J Forensic Sci, March 2007, Vol. 52, No. 2
doi:10.1111/j.1556-4029.2007.00395.x
Available online at: www.blackwell-synergy.com

FOR THE RECORD

Erin K. Hanson,[1,2] *M.S. and Jack Ballantyne,*[1,2,3] *Ph.D.*

Population Data for a Novel, Highly Discriminating Tetra-Local Y-Chromosome Short Tandem Repeat: DYS503

POPULATION: Population samples for gene diversity studies were obtained from the Virginia Division of Forensic Science (bloodstains), Richmond, VA. Ninety-eight Caucasian and 100 African American samples were included in this study.

KEYWORDS: forensic science, DNA typing, population genetics, Y-chromosome STR, DYS503, tetra-local, genotype frequencies, Richmond, VA, Caucasian, African American

During the construction of an annotated short tandem repeat (STR) physical map of the human Y chromosome (1), we identified a novel tetra-local locus, DYS503. As DYS 464, a previously characterized tetra-local locus, possesses high discriminatory capacity (>0.9) (2), population data from Caucasian Americans and African Americans were collected in order to determine whether DYS503 also exhibited an elevated diversity. Genotype frequencies and allele distributions were obtained in a sampling of 198 male individuals, from the Caucasian American ($n = 98$) and African American ($n = 100$) populations. The genotype diversity for DYS503 was 0.92 for the Caucasian American population and 0.75 for the African American population, which, for the Caucasian American population, is larger than any known single locus Y-STR. (Table 1)

The dried bloodstains used for the gene diversity studies were incubated overnight at 56°C in 400 µL of DNA extraction buffer (100 mM NaCl, 10 mM Tris-HCl, 25 mM ethylenediaminetetraacetic acid, 0.5% SDS) and 0.1 mg/mL Proteinase K. The swab pieces were placed in a spin ease basket and centrifuged at 14,000 *g* for 5 min. An equal volume of phenol/chloroform/isoamyl alcohol was added to the crude extract. The aqueous phase extracts containing the DNA were purified using Centricon 100™ concentrators (Millipore, Bedford, MA) according to the manufacturer's instructions. DNA was quantitated using ethidium bromide-induced fluorescence on a 1% agarose yield gel, using a reference set of DNA standards of known concentration (3).

The 25 µL reaction mix contained 3 ng of template DNA, 0.96 µM primers, 1 mM dNTPs, 1X PCR Buffer II (10 mM Tris-HCl, pH 8.3, 50 mM KCl), 2.0 mM $MgCl_2$, 10 µg of nonacetylated bovine serum albumin (Sigma-Aldrich, St. Louis, MO), and 1.5 U of AmpliTaq Gold DNA Polymerase (Applied Biosystems, Foster City, CA). Primer sequences for DYS503 were obtained from the

TABLE 1—*Genotype frequencies for DYS503 (expanded data) in the Caucasian and African American Populations.*

Genotype	*n*	Frequency (Caucasian)	*n*	Frequency (African American)
7-7-13-14	0	0.00	1	0.01
9-14-14-15	1	0.01	0	0.00
10-14-15-15	1	0.01	0	0.00
11-11-14-14	0	0.00	1	0.01
11-11-14-15	1	0.01	1	0.01
11-11-14-16	1	0.01	0	0.00
11-12-12-13	0	0.00	1	0.01
11-12-14-14	1	0.01	0	0.00
11-12-14-15	0	0.00	1	0.01
11-13-14-14	2	0.02	0	0.00
11-13-14-15	2	0.02	2	0.02
11-14-14-14	7	0.07	3	0.03
11-14-14-15	20	0.20	10	0.10
11-14-14-16	11	0.11	1	0.01
11-14-14-17	1	0.01	0	0.00
11-14-15-15	5	0.05	1	0.01
11-14-15-16	3	0.03	1	0.01
12-12-13-14	2	0.02	3	0.03
12-12-14-14	1	0.01	2	0.02
12-12-14-15	1	0.01	0	0.00
12-13-13-13	1	0.01	0	0.00
12-13-13-14	0	0.00	1	0.01
12-13-14-14	3	0.03	2	0.02
12-13-14-15	3	0.03	1	0.01
12-14-14-14	1	0.01	2	0.02
12-14-14-15	1	0.01	1	0.01
13-13-13-14	1	0.01	7	0.07
13-13-14-14	7	0.07	49	0.49
13-13-14-15	3	0.03	3	0.03
13-14-14-14	10	0.10	2	0.02
13-14-14-15	6	0.06	2	0.02
13-14-15-15	1	0.01	0	0.00
14-14-14-14	2	0.02	1	0.01
Genotype Diversity		0.92		0.75

[1]Graduate Program in Biomolecular Science, University of Central Florida, PO Box, 162366, Orlando, FL 32816-2366.
[2]National Center for Forensic Science, PO Box 162367, Orlando, FL 32816-2367.
[3]Department of Chemistry, University of Central Florida, PO Box 162366, Orlando, FL 32816-2366.

TABLE 2—*Allele sequences for DYS 503.*

Allele	Sequence
7	Y3S7r...$[TTA]_{7}$...Y3S7f
8	Y3S7r...$[TTA]_{8}$...Y3S7f
9	Y3S7r...$[TTA]_{9}$...Y3S7f
10	Y3S7r...$[TTA]_{10}$...Y3S7f
11	Y3S7r...$[TTA]_{11}$...Y3S7f
12	Y3S7r...$[TTA]_{12}$...Y3S7f
13	Y3S7r...$[TTA]_{13}$...Y3S7f
14	Y3S7r...$[TTA]_{14}$...Y3S7f
15	Y3S7r...$[TTA]_{15}$...Y3S7f
16	Y3S7r...$[TTA]_{16}$...Y3S7f
17	Y3S7r...$[TTA]_{17}$...Y3S7f

Y3S7r—reverse primer (www.gdb.org).
Y3S7f—forward primer (www.gdb.org).

Genome Database (www.gdb.org). The cycling conditions consisted of (1) 95°C 11 min, (2) 32 cycles of 96°C 30 sec, 60°C 1 min, 72°C 1 min, (3) and a final extension at 60°C for 30 min.

A 1.75 μL aliquot of the amplified product was added to 24 μL of deionized formamide and 1.0 μL GeneScan 500 LIZ internal lane standard. Tubes containing the above were heated at 95°C for 3 min and snap-cooled on ice for 3 min. Samples were injected onto an ABI Prism 310 Genetic Analyzer using Module G5 (5-s injection, 15 kV, 60°C) and analyzed with GeneScan Analysis Software v3.7 using Filter Set G5. A peak detection threshold of 50 RFUs was used for allele designation.

For an accurate determination of allele number, a reference sample was sequenced using standard Sanger sequencing (Genewiz, North Brunswick, NJ). Alleles were calibrated according to the genotype of the reference sample. From the sequencing data, the repeat structure for the DYS503 alleles was a simple TTA repeat unit. Sequences for the observed alleles are provided in Table 2.

All data were checked for duplicate samples. A small number of duplicate samples ($n = 4$) were removed from the data set after confirmation by the submitting laboratory of their duplicate status by autosomal short tandem repeat (STR) analysis. The following formulae were used: (1) discriminatory capacity = no. of individuals/no. of different haplotypes; (2) gene diversity (h), equivalent to the expected frequency of heterozygotes with autosomal diploid loci, was calculated as $(n/(n-1)) * (1 - \Sigma p_i^2)$, where p_i = allele frequency at the ith locus (4).

Genotypes for DYS503 consisted of one, two, three, or four alleles. Raw data (observed number of peaks/alleles from electropherograms) were converted to "expanded genotypes" to report the relative prevalence of the four alleles present for each sample. Expanded genotypes were determined by comparison of relative peak heights.

Thirty-three unique expanded genotypes were observed among the 198 male samples used in this study. The frequencies for each genotype within each population group are listed in. The overall genotype diversity was determined to be 0.92 for the Caucasian American population and 0.75 for the African American population. Significant differences in the essentially bimodal genotype distributions were observed between the two populations. The most frequent genotype in Caucasian Americans was 11-14-14-15 (20% of samples), whereas 13-13-14-14 (49% of samples) was the most common genotype observed in the African American population, further illustrating the significant stratification of the two populations obtained by DYS503 typing. Genotype diversity was higher in Caucasian Americans (0.92) than in African Americans (0.75), a situation observed with a number of other Y-STRs. For example, a comprehensive population survey of 90 YSTR loci indicated that *c.* 18% ($n = 16$) of the loci exhibited greater gene diversity within the Caucasian American than the African American populations (unpublished observations).

The DYS503 gene diversity for Caucasian Americans is greater than that reported for any single locus Y-STR, with the exception of DYS 643 (5). The latter data were obtained by typing only eight males (5) but subsequent testing has indicated that the actual gene diversity is significantly lower than this (unpublished data).

The complete data set is available to any interested researcher upon request.

Acknowledgments

This work was supported under Award number 1998-IJ-CX-K003 from the Office of Justice Programs, National Institute of Justice, Department of Justice. The points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice. The authors would like to thank Jeff Ban, Virginia Division of Forensic Sciences, for providing the population samples.

References

1. Hanson E, Ballantyne J. Comprehensive annotated STR physical map of the human Y chromosome: forensic implications. Leg Med (Tokyo) 2006;8(2):110–20.
2. Hanson EK, Ballantyne J. A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. J Forensic Sci 2004;49(1):40–51.
3. Kline MC, Duewer DL, Redman JW, Butler JM. NIST mixed stain study 3: DNA quantitation accuracy and its influence on short tandem repeat multiplex signal intensity. Anal Chem 2003;75(10):2463–9.
4. Nei M. Molecular evolutionary genetics. Columbia, NY: University Press, 1987.
5. Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, et al. A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 2004;74(6):1183–97.

Additional information and reprint requests:
Jack Ballantyne, Ph.D.
Associate Professor
Department of Chemistry
University of Central Florida
Bldg#5, 4000 Central Boulevard
Orlando, FL 32816-2366
E-mail: jballant@mail.ucf.edu