Sequence Variation in ZFX Introns in Human Populations

Wei Huang, Yun-Xin Fu,* Benny H.-J. Chang,* Xun Gu,* Lynn B. Jorde,† and Wen-Hsiung Li**

*Human Genetics Center, University of Texas at Houston; and †Department of Human Genetics, University of Utah

DNA variation in human populations was studied by examining the last intron of the ZFX gene (about 1,151 bp) with a worldwide sample of 29 individuals. Only one polymorphic site was found, which is located in an Alu sequence. This polymorphism is present at an intermediate frequency in all populations studied, and could be a shared polymorphism or due to migration among populations in Asia, Europe, and Africa. The nucleotide diversity is 0.04%, supporting the view that the level of nucleotide variation in nuclear DNA is very low in humans. From the sequence data, the age (T) of the most recent common ancestor of the sampled sequences is estimated: the mode of T is about 306,000 years, and the 95% confidence interval of T is 162,000–952,000 years. This mode estimate is considerably older than the estimates from Y-linked sequences.

Introduction

Earlier genetic studies of human populations were based on blood group types and protein polymorphisms detected by electrophoresis (e.g., Nei and Roychoudhury 1982; Cavalli-Sforza, Menozzi, and Mountain 1988). With the advances in DNA technology, a large-scale investigation of DNA polymorphism is now feasible. Compared to the former types of data, DNA sequence data are more informative for studying the evolutionary history of a population, including demographic history (e.g., Rogers and Harpending 1992; Donnelly et al. 1996; Fu and Li 1996; Ruvolo 1996; Brookfield 1997). Since mitochondrial DNA (mtDNA) is maternally inherited with no recombination, and since it has a high mutation rate, it has been extensively used in the study of human population genetics and evolution. Cann, Stoneking, and Wilson (1987) and Vigilant et al. (1989) studied the variation pattern in mtDNA in worldwide populations and concluded that Africa was the probable ancestral source of all human mtDNA genotypes and that the age T of the most recent common ancestor (MRCA) of human females was in the range of 140,000–290,000 years. In spite of some criticisms of the analysis (e.g., Maddison, Rovolo, and Swofford 1992; Templeton 1992; Thorne and Wolpoff 1992), the mtDNA data tend to support the African origin of modern humans, because in the majority of alternative trees, the deepest nodes lead to African sequences (see the review by Li and Sadler 1992). However, mtDNA is effectively only one locus and the estimated T represents only the maternal lineages. To increase our knowledge on human genetic variation and evolution, it is important to examine nuclear DNA variation.

Recently, several studies of nuclear DNA variation have been conducted on the nonrecombining genomic regions of the Y chromosome. Dorit, Akashi, and Gilbert (1995) surveyed the last intron of the ZFY gene (729 bp) in a worldwide sample of 38 individuals and found no variation, in contrast to the high sequence variation in mtDNA. Later, Hammer (1995) examined the sequence variation of a 2.6-kb fragment in a nonrecombining region of the Y chromosome in 16 humans and found three polymorphic nucleotide sites and one polymorphic Alu insertion. The ages (T) of the MRCA of the sampled sequences estimated by Dorit, Akashi, and Gilbert (1995) and Hammer (1995) were consistent with the estimate from mitochondrial DNA and were taken as evidence to support the Out of Africa model for the origin of modern humans. However, their methods are not accurate. Fu and Li (1996) and Fu (1996) developed a more rigorous method for estimating T and obtained lower estimates of the mean of T. Tishkoff et al. (1996) studied the global pattern of linkage disequilibrium at the CD4 locus and suggested a common, recent African origin of all non-African human populations; see also Jorde et al. (1997). However, there are data that are contradictory to this model (Xiong et al. 1991; Fullerton et al. 1994; Jorde et al. 1995). In particular, Harding et al.'s (1997) recent data from the beta globin region cannot be easily reconciled with a unidirectional migration out of Africa 100,000 years ago and with the assumption of total replacement of archaic populations in Asia.

As data on human genetic variation of nuclear DNA are still limited, to have a good estimate of human DNA variation and to resolve the debate about human evolution, more sequence data are needed. Since most data on the human origin are from mtDNA or Y-linked sequences, it is interesting to obtain data from autosomal or X-linked genomic regions. In this paper, we report sequence data from the last intron of the ZFX gene in a sample of 29 human individuals from different parts of the world.

Materials and Methods

Sample Collection

There are two factors that should be considered in sample collection. First, the sample size should be large enough that the gene tree (genealogy) of the sample will have a high probability of including the root (i.e., the most ancestral node) of the current population. For a random-mating population, the probability that a random sample of n sequences will include the root of the gene tree for the entire population is $p = 1 - 2/(n +$

Key words: sequence variation, ZFX intron, Alu, origin of modern humans.

Address for correspondence and reprints: Wen-Hsiung Li, Human Genetics Center, University of Texas, P.O. Box 20334, Houston, Texas 77225. E-mail: li@hgc.sph.uth.tmc.edu.

Mol. Biol. Evol. 15(2):138–142. 1998
© 1998 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

1) (Saunders, Travaré, and Watterson 1984). Thus, p is about 0.90–0.94 for a sample size of 20–30 sequences. Second, the DNA samples should cover at least the three major continental populations (i.e., Asians, Africans, and Europeans).

We collected a sample of 29 males (the number of individuals is given in parentheses): Europeans or European descendants: German (1), Russians (2), British (1), Canadian (1), and white Americans (3); Asians: Chinese (2), Japanese (4), and Taiwanese (3); and Africans: Pygmies (5) and Nigerians (7). The Nigerian DNA samples were kindly provided by Dr. R. Deka of the University of Pittsburg. The Pygmy samples and two of the Japanese samples were obtained from the cell lines of the Corriel Institute; the other two Japanese samples were obtained from two Japanese Americans in Utah (their Japanese ancestry has been verified). The other samples were extracted from peripheral blood using the DNA extraction procedure of Ellsworth, Rittenhouse, and Honeycutt (1993). Genomic DNA of three common chimpanzees (*Pan troglodytes*) were a gift from Dr. R. E. Ferrell (University of Pittsburgh). All samples were from males, so only one X-linked sequence was amplified by PCR.

PCR Amplification and Direct Sequencing

Two PCR primers (OLW 97, OLW 98) designed by Shimmin, Chang, and Li (1993) for the last intron of the human ZFX gene were synthesized. Human DNA samples were successfully amplified by PCR with these two primers in an Omnigene thermal cycler (Hybaid). Since this set of primers did not amplify chimpanzee DNA, another two X-specific primers (OLW 99: CTCGGCAGACTGGCTAAACAA; OLW 100: TTCGGGATGGTTTTTCATGTGC) were designed according to the human ZFX coding sequences (Schneider-Gädicke et al. 1989). Primer OLW 99 was from nucleotides 1168–1188, while OLW 100 was from nucleotides 1347–1326. The reaction mixture and PCR conditions were similar to those used by Shimmin, Chang, and Li (1993), except that the denaturation time was reduced to 30 s, and annealing temperature for chimpanzees was 59°C instead of 65°C. A negative control was used to check for contamination. The PCR products were purified by Wizards PCR prep DNA purification system (Promega).

Fourteen sequencing primers (Shimmin, Chang, and Li 1993) were used to sequence the last intron of the ZFX gene in humans and chimpanzees by using the *fmol* DNA cycle sequencing system (Promega). Different annealing temperatures were used according to the primer sequences. Normally, for cycle sequencing, the annealing temperature is estimated by $T_m = 4n_{GC} + 2n_{AT} - 5$, where n_{GC} and n_{AT} are the numbers of G and C nucleotides and of A and T nucleotides in the primer, respectively. Sequencing was conducted in both directions. If a polymorphic site was found among individuals, the region which includes this polymorphic site was sequenced again using a new PCR product. This verification step is to avoid PCR artifacts or gel misreading.

Data Analysis

The age (T) of the MRCA of the sampled sequences can be estimated from sequence variation data by the method of Fu and Li (1996) and Fu (1996). This method is based on the conditional probability $p(T|K)$, given that K polymorphic sites are observed. We used three estimators of T: (1) the mode, t_{mode}, which maximizes $p(T|K)$; (2) the mean, t_{mean}; and (3) the 95% confidence interval ($t_{2.5}$, $t_{97.5}$). The computer program was developed by Fu (1996).

To estimate T, we need to know the number of segregating sites (K), the mutation rate (v), the effective population size (N_e), and the sample size (n). In our study, we used chimpanzee or orangutan as an outgroup to estimate the mutation rate v (Dorit, Akashi, and Gilbert 1995). The evolutionary distance was computed by Jukes and Cantor's (1969) method. The divergence time (t_{div}) between human and chimpanzee is about 4~6 million years ago, while that between human and orangutan is 10~14 million years ago (for a review, see Ruvolo 1996). For long-term evolution, the effective population size of females has been estimated to be 5,000 (Takahata 1993).

Results

Nucleotide Sequences

We sequenced the last intron (1,151 bp) of the ZFX gene from 29 humans and 3 chimpanzees. There is one Alu insertion in this intron: it encompasses 322 bp from position 518 to position 839 in the alignment (data not shown). The GenBank accession number for the sequence data is AF022232.

In the sample of 29 humans, one polymorphic site was found. It is located in the Alu repeat (position 635), at which 17 of the 29 sequences show nucleotide G (17/29 = 58.6%), and the others show A (41.4%). At this site, all three chimpanzee sequences have A. Therefore, the ancestral nucleotide of the human population at this site was probably A, and the mutation (from A to G) probably arose after the human–chimpanzee divergence. This hypothesis is supported by the fact that nucleotide A was also found in orangutan, baboon and squirrel monkey (Shimmin, Chang, and Li 1993). The nucleotide at position 1141 is C in Shimmin, Chang, and Li's (1993) paper. However, in the 29 male individuals we sequenced, the nucleotide at this position is always T. This difference may be due to a PCR error or the presence of a rare variant in the sequence by Shimmin, Chang, and Li (1993).

The number of differences of the last intron of ZFX between human and chimpanzee is 16 or 15, depending on whether the nucleotide at the polymorphic site is G or A. Thus, the average difference is given by (16 × 58.6%) + (15 × 41.1%) ≈ 15.54, and the average proportion of differences is p = 15.54/1,151 = 1.35%. Then, the distance between human and chimpanzee is d = 1.35% by Jukes and Cantor's (1969) method. In the same manner, the distance between human and orangutan is d = 2.59%.

Table 1 shows the distribution of the polymorphism in various human populations. The frequency of nucle-

Table 1
The Distribution of Polymorphism in Different Populations

Population	n	n_G	n_A
Asian	9	4	5
Chinese	2	1	1
Taiwanese	3	2	1
Japanese	4	1	3
Caucasian	8	4	4
German	1	1	0
British	1	0	1
Russian	2	1	1
Canadian	1	1	0
American	3	1	2
African	12	9	3
Pygmy	5	5	0
Nigerian	7	4	3
Total	29	17	12

NOTE.—n is the number of individuals examined; n_G (n_A) is the number of individuals with nucleotide G (or A) at the polymorphic site.

otide G at the polymorphic site is 44.4% (4/9) in Asians and 50% (4/8) in Caucasians. By contrast, the frequency of nucleotide G in Africans is 75% (9/12). However, because of the small sample sizes, a chi-square test of the frequency difference between African and non-African populations is not significant ($\chi^2_{[1]} = 2.6$, $0.1 < P < 0.15$). The fact that G is present at an intermediate frequency in all populations studied (table 1) suggests a certain degree of migration between these populations, in agreement with the data of Hammer et al. (1997) and Harding et al. (1997). Of course, it is also possible that this polymorphism predates the divergence of these populations and has subsequently been carried by these populations.

The Parameter θ and the Mutation Rate *v*

To estimate *T*, we need to know the value of the parameter $\theta = 2N_e v$. Table 2 shows the results from various methods. The first two methods are to use an outgroup (chimpanzee or orangutan) for computing the mutation rate *v*. For example, consider the case with chimpanzee as an outgroup and $t_{div} = 5$ Myr between human and chimpanzee. Assuming that the human generation time is $g = 20$ years, the mutation rate (per sequence per generation) can be estimated by $v = (d/2t_{div}) \times g \times L = [1.35\%/(2 \times 5 \times 10^6)] \times 20 \times 1{,}151 = 0.32 \times 10^{-4}$, where *L* is the sequence length. Obviously, the outgroup method for estimating θ and *v* requires knowledge about the divergence time between human and the outgroup. To study the effects of outgroup selection and the assumed divergence time on the estimation of *T*, we estimated the mutation rate (*v*) from chimpanzee or from orangutan, and from a range of divergence times (see table 2).

For comparison, we also estimated θ by Watterson's (1975) method, i.e., $\hat{\theta} = K/a_n$, where *K* is the number of segragating sites and $a_n = 1 + 1/2 + \ldots + 1/(n - 1)$. In our ZFX data, $K = 1$ and $n = 29$, so $\hat{\theta} = 1/3.93 = 0.26$. The parameter θ can also be estimated by the average of pairwise differences (Tajima 1983), which gives $\hat{\theta} = 0.48$. However, these two methods may not be accurate when the observed number of polymorphic sites (*K*) is very small. An extreme example is the ZFY data (Dorit, Akashi, and Gilbert 1995), in which no variation was observed (i.e., $K = 0$). In this case, both methods give an estimate of $\theta = 0$, which is obviously unacceptable. For this reason, the latter two estimates will not be used in our estimation of *T*.

Estimation of *T* from a Sample of DNA Sequences

The estimates of *T* from the last intron of the ZFX gene are shown in table 3, where T_{mode} is the mode estimate of *T*, T_{mean} is the mean estimate of *T*, and $T_{2.5} \sim T_{97.5}$ is the 95% confidence interval of *T*. These calculations were performed using the computer program of Fu (1996). For example, if the divergence time between human and chimpanzee is 5 Myr, we obtain $T_{mode} = 296{,}000$ years, $T_{mean} = 413{,}000$ years, and a 95% confidence interval of 158,000–908,000 years. Figure 1 shows the conditional distribution of *T* (i.e., $p(T|K)$) in this case. If the divergence time between human and orangutan ($t_{div} = 12$ MYA) is used, we obtain $T_{mode} = 315{,}000$, $T_{mean} = 448{,}000$, and a 95% confidence interval of 167,000–1,022,000 years. The estimates of *T* from the two divergence times are essentially the same.

The divergence times between human and chimpanzee and between human and orangutan used above may not be accurate. To examine the effect of this uncertainty on the estimation of *T*, we considered the lower and upper bounds of t_{div} (table 3). As expected, the lower bound gives a more recent estimate of *T*. Roughly speaking, the effect of the uncertainty of the estimate of *T* is about 5%~10%.

Table 2
Estimation of Parameter θ and Mutation Rate *v* from the Last Intron of the ZFX Gene

Method	t_{div} (Myr)	d (%)	θ	v ($\times 10^{-4}$)
1. Chimpanzee	6.0–4.0 (5.0)	1.35	0.80–1.2 (0.96)	0.27–0.40 (0.32)
2. Orangutan	14.0–10.0 (12.0)	2.61	0.64–0.9 (0.75)	0.21–0.30 (0.25)
3. Tajima	—	—	0.48	0.16
4. Watterson	—	—	0.26	0.09

NOTE.—t_{div} is the divergence time between human and chimpanzee in method 1 or between human and orangutan in method 2; d is the evolutionary distance between species. In the first two methods, the parameter θ was estimated by $2N_e v$, where $N_e = 15{,}000$ for X-linked sequences. For the last two methods, the parameter θ was estimated by Tajima's (1983) and Watterson's (1975) methods, and v was estimated by $v = \theta/(2N_e)$.

Table 3
The Ages (T, $\times 10^3$ years) of the Most Recent Common Ancestors of the Sampled Sequences Based on the Last Introns of the ZFX and ZFY Genes

Outgroup	t_{div}	T_{mode}	T_{mean}	95% Confidence Interval of T
ZFX				
Chimpanzee	4,000	310	438	164~973
	5,000	296	413	158~908
	6,000	278	381	149~822
Orangutan.....	10,000	328	472	172~1,066
	12,000	315	448	167~1,022
	14,000	302	423	160~933
ZFY				
Chimpanzee	4,000	120	185	63~444
	5,000	119	183	62~437
	6,000	118	180	62~428
Orangutan.....	10,000	114	172	60~404
	12,000	113	168	59~394
	14,000	111	164	58~381

NOTE.—t_{div} is the divergence time (in 1,000 years) between human and the outgroup. T_{mode} is the mode estimate of T, and T_{mean} is the (posterior) mean estimate of T. The 95% confidence interval of the age T is from the 2.5% percentile to the 97.5% percentile.

It is interesting to compare the estimates of T from ZFX and ZFY sequences. For this purpose, the method of Fu (1996) was also applied to the ZFY intron sequence data of Dorit, Akashi, and Gilbert (1995) (table 3). It is clear that the T estimate from the last intron of ZFY is younger than that from the last intron of ZFX. Since the the difference in sequence variation between ZFX and ZFY is small ($K = 0$ for ZFY and $K = 1$ for ZFX), this difference in the estimates of T is largely due to the larger effective population size of an X-linked sequence than of a Y-linked sequence ($3N_e$ vs. N_e). The uncertainty of divergence time on the estimation of T from ZFY introns is less than 5%.

By averaging the mutation rates obtained from using different outgroups (chimpanzee or orangutan), the most probable age of the MRCA of the sampled ZFX intron sequences is estimated to be about 306,000 years ago. The 95% confidence interval is 162,000–952,000 years.

Discussion

In this study, the nucleotide variation in the last intron of the ZFX gene in humans was examined, with a worldwide sample of 29 individuals. Among the 1,151 nucleotides studied, only one polymorphic site was found, which is located on an Alu repeat. The nucleotide diversity (per site) is 0.04%, supporting the view that the sequence variation in nuclear DNA in humans is very low (Li and Sadler 1992; Dorit, Akashi, and Gilbert 1995).

One interesting quantity in studying human evolution is the age T of the MRCA of the DNA sequences in a sample. To estimate T, we used the new method of Fu and Li (1996) and Fu (1996), which is particularly useful in the case where the number of segregating sites (K) is very small or even zero. For the ZFX intron data,

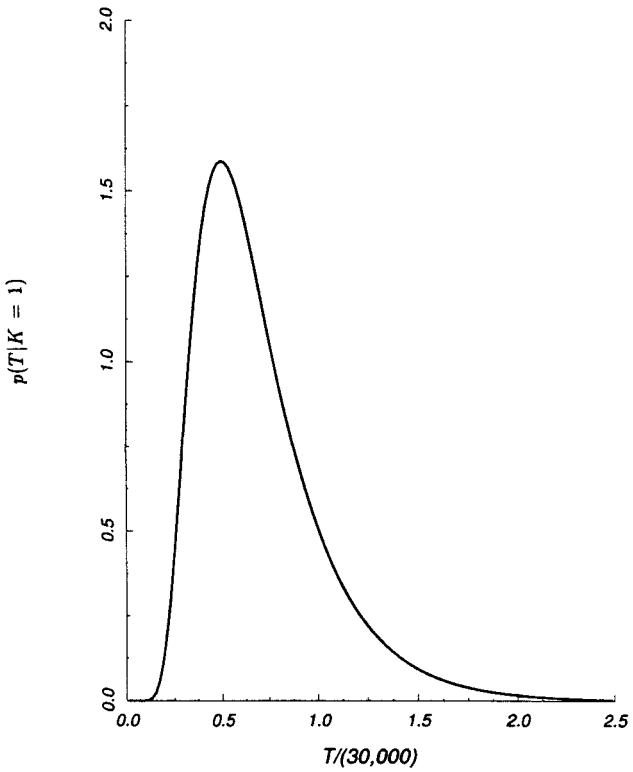

FIG. 1.—The distribution of $p(T|K = 1)$ for the last intron of the ZFX gene, with chimpanzee as an outgroup and $t_{div} = 5$ Myr.

we estimated that T_{mode} is about 306,000 years and that the 95% confidence interval of T is 162,000–952,000 years. We also estimated the age of the MRCA of the ZFY introns sampled by Dorit, Akashi, and Gilbert (1995); T_{mode} is about 116,000 years, and the 95% confidence interval of T is 61,000–416,000 years (table 3). There are several possible factors influencing the different estimates of T from ZFX and ZFY introns. The most important factor is that the effective population size of an X-linked gene is three times that of a Y-linked gene. As a consequence, the coalescent time of X-linked sequences is expected to be longer than that of Y-linked sequences. Second, T increases with the number of segragating sites. Thus, a larger T for the ZFX introns is expected because one segregating site was observed in the ZFX introns but none was found in the ZFY introns. Third, the mutation rate in a Y-linked gene is higher than that in an X-linked gene (Shimmin, Chang, and Li 1993; Chang, Hewett-Emmett, and Li 1996), which can counterbalance the effect of the first two factors to some degree. Fourth, because there is no recombination in the Y chromosome, selective sweeps may also play a role (Whitfield, Suiston, and Goodfellow 1995). These differences in population dynamics point to the need for using not only Y-linked sequences, but also X-linked and autosomal sequences when studying human evolution.

Acknowledgments

We thank Michael Bamshad for helpful comments. This study was supported by NIH grants GM50428

(Y.X.F.), GM30998, GM50798 (W.H.L.), and RR-00064 (L.B.J.) and NSF grants DBS-9310105 and SBR-9514733 (L.B.J.).

LITERATURE CITED

BROOKFIELD, J. F. Y. 1997. Importance of ancestral DNA ages. Nature **388**:134.

CANN, R. L., M. STONEKING, and A. C. WILSON. 1987. Mitochondrial DNA and human evolution. Nature **325**:31–36.

CAVALLI-SFORZA, L. L., P. MENOZZI, and J. MOUNTAIN. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. Proc. Natl. Acad. Sci. USA **85**:6002–6006.

CHANG, B. H.-J., D. HEWETT-EMMETT, and W.-H. LI. 1996. Male-to-female ratios of mutation rate in higher primates estimated from intron sequences. Zool. Stud. **35**:36–48.

DONNELLY, P., S. TAVARÉ, D. J. BALDING, and R. C. GRIFFITHS. 1996. The age of the common ancestor of human males estimated from ZFY intron sequence data. Science **272**: 1357–1359.

DORIT, R. L., H. AKASHI, and W. GILBERT. 1995. Absence of polymorphism at the ZFX locus on the human Y chromosome. Science **268**:1183–1185.

ELLSWORTH, D. L., K. D. RITTENHOUSE, and R. L. HONEYCUTT. 1993. Artificial variation in randomly amplified polymorphic DNA banding patterns. Biotechniques **14**:214–217.

FU, Y.-X. 1996. Estimating the age of the common ancestor of a DNA sample using the number of segregating sites. Genetics **144**:829–838.

FU, Y.-X., and W.-H. LI. 1996. The age of the common ancestor of human males estimated from ZFY intron sequence data. Science **272**:1356–1357.

FULLERTON, S. M., R. M. HARDING, A. J. BOYCE, and J. B. CLEGG. 1994. Molecular and population genetic analysis of allelic sequence diversity at the human beta-globin locus. Proc. Natl. Acad. Sci. USA **91**:1805–1809.

HAMMER, M. F. 1995. A recent common ancestry for human Y chromosomes. Nature **378**:376–378.

HAMMER, M. F., A. B. SPURDLE, T. KARAFET et al. (11 co-authors). 1997. The geographic distribution of human Y chromosome variation. Genetics **145**:787–805.

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.

JORDE, L. B., M. J. BAMSHAD, W. S. WATKINS, R. ENGER, A. E. FRALEY, P. A. KRAKOWIAK, K. D. CARPENTER, H. SOODYALL, T. JENKINS, and A. R. ROGERS. 1995. Origins and affinities of modern humans: a comparison of mitochondrial and nuclear genetic data. Am. J. Hum. Genet. **57**:523–538.

JORDE, L. B., A. R. ROGERS, M. BAMSHAD, W. S. WATKINS, P. KRAKOWIAK, S. SUNG, J. HERE, and H. C. HARPENDING. 1997. Microsatellite diversity and demographic history of modern humans. Proc. Natl. Acad. Sci. USA **94**:3100–3103.

JUKES, T. H., and C. R. CANTOR. 1969. Evolution of protein molecules. Pp. 21–132 *in* H. N. MUNRO, ed. Mammalian protein metabolism. Academic Press, New York.

LI, W.-H., and L. SADLER. 1992. DNA variation in humans and its implication for human evolution. Oxf. Surv. Evol. Biol. **8**:111–134.

MADDISON, D. R., M. ROVOLO, and D. L. SWOFFORD. 1992. Geographic origins of human mitochondrial DNA: phylogenetic evidence from control region sequences. Syst. Biol. **41**:111–124.

NEI, M., and A. K. ROYCHOUDHURY. 1982. Genetic relationship and evolution of human races. Evol. Biol. **14**:1–59.

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

RUVOLO, M. 1996. A new approach to studying modern human origins: hypothesis testing with coalescence time distributions. Mol. Phylogenet. Evol. **5**:202–219.

SAUNDERS, L. W., S. TRAVARÉ, and G. WATTERSON. 1984. On the genealogy of nested subsamples from haploid population. Adv. Appl. Probab. **16**:471–491.

SCHNEIDER-GÄDICKE, A., P. BEER-ROMERO, L. G. BROWN, G. MARDON, S.-W. LUOH, and D. C. PAGE. 1989. Putative transcription activator with alternative isoforms encoded by human ZFX gene. Nature **342**:708–711.

SHIMMIN, L. C., B. H.-J. CHANG, and W.-H. LI. 1993. Male-driven evolution of DNA sequences. Nature **362**:745–747.

TAJIMA, F. 1983. Evolution relationship of DNA sequences in finite populations. Genetics **105**:437–460.

TAKAHATA, N. 1993. Allelic genealogy and human evolution. Mol. Biol. Evol. **10**:2–22.

TEMPLETON, A. R. 1992. Human origins and analysis of mitochondrial DNA sequences. Science **255**:737.

THORNE, A. G., and M. H. WOLPOFF. 1992. The multiregional evolution of humans. Sci. Am. **266**:28–33.

TISHKOFF, S. A., E. DIETZSCH, W. SPEED et al. (15 co-authors). 1996. Global pattern of linkage disequilibrium at the CD4 locus and modern human origins. Science **271**:1380–1387.

VIGILANT, L., R. PENNINTON, H. HARPENDING, T. D. KOCHER, and A. C. WILSON. 1989. Mitochondrial DNA sequences in single hairs from a southern African population. Proc. Natl. Acad. Sci. USA **86**:9350–9354.

WATTERSON, G. 1975. On the number of segregation sites. Theor. Popul. Biol. **7**:256–276.

WHITFIELD, L. S., J. E. SUISTON, and P. N. GOODFELLOW. 1995. Sequence variation of the human Y chromosome. Nature **378**:379–380.

XIONG, W., W.-H. LI, I. POSNER, T. YAMAMURA, A. YAMAMOTO, A. M. GOTTO JR., and L. CHAN. 1991. No severe bottleneck during human evolution: evidence from two apolipoprotien C-II deficiency alleles. Am. J. Hum. Genet. **48**: 383–389.

NAOYUKI TAKAHATA, reviewing editor

Accepted October 21, 1997