

The American Journal of Human Genetics, Volume 94

Supplemental Data

**Genetic Origins of Lactase Persistence
and the Spread of Pastoralism in Africa**

**Alessia Ranciaro, Michael C. Campbell, Jibril B. Hirbo, Wen-Ya Ko, Alain Froment,
Paolo Anagnostou, Maritha J. Kotze, Muntaser Ibrahim, Thomas Nyambo, Sabah A.
Omar, and Sarah A. Tishkoff**

Figure S1: Haplotype network.

Haplotype network analysis based on 97 SNPs, identified in the three sequenced regions, and four fast-evolving microsatellites spanning a total of 197.88 kb for 509 individuals (Figure 1). Haplotypes are shown as circles and the size of each circle is proportional to the number of individuals with a given haplotype. We assigned the same weight to SNPs and STRs.



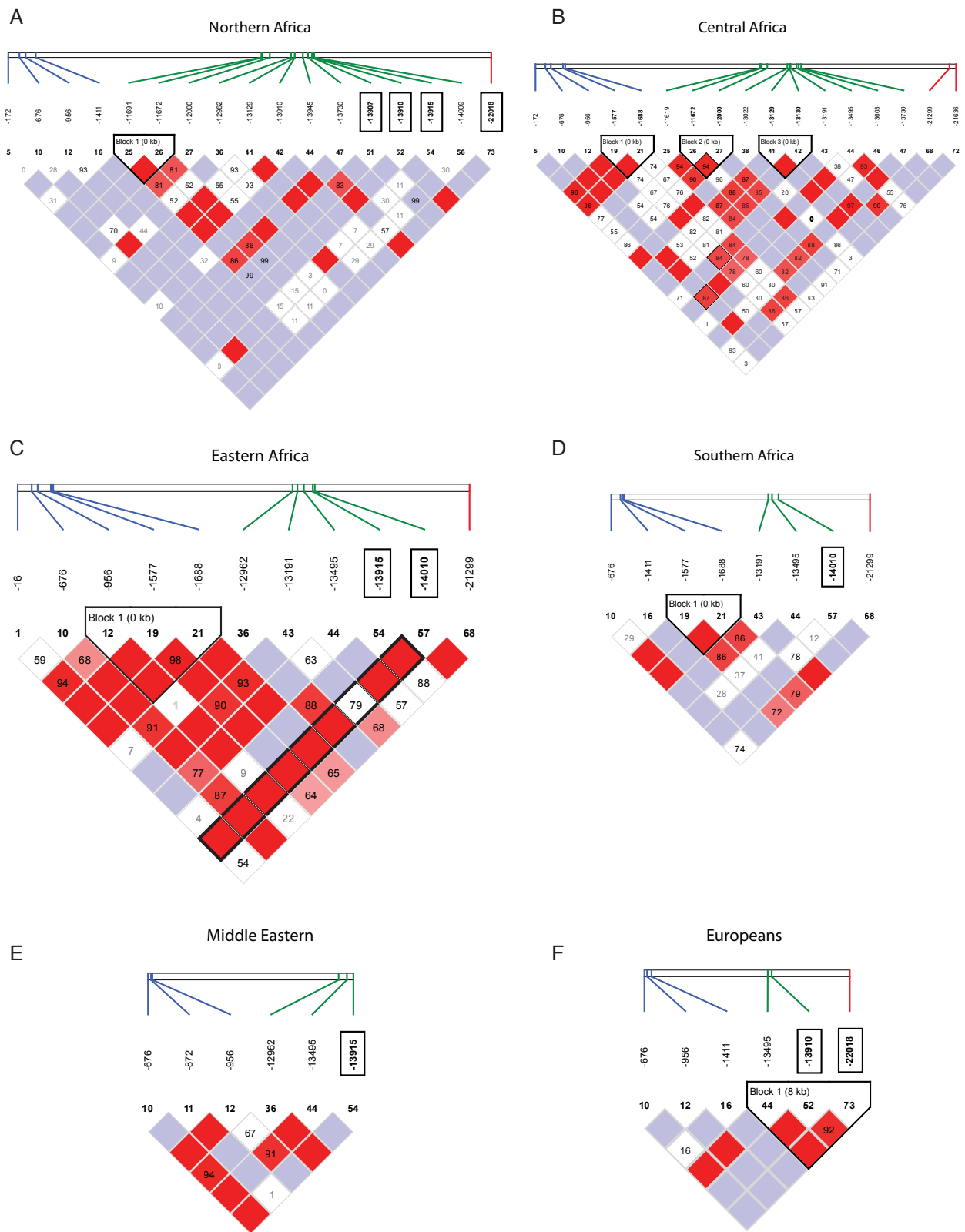


Figure S2: Linkage disequilibrium (LD) plots for major geographic regions analyzed.

Plots of the degree of LD between each pair of sequenced SNPs for major geographic regions analyzed. Markers in the promoter region of the *LCT* gene are indicated by blue lines, in intron 13 of the *MCM6* gene by dark green lines and in intron 9 of the *MCM6* gene by light green lines. SNPs associated with the LP trait are boxed in black (C -14010, G -13915, T -13910 and G -13907). A D' value < 1 with $\text{LOD} < 2$ is indicated by white. A D' value of 1 with $\text{LOD} < 2$ is indicated by blue. A D' value < 1 with $\text{LOD} > 2$ is indicated by shades of pink/red, a D' value of 1 with $\text{LOD} > 2$ is indicated by dark red.

We highlighted the long LD associated with C-14010 in the eastern Africans with a black box.

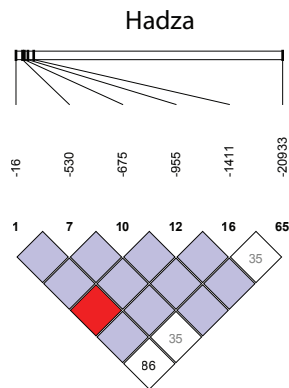
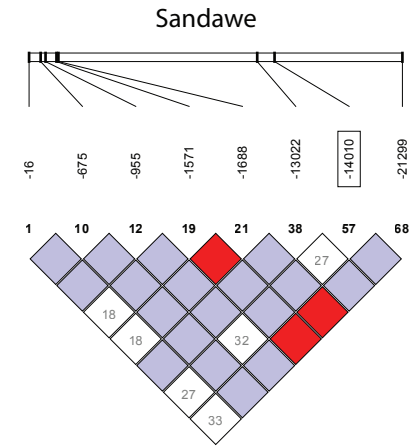
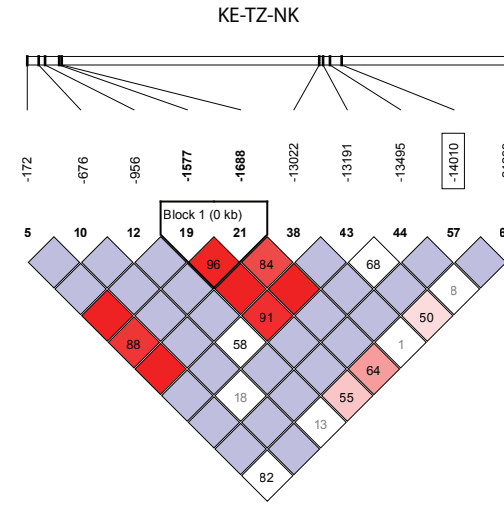
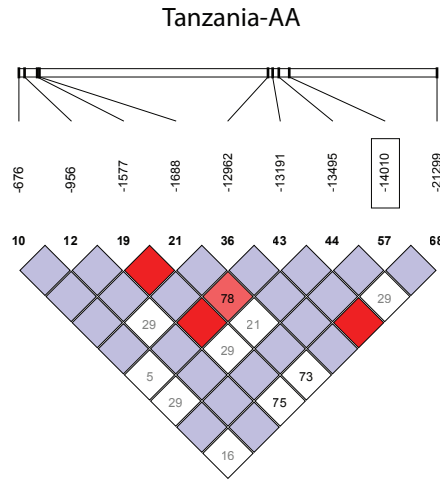
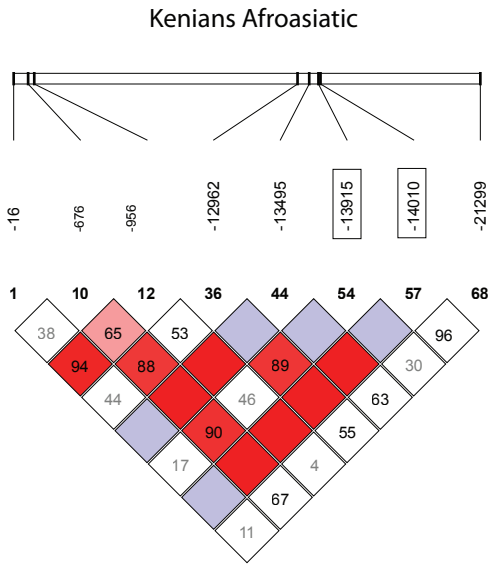
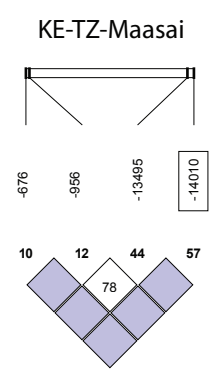
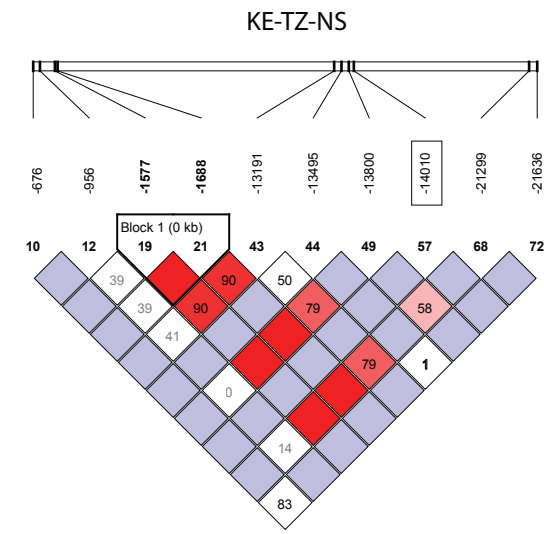
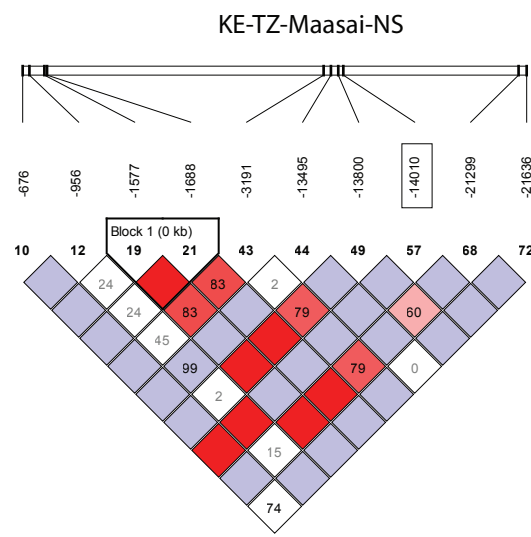
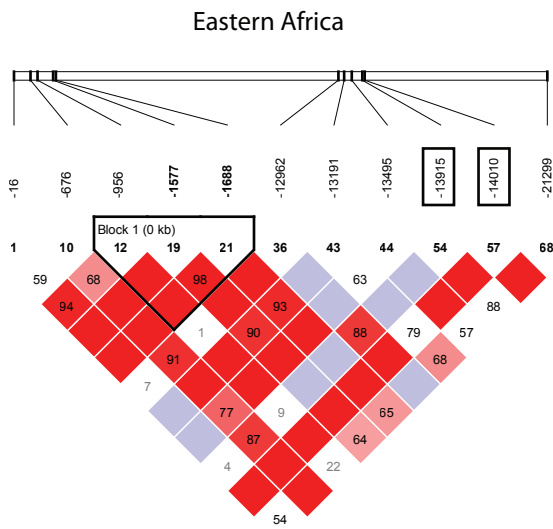


Figure S3: Detailed Linkage Disequilibrium (LD) plots for Eastern Africa populations. These plots have been created with the same criteria as Figure S2. Populations were grouped based on their genetic, linguistic and cultural affiliations (Table S1).

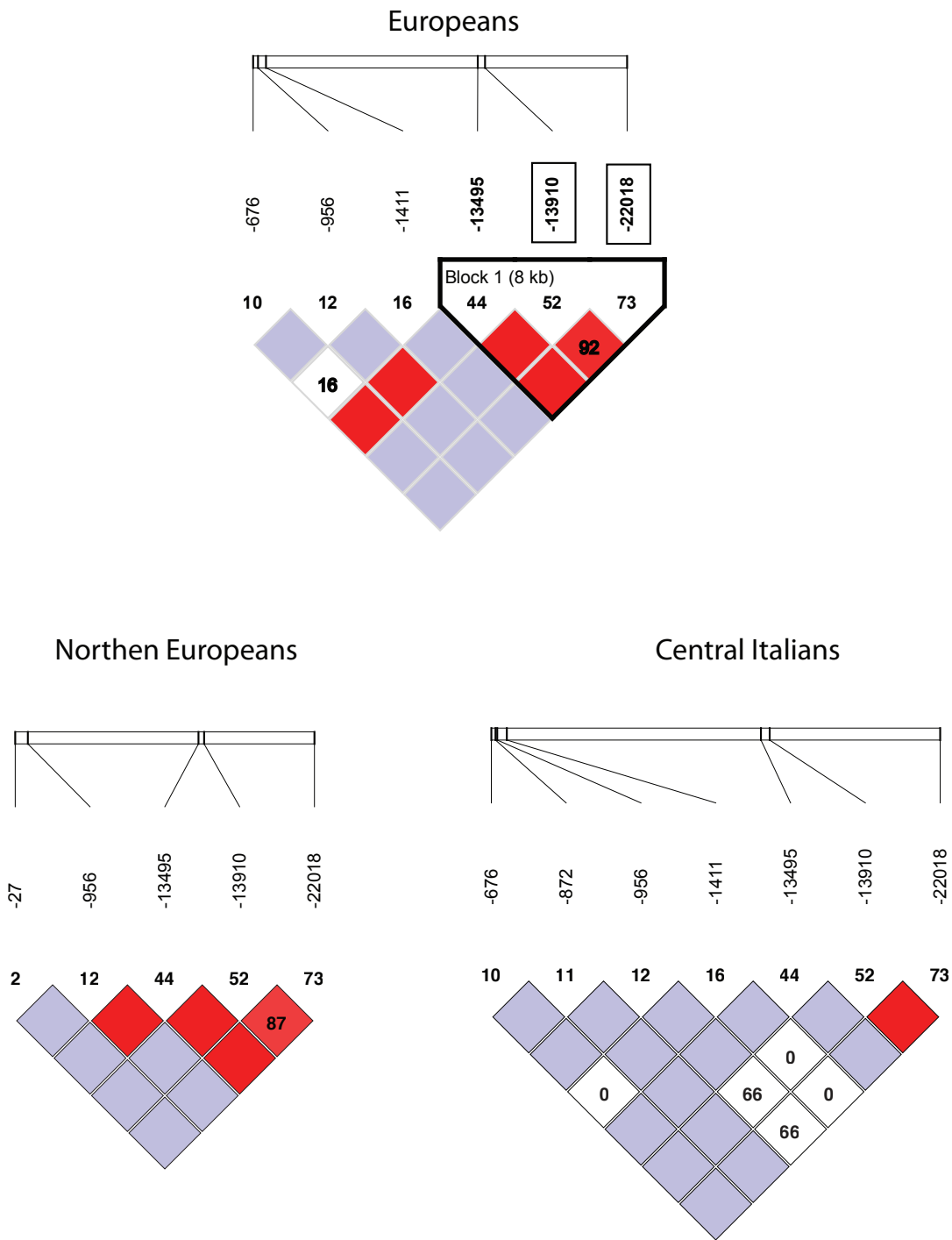


Figure S4: Detailed Linkage Disequilibrium (LD) plots for European populations.

These plots have been created with the same criteria as Figure S2. Populations were grouped based on their genetic, linguistic and cultural affiliations (Table S1).

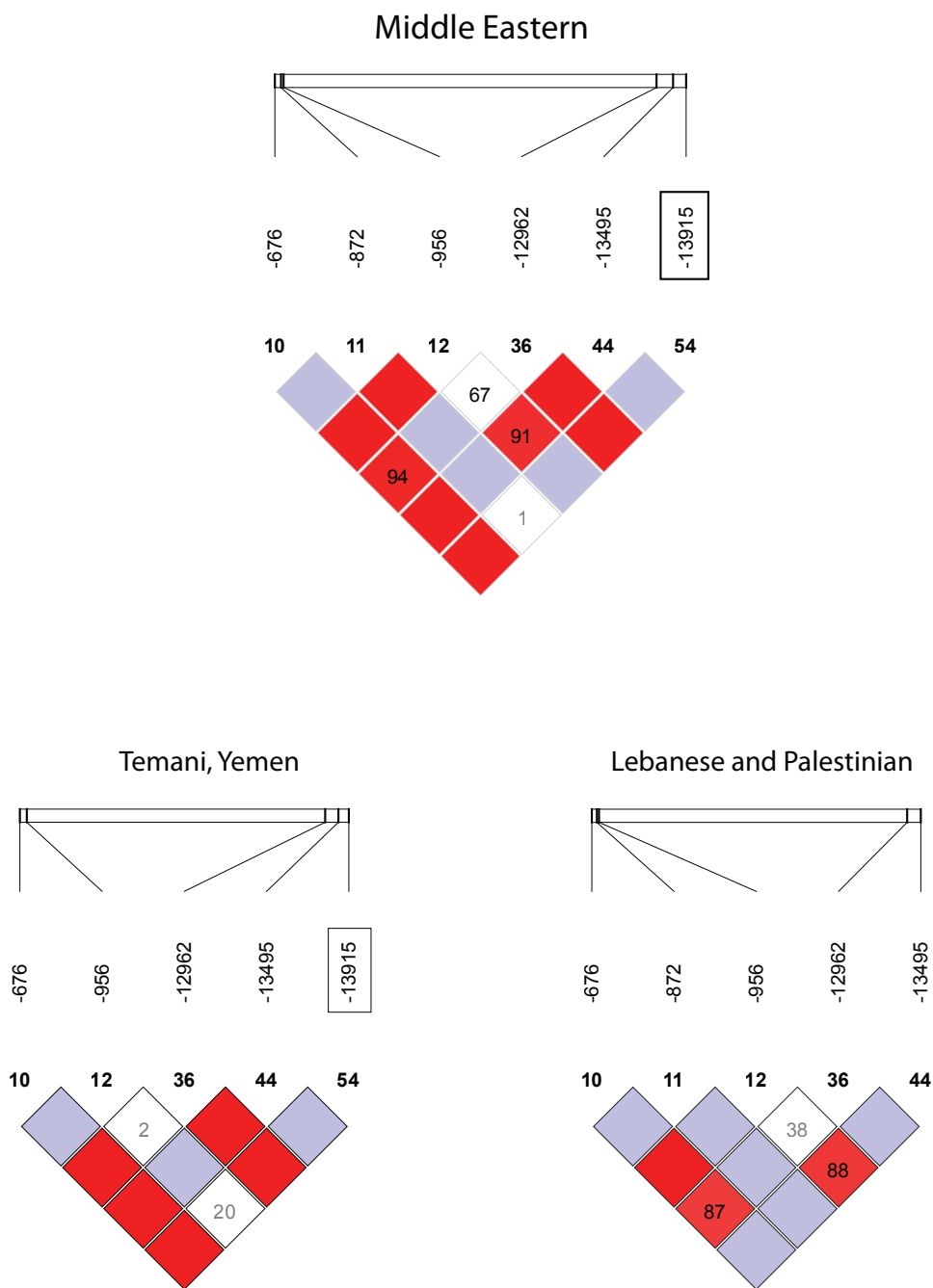
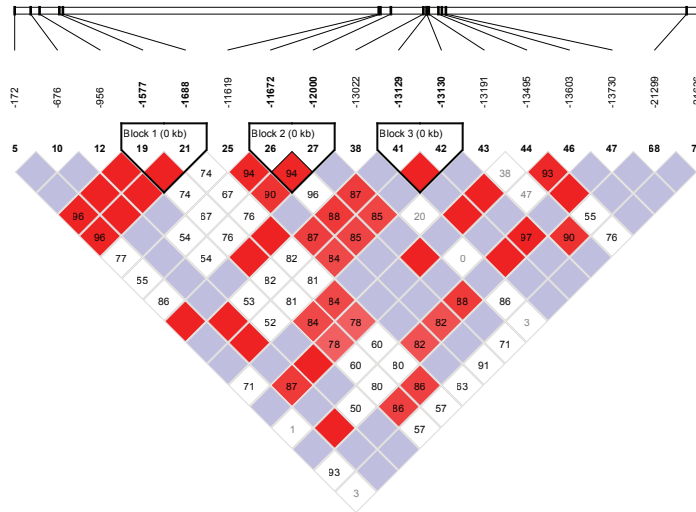


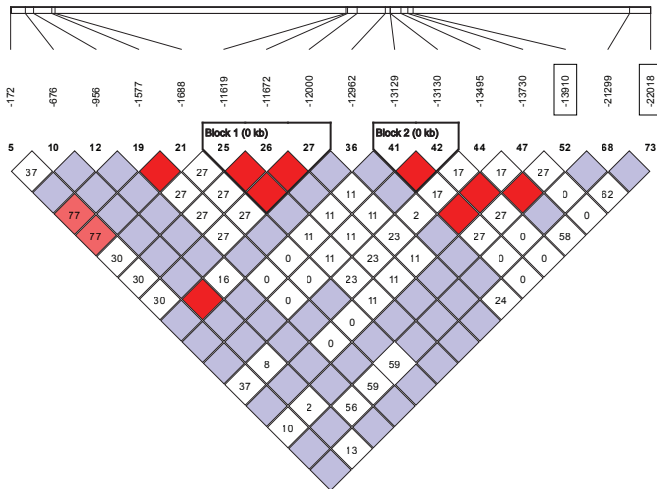
Figure S5: Detailed Linkage Disequilibrium (LD) plots for Middle East populations.

These plots have been created with the same criteria as Figure S2. Populations were grouped based on their genetic, linguistic and cultural affiliations (Table S1).

Central Africa



Arabic Baggara



Fulani

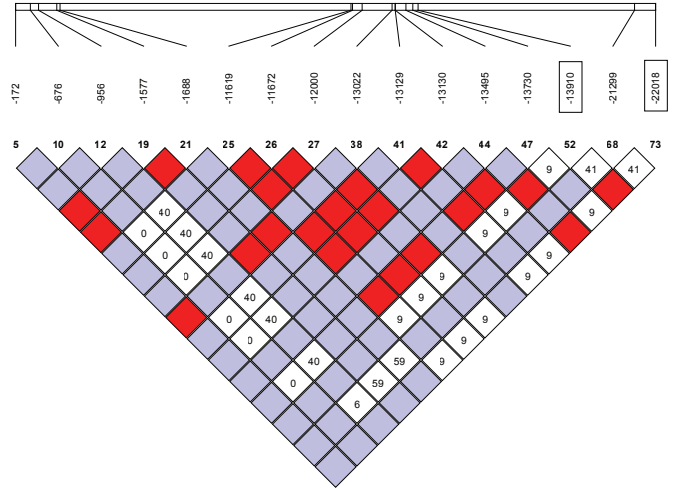
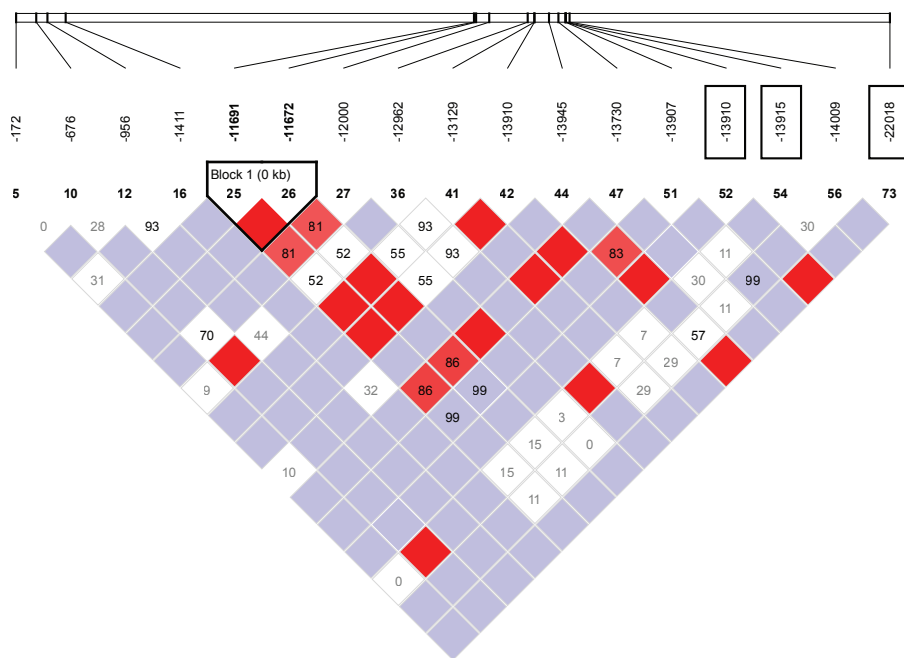


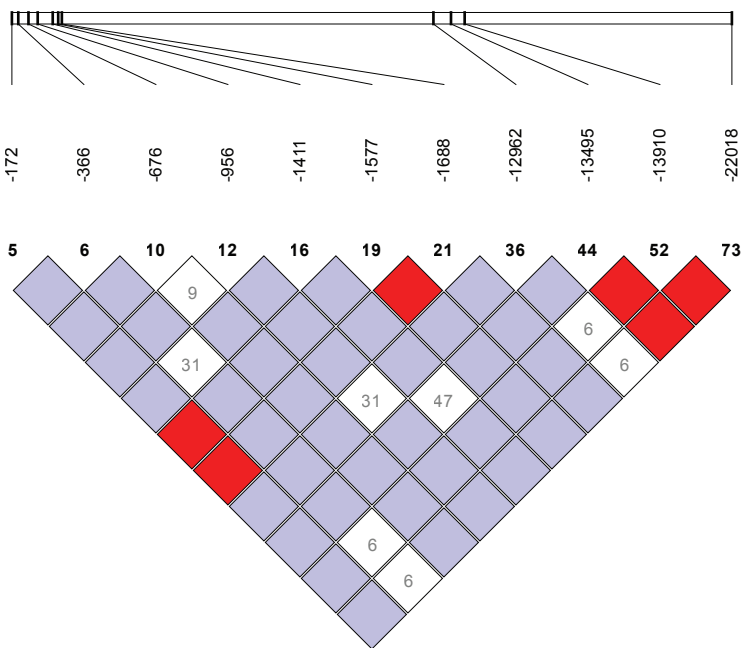
Figure S6: Detailed Linkage Disequilibrium (LD) plots for Central Africa populations.

These plots have been created with the same criteria as Figure S2. Populations were grouped based on their genetic, linguistic and cultural affiliations (Table S1).

Northern Africa



Mozabite



Sudanese Afroasiatic and Ethiopians

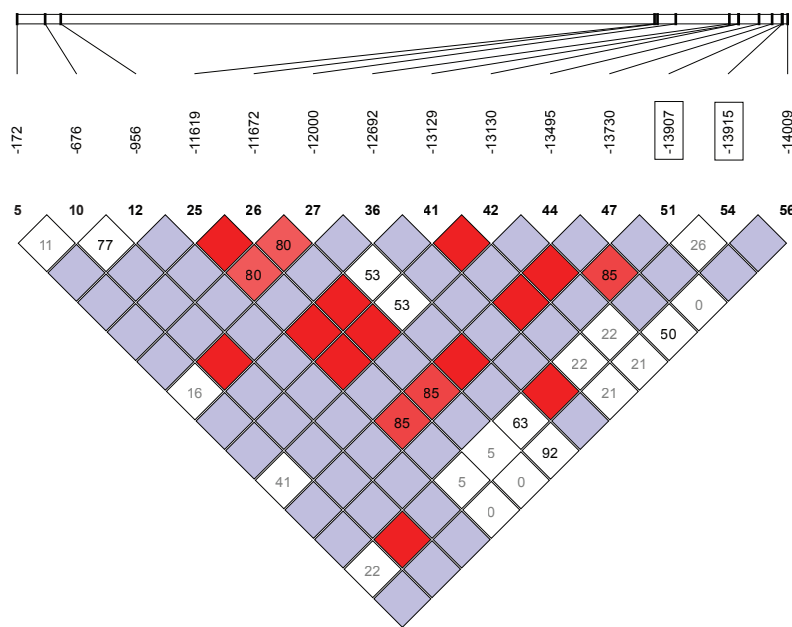
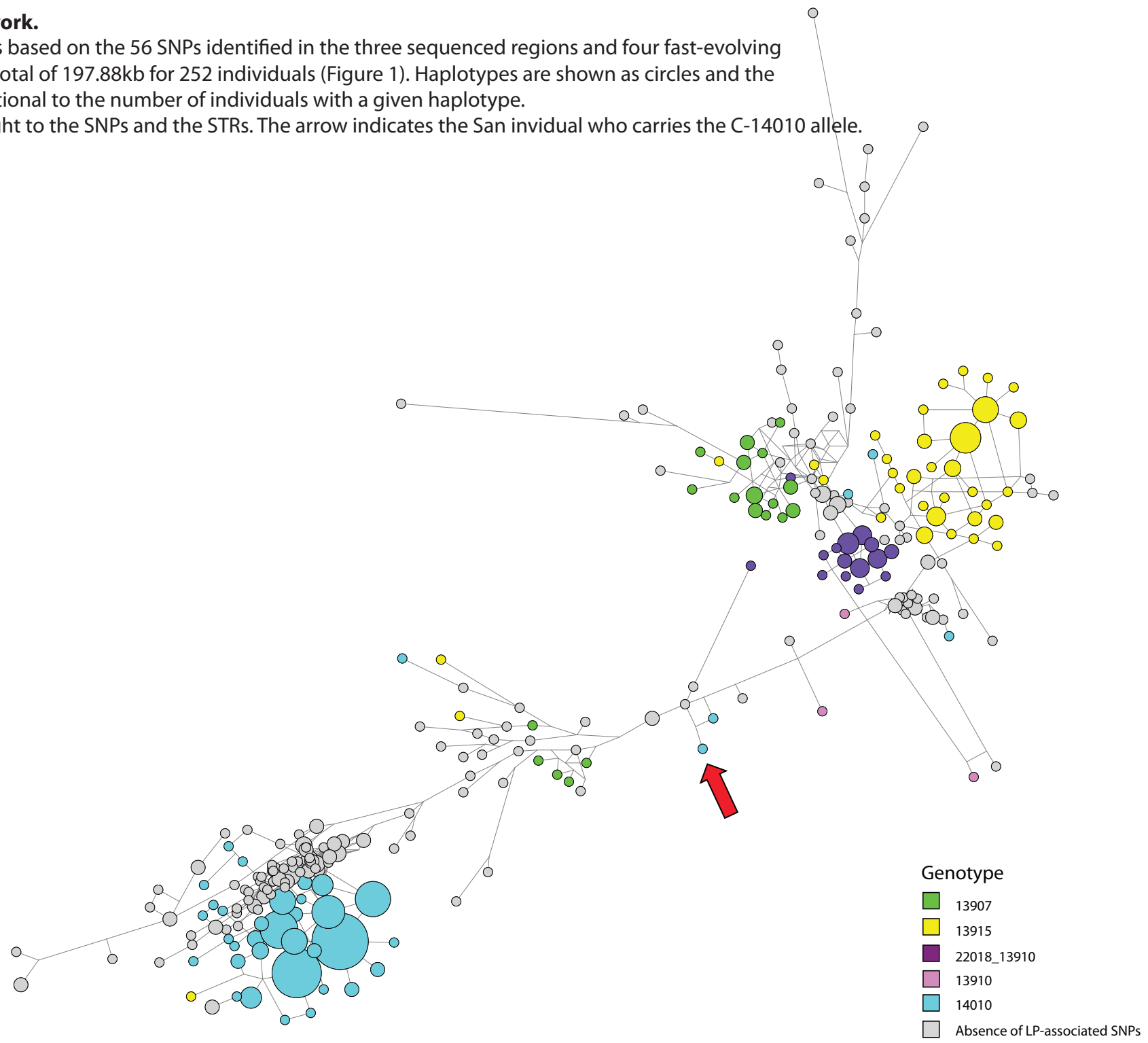


Figure S7: Detailed Linkage Disequilibrium (LD) plots for Northern Africa populations. These plots have been created with the same criteria as Figure S2. Populations were grouped based on their genetic, linguistic and cultural affiliations (Table S1).

Figure S8: Haplotype network.

Haplotype network analysis based on the 56 SNPs identified in the three sequenced regions and four fast-evolving microsatellites spanning a total of 197.88kb for 252 individuals (Figure 1). Haplotypes are shown as circles and the size of each circle is proportional to the number of individuals with a given haplotype.

We assigned the same weight to the SNPs and the STRs. The arrow indicates the San individual who carries the C-14010 allele.



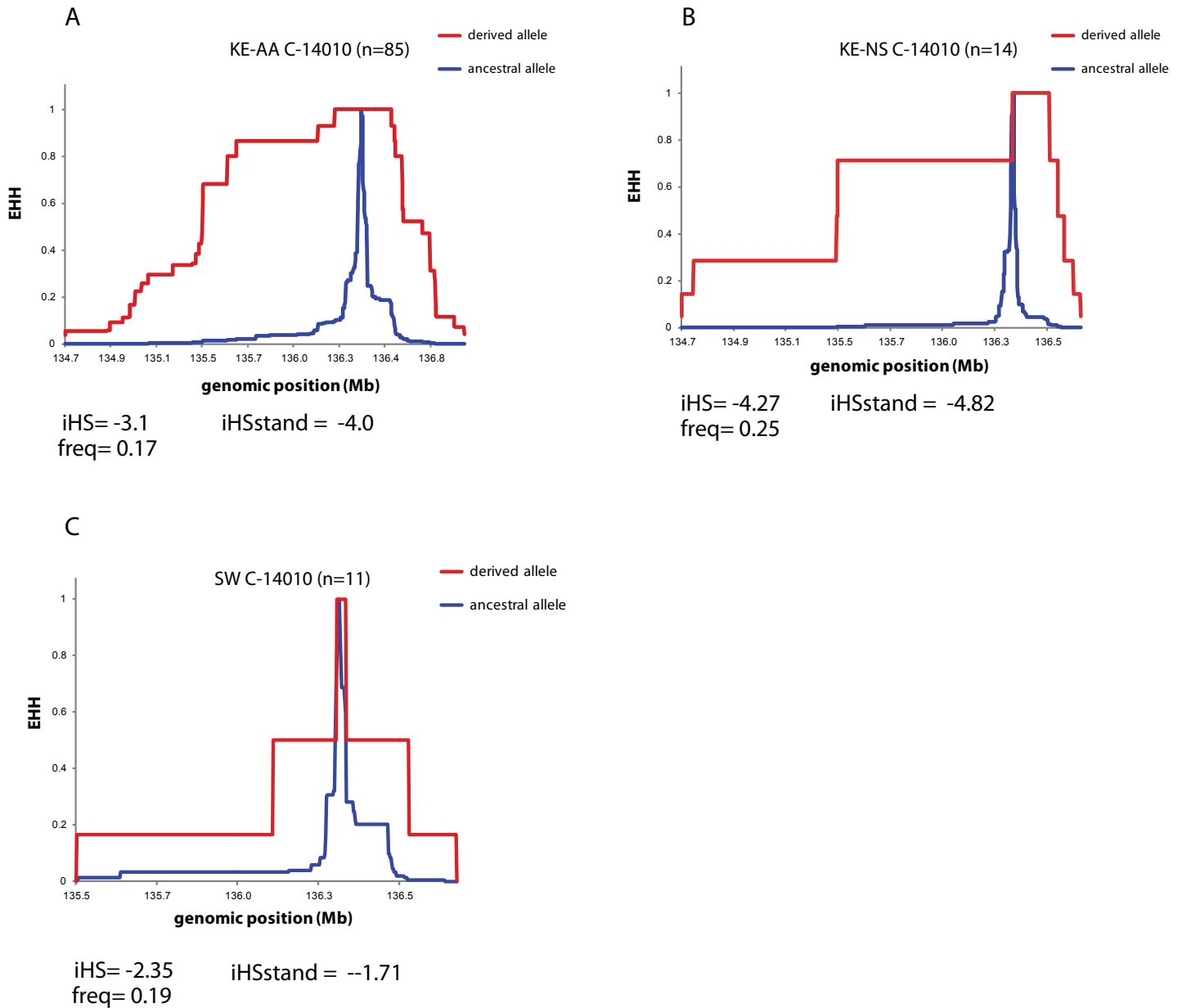


Figure S9: iHS values and EHH plots for the C-14010 LP associated SNP.

The area under the extended haplotype homozygosity (EHH) curve is expected to be much greater for a selected allele compared to a neutral allele, and we compute the integral of the observed decay of the EHH away from the core allele until the EHH reached the values of 0.05 (Voight et al. 2005). The area under the EHH curves for derived and ancestral alleles was then integrated separately and the unstandardized iHS test statistic was calculated as outlined in Voight et al. 2005. Here we present the decay of the (EHH) for chromosomes carrying the C-14010 allele in different ethno-linguistic groups from East and Central Africa: (A) Kenyan Afroasiatic (KE-AA); (B) Kenyan Nilo-Saharan (KE-NS); (C) Sandawe from Tanzanian. The ancestral allele is colored in blue, the LP-associated derived allele in red. Chromosomal positions are indicated on the X-axis and EHH values are indicated on the Y-axis. The number of individuals (n) is given for each population.

“iHS” stands for iHS unstandardized score

“iHS stand” stands for iHS standardized as described by Voight *et al.*, 2005

“freq” indicates the frequency of the derived allele at the core SNP

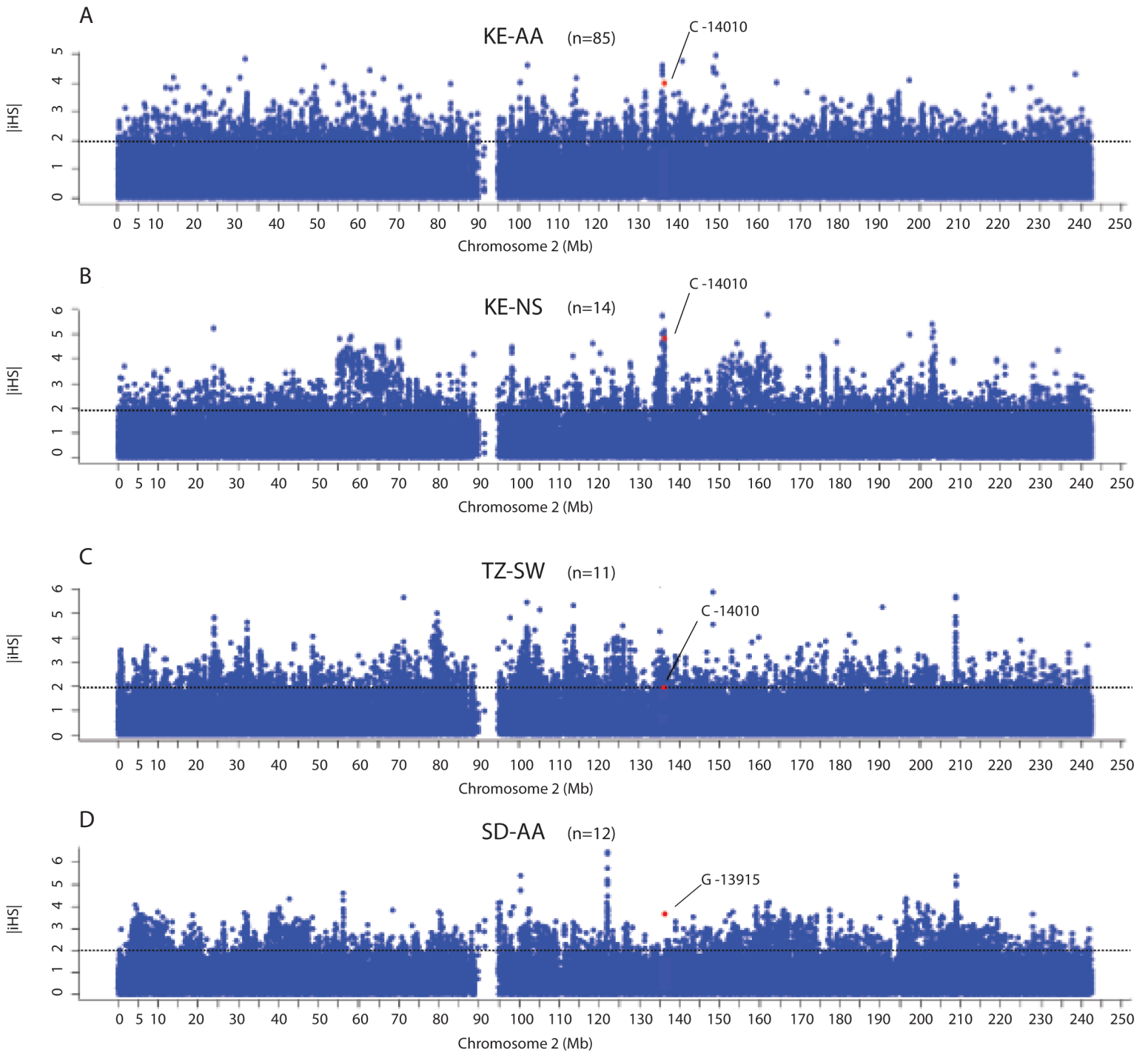


Figure S10: Plots of iHS values for SNPs identified in the promoter region and the adjacent *LCT* and *MCM6* chromosomal regions in African populations.

The absolute values of standardized iHS ($|iHS|$) are plotted against the genomic positions (NCBI Genome browser build 36.1) of SNPs identified in the *MCM6* gene and *LCT* promoter region and the neighboring genetic regions on chromosome 2 for a subset of our samples and populations from East and Central Africa: (A) Kenyan Afroasiatic (KE-AA); (B) Kenyan Nilo-Saharan (KE-NS); (C) Sandawe from Tanzania (TZ-SW); (D) Sudanese Afroasiatic (SD-AA). The dots colored in red are the $|iHS|$ estimates for the common SNPs associated with the LP trait. The 95% cutoff value (1.96) of the empirical distribution of standardized iHS is given in each plot (dashed line).

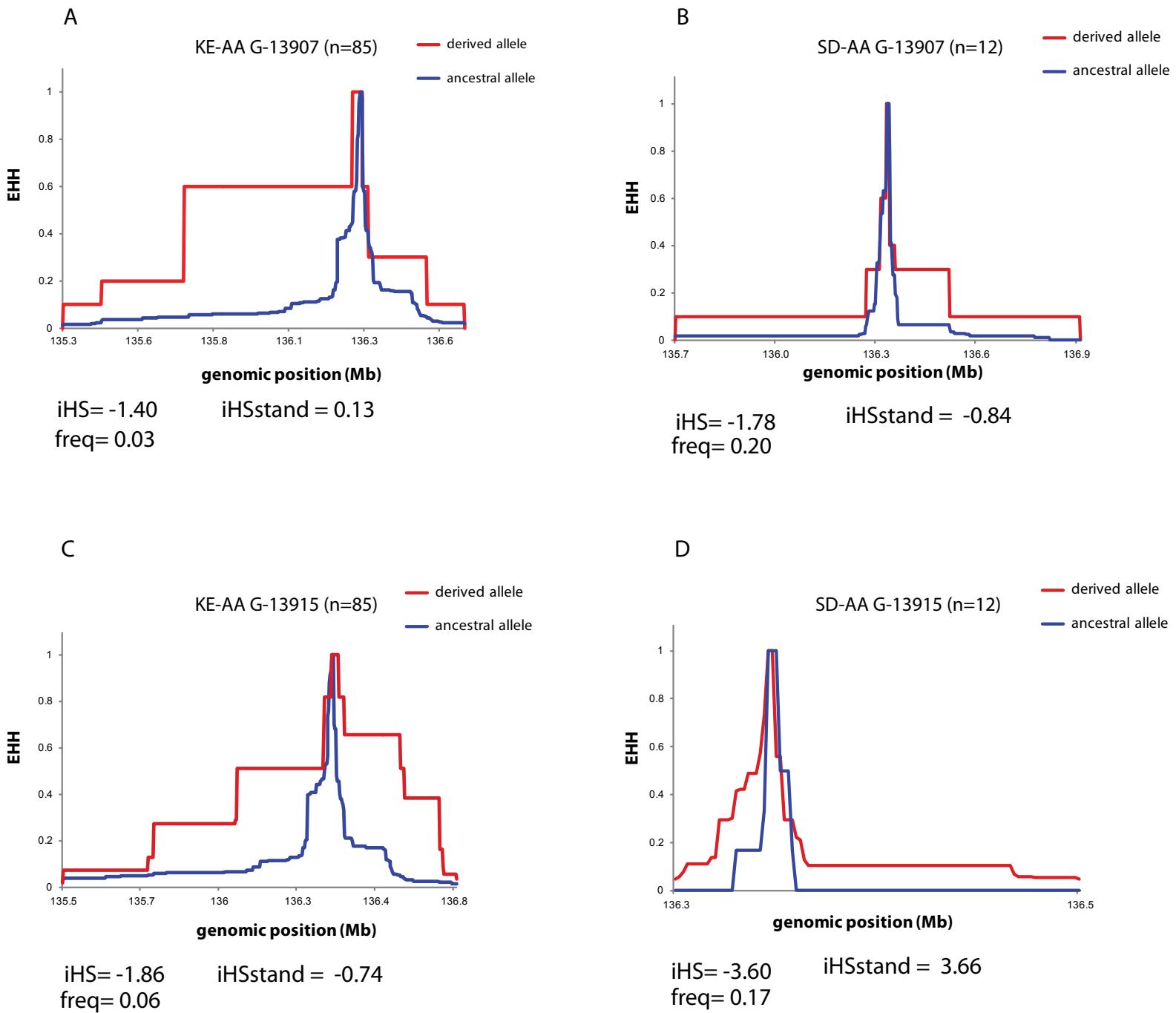


Figure S11: iHS values and EHH plots for the G-13915 and G-13907 LP associated SNPs

The decay of the EHH for the G-13915 and G-13907 variants in different ethno-linguistic groups from Eastern and Northern Africa: Kenyan Afroasiatic (KE-AA) (A and C), Sudanese Afroasiatic (SD-AA) (B and D). The ancestral allele is colored in blue, the derived LP associated allele in red. Chromosomal positions are shown on the X-axis and the EHH values are shown on the Y-axis.

“iHS” stands for iHS unstandardized score

“iHS stand” stands for iHS standardized as described by Voight *et al.*, 2005

“freq” indicates the frequency of the derived allele at the core SNP

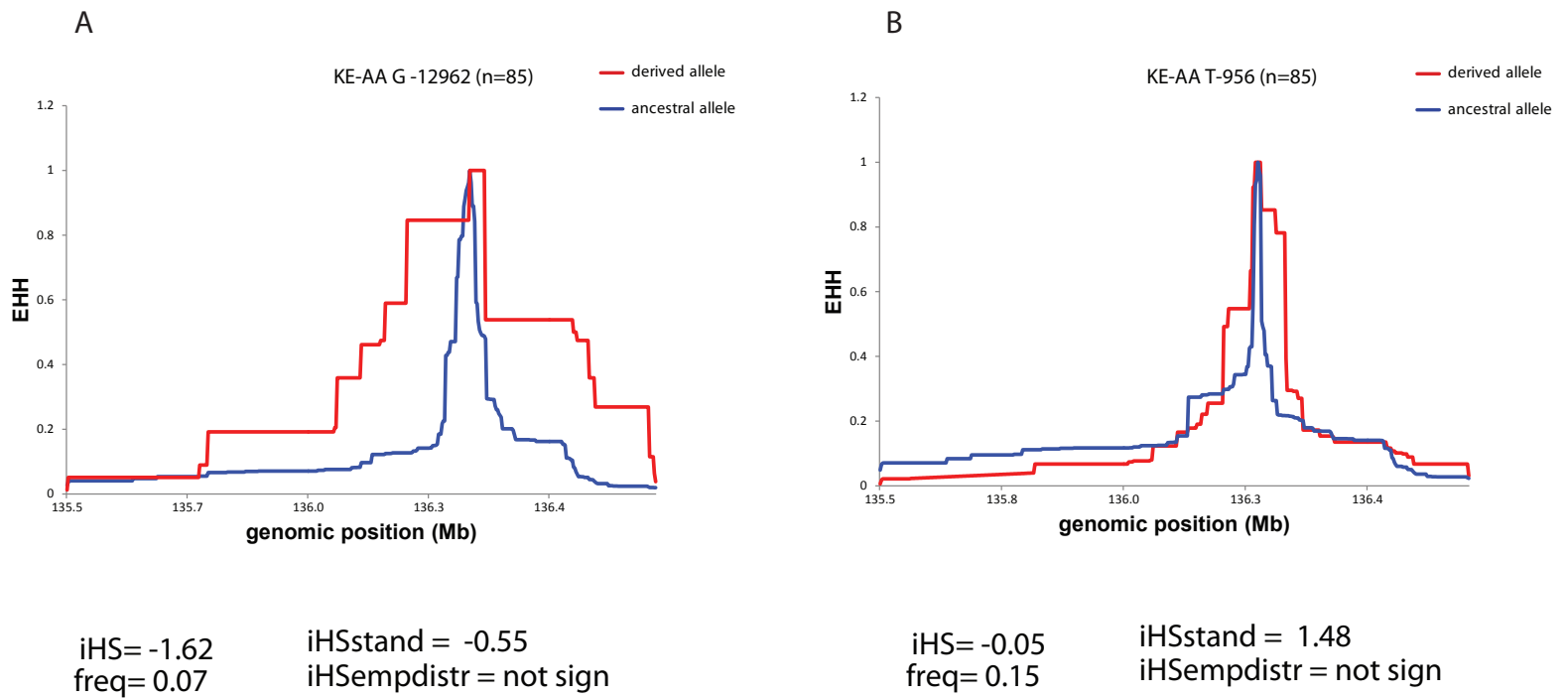


Figure S12: iHS values and EHH plots for the G-12692 and T -956 LP associated SNPs

The decay of the EHH for the G -12692 and T -956 variants in Kenyan Afroasiatic (KE-AA) speaking populations (Suppl. Figure 8A and B, respectively). The ancestral allele is colored in blue, the derived in red. Chromosomal positions are shown on the X-axis and the EHH values are shown on the Y-axis.

“iHS” stands for iHS unstandardized score

“iHS stand” stands for iHS standardized as described by Voight *et al.*, 2005

“freq” indicates the frequency of the derived allele at the core SNP