Biol. Chem., Vol. 378, pp. 923 – 927, August 1997 · Copyright © by Walter de Gruyter & Co · Berlin · New York

Short Communication

Multiplex DNA Typing of Short-Tandem-Repeat Loci on the Y Chromosome

Alan J. Redd*, Stephanie L. Clifford and Mark Stoneking

Department of Anthropology, The Pennsylvania State University, University Park, PA 16802, USA

* Corresponding author

To facilitate evolutionary and forensic studies of DNA polymorphisms on the Y chromosome, we devised a multiplex amplification procedure for short-tandem-repeat (STR) loci. Four tetranucleotide STR loci (DYS19, DYS390, DYS391, and DYS393) were simultaneously amplified with FAM-labeled primers and genotypes were determined with an automated DNA sequencer. We typed 162 males from three U.S. populations (African-Americans, European-Americans and Hispanics) and found that the haplotype diversities range from 0.920 to 0.969. This quadruplex system provides a facile means of genotyping these Y chromosome STRs, and should be useful in population genetic and forensic applications.

Key words: Humans / PCR / STR polymorphism.

DNA polymorphisms on the human Y chromosome are potentially very useful tools for evolutionary studies and human identification. The unilineal inheritance of the non-recombining portion of the Y chromosome results in the preservation of linked polymorphisms that can be used to trace the evolution of paternal lineages. Y chromosome polymorphisms will be valuable in a variety of forensic situations, such as rape cases where a mixture of male and female DNA is present, and paternity testing.

However, many researchers have noted a paucity of common classes of Y chromosome polymorphisms (Jakubiczka *et al.*, 1989; Malaspina *et al.*, 1990) including DNA sequence substitutions (Ellis *et al.*, 1990; Dorit *et al.*, 1995; Whitfield *et al.*, 1995). Hypotheses for explaining this reduction in genetic diversity include: selective sweeps; the smaller effective size of the Y chromosome gene pool; and mating patterns involving few successful males. Much of this apparent absence of Y chromosome polymorphism relative to autosomal polymorphism may be directly related to the reduced chromosomal effective population size, since newly characterized Y chromosome STR loci are nearly as variable as autosomal STR loci (Roewer *et al.*, 1992; Goldstein *et al.*, 1996; Underhill *et al.*, 1996). Y chromosome STR markers are gaining currency in both population (Mathias *et al.*, 1994; Cooper *et al.*, 1996; Jobling *et al.*, 1996; Roewer *et al.*, 1996; Deka *et al.*, 1996; Hammer *et al.*, 1997; De Knijff *et al.*, 1997) and forensic studies (Roewer *et al.*, 1992; Santos *et al.*, 1993; Kayser *et al.*, 1997).

Two recently published papers (Kayser *et al.*, 1997; Prinz *et al.*, 1997) include strategies for multiplexing Y chromosome STRs using an Applied Biosystems DNA

Table 1 Y Chromosome STR Loci and Their Summary Statistics.

Locus	DYS393	DYS390	DYS19 (DYS394)[a]	DYS391
Locus accession ID[b]	456649	366115	456738	366118
Number of alleles	6	8	6	6[c]
Number of repeats	12–17	20–28	12–17	9–13
Allele size range (bp)	120–140	199–231	238–258	279–295
Gene diversity[d]	0.48	0.78	0.67	0.53
Variance in repeat size	1.98	6.80	2.55	1.51
Allele sizing[e]:				
precision (%)	99.85	99.94	99.96	99.97
range (bp)	0–0.8	0.01–0.41	0–0.48	0.02–0.36
S.D. (bp)	0.18	0.12	0.09	0.09

[a] DYS394 primers detect variation at the DYS19 locus, see Jobling and Tyler-Smith (1995).
[b] Genome Database (http://gdbwww.gdb.org/).
[c] Number includes one triplication allele.
[d] Nei (1987), formulae 8.5 and 8.6.
[e] Kimpton et al. (1993).

sequencer (model 373A). In this paper, we describe another strategy using a unique combination of loci and PCR conditions. We also present genotyping results from 162 males from three U.S. populations (56 African-Americans, 56 Hispanics and 50 European-Americans). Characteristics of the STR loci (DYS19, DYS390, DYS391, and DYS393) used in this multiplex and summary statistics from our genotype data are given in Table 1.

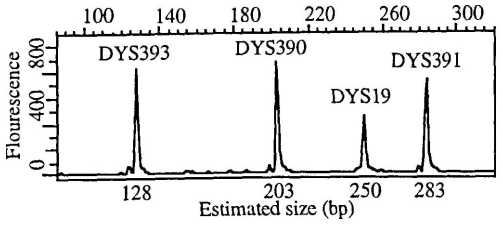

Fig. 1 Electrophoretogram Generated (Using Genescan 672 v 1.2.2-1) from Male DNA by Multiplex Amplification of Four STR Loci.

The amplification products of a single lane from a 6.5% denaturing polyacrylamide gel are depicted. The peaks indicate fluorescent intensities of dye-labeled PCR products from four loci: DYS393, DYS390, DYS19 and DYS391. Note that the peaks of the internal lane standard are not shown. Estimated size in base pairs is given along the x axis, and relative fluorescence is given along the y axis.

Methods: DNA Isolation: DNA was extracted from blood samples from unrelated individuals of African-American, Hispanic and European-American descent using an IsoQuick kit (MicroProbe) as described by Buchanan *et al.* (1993). The DNA samples were sexed with a PCR test of the amelogenin gene (Sullivan *et al.*, 1993).

PCR: Reactions were set up on ice using 5–10 ng of genomic DNA in a 12.5 μl reaction volume. The multiplex reaction included: TNK50 buffer [10 mM Tris-HCl, pH 8.3; 5 mM NH_4Cl; 50 mM KCl; 1.5 mM $MgCl_2$ (Blanchard *et al.*, 1993)], 1.0 mg/ml of Bovine Serum Albumin, 0.25 units of AmpliTaq® DNA polymerase (Perkin Elmer), 200 μM of each dNTP; and 0.32 μM DYS394 Forward (F) and Reverse (R) primers, 0.16 μM DYS390 F and R, 0.16 μM DYS391 F and R, and 0.1 μM DYS393 F and R. All primers were obtained from Perkin Elmer Applied Biosystems (PE-Applied Biosystems) and the forward primer in each set was 5′ end-labeled with the fluorescent dye FAM. PCR cycling conditions were 94 °C for 30 seconds, 59 °C for 30 seconds and 72 °C for 30 seconds for 28 cycles in a Perkin-Elmer 9600 Thermal Cycler.

Gel electrophoresis: A 1.5 μl aliquot of each multiplexed sample was combined with 4 fmol (0.5 μl) of GS-350 standard labeled with the fluorescent dye ROX (PE-Applied Biosystems), 2.5 μl of deionized formamide, and 0.5 μl of loading buffer [50 mg/ml blue dextran; 25 mM EDTA (PE-Applied Biosystems)]. The samples were then denatured for 2 minutes at 90 °C, and 3 μl loaded on a 6.5% denaturing polyacrylamide gel [7.5 M urea, 1 X TBE (0.009 M Tris, 0.089 M Boric Acid, and 0.002 M Na_2EDTA)]. Electrophoresis conditions were: 2000 V, 50 mA, and 35 W constant power using an Applied Biosystems DNA sequencer (model 373A) and 12 cm well-to-read (WTR) plates with 0.4 mm spacers.

Applied Biosystems 373A analysis: The data were analyzed with Genecan 672 software [(v. 1.2.2-1) PE-Applied Biosystems]. Allele size was estimated using the Local Southern Method. Alleles were processed with Genotyper software [(v.1.1) PE-Applied Biosystems].

PCR conditions of the multiplex were sought to maximize signal intensities and to minimize extraneous products. Multiplex conditions were optimized (see Figure 1 and the Methods section) by examining the effects of ion concentrations and temperature. The $MgCl_2$ concentration of the multiplex reaction was varied between 0.25 mM and 2.5 mM in steps of 0.25 mM. The optimal $MgCl_2$ concentration, in terms of amplification efficiency, for DYS390 and DYS391 was 0.75 mM, while the optimal concentration for DYS393 and DYS19 was 1.5 mM. Ion concentrations of the multiplex reaction were further examined with the buffer series of Blanchard *et al.* (1993), a PCR buffer system that is based on altering concentrations of potassium (KCl), thereby allowing the use of a single PCR profile and annealing temperature with many primer pairs. The amplification efficiency for DYS19, DYS390, and DYS391 was most efficient with 50 mM KCl, while the amplification efficiency for DYS393 was highest with 100 mM KCl. The theoretical annealing temperatures (Wu *et al.*, 1991) for all of the primers is approximately 66 °C, but annealing temperatures were varied between 56 °C and 62 °C in increments of one degree centigrade. An annealing temperature of 59 °C was sufficient to reduce spurious bands to background levels but not too stringent to compromise amplification efficiency. Primer concentrations were initially set at 0.2 μM for each locus and then adjusted empirically to balance the amplification efficiency across all loci.

We tested the multiplex using DNA concentrations of 20 ng, 10 ng, 5 ng, 1 ng, and 0.5 ng. The multiplex successfully amplified all four loci for each DNA concentration, but DNA concentrations of 5–10 ng seemed to be the most robust (data not shown). Furthermore, we used this multiplex to type over 450 individuals; the number of failures on the first attempt for one or more loci was less than 4% (data not shown).

It should be noted that the primers for DYS391 occasionally produced, in males, one non-specific product of approximately 256, 260, or 264 base pairs (bp). In addition, one or two bands in the same range were produced with female DNA, as has been found independently by Kayser *et al.* (1997). These artifact bands may represent an autosomal or X counterpart to this locus. Nevertheless, the fluorescent intensities of these spurious DYS391 bands were not high enough to hinder reliable genotyping with our multiplex.

The DYS394 primer sequences, rather than the DYS19 primers sequences, were used to detect variation at the DYS19 locus (see Jobling and Tyler-Smith, 1995), because the DYS394 primers have a higher annealing temperature. Also, these primers give a longer PCR product, which fits better with the multiplex. Since the variation detected with the DYS394 primers is at the DYS19 locus, we refer to this locus as DYS19. Alleles are designated on the basis of the number of repeats, which was determined by comparison with known standards (provided by M. Kayser).

None of the allele size ranges between the loci overlapped in our typings; however, there is potential overlap

between DYS390 and DYS19 (DYS394) based on a comparison between our typings and a worldwide compilation of the alleles observed at these loci (Kayser *et al.*, 1997). We found one Hispanic individual with allele 28 (231 bp) at locus DYS390 (allele size range = 191–231 bp), and Kayser *et al.* (1997) found one Romanian individual with allele 10 (230 bp) at locus DYS19 (allele size range = 230–266 bp). Given the possibility of allele size overlap at the extreme upper range of DYS390 and the extreme lower range of DYS19, and the fact that we used a single fluorescent-dye color (FAM) across loci, we typed by single locus PCR two Hispanic individuals, one with allele 28 (231 bp) at DYS390 and another with allele 12 (238 bp) at DYS19. The single locus PCR confirmed the allele assignments from the multiplex. Thus, it appears that there is only a remote possibility of overlap between DYS19 and DYS390 with our multiplex; furthermore, the locus assignment in any questionable instance can be verified by single locus PCR. Alternatively, this remote risk could be eliminated by labeling DYS19 or DYS390 with another dye marker since the 373A DNA sequencer is designed to run overlapping allele-size-ranges using different dyes.

To estimate the precision of the band sizing (Kimpton *et al.*, 1993), the multiplex PCR products from 30 individuals were electrophoresed and genotyped twice (Table 1). The allele sizing appears to be increasingly precise as the size of the PCR product increases, as noted by Ghosh *et al.* (1997). The values in Table 1 indicate that the allele sizing by this procedure is precise and highly reproducible.

Figure 2 shows the observed distributions of allele frequencies for the four loci for each population. The distributions are identical with respect to mode allele length,

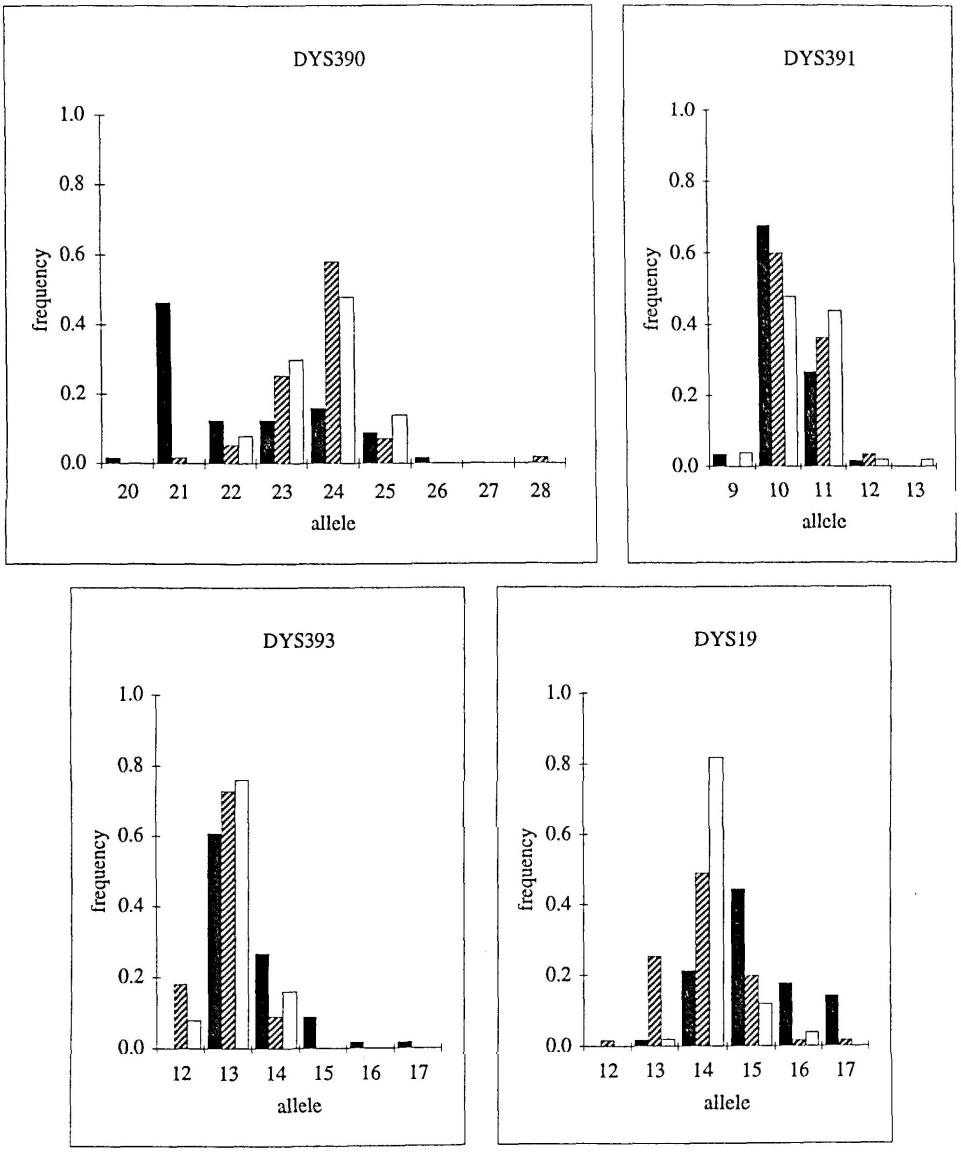

Fig. 2 Histograms of the Observed Allele Frequencies for DYS390, DYS391, DYS393 and DYS19 in Three Populations.
The ascending allele numbers indicate the number of repeats as inferred by comparison to DNA controls (provided by M. Kayser). Solid, hatched and open bars indicate African-American (N = 56), Hispanic (N = 55 – one individual with a DYS390 triplication allele was not included) and European-American (N = 50) populations, respectively.

and similar with respect to allele size range, to previous surveys in African, European and Hispanic populations (Cooper *et al.*, 1996; Deka *et al.*, 1996; Roewer *et al.*, 1996; Santos *et al.*, 1996; De Knijff *et al.*, 1997). For example, the modal allele length at DYS390 in the African-Americans is 21 repeats (46%), while the modal allele length is 24 repeats in the European-American (48%) and in the Hispanic (58%) populations. Likewise at DYS19, the most common allele in African-Americans is 15 repeats (45%), while the European-American (67%) and Hispanic (49%) populations have allele 14 in high frequency. The modal allele lengths are the same across the three populations for both DYS393 (allele 13) and DYS391 (allele 10). We found four new putative alleles: allele 16 (136 bp) and allele 17 (140 bp) at DYS393 in African-American individuals, allele 28 (231 bp) at DYS390 in a Hispanic individual, and a triplication allele at DYS390 in a Hispanic individual with alleles 23 (211 bp), 25 (219 bp) and 26 (223 bp). Biallelic patterns have been reported for DYS19 and DYS390 (Kayser *et al.*, 1997), while triplication alleles have been reported for DYS19 (Santos *et al.*, 1996) and DYS385 (Kayser *et al.*, 1997). Each of these putative new alleles will be sequenced to confirm the genotype results.

Table 1 gives the number of alleles, the allele size ranges, and various statistics for each of the four loci. The gene diversity values [or the probability that two alleles, chosen at random, are different (Nei 1987)] for the four loci are: 0.67, 0.78, 0.53, and 0.48 for DYS19, DYS390, DYS391, and DYS393 respectively. We also measure variance in repeat size (Jorde *et al.*, 1997), a genetic diversity measure that can be related to the single-step-mutation (SSM) model (Slatkin, 1995) of STRs. Under the SSM model the expected variance at equilibrium is $(N\text{-}1)\mu$, where N is the effective number of gametes, and μ is the mutation rate per generation (Goldstein *et al.*, 1996). The variance in repeat size is largest for DYS390, intermediate for DYS19 and relatively low for DYS393 and DYS391. The higher variances of DYS390 and DYS19 compared to the variances of DYS393 and DYS391 suggest that the mutation rates are higher for the former two loci. Given that Y chromosomes have a four-fold lower effective population size compared to autosomes, our finding that the average variance of the loci in this study is 3.6 is compatible with expectations; a sample of 30 autosomal STRs have an average variance of approximately 10 (Goldstein *et al.*, 1996).

Table 2 summarizes the haplotype variation detected within the three populations. For the four loci in the multiplex, 73 different haplotypes were detected among 162 individuals. Fifteen haplotypes were shared between populations: four haplotypes were shared across all three populations, while four more were shared by African-Americans and Hispanics only, five by African-Americans and European-Americans only, and two haplotypes by Hispanics and European-Americans only. More unique haplotypes (26) were detected in the African-American sample than in the Hispanic (18) and European-American (14) samples. In addition, African-Americans have the highest haplotype diversity, followed by Hispanics, and then European-Americans. The haplotype diversity values range from 0.920 to 0.969, suggesting that these linked loci are useful markers.

This quadruplex genotyping system is simple, reliable, and it provides PCR products from very low quantities of target DNA. Given the growing interest in polymorphisms on the Y chromosome and the high levels of STR polymorphism, this multiplex will be useful for evolutionary studies of human populations and for forensic applications.

Acknowledgements

We thank Kathy Lazaruk for technical support and Perkin Elmer Applied Biosystems for providing the fluorescent primers. We appreciate Manfred Kayser for supplying DNA standards. Mike Hammer provided useful advice and some independent verification of relative allele sizing. We acknowledge the technical assistance of Richard Thomas and MacGregor Fraser. We thank Henry Harpending, David Stock, Ken Weiss and two anonymous reviewers for their comments and suggestions. This research was supported by National Science Foundation SBR-9423118 (MS) and National Science Foundation Dissertation Improvement Award SBR-9617272 (AR).

References

Blanchard, M.M., Taillon-Miller, P., Nowotny, P., and Nowotny, V. (1993). PCR buffer optimization with uniform temperature regimen to facilitate automation. PCR Methods Applic. *2*, 234–240.

Buchanan, A.V., Sherry, S.T., Weiss, K.M., McGarvey, S.T., Neel, J.V., and Stoneking, M. (1993). Extraction of DNA from frozen red blood cells. Hum. Biol. *65*, 647–654.

Cooper, G., Amos, W., Hoffman, D., and Rubinsztein, D.C. (1996). Network analysis of human Y microsatellite haplotypes. Hum. Mol. Genet. *5*, 1759–1766.

De Knijff, P., Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B.,

Table 2 Summary of Haplotype Variation within Populations.

Population	n	Number of haplotypes	Numer of unique haplotypes	Haplotype diversity[a]
African-American	56	39	26	0.969 ± 0.009
Hispanic	56	28	18	0.944 ± 0.012
European-American	50	25	14	0.920 ± 0.016

[a] Melton et al. (1995) formulae 1 and 2.

Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Piccinini, A., Perez-Lezaun, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G., Roewer, L. (1997). Chromosome Y microsatellites: population genetic and evolutionary aspects. Int. J. Leg. Med. *110*, 134–149.

Deka, R., Jin, L., Shriver, M.D., Yu, L.M., Saha, N., Barrantes, R., Chakraborty, R., and Ferrell, R.E. (1996). Dispersion of human Y chromosome haplotypes based on five microsatellites in global populations. Genome Research *6*, 1177–1184.

Dorit, R.L., Akashi, H., and Gilbert, W. (1995). Absence of polymorphism a the ZYF locus on the human Y chromosome. Science *268*, 1183–1185.

Ellis, N., Taylor, A., Bengtsson, B.O., Kidd, J., Rogers, J., and Goodfellow, P. (1990). Population structure of the human pseudoautosomal boundary. Nature *344*, 663–665.

Goldstein, D.B., Zhivotovsky, L.A., Nayar, K., Linares, A.R., Cavalli-Sforza, L.L., and Feldman, M.W. (1996). Statistical properties of the variation at linked microsatellite loci: implications for the history of human Y chromosomes. Mol. Biol. Evol. *13*, 1213–1218.

Ghosh, S., Karanjawala, Z.E., Hauser, E.R., Ally, D., Knapp, J.I., Rayman, J.B., Musick, A., Tannenbaum, J., Te, C., Shapiro, S., Eldridge, W., Musick, T., Martin, C., Smith, J.R., Carpten, J.D., Brownstein, M.J., Powell, J.I., Whiten, R., Chines, P., Nylund, S.J., Magnuson, V.L., Boehnke, M., Collins, F.S., and the FUSION (Finland-U.S. Investigation of NIDDM Genetics) Study Group (1997). Methods for precise sizing, automated binning of alleles, and reduction of error rates in large-scale genotyping using fluorescently labeled dinucleotide markers. Genome Research *7*, 165–178.

Hammer, M.F., Spurdle, A.B., Karafet, T., Bonner, M.R., Wood, E.T., Novelletto, A., Malaspina, P., Mitchell, R.J., Horai, S., Jenkins, T., and Zegura, S.L. (1997). The geographic distribution of human Y chromosome variation. Genetics *145*, 787–805.

Jakubiczka, S., Arnemann, J., Cooke, H.J., Krawczak, M., and Schmidtke, J. (1989). A search for restriction fragment length polymorphism on the human Y chromosome. Hum. Genet. *84*, 86–88.

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

Jobling, M.A., Samara, V., Pandya, A., Fretwell, N., Bernasconi, B., Mitchell, R.J., Gerelsaikhan, T., Dashnyam, B., Sajantila, A., Salo, P.J., Nakahori, Y., Disteche, C.M., Thangaraj, K., Singh, L., Crawford, M.H., Tyler-Smith, C. (1996). Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Hum. Mol. Genet. *5*, 1767–1775.

Jorde, L.B., Rogers, A.R., Bamshad, M., Watkins, W.S., Krakowiak, P., Sung, S., Kere, J., and Harpending, H.C. (1997). Microsatellite diversity and the demographic history of modern humans. Proc. Natl. Acad. Sci. USA *94*, 3100–3103.

Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., Heidorn, F., Herrmann, S., Herzog, B., Hidding, M., Honda, K., Jobling, M., Krawczak, M., Leim, K., Meuser, S., Meyer, E., Oesterreich, W., Pandya, A., Parson, W., Piccinini, A., Perez-Lezaun, A., Prinz, M., Schmitt, C., Schneider, P.M., Szibor, R., Teifel-Greding, J., Weichhold, G., De Knijff, P., Roewer, L. (1997). Evaluation of Y chromosomal STRs: A Multicenter Study. Int. J. Leg. Med. *110*, 125–133.

Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.S., and Adams, M. (1993). Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Applic. *3*, 13–22.

Malaspina, P., Persichetti, F., Novelletto, A., Iodice, C., Terrenato, L., Wolfe, J., Ferraro, M., and Prantera, G. (1990). The human Y chromosome shows a low level of DNA polymorphism. Ann. Hum. Genet. *54*, 297–305.

Mathias, N., Bayés, M., and Tyler-Smith, C. (1994). Highly informative compound haplotypes for the human Y chromosome. Hum. Mol. Genet. *3*, 115–123.

Melton, T., Peterson, R., Redd, A.J., Saha, N., Sofro, A.S.M., Martinson, J., and Stoneking, M. (1995). Polynesian genetic affinities with Southeast Asian populations as identified by mtDNA analysis. Am. J. Hum. Genet. *57*, 403–414.

Nei, M. (1987). Molecular Evolutionary Genetics. (New York: Columbia University Press).

Prinz, M., Boll, K., Baum, H., Shaler, B. (1997). Multiplexing of Y chromosome specific STRs and performance for mixed samples. For. Sci. International *85*, 209–218.

Roewer, L., Arnemann, J., Spurr, N.K., Grzeschik, K.-H., and Epplen, J.T. (1992). Simple repeat sequences on the human Y chromsome are equally polymorphic as their autosomal counterparts. Hum. Genet. *89*, 389–394.

Roewer, L., Kayser, M., Dieltjes, P., Nagy, M., Bakker, E., Krawczak, M., de Knijff, P. (1996). Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations. Hum. Mol. Genet. *5*, 1029–1033.

Santos, F.R., Epplen, J.T., and Pena, S.D.J. (1993). Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism. In: DNA Fingerprinting: State of the Science. S.D.J. Pena, R. Chakraborty, J.T. Epplen and A.J. Jeffreys, eds. (Basel, Switzerland: Birkhäuser Verlag), pp. 261–265.

Santos, F.R., Gerelsaikhan, T., Munkhtuja, B., Oyunsuren, T., Epplen, J.T., Pena, S.D.J. (1996). Geographic differences in the allele frequencies of the human Y-linked tetranucleotide polymorphism DYS19. Hum. Genet. *97*, 309–313.

Slatkin, M. (1995). A measure of population subdivision based on microsatellite allele frequencies. Genetics *139*, 457–462.

Sullivan, K.M., Mannucci, A., Kimpton, C.P., and Gill, P. (1993). A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques *15*, 636–641.

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

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

Wu, D.Y., Ugozzoli, L., Pal, B.K., Qian, J., and Wallace, R.B. (1991). The effect of temperature and oligonucleotide primer length on the specificity and efficiency of amplification by the polymerase chain reaction. DNA Cell. Biol. *10*, 233–238.

Received May 13, 1997; accepted June 10, 1997