

## A New Characterized Y-STR and Its Allele Frequencies Distribution in 8 Chinese Populations

Yunliang Zhu, Guangzheng Zhang, Zhaoshu Zeng, Xiufeng Ge, Kemin Guo, Shuhong Zhang

Department of Forensic Medicine, Basic Medical College, Zhengzhou University,  
Zhengzhou, Henan 450052, China

**Abstract:** Y chromosome genomic DNA was used to search for new Y-specific short tandem repeats (Y-STRs) and a new Y-STR, DYS708, has been characterized. In 8 different Chinese populations, including Guangdong Han, Henan Han, Bai, Tibetan, Uygur, Tujia, Mongolia and Zhuang, 9 successive alleles have been found by PCR-based method. The sequences of different alleles showed that this locus included four blocks of AGAT and two blocks of AGAC and that length variations between existed not only in the largest block of AGAT, but also in the larger block of AGAC repeats. The name of each allele was the sum of the repeats number of these six repeats blocks ranging from 22 to 30 according to ISFG. The gene diversity was ranged between 0.7659 in Uygur and 0.6327 in Bai. Based on Nei's genetic distance, the genetic tree of these 8 populations was constructed. It was showed in the tree that Guangdong Han was closer to Bai than to Henan Han indicating gene flow between Guangdong Han and local minority, and Tujia was closer to Henan Han than to Guangdong Han implying that Tujia might come from northern part of China. These results agreed with the conclusion based on other genetic markers. All of the results indicated that DYS708 is a useful genetic marker and could be applied to human evolution study and forensic science. [Life Science Journal. 2006;3(2):61-65] (ISSN: 1097-8135).

**Keywords:** Y chromosome; short tandem repeats; gene diversity; human evolution; personal identification

**Abbreviations:** ISFG: the International Society for Forensic Genetics; NRY: non-recombining region of Y chromosome; PCR: polymerase chain reaction; Y-STRs: short tandem repeats on Y chromosome

### 1 Introduction

Human Y-STRs are tandemly repeated arrays of two to six base-pair units on NRY (also know as MSY, man specific region of Y chromosome). As genetic markers, just as its counterpart on the rest of chromosomes, Y-STRs are abundance, high degree of polymorphism and ease of scoring. It exists only in male and is transmitted only from father to son as a haplotype making it a useful tool in diverse fields. In forensic science<sup>[1]</sup>, to identify a male suspect, there is no need to separate the male part of sample mixtures derived from male and female. Y-STRs also can be used to identify male lineage<sup>[2]</sup>, which is difficult to use autosomal STRs. In anthropology Y-STRs can be used to trace male evolution<sup>[3-5]</sup>, just as its counterpart, mitochondria DNA, in female evolution. Some diseases have been found related to some haplotypes defined by Y-STRs<sup>[6]</sup> and other markers. Although many Y-STR have been reported, the additional well-characterized Y-STR is still necessary in order to increase the discriminating power and the chance of exclusion in forensic science and to get high-resolution haplotypes of Y chromosome for its full use in

human evolution and human genetics. The allele distribution of Y-STR is subject to natural selection, selective sweep, genetic draft, bottleneck events, population expansion and migration, so different population usually has different allele frequency distribution. This distribution is the basis of its use in many fields especially in forensic science. Here we described a new Y-STR, DYS708, which was identified from genome sequence of NRY. We also investigated its allele distribution in eight Chinese populations and constructed the genetic tree based on these genetic data.

### 2 Materials and Methods

#### 2.1 Blood samples and DNA extractions

Blood samples of Henan Han (95 individuals) were collected in Henan province of China, while that of Guangdong Han (175 individuals) in Guangdong province of China, Tujia (60 individuals) in Hunan province of China, Zhuang (31 individuals) in Guangxi Zhuang Autonomous Region of China, Uygur (104 individuals) in Xinjiang Autonomous Region of China, Mongolia (49 individuals) in Inner Mongolia Autonomous Region of China, Tibetan (52 individuals) and Bai (138 individ-

uals) in Yunnan province of China. The ethnic group to which the blood donor belongs was determined according to the self claim of the blood donor. DNA was extracted using the Chelex 100 and proteinase K protocol<sup>[7]</sup>.

## 2.2 Identification of novel Y-STR and PCR primer

Y-chromosomal DNA sequence data were obtained from GenBank (AC011298). Sequences were downloaded in FASTA format and used as input for the program Tandem Repeats Finder<sup>[8]</sup>. A Microsatellite with a 4-bp motif repeating 11 times was chosen from the output of this program as our research object. Primers were designed using Primer3 software. Unlabelled primers were synthesised by SBSBio Company (Beijing, China).

## 2.3 Amplification conditions and analysis of PCR products

The PCR was performed in a 20  $\mu$ l final volume containing 1  $\times$  Taq<sup>TM</sup> buffer (10 mmol/L Tris-HCl, pH 8.3, 1.5 mmol/L MgCl<sub>2</sub>, 50 mmol/L KCl), 200  $\mu$ mol/L dNTPs, 30 – 50 ng DNA, 1 U Taq<sup>TM</sup> (TaKaRa Biotechnology Co., Ltd, Dalian, Liaoning, China). Primer sequences were: left primer AGTGTATCCGCCATGGTAG C, right primer CTGCATTTTGGTACCCATA. In the TouchDown PCR protocol the DNA was initially denatured at 94  $^{\circ}$ C for 2 min. This was followed by 15 cycles starting at 94  $^{\circ}$ C for 0.5 min, 63  $^{\circ}$ C for 40 sec and 72  $^{\circ}$ C for 45 sec. The annealing temperature was decreased by 0.5  $^{\circ}$ C in each cycle. It was then followed by 20 cycles at 94  $^{\circ}$ C for 0.5 min, 55  $^{\circ}$ C for 40 sec and 72  $^{\circ}$ C for 45 sec. After a final extension step at 72  $^{\circ}$ C for 5 min the samples were kept at 4  $^{\circ}$ C until electrophoresis.

The amplification products (1  $\mu$ l) were electrophoresized on 6% polyacrylamide gels (PAG) in 1 $\times$  TBE buffer and the DNA bands were detected by silver stain<sup>[9]</sup>. The DNA bands at different length were eluted from gel and reamplified with the conditions previously described. The reamplified products were sequenced by BIOASIA Company.

## 2.4 Typing and nomenclature

Allele assignment of amplified respective DNA fragments was performed by side-to-side comparison with the allelic ladder consisting of sequenced alleles. The suggested repeat nomenclature of alleles follows the guidelines of the DNA commission of

the International Society for Forensic Genetics (ISFG)<sup>[10]</sup>. Two female samples were included as negative controls.

## 2.5 Statistical analysis

Gene frequencies of alleles were obtained by simple gene counting. Gene diversity and standard errors were calculated following Nei<sup>[11]</sup>. MEGA version 3.0<sup>[12]</sup> was used to construct genetic tree (neighbor-joining method<sup>[13]</sup>) based on  $D$  (genetic distance) and  $D$  equal to  $-\ln(J_{XY}/\sqrt{J_X J_Y})$  where  $J_{XY} = \sum x_i y_i$ ,  $J_X = \sum x_i^2$  and  $J_Y = \sum y_i^2$ .  $x_i$  and  $y_i$  are the frequencies of  $i$ th allele in population X and population Y respectively<sup>[14]</sup>.

## 3 Results

### 3.1 The repeat structure and nomenclature of alleles

DYS708 (dbSTS-Id: 340593; GenBank-Ac-cn: BV209666) primers amplified a DNA fragment specifically in male and no band was found in female. The amplified products have a sequence structure of AGTGTATCCGCCATGGTAGCATA ATAGAAATTTTATGAGTGGGAGAAATGGAT GACAGTAAAATGAAAACATTGCAATGTGTA TACTCAGAAACAAGGA (AGAT) $m$ GAT(A-GAT) $n$ (AGAC) $o$ (AGAT) $p$ (AGAC) $q$ (AGAT) $r$ AGAATATATTATGGGGTACCAAATGCAG (for the reference sequence in genbank, AC011289,  $m=3$ ,  $n=3$ ,  $o=1$ ,  $p=11$ ,  $q=8$  and  $r=2$ ). We have found 9 different alleles and at least one present sample from each allele was sequenced. There were no difference in  $m$ ,  $n$ ,  $o$  and  $r$  between the sequenced alleles, but  $p$  could be 6 to 13 and  $q$  could be 7, 8 or 9. The structures of sequenced alleles were summarized in Table 1. The suggested repeats nomenclature of alleles was the sum of  $m$ ,  $n$ ,  $o$ ,  $p$ ,  $q$  and  $r$  according to guidelines of the DNA Commission of the International Society for Forensic Genetics (ISFG).

### 3.2 Allele frequency distribution and gene diversity (GD)

In total 9 alleles have been found. 8 out of 9 alleles have been found in Henan Han population and only 4 alleles were found in each of Mongolia and Zhuang populations. The gene diversity ranged between 0.7659 (Uygur) and 0.6327 (Bai). The allele frequency and gene diversity were showed in Table 2.

**Table 1.** The structures of repeat blocks of DYS708 alleles

Allele (bp)	Repeat motif
23 (223)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)6(AGAC)8(AGAT)229bp
24 (227)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)7(AGAC)8(AGAT)229bp
25 (231)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)8(AGAC)8(AGAT)229bp 99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)9(AGAC)7(AGAT)229bp
26 (235)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)9(AGAC)8(AGAT)229bp
27 (239)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)10(AGAC)8(AGAT)229bp 99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)9(AGAC)9(AGAT)229bp
28 (243)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)11(AGAC)8(AGAT)229bp
29 (247)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)12(AGAC)8(AGAT)229bp
30 (251)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)13(AGAC)8(AGAT)229bp
31 (254)	99bp(AGAT)3GAT(AGAT)3(AGAC)1(AGAT)16(AGAC)8(AGAT)229bp

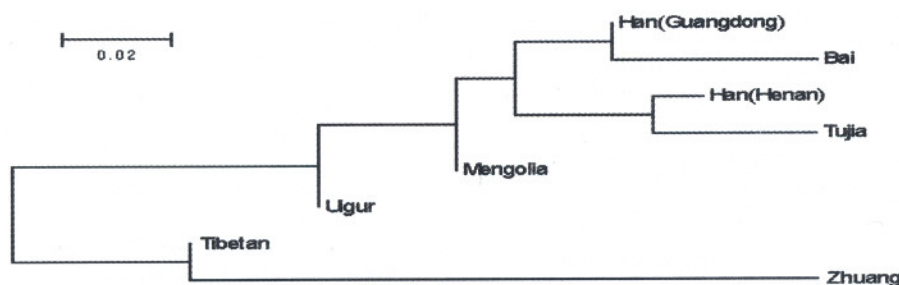
**Table 2.** The allele frequency and gene diversity in 8 Chinese populations

Allele (bp)	Guangdong Han	Henan Han	Bai	Tibetan	Uyгур	Tujia	Mongolia	Zhuang
23 (223)	0	0.0105	0	0	0	0	0	0
24 (227)	0	0.0105	0.0072	0	0.0192	0	0	
25 (231)	0.0057	0.0105	0	0	0.0096	0	0	0
26 (235)	0.0743	0.0948	0.0072	0.0385	0.1154	0.0167	0.0816	0.0645
27 (239)	0.4172	0.3263	0.5290	0.3077	0.3365	0.4333	0.3878	0.129
28 (243)	0.2743	0.4	0.2464	0.2692	0.2596	0.4	0.3265	0.4194
29 (247)	0.2114	0.1369	0.1739	0.3654	0.2115	0.0667	0.2041	0.3871
30 (251)	0.0171	0.0105	0.029	0.0192	0.0481	0.0833	0	0
31 (255)	0	0	0.0072	0	0	0	0	0
	GD:0.7042	GD:0.7128	GD:0.6327	GD:0.7111	GD:0.7659	GD:0.6514	GD:0.7091	GD:0.6752
	SE:0.0123	SE:0.0179	SE:0.0208	SE:0.0151	SE:0.0127	SE:0.0238	SE:0.0191	SE:0.0301
	n:175	n:95	n:138	n:52	n:104	n:60	n:49	n:31

**3.3 The genetic tree of populations**

Based on Nei's genetic distance, the genetic

tree of these 8 populations was constructed (Figure 1).



**Figure 1.** Genetic tree of 8 Chinese populations

**4 Discussion**

Human genome sequence data make it convenient to find novel polymorphic Y-STR just as did by Ayub<sup>[15]</sup> and Redd<sup>[16]</sup>. The euchromatin of Y chromosome is about 23 megabases (Mb) much of which contains sequences that shared with the X

and other chromosomes or repeated elsewhere on itself<sup>[17]</sup>, so the sequences on which Y-STR are searched for should be carefully selected in order to amplify Y specific sequence.

DYS708 could be classified as compound repeats since it consisted of two different motifs. The repeat region comprised four stretches of AGAT

and two stretches of AGAC. The variations of repeat number existed in the third block of AGAT repeats and the second block of AGAC repeats, but the former was more variable. Our nomenclature was based on the total number of repeats including six repeat blocks according to the recommendations of ISFG, not taking the sequence of GAT following the first block of AGAT into account, although GAT might be a AGAT motif having a deletion of A.

Compared with other reported Y-STRs, *DYS708* was a highly polymorphic Y-STR locus, at least in the populations we have investigated. That Uygur population had highest gene diversity might be due to gene flow along the history of the Silk Road. Since Xinjiang Uygur Autonomous Region is close to Central Asia, the result that Uygur population had highest gene diversity agreed with the conclusion made by Wells that Central Asia is an important reservoir of gene diversity<sup>[18]</sup>. From the genetic tree, some conclusions could be come to. First, Guangdong Han was closer genetically to Bai than to Henan Han, implying gene flow between Han and local minority. Second, Tujia, inhabiting in southern part of China, was closer genetically to Henan Han (northern part of China) than to Guangdong Han (southern part of China), indicating that Tujia might come from northern part of China. All of these results agreed well with the conclusion based on other genetic markers (red blood group, HLA, red blood enzyme, etc)<sup>[19]</sup>. Contrary to the previous findings, Zhuang lay in a separate branch with Tibetan, the reasons for which was not clear and need further investigation.

Since Y chromosome is transmitted from father to son as a haplotype, its usefulness should be viewed as haplotype diversity. Although further investigations of its properties, such as its haplotype diversity with other Y chromosome genetic markers and its mutation rate, should be made before its full use in many fields, we still anticipated that this STR is a useful genetic marker and can be used in combination with other genetic markers in the studies on human evolution, the association study of some diseases and forensic science.

#### Correspondence to:

Yunliang Zhu

Department of Forensic Medicine

Basic Medical College

Zhengzhou University

Zhengzhou, Henan 450052, China

Telephone: 86-371-6665-8173 (office)

Email: zhuyunliang@hotmail.com

#### References

1. Kayser M, Caglia 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, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Roewer L. Evaluation of Y-chromosomal STR: a multicenter study. *Int J Legal Med* 1997; 110(3): 125–33.
2. Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet R, Zerjal T, Tyler-Smith C. Jefferson fathered slave's last child. *Nature* 1998; 396(6706):27–8.
3. Ruiz Linares A, Nayar K, Goldstein DB, Hebert JM, Seielstad MT, Underhill PA, Lin AA, Feldman MW, Cavalli Sforza LL. Geographic clustering of human Y-chromosome haplotypes. *Ann Hum Genet* 1996; 60: 401–8.
4. Zerjal T, Dashnyam B, Pandya A, Kayser M, Roewer L, Santos FR, Schiefenhover W, Fretwell N, Jobling MA, Harihara S, Shimizu K, Semjiddmaa D, Sajantila A, Salo P, Crawford MH, Ginter EK, Evgrafov OV, Tyler-Smith C. Genetic relationships of Asians and Northern Europeans, revealed by Y-chromosomal DNA analysis. *Am J Hum Genet* 1997; 60(5): 1174–83.
5. de Knijff P, Kayser M, Caglia 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, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Roewer L. Chromosome Y microsatellites: population genetic and evolutionary aspects. *Int J Legal Med* 1997; 110(3):134–49.
6. Jobling MA, Tyler-Smith C. New use for new haplotypes. *Trends in Genetics* 2000; 16(8):356–62.
7. Walsh PS, Metzger DA, Higuchi R. Chelex 100 as a medium for the simple extraction of DNA for PCR-based typing from forensic materials. *Biotechniques* 1991; 10(4):506–13.
8. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* 1999; 27(2):573–80.
9. Budowle B, Chakraborty R, Giusti AM, Eisenberg AJ, Allen RC. Analysis of the VNTR locus *DIS80* by the PCR followed by high-resolution PAGE. *Am J Hum Genet* 1991; 48(1):137–44.
10. Gill P, Brenner C, Brinkmann B, Budowle B, Carracedo A, Jobling MA, de Knijff P, Kayser M, Krawczak M, Mayr WR, Morling N, Olaisen B, Pascali V, Prinz M, Roewer L, Schneider PM, Sajantila A, Tyler-Smith C. DNA commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STR. *Int J Legal Med* 2001; 114(6):30–9.
11. Nei M. *Molecular Evolutionary Genetics*. Columbia University Press, New York 1987; 176–81.
12. Kumar S, Tamura K, Nei M. MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. *Briefings in Bioinformatics* 2004; 5(2):150–63.
13. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol*

- Evol 1987; 4 (4): 406–26.
14. Nei M. *Molecular Population Genetics and Evolution*. North-Holland Publishing Company, Oxford 1975:177.
  15. Ayub Q, Mohyuddin A, Qamar R, Mazhar K, Zerjal T, Mehdi SQ, Tyler-Smith C. Identification and characterization of novel human Y-chromosomal microsatellites from sequence database information. *Nucleic Acids Res* 2000; 28(2):E8–e8.
  16. Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, de Knijff P, Butler J M, Hammer M F. Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci Int* 2002; 130(2–3):97–111.
  17. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, Chinwalla A, Delehaunty A, Delehaunty K, Du H, Fewell G, Fulton L, Fulton R, Graves T, Hou SF, Latrielle P, Leonard S, Mardis E, Maupin R, McPherson J, Miner T, Nash W, Nguyen C, Ozersky P, Pepin K, Rock S, Rohlfling T, Scott K, Schultz B, Strong C, Tin-Wollam A, Yang SP, Waterston RH, Wilson RK, Rozen S, Page DC. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 2003; 423(6942):825–37.
  18. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamar-jashvili E, Gambarov S, Nikbin B, Dostiev A, Ak-nazarov O, Zalloua P, Tsoy I, Kitaev M, Mirrakhimov M, Chariev A, Bodmer WF. The Eurasian heartland: a continental perspective on Y-chromosome diversity. *Proc Natl Acad Sci USA*. 2001;98(18):10244–9.
  19. Du RF, Xiao CJ. The genetic origin and evolution of Chinese population. *Chinese Social Science* 1997; 4: 139–46.

*Received March 21, 2006*