Age of the ΔF508 cystic fibrosis mutation

Sir — In a recent *Nature Genetics* paper, Morral *et al.* suggested that the ΔF508 mutation, responsible for 70% of the cases of cystic fibrosis (CF), is at least 2,600 generations old. This is far greater than the usually accepted figure of a few hundred generations[2].

Morral *et al.* constructed an evolutionary tree for a sample of ΔF508 chromosomes and inferred the number n of mutations per lineage in this tree for three microsatellite markers (IVS8CA, IVS17BTA, IVS17BCA) intragenic to the CF gene. They also inferred the mutation rate μ at these markers, and estimated the number of generations G since the ancestral ΔF508 mutation from the relation $n=G\mu$. We believe that their estimate of n is too large, and the μ value they used is too small: both have the effect of increasing G.

As Morral *et al.* point out, it is difficult to estimate very low mutation rates. In 3,000 meioses they observed no mutation events at any of the three markers and so gave an upper confidence limit for the mutation rate. If the rate is μ_i for the ith marker, the probability that none of the markers mutates in 3,000 meioses is approximately $e^{-3000(\mu_1+\mu_2+\mu_3)}$. Setting this to 0.05 gives an upper 95% confidence limit on $\Sigma_i \mu_i$ of 9.99×10^{-4}. There is no need for separate values for the μ_i's, although Morral et al. assumed equal rates and gave $\mu_i = 3.33 \times 10^{-4}$, $i=1,2,3$. In the G generations since the original ΔF508 mutation, the total expected number of changes over the three markers per lineage is $G\Sigma_i \mu_i$. We concur with the parsimony tree for the 1705 ΔF508 chromosomes given by Morral *et al.*, and we believe that they assumed 1,478 marker allele changes over this tree to arrive at their estimate of n of 1,478/1,705=0.867. This provides $G=0.867/(9.99\times10^{-4})=868$. Although Morral *et al.* used the total number of mutational changes over three loci, they used a per-locus mutation rate, $\mu_i=3.33\times10^{-4}$, to estimate $G=2,627$. This is why we consider their mutation rate μ to be too low.

The 1,705 disease chromosomes[1] contained 47 distinct haplotypes. One of the 47 is 23–31–13, the haplotype that is inferred to be the one on which the ΔF508 mutation occurred, and this is present 711 times. The other 994 chromosomes have 1, 2 or 3 mutational changes from 23–31–13. If the number of changes is multiplied by the number of chromosomes having that number, and these products added together, the figure of 1,478 is obtained. This total represents 898, 571 and 9 changes at the three markers, and this seems to be inconsistent with an assumption of equal rates of change at the three. Now there are 402 copies of the 17–31–13 haplotype, for example, and this is one mutation away from the 23-31-13 haplotype. Rather than regarding these as contributing 402 × 1 to 1,478, we consider it likely that many of the 402 chromosomes descend from a single 17–31–13 haplotype subsequent to the mutation from 23–31–13, and quite unlikely that there were 402 separate mutations of this type. The coalesence of some of the 402 chromosomes is likely to outweigh the effects of multiple mutations resulting in a net single change from the ancestral haplotype. For this reason we consider the 1,478, and hence the n, obtained by Morral *et al.* to be too high. The minimum number of changes needed to explain the parsimony tree is 46, consisting of 14, 26 and 6 changes at the three loci. Of course the actual number of changes is unknown.

The fact that the marker alleles on ΔF508 haplotypes are rare in the general European population persuaded Morral *et al.* that the disease arose in a population ancestral to the one that later spread throughout Europe. We note that the disease mutation may simply have arisen on a rare marker haplotype.

Morral *et al.* do not discuss the implications of an age of 2,600 generations for the observed linkage disequilibrium of CF with markers XV2C and KM19[2]. As we detail elsewhere[3], this disequilibrium is probably due to a lack of recombination, between ΔF508 and the markers, preventing the breakup of ancestral haplotypes. For a recombination fraction c between a disease and a marker locus, and G generations since the origin of the disease, the probability that a disease chromosome has the ancestral marker allele is approximately e^{-cG}. Kerem *et al.*[4] reported distances from ΔF508 to XV2C and KM19 of 285 kb and 220 kb, respectively, which convert to recombination values of 0.00285 and 0.00220 assuming 1 cM ~ 1,000 kb. However, it seems likely that recombination is restricted in this region[5], so we halved these recombination values to 0.00143 and 0.00110 (Farrall *et al.*[5] suggested, without real evidence, 0.001 as the value for KM19, whereas Beaudet *et al.*[6] suggested 0.006). With these values, the probability that a disease chromosome has the ancestral marker alleles 2,627 generations after the disease mutation is approximately 0.02 and 0.06 for markers XV2C and KM19. In other words, it would be unusual to find a disease chromosome carrying the ancestral marker alleles, whereas Kerem *et al.* found almost all of the 77 sampled ΔF508 chromosomes carried the same (presumed ancestral) marker alleles at these loci. Taking G to be 868, however, raises the probabilities to 0.29 and 0.38.

N. L. Kaplan
Statistics and Biomathematics Branch, National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, USA

P. O. Lewis
B. S. Weir
Program in Statistical Genetics, Department of Statistics, North Carolina State University, Raleigh, North Carolina 27695-8203, USA

In Reply — The use of variation at microsatellite markers to trace the evolution of a disease mutation has no precedent in molecular population genetics. Thus, we are dealing with a new situation which may be difficult to understand and which will benefit from further experience with other loci and diseases, as well as new theoretical approaches.

In our analysis of more than 3,000