Polymorphism Profile of Nine Short Tandem Repeat Loci in the Han Chinese

Shuangding Li[1,3*], Chunxia Yan[1,2,3*], Yajun Deng[1,2,3], Ruilin Wang[2,3], Jian Wang[2], Huanming Yang[2], and Shengbin Li[1,2,3#]

[1] *Key Laboratory of National Ministry of Health for Forensic Sciences, Xi'an Jiaotong University, Xi'an 710061, China;* [2] *Beijing Genomics Institute, Chinese Academy of Sciences, Beijing 101300, China;* [3] *Forensic Center in Beijing Genomics Institute, Beijing 101300, China.*

Nine short tandem repeat (STR) markers (D3S1358, VWA, FGA, THO1, TPOX, CSFIPO, D5S818, D13S317, and D7S820) and a sex-identification marker (Amelogenin locus) were amplified with multiplex PCR and were genotyped with a four-color fluorescence method in samples from 174 unrelated Han individuals in North China. The allele frequencies, genotype frequencies, heterozygosity, probability of discrimination powers, probability of paternity exclusion and Hardy-Weinberg equilibrium expectations were determined. The results demonstrated that the genotypes at all these STR loci in Han population conform to Hardy-Weinberg equilibrium expectations. The combined discrimination power (DP) was 1.05×10^{-10} within nine STR loci analyzed and the probability of paternity exclusion (EPP) was 0.9998. The results indicate that these nine STR loci and the Amelogenin locus are useful markers for human identification, paternity and maternity testing and sex determination in forensic sciences.

Key words: short tandem repeats, polymorphism, Han population

Introduction

Most DNA polymorphisms are variations in the DNA sequence that usually have no obvious effect on the phenotype of an individual and are inherited from generation to generation according to Mendelian rules. They contribute significantly to the variability in the DNA sequence of individuals of a species. Among the variations, there are many short tandem repeats (STRs) in the human genome. STRs appear to be abundant throughout the human genome, occurring on average every 6-10 Kb (*1*). Although generally small, 100-350 bp in size and simple repeated sequences of 2-4 bp in length, STRs exhibit a high degree of length and sequence polymorphism and provide a source of highly informative loci that can be used for the identification of individuals. Therefore, STR markers are of great value for genetic studies and forensic identification (*2-6*).

*** These authors contributed equally to this work.**
 # Corresponding author.
E-mail: lishengbin@genomics.org.cn

By using four-color fluorescence DNA sequencing platform (Applied Biosystems Automated DNA Sequencer Model 377) and multiplex PCR, we analyzed nine STR loci (D3S1358, VWA, FGA, THO1, TPOX, CSFIPO, D5S818, D13S317, and D7S820) and a sex-typing locus (Amelogenin) with samples collected from 174 unrelated Han individuals in Chang'an County, North China.

Results and Discussion

The allelic frequencies of the nine STR loci and the Amelogenin locus in the Han Chinese were summarized and compared with those from African-American and Caucasian populations (Table 1). We used pertinent data to determine the discrimination power (DP) and probabilities of paternity exclusion (EPP) of the Han Chinese for nine STR loci (Table 2). These STR data conformed to the predictions of Hardy-Weinberg equilibrium.

Three fluorescent labels and size separation provided by DNA sequencer Model 377 enabled us to distinguish many independent loci in the same lane. To exploit the potential to increase throughput with

ABI prism multicolor fluorescence technology, we performed multiplex electrophoresis by co-loading the products of multiple PCR reactions in the same lane (Figure 1).

Table 1 Allele Frequencies for Nine STR Loci

Allele size range	Han Chinese (n=174)	African-American (n=195)	Caucasian (n=200)	Allele size range	Han Chinese (n=174)	African-American (n=195)	Caucasian (n=200)	Allele size range	Han Chinese (n=174)	African-American (n=195)	Caucasian (n=200)
D3S1358				**THO1**				**D5S818**			
9	*	0.26	*	5	*	0.26	*	7	*	0.26	0.25
10	*	*	*	6	17.26	13.08	25.25	8	1.28	5.13	0.50
11	*	0.26	0.25	7	28.57	40.00	15.00	9	8.97	2.05	2.25
12	*	0.51	0.25	8	4.76	22.82	9.25	10	11.54	7.44	6.75
13	*	*	0.50	9	46.43	13.08	13.75	11	41.03	25.39	39.25
14	4.17	11.80	11.25	9.3	2.38	10.00	35.00	12	25.64	32.56	33.25
15	35.17	27.95	28.25	10	0.6	0.77	0.75	13	11.54	24.87	16.50
15.2	*	0.26	*	TPOX				14	*	2.05	1.00
16	26.19	32.31	2.25	6	*	8.21	*	15	*	0.26	0
17	20.83	21.80	22.25	7	0.60	3.59	*	16	*	*	0.25
18	10.71	4.62	14.50	8	48.19	34.36	58.25	D13S317			
19	2.38	0.26	0.50	9	15.06	20.00	11.25	5	3.25	*	0.25
VWA				10	0.60	9.23	4.75	8	24.03	3.59	11.50
11	*	*	*	11	32.53	21.54	22.50	9	13.64	2.31	7.75
12	*	0.54	*	12	3.01	3.08	3.00	10	12.99	2.31	6.75
13	*	1.54	*	13	*	*	0.25	11	23.38	27.18	31.25
14	25.30	7.70	8.50	FGA				12	20.13	44.36	28.25
15	1.81	22.05	8.20	16.2	*	0.26	*	13	2.6	14.10	9.75
16	17.47	26.92	19.75	17	*	0.26	*	14	*	6.15	4.25
17	34.34	16.92	25.00	18	0.62	0.51	1.50	15	*	*	0.25
18	15.66	13.85	11.00	19	4.32	4.62	6.25	D7S820			
19	5.42	8.46	11.00	19.2	*	0.26	*	6	0.66	0.51	*
20	*	2.06	1.50	20	3.7	4.36	16.25	6.3	*	*	0.25
21	*	*	0.25	20.2	*	*	0.75	7	*	*	2.50
22	*	0.26	*	21	9.26	13.33	17.75	8	13.16	17.95	17.50
CSP1PO				21.2	2.08	0.26	*	9	4.61	11.80	13.00
6	*	0.26	*	22	16.05	18.97	16.50	10	15.79	33.59	24.00
7	*	7.18	*	22.2	1.85	0.26	6.50	11	31.58	22.82	23.82
8	1.28	7.44	0.25	23	24.69	18.97	14.00	12	27.63	9.49	16.00
9	5.77	3.85	2.75	24	20.99	16.41	13.25	13	6.58	3.59	2.75
10	23.08	30.51	26.75	25	13.58	12.31	11.25	14	*	0.26	0.75
10.3	*	*	0.26	26	4.70	4.10	1.50	15	*	*	0.25
11	22.44	21.28	25.75	27	1.23	4.10	0.50				
12	40.38	22.05	35.75	28	*	0.77	*				
13	6.41	5.92	7.00	29	*	0.26	*				
14	*	0.51	1.50	30	*	*	*				
15	0.64	*	*								

* Alleles were not found in these population samples. The highest values of allele frequencies are underlined.

Table 2 Probability of DP and EPP for STR Loci

Loci	DP			Loci	EPP		
	Han Chinese	African-American	Caucasian		Han Chinese	African-American	Caucasian
D3S1358	0.1110	0.102	0.078	D3S1358	0.4370	0.5250	0.5797
VWA	0.0854	0.058	0.065	VWA	0.6421	0.6394	0.6170
FGA	0.0598	0.035	0.036	FGA	0.6567	0.7202	0.7173
THO1	0.1553	0.102	0.094	THO1	0.5063	0.5250	0.5418
TPOX	0.1905	0.081	0.211	TPOX	0.4796	0.5764	0.3589
CSFIPO	0.0870	0.070	0.122	CSFIPO	0.6085	0.6048	0.4854
D5S818	0.1030	0.097	0.140	D5S818	0.6602	0.5375	0.4554
D13S317	0.0813	0.031	0.074	D13S317	0.7222	0.4725	0.5940
D7S820	0. 0862	0.081	0.061	D7S820	0.7366	0.5742	0.6307
Combined	1.05×10^{-10}	1.23×10^{-10}	2.79×10^{-10}	Combined	0.9998	0.9996	0.9994

Allele definitions for STRs are based on the fragment length of PCR products estimated by slab gel or capillary electrophoresis. Experimental variations in fragment size have led to the practice of "binned" alleles, grouping allele fragments belonging to a particular size range, centered around the average size within an accepted limit. For a fragment with a length of 169 bp, the corresponding allele binning definition would be 169.00±0.50 bp. This method allows us to define alleles based on sample sizes. For example, allele THO1-9 in our samples was different from allele THO1-9.3 because the fragment length of THO1-9 was 185 bp while the length of THO1-9.3 was 189 bp. The difference in size of over 2-bp between two individuals provides a reliable measure for the different definitions (Figure 2).

Most GeneScan fragment analyses depend greatly on PCR amplification. A successful amplification is governed by the quality of DNA template and other relevant factors in the PCR system. Either overloading or underloading template can affect the results of GeneScan. For this reason, we recommend using 1.0-1.5 ng DNA to avoid insufficient templates and PCR inhibitors (for example, heme compounds, EDTA, or certain dyes). To improve the specificity and sensitivity of PCR amplification, we used AmpliTaq Gold DNA polymerase at 95°C for 5 min prior to the PCR cycles.

Compared with the conventional methods of DNA identification, such as RFLP, VNTR (*7, 8*), STR GeneScan analysis is able to detect small alleles in highly degraded samples, decreasing assay time and labor, and also non-isotopic labeling (*9*). These combined qualities make this technique an extremely robust tool for criminal investigations, the study of human origin, gene diagnosis, archaeology, genomic mapping, and also in the protection of the gene resources.

Table 3 GeneScan Loci Designated by GenBank Locus Name and Motifs

GenBank Locus Name	Chromosome Location	Sequence Repeats	Allele Size Range	Known Alleles
Amelogenin	Xp22.1-22.3 Y	NA	107,113	NA
D3S1358	3P	TCTA	114∼142	12, 13, 14, 15, 16, 17, 18, 19
D5S818	5q	TGAT	135∼171	7, 8, 9, 10, 11, 12, 13, 14, 15, 16
VWA	12P12p	AGAT	157∼197	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21
D13S317	13q22∼q31	GATA	206∼234	8, 9, 10, 11, 12, 13, 14, 15
THO1	11P15.5	AATG	169∼189	5, 6, 7, 8, 9, 9.3, 10
TPOX	2p23∼2p	AATG	218∼242	6, 7, 8, 9, 10, 11, 12, 13
FGA	4q	TTTC	219∼267	18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30
D7S820	7q	GATA	258∼294	6, 7, 8, 9, 10, 11, 12, 13, 14, 15
CSF1PO	5q33.3∼34	AGAT	281∼317	6, 7, 8, 9, 10, 11, 12, 13, 14, 15

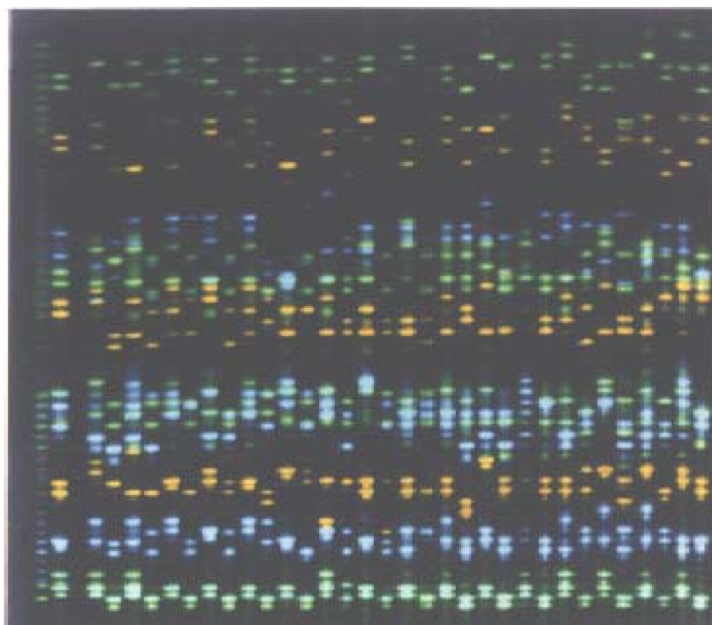

Fig. 1 Fluorescence image of nine STR loci and Amelogenin locus in the ABI prism 377 DNA sequencer. The left lane is allele ladder and other lanes are samples. Yellow-NED-labeled loci are D5S818, D13S317, and D7S820. Blue-S-FAM-labeled loci are D3S1358, VWA and FGA. Green-JOE-labeled loci are Amelogenin, THOE, TPOX, and CSF1PO.

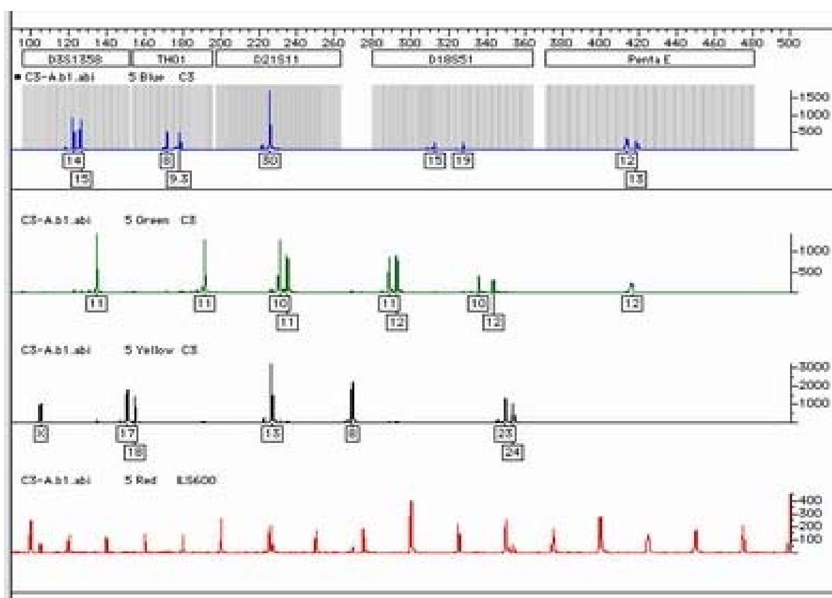

Fig. 2 Alleles identified by Genotyper 2.5.2. DNA fragment sizes are indicated with a ruler. The blue, green and black peaks are loci labeled with different fluorescence, and the red peak is ROX for the GeneScan internal lane size standards. The boxes numbers are alleles of different loci. The vertical scales indicate the relative fluorescence unit (RFU).

Materials and Methods

Samples and markers

Blood samples were taken from unrelated volunteer donors of the Han Chinese who had settled in Chang'an County at least three generations before sample collecting. We divided 1 mL fresh blood into three aliquots, stored at -80°C, at room temperature, and in filter swabs, respectively.

The fragment size of electrophoresis bands was au-

tomatically determined by a calibration curve that was obtained from a co-electrophorized standard size ladder by using the GeneScan software. The pertinent information for these STRs is given in Table 3 (*3*, *4*).

DNA extraction and quantification

DNA was extracted from 50μL fresh blood or bloodstains by using the extraction buffer. After ethanol precipitation, we dried the DNA in a vacuum centrifuge and then added 100 μL deionized water for future use. We diluted DNA in order to add the minimum necessary volume for sequencing. According to the results of Yield Gel, we diluted to 1.0-1.5 ng/μL DNA with deionized water.

DNA amplification and electrophoresis

Multiplex PCR is a technique of simultaneously amplifying multiple DNA targets using multiple primer pairs at one PCR reaction. For each reaction, we added 2 μL template DNA, and 3 μL Master Mix to a separate tube. The total volume was 5μL. The samples were amplified in a Perkin Elmer DNA Thermal cycler 9700 for 30 cycles using the following parameters: 94°C, 40 s for denaturing; prime anealing at 59°C, 30 s; extension at 60°C, 20 min. Before carrying out the cycling, we denatured the DNA at 95°C for 5 min. The parameters of the pre-run were as follows: time 20 min, power 168 W, voltage 3.0 KV, current 48 mA. The parameters of the run were: time 90 min, power 200 W, voltage 3.0 KV, current 60 mA. At first, we added 1.3 μL FLS/GS350 to the sample and denatured it at 95°C for 2 min. Then we placed the sample on the ice for future use. The markers were dealt with the same way. Then we loaded 1 μL volume of samples onto the gel comb cannels and performed a run. The raw data was stored in a gel file along with sample information from the run sheet.

GeneScan techniques and allele designation

The GeneScan Analysis Software analyzed the data collected by the ABI prism 377 DNA sequencers to determine the size and quantity of DNA fragments. GeneScan analyses of sample files included establishing a baseline, adjusting for spectral overlap of the dye, peak detection, and size calling. In our tests, we used the same GeneScan size standard, labelled with the same dye for all samples. We compared peak areas, heights, and sizes in nucleotide bases only for those fragments that were labelled with the same dye. There were four colors in the electropherogram. Different DNA fragments labelled with different color fluorescent dyes. Red-350 [ROX] was reserved for the GeneScan Internal Lane Size Standards. These fluorescently labelled DNA ladders were loaded in the same lane as experimental samples. The fragment lengths were 50, 75, 100, 139, 150, 160, 200, 250, 300, 340, and 350 bp, respectively.

According to the repetitive frequency of the nine STR loci, Genotyper computes the allele definitions using allele-binning method. For example, we designated THO1 locus as 5, 6, 7, 8, 9, 3, and 10, depending on the repetitive frequency of this allele.

References

1. Cib, C., *et al.* 1996. A comprehensive genetic map of the human genome based on 5264 microsatellites. *Nature* 380:152-154.
2. Thomas, J., *et al.* 1995. An STS-based map of the human genome. *Science* 270: 1945-1954.
3. Marino, M.A. 1998. Sequencing using capillary electrophoresis of short tandem repeat alleles separated and purified by HPLC. *Electrophoresis* 19: 108-112.
4. Calafell, F, *et al.* 1998. Short tandem report polymorphism evolution in humans. *Europ. J. Hum. Genetics* 6: 38-49.
5. Amar, A., *et al.* 1999. Genetic Variation of three tetrameric tandem repeats in four distinct Israel ethnic groups. *J. Forensic Sci.* 44: 983-986.
6. Christian, M., *et al.* 2001. STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucl. Acids Res.* 29: 320-322.
7. Rosane, S., *et al.* 1998. Allelic frequency distribution for three VNTR markers D6S132, D7S467, D17S26-in Rio de Janeiro population, Brazil. *Forensic Sci. Int.* 94: 33-38.
8. Sacchetti, L., *et al.* 1999. Efficiency of two different nine-loci short tandem repeat systems for DNA typing purpose. *Clin. Chem.* 45: 178-183.
9. Butler, J.M., *et al.* Overview and history of DNA typing. In *Forensic DNA typing-Biology and technology behind STR markers.* pp 1-13. Academic Press, London, United Kingdom.

Received: 28 April 2003
Accepted: 11 May 2003