Letter to the Editor

A Human Population Bottleneck Can Account for the Discordance Between Patterns of Mitochondrial Versus Nuclear DNA Variation

Justin C. Fay* and Chung-I Wu*†

*Committee on Genetics and †Department of Ecology and Evolution, University of Chicago

Whether or not humans have experienced a reduction in population size in the recent past is a controversial issue germane to the origin and colonization of our own species (Stringer and Andrews 1988). A change in population size can result in deviations from the neutral patterns of nucleotide variation expected at equilibrium. Using the frequency distribution of mutations segregating in extant populations, the magnitude of a deviation can be measured by Tajima's (1989a) D statistic or by a number of alternative measures (Fu and Li 1993; Fu 1996). In a population of constant size, variation at a neutrally evolving locus is expected to have a D value of approximately zero. Following a reduction in population size, rare frequency mutations are lost more readily than are common mutations (Nei, Maruyama, and Chakraborty 1975), and transient positive D values are expected (Tajima 1989b). Following an increase in population size, there is a temporary excess of new mutations segregating at rare frequencies, and negative D values are expected. The sign of Tajima's D subsequent to a population bottleneck can be positive, negative, or zero depending on the length of time since the bottleneck and the severity of the bottleneck. If a bottleneck is so severe that all variation is eliminated or lasts so long that the population reaches a new equilibrium, Tajima's D follows the pattern produced by an expansion in population size. However, following an incomplete bottleneck, Tajima's D is transiently positive before becoming negative and eventually approaching its equilibrium (Tajima 1989b).

In humans, mitochondrial variation is characterized by an excess of rare frequency mutations and a negative D value, which has been interpreted as the result of a recent expansion in population size (Merriwether et al. 1991; Rogers and Harpending 1992). In contrast, most nuclear loci are characterized by the opposite pattern, an excess of common mutations and positive D values, which can result from a recent reduction in population size (Hey 1997; Harding et al. 1997; Clark et al. 1998; Zietkiewicz et al. 1998). The conflicting profiles of mitochondrial and nuclear variation have led to the suggestion that these patterns cannot be simultaneously accounted for by human population history, which must be shared by both genomes (Hey 1997).

Abbreviation: mtDNA, mitochondrial DNA.

Key words: population bottleneck, human evolution, *Homo sapiens.*

Address for correspondence and reprints: Justin Fay, Committee on Genetics, 1101 East 57th Street, Chicago, Illinois 60637. E-mail: jfay@midway.uchicago.edu.

Since the mitochondrial genome has an effective population size one quarter that of an average nuclear locus, its patterns of variation may be quite different following a change in population size. While current explanations must invoke complex demographic histories or selection, we have found using computer simulations that the apparently incongruent patterns of mitochondrial and nuclear variation are in fact quite compatible with a recent human population bottleneck.

Patterns and levels of nucleotide variation are commonly summarized by $\hat{\pi}$ and $\hat{\theta}$, two estimators of the population parameter θ, equal to four times the product of the effective population size and the mutation rate. In a sample of n genes, $\hat{\pi}$ is the average number of pairwise differences, and $\hat{\theta} = S/a_n$, where S is the number of segregating sites and a_n is the sum of $1 + 1/2 + 1/3 + \ldots + 1/(n-1)$ (Watterson 1975). Tajima's D measures the difference between $\hat{\pi}$ and $\hat{\theta}$ divided by the variance of the difference:

$$D = \frac{\hat{\pi} - \hat{\theta}}{\sqrt{e_1 S + e_2 S(S-1)}}, \qquad (1)$$

where e_1 and e_2 are constants derived from the sample size (Tajima 1989a). In a population of constant size, the expected difference between $\hat{\pi}$ and $\hat{\theta}$ and the corresponding D value is just below zero, assuming neutral evolution (Tajima 1989a).

We used a coalescence algorithm (following that of Simonsen, Churchill, and Aquadro 1995) and Tajima's (1989b) analytical equations to generate the expected reduction in $\hat{\pi}$ and $\hat{\theta}$ and the corresponding D values for human mitochondrial and nuclear parameters following a population bottleneck. To explore the effects of different parameters, we used Tajima's (1989b) equations for the expectation of $\hat{\pi}$ and $\hat{\theta}$ following a change in population size. However, these equations do not provide the covariance between $\hat{\pi}$ and $\hat{\theta}$, and thus result in biased estimates of the variance in D. Therefore, all of the presented results were generated using 5,000 iterations of the coalescence algorithm for a given set of bottleneck and population parameters (with sample size $n = 50$). We assume the mitochondrial genome has one quarter the effective population size of the nuclear genome, and each experiences the same percent reduction in population size during a bottleneck.

A population bottleneck affects the subsequent patterns of DNA variation. In a simple stepwise bottleneck (fig. 1A), a population at equilibrium of size N_0 is reduced to N_1 for T generations. The severity of the bottleneck is determined by the reduction in population size, N_0/N_1, and the duration of the bottleneck, T/N_0 (measured in units of N_0 generations) (Tajima 1989b). Over a range of bottleneck severities, nearly the same
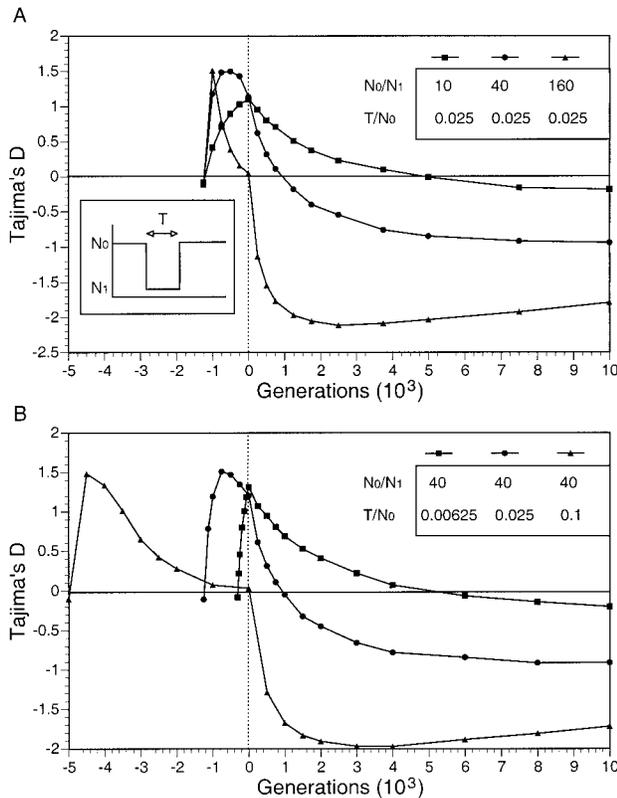
Fig. 1.—Similar *D* values result from bottlenecks (*A*) of different reductions in population size or (*B*) different bottleneck lengths. Each point is the mean of 5,000 iterations of a coalescence algorithm, where $N_0 = 2.5 \times 10^4$ and $\theta = 20$, and the end of the bottleneck (dotted line) is at generation zero (see text for details).

Fig. 2.—The mitochondrial (open circles) and nuclear (closed circles) *D* values during and subsequent to a population bottleneck. The reduction in population size is $N_0/N_1 = 20$ and the duration is for 1,500 generations, as indicated by the vertical dotted line. The nuclear and mitochondrial parameter values are, respectively, $N_0 = 3 \times 10^4$ and 7.5×10^3, $T/N_0 = 0.025$ and 0.1, and $\theta = 30$ and 150 (mtDNA is assumed to have one quarter the population size and 20 times the mutation rate of nuclear DNA). Horizontal dotted lines indicate the observed mitochondrial *D* and the range of three nuclear *D* values (see text).

D values are produced for a bottleneck that is four times as long and one quarter the N_0/N_1 ratio (comparing the corresponding curves in fig. 1*A* and *B*). Hence, the severity of a bottleneck is approximately proportional to the product of T/N_0 and N_0/N_1, or T/N_1, when comparing bottlenecks of intermediate severity ($0.25 < T/N_1 < 4$).

Passing through the same bottleneck, the mitochondrial genome experiences a greater reduction in levels of variation and subsequently lower *D* values, but recovers more quickly than the nuclear genome. The difference in severity is due only to the mitochondrial genome's smaller population size, which determines the effective duration, T/N_0, of the bottleneck (note that the ratio N_0/N_1 is the same for both genomes). The mutation rate, which is 10–30 times higher in mitochondrial DNA (Horai et al. 1995; Takahata and Satta 1997), does not contribute to the difference between patterns of mitochondrial and nuclear variation, since, like the sequence length or sample size, it only influences our ability to estimate the difference between $\hat{\pi}$ and $\hat{\theta}$. Using reasonable parameters for a human population bottleneck, simulated *D* values (fig. 2) are similar to the observed mitochondrial *D* value, -2.13 (Merriwether et al. 1991), and the range of *D* values obtained from three nuclear genes: β-globin, 1.06 (Harding et al. 1997); lipoprotein lipase, 0.91 (Clark et al. 1998); and dystrophin, 0.96, which is on the X chromosome (Zietkiewicz et al. 1998).
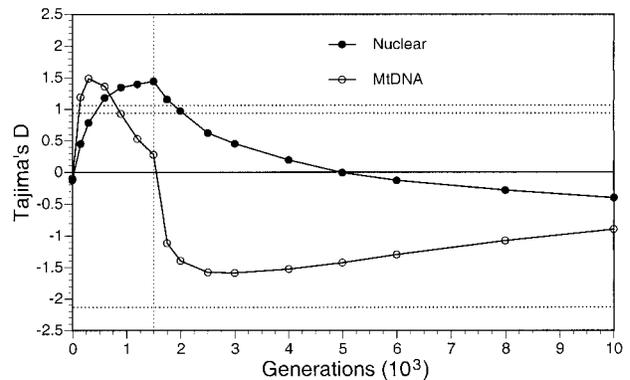
Although the observed and simulated *D* values are not statistically comparable due to sampling differences, the standard deviations of *D* calculated from the simulation were 0.87 at generation zero (equilibrium) and 1.39 and 1.05 at generation 3,000 for the nuclear and mitochondrial loci, respectively.

The difference between patterns of mitochondrial and nuclear variation following a bottleneck depends, of course, on the severity of the bottleneck. Following a severe bottleneck, both mitochondrial and nuclear *D* values tend to be more negative, and following a mild bottleneck, both tend to be closer to zero. Since numerous combinations of N_0, N_1, and T exist for any given severity, a recent population bottleneck provides a simple explanation for simultaneous negative mitochondrial and positive nuclear *D* values.

In conclusion, the opposite skews in the frequency distribution of mitochondrial and nuclear variation found in extant human populations are not necessarily incompatible with a common history shared by the two genomes, contrary to previous claims (e.g., Hey 1997). An incomplete bottleneck can produce *D* values which range from positive to negative, depending on values of N_0, N_1, and T that are not unrealistic by our current understanding of human history. In contrast, a population expansion is expected to produce negative *D* values for both mitochondrial and nuclear loci. Selection could certainly increase or decrease any *D* value, but a specific model of selection would be needed to explain not only the contrast between the mitochondrial and nuclear genomes, but also the generally positive *D* values of the several nuclear genes extensively studied so far (Harding et al. 1997; Clark et al. 1998; Zietkiewicz et al. 1998).

Genome-wide studies are particularly useful for inferring a population's history since any observed pattern cannot be explained by locus-specific effects. In a study of 60 microsatellite loci, patterns of variation were also

found to be compatible with a recent human population bottleneck (Kimmel, Chakraborty, and King 1998). In addition, the strongest signature of a population bottleneck was found in Asian populations, followed by Caucasian and African populations. Interestingly, a comparison of mitochondrial and nuclear *D* values across populations shows a similar pattern: non-African populations have lower mitochondrial *D* values and higher nuclear *D* values when compared to African populations (Merriwether et al. 1991; Harding et al. 1997; Clark et al. 1998; Zietkiewicz et al. 1998), and this pattern is most striking in the comparison of African and Asian populations. While in reality the history of human populations must be much more complex, with geographical differentiation, migration, and demographics specific to certain populations (Templeton 1997), a population bottleneck is at least a parsimonious explanation for the seemingly incongruent observations of mitochondrial and nuclear variation.

Acknowledgments

LITERATURE CITED

CLARK, A. G., K. M. WEISS, D. A. NICKERSON et al. (11 co-authors). 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am. J. Hum. Genet. **63**:595–612.

FU, Y. X. 1996. New statistical tests of neutrality for DNA samples from a population. Genetics **143**:557–570.

FU, Y. X., and W. H. LI. 1993. Statistical tests of neutrality of mutations. Genetics **133**:693–709.

HARDING, R. M., S. M. FULLERTON, R. C. GRIFFITHS, J. BOND, M. J. COX, J. A. SCHNEIDER, D. S. MOULIN, and J. B. CLEGG. 1997. Archaic African and Asian lineages in the genetic ancestry of modern humans. Am. J. Hum. Genet. **60**:772–789.

HEY, J. 1997. Mitochondrial and nuclear genes present conflicting portraits of human origins. Mol. Biol. Evol. **14**:166–172.

HORAI, S., K. HAYASAKA, R. KONDO, K. TSUGANE, and N. TAKAHATA. 1995. Recent African origin of modern humans revealed by complete sequences of hominoid mitochondrial DNAs. PNAS **92**:532–536.

KIMMEL, M., R. CHAKRABORTY, J. P. KING, M. BAMSHAD, W. S. WATKINS, and L. B JORDE. 1998. Signatures of population expansion in microsatellite repeat data. Genetics **148**:1921–1930.

MERRIWETHER, D. A., A. G. CLARK, S. W. BALLINGER, T. G. SCHURR, H. SOODYALL, T. JENKINS, S. T. SHERRY, and D. C. WALLACE. 1991. The structure of human mitochondrial DNA variation. J. Mol. Evol. **33**:543–555.

NEI, M., T. MARUYAMA, and R. CHAKRABORTY. 1975. The bottleneck effect and genetic variability in populations. Evolution **29**:1–10.

ROGERS, A. R., and H. HARPENDING. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. **9**:552–569.

SIMONSEN, K. L., G. A. CHURCHILL, and C. F. AQUADRO. 1995. Properties of statistical tests of neutrality for DNA polymorphism data. Genetics **141**:413–429.

STRINGER, C. B., and P. ANDREWS. 1988. Genetic and fossil evidence for the origin of modern humans. Science **239**:1263–1268.

TAJIMA, F. 1989*a*. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585–595.

———. 1989*b*. The effect of change in population size on DNA polymorphism. Genetics **123**:597–601.

TAKAHATA, N., and Y. SATTA. 1997. Evolution of the primate lineage leading to modern humans: phylogenetic and demographic inferences from DNA sequences. PNAS **94**:4811–4815.

TEMPLETON, A. R. 1997. Out of Africa? What do genes tell us? Curr. Opin. Genet. Dev. **7**:841–847.

WATTERSON, G. A. 1975. On the number of segregating sites in genetic models without recombination. Theor. Popul. Biol. **7**:256–276.

ZIETKIEWICZ, E., V. YOTOVA, M. JARNIK et al. (11 co-authors). 1998. Genetic structure of the ancestral population of modern humans. J. Mol. Evol. **47**:146–155.