
		Hispanic	228	73	154			1
		Other/unknown	470	198	268			4
	Canada	Europe	98	95				3
		N/S/E Asia	7	6				1
		Western Asia	1	1				
		Hispanic	1	1				
		Africa*	1	1				
		Other/unknown	12	11				1
	UK -Anthony Nolan	Europe	32	2	30			
		N/S/E Asia	4		4			
		Africa	5	1	4			
		Western Asia	7	1	6			
		Other/unknown	16		15	1		
	Italy	(Europe)	90	5	82			3
	UK-BBMR	(Europe)	86		80	5	1	
	USA-Gift of Life	(Europe)	65	19	44			2
	France	(Europe)	49	20	29			
	Cyprus BMDR	(Europe)	21	6	14			1
	Poland-DKMS	(Europe)	19	19				
	UK-Wales	(Europe)	17	9	6	2		
	Belgium	(Europe)	7		6			1
	Australia	(Europe)	7	2	5			
	USA -UCLA	(Europe)	6	1	5			
	Norway	(Europe)	5		4	1		
	Spain	(Europe)	3		3			
	Portugal	(Europe)	3	2	1			
	Austria	(Europe)	2		2			
	Switzerland	(Europe)	2	1	1			
	Czechia-Central BMDR	(Europe)	2	1	1			
	Poland-POLTransplant	(Europe)	2		2			
	Ireland	Europe	1		1			
	Lithuania	(Europe)	1		1			
	Sweden	(Europe)	1	1				
	Israel-Hadassah	(Western Asia)	3		2			1
	Israel-Ezer Mizion	(Western Asia)	21	4	15		1	1
	Turkey#	(Western Asia)	8	8				
	Singapore	N/S/E Asia	1		1			
Published studies	dbMHC	Western Asia	6	1			2	3
		Africa	3	3				
		N/S/E Asia	2	1			1	
		Hispanic	1	1				
	Uganda -ref. 1	Africa	1	1				
	Kenya/Tanzania -ref. 2	Africa	6	6				
	Sudan -ref. 3	Africa	3	3				
	African American -ref. 2	Africa*	1	1				
	Rwanda -ref. 4	Africa	1		1			
	Cameroon Bamileke -ref. 5	Africa	1	1				
	Cameroon Beti -ref. 5	Africa	4	4				
	Parsi -ref. 6	Western Asia	9		9			
	Iran Baloch -ref. 7	Western Asia	1	1				
	Oman -ref. 8	Western Asia	2	2				

Fig. S7. _HLA-C_ characteristics of individuals with _HLA-B*73_. To investigate how frequently the _B*73_ haplotypes also carry the _HLA-C*15:05_ allele, we summarized data available in the literature and obtained genotype data from a large number of bone marrow donors with _B*73_. Parentheses in the 'Ethnicity' column indicate that no ethnicity data were available: the region where the bone marrow registry is located was used instead. 'Allele string', lower resolution _HLA_-typing data. * African-American(s). ** West Asia or north African coast. # _C*15:05_ and _C*15:22_ could not be distinguished. References: dbMHC(50), ref. 1 (80), ref.2 (this study), ref. 3(12), ref. 4(81) , ref. 5(82), ref. 6(78), ref. 7(83), ref. 8(84).

Figure S8
HLA-B allelic groups in the African populations that diverged before the Out-of-Africa migration



		Populations				
		Hadza n=44	Khoisan n=56	Mbenzele n=36	Bakolas n=50	Baka n=10
		Ref. 1		Ref. 2		Ref. 3
HLA-B alleles represented	🟩 B*07	●	●	●		
	🟩 B*08	●		●	●	●
	🟩 B*13	●	●	●	●	
	🟩 B*14	●	●		●	
	🟩 B*15	●	●	●	●	●
	🟦 B*18		●			
	🟩 B*27		●	●		
	🟩 B*35		●	●		
	🟩 B*37			●	●	●
	🟦 B*38		●			
	🟩 B*39			●	●	
	🟩 B*40		●	●		
	🟩 B*41	●	●	●	●	●
	🟩 B*42		●	●	●	
	🟩 B*44	●	●	●	●	●
	🟩 B*45			●	●	●
	B*46					
	🟩 B*47		●	●		
	B*48					
	🟦 B*49	●				
	B*50					
	🟦 B*51		●			
	B*52					
	🟩 B*53	●		●	●	●
	B*54					
	B*55					
	B*56					
	🟩 B*57	●	●		●	
	🟩 B*58	●	●	●	●	●
	B*59					
	🟦 B*67		●			
	B*73					
	B*78					
	🟦 B*81				●	
	🟦 B*82		●			
	🟦 B*83			●		

Fig. S8. *HLA-B* allelic groups in the African populations that diverged before the Out-of-Africa migration. *HLA-B* content is summarized for two Khoisan-speaking (Hadza and Khoisan) and three pygmy (Mbenzele, Bakolas and Baka) populations: separations between these populations and others represent the earliest splits between modern human populations (*14*, *85*). Alleles groups colored green are present in more than one population, while those colored blue are present in only one of the five populations. References are: ref. 1 (*86*), ref. 2 (*87*) and ref. 3 (*82*).

Figure S9
HLA-B*73 and _HLA-C*15:05_ frequencies used for simulations

A

Region	Population	Frequency Data		Database	Summary		
		n (individuals)	_B*73_ n (alleles)		n (individuals)	_B*73_ n (alleles)	_B*73_ AF
West Asia	Israel Ashkenazi and Non Ashkenazi Jews	109	5	Both	1,232	27	1.09
	Israel Arab Druse	100	1	Both			
	Jordan Amman	146	1	AFN			
	Oman	121	2	Both			
	Saudi Arabia pop 2	383	3	AFN			
	United Arab Emirates	373	15	AFN			
Africa	Algeria pop 2	106	1	AFN	4,503	25	0.28
	Burkina Faso Fulani	49	0	AFN			
	Burkina Faso Mossi	53	1	AFN			
	Burkina Faso Rimaibe	47	0	AFN			
	Cameroon Bamileke	77	1	AFN			
	Cameroon Beti	174	4	AFN			
	Cameroon Sawa	13	0	AFN			
	Kenya Luo	265	1	Both			
	Kenya Nairobi Sex workers	1481	8	This study			
	Kenya Nairobi	143	0	dbMHC			
	Kenya Nandi	240	1	Both			
	Mali Bandiagara	138	0	Both			
	Morocco Casablanca	100	2	AFN			
	Morocco Nador Metalsa	68	0	Both			
	Morocco Settat	68	0	dbMHC			
	Rwanda	280	1	AFN			
	Senegal	94	0	dbMHC			
	South African Zulu	201	0	dbMHC			
	Sudanese	200	3	AFN			
	Tunisia	100	0	AFN			
	Uganda Kampala	161	0	Both			
	Uganda Kampala pop 2	175	1	AFN			
	Zambia Lusaka	44	0	Both			
	Zimbabwe Harare Shona	226	1	Both			

B

Region	Population	Frequency Data		Database	Summary		
		n (individuals)	_C*15:05_ n (alleles)		n (individuals)	_C*15:05_ n (alleles)	_C*15:05_ AF
West Asia	Israel Ashkenazi and Non Ashkenazi Jews	94	5	Both	504	29	2.88
	Israel Arab Druse	100	10	Both			
	Lebanon	97	1	AFN			
	Saudi Arabia Guraiat and Hail	213	13	AFN			
Africa	Cameroon Bamileke	77	1	AFN	2,254	39	0.87
	Cameroon Beti	174	8	AFN			
	Cameroon Sawa	13	0	AFN			
	Kenya	113	1	Both			
	Kenya Luo	265	3	Both			
	Kenya Nandi	240	4	Both			
	Mali Bandiagara	129	0	Both			
	Morocco Nador Metalsa	73	2	AFN			
	South African Zulu	96	0	dbMHC			
	Senegal Niokholo Mandenka	165	0	AFN			
	Sudanese	200	11	AFN			
	Tunisia	100	3	AFN			
	Uganda Kampala	163	1	Both			
	Uganda Kampala pop 2	175	4	AFN			
	Zambia Lusaka	45	0	Both			
	Zimbabwe Harare Shona	226	1	Both			

Fig. S9. _HLA-B*73_ and _HLA-C*15:05_ frequencies used for simulations. (**A**-**B**) _HLA-B*73_ (**A**) and _C*15:05_ (**B**) frequency data in African and west Asian populations were obtained from the Allele Frequency Net (AFN) website (_12_) or from dbMHC (_50_).

Figure S10
Haplotype data used for the *B*73* population simulations

Region	Haplotypes									
	Type	Model 1: LD 100%	Model 2: LD 95%	Model 3: LD 90%	Model 4: LD 95% / 100%	Model 5: LD 100% / 95%	Model 6: LD 90% / 100%	Model 7: LD 90% / 95%	Model 8: LD 100% / 90%	Model 9: LD 95% / 90%
W Asia	B*73-C*15:05	27	26	24	26	27	24	24	27	26
	B*73-No C*15:05	0	1	3	1	0	3	3	0	1
	No B*73-C*15:05	44	45	47	45	44	47	47	44	45
	No B*73-No C*15:05	2,393	2,392	2,390	2,392	2,393	2,390	2,390	2,393	2,392
Africa	B*73-C*15:05	25	24	23	25	24	25	24	23	23
	B*73-No C*15:05	0	1	3	0	1	0	1	3	3
	No B*73-C*15:05	53	54	55	53	54	53	54	55	55
	No B*73-No C*15:05	8,928	8,927	8,926	8,928	8,927	8,928	8,927	8,926	8,926

Fig. S10. Haplotype data used for the *B*73* population simulations. Haplotype data for the population simulations were generated for nine different models of linkage disequilibrium (LD) using the *B*73* and *C*15:05* allele frequencies of figure S9. For models 1-3, LD is set at the same percentage for both regions (Africa and west Asia), while for models 4-9, a different value is used for each region (west Asia/Africa).

Figure S11
Results of *HLA-B*73* population simulations

A

Models		Observed parameters (*100)				Derived Fraction (DF) statistic (99% tolerance) Linkage Disequilibrium (LD) statistic (50% tolerance)							
						Double Accepts		Runs $(/10^6)$	α (%)	Range tolerated (*100)			
Name	LD (%) W.Asia/Africa	DF1	DF2	LD1	LD2	Number	Ratio			DF1	DF2	LD1	LD2
Ma1 Mb1	100 / 100	1.10	0.28	37.34	31.86	14,823 85	174	10 10	0.57	0.01-2.18	0.003-0.55	18.67-53.27	15.93-47.79
Ma2 Mb2	95 / 95	1.06	0.27	34.55	29.34	7,132 44	162	5 5	0.61	0.01-2.10	0.003-0.53	17.27-49.22	14.67-44.01
Ma3 Mb3	90 / 90	0.97	0.26	29.29	25.87	6,104 44	139	5 5	0.72	0.01-1.94	0.003-0.51	14.64-42.22	12.94-38.81
Ma4 Mb4	95 / 100	1.06	0.28	34.55	31.86	15,008 75	200	10 10	0.50	0.01-2.10	0.003-0.55	17.27-50.48	15.93-47.79
Ma5 Mb5	100 / 95	1.10	0.27	37.34	29.34	7,301 35	209	5 5	0.48	0.01-2.18	0.003-0.53	18.67-52.01	14.67-44.01
Ma6 Mb6	90 / 100	0.97	0.28	29.29	31.86	7,247 42	173	5 5	0.58	0.01-1.94	0.003-0.55	14.64-45.22	15.93-47.79
Ma7 Mb7	90 / 95	0.97	0.27	29.29	29.34	6,985 58	120	5 5	0.82	0.01-1.94	0.003-0.53	14.64-43.96	14.67-44.01
Ma8 Mb8	100 / 90	1.10	0.26	37.34	25.87	6,404 41	156	5 5	0.64	0.01-2.18	0.003-0.51	18.67-50.28	12.94-38.81
Ma9 Mb9	95 / 90	1.06	0.26	34.55	25.87	6,409 37	173	5 5	0.57	0.01-2.10	0.003-0.51	17.27-47.49	12.94-38.81

B

Models		Observed parameters (*100)				Derived Fraction (DF) statistic (50% tolerance) Linkage Disequilibrium (LD) statistic (25% tolerance)							
						Double Accepts		Runs $(/10^6)$	α (%)	Range Tolerated (*100)			
Name	LD (%) W.Asia/Africa	DF1	DF2	LD1	LD2	Number	Ratio			DF1	DF2	LD1	LD2
Ma4 Mb4	95 / 100	1.06	0.28	34.55	31.86	1,103 -	> 3,309	10 30	<0.1	0.53-1.58	0.14-0.42	25.91-43.18	23.90-39.83

C

Models		Simulation			α (%)
		Double Accepts		Runs $(/10^6)$	
Name	LD (%) W Asia/Africa	Number	Ratio		
Ma4 Mb4	95 / 100	15,008 75	200	10 10	0.497
Ma4_H Mb4	95 / 100	1,660 75	2,354	0.09 10	0.041

Fig. S11. Results of *HLA-B*73* population simulations. (**A**-**C**) Comparison of two competing models for the history of *B*73*: archaic admixture (model a) or African origin (model b). Shown are the summary statistics obtained following simulations with both models using the parameters indicated at the left. 'LD' is linkage disequilibrium between the *HLA-B* and *HLA-C* alleles. Panel (**A**) shows simulations 1-9, the input data for which are detailed in figure S10, the central columns show the result and the columns at the right show the tolerated range. Panel (**B**) shows simulations with tolerance values halved. (**C**) Simulations using a novel model 'Ma4_H', for which an already formed *B*73-C*15:05* haplotype is introduced by admixture. α: confidence level to reject model 'b'. (**B**-**C**) Because no significance difference was observed between the nine models of LD in the analyses of panel (**A**), Model 4 was used in these analyses as it best represents the observed *B*73-C*15* LD in Africa and west Asia (Fig. 1).

Figure S12
Characterization of the two Denisovan *HLA-B* alleles

	Model 1		Model 2		Model 3		Model 4	
	Modern Alleles	Differences	Modern Alleles	Differences	Modern Alleles	Differences	Modern Alleles	Differences
Denisova Allele 1	B*35:63	4	B*35:63	5	B*35:63	0	B*35:63	1
	B*35:03:01	9	B*35:03:01	10	B*35:05:01	6	B*35:05:01	7
	B*35:05:01	10	B*35:05:01	11	B*35:01:22	7	B*35:01:22	8
	B*35:01:22	11	B*52:01:01:01	11	B*35:03:01	7	B*35:03:01	8
	B*35:02:01	11	B*35:01:22	12	B*35:08:01	7	B*35:08:01	8
	B*35:08:01	11	B*35:08:01	12	B*35:01:01:02	8	B*35:01:01:02	9
	B*52:01:01:01	11	B*35:02:01	12				
			B*52:01:01:02	12				
Denisova Allele 2	B*46:01:05	3	B*15:01:01:01	4	B*15:58	3	B*15:58	4
	B*15:01:01:01	3	B*15:03:01	4	B*15:17:01:01	4	B*15:17:01:01	5
	B*15:03:01	3	B*15:11:01	4	B*15:17:01:02	4	B*15:17:01:02	5
	B*15:11:01	3	B*15:18:01	4	B*15:01:01:01	5	B*15:01:01:01	6
	B*15:18:01	3	B*15:66	4	B*15:03:01	5	B*15:03:01	6
	B*46:01:01	3	B*46:01:01	4	B*15:11:01	5	B*15:11:01	6
	B*15:66	3	B*46:01:05	4	B*15:18:01	5	B*15:18:01	6
	B*15:07:01	4	B*15:07:01	5	B*15:66	5	B*15:66	6
	B*56:03	4	B*56:03	5	B*46:01:01	5	B*46:01:01	6
	B*15:108	5	B*15:108	6	B*46:01:05	5	B*46:01:05	6
	B*15:27:01	5	B*15:32	6	B*15:07:01	6	B*15:07:01	7
	B*15:32	5	B*15:27:01	6	B*56:03	6	B*56:03	7
	B*15:58	5	B*15:58	6	B*15:02:01	7	B*15:13	8
	B*15:02:01	6	B*15:02:01	7	B*15:108	7	B*15:27:01	8
	B*15:13	6	B*15:13	7	B*15:25:01	7	B*15:02:01	8
	B*15:17:01:01	6	B*15:17:01:01	7	B*15:32	7	B*15:108	8
	B*15:17:01:02	6	B*15:17:01:02	7	B*15:13	7	B*15:25:01	8
	B*15:25:01	6	B*15:25:01	7	B*15:27:01	7	B*15:32	8
Mininum differences	7		9		3		5	

Fig. S12. Characterization of the two Denisovan *HLA-B* alleles. To analyze several Denisovan *HLA-B* sequence reads that could belong to either of the two alleles, we investigated all possible combinations and assessed the number of differences with modern alleles. The combination that produced the lowest number of differences with modern sequences (Model 3) is presented in the main text.

Figure S13
Putative Denisovan *HLA-A-C* haplotype frequencies in modern populations

		Denisovan *HLA class I* loci		Modern populations with haplotypes resembling the Denisovan *HLA class I* haplotypes					
		HLA-A	*HLA-C*	Name	Haplotype			HF	n
Putative Denisovan *HLA-A-C* haplotypes	1	*A*02*	*C*15*	Australia Kimberley aborigines	A*02:01	C*15:02	B*40:02	8.3	24
				Venezuela Sierra de Perija Yucpa	A*02:04	C*15:03	B*52:01	5.8	73
				Pakistan Sindhi	A*02	C*15	B*40	4.3	101
				Brazil Guarani-Nandewa	A*02:01	C*15:03	B*51:01	2.9	51
				Albania	A*02	C*15	B*51	2.5	160
				Australia Yuendumu Aborigines	A*02:01	C*15:02	B*40:02	2.5	190
				Venezuela Bari	A*02:22	C*15:02	B*51:01	2.2	23
				Venezuela Bari	A*02:22	C*15:02	B*51:04	2.2	23
				Venezuela Bari	A*02:22	C*15:07	B*52:01	2.2	23
				China Yunnan Hani	A*02:03	C*15:02	B*15:01	2.0	150
				Israel Druze	A*02:01	C*15:05	B*07:02	2.0	100
				Taiwan Paiwan	A*02:01	C*15:02	B*40:01	2.0	51
				Brazil Guarani-Nandewa	A*02:01	C*15:03	B*51:04	2.0	51
				India New Delhi	A*02:01	C*15:07	B*40:06	1.9	53
				Brazil Guarani-Nandewa	A*02:04	C*15:03	B*51:04	1.8	51
				Papua New Guinea Goroka	A*02:06	C*15:02	B*40:02	1.6	32
				Brazil Guarani-Kaiowa	A*02:04	C*15:03	B*51:04	1.5	142
				Brazil Guarani-Kaiowa	A*02:01	C*15:03	B*51:04	1.2	142
				Finnish	A*02:01	C*15:02	B*51:01	1.1	90
				American Samoa	A*02:01	C*15:05	B*40:08	1.0	50
				Taiwan Rukai	A*02:01	C*15:02	B*56:01	1.0	50
				Israel Druze	A*02:01	C*15:05	B*44:02	1.0	100
				Irish	A*02:01	C*15:02	B*51:01	0.6	1000
		*A*11*	*C*12:02*	China Yunnan Hani	A*11:01	C*12:02	B*15:07	9.0	150
				Taiwan Pazeh	A*11:02	C*12:02	B*27:04	8.2	55
				India New Delhi	A*11:01	C*12:02	B*52:01	5.7	53
				Papua New Guinea Haruai	A*11:01	C*12:02	B*27:04	5.6	54
				Taiwan Atayal	A*11:02	C*12:02	B*27:04	3.8	106
				Papua New Guinea Madang	A*11:01	C*12:02	B*27:04	3.7	54
				China Yunnan Han	A*11:02	C*12:02	B*27:04	3.5	101
				Philippines Ivatan	A*11:02	C*12:02	B*15:25	3.0	50
				Taiwan Yami	A*11:02	C*12:02	B*27:04	3.0	50
				Israeli Jews	A*11:01	C*12:02	B*52:01	3.0	67
				Taiwan Siraya	A*11:02	C*12:02	B*27:04	2.9	51
				Taiwan Siraya	A*11:02	C*12:02	B*40:02	2.9	51
				Taiwan Puyuma	A*11:02	C*12:02	B*27:04	2.7	50
				China Yunnan Bulang	A*11:01	C*12:02	B*15:07	2.5	116
				Papua New Guinea Wosera	A*11:01	C*12:02	B*27:04	1.6	95
				Russia Tuva	A*11:01	C*12:02	B*52:01	1.5	166
				Papua New Guinea Wosera	A*11:01	C*12:02	B*39:01	1.1	95
				Taiwan Ami	A*11:02	C*12:02	B*27:04	1.0	98
				Philippines Ivatan	A*11:02	C*12:02	B*27:04	1.0	50
	2	*A*02*	*C*12:02*	Taiwan Puyuma	A*02:06	C*12:02	B*27:04	2.8	50
		*A*11*	*C*15*	Papua New Guinea Wosera	A*11:01	C*15:02	B*40:02	20.5	95
				Australia Groote Eylandt aborigines	A*11:01	C*15:02	B*40:02	7.5	73
				Papua New Guinea Haruai	A*11:01	C*15:02	B*40:02	7.4	54
				Australia Yuendumu aborigines	A*11:01	C*15:02	B*40:02	6.2	190
				Pakistan Brahui	A*11	C*15	B*40	6.0	104
				Pakistan Baloch	A*11	C*15	B*40	5.4	66
				India New Delhi	A*11:01	C*15:07	B*40:06	4.7	53
				Iran Baloch	A*11:01	C*15:02	B*40:06	3.4	100
				Israel Druze	A*11:01	C*15:02	B*51:01	3.0	100
				Pakistan Sindhi	A*11	C*15	B*40	2.6	101
				New Caledonia	A*11:01	C*15:02	B*40:02	2.6	39
				Australia Kimberley aborigines	A*11:01	C*15:02	B*40:02	2.1	24
				Papua New Guinea Goroka	A*11:01	C*15:02	B*40:02	1.6	32
				Taiwan Pazeh	A*11:01	C*15:02	B*40:01	1.1	55
				Papua New Guinea Haruai	A*11:01	C*15:02	B*15:06	0.9	54

Fig. S13. Putative Denisovan *HLA-A-C* haplotype frequencies in modern populations. Shown are the frequencies in modern humans of haplotypes carried by the Denisovan individual. Although phase could not be unambiguously determined, there are only two possible combinations (1 and 2, left column), and thus four possible haplotypes of the Denisovan *HLA-A* and *-C*. All four haplotypes exist in modern human populations. They are shown together with the specific *HLA-B* allele associated with that haplotype in each population. HF gives the allele frequency of the haplotype. Some populations shown here were too small (n<40) to be included in the distribution maps.

Figure S14
Distribution of the Denisovan *HLA class I* alleles

A

Allele	Population	Freq. (%)
*A*11*	Papua New Guinea Madang	63.6
	China Yunnan Province Hani	61.3
	China Yunnan Province Wa	58.4
	Papua New Guinea West Schrader Ranges Haruai	55.0
	China Yunnan Province Bulang	54.3
	China Yunnan Province Nu	51.9
	China Yunnan Province Lisu	51.2
	China Guangdong Province Meizhou Han	48.6
	China Southwest Dai	41.9
	China Guizhou Province Miao	41.8
*C*15*	Australia Kimberly Aborigine	28.6
	Papua New Guinea Wosera Abelam	23.2
	Australia Yuendumu Aborigine	20.3
	Pakistan Baloch	18.9
	Pakistan Brahui	18.3
	Iran Baloch	17.7
	India Kerala Hindu Namboothiri	16.3
	India Khandesh Region Pawra	16.0
	India Kerala Hindu Nair	14.6
	India New Delhi	14.3
*C*12:02*	India Khandesh Region Pawra	15.0
	India West Bhil	14.0
	Japan (4 pops)	10.4-12.1
	Papua New Guinea Wosera Abelam	11.6
	Taiwan Pazeh	10.9
	China Yunnan Province Hani	10.3
	Pakistan Sindhi	10.2
	Pakistan Burusho	10.2
	Italy Sardinia	10.0
	Iran Tehran	9.8

B

Denisovan allele	Observations (IMGT/HLA)	
	Number	Country
*B*15:58-like*	4	China (3), Korea (1)
*B*35:63*	1	Korea

Fig. S14. Distribution of the Denisovan *HLA class I* alleles. (**A**) Denisovan *A*11*, *C*15* and *C*12:02* are common in modern humans. Ten populations with the highest allele frequencies are shown. (**B**) While Denisovan *HLA-B* alleles are absent or rare in modern humans, their closest modern genomic relatives are characteristically Asian. Number in parentheses indicate the number of observations reported in the IMGT-HLA Database (*10*). *B*35:63* is a recombinant with *B*40:01-like* exon 2 and *B*35:01* background.

Figure S15
Emergence of *A*11*, *C*15* and *C*12:02*

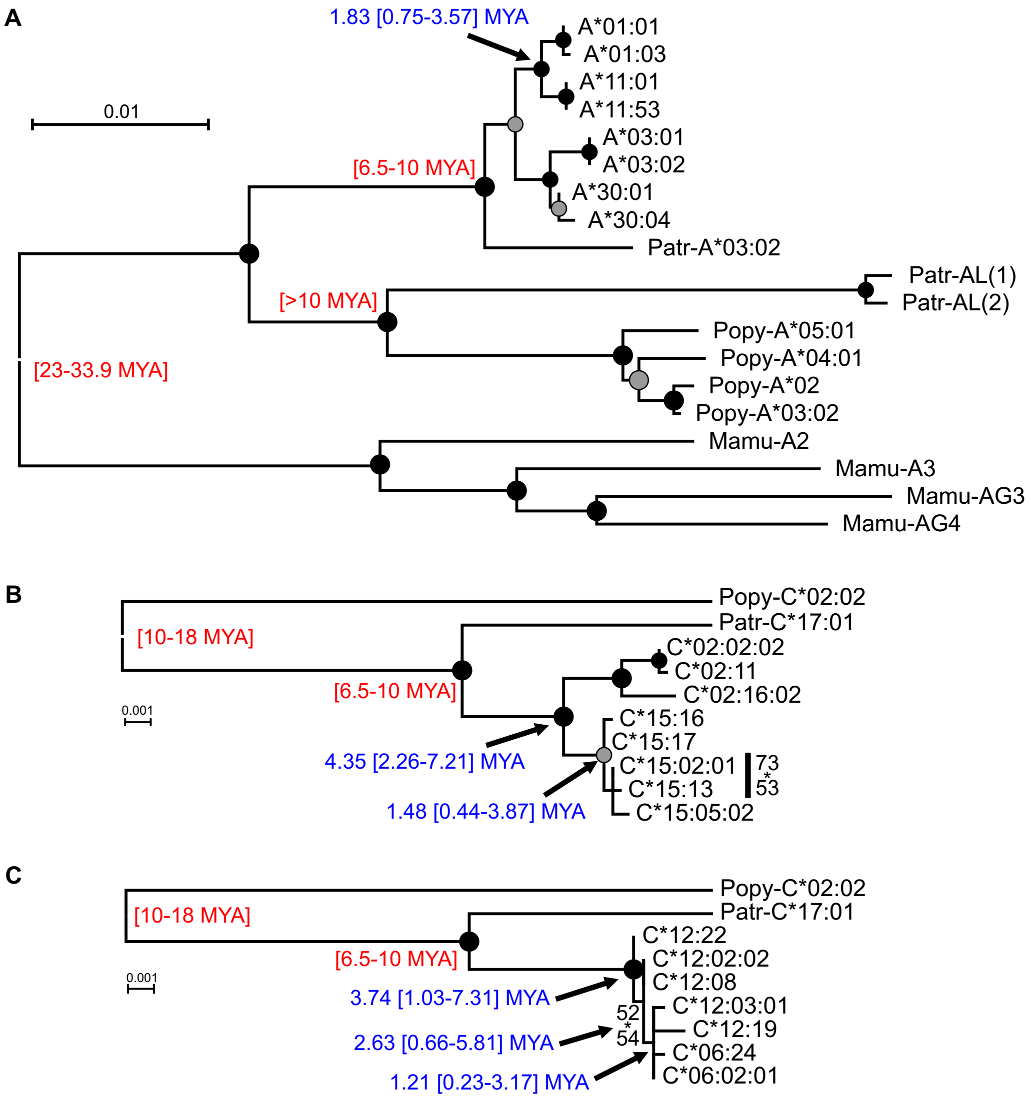


Fig. S15. Emergence of *A*11*, *C*15* and *C*12:02*. (**A**-**C**) Divergence time estimates for the emergence of *A*11* (**A**), *C*15* (**B**) and *C*12:02* (**C**). Phylogenetic analyses and display are as described in the legend to figure S3. At nodes, red characters indicate the fossil calibrations used for these analyses, while blue characters give the estimated age of the nodes (mean and 95% credibility interval). MYA, million years ago.

Figure S16
Haplotype diversity of *A*11*, *C*15* and *C*12:02*

A

	Region	Haplotypes		Pops	Max
		HLA-A	*HLA-C*	n	HF
*A*11*	Asia	**A*11[:01]**	C*01[:02]	12	7.5
		A*11[:01/02]	C*03[:02/03/04]	14	12.0
		A*11[:01/02]	C*04[:01/03]	25	33.8
		A*11[:01/02]	C*07[:01/02/04]	19	20.2
		A*11[:01/02/04]	C*08[:01]	17	22.9
		A*11[:01/02]	C*12[:02/03]	16	14.0
		A*11[:01]	C*14[:02]	5	3.9
		A*11[:01]	C*15[:02/07]	7	6.0
	Pacific Islanders / Australia	**A*11[:01]**	C*01[:02]	7	13.9
		A*11[:01]	C*03[:03/04]	6	9.3
		A*11[:01]	C*04[:01/03]	8	32.5
		A*11[:01]	C*07[:01/02]	5	14.8
		A*11[:01]	C*08[:01]	2	2.5
		A*11[:01]	C*12[:02/03]	4	5.6
		A*11[:01]	C*15[:02]	4	20.5
	Europe	**A*11[:01]**	C*03[:03/04]	2	2.2
		A*11[:01]	C*04[:01]	3	4.0
		A*11[:01]	C*05[:01]	1	0.4
		A*11[:01]	C*07[:01/02]	1	2.1
	N.Africa	**A*11[:01]**	C*01	1	2.0

B

	Region	Haplotypes		Pops	Max
		HLA-C	*HLA-B*	n	HF
*C*15*	Asia/Americas	**C*15[:05]**	B*07[:02/05]	3	4.7
		C*15[:02/03]	B*15[:01/04]	3	6.0
		C*15[:02]	B*27[:07]	1	2.5
		C*15[:04/05]	B*35[:01]	2	1.5
		C*15[:02/07]	B*40[:01/02/06]	13	11.1
		C*15[:05]	B*44[:02]	1	1.0
		C*15[:02/03/05]	B*51[:01/04]	12	9.5
		C*15[:03]	B*52[:01]	1	12.8
		C*15[:02]	B*56[:01]	3	5.0
	Pacific Islanders/ Australia	**C*15[:02]**	B*15[:06/25]	2	3.2
		C*15[:02/05/07]	B*40[:02/08/12]	6	20.5
	Europe	**C*15[:02]**	B*51:[01]	3	2.5
	N. Africa	**C*15[:02]**	B*51[:01]	1	2.5
	African American	**C*15[:05]**	B*07[:02]	1	1.1
		C*15[:02]	B*51[:01]	1	0.4
*C*12:02*	Asia	**C*12[:02]**	B*15[:07/25]	3	9.0
		C*12[:02]	B*27[:04]	9	10.9
		C*12[:02]	B*40[:02/06]	2	4.9
		C*12[:02]	B*44[:04]	1	1.4
		C*12[:02]	B*52[:01]	7	10.7
		C*12[:02]	B*53[:03]	1	5.8
		C*12[:02]	B*56[:01]	1	1.0
	Pacific Islanders/ Australia	**C*12[:02]**	B*27[:04]	3	5.6
		C*12[:02]	B*39[:01]	1	1.1
	N. Africa	**C*12[:02]**	B*52[:01]	1	2.0
	African American	**C*12[:02]**	B*52[:01]	1	0.1

Fig. S16. Haplotype diversity of *A*11*, *C*15* and *C*12:02*. Shown are all of the haplotypes in modern populations that contain one of the three *HLA* alleles, *A*11*, *C*15* or *C*12:02,* identified in the Denisovan individual (**A**) *HLA-A/C*. (**B**) *HLA-B/C*: *C*15* and *C*12:02* are found at low frequencies in Africa but because their haplotype frequencies are <1%, their haplotype structures in Africa are not reported in databases. To investigate *HLA-B/C* haplotype structures for these two alleles in Africa, we used data from a study of ~2,400 African-American donors, in which all the haplotypes were defined and reported (*62*). Allele subtypes listed in brackets were all observed.

Figure S17
HLA-A, *B*, and *C* alleles of three Neandertals

Individuals	Neandertal *HLA class I* loci				Most closely related modern allele group
	Name	Reads	Specific gene coverage		Allele(s)
			Allele #	Reads	
SLVi33.16	*HLA-A*	49	1	25	A*02
			2	5	A*26/*66
	HLA-B	57	1	14	B*07/*48
			2	21	B*51/*52/*78
	HLA-C	75	1	30	C*07:02
			2	19	C*16:02
SLVi33.25	*HLA-A*	35	1	7	A*02
			2	8	A*26/*66
	HLA-B	28	1	5	B*07/*48
			2	10	B*51/*53
	HLA-C	22	1	10	C*07:02
			2	5	C*16
SLVi33.26	*HLA-A*	28	1	6	A*02
			2	3	A*26/*31/*34/*66
	HLA-B	45	1	10	B*07/*48
			2	12	B*51/*52/*78
	HLA-C	35	1	12	C*07:02
			2	7	C*16

Fig. S17. *HLA-A*, *B*, and *C* alleles of three Neandertals. *HLA-A*, *B*, and *C* types were investigated for each of the three Neandertal individuals whose genome was sequenced (*3*). For each individual a precise type was obtained for two *HLA-C* alleles and one *HLA-A* allele. For the other alleles, the type is given as a string, listing all possible types.

Figure S18
Modern populations with two high-frequency *HLA-C* alleles

Population		*HLA-C* alleles		p of having three individuals with the same two alleles	
Name	n	Name	Frequency	Two related	All unrelated
India Kerala Kattunaikka	17	C*04	0.412	0.0475	0.0391
		C*14	0.412		
Central African Republic Mbenzele Pygmy	36	C*02	0.520	0.0365	0.0272
		C*07	0.289		
India Kerala Paniya	10	C*12	0.300	0.0360	0.0270
		C*14	0.500		
Venezuela Perja Mountain Bari	55	C*03	0.409	0.0219	0.0152
		C*07	0.303		
Mexico Oaxaca Mixtec	103	C*04	0.373	0.0218	0.0153
		C*07	0.333		
India Kerala Kuruma	15	C*04	0.300	0.0202	0.0138
		C*14	0.400		
Papua New Guinea Eastern Highlands Goroka Asaro	57	C*01	0.302	0.0200	0.0137
		C*04	0.396		
Taiwan Tao	50	C*03	0.290	0.0198	0.0134
		C*04	0.410		
Mexico Oaxaca Zapotec	90	C*04	0.306	0.0187	0.0127
		C*07	0.381		
Taiwan Tsou	51	C*03	0.402	0.0165	0.0108
		C*07	0.275		
Australia Kimberly Aborigine	41	C*01	0.375	0.0152	0.0099
		C*15	0.286		
Malaysia Negeri Sembilan Minangkabau	34	C*04	0.340	0.0144	0.0094
		C*07	0.310		
Papua New Guinea Madang	65	C*04	0.345	0.0140	0.0090
		C*07	0.302		
Taiwan Taroko	55	C*03	0.327	0.0139	0.0090
		C*07	0.318		
Mexico Oaxaca Mixe	55	C*04	0.378	0.0139	0.0088
		C*07	0.273		
Australia Groote Eylandt Aborigine	75	C*01	0.267	0.0131	0.0082
		C*04	0.377		
Cameroon Baka Pygmy	10	C*02	0.250	0.0130	0.0080
		C*07	0.400		
Taiwan Puyuma	50	C*03	0.320	0.0114	0.0071
		C*08	0.300		
Malaysia Kelantan	25	C*04	0.280	0.0096	0.0058
		C*07	0.320		
India Kerala Malapandaram	10	C*03	0.350	0.0092	0.0054
		C*14	0.250		
Sudan East Rashaida	27	C*06	0.280	0.0086	0.0050
		C*07	0.306		
Norway Sami	200	C*03	0.251	0.0080	0.0045
		C*07	0.330		
Taiwan Atayal	106	C*03	0.283	0.0076	0.0043
		C*07	0.288		
Sudan South Nuba	46	C*04	0.293	0.0076	0.0043
		C*07	0.278		

Fig. S18. Modern populations with two high-frequency *HLA-C* alleles. For modern human populations, the probability of three randomly selected, unrelated individuals having the same two *HLA-C* alleles is small. Displayed are the 24 populations (out of ~250) with the highest probabilities of producing three individuals with the same two *HLA-C* allotypes as they have two high frequency *HLA-C* alleles. Of these the top 10 have probabilities of 1-5%. Even if two of the three individuals are maternally-related (3), this number would only increase from 10 to 18. The probability calculation was made as follows: one individual randomly selected from a population having two alleles of frequencies x and y will have a probability p=2*(x*y) to have one copy of each allele. The probability of three unrelated individuals having the same allelic content is $(2*(x*y))^3$.

Figure S19
Putative Neandertal *HLA-A-C* haplotype frequencies in modern populations

HLA-A-C haplotype		Population	HF	n
HLA-A	*HLA-C*			
A*02	C*07:02	Venezuela Yucpa	26.9	73
		USA Canoncito Navajo	13.1	41
		China Yunnan Lisu	10.4	111
		Ireland Northern	5.1	1,000
		Philippines Ivatan	4.7	50
		China Yunnan Bulang	4.3	116
		China Yunnan Nu	4.1	107
		Finn	3.9	90
		China Guizhou Bouyei	3.3	109
		China Yunnan Hani pop 2	3.0	150
		Russia Bering Island Aleut	2.8	85
		Taiwan Pazeh	2.7	55
		China Southwest Dai	2.6	124
		Taiwan Saisiat	2.6	51
		Taiwan Paiwan	2.0	51
		Brazil Guarani-Nandewa	2.0	51
		Georgian	1.9	104
		India New Delhi	1.9	53
		Taiwan Siraya	1.8	51
		Taiwan Ami	1.5	98
		Vietnam Hanoi Kinh pop 2	1.5	170
		Czech	1.4	104
		Russia Tuva	1.4	166
		Poland	1.3	200
		Australia Yuendumu	1.2	190
		Zambian	1.2	43
		Iran Baloch	1.1	100
		American Samoa	1.0	50
		Taiwan Puyuma	1.0	50
		Taiwan Rukai	1.0	50
A*26	C*16:02	Europe, Asia	Rare[#]	
A*02	C*16:02	Pakistan Pathan	3.4	100
		Georgian	1.4	104
A*26	C*07:02	Taiwan Taroko	10.8	55
		Pakistan Burusho	6.0	92
		Taiwan Rukai	5.4	50
		Taiwan Bunun	5.3	101
		Pakistan Sindhi	5.0	101
		Taiwan Atayal	4.3	106
		Pakistan Pathan	3.8	100
		India New Delhi	3.8	53
		Taiwan Siraya	2.0	51
		Taiwan Saisiat	1.2	51
		Iran Baloch	1.1	100
		Filipino	1.1	94
		American Samoa	1.0	50

Fig. S19. Putative Neandertal *HLA-A-C* haplotype frequencies in modern populations. The two *HLA-A* and two *HLA-C* alleles of the three Neandertals can form four possible haplotypes: their frequencies (HF) in modern human populations are shown. #, haplotype was observed in large cohorts of registry donors (*62*) but not in anthropology studies with HF>1%.

Figure S20
Allele frequency distributions for *HLA-A*02* and *A*26/*66*

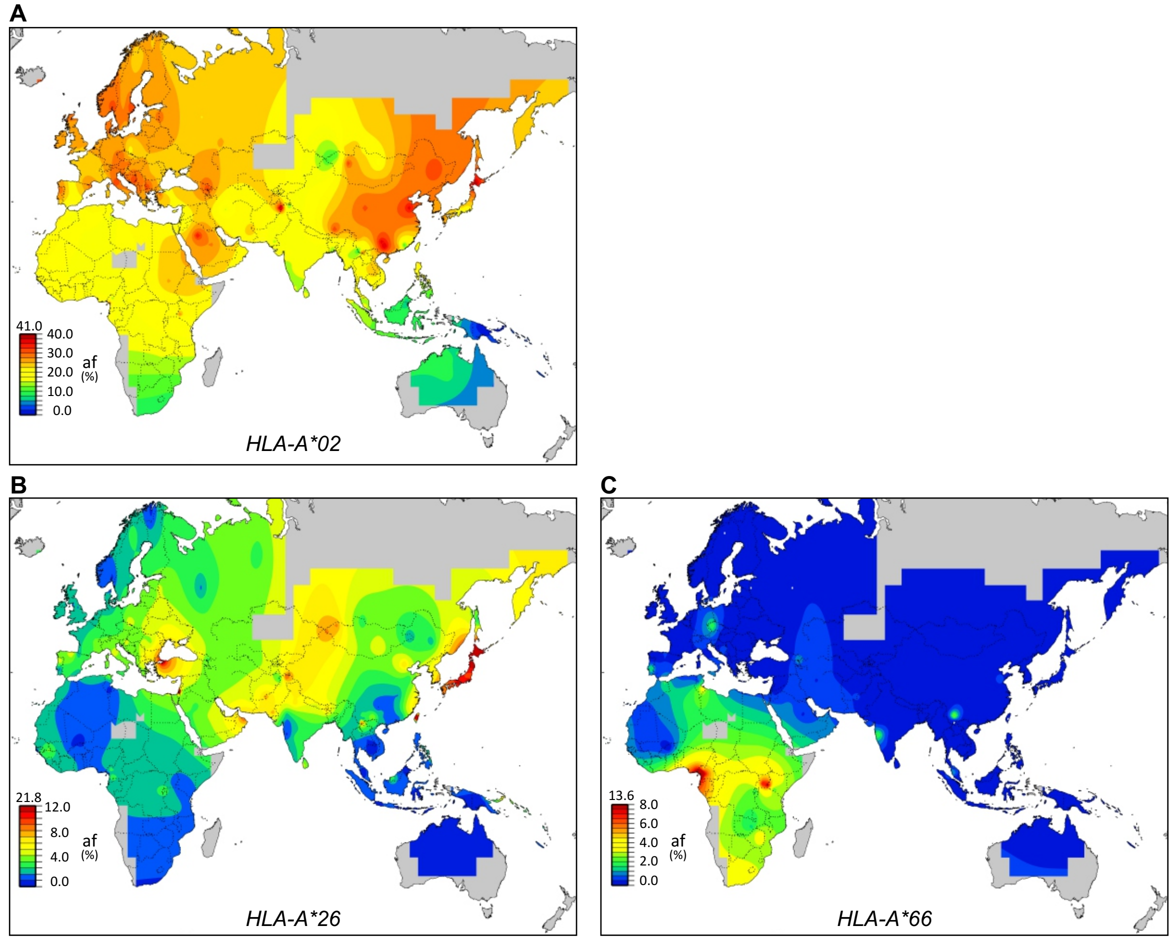


Fig. S20. Allele frequency distributions for *HLA-A*02* and *A*26/*66*. (**A**-**C**) In addition to *A*02*, a common allele group worldwide (**A**), the *HLA-A* genotype of one Neandertal allele could not be determined precisely enough for careful distribution analyses: the genotype of this *HLA-A* allele was narrowed down to two possibilities (*A*26/*66*) but while *A*26* is most common in Eurasia (**B**), *A*66* is more common in Africa (**C**).

Figure S21
Distribution of Neandertal *HLA class I* alleles

Allele	Population	Freq. (%)
*A*26*	Taiwan Aborigines (3 pops)	14-21.8
	Japan (8 pops)	10.8-19
	Israel Ashkenazi Jews	15.4
	Turkey	13.5
	Pakistan Baloch	10.3
	Oman	10.2
	Pakistan Burusho	9.8
	China Yunnan Province Jinuo	9.6
	Mongolia Hoton	9.4
	Mongolia Tarialan Khoton	9.4
*B*07:02/03/06*	Sweden Northern Sami	19.0
	Sweden Southern Sami	18.6
	Ireland Northern	17.3
	England North West	15.3
	Finland	14.4
	Austria	12.9
	Poland	12.1
	Russia Bering Island Aleut	10.0
	Croatia	9.7
	Azores Santa Maria and Sao Miguel	9.0
*C*07:02*	Venezuela Sierra de Perija Yucpa	71.4
	Taiwan Saisiat	66.7
	Papua New Guinea East New Britain Rabaul	57.1
	Mexico Oaxaca Mixtec	33.3
	China Yunnan Province Lisu	32.9
	Taiwan Taroko	31.8
	China Yunnan Province Nu	30.7
	Venezuela Perja Mountain Bari	30.3
	Papua New Guinea Madang	30.2
	Taiwan Atayal	28.8

Allele	Population	Freq. (%)
*A*66*	Cameroon Bakola Pygmy	14.5
	Kenya Luo	7.7
	Uganda Kampala pop 2	7.5
	Cameroon Yaounde	7.1
	Cameroon Beti	5.5
	Kenya Nandi	5.4
	Cameroon Bamileke	5.1
	Tunisia Ghannouch	4.3
	Zimbabwe Harare Shona	4.2
	India West Bhil	4.0
*B*51:01/08*	Bulgaria	20.9
	Japan Hokkaido Ainu	19.0
	Oman	17.8
	China Tibet Region Tibetan	16.8
	Georgia Tibilisi	16.7
	China North Han	15.8
	Georgia Svaneti Region Svan	15.6
	India Tamil Nadu Nadar	15.6
	Saudi Arabia Guraiat and Hail	15.6
	India Andhra Pradesh Golla	14.4
*C*16:02*	Pakistan Pathan	5.5
	Iran Tehran	4.9
	Georgia Tibilisi	4.2
	Spain Andalusia	3.7
	Tunisia	3.0
	Pakistan Kalash	2.9
	Iran Baloch	2.6
	India Delhi	2.5
	Italy North	2.3
	Jordan Amman	2.1

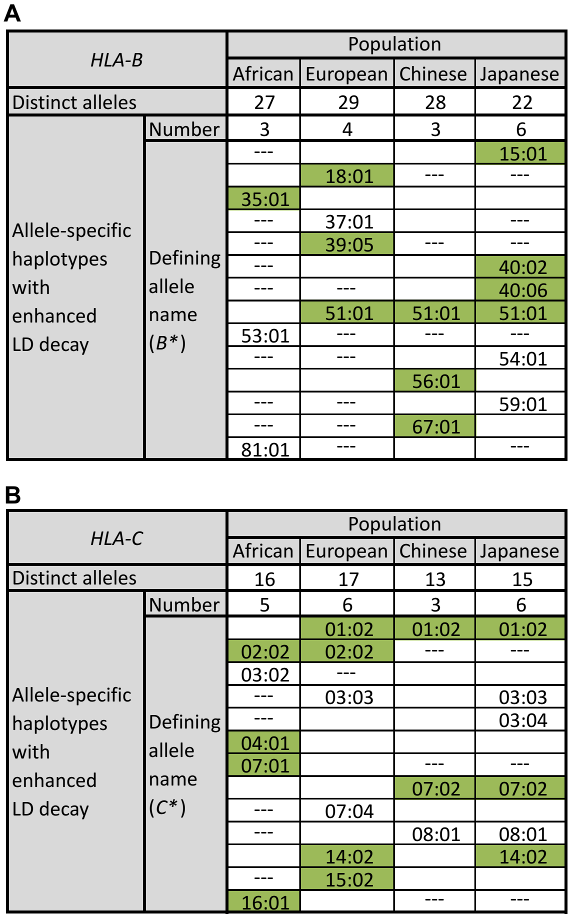
Fig. S21. Distribution of Neandertal *HLA class I* alleles. Neandertal alleles are common in modern humans. For each assigned allele or candidate allele the 10 populations with the highest allele frequencies are given.

Figure S22
Haplotype diversity of *B*07:02/03/06*, *B*51:01/08*, *C*07:02* and *C*16:02*

	Region	Haplotypes		Pops	Max
		HLA-C	*HLA-B*	n	HF
*B*51:01/08*	Asia / Americas	C*03[:04]	**B*51[:01]**	2	9.0
		C*04[:03]	**B*51[:01]**	1	2.0
		C*07[:01]	**B*51[:01]**	1	2.8
		C*14[:02]	**B*51[:01]**	12	9.1
		C*15[:02/03/05]	**B*51[:01]**	10	4.9
		C*16[:02]	**B*51[:01/08]**	2	1.4
	Europe	C*01[:02]	**B*51[:01]**	1	2.8
		C*14[:02]	**B*51[:01]**	1	1.8
		C*15[:02]	**B*51[:01]**	2	1.1
	N.Africa	C*15[:02]	**B*51[:01]**	1	2.5
		C*16[:02]	**B*51[:01]**	1	2.0
	Sub-Saharan Africa	C*16[:01/02]	**B*51[:01]**	4	4.0
*B*07:02/03/06*	Asia / Americas /Oceania	C*04[:01]	**B*07[:02]**	1	1.6
		C*07[:02/04]	**B*07[:02]**	16	6.5
		C*15[:05]	**B*07[:02]**	1	2.0
	Europe	C*07[:02]	**B*07[:02]**	4	17.0
	Sub-Saharan Africa	C*07[:01/02]	**B*07[:02]**	8	4.3
*C*07:02*	Asia / Americas /Oceania	**C*07[:02]**	B*07[:02/05]	19	6.5
		C*07[:02]	B*08[:01]	6	14.8
		C*07[:02]	B*13[:01]	3	1.6
		C*07[:02]	B*15[:03/25/32/35/36]	7	5.9
		C*07[:02]	B*18[:01]	1	0.9
		C*07[:02]	B*27[:04]	1	0.9
		C*07[:02]	B*35[:01]	2	6.0
		C*07[:02]	B*38[:01/02]	14	14.9
		C*07[:02]	B*39[:01/02/03/05/06/08/09/11]	26	71.1
		C*07[:02]	B*40[:01/02/06]	20	11.8
		C*07[:02]	B*46[:01]	3	1.8
		C*07[:02]	B*48[:01/03]	2	1.5
		C*07[:02]	B*52[:01]	4	17.6
		C*07[:02]	B*55[:02]	2	3.0
		C*07[:02]	B*56[:01/02]	7	22.2
		C*07[:02]	B*58[:01]	2	2.7
		C*07[:02]	B*67[:01]	1	1.1
	Europe	**C*07[:02]**	B*07[:02]	4	17.0
		C*07[:02]	B*39[:01]	1	1.1
	N.Africa	**C*07[:02]**	B*08[:01]	1	2.0
	Sub-Saharan Africa	**C*07[:02]**	B*07[:02]	8	4.3
		C*07[:02]	B*08[:01]	2	1.6
*C*16:02*	Asia	**C*16[:02]**	B*40[:06]	1	1.1
		C*16[:02]	B*51:[01/08]	5	5.9
	N.Africa	**C*16[:02]**	B*51[:01]	1	2.0
	Sub-Saharan Africa	**C*16[:02]**	B*51[:01]	1	1.3

Fig. S22. Haplotype diversity of *B*07:02/03/06*, *B*51:01/08*, *C*07:02*, and *C*16:02*. Diversity of the *HLA-B-C* haplotypes carrying *B*07:02/03/06*, *B*51:01/08*, *C*07:02*, and *C*16:02* in modern populations. Allele subtypes listed in brackets were all observed.

Figure S23
Enhanced *HLA class I* LD decay significantly correlates with archaic ancestry



A

HLA-B		Population			
		African	European	Chinese	Japanese
Distinct alleles		27	29	28	22
Allele-specific haplotypes with enhanced LD decay	Number	3	4	3	6
	Defining allele name (B*)	---			15:01
			18:01	---	---
		35:01			
		---	37:01		---
		---	39:05	---	---
		---			40:02
		---	---		40:06
			51:01	51:01	51:01
		53:01	---	---	---
		---	---		54:01
				56:01	
		---	---		59:01
		---	---	67:01	
		81:01	---		---

B

HLA-C		Population			
		African	European	Chinese	Japanese
Distinct alleles		16	17	13	15
Allele-specific haplotypes with enhanced LD decay	Number	5	6	3	6
	Defining allele name (C*)		01:02	01:02	01:02
		02:02	02:02	---	---
		03:02	---		
		---	03:03		03:03
		---			03:04
		04:01			
		07:01		---	---
				07:02	07:02
		---	07:04		
		---		08:01	08:01
			14:02		14:02
		---	15:02		
		16:01		---	---

Fig. S23. Enhanced *HLA class I* LD decay significantly correlates with archaic ancestry. (**A-B**) Shown for each HapMap population are (top row) the number of distinct *HLA-B* or *HLA-C* alleles present and (second row) the number exhibiting enhanced LD decay (all allele-defining SNPs (r^2>0.2) are within 500kb of *HLA-A* (*31*)). The allele names are listed (rows 3-14) and colored green when observed in archaic humans (Figs 2-3) or associated with archaic-origin haplotypes (fig. S25). *HLA-A* is shown in Fig. 4. ---, absent in the population.

Figure S24
HLA-C alleles linked to _B*73_ when _C*15_ is absent

HLA-C-B haplotype	Frequency: all haplotypes ± SE (%) (n=92)	Frequency: _B*73_ haplotypes ± SE (%) (n=47)
_C*12:02 - B*73:01_	14.9±3.7	29.2±7.3
_C*02:02 - B*73:01_	9.8±3.1	19.1±6.1
_C*06:02 - B*73:01_	5.4±2.4	10.6±4.6
_C*16:01 - B*73:01_	3.3±1.9	6.4±3.6
_C*04:01 - B*73:01_	3.1±1.8	6.1±3.6
_C*07:01 - B*73:01_	2.8±1.7	5.4±3.3

Fig. S24. _HLA-C_ alleles linked to _B*73_ when _C*15_ is absent. _HLA-C-B_ haplotype structures were determined for 46 _B*73_$^{+}$ individuals lacking _C*15_ (fig. S7), and _B*73_ haplotypes observed at least twice are listed with their estimated frequency (n=92, all haplotypes; n=47, _B*73_ haplotypes only). SE, standard error.

Figure S25
Alleles associated with Denisovan/Neandertal-like haplotypes

Main haplotype	Archaic human-like haplotypes		
2-locus haplotype	Denisovan *HLA-A-C*	Neandertal *HLA-A-B*	Neandertal *HLA-A-C*
Third locus	*HLA-B*	*HLA-C*	*HLA-B*
Alleles of the third locus present on modern human haplotypes	*07:02	*01:02	*07:02/05
	*15:01/06/07/25	*02:02	*08:01
	*27:04	*04:01	*15:01/03/25
	*39:01	*07:01/02	*18:01
	*40:01/02/06/08	*14:02	*35:01
	*44:02	*15:02/03/05	*38:01/02
	*51:01/04	*16:01/02	*39:01/03/05/06
	*52:01		*40:01/06
	*56:01		*48:01
			*51:01
			*52:01
			*56:01
			*67:01

Fig. S25. Alleles associated with Denisovan/Neandertal-like haplotypes. Shown is each category of two-locus archaic haplotypes we inferred, together with the allele at the third locus that is present on the equivalent modern-human haplotype. Every *HLA-B* allele that is found in modern humans on at least one of the four Denisovan-like *HLA-A-C* haplotypes is shown in column 3. The third and fourth columns give the *HLA-C* and *HLA-B* alleles found on the Neandertal-like *HLA-A-B* and *HLA-A-C* haplotypes, respectively.

Figure S26
Population recombination rates for the *HLA class I* region

Genomic region	Population	ρ (*10^4)	
		Mean	SD
HLA-A-B (~1.4Mb)	African	1.00	0.06
	European	0.70	0.06
	Japanese	1.21	0.13
	Chinese	1.33	0.13
HLA-A-C (~1.3Mb)	African	0.96	0.08
	European	0.70	0.07
	Japanese	1.14	0.11
	Chinese	1.18	0.12
HLA-C-B (~85kb)	African	1.54	0.31
	European	2.13	0.48
	Japanese	3.12	0.64
	Chinese	4.88	1.00

Fig. S26. Population recombination rates for the *HLA class I* region. Population recombination rates (ρ) for four populations and three *HLA class I* intervals.