



Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations.

R Deka, L Jin, M D Shriver, et al.

Genome Res. 1996 6: 1177-1184

Access the most recent version at doi:[10.1101/gr.6.12.1177](https://doi.org/10.1101/gr.6.12.1177)

References This article cites 22 articles, 5 of which can be accessed free at:
<http://genome.cshlp.org/content/6/12/1177.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © Cold Spring Harbor Laboratory Press

RESEARCH

Dispersion of Human Y Chromosome Haplotypes Based on Five Microsatellites in Global Populations

Ranjan Deka,^{1,5} Li Jin,² Mark D. Shriver,¹ Ling Mei Yu,¹ Nilmani Saha,¹ Ramiro Barrantes,³ Ranajit Chakraborty,⁴ and Robert E. Ferrell¹

¹Department of Human Genetics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261; ²Department of Genetics, Stanford University School of Medicine, Stanford, California 94305; ³Istituto de Investigaciones en Salud, Universidad de Costa Rica, San Jose, Costa Rica; ⁴Human Genetics Center, University of Texas Health Science Center, Houston, Texas 77225

We have analyzed five microsatellite loci from the nonrecombining portion of the human Y chromosome in 15 diverse human populations to evaluate their usefulness in the reconstruction of human evolution and early male migrations. The results show that, in general, most populations have the same set of the most frequent alleles at these loci. Hypothetical ancestral haplotypes, reconstructed on the basis of these alleles and their close derivatives, are shared by multiple populations across racial and geographical boundaries. A network of the observed haplotypes is characterized by a lack of clustering of geographically proximal populations. In spite of this, few distinct clusters of closely related populations emerged in the network, which are associated with population-specific alleles. A tree based on allele frequencies also shows similar results. Lack of haplotypic structure associated with the presumed ancestral haplotypes consisting of individuals from almost all populations indicate a recent common ancestry and/or extensive male migration during human evolutionary history. The convergent nature of microsatellite mutation confounds population relationships. Optimum resolution of Y chromosome evolution will require the use of additional microsatellite loci and diallelic genetic markers with lower mutation rates.

Reconstruction of the evolutionary histories of human populations from genetic data is intrinsically complex because most contemporary populations have exchanged genes during prehistoric/historic migrations. Nevertheless, a variety of polymorphic systems have been used to study human evolutionary history, each of which has its own inherent shortcomings. Apparently identical regions of the X chromosome, autosomes, as well as the pseudoautosomal portion of the Y chromosome, have multiple ancestors because of recombination. Therefore, it is theoretically difficult to trace the migrations of single haplotypes. The mitochondrial (mtDNA) genome has a single maternal ancestor and its utility in the study of human evolution has been demonstrated (Cann et al. 1987; Stoneking 1993). However, the high mutation rate of mtDNA (Stoneking 1993) results in frequent recurrent mutations

that may obscure the migratory histories of populations. Highly polymorphic autosomal microsatellites have been used recently to reconstruct phylogenetic relationships of human populations (Bowcock et al. 1994; Deka et al. 1995; Jorde et al. 1995). These loci, however, are subject to recombination and high mutation rates (Weber and Wong 1993), and the convergent nature of their mutational processes (Shriver et al. 1993) make them less efficient for tracing human migration. The usefulness of the nonrecombining portion of the human Y chromosome in studies of human origins has been shown recently (Hammer 1995; Jobling and Tyler-Smith 1995; Pena et al. 1995; Whitfield et al. 1995; Santos et al. 1996; Underhill et al. 1996). However, only a limited number of polymorphic loci on the human Y chromosome have been identified and only a few systematic population genetic studies on Y-specific microsatellites have been reported (Roewer et al. 1996). In an attempt to evaluate the usefulness of the Y-specific microsat-

⁵Corresponding author.
E-MAIL rdeka@helix.hgen.pitt.edu; FAX (412) 624-3020.

DEKA ET AL.

ellites in detecting evidence of early migration during human evolution, we have analyzed five such loci in 15 ethnically and geographically diverse human populations.

RESULTS

Populations Share the Most Frequent Alleles at Y Microsatellites

Allele frequency distributions at five Y-specific microsatellite loci are given in Table 1, where the most frequent allele (MFA) for each population is indicated by boldface type. Fourteen out of the 15 populations share the 283-bp MFA at DYS391, the exception being Guyami who are monomorphic for the 287-bp allele. At DYS388, 12 populations share the 129-bp MFA. At DYS395, 11 populations have the 123-bp MFA. The MFAs at the other four populations at this locus are only one step away from the 123-bp allele. At DYS394 (also known as DYS19; Jobling and Tyler-Smith 1995), eight populations share the 251-bp MFA, and four have MFAs one step away from this allele. Interestingly, three South American Indian populations (Guyami, Bribri, and Pehuenche) display an extremely high frequency of the 243-bp allele at this locus. DYS390 displays the largest number of alleles. At this locus the MFA for 12 populations is either the 211-bp allele (four populations) or the 215-bp allele (eight populations), whereas the MFA in the two African populations (Benin and Sokoto) is the 203-bp allele, which is rare in all non-African populations.

Under a stepwise mutation model, the frequency distribution of alleles labeled by repeat size can drift away from the original distribution and may never reach a stationary distribution (Moran 1975; Shriver et al. 1993; Valdes et al. 1993). For isolated populations, the mode of the allele frequency distributions is expected to vary. At least two mechanisms can be responsible for the observation of common MFAs over several loci in different populations—common ancestry and/or migration among populations. In the first case, the observation of common MFAs across populations indicates that the differentiation among populations is not great and the common ancestral alleles are still the MFAs. On this basis, common haplotypes consisting of the MFAs at five loci should be CCCCH, CCCCG, CCCBH, CCCBG, and their derivatives. We observed a total of 125 haplotypes (see Fig. 1). With the exception of CCCCH, which was not observed in our

sample, the other three haplotypes are shared by multiple populations: CCCCG (J/M/N/V), CCCBH (D/G/O/V/E), CCCBG (E/G). Several one-step derivatives of these haplotypes (i.e., differ by only one repeat unit at one of the five loci) are also observed across racial and geographical boundaries, e.g., CCCCF (J/S), CCCCI (A/G/S), CCDCH (C/V), CDCCG (N/V/C), CDCCH (N/S/V/M/O), CCDCF (A/E/V), DCCBG (C/E/P), and DCCBH (E/G/P).

Haplotype Network

The convergent nature of mutations at microsatellite loci may obscure the true phylogenetic relationships of haplotypes. Two haplotypes (or sequences) can be identical by state but not by descent. Therefore, two identical haplotypes may not necessarily be derived from a recent common ancestral haplotype and a haplotype can mutate to another by multiple pathways. Although these haplotypes can be related as phylogenetic trees, given these concerns, a more reasonable way to display haplotype relationships is to construct a network of haplotypes. Such a network is shown in Figure 1. The hypothetical ancestral (common) haplotypes and their derivatives are indicated by dark circles that generally lie in the middle of the network and are surrounded by haplotypes observed in diverse populations. The network is characterized by a lack of clustering of haplotypes from the same population or geographically neighboring populations. Nevertheless, few distinctive clusters emerge that are ethnically specific. Several haplotypes from the Native American populations form a cluster in the upper right corner of the network. This may be consistent with the Native American-specific Y haplotype (Pena et al. 1995; Underhill et al. 1996). Some haplotypes observed in the Kacharis also cluster indicating an independently derived lineage. Haplotypes observed among the Samoans (shown at the lower left corner) form another distinct cluster. Although a few of the haplotypes observed among the Africans appear to be among the ancestral ones, the majority of the haplotypes found in Africa form a distinct cluster in the lower right corner of the network, along with a few haplotypes from Japan. This observation is consistent with African-specific Y haplotypes characterized by the YAP (Hammer 1995) and/or G allele in DYS271 (Seielstad et al. 1994).

The data presented can also be analyzed as simple frequency data using a measure of genetic

Y CHROMOSOME HAPLOTYPES IN HUMAN POPULATIONS

distance appropriate for microsatellite loci. For this analysis we used the stepwise weighted genetic distance D_{SW} , which takes into account the stepwise mutational mechanism of these loci (Shriver et al. 1995). When we use allele frequencies, the information contained in the haplotypes attributable to associations among loci is not used, and, therefore, the additional variance attributable to repeated creation of particular haplotypes is not present. We can, therefore, expect a simplified picture of the relationships of Y chromosomes among populations. The D_{SW} tree created using the Neighbor-joining method (Saitou and Nei 1987) is shown in Figure 2. Because other genetic data have established the root of human phylogenetic trees in the branch connecting African and all non-African populations (see Nei and Takezaki 1996), we have placed the root between Africans and all other populations. Asian populations form a monophyletic group in this tree within which another cluster is found that contains the American Indians (Dogrib, Guyami, Bribri, and Pehuenche). The German and the CEPH lie outside the Asian group and the Samoan and African populations outside of these branchings. Although the longer branches for the Amerindian populations could possibly be influenced by relatively smaller samples of the individual populations, most likely they reflect founder effects and genetic drift in these populations, which is supported by evidence from other studies (e.g., Underhill et al. 1996).

DISCUSSION

We have used a network of Y-specific microsatellite haplotypes in samples from 15 populations to search for insights into human population history. The lack of clustering of haplotypes among ethnically and geographically proximal populations may result from several factors. First, a larger number of loci may be needed to resolve population clusters given the recent divergence of human populations. Second, past male migration among human populations and/or racial groups may have been sufficient to erase any distinctive haplotype clustering within the same populations or between geographically neighboring groups. Third, the convergent nature of the mutational process and high mutation rate associated with microsatellites may have randomized the haplotypic structures. The first problem can be resolved by typing additional microsatellite loci. The second factor is directly related to the

evolutionary histories of human populations. The third factor is an inherent characteristic of the loci in question. Therefore, the utility of Y-specific microsatellites is influenced by stochastic noise introduced by recurrent mutations at microsatellite loci.

A population-specific mutation is distinctive only when the mutant allele is either absent or at low frequency in other populations. However, the allele frequency distributions at the five microsatellite loci studied here are, in general, unimodal in all populations, with the mode centered around the same common allele. However, the few distinct clusters observed in our network are associated with the presence of population-specific alleles. (1) The American Indian cluster is characterized by high frequency of the smallest allele (243 bp) at DYS394. This supports the reported geographical localization of unique Y-specific polymorphisms in the American Indians (Pena et al. 1995; Underhill et al. 1996). (2) Similarly, the haplotypes found in the Kachari cluster are characterized by the presence of the 123-bp allele at DYS388, which is rare in other populations except the Chinese. This reflects the known Mongoloid lineage of the Kachari. However, their divergence is reflected in the fact that haplotypes in the Kachari are more than one mutational step away from any of the presumed ancestral haplotypes. (3) The Samoans present a very different picture—at three of the five loci, their MFAs are not the MFAs observed in any other population. Eleven of their haplotypes out of 20 form a contiguous link at the lower left of the network. Interestingly, these haplotypes are connected by a broken line to a New Guinea haplotype, which is connected to haplotypes observed in other Asian populations, for example, Javanese, Malay, and Chinese. Six other Samoan haplotypes are also shared by these Asian populations. This pattern is consistent with the proposed migration history and frequent population bottlenecks, which are thought to characterize the settlement of the Pacific. (4) The haplotypes found in Japan are very heterogeneous. Some of them cluster with Africans, which is interesting given reports of the African YAP insertion in a previous Japanese sample (Hammer and Horai 1995). A few others cluster with Javanese, Malay, and New Guinean. (5) Most haplotypes found in Africa form a distinct cluster characterized by the MFA (203 bp) at DYS390, which is rare or absent in other populations. Together with this, the association of the rest with the ancestral haplotypes or their close

DEKA ET AL.

Table 1. Allele Frequency Distribution of Five Microsatellites in Global Populations

Locus Allele	Size (bp)	Populations ^a																
		B	S	G	E	J	C	A	M	V	D	R	I	P	N	O		
DYS391		26 ^b	17	22	20	20	20	27	14	20	15	8	7	16	19	20		
A	275					.05												
B	279		.06			.05			.29	.10	.47			.13		.05		
C	283	.85	.59	.82	.65	.75	.80	.85	.57	.80	.53	.86	.75	.79	.85	.85		
D	287	.15	.35	.18	.30	.15	.15	.11	.14	.10		.14	.13	.21	.10			
E	291							.04										
F	295				.05													
DYS388		26	17	22	20	19	19	27	11	17	17	8	7	16	18	20		
A	123		.06		.05		.32		.67									
B	126		.06															
C	129	1.00	.76	.64	.60	.74	.58	.33	.64	.59	.53	1.00	.71	.44	.50	.10		
D	132		.06	.27	.15	.26	.11		.18	.35	.47	.29	.56	.39	.35	.35		
E	135		.06	.09	.05				.09	.06				.11	.05	.05		
F	138				.10													
G	141				.05											.50		
DYS395		26	17	22	20	20	20	27	13	20	15	8	7	16	20	20		
A	115							.11										
B	119			.18	.15	.25		.55	.31	.20	.47	.75		.05				
C	123	.46	.65	.59	.50	.55	.20	.15	.69	.55	.47	1.00	.81	.70	.45			
D	127	.31	.24	.09	.30	.20	.20	.19	.10	.10	.19	.25	.19	.25	.25			
E	131	.23	.12	.14	.05	.05	.05		.15	.07	.07							

Y CHROMOSOME HAPLOTYPES IN HUMAN POPULATIONS

Table 1. (Continued)

Locus Allele	Size (bp)	Populations ^a																		
		B	S	G	E	J	C	A	M	V	D	R	I	P	N	O				
DYS394		26	17	22	20	20	20	20	27	14	16	15	8	7	16	17	20			
A	243		.06		.05	.15	.05	.05	.05		.19	.40	.88	1.00	.88					
B	247	.12	.06	.45	.50	.05	.30	.59		.19	.13	.13			.13	.18	.10			
C	251	.42	.53	.27	.35	.45	.45	.26		.56	.47					.76	.40			
D	255	.35	.24	.23	.05	.15	.15	.11		.25						.06	.50			
E	259	.12	.12	.05	.05	.20	.05	.04												
DYS390		26	17	22	20	19	20	27	13	17	15	8	7	16	18	20				
A	187															.05				
C	195															.10				
D	199	.04			.05											.35				
E	203	.92	.53		.05											.05				
F	207	.04	.06	.18	.20	.32		.07		.12						.05				
G	211		.06	.23	.15	.11	.45	.63	.32	.29					.25	.44				
H	215		.18	.32	.45	.32	.30	.07	.38	.35	.80	.75	.57	.19	.22	.35				
I	219		.18	.23	.15	.21	.25	.22	.31	.24		.25	.14	.56	.06	.05				
J	223																			
K	227					.05									.28					

^aPopulation abbreviations used in Table 1 and Fig. 1: (B) Benin; (S) Sokoto; (G) German; (E) CEPH parents; (I) Japanese; (C) Chinese; (A) Kachari; (M) Malay; (V) Javanese; (D) Dogrib; (R) Guyami; (I) Bribri; (P) Pehuenche Indian; (N) New Guinea Highlander; (O) Samoan.

^bNumber of individuals examined at each locus in each population.

DEKA ET AL.

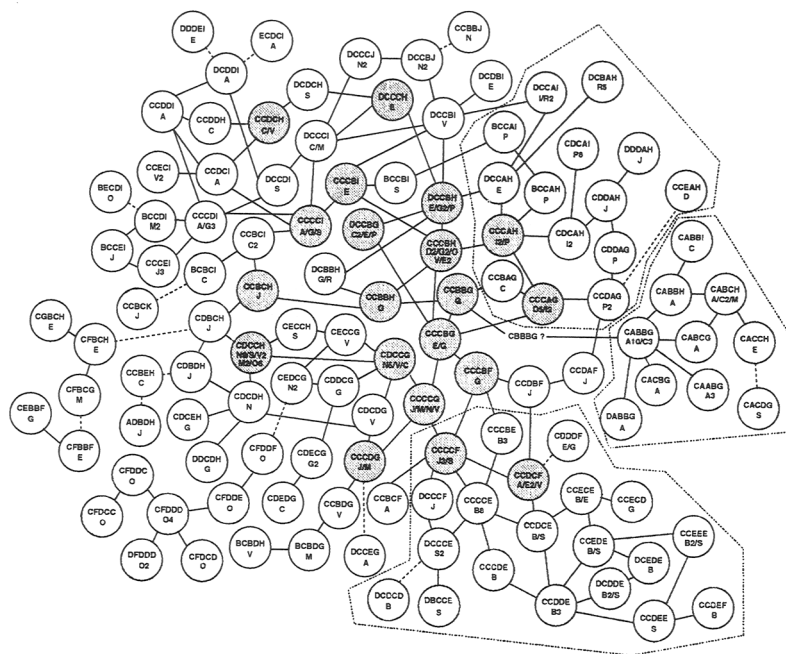


Figure 1 Network of haplotypes where each circle represents one haplotype. Solid lines connect two haplotypes differing by a single repeat unit at one of the five microsatellite loci. Broken lines connect two haplotypes differing by more than one step change. The haplotypes are designated by five letters in a circle, representing alleles at each locus in the order shown in Table 1. Population abbreviations (see Table 1) are given under each observed haplotype.

derivatives may be suggestive of uniqueness of the Africans. (6) The haplotypes found in European populations are distributed over the entire network, and their frequent appearance with the ancestral haplotypes is striking.

The ancestral haplotypes, located in the center, consist of individuals from almost all ethnic groups. This suggests that male migration has been a common feature of human evolution. The lack of haplotypic structure in the network is probably caused by a high level of male migration and frequent recurrent mutation at microsatellite loci. Notably, this observation is also consistent with a recent common ancestry of modern human populations.

The D_{SW} tree shows some of the same trends seen in the network. There is a very clear clustering of populations with Asian ancestry. In particular, all the Amerindian populations cluster together, which reflects the high frequency of the 243-bp allele at *DYS394*. The two African populations also cluster and are separated by a relatively long branch from most populations. In addition, as seen in the haplotype network, most

populations are not very distinct, owing to the small number of loci typed, the recurrent nature of microsatellite evolution, and extensive admixture before or after the geographical separation of the ancestors of contemporary human populations. As with other genetic systems used to infer human population history, recurrent mutation appears to be a major confounding factor. Optimum resolution of Y chromosome evolution will require the use of additional microsatellite loci and diallelic restriction fragment length polymorphisms (RFLPs), with presumably lower mutation rates.

METHODS

Population Samples

The 15 human populations studied belong to five major groups, namely African, Caucasian, Mongoloid, American Indian, and Pacific Island. The Africans are represented by two linguistically diverse populations from Nigeria, the Benin (B) and Sokoto (S). The Caucasians are represented by German (G) and unrelated parents of the CEPH cohort (E). The Mongoloid populations are Japanese (J), Chinese (C), Kachari from Northeast India (A), Malay (M), and Javanese (V). The American Indian populations consist of Dogrib (D) from Canada, Guyami (R) and Bribri (I) from Costa Rica, and Pehuenche (P) from Chile. The Pacific Island populations are New Guinea Highlander (N) and Samoan (O). Detailed description of these populations are presented elsewhere (Deka et al.

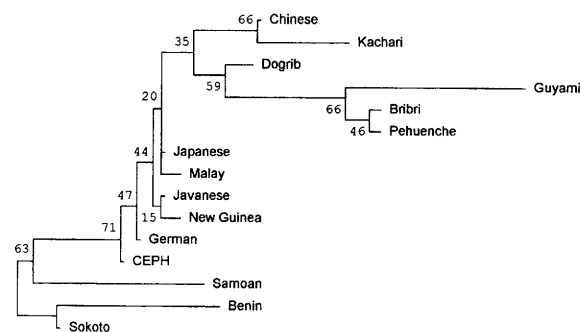


Figure 2 A neighbor-joining tree based on stepwise-weighted genetic distance D_{SW} . Bootstrap values, indicating the degree of support for each branch point, are shown as the percent of all replicates consistent with each branch point.

Y CHROMOSOME HAPLOTYPES IN HUMAN POPULATIONS

1996). The DNA samples analyzed are drawn from unrelated males.

DNA Analysis

Locus and primer descriptions are given in Jobling and Tyler-Smith (1995). PCR analysis of microsatellites was performed by amplification of 50 ng of genomic DNA in a 25- μ l reaction mixture containing standard PCR buffer, 200 μ M each dNTP, and 1 unit of *Taq* polymerase. The forward primer was end-labeled using [γ -³²P]ATP. The amplified products were separated on 6% denaturing polyacrylamide gels. Allelic fragments were visualized by autoradiography and fragment sizes were determined by comparison with an M13 sequence ladder.

ACKNOWLEDGMENTS

We thank Drs. E.J.E. Szathmary, S.T. McGarvey, P.E. Smouse, J.C. Long, G. Flatz, F. Rothhammer, S.-J. Yin, T. Miki, and C.H. Bunker for providing DNA samples from various populations. This work was supported by grants from the National Institutes of Health (NIH) (GM45861, GM41399, AG09375), grant SBR-9600910 from the National Science Foundation, and in part by a grant from the W.M. Keck Foundation for Training in Computational Biology. L.J. was supported by NIH Training Grant T32-GS08404.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Bowcock, A.M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J.R. Kidd, and L.L. Cavalli-Sforza. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* **368**: 455–457.

Cann, R.L., M. Stoneking, and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* **325**: 31–36.

Deka, R., L. Jin, M.D. Shriver, L.M. Yu, S. DeCroo, J. Hundrieser, C.H. Bunker, R.E. Ferrell, and R. Chakraborty. 1995. Population genetics of dinucleotide (dC-dA)_n-(dG-dT)_n polymorphisms in world populations. *Am. J. Hum. Genet.* **56**: 461–474.

Deka, R., P.P. Majumder, M.D. Shriver, D.N. Stivers, Y. Zhong, L.M. Yu, R. Barrantes, S.J. Yin, T. Miki, J. Hundrieser, C.H. Bunker, S.T. McGarvey, S. Sakallah, R.E. Ferrell, and R. Chakraborty. 1996. Distribution and evolution of CTG repeats at the myotonin protein kinase gene in human populations. *Genome Res.* **6**: 142–154.

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

Hammer, M.F. and S. Horai. 1995. Y chromosomal DNA

variation and the peopling of Japan. *Am. J. Hum. Genet.* **56**: 951–962.

Jobling, M.A. and C. Tyler-Smith. 1995. Fathers and sons: Y chromosome and human evolution. *Trends Genet.* **11**: 449–456.

Jorde, L.B., M.J. Bamshad, W.S. Watkins, R. Zenger, A.E. Farley, 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.

Moran, P.A.P. 1975. Wandering distributions and the electrophoretic profile. *Theor. Popul. Biol.* **8**: 318–330.

Nei, M. and N. Takezaki. 1996. The root of the phylogenetic tree of human populations. *Mol. Biol. Evol.* **13**: 170–177.

Pena, S.D.J., F.R. Santos, N.O. Bianchi, C.M. Bravi, C.M. Carnese, F. Rothhammer, T. Gerelsaikhan, B. Munkhtuja, and T. Oyunsuren. 1995. A major founder Y-chromosome haplotype in Amerindians. *Nature Genet.* **11**: 15–16.

Roewer, L., M. Kayser, P. Dieltjes, M. Nagy, E. Bakker, M. Krawczak, and P. deKnijff. 1996. Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. *Hum. Mol. Genet.* **7**: 1029–1033.

Saitou, N. and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**: 406–425.

Santos, F.R., N.O. Bianchi, and S.D.J. Pena. 1996. Worldwide distribution of human Y-chromosome haplotypes. *Genome Res.* **6**: 601–611.

Seielstad, M.T., J.M. Hebert, A.A. Lin, P.A. Underhill, M. Ibrahim, D. Vollarth, and L.L. Cavalli-Sforza. 1994. Construction of human Y-chromosomal haplotypes using a new polymorphism A to G transition. *Hum. Mol. Genet.* **3**: 2159–2161.

Shriver, M.D., L. Jin, R. Chakraborty, and E. Boerwinkle. 1993. VNTR allele frequency distributions under the stepwise mutation model: A computer simulation approach. *Genetics* **134**: 983–993.

Shriver, M.D., L. Jin, E. Boerwinkle, R. Deka, R.E. Ferrell, and R. Chakraborty. 1995. A novel measure of genetic distance for highly polymorphic tandem repeat loci. *Mol. Biol. Evol.* **12**: 914–920.

Stoneking, M. 1993. DNA and recent human evolution. *Evol. Anthropol.* **2**: 60–73.

Underhill, P.A., L. Jin, R. Zemans, P.J. Oefner, and L.L. Cavalli-Sforza. 1996. A pre-Columbian Y chromosome-specific transition and its implications for human evolutionary history. *Proc. Natl. Acad. Sci.* **93**: 196–200.

DEKA ET AL.

Valdes, A.M., M. Slatkin, and N.B. Freimer. 1993. Allele frequencies at microsatellite loci: The stepwise mutation model revisited. *Genetics* **133**: 737–749.

Weber, J.L. and C. Wong. 1993. Mutation of human short tandem repeats. *Hum. Mol. Genet.* **2**: 1123–1128.

Whitfield, L.M., J.E. Sulston, and P.N. Goodfellow. 1995. Sequence variation of the human Y chromosome. *Nature* **378**: 379–380.

Received July 29, 1996; accepted in revised form September 18, 1996.