

# A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history

(DNA polymorphism/nucleotide diversity/comparative DNA sequencing/denaturing high-performance liquid chromatography/microsatellite)

PETER A. UNDERHILL\*<sup>†</sup>, LI JIN\*, RACHEL ZEMANS\*, PETER J. OEFNER<sup>‡</sup>, AND L. LUCA CAVALLI-SFORZA\*

Departments of \*Genetics and <sup>‡</sup>Biochemistry, Stanford University, Stanford, CA 94305

Contributed by L. Luca Cavalli-Sforza, September 14, 1995

**ABSTRACT** A polymorphic C → T transition located on the human Y chromosome was found by the systematic comparative sequencing of Y-specific sequence-tagged sites by denaturing high-performance liquid chromatography. The results of genotyping representative global indigenous populations indicate that the locus is polymorphic exclusively within the Western Hemisphere. The pre-Columbian T allele occurs at >90% frequency within the native South and Central American populations examined, while its occurrence in North America is ≈50%. Concomitant genotyping at the polymorphic tetranucleotide microsatellite *DYS19* locus revealed that the C → T mutation displayed significant linkage disequilibrium with the 186-bp allele. The data suggest a single origin of linguistically diverse native Americans with subsequent haplotype differentiation within radiating indigenous populations as well as post-Columbian European and African gene flow. The mutation may have originated either in North America at a very early time during the expansion or before it, in the ancestral population(s) from which all Americans may have originated. The analysis of linkage of the *DYS199* and the *DYS19* tetranucleotide loci suggests that the C → T mutation may have occurred around 30,000 years ago. We estimate the nucleotide diversity over 4.2 kb of the nonrecombining portion of the Y chromosome to be 0.00014. Compared to autosomes, the majority of variation is due to the smaller effective population size of the Y chromosome rather than selective sweeps. There begins to emerge a pattern of pronounced geographical localization of Y-specific nucleotide substitution polymorphisms.

The apparently nonrecombining (Y specific) portion of the human Y chromosome provides a unique system for the study of human origins, migration, and admixture (1). However, few such polymorphisms have been identified to date, presumably because of a smaller effective population size relative to autosomes and reduction of variation on the entire chromosome due to selection at a single locus (2–7). Moreover, some of the previously described loci are not amenable to convenient PCR genotyping techniques (e.g., see refs. 8 and 9).

Some Y-specific polymorphisms such as the Y *Alu* polymorphic element (10, 11) and an A to G transition (12), which have apparently arisen once in human evolution, have underscored the great potential of such loci in creating readily interpretable Y-chromosomal haplotypes for evolutionary analysis. Such data are of particular value in the study of human migration. A recent report (4) indicated an absence of polymorphism over 729 bp on the human Y chromosome. Here we report on the discovery of a C → T point mutation in association with representative genotype results from globally diverse populations. Our observations of a locus polymorphic only within the Western Hemisphere, which shows the pattern of linkage

disequilibrium with a previously described microsatellite locus (13), indicate that the C → T mutation may have been introduced exclusively during the early stages of human habitation of the Americas. Its potential utility in understanding the pre-Columbian peopling of the New World as well as inferring the degree of admixture within indigenous populations are discussed. The comparative sequencing of >4.2 kb of the nonrecombining portion of the human Y chromosome reveals detectable amounts of base substitution polymorphisms.

## MATERIALS AND METHODS

Previously described (14) sequence-tagged sites (STSs) were amplified by using a “touchdown” PCR regime (12, 15). Amplification primers were obtained from Research Genetics (Huntsville, AL). Reaction products were analyzed by agarose gel electrophoresis. We have developed a method of comparative DNA sequencing (16) based on the capability of ion-pair reverse-phase liquid chromatography on alkylated nonporous poly(styrene divinylbenzene) particles (17) to resolve homoduplex molecules under conditions of partial denaturation. We refer to this method as denaturing high-performance liquid chromatography (DHPLC). Duplexes for analysis were formed upon mixing, denaturing, and reannealing Y-specific PCR amplified products from various male genomic DNA samples.

Individual samples, which had displayed heteroduplex signature patterns, were purified with Qiagen (Chatsworth, CA) QIAquick spin columns and both strands were sequenced to confirm the polymorphism using the amplification oligonucleotides as sequencing primers and DyeDeoxy terminator cycle sequencing reagents (Perkin-Elmer) following the manufacturer’s protocols. Products of sequencing reactions were purified by Centri-Sep spin columns (Princeton Separations, Adelphia, NJ) and analyzed on an Applied Biosystems DNA sequencer (model 373A). The reliability of DHPLC to efficiently identify polymorphic STSs was corroborated by sequencing the same STSs by conventional fluorescent DNA sequencing methods.

Genotyping of the *DYS199* locus was done by allele-specific PCR. Two amplification reactions were performed for each sample in which the *DYS199* forward primer (Fig. 1) was combined with one or the other allele-specific reverse primer. The C-specific reverse primer sequence was 5′-GGTAC-CAGCTCTTCCTAATTG-3′ and the T-specific primer was 5′-GGTACCAGCTCTTCCTAATTA-3′. Each 15-μl allele reaction mixture contained 50 ng of genomic DNA, 5 pmol of each primer, 100 μM each dNTP, 1.5 mM MgCl<sub>2</sub>, 10 mM Tris-HCl (pH 9.0), 50 mM KCl, and 0.375 unit of *Taq* polymerase. Thirty cycles of allele-specific PCR were performed in

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: STS, sequence-tagged site; DHPLC, denaturing high-performance liquid chromatography.

<sup>†</sup>To whom reprint requests should be addressed.

5' - TAATCAGTCTCCTCCAGCAAGTGATATGCAACTGAGATT  
 CCTTATGACACATCTGAACACTAGTGGATTGCTTTGTAGTAG  
 GAACAAGGTACATTGCGGGGATAAATGTGGCCAAGTTTTATC  
 TGCTGCCAGGGCTTTCAAATAGTTGACCTGACAATGGGTCA  
 CCTCTGGGACTGA(C)AATTAGGAAGACTGGTACCTAAATG  
 AAAGATGCCCTTAAATTCAGATTCCACAATTTT - 3'

FIG. 1. Complete 241-bp *DYS199* nucleotide sequence. Polymorphic nucleotide position 181 is indicated by parentheses. Regions used for primer design are underlined.

a Perkin-Elmer model 9600 thermal cycler using the following two-step PCR regime: 94°C for 30 s and 61°C for 20 s. Each reaction was analyzed on an ethidium bromide-stained agarose gel and the specific 201-bp product was UV visualized to determine allelic state (Fig. 2). Positive controls of both C and T allele individuals (whose genotypes were previously confirmed by sequencing) were always included in the allele-specific PCR genotyping experiments.

The *DYS19* tetranucleotide microsatellite was genotyped with a fluorescently labeled primer and one unlabeled primer (1 pmol each) in a 5- $\mu$ l touchdown PCR mixture (see above) containing 20 ng of genomic DNA, 200  $\mu$ M dNTPs, 0.25 unit of *Taq* polymerase, and 3 mM MgCl<sub>2</sub>. A single microliter of amplified product was combined with a fluorescently labeled internal size standard ladder and analyzed on a 6% denaturing gel in an Applied Biosystems DNA sequencer (model 373A) using Genescan 672 fragment sizing software (Perkin-Elmer). Subsequent allele calling was accomplished with Genotyper software (Perkin-Elmer).

Nucleotide diversity ( $\pi$ ) was calculated according to ref. 18 as

$$\pi = \sum_{ij} x_i x_j d_{ij},$$

where  $x_i$  and  $x_j$  are, respectively, the frequencies of the  $i$ th and  $j$ th types and  $d_{ij}$  is the proportion of nucleotide differences between the  $i$ th and  $j$ th types of DNA sequences.

## RESULTS

Expanding our initial search of 7 STSs for Y-specific polymorphic loci, in which a single A  $\rightarrow$  G transition had been

Table 1. Identity, locus, and size of Y-specific STSs sequenced

STS	Locus	Size, bp
sY-17	<i>DYS250</i>	289
sY-65	<i>DYS260</i>	269
sY-67	<i>DYS262</i>	78
sY-84	<i>DYS273</i>	286
sY-85	<i>DYS274</i>	295
sY-87	<i>DYS275</i>	212
sY-95	<i>DYS280</i>	263
sY-98	<i>DYS282</i>	226
sY-103	<i>DYS199</i>	201
sY-105	<i>DYS201</i>	261
sY-117	<i>DYS209</i>	222
sY-161	<i>DYS243</i>	276

The reported size includes only the length scanned for polymorphism and not the primer sequence.

observed in 1357 bp (12), an additional 12 loci (Table 1) were surveyed by comparatively sequencing 20 individuals of diverse geographic origin. This was achieved by crossing PCR products with an arbitrary reference African male. Duplex analysis was then performed by means of DHPLC under partially denaturing conditions. As shown in Fig. 3, heteroduplexes were readily resolved within minutes from their corresponding homoduplexes. Screening of these additional 2878 bp revealed a single polymorphic nucleotide substitution, a C  $\rightarrow$  T transition, at base position 181 of locus *DYS199* (Fig. 1).

A total of 173 humans from five continents as well as 6 nonhuman primates were subsequently genotyped at this biallelic locus (Table 2) using allele-specific PCR techniques. The presence of one allele or the other was unequivocally detected in all cases (Fig. 2), including nonhuman primates. The T allele occurred exclusively in 90.5% (i.e., 38/42) of the South and Central American populations examined, while all other human individuals ( $n = 123$ ) outside the Americas and the nonhuman primates displayed only the C allele. The T allele was also observed in indigenous North American Navajo and Eskimo populations, but each at a lower frequency of 50% and 67%, respectively.

A subset of 133 males was also genotyped at the microsatellite locus *DYS19* (13, 19), which permitted the construction of Y-chromosomal haplotypes (Table 3). The T allele was

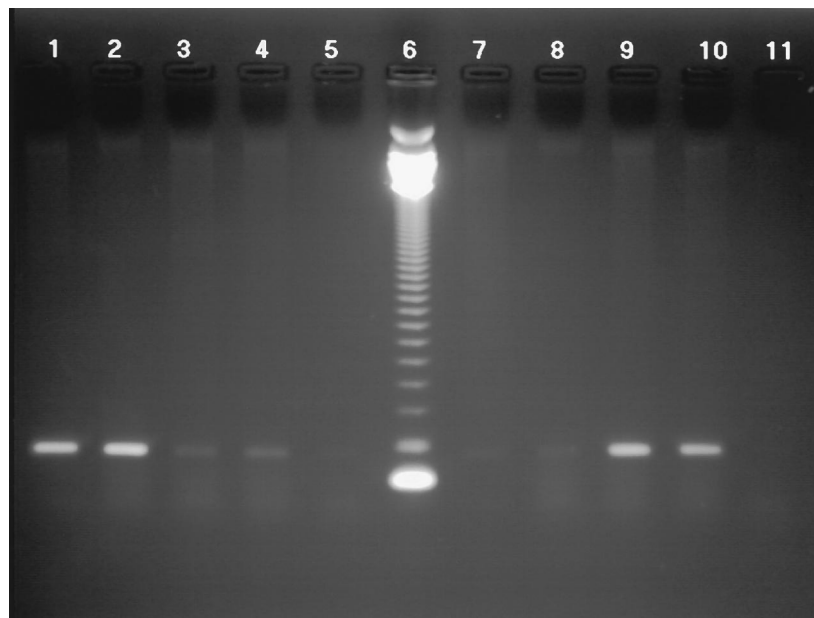


FIG. 2. Representative 201-bp allele-specific PCR products analyzed on an ethidium bromide-stained agarose gel. Lanes 1-5 and 7-11, C- and T-specific reactions, respectively. Lanes 1, 2, 7, and 8, C males; lanes 3, 4, 9, and 10, T males; lanes 5 and 11, females; lane 6, 123-bp size ladder.

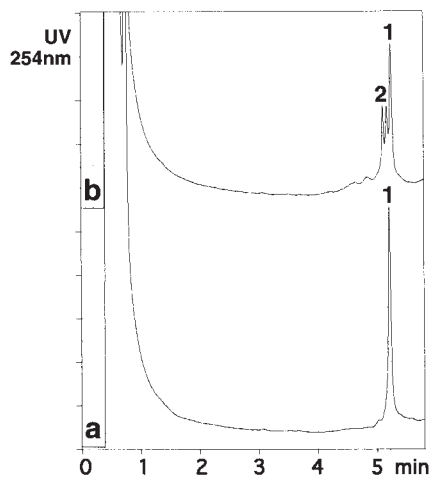


FIG. 3. Representative chromatograms of homo- (peak 1) and heteroduplexes (peak 2) created upon reannealing 241-bp PCR products of a male reference with an Asian (*a*) and an American Indian (*b*) individual. Heteroduplexes reflect a single-base substitution. Column, poly(styrene divinylbenzene) C<sub>18</sub>, 2.1  $\mu$ m, 50  $\times$  4.6 mm i.d.; linear gradient, 9.25–15.75% (vol/vol) acetonitrile in 0.1 M triethylammonium acetate (pH 7.0) in 5.5 min at a flow rate of 1 ml/min; column temperature, 55°C.

found to be in strong linkage disequilibrium with the 186-bp microsatellite allele ( $\chi^2 = 68.7$  with 1 df). Specifically, the 186-bp allele was exclusively associated with the *DYS199* T allele in all six American populations studied, while it is relatively rare (6.5%) in non-American populations and is, in these cases, associated only with the *DYS199* C allele. Conversely, with the exception of the Eskimos, the *DYS199* C allele is predominantly associated with the *DYS19* microsatellite 190-bp and 194-bp alleles, especially in non-American populations. In the American populations, five of the six individuals (two Mayans, one Surui, and two Navajos) having the *DYS199* C allele also had the *DYS19* 190-bp allele. The origin of these five individuals cannot be assessed unequivocally, although, given the high frequency (7/9 = 78%) of this haplotype in Europe, European admixture is likely. However, we still cannot exclude completely the possibility that some of the *DYS199* C

Table 2. *DYS199* genotypes by population

Continent	Population	C	T	<i>n</i>
Human				
Africa	Lisongo	2		2
	CAR pygmy	20		20
	Zaire pygmy	8		8
	Bushman	1		1
Asia	Chinese	16		16
	Cambodian	18		18
	Japanese	14		14
Oceania	Melanesian	4		4
	New Guinean	12		12
	Australian	2		2
Europe	Caucasian	22		22
America	Karitiana		15	15
	Surui	1	16	17
	Mayan	3	5	8
	Colombian		2	2
	Eskimo	2	4	6
	Navajo	3	3	6
Nonhuman primate				
	Chimpanzee	4		4
	Orangutan	2		2
	Total	134	45	179

Table 3. *DYS199* and *DYS19* haplotypes

	C allele						T allele			<i>n</i>
	182	186	190	194	198	202	186	190	194	
Africa	1	3	3	5	10	4	0	0	0	26
Asia	0	3	9	19	5	4	0	0	0	40
Oceania	0	0	3	8	1	1	0	0	0	13
Europe	0	0	7	2	0	0	0	0	0	9
America										
Karitiana	0	0	0	0	0	0	8	3	0	11
Surui	0	0	1	0	0	0	13	0	0	14
Mayan	0	0	2	1	0	0	2	0	1	6
Colombia	0	0	0	0	0	0	1	0	1	2
Eskimo	0	2	0	0	0	0	4	0	0	6
Navajo	0	0	3	0	0	0	2	1	0	6
Total	1	8	28	35	16	9	30	4	2	133

alleles in the American populations were pre-Columbian. One Mayan male, previously shown (12) to have an African Y chromosome, had the 194-bp C haplotype. Within native Americans, the 186-bp C haplotype was observed in Eskimos, Navajos, and some other North American Indian populations (data not shown).

Two polymorphic sites have now been found by scanning 4235 bp of Y-specific sequence in 20 human males from around the world. An estimate of nucleotide diversity ( $\pi$ ), which is the average pairwise difference for all pairs of sequences randomly drawn from a population (16), was calculated for the nonrecombining portion of the human Y chromosome to be 0.00014 per nucleotide. Using a previously compiled composite estimate of 0.0011 for autosomal nucleotide diversity (20), we estimated a 7.9-fold (i.e., 0.0011/0.00014) reduction in the predicted polymorphism frequency relative to an autosomal locus.

## DISCUSSION

The clarity of the human historical content of the Y chromosome resides in the absence of recombination on the Y-specific region such that haplotype information is preserved. Our discovery of another base substitution, this time within the Western Hemisphere and thus relevant to the colonization of the New World, underscores the efficacy of this approach.

The results of genotyping representative continental populations show that *DYS199* is apparently polymorphic only within the Western Hemisphere. This observation indicates the discovery of a pre-Columbian allele within the New World that can be used to assess historical origins and migrations of early native Americans. The C  $\rightarrow$  T mutation can be localized to the American continent if our inability to detect the T allele outside the Americas is confirmed by further investigation. However, we cannot yet exclude that the mutation occurred in the Northeast Asian population(s) from which all or most American natives originated. Furthermore, the presence of the T allele in Navajo and Eskimo populations as well as other North American groups (unpublished data) suggests that the mutation occurred prior to migration into Central and South America, although back migrations cannot be discounted. Obviously, by genotyping additional contemporary Asiatic and native American populations it should be possible to trace the genetic migratory trail from Asia and the peopling of the New World.

The relative age of both the *DYS199* and *DYS19* polymorphisms can be estimated. Since nonhuman primates possess the *DYS199* C allele, it is ancestral to the T allele, assuming the probability of recurrent C  $\rightarrow$  T mutations is extremely low. Furthermore, since the *DYS19* 186-bp allele is far more frequent than the other two observed 190-bp and 194-bp alleles in *DYS199* T haplotype Y chromosomes, it is reasonable

to hypothesize that the T allele was initially introduced on precursor chromosomes bearing the *DYS19* 186-bp allele and that the 190- and 194-bp T haplotypes are more recently derived. The underlying rationale is that the most frequent genotype is most often the oldest under either the infinite allele model (21) or a stepwise mutation model (22). The relative timing is further supported by the observation that the 186-bp T haplotype is the most frequent in the New World populations studied and by the fact that microsatellite loci display much higher mutation rates (23) than those of single nucleotide substitutions. The exclusive localization of the *DYS199* T allele in the Americas, closely coupled with a particular Y-specific microsatellite allele (186 bp), strengthens the claim of a specific haplotype associated with pre-Columbian New World populations. In fact, the presence of the 186-bp C haplotype in Eskimos and Navajos further supports the hypothesis that the 186-bp T allele might be derived from the 186-bp C allele, and both haplotypes could comprise the ancestral population(s) that peopled the Americas. Interestingly, the T allele appears in all three major linguistic groups—namely, Amerind, Na-Dene, and Eskimo-Aleut (24).

Haplotype analysis of the mtDNA locus, coupled with anthropological and linguistic data, has been extensively used to study the peopling of the Americas (25). Suitable Y chromosome variation is desirable to complement such studies. Previous studies of population substructure in four Amerind populations using both mtDNA and Y-chromosome restriction fragment length polymorphism haplotypes were reported (26). Unfortunately, their Y-chromosomal data were less informative as an index of substructure, since many haplotypes were shared in both Old and New World populations. Geographically specific polymorphic loci such as *DYS199* are more useful in dissecting post-Columbian admixture. This has been independently realized with the description of another Y-chromosome haplotype based on the microsatellite *DYS19* and sequence variation in alphoid repeats (27). Although additional research is necessary to unravel the peopling of the New World prior to the Age of Exploration, this C → T transition together with the *DYS19* locus can be exploited immediately in assessing the level of admixture within native American and Hispanic populations throughout the Americas. Even if some of the C allele present in contemporary American populations is autochthonous, the maximum level of male-mediated post-Columbian admixture can still be estimated by simply using the percentage of male individuals displaying the C allele. Accurate assessment of admixture will require the discovery of additional Y-specific polymorphic loci that display alleles specific to African or Caucasian populations as well as possibly analyzing pre-Columbian human remains.

The moment at which the C → T transition occurred can be estimated from the ratio of non-186-bp T haplotypes to all haplotypes associated with the T allele in American populations, which is 0.167 (6/36) assuming there was no selection. Luria and Delbrück (28), in their analysis of the dynamics of bacterial mutations, showed that the proportion of mutants in a population is expected to be equal to the product of mutation rate and the time since the first appearance of the T allele. Based on a recent estimate of an average mutation rate of  $1.5 \times 10^{-4}$  (29) derived from  $\approx 700$  tetranucleotide loci using the distribution of allelic differences in terms of number of repeats, and an average generation time of 27 years (30), the C → T transition should have occurred 30,000 years ago. This value is similar to an earlier estimate of the time of entry to America (32,000 years) based on classical genetic markers (31). However, in contrast, the mutation rate of  $2.1 \times 10^{-3}$  reported by Weber and Wong (23) would indicate that the C → T transition had occurred only 2147 years ago, which obviously represents an underestimate. Knowledge of the mutation rate

of this particular tetranucleotide would be important for validating this dating.

While our observation of sparse variation on the Y chromosome is consistent with a recent report (4), we are now able to present a preliminary estimate of nucleotide diversity ( $\pi$ ) on the nonrecombining portion of the human Y chromosome that is based on actually observed polymorphisms. Our estimate of  $\pi = 0.00014$  differs from a compiled composite value for autosomal nucleotide diversity of 0.0011 (20). Comparing these values yields an estimated 7.9-fold reduction in the predicted polymorphism frequency of the Y chromosome relative to an autosomal locus. Since the nucleotide diversity of human Y-chromosome loci is expected to be reduced 4-fold because of its smaller effective population size, much of the observed reduction can be attributed to this factor alone. Additional factors such as polygyny would also contribute to the lower amount of variation observed relative to autosomal loci. These results indicate that it may not be necessary to invoke homogenization by rapid selective sweeps (4) to explain the reduced but detectable amount of variation that is apparently geographically maintained on the nonrecombining regions of the Y chromosome. Our estimate of nucleotide diversity should be considered preliminary given that only 4.2 kb and a small number of individuals have been surveyed. The discovery of additional single base substitutions, which will be accelerated by using DHPLC for comparative sequencing, will yield a more accurate estimate. Nonetheless, it is only when numerous Y-specific chromosome polymorphic DNA markers, as well as additional mutation rate data from nonhuman primates, are available that the rigorous phylogenetic assessment of human evolution based on the Y chromosome can be fully realized. The discovery of this single nucleotide substitution covering a unique geographic area is remarkably similar to our earlier finding of another such substitution localized in Africa. It indicates that Y chromosome variation appears to have a pattern of geographic distribution quite unlike mtDNA and autosomes.

We especially thank those donor individuals who provided material for DNA analysis. The Eskimo samples were generously provided by Mary Ann Walkinshaw, Department of Public Safety, Alaska, while Navajo DNA was kindly donated by Gary Troup, University of New Mexico School of Medicine. We thank Douglas Vollrath for providing information concerning the Y-chromosome STSs and Marc Feldman for constructive suggestions. We thank Alice Lin for cell line culture expertise and R. Spencer Wells for advice concerning allele-specific PCR. Additional technical assistance was provided by the following undergraduate students: Bryan Sun, Marybell Ang, and Vishal Doctor. Support was provided by National Institutes of Health Grants 20467 and 28428 to L.L.C.-S. and by National Institutes of Health Grant 20505 to Ronald Davis as well as National Institutes of Health Training Grant T32-GS08404 to L.J.

1. Casanova, M., Leroy, P., Boucekine, C., Weissenbach, J., Bishop, C., Fellous, M., Purrello, M., Fiori, G. & Siniscalco, M. (1985) *Science* **230**, 1403–1406.
2. Malaspina, P., Persichetti, F., Novelletto, A., Iodice, C., Terrenato, L., Wolfe, J., Ferraro, M. & Prantero, G. (1990) *Ann. Hum. Genet.* **54**, 297–305.
3. Berry, A. J., Ajoika, J. W. & Kreitman, M. (1991) *Genetics* **129**, 1111–1118.
4. Dorit, R. L., Akashi, H. & Gilbert, W. (1995) *Science* **268**, 1183–1185.
5. Pääbo, S. (1995) *Science* **268**, 1141–1142.
6. Tyler-Smith, C. & Taylor, L. (1988) *J. Mol. Biol.* **203**, 837–848.
7. Spurdle, A. & Jenkins, T. (1992) *Am. J. Hum. Genet.* **50**, 107–125.
8. Ngo, K. Y., Vergnaud, G., Johnsson, C., Lucotte, G. & Weissenbach, J. (1986) *Am. J. Hum. Genet.* **38**, 407–418.
9. Mathias, N., Bayes, M. & Tyler-Smith, C. (1994) *Hum. Mol. Genet.* **3**, 115–123.
10. Hammer, M. F. (1994) *Mol. Biol. Evol.* **11**, 749–761.

11. Hammer, M. F. & Horai, S. (1995) *Am. J. Hum. Genet.* **56**, 951–962.
12. Seielstad, M. T., Herbert, J. M., Lin, A. A., Underhill, P. A., Ibrahim, M., Vollrath, D. & Cavalli-Sforza, L. L. (1994) *Hum. Mol. Genet.* **12**, 2159–2161.
13. Roewer, L., Arnemann, J., Spurr, N. K., Grzeschik, K.-H. & Epplen, J. T. (1992) *Hum. Genet.* **89**, 389–394.
14. Vollrath, D., Foote, S., Hilton, A., Brown, L. G., Beer-Romano, P., Bogan, J. S. & Page, D. C. (1992) *Science* **258**, 52–59.
15. Don, R. H., Cox, P. T., Wainwright, B. J., Baker, K. & Mattick, J. S. (1991) *Nucleic Acids Res.* **19**, 4008.
16. Oefner, P. J. & Underhill, P. A. (1995) *Am. J. Hum. Genet.* **57**, Suppl., 1547 (abstr.).
17. Huber, C. G., Oefner, P. J. & Bonn, G. K. (1993) *Anal. Biochem.* **212**, 351–358.
18. Nei, M. & Li, W.-H. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5269–5273.
19. Santos, F. R., Pena, S. D. J. & Epplen, J. T. (1993) *Hum. Genet.* **90**, 655–656.
20. Li, W.-H. & Sadler, L. A. (1991) *Genetics* **129**, 513–523.
21. Watterson, G. A. & Guess, H. A. (1977) *Theor. Popul. Biol.* **11**, 141–160.
22. Chakraborty, R. (1977) *J. Mol. Evol.* **9**, 313–322.
23. Weber, J. L. & Wong, C. (1993) *Hum. Mol. Genet.* **2**, 1123–1128.
24. Greenberg, J. H. (1987) *Language in the Americas* (Stanford Univ. Press, Stanford, CA).
25. Szathmary, E. J. E. (1993) *Am. J. Hum. Genet.* **53**, 793–799.
26. Torroni, A., Chen, Y.-S., Semino, O., Santachiara-Benecere, A. S., Scott, C. R., Lott, M. T., Winter, M. & Wallace, D. C. (1994) *Am. J. Hum. Genet.* **54**, 303–318.
27. Pena, S. D. J., Santos, F. R., Bianchi, N. O., Bravi, C. M., Carnese, F. R., Rothhammer, F., Gerelsaikhon, T., Munkhtuja, B. & Oyunsuren, T. (1995) *Nat. Genet.* **11**, 15–16.
28. Luria, S. E. & Delbrück, M. (1943) *Genetics* **28**, 491–511.
29. Jin, L., Zhong, Y., Shriver, M. D., Deka, R. & Chakraborty, R. (1994) *Am. J. Hum. Genet.* **55**, Suppl., 39 (abstr.).
30. Weiss, K. (1973) *Am. Antiq.* **38**, 1–186.
31. Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. (1995) *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, NJ), p. 320–326.