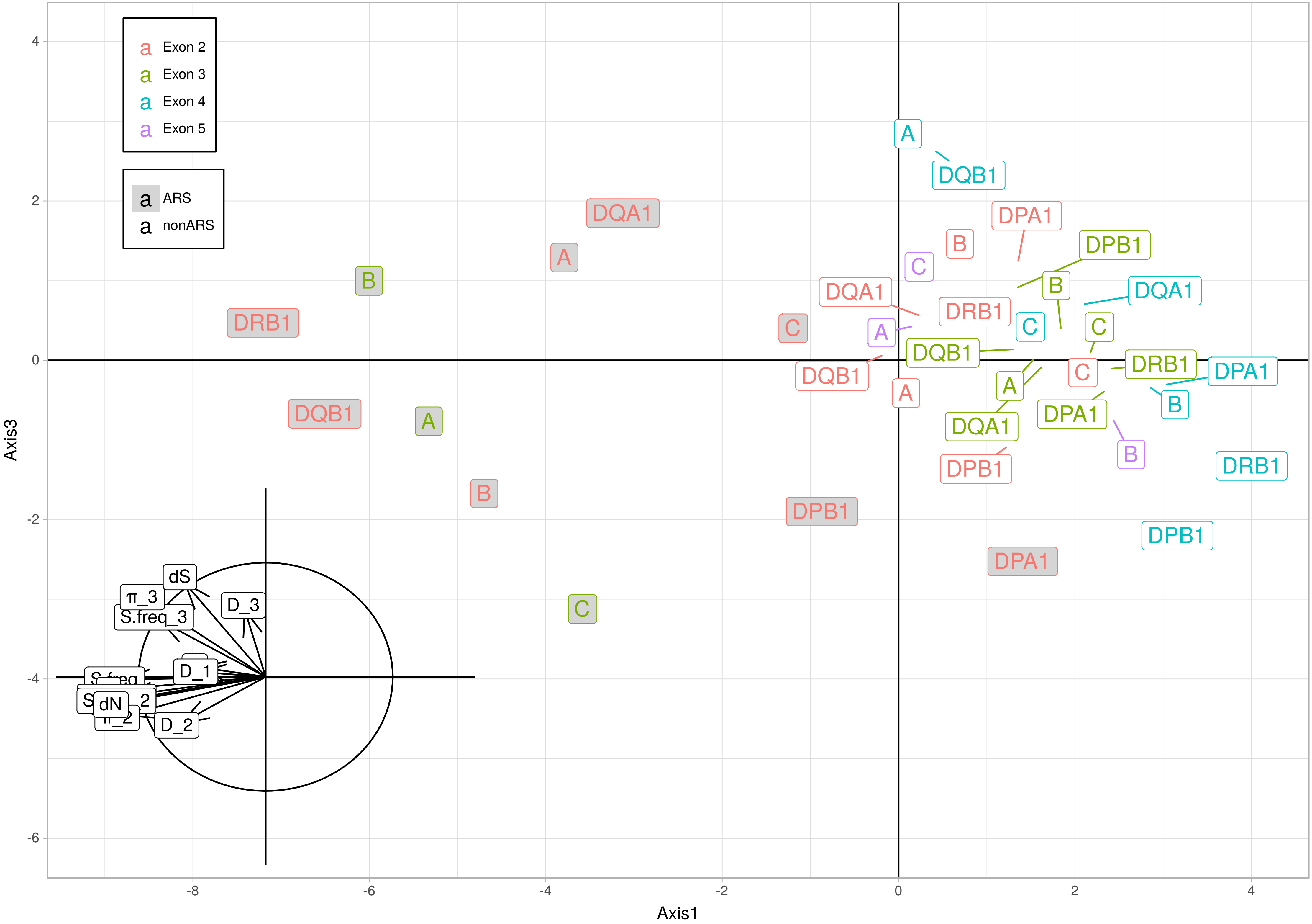
Supplementary Information S06

Principal Component Analysis, axes 1 and 3

Principal component analysis (axes 1 and 3, explaining respectively 60% and 11% of the total variance) based on Tajima's D, nucleotide diversity π, frequency of segregating sites S.freq, number of non-synonymous dN and synonymous dS nucleotides at exons 2, 3 and 4 of loci HLA-A, -B, - C, -DRB1, -DQA1, -DQB1, -DPA1, -DPB1 and exons 5 of loci HLA-A, -B, -C. Symbols D, D_1, D_2, D_3 represent Tajima's D estimated on the whole gene region and at the 1st, 2nd and 3rd nucleotide of each codon, respectively. Similarly, π , π_1, π_2 and π_3 as well as S.freq, S.freq_1, S.freq_2 and S.freq_3 represent the nucleotide diversity π and the frequency of segregating sites S estimated on the whole gene region and at nucleotide positions 1, 2 and 3 of each codon, respectively. Grey boxes correspond to ARS codons, white boxes to non-ARS codons. The inset graph at the bottom left represents the correlations between the projections of the variables (for each pair of variables, the correlation is measured by the cosine of the angle of the two variable vectors) on the plan of the PCA.


a Exon 2
a Exon 3
a Exon 4
a Exon 5
a ARS
a nonARS
A
DQB1
DQA1
DPA1
B
A
DPB1
C
B
DQA1
DQA1
B
DRB1
DRB1
C
C
C
A
DQB1
DRB1
C
DPA1
DQB1
A
A
DQB1
A
DPA1
B
DQA1
B
DPB1
DRB1
B
DPB1
DPB1
DPA1
C
dS
π_3
D_3
S.freq_3
D_1
S.freq
S
dN
_2
π_2
D_2
4
2
0
-2
-4
-6
-8
-6
-4
-2
0
2
4
Axis3
Axis1