GMR Genetics and Molecular Research
Online Journal
ISSN:1676-5680

# Structure and polymorphism of 16 novel Y-STRs in Chinese Han Population

**G.Q. Zhang, S.Y. Yang, L.L. Niu and D.W. Guo**

School of Forensic Medicine, Shanxi Medical University,
Taiyuan, Shanxi, P.R. China

Corresponding author: G.Q. Zhang
E-mail: sxmulegmed@163.com

Genet. Mol. Res. 11 (4): 4487-4500 (2012)
Received November 7, 2011
Accepted June 20, 2012
Published October 11, 2012
DOI http://dx.doi.org/10.4238/2012.October.11.1

**ABSTRACT.** Y-chromosome short tandem repeats (Y-STRs) are useful tools for identifying paternity origin and male-female mixed samples because of their male-specificity, haploid inheritance and relatively simplicity. We focused on novel Y-STRs deposited in the human Genome database from DYS708 to DYS726. We typed 16 male-specific Y-STRs from males of a Chinese Han population residing in Shanxi Province (north China), including DYS708-719, DYS721-723, and DYS726, but failed in typing DYS720, DYS724 and DYS725. The 16 Y-STRs, with mean gene diversity (GD) of 0.79, included three trinucleotide Y-STRs (711, 718, 719), nine tetranucleotide STRs (708, 709, 710, 712, 713, 715, 722, 723, 726) and four pentanucleotide repeat STRs (714, 716, 717, 721). DYS712, consisting of eight alleles, was the most informative STR in our population, with a GD of 0.843. The STRs were classified as simple STRs and complex STRs, according to their structures based on sequencing. Genetic indexes, including allele frequencies, haplotype distribution and male-specificity were determined. The Y-STRs, especially those male-specific, tetra- and penta-nucleotide, with only one copy on Y-chromosome, and relative simple structures, such as DYS709, DYS714, DYS715, DYS716, DYS718, DYS719, and DYS726, were suggested for the future forensic

 

DNA analysis, while DYS724 and DYS725 were not recommended for their multi-copy distribution. The population data provided putative Y-STRs for future genetic and forensic applications.

**Key words:** Y-chromosome short tandem repeats (Y-STRs); Shanxi; Population data; Chinese Han

## INTRODUCTION

The human Y chromosome is male-specific. The short tandem repeats (STRs) located on the Y chromosome, especially in the non-recombining region of the chromosome are passed on as a haplotype in a paternal mode of inheritance, providing discriminatory power in determining the male lineage of individuals. Being male-specific genetic markers, they are useful tools to distinguish the male DNA in material from sexual assault cases. However, Y-chromosome STRs (Y-STRs) are less polymorphic compared to autosome STRs because absence of recombination, thus making it less powerful for individual identification. This genetic property requires the establishment of genetic databases based on haplotypes of a group of Y-STRs in different populations in order to increase discrimination power. Y-chromosome polymorphisms are readily influenced by genetic drift processes and offer useful markers for genetic differences between related populations (Perez-Lezaun et al., 1999; Ge et al., 2010). Although over 700 Y-STRs have been named by now, only over 20 Y-STRs have been fully investigated and are commercially used, such as the Ampfl Y-filer kit (Applied Biosystems, USA) and the PowerPlex Y system (Promega). Because of the limitations of Y-STRs' specificity in geographic or population polymorphisms, more Y-STRs with more polymorphism and less population specificity are needed for forensic investigation purposes. Here, we investigated the "novel" released Y-STRs in Genome databank (GDB, www.gdb.org, not available now) from DYS708 to DYS726. The Y-STRs were selected by the following criteria: 1) it was an STR system with small allele size ranging from 100 to 400 bp, which allows for multiplexing and degraded materials; 2) relatively simple structures, which would be easily typed by different laboratories; 3) the repeat-unit size should be ≥3 and a repeat count ≥10, because less stutter bands and higher probability of variation would be expected as this criterium was set; 4) Y-chromosome specific, i.e., only one band appeared in male and no band in female. In this study, we collected blood samples for genotyping in a Chinese population from the Shanxi Province, located in the northern central part of China. The sequence structures and frequencies of each allele of the Y-STR were also confirmed in a Han population from the Shanxi Province. The polymorphisms of 16 Y-STRs from DYS708 to DYS726 were determined and their potential application in forensic casework was tested.

## MATERIAL AND METHODS

### Population

Whole blood samples were obtained from 120 unrelated healthy Chinese Han males living in Shanxi, a northern central Province of China, whose ancestors had lived in this region for at least three generations. Ethnic origin was enrolled by self-declaration. Blood of 20 female volunteers was collected to test Y-STR specificity.

 

## DNA extraction and quantitation

Genomic DNA was extracted using the hydroxybenzene-chloroform protocol (Zhang et al., 2004). Quantitation was done with a UV spectrometer, and DNA was stored at -20°C until used.

## PCR amplification and typing

The sequences of PCR primers were from GDB or redesigned using an online primer designing tool (http://frodo.wi.mit.edu/primer3) and synthesized by Takara (TaKaRa Bio Inc., Japan; primer sequences are listed in Table 1). The amplification reactions were performed in a total of 25 µL containing 0.5-1 µg DNA, 2.5 µL 10X PCR buffer, 200 µM dNTPs, 0.3 µM of each primer, and 1.5 U AmpliTaq DNA polymerase (5 U/µL, TaKaRa Bio Inc.). PCR amplification was carried out using a 9700 Thermal Cycler (Applied Biosystems). PCR was performed with the program as described: 95°C for 5 min, then 32 cycles at 95°C for 50 s, optimal annealing temperature for 45 s, 72°C for 45 s, and a final extension of 10 min at 72°C. The optimal annealing temperature was 59°C for DYS711, DYS712, DYS713, DYS719, DYS721, DYS723, DYS726, and 55°C for the rest of Y-STRs.

**Table 1.** Primers and structures of 16 Y-STRs in the Chinese Population from Shanxi.

| Y-STR loci | Primers | Repeat-units | Repeats | Alleles |
|---|---|---|---|---|
| DYS708 | F: AGTGTATCCGCCATGGTAGC<br>R: CTGCATTTTGGTACCCCATA | *(GATA)$_m$ (GACA)$_n$*; *m* = 9-11, *n* = 7-10 | Tetra | 6 |
| DYS709 | F: GTTGCCATGGTTTCTTGCTT<br>R: CGAACCTGCAAATTGTTCAC | (TTCT)$_4$CTCT(TTCT)$_2$(CTTTTCT)$_2$CTT*(TTCT)$_{13\text{-}17}$* | Tetra | 5 |
| DYS710 | F: GAGGTCAAGGCTGCAAGAATCTATGA<br>R: CATACTCTCTCCCTCCCTCTCTTTTTTC | *(GAAA)$_m$*(*GAGA*)$_n$GA*(GAAA)$_q$*; *m* = 16-17, *n* = 6-8, *q* = 11-14 | Tetra | 5 |
| DYS711 | F: CAGAGCCCAGCACCTAGGTTAAGT<br>R: TGCTGTCATTGTATCTCTTCACTCC | *(CTT)$_m$*(CTC)$_3$(CTT)CTC(CTT)$_3$(CTCCTT)$_4$ CTCCTA*(CTT)$_n$*; *m* = 26-34, *n* = 19-26 | Tri | 6 |
| DYS712 | F: CAAGAACAGCCTGGGTAACAGTG<br>R: TATATGGTACAGCCCATGAACACTT | *(AGAT)$_m$(AGAC)$_n$*; *m* = 13-17, *n* = 5-11 | Tetra | 8 |
| DYS713 | F: GTGCAAGCCAAGGGCTTTATAAGT<br>R: CCTGGGTGACAGACTCCATCTTAAA | *(TCTT)$_m$*TC(TCTT)$_2$(TCTG)$_1$(TCTT)TTT (TCTT)TC*(TCTT)$_n$*TT(TCTT)$_5$CCT(TCTT)TC (TCTT)T(TCTT); *m* = 16-22, *n* = 12-16 | Tetra | 6 |
| DYS714 | F: TATTAGGCCATCTTGCCAGC<br>R: TTTTCTACCTATGATGCCCTTTG | *(TTTTC)$_m$*(TCTTC)$_2$(TTTTC)$_2$(TCTTC)$_2$ (TTTTC)$_2$; *m* = 10-14 | Penta | 5 |
| DYS715 | F: ATGGTTGGAAGAAAGCATTGATGA<br>R: CTAGGTAATTAGCTACCTAGTTAG | *(TAGA)$_{12\text{-}16}$* | Tetra | 5 |
| DYS716 | F: TAAATCAGAATTCCTTTCCAATCCA<br>R: TCTGGGTTTCAGAGTGGGATAATT | *(CACTC)$_{6\text{-}7}$(CATTC)$_{8\text{-}12}$* | Penta | 5 |
| DYS717 | F: GGCCGAGAGAATGGAATTGAT<br>R: CCCGAACTTCAGCACTATGAAATG | (TGTAT)$_2$TAT(TGTAT) *(TGTAT)$_m$(TGTAT )$_n$*; *m* = 5-7, *n* = 10-11 | Penta | 5 |
| DYS718 | F: CACAATGCAAAAAGAAAGAAGA<br>R: GGAAGCAGCACACCAGCTT | *(TTA)$_{14\text{-}17}$* | Tri | 4 |
| DYS719 | F: GAATGGGGAGGGATAACAA<br>R: GGAGAAAATTCAATGCAGA | *(ATA)$_{12\text{-}16}$* | Tri | 5 |
| DYS721 | F: GGGTGATAGAGGGAGGCTTCT<br>R: CGGGCATGAGCTATTGAGTC | *(AAGGG)$_m$*N$_{10}$(AAGGG)$_2$N$_7$(AAGCA)$_6$; *m* = 9-13 | Penta | 5 |
| DYS722 | F: GCAAAATTGTGAACTGACCAGCAAA<br>R: GTGAACCCTCTGCCAACC | *(GAAA)$_m$*AAGA(GAAA)$_2$A(GAAA)$_2$GAGA (GAAA)$_2$; *m* = 13-17 | Tetra | 5 |
| DYS723 | F: GACAGGTGGATGCATAAATGG<br>R: CCTATCTGGCATCTGTCTGC | (GATA)$_2$TAT*(GATA)$_m$*GAT(GATA)$_1$ GAT(GATA)$_7$; *m* = 10-14 | Tetra | 5 |
| DYS726 | F: GGGTAAACCTCTGAAGACCATAC<br>R: GAATGACAGACCAAGACTCTCTC | (CTTC)$_3$N$_{13}$ *(CTTC)$_m$*N$_{21}$(CTTT)$_3$ T(CTTT)$_2$; *m* = 11-14 | Tetra | 4 |

Italic repeat motif sequences indicate the variable repeats in our population.

PCR products were detected by 6% polyacrylamide gel electrophoresis (PAGE). The allelic ladder was made by mixing different PCR products from the same Y-STR. STR structures were confirmed by sequencing of each allele of the ladder. Allele name was assigned according to the International Society of Forensic Genetics (ISFG) guidelines for forensic Y-STR analysis (Gill et al., 2001; Gusmao et al., 2006).

## Statistical analyses

Haplotypes and allele frequencies were calculated by direct counting. Allele diversity [gene diversity (GD)] was calculated according to the Equation:

$$h = n\,(1-\sum x^2) / (n-1), \qquad \text{(Equation 1)}$$

where $n$ is the number of individuals, and $x$ the allele frequency in the population. Standard error (SE) was calculated as follows: $SE = \{(2/n)[\sum x^3 - (\sum x^2)]^2\}^{1/2}$. Haplotype diversity was calculated as $D = 1-\sum P_i^2$, where $P_i$ represents the haplotype frequency. Bias correction was not done because the possibility that when each haplotype is observed once in a database, the corrected bias $D$ would equal to 1 (Nei and Tajima, 1981; Dai et al., 2004; Budowle et al., 2009).

# RESULTS

## Informative loci and allele definition

A total of 84 alleles were found in our population group, ranging from 4 to 8 alleles for each Y-STR. Allelic frequencies of 16 Y-STRs are summarized in Table 2. The allelic number varied from four (DYS718, DYS726) to eight (DYS712). The mean GD of the 16 Y-STR was GD = 0.79 ± 0.03 (mean ± SD), covering a range of 0.713 in DYS718 to 0.843 in DYS712.

GD in 14 Y-STRs was over 75%. The most frequent allele was allele 16 in DYS718 with a frequency of 0.358, followed by allele 19 in DYS726 (with a frequency of 0.325) and allele 14 in DYS719 (with a frequency of 0.308). Each male had one unique haplotype (Table 3).

Analysis of allelic sequences indicated that DYS709, DYS714, DYS715, DYS718, DYS719, DYS721, DYS722, DYS723, and DYS726 are simple Y-STRs, while DYS708, DYS710, DYS711, DYS712, DYS713, DYS716, and DYS717 are complex STR systems. The sequence data are shown in Table 4. They are identical to those in GDB and Genbank databases except for the motif repeats.

## Structures of the 16 Y-STRs

Among the 16 Y-STRs, there were three trinucleotide, nine tetranucleotide and four pentanucleotide STRs. DYS711, DYS718 and DYS719 belonged to the trinucleotide STRs, with the repeat sequence being (TTC), (TTA) and (ATA), respectively. DYS708, DYS709, DYS710, DYS712, DYS713, DYS715, DYS722, DYS723, and DYS726 belonged to the tetranucleotides, and their repeat sequences are listed in Tables 1 and 4. DYS714, DYS716, DYS717, and DYS721 were classified as pentanucleotide STRs. According to the structure of STR, they were classified as simple STRs, and complex STRs in accordance with Kayser's suggestions (Kayser et al., 2004). The numbers of repeat motif needed in the nomenclature system are shown Table 4.

**Table 2.** Allelic frequencies and statistical parameters regarding the 16 Y-STR loci of the Chinese Han population in Shanxi Province (N = 120).

| Allele | DYS708 | DYS709 | DYS710 | DYS711 | DYS712 | DYS713 | DYS714 | | |
|---|---|---|---|---|---|---|---|---|---|
| 18 | | | | | | | 0.192 | | |
| 19 | | | | | 0.033 | | 0.208 | | |
| 20 | | | | | 0.158 | | 0.217 | | |
| 21 | | 0.183 | | | 0.092 | | 0.225 | | |
| 22 | | 0.283 | | | 0.217 | | 0.158 | | |
| 23 | | 0.217 | | | 0.233 | | | | |
| 24 | | 0.175 | | | 0.133 | | | | |
| 25 | 0.125 | 0.142 | | | 0.092 | | | | |
| 26 | 0.142 | | | | 0.042 | | | | |
| 27 | 0.267 | | | | | | | | |
| 28 | 0.208 | | | | | | | | |
| 29 | 0.150 | | | | | | | | |
| 30 | 0.108 | | | | | | | | |
| 34 | | | 0.167 | | | | | | |
| 35 | | | 0.217 | | | | | | |
| 36 | | | 0.308 | | | | | | |
| 37 | | | 0.183 | | | | | | |
| 38 | | | 0.125 | | | | | | |
| 40 | | | | | | 0.100 | | | |
| 42 | | | | | | 0.150 | | | |
| 44 | | | | | | 0.192 | | | |
| 46 | | | | | | 0.208 | | | |
| 48 | | | | | | 0.225 | | | |
| 50 | | | | | | 0.125 | | | |
| 54 | | | | 0.167 | | | | | |
| 57 | | | | 0.167 | | | | | |
| 59 | | | | 0.175 | | | | | |
| 62 | | | | 0.242 | | | | | |
| 64 | | | | 0.133 | | | | | |
| 68 | | | | 0.117 | | | | | |
| GD | 0.823 | 0.795 | 0.786 | 0.831 | 0.843 | 0.826 | 0.804 | | |
| SE | 0.008 | 0.006 | 0.008 | 0.005 | 0.008 | 0.005 | 0.003 | | |
| Allele | DYS715 | DYS716 | DYS717 | DYS718 | DYS719 | DYS721 | DYS722 | DYS723 | DYS726 |
| 10 | | | | | | | | | |
| 11 | | | | | | | | | 0.250 |
| 12 | 0.142 | | | | 0.133 | | | | 0.325 |
| 13 | 0.292 | | | | 0.158 | | | | 0.250 |
| 14 | 0.233 | | | 0.167 | 0.233 | | | | 0.175 |
| 15 | 0.200 | 0.255 | | 0.292 | 0.308 | | | | |
| 16 | 0.133 | 0.183 | | 0.358 | 0.167 | | | | |
| 17 | | 0.158 | | 0.183 | | 0.150 | | | |
| 18 | | 0.192 | 0.267 | | | 0.225 | | | |
| 19 | | 0.242 | 0.250 | | | 0.267 | 0.167 | | |
| 20 | | | 0.150 | | | 0.217 | 0.233 | 0.192 | |
| 21 | | | 0.192 | | | 0.142 | 0.275 | 0.192 | |
| 22 | | | 0.141 | | | | 0.200 | 0.217 | |
| 23 | | | | | | | 0.125 | 0.233 | |
| 24 | | | | | | | | 0.167 | |
| 25 | | | | | | | | | |
| GD | 0.785 | 0.802 | 0.793 | 0.731 | 0.787 | 0.795 | 0.793 | 0.803 | 0.745 |
| SE | 0.008 | 0.004 | 0.006 | 0.1 | 0.009 | 0.006 | 0.006 | 0.003 | 0.007 |

GD = gene diversity; SE = standard error.

## Simple STR

Simple microsatellites are those loci that consist of one type of repeated unit without interruptions of variant area. DYS715, DYS718 and DYS719 showed the simplest structures in this study, which present (TAGA), (TTA) and (ATA) as the repeat-unit, respectively. The loci DYS709, DYS714, DYS721, DYS722, DYS723, and DYS726 have non-variant repeat block adjacent to the variant motif. The sequenced alleles of the six microsatellites showed that the variations were always due to the largest repeat block. Alleles of these simple STRs were named based on the total number of the repeat-unit located in variant and non-variant blocks (Table 4).

**Table 3.** Haplotypes of 16 Y-STR in the Shanxi Chinese population (N = 120).

| No. | DYS708 | DYS709 | DYS710 | DYS711 | DYS712 | DYS713 | DYS714 | DYS715 | DYS716 | DYS717 | DYS718 | DYS719 | DYS721 | DYS722 | DYS723 | DYS726 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 25 | 21 | 37 | 54 | 23 | 50 | 21 | 15 | 18 | 20 | 15 | 15 | 21 | 20 | 25 | 11 |
| 2 | 25 | 21 | 37 | 59 | 26 | 40 | 18 | 14 | 16 | 21 | 16 | 16 | 19 | 23 | 22 | 11 |
| 3 | 25 | 22 | 34 | 59 | 22 | 48 | 22 | 15 | 18 | 21 | 14 | 14 | 19 | 23 | 23 | 12 |
| 4 | 25 | 22 | 35 | 62 | 19 | 46 | 19 | 13 | 16 | 19 | 16 | -14 | 19 | 22 | 24 | 12 |
| 5 | 25 | 22 | 37 | 57 | 25 | 44 | 20 | 13 | 15 | 22 | 17 | 14 | 17 | 21 | 22 | 12 |
| 6 | 25 | 22 | 38 | 57 | 23 | 48 | 18 | 15 | 17 | 18 | 17 | 14 | 20 | 22 | 23 | 11 |
| 7 | 25 | 22 | 38 | 59 | 25 | 42 | 20 | 12 | 19 | 21 | 16 | 13 | 18 | 23 | 25 | 13 |
| 8 | 25 | 22 | 38 | 64 | 20 | 40 | 20 | 15 | 18 | 19 | 14 | 13 | 17 | 21 | 23 | 11 |
| 9 | 25 | 22 | 38 | 68 | 22 | 50 | 18 | 13 | 15 | 20 | 15 | 15 | 20 | 21 | 21 | 11 |
| 10 | 25 | 23 | 36 | 54 | 25 | 46 | 19 | 14 | 18 | 19 | 15 | 13 | 19 | 21 | 23 | 11 |
| 11 | 25 | 23 | 36 | 64 | 24 | 50 | 21 | 13 | 19 | 19 | 17 | 14 | 19 | 20 | 23 | 12 |
| 12 | 25 | 23 | 36 | 64 | 26 | 44 | 19 | 13 | 15 | 20 | 15 | 14 | 17 | 19 | 20 | 11 |
| 13 | 25 | 24 | 35 | 62 | 23 | 44 | 21 | 14 | 15 | 20 | 17 | 15 | 20 | 22 | 24 | 14 |
| 14 | 26 | 21 | 34 | 59 | 23 | 40 | 19 | 12 | 17 | 18 | 16 | 16 | 20 | 19 | 21 | 13 |
| 15 | 26 | 21 | 34 | 59 | 21 | 42 | 19 | 12 | 15 | 18 | 16 | 15 | 18 | 22 | 21 | 13 |
| 16 | 26 | 21 | 36 | 57 | 23 | 44 | 22 | 16 | 16 | 21 | 16 | 13 | 19 | 23 | 24 | 12 |
| 17 | 26 | 22 | 34 | 54 | 24 | 48 | 22 | 15 | 19 | 18 | 17 | 13 | 21 | 19 | 25 | 11 |
| 18 | 26 | 22 | 35 | 62 | 23 | 42 | 20 | 14 | 19 | 20 | 14 | 15 | 18 | 20 | 24 | 13 |
| 19 | 26 | 22 | 36 | 59 | 23 | 50 | 18 | 12 | 19 | 22 | 17 | 16 | 21 | 19 | 21 | 14 |
| 20 | 26 | 22 | 37 | 54 | 26 | 48 | 21 | 16 | 19 | 20 | 16 | 15 | 20 | 21 | 20 | 14 |
| 21 | 26 | 22 | 38 | 62 | 22 | 44 | 21 | 14 | 19 | 19 | 15 | 12 | 18 | 23 | 21 | 14 |
| 22 | 26 | 23 | 35 | 64 | 22 | 42 | 21 | 14 | 19 | 19 | 15 | 15 | 18 | 20 | 25 | 13 |
| 23 | 26 | 23 | 36 | 59 | 20 | 46 | 19 | 15 | 16 | 21 | 15 | 16 | 19 | 23 | 23 | 11 |
| 24 | 26 | 23 | 37 | 62 | 21 | 50 | 22 | 13 | 18 | 19 | 15 | 16 | 19 | 20 | 21 | 13 |
| 25 | 26 | 24 | 36 | 62 | 25 | 46 | 18 | 15 | 16 | 21 | 15 | 15 | 19 | 23 | 25 | 11 |
| 26 | 26 | 24 | 37 | 54 | 22 | 48 | 22 | 15 | 19 | 18 | 14 | 15 | 18 | 21 | 25 | 12 |
| 27 | 26 | 24 | 37 | 62 | 22 | 46 | 20 | 12 | 18 | 21 | 16 | 12 | 20 | 22 | 24 | 11 |
| 28 | 26 | 24 | 37 | 64 | 23 | 44 | 20 | 12 | 18 | 19 | 15 | 12 | 17 | 21 | 22 | 14 |
| 29 | 26 | 25 | 34 | 62 | 23 | 50 | 22 | 16 | 16 | 22 | 16 | 14 | 19 | 22 | 25 | 12 |
| 30 | 26 | 25 | 35 | 57 | 22 | 44 | 22 | 16 | 16 | 18 | 16 | 16 | 20 | 22 | 23 | 12 |
| 31 | 26 | 25 | 37 | 62 | 19 | 42 | 20 | 13 | 18 | 19 | 14 | 13 | 18 | 22 | 23 | 11 |
| 32 | 26 | 25 | 38 | 59 | 20 | 46 | 21 | 15 | 18 | 21 | 16 | 15 | 19 | 21 | 24 | 14 |
| 33 | 27 | 21 | 34 | 68 | 24 | 44 | 21 | 13 | 19 | 19 | 15 | 14 | 18 | 22 | 22 | 12 |
| 34 | 27 | 21 | 35 | 59 | 21 | 50 | 18 | 16 | 17 | 18 | 15 | 12 | 19 | 21 | 20 | 13 |
| 35 | 27 | 21 | 36 | 64 | 24 | 48 | 20 | 13 | 16 | 21 | 16 | 14 | 20 | 20 | 24 | 12 |
| 36 | 27 | 21 | 37 | 57 | 23 | 46 | 21 | 12 | 17 | 18 | 16 | 13 | 21 | 22 | 23 | 12 |
| 37 | 27 | 21 | 37 | 62 | 25 | 46 | 21 | 13 | 18 | 19 | 16 | 15 | 19 | 19 | 20 | 14 |
| 38 | 27 | 21 | 38 | 62 | 23 | 46 | 19 | 15 | 18 | 19 | 15 | 13 | 18 | 19 | 22 | 13 |
| 39 | 27 | 22 | 34 | 62 | 20 | 48 | 22 | 15 | 18 | 19 | 16 | 16 | 19 | 22 | 24 | 11 |
| 40 | 27 | 22 | 35 | 57 | 21 | 48 | 21 | 13 | 19 | 21 | 16 | 15 | 21 | 19 | 21 | 13 |
| 41 | 27 | 22 | 36 | 54 | 21 | 46 | 18 | 15 | 15 | 18 | 17 | 14 | 21 | 19 | 22 | 12 |
| 42 | 27 | 22 | 36 | 59 | 23 | 42 | 19 | 16 | 18 | 22 | 16 | 14 | 20 | 22 | 22 | 14 |

Continued on next page

**Table 3.** Continued.

| No. | DYS708 | DYS709 | DYS710 | DYS711 | DYS712 | DYS713 | DYS714 | DYS715 | DYS716 | DYS717 | DYS718 | DYS719 | DYS721 | DYS722 | DYS723 | DYS726 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 43 | 27 | 22 | 37 | 68 | 23 | 40 | 18 | 15 | 19 | 21 | 15 | 13 | 21 | 21 | 23 | 11 |
| 44 | 27 | 22 | 38 | 57 | 21 | 48 | 22 | 14 | 19 | 21 | 16 | 13 | 19 | 20 | 22 | 11 |
| 45 | 27 | 23 | 34 | 64 | 24 | 40 | 21 | 16 | 19 | 22 | 16 | 15 | 20 | 21 | 22 | 14 |
| 46 | 27 | 23 | 34 | 68 | 24 | 48 | 19 | 16 | 16 | 22 | 16 | 16 | 18 | 22 | 25 | 14 |
| 47 | 27 | 23 | 35 | 62 | 26 | 48 | 19 | 16 | 18 | 19 | 15 | 15 | 19 | 23 | 22 | 14 |
| 48 | 27 | 23 | 37 | 68 | 25 | 48 | 19 | 12 | 18 | 22 | 14 | 14 | 20 | 23 | 20 | 12 |
| 49 | 27 | 23 | 38 |  | 22 | 48 | 20 | 15 | 16 | 21 | 16 | 15 | 18 | 22 | 22 | 13 |
| 50 | 27 | 24 | 34 | 54 | 20 | 46 | 20 | 13 | 16 | 19 | 15 | 14 | 19 | 19 | 20 | 13 |
| 51 | 27 | 24 | 35 | 62 | 22 | 48 | 18 | 15 | 19 | 19 | 14 | 13 | 20 | 21 | 23 | 12 |
| 52 | 27 | 24 | 35 | 64 | 20 | 44 | 20 | 13 | 18 | 19 | 15 | 14 | 19 | 21 | 21 | 12 |
| 53 | 27 | 25 | 35 | 57 | 22 | 50 | 18 | 12 | 19 | 20 | 17 | 15 | 20 | 21 | 21 | 13 |
| 54 | 27 | 25 | 36 | 68 | 23 | 48 | 20 | 12 | 18 | 22 | 16 | 15 | 21 | 22 | 25 | 12 |
| 55 | 27 | 25 | 36 | 68 | 26 | 44 | 19 | 12 | 19 | 19 | 15 | 15 | 18 | 22 | 21 | 13 |
| 56 | 27 | 25 | 37 | 57 | 22 | 48 | 19 | 13 | 17 | 18 | 17 | 15 | 21 | 19 | 21 | 12 |
| 57 | 28 | 21 | 35 | 57 | 21 | 46 | 18 | 15 | 15 | 20 | 17 | 15 | 17 | 19 | 23 | 12 |
| 58 | 28 | 21 | 35 | 62 | 20 | 48 | 20 | 15 | 19 | 18 | 14 | 12 | 19 | 20 | 23 | 11 |
| 59 | 28 | 21 | 36 | 62 | 25 | 46 | 18 | 12 | 15 | 20 | 17 | 16 | 20 | 22 | 25 | 12 |
| 60 | 28 | 21 | 36 | 62 | 23 | 46 | 20 | 13 | 18 | 22 | 15 | 12 | 19 | 20 | 24 | 11 |
| 61 | 28 | 21 | 38 | 64 | 20 | 42 | 21 | 15 | 16 | 19 | 16 | 12 | 19 | 22 | 22 | 11 |
| 62 | 28 | 22 | 34 | 62 | 24 | 50 | 18 | 12 | 15 | 21 | 17 | 15 | 19 | 19 | 25 | 12 |
| 63 | 28 | 22 | 34 | 62 | 22 | 44 | 21 | 14 | 17 | 21 | 17 | 14 | 17 | 20 | 24 | 12 |
| 64 | 28 | 22 | 35 | 54 | 24 | 42 | 18 | 14 | 17 | 20 | 15 | 16 | 17 | 22 | 21 | 14 |
| 65 | 28 | 22 | 35 | 59 | 23 | 44 | 21 | 15 | 15 | 18 | 17 | 15 | 20 | 21 | 22 | 14 |
| 66 | 28 | 22 | 36 | 59 | 22 | 48 | 19 | 14 | 18 | 21 | 16 | 15 | 20 | 23 | 22 | 12 |
| 67 | 28 | 22 | 36 | 64 | 25 | 46 | 22 | 13 | 19 | 19 | 15 | 15 | 17 | 20 | 25 | 13 |
| 68 | 28 | 22 | 37 | 62 | 22 | 46 | 22 | 13 | 17 | 19 | 15 | 15 | 18 | 20 | 22 | 13 |
| 69 | 28 | 22 | 38 | 68 | 23 | 50 | 21 | 13 | 15 | 18 | 16 | 16 | 17 | 20 | 22 | 11 |
| 70 | 28 | 23 | 34 | 59 | 24 | 50 | 21 | 14 | 15 | 20 | 16 | 14 | 18 | 21 | 23 | 12 |
| 71 | 28 | 23 | 35 | 68 | 22 | 44 | 21 | 13 | 15 | 20 | 17 | 12 | 21 | 20 | 22 | 12 |
| 72 | 28 | 23 | 36 | 54 | 23 | 44 | 19 | 13 | 15 | 22 | 17 | 14 | 19 | 21 | 23 | 13 |
| 73 | 28 | 23 | 36 | 57 | 24 | 44 | 22 | 13 | 16 | 20 | 16 | 13 | 21 | 23 | 23 | 13 |
| 74 | 28 | 23 | 36 | 62 | 20 | 42 | 21 | 14 | 15 | 18 | 16 | 15 | 17 | 20 | 22 | 12 |
| 75 | 28 | 23 | 36 | 62 | 22 | 44 | 19 | 14 | 15 | 18 | 16 | 12 | 18 | 19 | 23 | 12 |
| 76 | 28 | 23 | 37 | 57 | 22 | 50 | 18 | 16 | 15 | 20 | 16 | 14 | 17 | 19 | 22 | 11 |
| 77 | 28 | 23 | 38 | 59 | 23 | 44 | 18 | 13 | 17 | 20 | 17 | 16 | 17 | 21 | 24 | 13 |
| 78 | 28 | 23 | 38 | 62 | 21 | 46 | 19 | 16 | 15 | 18 | 14 | 14 | 20 | 19 | 23 | 11 |
| 79 | 28 | 24 | 36 | 54 | 20 | 40 | 18 | 13 | 15 | 20 | 16 | 13 | 17 | 20 | 24 | 13 |
| 80 | 28 | 24 | 36 | 54 | 23 | 42 | 22 | 14 | 19 | 18 | 14 | 12 | 17 | 21 | 23 | 11 |
| 81 | 28 | 24 | 36 | 54 | 20 | 48 | 19 | 16 | 15 | 18 | 17 | 16 | 20 | 20 | 21 | 12 |
| 82 | 28 | 24 | 36 | 62 | 22 | 42 | 19 | 13 | 15 | 19 | 14 | 16 | 18 | 23 | 23 | 11 |
| 83 | 28 | 24 | 37 | 59 | 22 | 46 | 22 | 12 | 15 | 18 | 16 | 14 | 18 | 19 | 20 | 12 |
| 84 | 28 | 24 | 37 | 64 | 19 | 42 | 22 | 16 | 18 | 19 | 15 | 15 | 19 | 20 | 21 | 14 |

Continued on next page

**Table 3.** Continued.

| No. | DYS708 | DYS709 | DYS710 | DYS711 | DYS712 | DYS713 | DYS714 | DYS715 | DYS716 | DYS717 | DYS718 | DYS719 | DYS721 | DYS722 | DYS723 | DYS726 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 85 | 28 | 24 | 38 | 62 | 24 | 44 | 20 | 13 | 16 | 18 | 16 | 12 | 19 | 20 | 22 | 12 |
| 86 | 28 | 24 | 38 | 68 | 22 | 44 | 20 | 15 | 17 | 19 | 15 | 16 | 19 | 21 | 23 | 12 |
| 87 | 28 | 25 | 36 | 54 | 22 | 50 | 20 | 14 | 15 | 20 | 16 | 15 | 17 | 21 | 22 | 14 |
| 88 | 29 | 21 | 34 | 64 | 20 | 44 | 19 | 13 | 19 | 19 | 15 | 13 | 18 | 19 | 23 | 11 |
| 89 | 29 | 21 | 36 | 57 | 23 | 48 | 20 | 15 | 19 | 18 | 14 | 14 | 18 | 20 | 21 | 11 |
| 90 | 29 | 21 | 36 | 57 | 24 | 42 | 20 | 14 | 17 | 20 | 14 | 12 | 18 | 21 | 21 | 13 |
| 91 | 29 | 21 | 37 | 54 | 22 | 40 | 21 | 13 | 18 | 21 | 15 | 16 | 19 | 20 | 20 | 11 |
| 92 | 29 | 21 | 38 | 57 | 25 | 48 | 18 | 14 | 17 | 22 | 17 | 12 | 18 | 20 | 23 | 14 |
| 93 | 29 | 22 | 34 | 54 | 22 | 46 | 21 | 12 | 17 | 19 | 15 | 12 | 21 | 22 | 22 | 13 |
| 94 | 29 | 22 | 35 | 64 | 20 | 50 | 20 | 14 | 15 | 19 | 14 | 16 | 18 | 19 | 21 | 11 |
| 95 | 29 | 22 | 36 | 64 | 25 | 40 | 22 | 14 | 16 | 22 | 14 | 14 | 21 | 21 | 23 | 12 |
| 96 | 29 | 22 | 37 | 54 | 21 | 48 | 20 | 14 | 19 | 18 | 16 | 12 | 18 | 19 | 21 | 11 |
| 97 | 29 | 22 | 37 | 57 | 25 | 46 | 22 | 14 | 16 | 22 | 16 | 15 | 19 | 23 | 22 | 14 |
| 98 | 29 | 22 | 38 | 54 | 22 | 48 | 19 | 14 | 15 | 18 | 17 | 13 | 17 | 20 | 24 | 13 |
| 99 | 29 | 23 | 35 | 59 | 22 | 42 | 19 | 13 | 16 | 21 | 15 | 15 | 20 | 21 | 23 | 12 |
| 100 | 29 | 24 | 35 | 57 | 23 | 44 | 20 | 12 | 17 | 21 | 17 | 14 | 17 | 20 | 24 | 12 |
| 101 | 29 | 24 | 36 | 68 | 23 | 40 | 21 | 14 | 18 | 18 | 16 | 15 | 19 | 21 | 23 | 12 |
| 102 | 29 | 24 | 37 | 57 | 21 | 42 | 20 | 15 | 19 | 21 | 14 | 15 | 20 | 21 | 23 | 14 |
| 103 | 29 | 25 | 36 | 59 | 23 | 42 | 21 | 13 | 17 | 19 | 15 | 15 | 18 | 22 | 24 | 13 |
| 104 | 29 | 25 | 36 | 59 | 20 | 46 | 19 | 14 | 16 | 21 | 14 | 15 | 19 | 20 | 25 | 13 |
| 105 | 29 | 25 | 36 | 62 | 20 | 48 | 22 | 16 | 16 | 19 | 15 | 14 | 17 | 21 | 20 | 11 |
| 106 | 30 | 21 | 34 | 57 | 21 | 44 | 18 | 14 | 19 | 19 | 16 | 14 | 18 | 20 | 21 | 13 |
| 107 | 30 | 22 | 35 | 59 | 23 | 48 | 22 | 14 | 17 | 18 | 16 | 14 | 21 | 22 | 24 | 12 |
| 108 | 30 | 22 | 35 | 68 | 23 | 46 | 20 | 15 | 18 | 18 | 16 | 15 | 20 | 19 | 24 | 13 |
| 109 | 30 | 22 | 36 | 54 | 20 | 46 | 18 | 13 | 16 | 20 | 15 | 12 | 19 | 23 | 20 | 11 |
| 110 | 30 | 23 | 34 | 62 | 24 | 40 | 19 | 14 | 18 | 19 | 15 | 13 | 18 | 20 | 23 | 11 |
| 111 | 30 | 23 | 35 | 54 | 24 | 48 | 19 | 16 | 19 | 22 | 16 | 16 | 21 | 21 | 25 | 14 |
| 112 | 30 | 23 | 35 | 59 | 20 | 40 | 21 | 12 | 16 | 21 | 16 | 13 | 20 | 21 | 20 | 12 |
| 113 | 30 | 23 | 35 | 64 | 23 | 46 | 18 | 16 | 18 | 18 | 16 | 16 | 20 | 21 | 20 | 14 |
| 114 | 30 | 23 | 36 | 64 | 22 | 50 | 21 | 13 | 17 | 18 | 15 | 16 | 21 | 21 | 20 | 13 |
| 115 | 30 | 24 | 35 | 57 | 20 | 42 | 20 | 15 | 19 | 22 | 14 | 16 | 20 | 21 | 23 | 12 |
| 116 | 30 | 24 | 36 | 54 | 20 | 46 | 18 | 13 | 15 | 20 | 16 | 13 | 20 | 23 | 22 | 13 |
| 117 | 30 | 25 | 34 | 59 | 23 | 48 | 21 | 13 | 19 | 22 | 14 | 13 | 18 | 21 | 23 | 14 |
| 118 | 30 | 25 | 34 | 68 | 24 | 42 | 20 | 14 | 16 | 18 | 14 | 13 | 20 | 22 | 20 | 12 |
| 119 | 30 | 25 | 34 | 68 | 19 | 44 | 18 | 13 | 19 | 19 | 15 | 14 | 19 | 20 | 24 | 13 |
| 120 | 30 | 25 | 35 | 62 | 24 | 44 | 21 | 13 | 15 | 18 | 17 | 16 | 21 | 22 | 22 | 12 |
| GD | 0.992 | | | | | | | | | | | | | | | |

GD = gene diversity.

**Table 4.** Allele sequences of the 16 Y-STR loci [P(1) and P(2) are the forward and reverse primers, respectively].

| Loci (ID in GenBank) | Allele (bp) | Sequences | Nomenclature |
|---|---|---|---|
| DYS708 [NT_011875.12 (3476255-3476499)] | Consensus structure | P1(19bp)ataatagaaattttatgagtgggagaaatggatgacagtaaaatgaaaacattgcaatgtgtatactcagaaacaaggaaa(gata)$_2$ gat(gata)$_4$(gaca)(gata)$_m$(gaca)$_n$(gata)$_2$gaatatatP2(20bp) | 2+4+1+m+n+2 |
| | 25 (233) | P1(19bp)(81bp)(gata)$_2$gat(gata)$_4$(gaca)(gata)$_9$(gaca)$_2$(gata)$_2$(8bp)P2(20bp) | |
| | 26 (237) | P1(19bp)(81bp)(gata)$_2$gat(gata)$_4$(gaca)(gata)$_9$(gaca)$_3$(gata)$_2$(8bp)P2(20bp) | |
| | 27 (241) | P1(19bp)(81bp)(gata)$_2$gat(gata)$_4$(gaca)(gata)$_9$(gaca)$_4$(gata)$_2$(8bp)P2(20bp) | |
| | 28 (245) | P1(19bp)(81bp)(gata)$_2$gat(gata)$_4$(gaca)(gata)$_{10}$(gaca)$_9$(gata)$_2$(8bp)P2(20bp) | |
| | 29 (249) | P1(19bp)(81bp)(gata)$_2$gat(gata)$_4$(gaca)(gata)$_{10}$(gaca)$_{10}$(gata)$_2$(8bp)P2(20bp) | |
| | 30 (253) | P1(19bp)(81bp)(gata)$_2$gat(gata)$_4$(gaca)(gata)$_{11}$(gaca)$_{10}$(gata)$_2$(8bp)P2(20bp) | |
| DYS709 [AC006989.3 (34781-34978)] | Consensus structure | P1(20bp)ctttctttccaatgaccaagacgtgc(ttct)$_4$ctct(ttct)$_2$ctt(ttct)ctt(ttct)tcct(ttct)$_m$ ttttattatactttaagttttagggtacatP2(20bp) | 4+2+1+1+m |
| | 21 (194) | P1(20bp)(26bp)(ttct)$_4$ctct(ttct)$_2$ctt(ttct)ctt(ttct)tcct(ttct)$_{13}$(30bp)P2(20bp) | |
| | 22 (198) | P1(20bp)(26bp)(ttct)$_4$ctct(ttct)$_2$ctt(ttct)ctt(ttct)tcct(ttct)$_{14}$(30bp)P2(20bp) | |
| | 23 (202) | P1(20bp)(26bp)(ttct)$_4$ctct(ttct)$_2$ctt(ttct)ctt(ttct)tcct(ttct)$_{15}$(30bp)P2(20bp) | |
| | 24 (206) | P1(20bp)(26bp)(ttct)$_4$ctct(ttct)$_2$ctt(ttct)ctt(ttct)tcct(ttct)$_{16}$(30bp)P2(20bp) | |
| | 25 (210) | P1(20bp)(26bp)(ttct)$_4$ctct(ttct)$_2$ctt(ttct)ctt(ttct)tcct(ttct)$_{17}$(30bp)P2(20bp) | |
| DYS710 [NT_011875.12 (5638035-5638259)] | Consensus structure | P1(26bp)tcacagcacttttctgaatcctggacaagtgagaaaa(gaaa)$_m$(gaga)$_n$ga(gaaa)$_q$P2(28bp) | m+n+q |
| | 34 (225) | P1(26bp)(37bp)(gaaa)$_{16}$(gaga)$_6$ga(gaaa)$_{12}$P2(24bp) | |
| | 35 (229) | P1(26bp)(37bp)(gaaa)$_{17}$(gaga)$_7$ga(gaaa)$_{11}$P2(24bp) | |
| | 36 (233) | P1(26bp)(37bp)(gaaa)$_{17}$(gaga)$_7$ga(gaaa)$_{12}$P2(24bp) | |
| | 37 (237) | P1(26bp)(37bp)(gaaa)$_{17}$(gaga)$_7$ga(gaaa)$_{13}$P2(24bp) | |
| | 38 (241) | P1(26bp)(37bp)(gaaa)$_{16}$(gaga)$_8$ga(gaaa)$_{14}$P2(24bp) | |
| DYS711 [AC016991.5 (23057-23414)] | Consensus structure | P1(24bp)gac(ctt)$_5$t(ctt)$_m$(ctc)$_3$(ctt)ctc(ctt)$_3$(ctcctt)$_4$ctccta(ctt)$_n$ctcctc(ctt)$_3$tatt(ctt)ctgtctttgtttttgttttgagacagagtcttgcactg ttgctcgggctP2(25bp) | 5+m+1+3+n |
| | 54 (328) | P1(24bp)gac(ctt)$_5$t(ctt)$_{25}$(ctc)$_3$(ctt)ctc(ctt)$_3$(ctcctt)$_4$ctccta(ctt)$_{20}$ctcctc(ctt)$_3$tatt(ctt) (49bp)P2(25bp) | |
| | 57 (337) | P1(24bp)gac(ctt)$_5$t(ctt)$_{29}$(ctc)$_3$(ctt)ctc(ctt)$_3$(ctcctt)$_4$ctccta(ctt)$_{19}$ctcctc(ctt)$_3$tatt(ctt) (49bp)P2(25bp) | |
| | 59 (379) | P1(24bp)gac(ctt)$_5$t(ctt)$_{31}$(ctc)$_3$(ctt)ctc(ctt)$_3$(ctcctt)$_4$ctccta(ctt)$_{19}$ctcctc(ctt)$_3$tatt(ctt) (49bp)P2(25bp) | |
| | 62 (388) | P1(24bp)gac(ctt)$_5$t(ctt)$_{27}$(ctc)$_3$(ctt)ctc(ctt)$_3$(ctcctt)$_4$ctccta(ctt)$_{26}$ctcctc(ctt)$_3$tatt(ctt) (49bp)P2(25bp) | |
| | 64 (397) | P1(24bp)gac(ctt)$_5$t(ctt)$_{31}$(ctc)$_3$(ctt)ctc(ctt)$_3$(ctcctt)$_4$ctccta(ctt)$_{24}$ctcctc(ctt)$_3$tatt(ctt) (49bp)P2(25bp) | |
| | 68 (406) | P1(24bp)gac(ctt)$_5$t(ctt)$_{33}$(ctc)$_3$(ctt)ctc(ctt)$_3$(ctcctt)$_4$ctccta(ctt)$_{26}$ctcctc(ctt)$_3$tatt(ctt) (49bp)P2(25bp) | |
| DYS712 [NT_011875.12 (1759829-1759997)] | Consensus structure | P1(23bp)tg(agat)$_m$(agac)$_n$aggatgtcatttgtttatatcatttaattatP2(25bp) | m+n |
| | 19 (157) | P1(23bp)tg(agat)$_{13}$(agac)$_4$(31bp)P2(25bp) | |
| | 20 (161) | P1(23bp)tg(agat)$_{13}$(agac)$_7$(31bp)P2(25bp) | |
| | 21 (165) | P1(23bp)tg(agat)$_{15}$(agac)$_4$(31bp)P2(25bp) | |
| | 22 (169) | P1(23bp)tg(agat)$_{17}$(agac)$_5$(31bp)P2(25bp) | |
| | 23 (173) | P1(23bp)tg(agat)$_{13}$(agac)$_{10}$(31bp)P2(25bp) | |
| | 24 (177) | P1(23bp)tg(agat)$_{14}$(agac)$_{10}$(31bp)P2(25bp) | |
| | 25 (181) | P1(23bp)tg(agat)$_{14}$(agac)$_{11}$(31bp)P2(25bp) | |
| | 26 (185) | P1(23bp)tg(agat)$_{15}$(agac)$_{11}$(31bp)P2(25bp) | |
| DYS713 [NT_011896.9 (5182467-5182786)] | Consensus structure | P1(24bp)tatctgaggctgttatatctttaaacttttttttaccagaaatagatttattcacgcttg(tctt)$_m$tc(tctt)$_2$(tctg)$_1$(tctt)ttt(tctt)tc(tctt)$_n$tt(tctt)$_5$ cct(tctt)tc(tctt)t(tctt)tttttattatttatttatttattP2(25bp) | m+2+1+1+n+5+1+1+1 |
| | 40 (312) | P1(24bp)(61bp)(tctt)$_{16}$tc(tctt)$_2$tctg(tctt)ttt(tctt)tc(tctt)$_{12}$tt(tctt)$_5$cct(tctt)tc(tctt)t(tctt)(23bp)P2(25bp) | |
| | 42 (320) | P1(24bp)(61bp) (tctt)$_{17}$tc(tctt)$_2$tctg(tctt)ttt(tctt)tc(tctt)$_{13}$tt(tctt)$_5$cct(tctt)tc(tctt)t(tctt)(23bp)P2(25bp) | |
| | 44 (328) | P1(24bp)(61bp) (tctt)$_{19}$tc(tctt)$_2$tctg(tctt)ttt(tctt)tc(tctt)$_{13}$tt(tctt)$_5$cct(tctt)tc(tctt)t(tctt)(23bp)P2(25bp) | |
| | 46 (336) | P1(24bp)(61bp) (tctt)$_{20}$tc(tctt)$_2$tctg(tctt)ttt(tctt)tc(tctt)$_{14}$tt(tctt)$_5$cct(tctt)tc(tctt)t(tctt)(23bp)P2(25bp) | |

Continued on next page

**Table 4.** Continued.

| Loci (ID in GenBank) | Allele (bp) | Sequences | Nomenclature |
|---|---|---|---|
| | 48 (344) | P1(24bp)(61bp) $(tctt)_{21}$tc$(tctt)_2$tctg(tctt)ttt(tctt)tc$(tctt)_{15}$tt$(tctt)_5$cct(tctt)tc(tctt)t(tctt)(23bp)P2(25bp) | |
| | 50 (352) | P1(24bp)(61bp) $(tctt)_{22}$tc$(tctt)_2$tctg(tctt)ttt(tctt)tc$(tctt)_{16}$tt$(tctt)_5$cct(tctt)tc(tctt)t(tctt)(23bp)P2(25bp) | |
| DYS714 [AC009233.3 (134512-134704)] | Consensus structure | P1(20bp)aatctttc$(ttttc)_m$(tcttc)$_2$(tttttc)$_2$(tcttc)$_2$(ttttc)$_2$tttttttttttaP2(23bp) | m+2+2+2+2 |
| | 18 (154) | P1(20bp)(8bp)$(ttttc)_{10}$(tcttc)$_2$(ttttc)$_2$(tcttc)$_2$(ttttc)$_2$(13bp)P2(23bp) | |
| | 19 (159) | P1(20bp)(8bp)$(ttttc)_{11}$(tcttc)$_2$(ttttc)$_2$(tcttc)$_2$(ttttc)$_2$(13bp)P2(23bp) | |
| | 20 (164) | P1(20bp)(8bp)$(ttttc)_{12}$(tcttc)$_2$(ttttc)$_2$(tcttc)$_2$(ttttc)$_2$(13bp)P2(23bp) | |
| | 21 (169) | P1(20bp)(8bp)$(ttttc)_{13}$(tcttc)$_2$(ttttc)$_2$(tcttc)$_2$(ttttc)$_2$(13bp)P2(23bp) | |
| | 22 (174) | P1(20bp)(8bp)$(ttttc)_{14}$(tcttc)$_2$(ttttc)$_2$(tcttc)$_2$(ttttc)$_2$(13bp)P2(23bp) | |
| DYS715 [NG_001197.3 (3143-3338)] | Consensus structure | P1(24bp)atggatggatagg$(taga)_m$taaagatgatatagatgtgatggatggatggatggatggatggatggatggatggaatagatgatagP2(24bp) | m |
| | 12 (180) | P1(24bp)(13bp)$(taga)_{12}$(71bp)P2(24bp) | |
| | 13 (184) | P1(24bp)(13bp)$(taga)_{13}$(71bp)P2(24bp) | |
| | 14 (188) | P1(24bp)(13bp)$(taga)_{14}$(71bp)P2(24bp) | |
| | 15 (192) | P1(24bp)(13bp)$(taga)_{15}$(71bp)P2(24bp) | |
| | 16 (196) | P1(24bp)(13bp)$(taga)_{16}$(71bp)P2(24bp) | |
| DYS716 [AC069323.5 (35539-35739)] | Consensus structure | P1(26bp)taaactcg$(cactc)_m$$(cattc)_n$catcctattctattgaactc(cattc)$(cactc)_2$(cattc)aattc(cattc)cactt(cattc)catttgatttctttctttcgP2(25bp) | m+n |
| | 15 (214) | P1(26bp)(8bp)$(cactc)_7$$(cattc)_8$(20bp)(cattc)$(cactc)_2$(cattc)(5bp)(cattc)(5bp)(cattc)(20bp)P2(25bp) | |
| | 16 (219) | P1(26bp)(8bp)$(cactc)_6$$(cattc)_{10}$(20bp)(cattc)$(cactc)_2$(cattc)(5bp)(cattc)(5bp)(cattc)(20bp)P2(25bp) | |
| | 17 (224) | P1(26bp)(8bp)$(cactc)_6$$(cattc)_{11}$(20bp)(cattc)$(cactc)_2$(cattc)(5bp)(cattc)(5bp)(cattc)(20bp)P2(25bp) | |
| | 18 (229) | P1(26bp)(8bp)$(cactc)_7$$(cattc)_{10}$(20bp)(cattc)$(cactc)_2$(cattc)(5bp)(cattc)(5bp)(cattc)(20bp)P2(25bp) | |
| | 19 (234) | P1(26bp)(8bp)$(cactc)_6$$(cattc)_{12}$(20bp)(cattc)$(cactc)_2$(cattc)(5bp)(cattc)(5bp)(cattc)(20bp)P2(25bp) | |
| DYS717 [AC010972.3 (27232-27458)] | Consensus structure | P1(21bp)tctcagagtatcatccctgaaggatggatggggaattcccaaggttatgataaagtgaaggcc$(tgtat)_2$tat$(tgtat)(tgtac)_m(tgtat)_n$ atatttagtggatacaagtgcagttatgttgcatgggtaP2(24bp) | 2+1+m+n |
| | 18 (240) | P1(21bp)(63bp)$(tgtat)_2$tat$(tgtat)(tgtac)_5(tgtat)_{10}$(39bp)P2(24bp) | |
| | 19 (245) | P1(21bp)(63bp)$(tgtat)_2$tat$(tgtat)(tgtac)_6(tgtat)_{10}$(39bp)P2(24bp) | |
| | 20 (250) | P1(21bp)(63bp)$(tgtat)_2$tat$(tgtat)(tgtac)_6(tgtat)_{11}$(39bp)P2(24bp) | |
| | 21 (255) | P1(21bp)(63bp)$(tgtat)_2$tat$(tgtat)(tgtac)_7(tgtat)_{11}$(39bp)P2(24bp) | |
| | 22 (260) | P1(21bp)(63bp)$(tgtat)_2$tat$(tgtat)(tgtac)_7(tgtat)_{12}$(39bp)P2(24bp) | |
| DYS718 [AC011289.4 (59287-59397)] | Consensus structure | P1(22bp)ataaacactattatctgtgttagtattatcttaggagaaaattcaatgcagttacctatttc$(tta)_m$tactttacgttttagggtacata tgcacaatatgcaggttagttacaaatgtataaatgtgccP2(19bp) | m |
| | 14 (208) | P1(22bp)(62bp)$(tta)_{14}$(63bp)P2(19bp) | |
| | 15 (211) | P1(22bp)(62bp)$(tta)_{15}$(63bp)P2(19bp) | |
| | 16 (214) | P1(22bp)(62bp)$(tta)_{16}$(63bp)P2(19bp) | |
| | 17 (217) | P1(22bp)(62bp)$(tta)_{17}$(63bp)P2(19bp) | |
| DYS719 [AC006989.3 (8685-8889)] | Consensus structure | P1(18bp)ttaggagatatatctaatgctaaatgacgagttaatgggtgcagcataccagcatgacacatgtatacatatgtaactaacctgcacattgtgc acatgtaccctaaaacttaaagt$(ata)_m$aaaagaaataggtaP2(19bp) | m |
| | 12 (204) | P1(18bp)(117bp) $(taa)_{12}$(14bp)P2(19bp) | |
| | 13 (207) | P1(18bp)(117bp) $(taa)_{13}$(14bp)P2(19bp) | |
| | 14 (210) | P1(18bp)(117bp) $(taa)_{14}$(14bp)P2(19bp) | |
| | 15 (213) | P1(18bp)(117bp) $(taa)_{15}$(14bp)P2(19bp) | |
| | 16 (216) | P1(18bp)(117bp) $(taa)_{16}$(14bp)P2(19bp) | |
| DYS721 [AC011289.4 (70142-70432)] | Consensus structure | P1(21bp)tctcagaaaaaaaaagaaaagaa$(aaggg)_m$aaaggaagag$(aaggg)_2$aaaggag$(aagca)_n$aagatgaaaaacacaaacatt aactttgcaattccaaatggttgaaatcaaatcccttcagtttatctgggactcattaaaaactgttagaattttagatgaaatcgactaggctggccaP2(20bp) | m+2+n |
| | 17 (286) | P1(21bp)(23bp)$(aaggg)_9$(10bp)$(aaggg)_2$(7bp)$(aagca)_6$(120bp)P2(20bp) | |

Continued on next page

**Table 4.** Continued.

| Loci (ID in GenBank) | Allele (bp) | Sequences | Nomenclature |
|---|---|---|---|
| | 18 (291) | P1(21bp)(23bp)(aaggg)$_{10}$(10bp)(aaggg)$_2$(7bp)(aagca)$_6$(120bp)P2(20bp) | |
| | 19 (294) | P1(21bp)(23bp)(aaggg)$_{11}$(10bp)(aaggg)$_2$(7bp)(aagca)$_6$(120bp)P2(20bp) | |
| | 20 (299) | P1(21bp)(23bp)(aaggg)$_{12}$(10bp)(aaggg)$_2$(7bp)(aagca)$_6$(120bp)P2(20bp) | |
| | 21 (304) | P1(21bp)(23bp)(aaggg)$_{13}$(10bp)(aaggg)$_2$(7bp)(aagca)$_6$(120bp)P2(20bp) | |
| DYS722 [AC011289.4 (72629-72898)] | Consensus structure | P1(25bp)aatgttaaacatccaaaaaaaaaagaaggaagaaaggaagaagaac(gaaa)$_m$aaga(gaaa)$_2$a(gaaa)$_2$gaga(gaaa)$_2$ gctgaacacagtattaaggaaacaacttgaatgaatttctgagtttggaattgcaatcttgtaggtctcagacaacattgctP2(18bp) | m+2+2+2 |
| | 19 (258) | P1(25bp)(47bp)(gaaa)$_{13}$aaga(gaaa)$_2$a(gaaa)$_2$gaga(gaaa)$_2$(83bp)P2(18bp) | |
| | 20 (262) | P1(25bp)(47bp)(gaaa)$_{14}$aaga(gaaa)$_2$a(gaaa)$_2$gaga(gaaa)$_2$(83bp)P2(18bp) | |
| | 21 (266) | P1(25bp)(47bp)(gaaa)$_{15}$aaga(gaaa)$_2$a(gaaa)$_2$gaga(gaaa)$_2$(83bp)P2(18bp) | |
| | 22 (270) | P1(25bp)(47bp)(gaaa)$_{16}$aaga(gaaa)$_2$a(gaaa)$_2$gaga(gaaa)$_2$(83bp)P2(18bp) | |
| | 23 (274) | P1(25bp)(47bp)(gaaa)$_{17}$aaga(gaaa)$_2$a(gaaa)$_2$gaga(gaaa)$_2$(83bp)P2(18bp) | |
| DYS723 [AC011289.4 (76324-76517)] | Consensus structure | P1(21bp)tagacatagtagaa(gata)$_2$tat(gata)$_m$gat(gata)$_2$gat(gata)$_2$taacagatgaatgg attaataaataaagagcatagatggatgcataaatatP2(20bp) | 2+m+1+7 |
| | 20 (194) | P1(21bp)(13bp)(gata)$_2$tat(gata)$_{10}$gat(gata)$_2$gat(gata)$_2$(51bp)P2(20bp) | |
| | 21 (198) | P1(21bp)(13bp)(gata)$_2$tat(gata)$_{11}$gat(gata)$_2$gat(gata)$_2$(51bp)P2(20bp) | |
| | 22 (202) | P1(21bp)(13bp)(gata)$_2$tat(gata)$_{12}$gat(gata)$_2$gat(gata)$_2$(51bp)P2(20bp) | |
| | 23 (206) | P1(21bp)(13bp)(gata)$_2$tat(gata)$_{13}$gat(gata)$_2$gat(gata)$_2$(51bp)P2(20bp) | |
| | 24 (210) | P1(21bp)(13bp)(gata)$_2$tat(gata)$_{14}$gat(gata)$_2$gat(gata)$_2$(51bp)P2(20bp) | |
| S726 [AC134879.3 (146076-146273)] | Consensus structure | P1(22bp)aatttagggaacactaacacctcagaaaaactctgtaaa(cttc)$_2$ctttttctccctccc(cttc)$_m$ttt(cttc)cctctcctctctct(cttt)$_3$t(cttt)$_2$gttcatttttP2(23bp) | m |
| | 11 (205) | P1(22bp)(39bp)(cttc)$_3$(13bp)(cttc)$_{11}$ttt(cttc)(14bp)(cttt)$_3$t(cttt)$_2$(10bp)P2(23bp) | |
| | 12 (209) | P1(22bp)(39bp)(cttc)$_3$(13bp)(cttc)$_{12}$ttt(cttc)(14bp)(cttt)$_3$t(cttt)$_2$(10bp)P2(23bp) | |
| | 13 (213) | P1(22bp)(39bp)(cttc)$_3$(13bp)(cttc)$_{13}$ttt(cttc)(14bp)(cttt)$_3$t(cttt)$_2$(10bp)P2(23bp) | |
| | 14 (217) | P1(22bp)(39bp)(cttc)$_3$(13bp)(cttc)$_{14}$ttt(cttc)(14bp)(cttt)$_3$t(cttt)$_2$(10bp)P2(23bp) | |

## Complex STR

Complex STR are those loci that contain more than one type of repeat-unit with/or without interruptions. DYS711, DYS713 and DYS717 comprised two variant motifs, which share the same repeat-unit, with/without the adjacent non-variant blocks. The aggregate number of the repeated unit in the interrupted variant motif and non-variant blocks was used for the allele name. With the repeat-unit of (CTT) for DYS711, (TCTT) for DYS713 and (TGAG) for DYS717, the variant and non-variant portions that comprised the allele diversity are listed in Tables 1 and 4. DYS708, DYS710, DYS712, DYS716 are more complicated STRs because they consist of different repeat-unit with (DYS710, DYS716) or without (DYS708, DYS712) interruptions. These four STRs comprise different repeat-unit with identical size, i.e., DYS708 consists of tetranucleotide variable repeat-unit of (GATA) and (GACA); DYS712 consists of repeated (AGAT) and (AGAC); DYS710 is comprised by three variant motifs $(GAAA)_m(GAGA)_nGA(GAAA)_q$, m/n/q stands for 16-17, 6-8 and 11-14, respectively; DYS716 includes two pentanucleotide motifs of (CACTC) and (CATTC). Since the number of non-variant repeated units are less than three and separated by 20 nucleotides from the core variable motif in DYS726, they are excluded in the allele nomenclature. Alleles of DYS710 and DYS712 are named simply by the aggregation number of variable different repeat-unit. Non-variant repeats of DYS708 are considered for the allelic name because they are adjacent to the core variable motif blocks. We summarized the sequences comprising allele diversity and nomenclature of alleles by number of variable and non-variable repeat-unit in Table 4.

## Haplotypes and male specificity of 16 distinct STRs

A total of 120 distinct haplotypes consisting of the 16 Y-STRs are listed in Table 3. Since the haplotype of each male is unique, the unrelated males can be identified. The specificity for human males was determined by comparing DNA typing from 20 females and animals including pig, dog, chicken, rabbit, and rat. No PCR products were found using our primers and PCR parameters.

## Y-STRs failed to analysis

One typical Y-STR should show only one band when analyzed by PAGE due to its Y-chromosome-specific character. In our data, DYS724 and DYS725 showed more than two bands in their PCR products. When the products of DYS724 were visualized on PAGE, 32 individuals showed one band, 68 showed two bands, and 20 showed three bands. Besides, the polymorphism of DYS724 was limited, so we did not investigate it further. DY725 also showed two bands and non-polymorphism in our population (data not shown). The information of DYS720 was lacking because we did not find suitable primers and amplification conditions for this STR. Due to the above-mentioned reasons, we did not conduct genotyping of these 3 Y-STRs in our collected samples.

# DISCUSSION

In the present study, we report the detailed information on allele frequencies and struc-

tures of novel Y-STRs deposited in Genome databank, which were named DYS708 to DYS726. They were successfully amplified and analyzed by designed primers and sequencing, except for DYS720, DYS724 and DYS725. The multiple bands may stem from duplication or multiple copies on the chromosomes. The Y-STRs with variable multiple bands mean that they are likely to be more informative than those STRs containing only one band, just like DYS389 and DYS385 (Komuro et al., 1998; Rolf et al., 1998; Niederstatter et al., 2005). However, multi-copies Y-STR would complicate the analysis of materials from sexual assault cases in forensic applications, such as the determination of the number of male suspects involved in the case. As a result, we do not suggest the two STRs, DYS724 and DYS725, to be involved into future Y-STR genotyping system because of their irregular allele numbers. In our population data, DYS712 had the highest GD, i.e., 0.843. Seven Y-STRs possessed a GD over 0.80, including DYS708, DYS711, DYS712, DYS713, DYS714, DYS716, and DYS723. Generally speaking, the alleles are relatively evenly distributed in this population since no allele frequency over 0.5 was found. These 16 Y-STR provide a highly haplotype diversity and would be more powerful if they are applied in groups or incorporated into commercial multiplex kits. Although the allele diversity and haplotypes of these STRs are not reported in other populations, the highly polymorphic could be expected in populations with similar genetic background, such as in Asia populations. Because Y-chromosome genetic markers are highly specific in a population group and dwelling region, the polymorphic data need to be studied in other ethnic populations. In fact, allele frequency information is one key discrepancy in different ethnic populations (Jarve et al., 2009).

The STRs were classified as simple STRs and complex STRs based on their structures in order to facilitate our analysis. Such classification may be changed in the future if more complex structure are found in other populations, just like the D21S11 (Urquhart et al., 1994). Even so, simple STRs are always preferred in future lineage analysis for their unambiguous typing. We named each STR allele according to ISFG's rules, as listed in Table 4. How to interpret repeated motifs is crucial to name an allele. One principle that should be followed is that allele name should have a mathematical relationship to the allele length. The most majority of Y-STR in this study were comprised of repeat-unit with the same size, except for DYS710 and DYS711. In this study, DYS710 was regarded as a tetranucleotide STR consisting of $(GAAA)_m(GAGA)_nGA(GAAA)_q$ instead of $(GAAA)_m(GA)_n(GAAA)_q$ because the sequences showed allele diversity with integral number of (GAGA) in this population. DYS711 was considered as a trinucleotide STR although it was composed by (CTT), (CTC) and (CTCCTT), because only $(CTT)_m$ and $(CTT)_n$ are related to the allele length.

To sequence each sample for each STR is an unacceptable huge work. However, compared with fully sequencing, high through-put automatic fragment analysis in forensic casework would compromise the STR polymorphism, because alleles with equivalent length but consisting of different sequences cannot be distinguished, particularly in those complex STR that have different repeat-units. To better understand the properties and sequence diversity of these STR for future genetic applications, thorough investigations in other populations are needed.

It should be noted that the primers of DYS709, SYS719, DYS722, and DYS723 have a good BLAST value of Pan troglodytes in Genbank. Although we did specificity test in pigs, dogs, chicken, rabbit, and rats, the species specificity of these STR in other species should be addressed in future studies, especially for forensic applications.

In summary, we identified 16 one-copy Y-STRs and 2 multi-copies Y-STRs with the allele diversities and structures in human males in the Chinese population. These Y-STRs, es-

pecially tetra- and penta-nucleotide Y-STRs, such as DYS709, DYS714, DYS715, DYS716, DYS718, DYS719, and DYS726 are suggested to be integrated in the multiplex systems after fully investigations in other populations and serve as the new members for forensic and genetic applications.

## Limitations of this study

As a preliminary screening study for suitable Y-STRs for forensic purposes, it did not involve a substantial population number and different ethnicities.

## ACKNOWLEDGMENTS

Research supported by the Shanxi Natural Science Foundation (#20051046, #200911055-2) and the National Natural Science Foundation of China (#30900593).

## REFERENCES

Budowle B, Ge J, Low J, Lai C, et al. (2009). The effects of Asian population substructure on Y STR forensic analyses. *Leg. Med.* 11: 64-69.

Dai HL, Wang XD, Li YB, Wu J, et al. (2004). Characterization and haplotype analysis of 10 novel Y-STR loci in Chinese Han population. *Forensic Sci. Int.* 145: 47-55.

Ge J, Budowle B, Planz JV, Eisenberg AJ, et al. (2010). US forensic Y-chromosome short tandem repeats database. *Leg. Med.* 12: 289-295.

Gill P, Brenner C, Brinkmann B, Budowle B, et al. (2001). DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs. *Int. J. Legal Med.* 114: 305-309.

Gusmao L, Butler JM, Carracedo A, Gill P, et al. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci. Int.* 157: 187-197.

Jarve M, Zhivotovsky LA, Rootsi S, Help H, et al. (2009). Decreased rate of evolution in Y chromosome STR loci of increased size of the repeat unit. *PLoS One* 4: e7276.

Kayser M, Kittler R, Erler A, Hedman M, et al. (2004). A comprehensive survey of human Y-chromosomal microsatellites. *Am. J. Hum. Genet.* 74: 1183-1197.

Komuro T, Tsutsumi H, Mukoyama R and Nakamura M (1998). Repeat structure of DYS389 locus. *Nihon Hoigaku Zasshi* 52: 227-232.

Nei M and Tajima F (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145-163.

Niederstatter H, Berger B, Oberacher H, Brandstatter A, et al. (2005). Separate analysis of DYS385a and b versus conventional DYS385 typing: is there forensic relevance? *Int. J. Legal Med.* 119: 1-9.

Perez-Lezaun A, Calafell F, Comas D, Mateu E, et al. (1999). Sex-specific migration patterns in Central Asian populations, revealed by analysis of Y-chromosome short tandem repeats and mtDNA. *Am. J. Hum. Genet.* 65: 208-219.

Rolf B, Meyer E, Brinkmann B and de KP (1998). Polymorphism at the tetranucleotide repeat locus DYS389 in 10 populations reveals strong geographic clustering. *Eur. J. Hum. Genet.* 6: 583-588.

Urquhart A, Kimpton CP, Downes TJ and Gill P (1994). Variation in short tandem repeat sequences - a survey of twelve microsatellite loci for use as forensic identification markers. *Int. J. Legal Med.* 107: 13-20.

Zhang GQ, Wang Y, Zhang YX, Xu XL, et al. (2004). Study of polymorphism at new Y-STR DYS605 in a Chinese Han population of Shanxi. *Yi Chuan* 26: 295-297.