Supplemental Figures

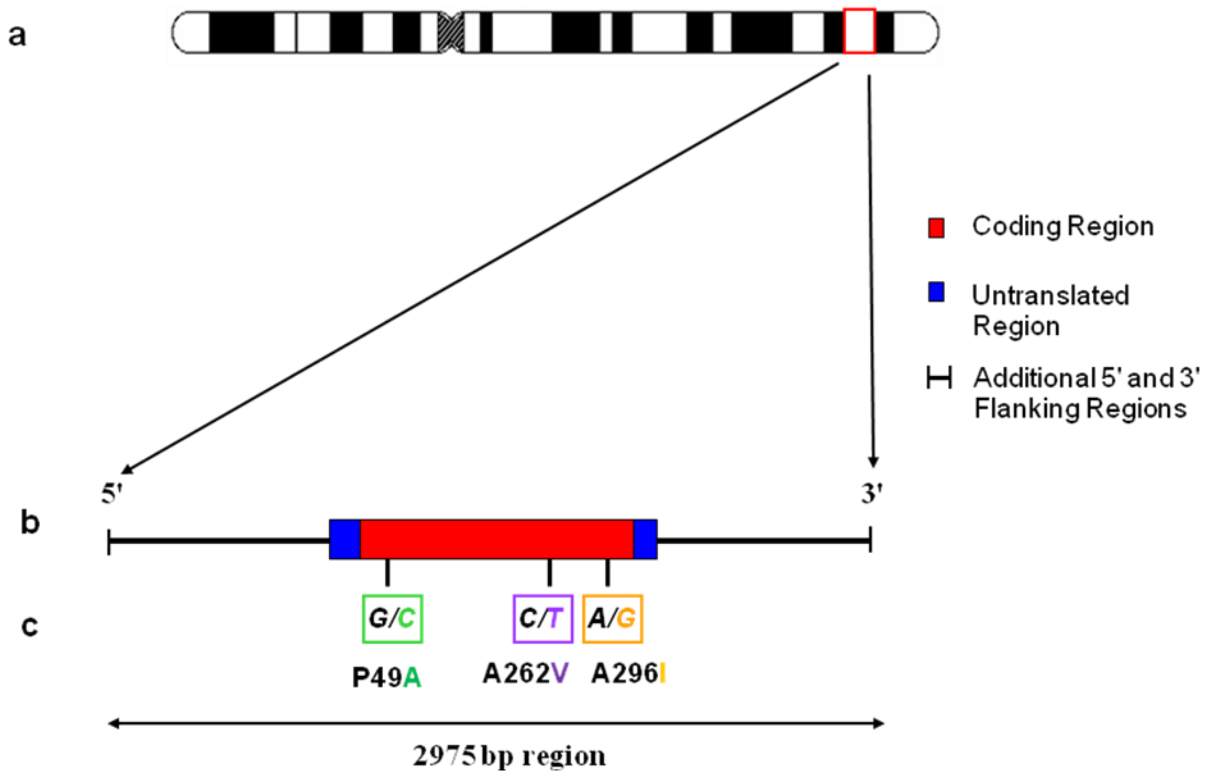


Figure S1. ***TAS2R38 g*****ene structure and genomic location**. (a) The *TAS2R38* gene is located on the long arm of chromosome (7q34) (b) *TAS2R38* consists of a single 1,002 bp coding exon (depicted in red), and two untranslated regions (in blue) on either side of the coding exon. The horizontal lines represent the additional non-coding regions sequenced upstream and downstream of the gene (c) the colored boxes indicate the location of common non-synonymous SNPs coding for amino acids substitutions at positions 49, 262 and 296. The horizontal arrow indicates the total 2,975 bp region sequenced.

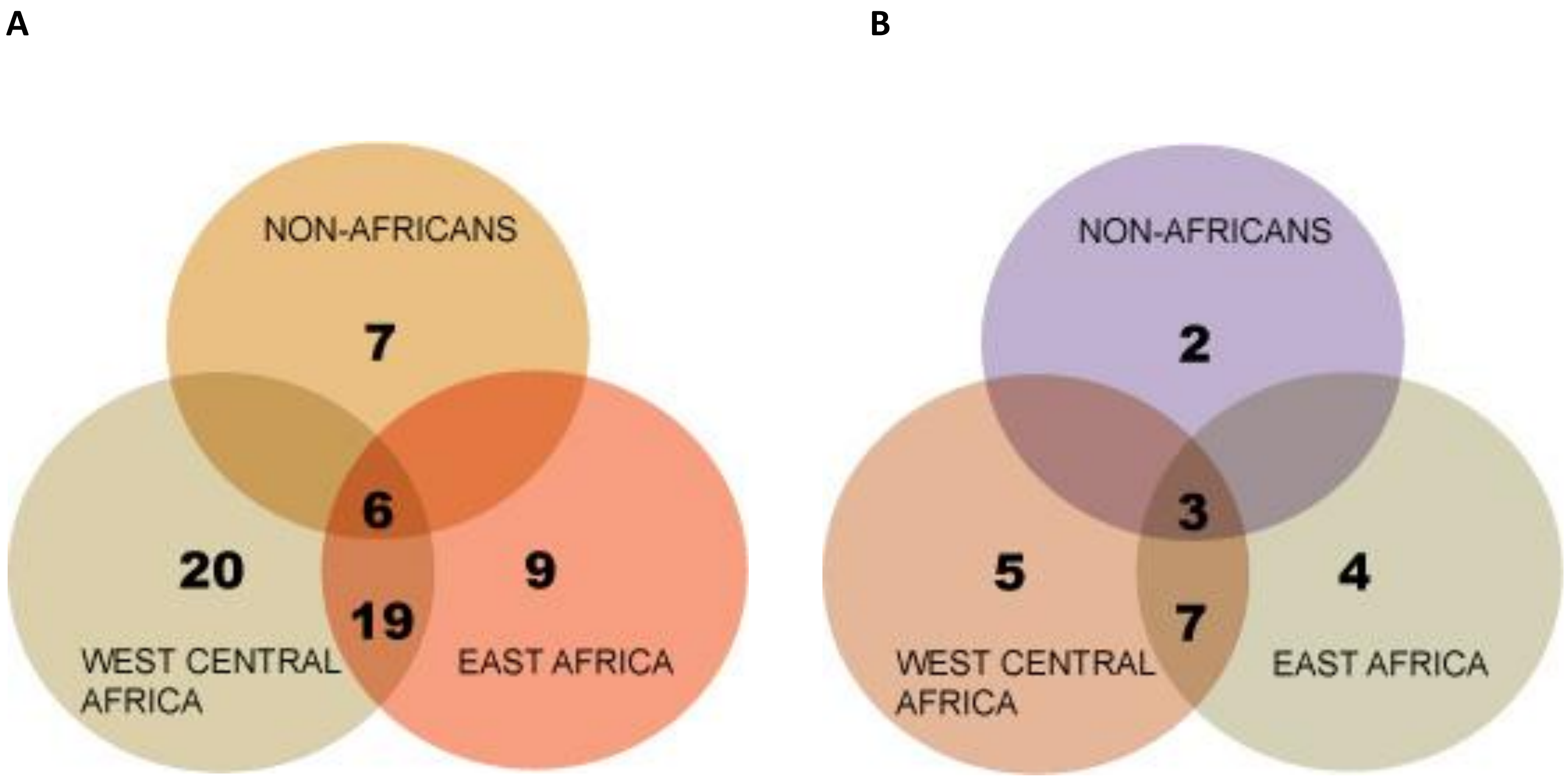
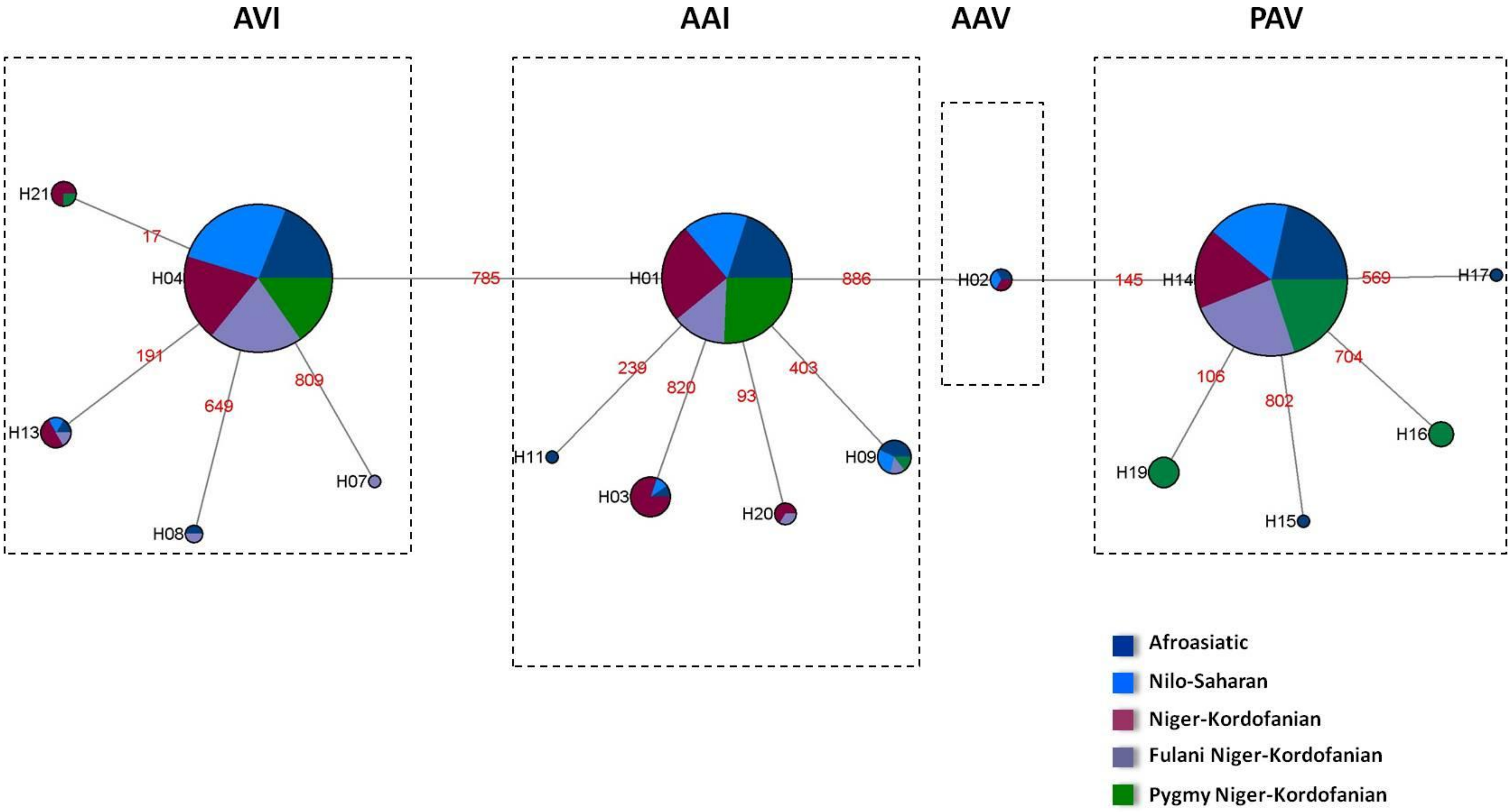


Figure S2. Distribution of Nucleotide Diversity (A) in 2,975 bp encompassing coding and non coding regions of TAS2R38, and (B) in the single coding exon of *TAS2R38*.

(A) West Central Africa



(B) East Africa

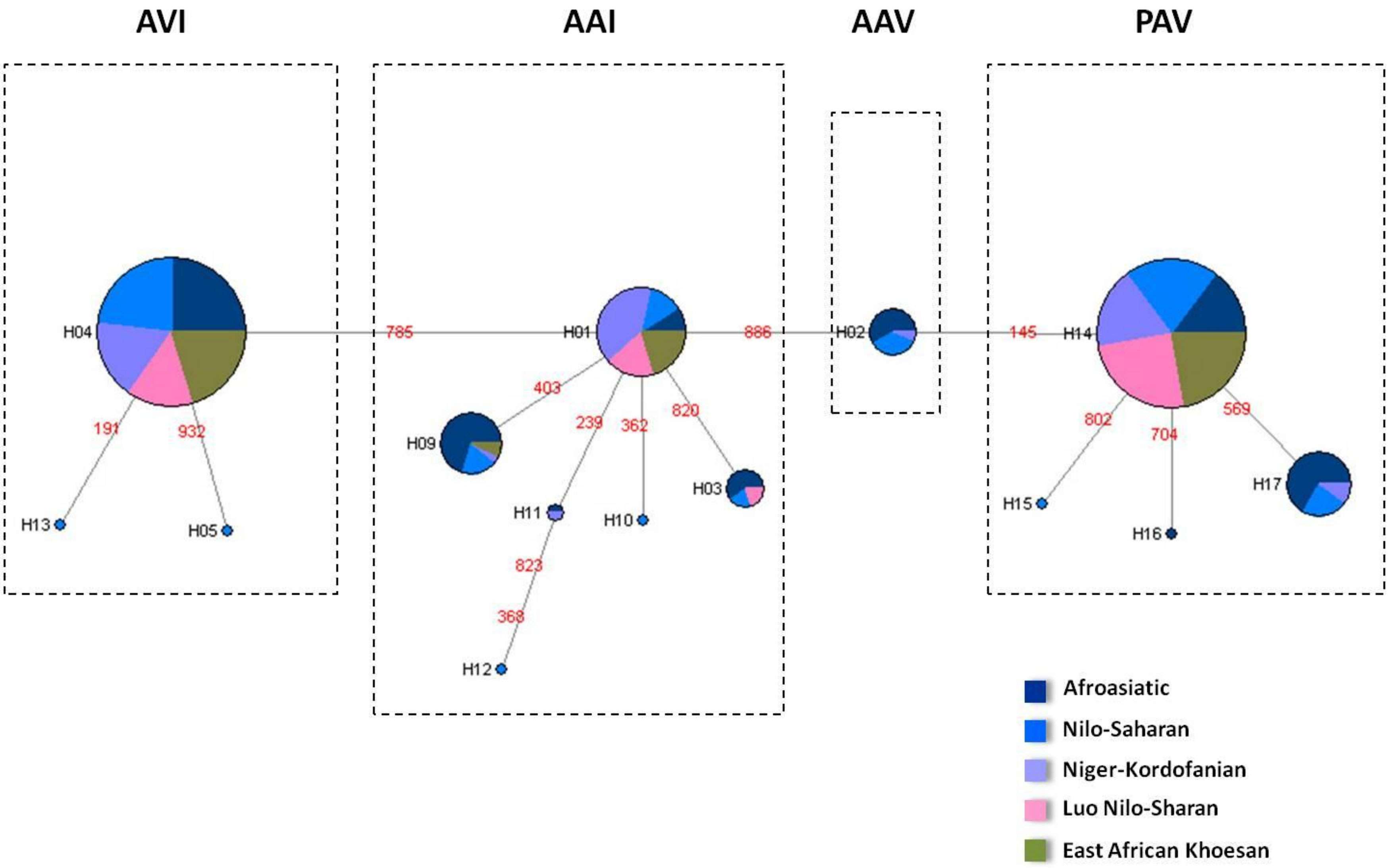
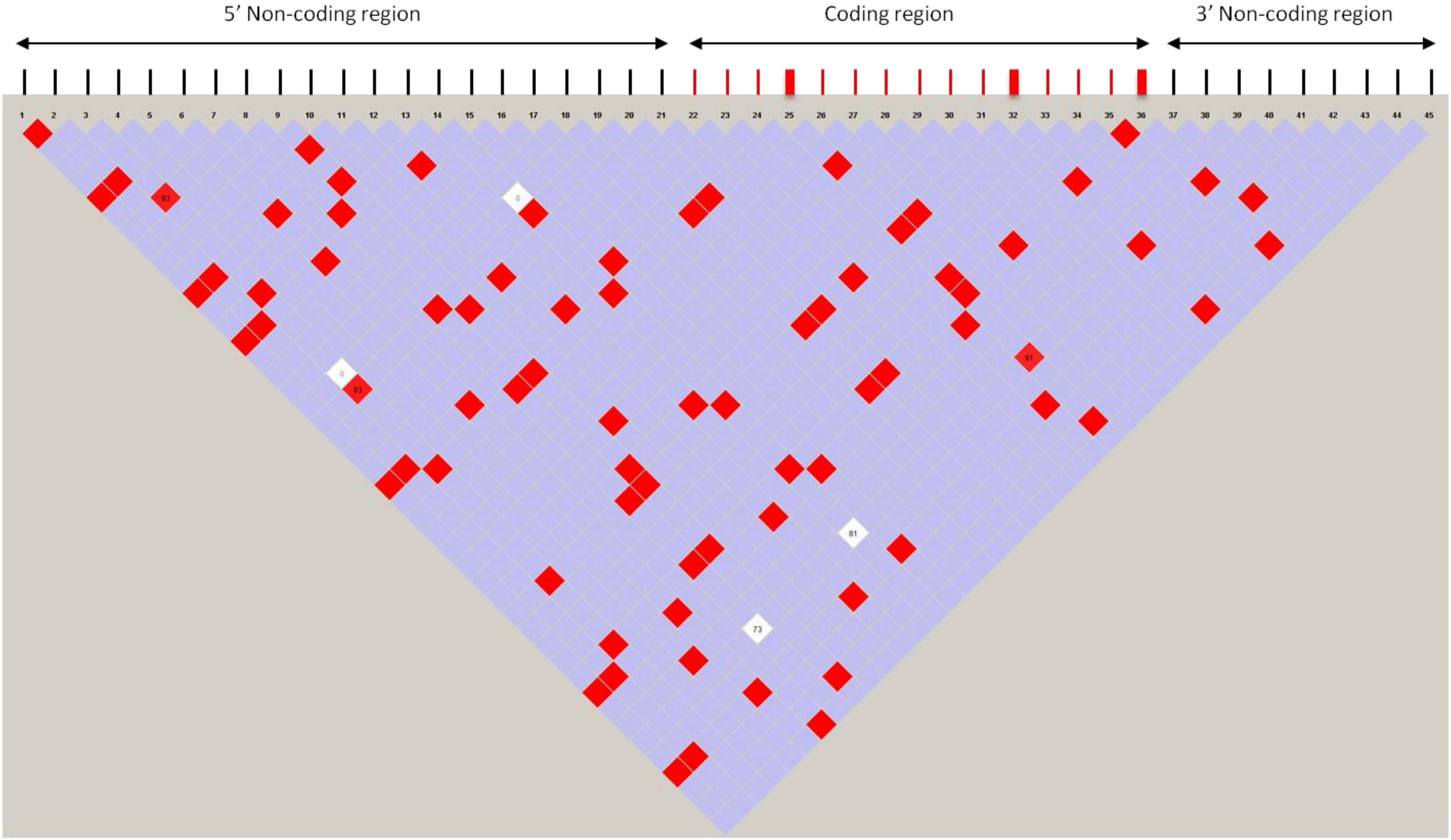
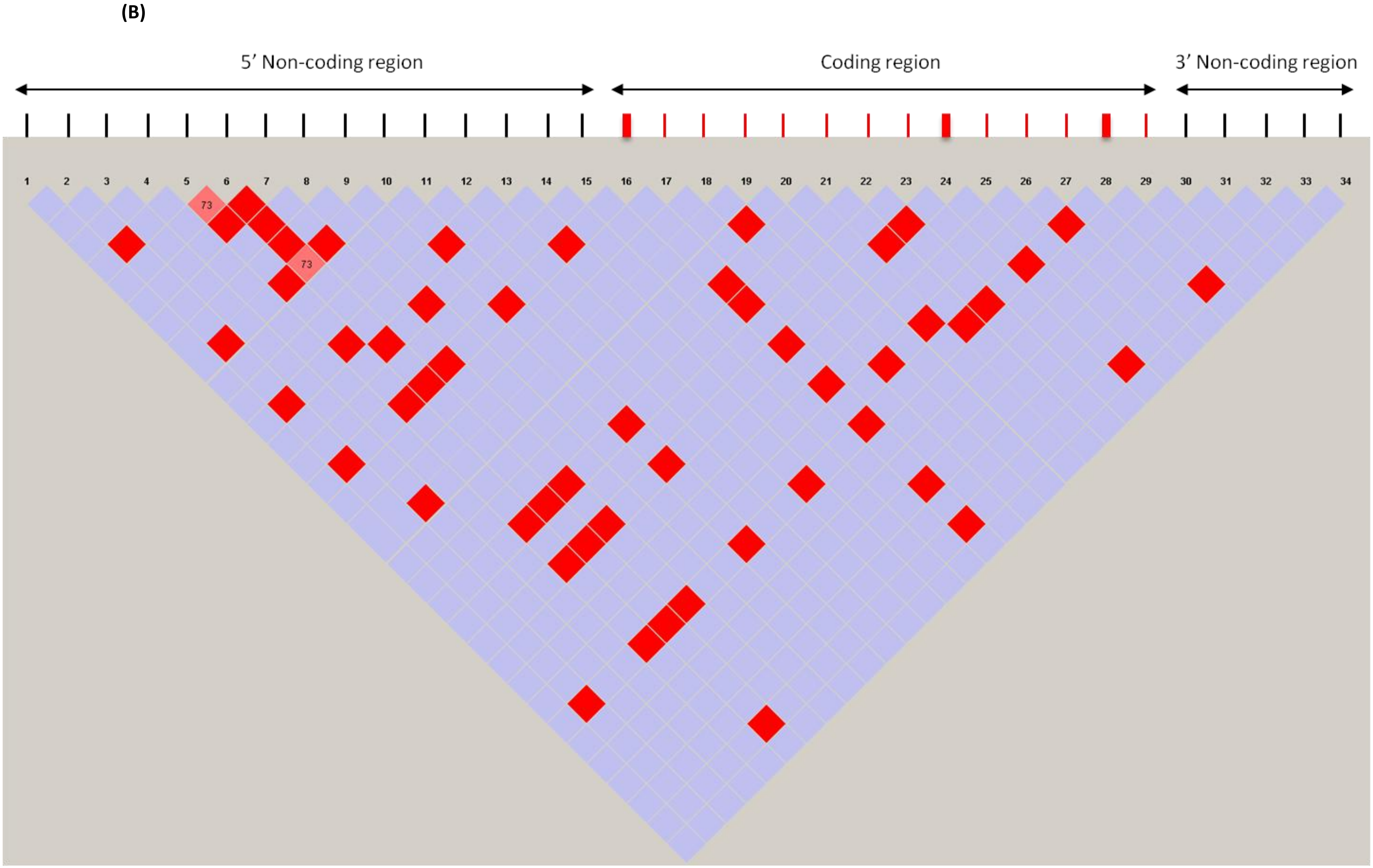


Figure S3. Networks of inferred haplotype relationships based on African data. Genealogical relationships of haplotypes were constructed using a maximum parsimony approach for (A) West Central Africa and (B) East Africa. Circles represent amino acid haplotypes, and the size of the circles is proportional to the number of chromosomes with a given haplotype. Colors within each haplotype represent the proportion of individuals within a population or geographic region with that given haplotype. The lines and numbers are the mutational changes between haplotypes.

(A)


5′ Non-coding region
Coding region
3′ Non-coding region
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
93
0
0
93
91
81
73


(B)
5′ Non-coding region
Coding region
3′ Non-coding region
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
73
73

(C)

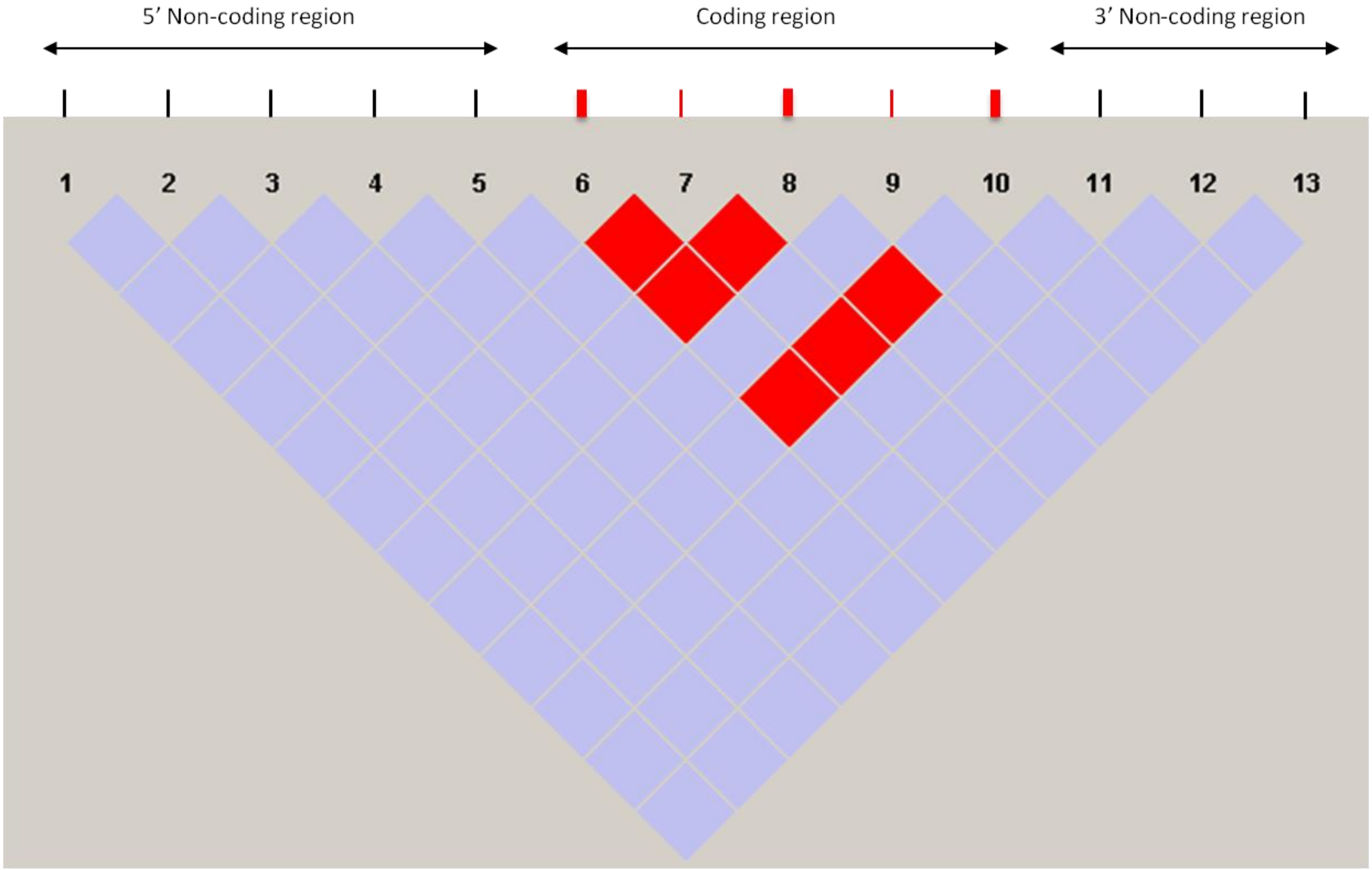

5’ Non-coding region
Coding region
3’ Non-coding region
1
2
3
4
5
6
7
8
9
10
11
12
13

Figure S4. LD plots for West Central Africans, East Africans, and non-Africans. Pairwise plots of LD across the 2,975 bp region were constructed using the Haploview program (Barrett et al. 2005) for West Central African (A), East African (B) and Non-African (C) populations. The color of each square signifies the strength of the relationship between SNP alleles. For example, bright red squares indicate statistically significant LD between a given pair of SNP loci (D'=1; LOD > 2), while shades of pink/white squares indicate little evidence of LD. Purple squares indicate high LD but with little statistical support (low LOD). Vertical red lines in bold in the coding region at the top of each figure represent the common variants associated with PTC bitter taste sensitivity. Because most of the variants within the 2,975 bp region occur at very low frequency ($q < \%5$), where q represents the minor allele, traditional measures of LD, such as D' or r^2, may not be sufficient to fully detect LD blocks across this genomic region (Goddard et al. 2000).

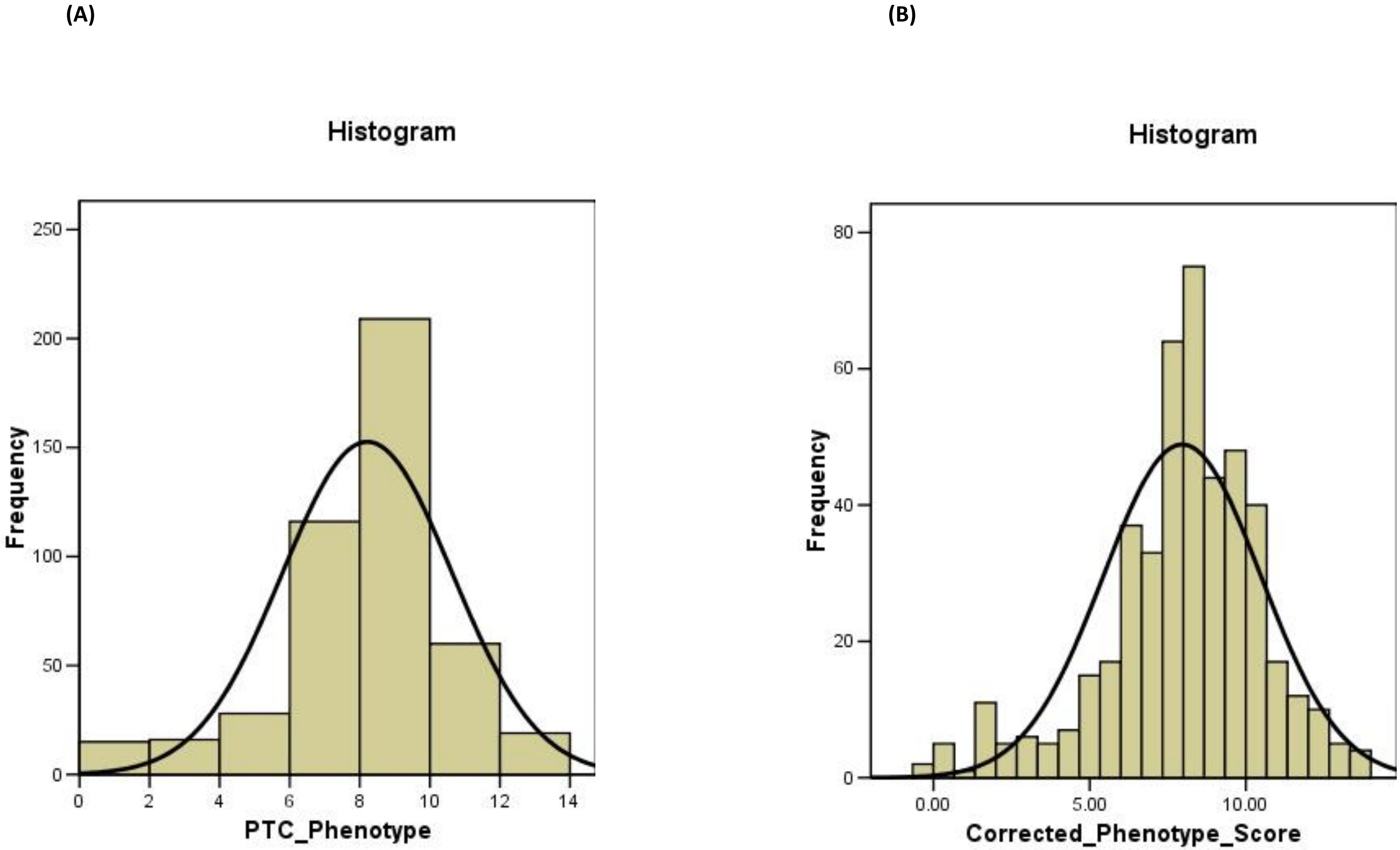
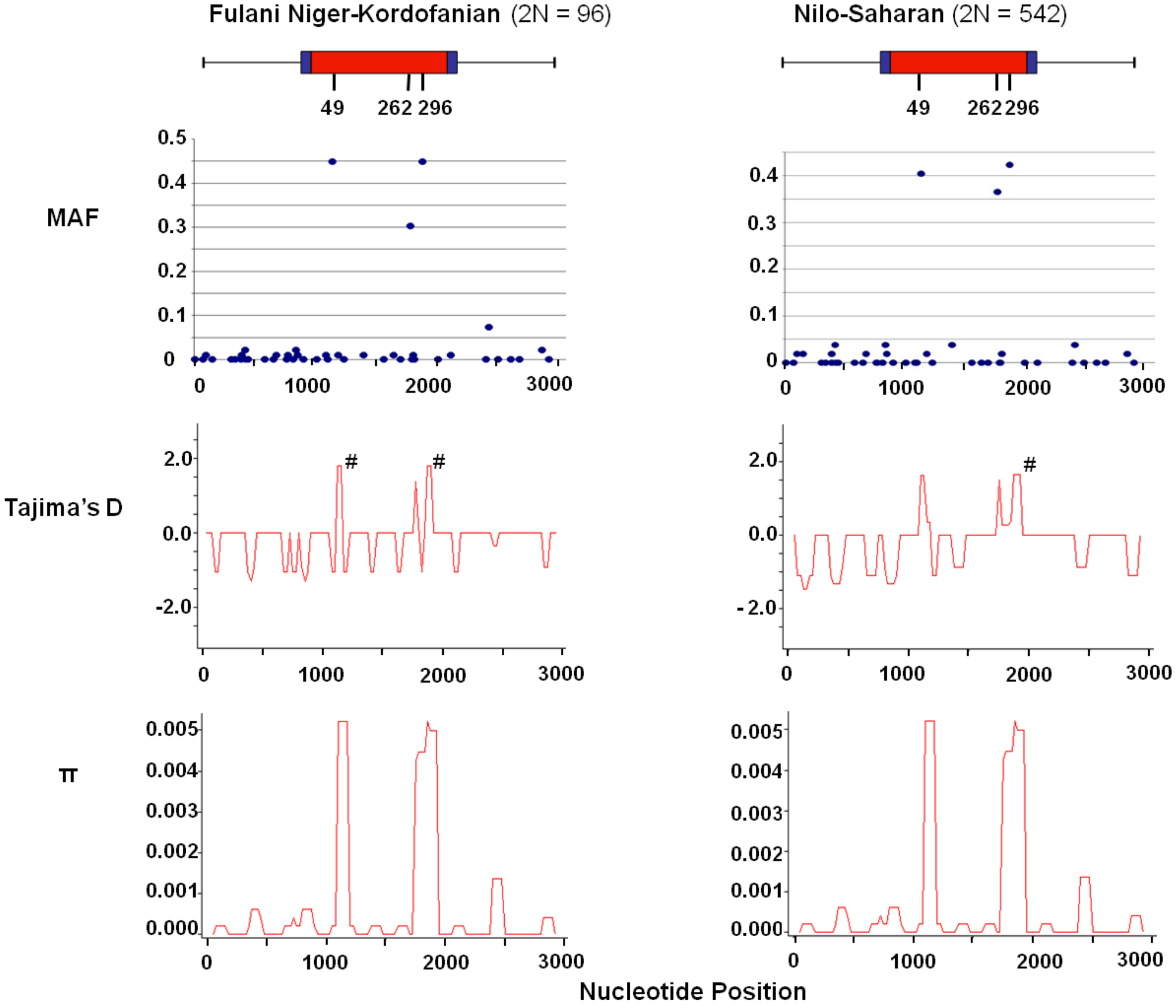
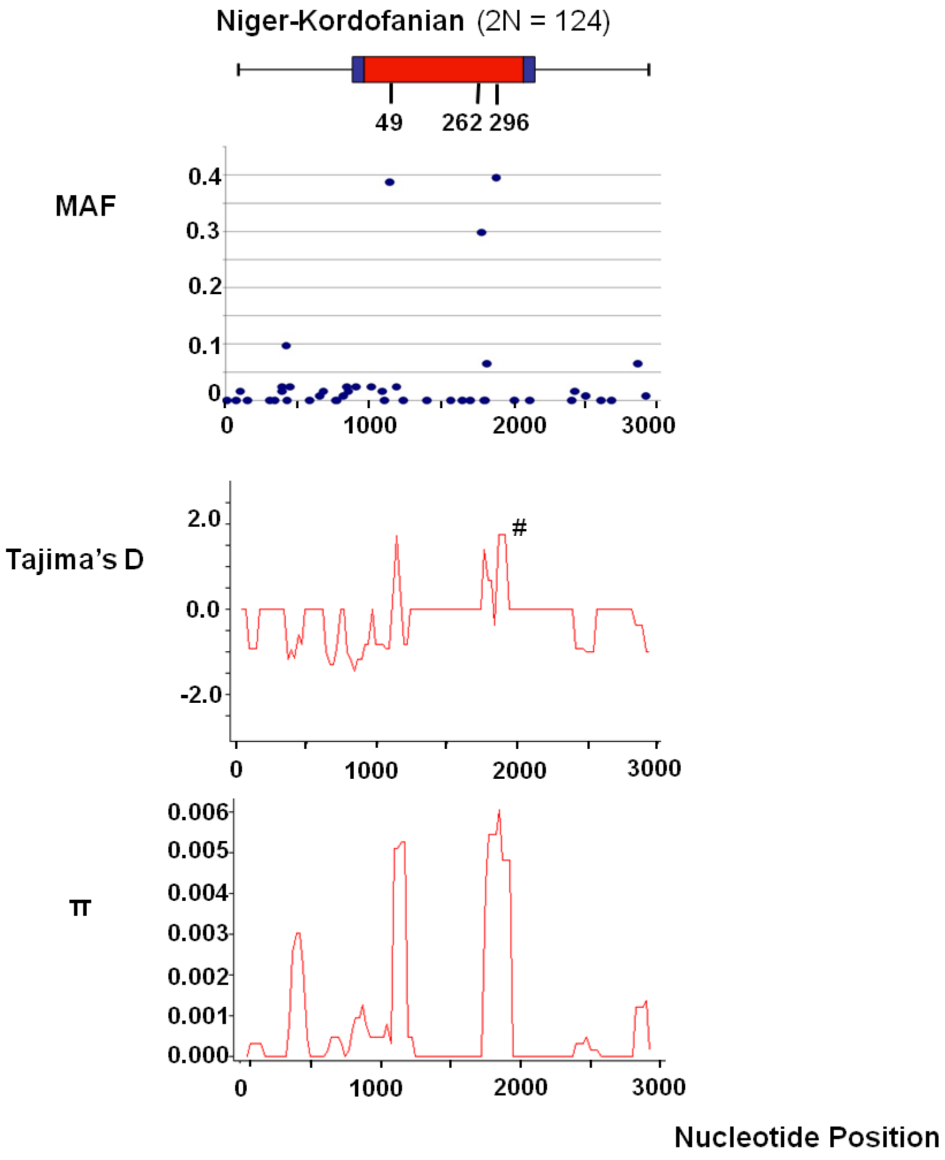


Figure S5. Distribution of PTC scores in 463 African individuals (A) raw PTC scores (B) PTC scores adjusted for age and sex. The solid line represents a normal curve drawn over the observed distribution of PTC score in Africans.

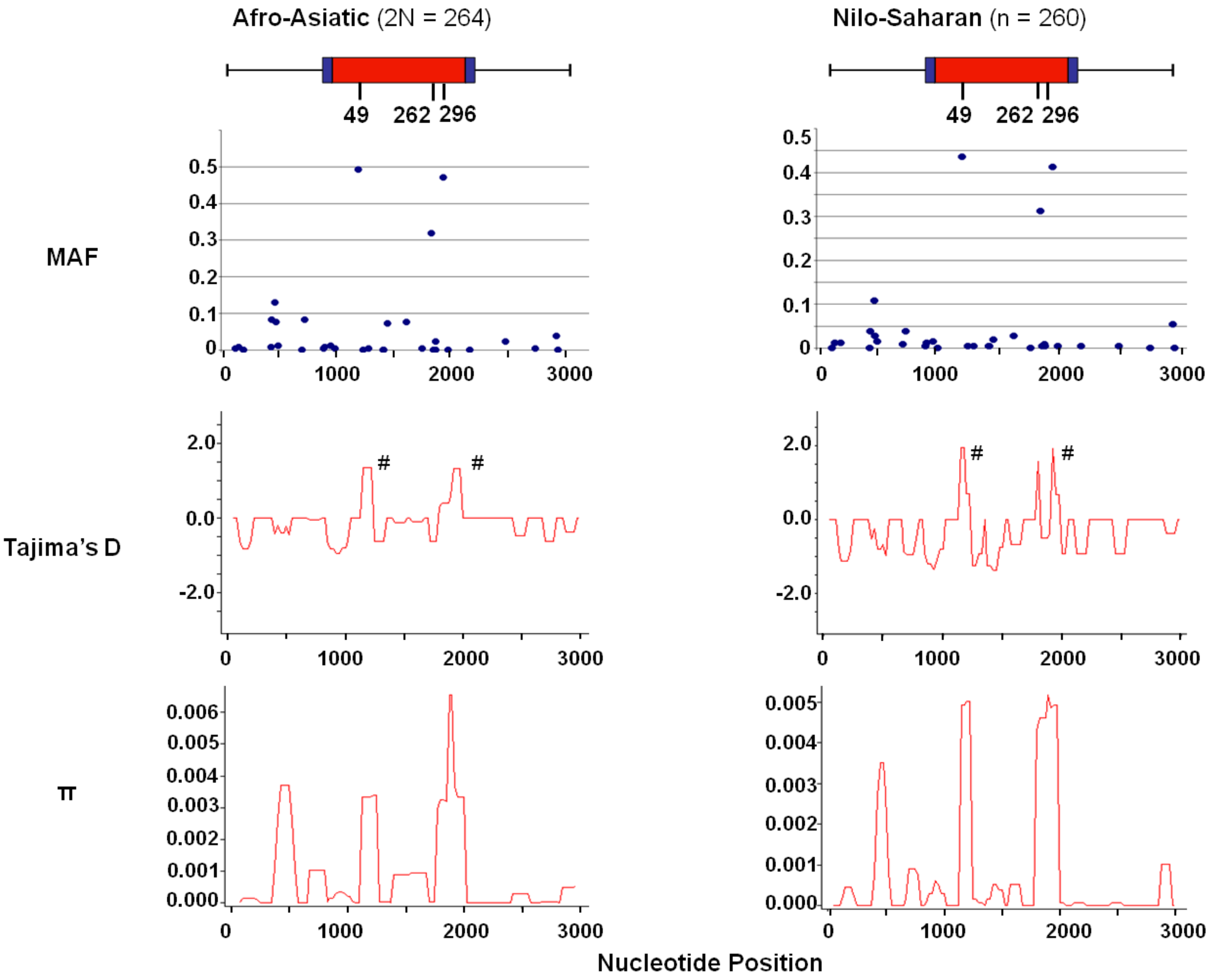
WEST CENTRAL AFRICA



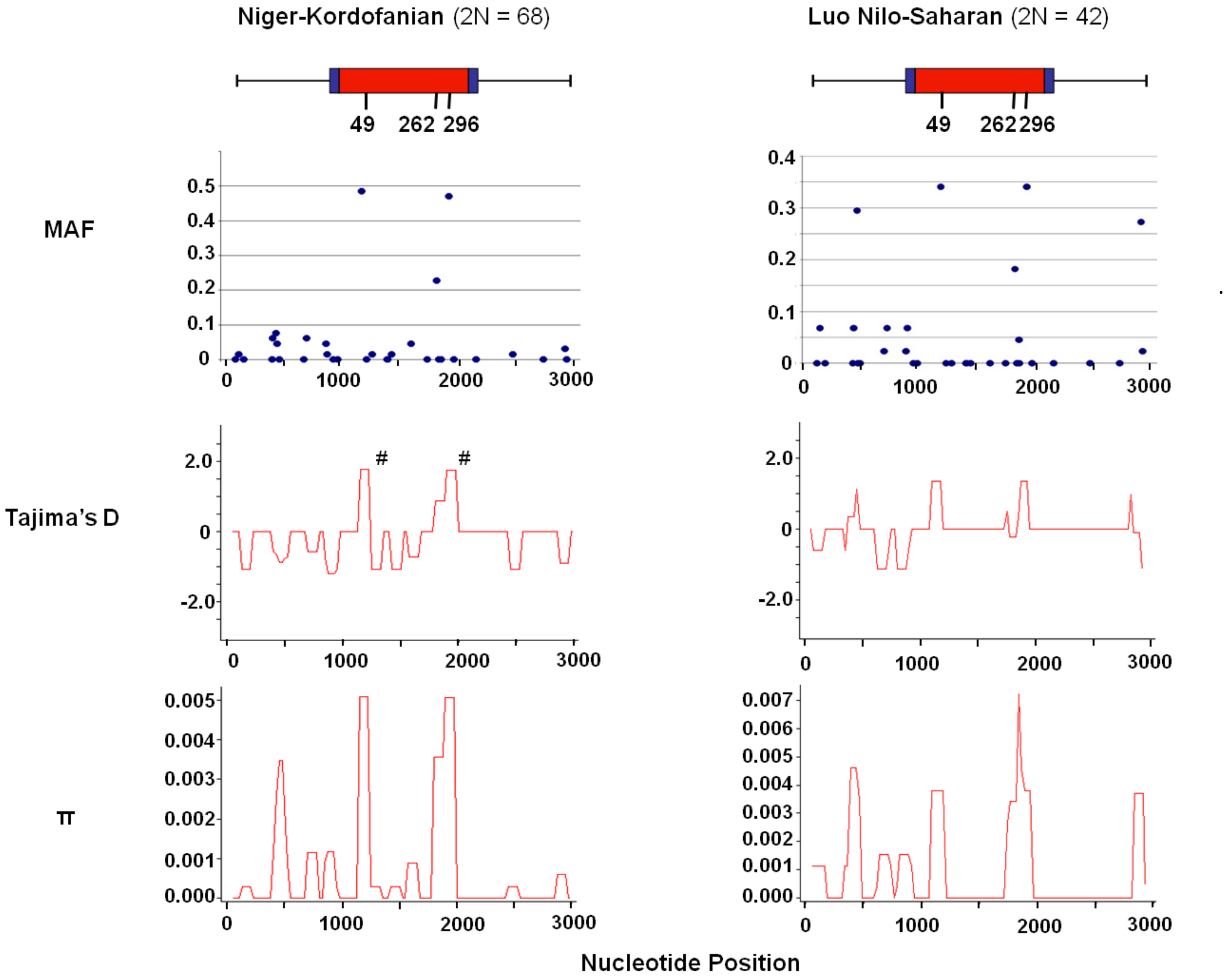
WEST CENTRAL AFRICA



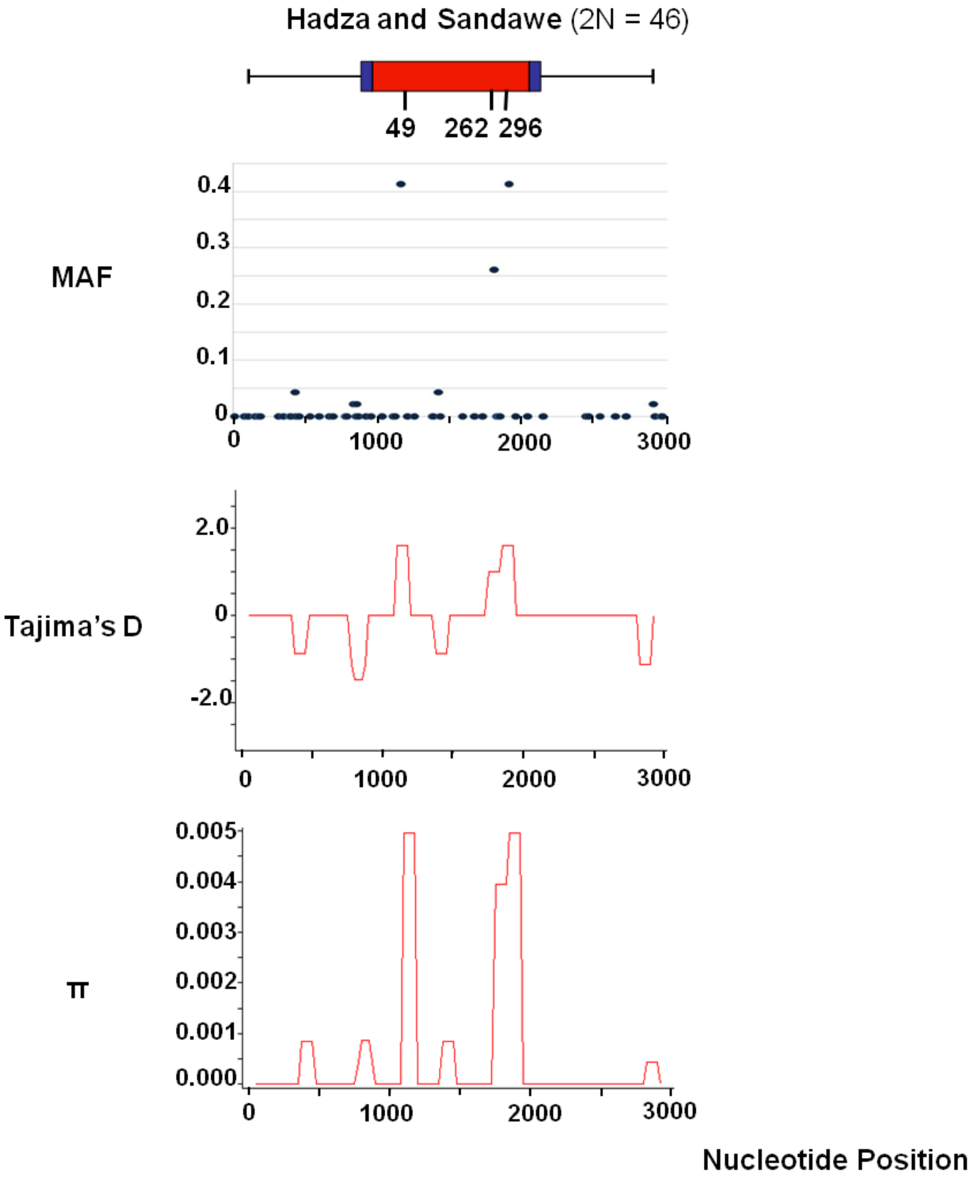
EAST AFRICA


Afro-Asiatic (2N = 264)
Nilo-Saharan (n = 260)
49
262
296
MAF
0.5
0.4
0.3
0.2
0.1
0
0
1000
2000
3000
Tajima's D
2.0
0.0
-2.0
#
#
π
0.006
0.005
0.004
0.003
0.002
0.001
0.000
Nucleotide Position

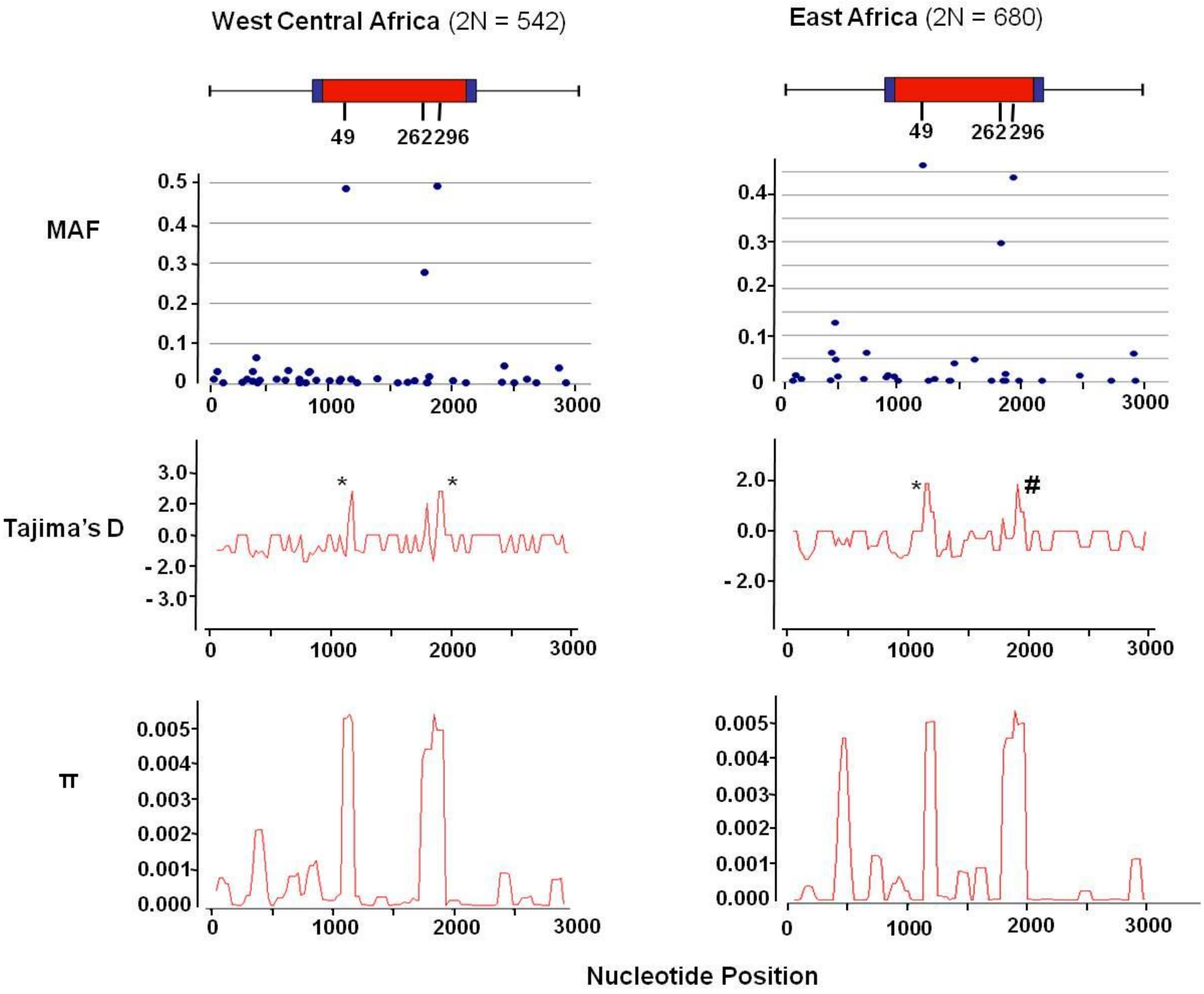
EAST AFRICA


Niger-Kordofanian (2N = 68)
Luo Nilo-Saharan (2N = 42)
49
262
296
49
262
296
MAF
0.5
0.4
0.3
0.2
0.1
0
0
1000
2000
3000
0.4
0.3
0.2
0.1
0
0
1000
2000
3000
Tajima's D
2.0
0
-2.0
#
#
0
1000
2000
3000
2.0
0
-2.0
0
1000
2000
3000
π
0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
0.007
0.006
0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
Nucleotide Position

EAST AFRICA


Hadza and Sandawe (2N = 46)
49
262
296
MAF
0.4
0.3
0.2
0.1
0
0
1000
2000
3000
Tajima's D
2.0
0
-2.0
0
1000
2000
3000
π
0.005
0.004
0.003
0.002
0.001
0.000
0
1000
2000
3000
Nucleotide Position

POOLED POPULATIONS



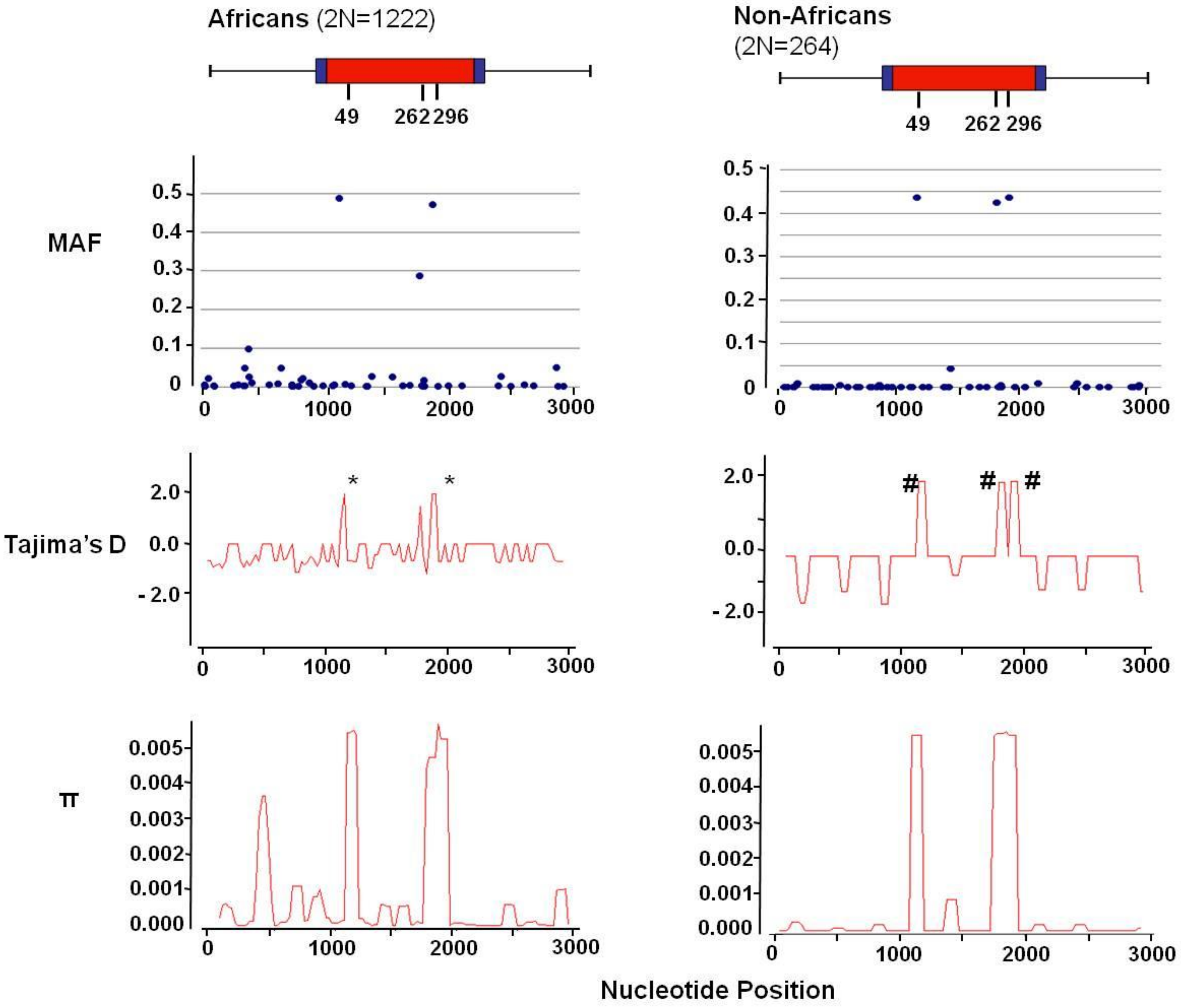

Africans (2N=1222)
49
262
296
Non-Africans (2N=264)
49
262
296
MAF
0.5
0.4
0.3
0.2
0.1
0
0
1000
2000
3000
Tajima's D
2.0
0.0
- 2.0
*
*
#
#
#
π
0.005
0.004
0.003
0.002
0.001
0.000
Nucleotide Position

Figure S6. Sliding window analysis of coding and flanking noncoding regions of *TAS2R38*. Tajima’s D and average pairwise differences between sequences (π) were calculated across the 2,975 bp region in diverse populations from West Central Africa (Cameroon), East Africa (Kenya and Tanzania), and non-Africans. The minor allele frequency (MAF) was also calculated for each identified SNP within the 2,975 bp region. The gene structure of *TA2R38* is located at the top of each figure. Statistically significant p-values ($P<0.05$) are indicated by *, while marginally significant p-values ($0.05<p<0.1$) are indicated by #.

Supplemental References

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263-265.

Goddard, K. A. B., P. J. Hopkins, J. M. Hall, and J. S. Witte. 2000. Linkage disequilibrium and allele-frequency distributions for 114 single-nucleotide polymorphisms in five populations. American Journal of Human Genetics 66:216-234.