

Additional methods description.

This section is part of the Supplementary Data File 1.

Detection of selection using PAML

We also estimated the ratio of pN/pS by gene in each population using the PAML package v4.10.6 (147). For this analysis, we derived 539,298 gene nucleotide alignments with the vcf2fasta package (<https://github.com/santiagosnchez/vcf2fasta>) from protein coding genes with at least 1 SNP in each population. Then, each alignment was translated into amino acids sequences with transeq from the EMBOSS package v6.6.0.0 (148), and their corresponding codon alignments were created with the pal2nal.pl program v14 (149). A total of 162 protein coding genes were removed from this analysis due to the presence of multiples indels affecting the positions of codons, resulting in a final dataset of 539,136 codon alignments. A ML phylogenetic tree was reconstructed for each codon alignment with FastTree v2.1 (150) and the GTR+GAMMA model. The one-ratio model (M0) was used to calculate the pN/pS ratio average for the whole gene over all branches in the phylogeny with the codeml program from the PAML package (147). The pN/pS ratio of a gene was the average pN/pS ratios calculated within the gene.

PAML detected 2.5-fold times more sites under selection than the improved K2P Li's method. However, ~37% of the pN/pS ratios detected with PAML exhibit standard deviations >10, suggesting either a high divergence among individuals within a population or an overestimation of the pN/pS ratio per gene. The latter is more probable because the PAML-based pN/pS ratio method ignores SNPs segregating within a population and transient SNPs within divergent populations (151–153). Based on these results, we further analyzed and discussed only the pN/pS ratios estimated with the improved K2P Li's method.