

FOR THE RECORD

Anamika Singh,¹ M.Sc.; R. Trivedi,¹ Ph.D.; and V. K. Kashyap,^{1,*} Ph.D.

**Haplotype Diversity at 20 Y-Chromosome
Microsatellites in an Indo-Caucasian Population
of Bengal, India**

POPULATION: Sixty-one Indo-Caucasian individuals from eastern India.

KEYWORDS: forensic science, DNA typing, population genetics, Y-STR, DYS426, DYS392, DYS388, DYS389I, DYS389II, DYS439, DYS437, DYS391, DYS385a, DYS385b, DYS390, DYS393, DYS19, DYSH4, DYS460, DYS447, DYS438, YCAIIa, YCAIIb, DYS448, haplotype, Indo-Caucasian, Bengal, eastern-India

This study presents the distribution of Y-chromosomal haplotypes at 20 short tandem repeats (STRs) in a caste group of eastern India. The studied STRs comprise three trinucleotide repeats (DYS426, DYS392 & DYS388), 12 tetranucleotide repeats (DYS389I, DYS389II, DYS439, DYS437, DYS391, DYS385a, DYS385b, DYS390, DYS393, DYS19, DYSH4, DYS460), two pentanucleotide (DYS447, DYS438), two dinu-

cleotide markers (YCAIIa, YCAIIb), and one hexanucleotide marker, DYS448. Ethnically, all the individuals belong to the Indo-caucasian group and speak Bengali—a language of the Indo-European linguistic family. Almost all the individuals either depend on agriculture or work as palanquin-bearers and boat-sailors, for their livelihood. This is the first study involving 20 Y-STR markers in the population of this geographic area and

TABLE 1—Allele frequency for the 16 Y-chromosomal short tandem repeats in an Indo-Caucasian population of Bengal, India.

Alleles	DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS438	DYS439	DYS388	DYS426	DYS437	DYS447	DYS448	DYS460	DYSH4
7					0.016											
8					0.18			0.033								
9					0.377	0.088		0.23							0.377	
10					0.295	0.526		0.23							0.524	0.164
11					0.115	0.053	0.23	0.507		0.066	0.426				0.066	0.442
12		0.59				0.298	0.492			0.197	0.082				0.033	0.328
13		0.361					0.196			0.066			0.754			0.066
14	0.279	0.049				0.035	0.082						0.131			
15	0.475												0.115			
16	0.213				0.017				0.049	0.049						
17	0.033			0.017					0.131	0.622						
18									0.656							
19									0.148							
20									0.016							
21				0.017											0.164	
22				0.262								0.066		0.525		
23				0.311								0.31		0.279		
24				0.098								0.148		0.016		
25				0.262								0.23		0.016		
26				0.033								0.18				
27			0.016													
28			0.115													
29			0.344													
30			0.328													
31			0.18													
32			0.017													

¹National DNA Analysis Center, Central Forensic Science Laboratory, 30, Gorachand Road, Kolkata-700014, India.

*Present address: National Institute of Biologicals, A-32, Sector 62, Institutional Area, Noida-201307, India.
Higher values are signified in bold.

TABLE 2—Frequency of allele pattern for the *DYS385* and *YCAII*STR marker in an Indo-Caucasian population of Bengal, India.

Allelic Combination Observed at <i>DYS385a&b</i>	Frequency	Allelic Combination Observed at <i>YCAIIa&b</i>	Frequency
8–12	0.016	11–18	0.016
8–13	0.016	14–16	0.016
8–16	0.016	14–17	0.049
9–11	0.049	14–19	0.016
9–12	0.082	15–16	0.098
9–13	0.032	15–17	0.229
9–17	0.016	15–18	0.032
10–11	0.032	15–19	0.082
10–12	0.016	15–21	0.049
10–13	0.016	16–21	0.032
10–14	0.016	17–17	0.016
10–15	0.016	17–18	0.032
10–16	0.016	17–19	0.131
11–14	0.016	17–20	0.082
11–15	0.016	17–21	0.016
11–16	0.115	18–19	0.098
12–17	0.016	—	—
13–14	0.115	—	—
13–15	0.098	—	—
13–16	0.016	—	—
13–17	0.065	—	—
14–15	0.098	—	—
14–16	0.049	—	—
14–17	0.016	—	—
15–19	0.016	—	—
17–17	0.016	—	—

shows a high polymorphism at all the loci. There are 61 different unique haplotypes detected with a frequency 0.016. The haplotype diversity for the studied population is 1.000 ± 0.003 , whereas the mean number of pairwise difference is calculated as 12.632 ± 5.775 . There is no haplotype sharing observed in the studied population.

Blood samples were collected from 61 healthy consenting unrelated male individuals. DNA was isolated from the collected samples by the organic extraction method (1). The extracted DNA was quantified through the slot–blot hybridization technique using the QuantiBlot[®] Human DNA Quantification Kit (Applied Biosystems, Foster City, CA). The 20 Y-STRs mentioned above were coamplified using 10 ng of template DNA and fluorescent-labeled primers as described earlier (2,3) in a single tube twentyplex PCR reaction under laboratory-standardized conditions. Amplification products were run on a 5% polyacrylamide gel using LIZTM-500 (PE Applied Biosystems) as the internal lane standard on an ABI Prism 3100 Genetic AnalyzerTM (Applied Biosystems, Foster City, CA). Sizing of the products was performed using the ABI 377 Genescan Analysis Software version (3.1).

The alleles were named according to the number of repeat units that they contain. The allele frequency for each locus was calculated by simple gene-count method. Haplotype diversity was calculated according to the formula (4),

$$h = n(1 - \sum x_i^2)/(n - 1)$$

where n represents the number of chromosomes sampled and x_i is the frequency of the i th allele or haplotype. The mean number of pairwise difference and average gene diversity was calculated using the ARLEQUIN ver. 2.0 (5) software package. The allele frequency and relative haplotype frequency were also calculated.

The results of this extensive study are shown in Tables 1–4.

TABLE 3—Distribution of Y-Chromosomal Haplotypes at 20 Y-chromosomal STRs in an Indo-Caucasian population of Bengal, India.

Haplotype	<i>DYS438</i>	<i>DYS389II</i>	<i>DYS439</i>	<i>DYS437</i>	<i>DYS389I</i>	<i>DYS391</i>	<i>DYS385a</i>	<i>DYS385b</i>	<i>DYS390</i>	<i>YCAa</i>	<i>YCAb</i>	<i>DYS393</i>	<i>DYS426</i>	<i>DYS392</i>	<i>DYS19</i>	<i>DYS388</i>	<i>DYS44</i>	<i>DYS460</i>	<i>DYS448</i>	<i>DYS447</i>	Absolute Frequency
H1	10	29	18	13	13	11	13	17	25	17	20	14	11	12	15	12	11	10	21	28	1
H2	11	30	18	13	13	10	10	16	25	17	21	13	12	10	15	12	12	12	22	25	1
H3	9	30	19	13	13	10	14	15	22	17	19	12	11	10	15	12	12	11	22	24	1
H4	9	29	16	13	13	10	13	16	22	17	19	13	11	10	15	12	12	10	21	23	1
H5	11	28	18	13	13	10	10	14	23	18	19	13	11	10	17	12	11	10	25	23	1
H6	10	29	18	13	13	10	13	17	25	17	20	14	11	12	15	12	11	10	21	28	1
H7	10	28	18	14	12	10	11	14	23	17	19	13	11	10	16	13	12	10	22	25	1
H8	11	27	17	13	12	10	10	13	23	17	19	12	11	10	16	11	11	10	24	23	1
H9	10	29	18	13	12	10	13	14	22	17	19	14	11	10	14	12	11	11	22	22	1
H10	9	28	17	14	12	11	14	15	22	18	19	12	11	12	15	12	12	10	22	23	1
H11	11	28	18	13	13	10	10	15	23	18	19	13	11	10	16	13	11	10	23	24	1
H12	10	29	18	13	13	11	13	17	25	17	19	14	11	12	15	12	11	10	21	28	1
H13	10	29	18	13	13	11	13	17	25	18	19	14	11	12	15	12	11	10	21	28	1
H14	9	29	17	13	13	10	13	15	22	17	19	12	12	10	15	11	12	10	22	23	1
H15	11	30	18	13	14	10	10	11	25	17	20	13	12	10	15	13	12	12	22	25	1
H16	11	30	18	13	14	10	10	11	25	17	20	13	12	10	15	13	12	11	22	25	1
H17	9	28	17	14	12	11	14	15	22	17	19	12	11	10	15	11	12	10	22	23	1
H18	9	28	17	14	12	11	14	15	22	18	19	12	11	10	15	11	12	10	22	23	1
H19	9	28	17	14	12	11	14	15	22	18	19	12	11	10	15	11	12	10	22	23	1
H20	11	31	16	13	14	10	12	15	25	17	20	13	12	10	15	12	12	10	22	26	1
H21	10	29	18	13	14	12	13	15	23	16	21	12	10	10	14	17	10	9	22	25	1
H22	11	31	18	13	13	8	11	16	23	15	17	12	11	10	14	17	11	10	23	26	1

TABLE 3—Continued.

Haplotype	DYS438	DYS389II	DYS439	DYS437	DYS389I	DYS385a	DYS385b	DYS390	YCAa	YCab	DYS393	DYS426	DYS392	DYS19	DYS388	DYS4	DYS460	DYS448	DYS447	Absolute Frequency		
H23	10	30	18	13	12	8	11	16	25	11	18	12	10	10	15	17	17	10	9	21	25	1
H24	11	30	18	14	13	9	8	16	25	15	16	11	10	12	16	17	17	10	10	21	26	1
H25	11	31	18	13	13	8	9	13	25	15	17	12	10	10	15	17	17	11	10	23	24	1
H26	9	31	18	13	13	8	13	15	23	15	16	11	10	0	15	17	10	9	21	25	1	
H27	11	31	18	13	12	9	9	13	23	14	17	12	11	14	15	17	12	10	23	24	1	
H28	11	30	18	13	13	8	11	16	22	15	17	11	11	12	14	17	11	10	22	26	1	
H29	8	29	19	13	12	9	13	14	23	16	21	11	10	15	17	10	10	22	24	1		
H30	11	30	16	13	13	9	11	16	23	15	17	13	11	9	14	17	11	10	22	26	1	
H31	9	32	18	13	12	9	14	16	22	15	16	11	10	12	15	17	11	10	23	25	1	
H32	10	31	17	15	13	10	9	17	23	15	16	12	10	12	17	17	11	9	21	23	1	
H33	11	30	18	14	12	9	11	16	23	15	19	11	10	14	17	12	9	23	23	1		
H34	11	30	18	13	13	8	11	16	22	15	17	12	9	9	15	17	11	9	23	24	1	
H35	11	30	18	13	12	10	8	13	25	15	19	12	11	10	16	17	11	10	23	25	1	
H36	11	30	18	15	12	9	14	16	23	15	18	11	11	10	15	17	11	9	23	26	1	
H37	8	30	19	14	12	8	15	19	26	14	17	11	10	10	15	17	10	9	22	24	1	
H38	11	30	18	13	13	8	11	16	23	14	16	11	11	9	14	16	11	10	22	26	1	
H39	11	30	18	13	12	10	8	12	23	15	19	11	11	10	16	16	11	10	23	25	1	
H40	11	31	18	13	12	10	9	12	23	15	19	11	11	10	15	16	11	10	23	24	1	
H41	11	30	18	15	13	9	13	15	22	15	17	11	10	10	15	17	12	10	22	23	1	
H42	9	29	17	13	12	9	13	15	22	15	17	11	10	10	15	17	12	10	22	23	1	
H43	11	30	18	13	12	10	9	12	24	15	21	12	11	10	16	17	13	10	23	23	1	
H44	9	29	19	13	12	9	13	15	21	15	17	12	10	10	15	17	12	9	22	23	1	
H45	11	29	18	15	12	8	9	12	17	15	17	13	10	12	14	17	11	9	22	26	1	
H46	10	29	19	13	12	9	13	14	25	15	16	12	10	12	14	17	10	9	22	22	1	
H47	11	31	18	13	13	9	9	11	24	17	18	12	11	12	16	17	13	9	22	23	1	
H48	11	29	18	15	12	9	12	17	23	15	17	13	11	12	14	17	11	9	22	26	1	
H49	9	30	18	13	12	9	14	16	22	15	17	12	10	0	15	17	12	9	23	22	1	
H50	11	31	19	13	12	9	9	11	24	0	17	12	11	12	16	17	13	9	22	23	1	
H51	11	29	18	15	12	9	11	15	22	14	19	12	10	0	14	17	11	9	22	25	1	
H52	11	30	18	13	12	8	9	12	24	15	21	12	11	0	16	17	11	10	23	23	1	
H53	9	29	19	13	12	9	13	14	22	15	17	12	10	10	14	17	11	9	22	23	1	
H54	10	29	18	13	12	9	13	14	25	15	17	12	10	12	14	17	12	10	22	25	1	
H55	9	30	18	13	12	9	14	15	22	15	17	11	10	11	14	17	11	9	23	22	1	
H56	10	29	18	13	12	9	14	17	25	15	18	12	10	12	16	17	10	9	21	26	1	
H57	11	31	18	13	13	16	9	11	24	17	18	12	10	12	16	17	13	9	22	23	1	
H58	11	29	19	15	12	9	0	17	23	14	17	13	10	9	14	17	11	9	22	26	1	
H59	10	29	18	13	12	9	13	14	24	15	21	12	10	11	14	17	10	9	22	25	1	
H60	10	29	19	13	12	7	13	14	25	15	16	12	10	11	14	17	10	9	22	25	1	
H61	11	31	20	13	12	8	9	12	26	15	19	12	11	9	16	17	12	10	23	24	1	

TABLE 4—Haplotype diversity, mean number of pairwise differences, and average gene diversity in an Indo-Caucasian population of Bengal, India.

Haplotype diversity	1.000 ± 0.003
Mean number of pairwise differences	12.632 ± 5.775
Average gene diversity	0.665 ± 0.337

Sixty-one unique haplotypes were detected in this study, the relative haplotype frequency value for each being 0.0164. When compared with the European Y-STR Haplotype Reference Database (6), using the set of minimal haplotype markers, none of our haplotypes had a match with the 3423 minimal haplotypes present in the database. Similarly, there was no haplotype match for the U.S. population as well as Asian population haplotypes present in the respective databases (7,8). The analysis shows a haplotype diversity value (PD) of 1.000 ± 0.003 and an average gene diversity of 0.665 ± 0.337 . The most frequent allele in the studied population was allele 13 at locus DYS437, followed by allele 18 at DYS439 and 17 at locus DYS388. However, none of the analyzed microsatellite markers was found to be monomorphic. The data generated showed that the 20-YSTR haplotype system is highly polymorphic and discriminatory. This system is, therefore, extremely significant in determining the paternal lineage and differentiating mixture of specimens from various male individuals. This work is also intended to supplement the ever-increasing information on the Indian population database (9).

The complete dataset is available upon request from the corresponding author.

Acknowledgments

The authors are grateful to the participants for providing their blood samples for this study and also to the Ministry of Home Affairs, Government of India, for providing financial support to

Central Forensic Science Laboratory, Kolkata, to carry out the research work. We are also grateful to Ms. Sanghamitra Sahoo for providing valuable suggestions and technical support.

This study is supported by a grant from BPR&D, MHA, Government of India.

References

1. Sambrook J, Fritsch EF, Maniatis T. Molecular cloning. A laboratory manual. 2nd ed. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press, 1989.
2. Butler JM, Schoske R, Vallone PM, Kline MC, Redd AJ, Hammer MF. A novel multiplex for simultaneous amplification of 20 Y-chromosome STR markers. *Forensic Sci Int* 2002;129:10–24.
3. Schneider PM, Meuser S, Waiyawuth W, Rittner C. Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations. *Forensic Sci Int* 1998; 97:61–70.
4. Nei M. Molecular evolutionary genetics. New York, NY: Columbia University Press, 1987.
5. Schneider S, Roessli D, Excoffier L. ARLEQUIN ver 2.0. A software for population genetics data analysis. Geneva, Switzerland: Genetics and Biometry Laboratory, University of Geneva, 2000.
6. European Y-STR Haplotype Reference Database. <http://www.ystr.org/>
7. Y-STR Haplotype Reference Database for US Populations. <http://www.ystr.org/usa>
8. Y-STR Haplotype Reference Database for Asian Populations. <http://www.ystr.org/asia>
9. Sahoo S, Chainy GBN, Kashyap VK. Allele frequency of eight Y-chromosome STR loci in Oriya population of India. *J Forensic Sci* 2003; 48(1):245–52.

Additional information and reprint requests:

V. K. Kashyap, Ph.D.

National Institute of Biologicals A-32

Sector 62, Institutional Area

Noida-201307

India

E-mail: vkk2k@hotmail.com