Supplementary methods

Linkage disequilibrium analysis

Six-marker haplotypes (D11S1325, AFM254zd5, $[CA]_n$; D11S1781, AFMa115zc1, $[CA]_n$; D11S1343, AFM296yd9, $[CA]_n$; D11S4206, AFMa102xf9, $[CA]_n$; D11S1391, GATA4E01, $[GATA]_n$; and D11S1793, AFM312vc9, $[CA]_n$: average heterozygosity: 0.53, 0.34, 0.55, 0.72, 0.78, and 0.83, respectively) encompassing the c.622C>T mutation at the *ACAT1* locus were inferred either by phased genotypes or by use of the expectation–maximization algorithm as implemented in Haploview (v. 4.0; Broad Institute, Cambridge, MA) (Barrett et al, 2005). Linkage disequilibrium (LD) was calculated according to the method of Bengtsson and Thompson (1981). Physical and genetic map distances were obtained from NCBI Map Viewer (Annotation Release 107; https://www.ncbi.nlm.nih.gov/projects/mapview).

Age was first estimated by two moment methods, that of Risch et al. (1995) and one introduced by Reich and Goldstein (1999), both of which are based on the "genetic clock". Its equation, $\ln Q = -\theta g$, relates the time (in generations [g]) since the most recent common ancestor (MRCA) of mutant chromosomes, the frequency of recombination between the disease locus and marker (θ), and the probability that a marker's allele on a disease chromosome is the ancestral one (Q). An unbiased estimate of Q is the proportion of observed haplotypes that are ancestral. The genetic clock was set following the approach of Luria and Delbrück (1943), which takes into account the population growth rate (d). Accordingly, the estimated age was corrected by adding

$g_0 = -(1/d)\ln(\theta/d)$ to the g value (Labuda et al, 1996). The second moment method (Reich and Goldstein, 1999) generates a Markov transition matrix, which gives the probability that, in a single generation, any one haplotype will be transformed into any other one. Under this method, it was possible to correct the decay of LD over generations for the mutation rate (μ = 0.00056 and 0.0021 for dinucleotide and tetranucleotide repeats, respectively [Weber and Wong 1993])

at the marker loci. LD data were also used to estimate MRCA age by the Bayesian Markov chain–Monte Carlo method of Rannala and Reeve (2001) implemented in DMLE+ v. 2.3 software (www.dmle.org).

Throughout the calculations, we assumed that mating was random in the studied population, the generations were non-overlapping, there is no population substructure, the mutant allele is selectively neutral, and there is no further recurrent mutation at the disease locus; the first two assumptions are commonly accepted for this kind of genetic inference. The origin of the Kinh population remains controversial (Vu-Trieu et al, 1997), and the presence of subpopulations cannot be excluded. It is likely that Vietnamese people originated from one of the many Viet tribes living in southern China that migrated further south to the Indochina peninsula, where they met and mixed with a local Indonesian population of the Red River Delta around the second millennium BCE. Archaeological, anthropological, and genetic evidence is consistent with this hypothesis (Vu-Trieu et al, 1997; Bezacier, 1972). However, if such genetic heterogeneity exists, it would have little or no bearing on our analysis because its origin predates, or is contemporary with, the estimated age of the c.622C>T mutation. According to Kaplan et al. (1995) and Guo and Xiong (1997), in the absence of any evidence to the contrary, it is reasonable to assume that a mutation causing a rare autosomal recessive disorder is selectively neutral. Finally, the absence of further recurrent mutations at the same locus rests on the hypothesis that the *ACAT1* c.622C>T mutation in the Kinh population originates from a single founder, as suggested by the highly conserved seven-marker 11q22.3 haplotype of the mutation-bearing chromosomes.

References

Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview analysis and visualization of LD and haplotype maps. *Bioinformatics* 21: 263–265.

Bezacier L (1972) *Le Viet-Nam*. Paris: Picard.

Bengtsson BO, Thomson G (1981) Measuring the strength of associations between HLA antigens and diseases. *Tissue Antigens* 18: 356–363.

Guo SW, Xiong M (1997) Estimating the age of mutant disease alleles based on linkage disequilibrium. *Hum Hered* 47: 315–337.

Kaplan NL, Hill WG, Weir BS (1995) Likelihood methods for locating disease genes in nonequilibrium populations. *Am J Hum Genet* 56: 18–32.

Labuda M, Labuda D, Korab-Laskowska M et al (1996) Linkage disequilibrium analysis in young populations: pseudo-vitamin D-deficiency rickets and the founder effect in French Canadians. *Am J Hum Genet* 59: 633–643.

Luria SE, Delbrück M (1943) Mutations of bacteria from virus sensitivity to virus resistance. *Genetics* 28: 491–511.

Rannala B, Reeve JP (2001) High-resolution multipoint linkage-disequilibrium mapping in the context of a human genome sequence. *Am J Hum Genet* 69: 159–178.

Risch N, de Leon D, Ozelius L et al (1995) Genetic analysis of idiopathic torsion dystonia in Ashkenazi Jews and their recent descent from a small founder population. *Nat Genet* 9: 152–159.

Reich DE, Goldstein DB (1999) Estimating the age of mutations using variation at linked markers. In Goldstein GB, Schlötterer C, eds. *Microsatellites: Evolution and Applications*. Oxford: Oxford University Press, 129–138.

Vu-Trieu A, Djoulah S, Tran-Thi C et al (1997) HLA-DR and -DQB1 DNA polymorphisms in a Vietnamese Kinh population from Hanoi. *Eur J Immunogenet* 24: 345–356.

Weber JL, Wong C (1993) Mutation of human short tandem repeats. *Hum Mol Genet* 2: 1123–1128.