

The evolution and population genetics of the *ALDH2* locus: random genetic drift, selection, and low levels of recombination

[Hiroki Oota](#), [Andrew J. Pakstis](#), [Batsheva Bonne-Tamir](#), [David Goldman](#), [Elena Grigorenko](#), [Sylvester L. B. Kajuna](#), [Nganyirwa J. Karoma](#), [Selemani Kungulilo](#), [Ru-Band Lu](#), [Kunle Odunsi](#), [Friday Okonofua](#), [Olga V. Zhukova](#), [Judith R. Kidd](#) and [Kenneth K. Kidd](#)

Summary

The catalytic deficiency of human aldehyde dehydrogenase 2 (ALDH2) is caused by a nucleotide substitution (G1510A; Glu487Lys) in exon 12 of the *ALDH2* locus. This SNP, and four non-coding SNPs, including one in the promoter, span 40 kb of *ALDH2*; these and one downstream STRP have been tested in 37 worldwide populations. Only four major SNP-defined haplotypes account for almost all chromosomes in all populations. A fifth haplotype harbours the functional variant and is only found in East Asians. Though the SNPs showed virtually no historic recombination, LD values are quite variable because of varying haplotype frequencies, demonstrating that LD is a statistical abstraction and not a fundamental aspect of the genome, and is not a function solely of recombination. Among populations, different sets of tagging SNPs, sometimes not overlapping, can be required to identify the common haplotypes. Thus, solely because haplotype frequencies vary, there is no common minimum set of tagging SNPs globally applicable. The F_{st} values of the promoter region SNP and the functional SNP were about two S.D. above the mean for a reference distribution of 117 autosomal biallelic markers. These high F_{st} values may indicate selection has operated at these or very tightly linked sites.