A Potential Marker DYF371 for Differentiating Han Population from Non-Han Population in Chinese

Hailun Nan, Dejian Lu
Faculty of Forensic Medicine, Zhongshan Medical School, Sun Yat-Sen University, Guangzhou, China

Abstract
DYF371 locus and single-nucleotide polymorphisms (SNPs) defined as DYF371C or DYF371T were simultaneously examined for 142 Chinese unrelated male individuals, including 90 Han population and 52 minorities. In total, 37 SNP-short tandem repeat haplotypes were detected. Nine of these 37 haplotypes were unique (24.22%). The gene diversity and discrimination capacity values are 0.8321 and 0.2606, respectively. Furthermore, only five males (5.55%) in the Han samples were determined that they had one copy of allele with SNP T-type and three copies of alleles with SNP C-types, whereas 14 individuals (26.92%) were observed in minorities’ samples. The genotype proportion comprising three distinguishing copies of the allele with SNP C-types in Han sample was greatly lower than that in the sample of minorities. Similar results were shown for allele 11C and 10C. Therefore, the alleles of DYF371 with SNP C-type have the potential to differentiate the Han and non-Han Chinese populations.

Keywords: DYF371, multi-copy, multiplex polymerase chain reaction, single-nucleotide polymorphism, polymorphism-short tandem repeat

Introduction
Y chromosome is different from autosome since the characteristic of Y chromosome is non-recombination in most parts of it.[1] Therefore, the genetic diversity of the Y chromosome is relatively low.

The Y chromosome short tandem repeats (STRs) located on Y chromosome palindrome P1 and P5 [Supplement Figure 1]. Single-nucleotide polymorphism (SNP) sites rs1250403754 and rs1556285687 are found in the flanking regions of DYF371. Therefore, the alleles of DYF371 locus have two SNP types which are T-type or C-type [Supplement Figure 1]. Recently, several papers have reported gene diversity for DYF371 locus in different population groups. [4-7] Their data demonstrated DYF371 locus had the potential for identifying male individuals and analyzing genetic genealogy in forensic field. However, few of the allele frequencies for DYF371 have been investigated in Chinese population. In addition, the SNP-types had not been detected previously. The distribution of haplotypes of SNP-STRs is still unclear. Therefore, in this study, we will explore the length polymorphism of DYF371 and its flanking SNP types in the Chinese Han population and minority groups.

Materials and Methods
DNA samples
Bloodstain samples were collected from 142 Chinese males. The samples included 90 individuals from Han population and 52 unrelated males from nine minority groups: Kazakh (5), Mongolian (4), Hui (4), Yi (5), Shui (5), Tujia (5), Zhuang (5), and Tibetan (20). The study was approved by the Human Subjects Committee at Zhongshan School of Medicine, Sun Yat-Sen University. Informed consent was obtained from the participants. Ethical clearance was obtained from the Ethics Committee of Sun Yat-Sen University.

Address for correspondence:
Dr. Dejian Lu,
Faculty of Forensic Medicine, Zhongshan Medical School, Sun Yat-Sen University, 74# Zhongshan Road II, Guangzhou, 510089, China.
E-mail: lvdejian@mail.sysu.edu.cn

Submitted: 01-Jan-2020 Revised: 24-Feb-2020 Accepted: 24-Feb-2020 Published: 17-Mar-2020

This is an open access journal, and articles are distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as appropriate credit is given and the new creations are licensed under the identical terms.

For reprints contact: reprints@medknow.com

How to cite this article: Nan H, Lu D. A potential marker DYF371 for differentiating Han population from non-Han population in Chinese. J Forensic Sci Med 2020;6:1-4.
Multiplex polymerase chain reaction amplification

The SNP-STR genotypes DYF371T-type and DYF371C-type were analysed using the primers described in the previous study.[8] The forward primer sequence was 5'- AGTAATTCTGGAGGTAAAATGG3'. The backward primer sequence for DYF371C-type was: ROX-5'-TGGAGAGAAGAAGAGAGAC-3', and for DYF371T-type was: FAM-5'-TGGAGAGAAGAAGAGAGAAAT-3' [Figure 1]. The multiplex PCR amplification was carried out using a 1.2 mm punched bloodstain on filter paper, without DNA extraction. PCR strategy was performed in a total volume of 10 μl containing 2 μl Golden 2 × PCR Master Mix (Peoplespot Inc., Beijing, China), 2 μl PCR primers mix, and 6 μl H2O. Cycling conditions were as follows: initial denaturation at 95°C for 11 min; then 30 cycles of 30 s at 94°C, 75 s at 62°C, 60 s at 72°C; and final extension of 45 min at 61°C. PCR was carried out in a GeneAmp® 9700 PCR System (Applied Biosystems, Foster City, USA).

Single-nucleotide polymorphism-short tandem repeat of DYF371 typing

PCR products were prepared for capillary electrophoresis (CE) by combining 1 μl of amplified product with 9 μl of a 30:1 mixture of deionized Hi-Di formamide (Applied Biosystems, Foster City, USA) and Internal Lane Standard ORG (Peoplespot Inc., Beijing, China). PCR products were separated and detected on an ABI PRISM® 3500xL Genetic Analyzer (Applied Biosystems, Foster City, USA) and analyzed by GeneMapper ID v. 1.4 software (Applied Biosystems, Foster City, USA) according to the manufacturer’s instructions. Control DNA 007 (Applied Biosystems, Foster City, USA) or 9948 (Applied Biosystems, Foster City, USA) were genotyped as a standard reference.

Cloning sequencing analysis of the alleles of DYF371 were performed by the Beijing Genomics Institute (BGI, Shenzhen China).

Data analysis

We counted the genotype frequencies of the DYF371 locus. The gene diversity (GD) was calculated using the formula:
\[ GD = \frac{[n(1-\sum p_i^2)]}{(n-1)} \]
where \( n \) is the number of haplotypes, \( p_i \) is the frequency of the \( i \)th haplotype. Haplotype discrimination capacity (DC) was determined as DC = \( h/N \), where \( h \) is the whole number of different haplotypes and \( N \) is the total number of individuals in the samples.[9]

RESULTS AND DISCUSSION

Description of the multiplex polymerase chain reaction system

Normally, DYF371 shows four allele copies on the Y chromosome [Supplementary Figure 1].[10] One copy of the allele on palindromes P1 or P5 has a tightly linked SNP site rs1250403754 or rs1556285687, respectively [Supplementary Figure 1]. The SNPs involving T and C-type were defined as DYF371C or DYF371T here. The multiplex PCR assay described here was constructed to simultaneously amplify the STR of DYF371 and SNP sites DYF371C and DYF371T. The number of STR repeats of DYF371 and SNP in the flanking regions of the STR locus could be detected by PCR products produced from above SNP-STR primers [Figure 1]. A representative electropherograms of control DNA 9948 and 007 for our multiplex PCR are shown in Figure 2.

Genetic diversity of DYF371C and DYF371T

The result from sequencing analysis showed that the repeat structure of DYF371 is (GTT),\( ^n \). The sequence of allele 10C

![Figure 1:](image1.png)

![Figure 2:](image2.png)
of control DNA 9948 is presented in Supplementary Figure 2. The allele lengths varying from 9 to 17 repeats were found at DYF371 locus in this study. Four different alleles with T-type SNP ranging from 10T to 13T repeats and nine alleles with C-type SNP ranging from 9C to 17C repeats were observed. When repeat number of STR was combined with the SNP types, the number of alleles at DYF371 locus has reached up to 13, which increased 4 alleles than those only counted for sizes of alleles. Hence, the addition of SNP information can result in an extension in the number of allele types and improve the discriminating power of DYF371 locus.

Depending on the result of electropherograms, 37 different genotypes were found in the 142 Chinese male individuals. Table 1 shows the number and frequency of different haplotypes at DYF371 locus including alleles of repeat number and SNP-type (C-type and T-type). Among them, 11T, 13C, 15C haplotype showed 48 times (20.69%) which is the most frequent haplotype. Nine of these 37 genotypes were unique (24.32%). The GD value and DC value are 0.8321 and 0.2606, respectively. The results demonstrate that the high polymorphism of DYF371 locus is suitable for differentiating male individuals.

Redd et al. had reported the gene diversity (GD) of DYF371 was 0.929 in the Asians, slightly more than the GD value of 0.814 in the Europeans/Middle Easterners and our value of 0.8321. The GD values were only 0.623 in the Africans and 0.589 in European–Americans from South Dakota, respectively, which were significantly lower than our result. Wang et al. discovered 23 alleles ranged from 10 to 15 repeats for DYF371 in Guizhou Gelao ethnic group from Western China, and the GD value was 0.8983.[10] However, the alleles of DYF371 were not differentiated by SNP-type in their assay. Therefore, their result disagreed with our finding. On the whole, the genetic variant of DYF371 locus is different in various population. This indicates that the DYF371 is a suitable marker for paternal bio-geographic ancestry inference worldwide. Unfortunately, the genetic distances cannot be estimated since the haplotype frequencies are unavailable.

### The different genotype of DYF371 locus between Chinese Han population and minorities

It is well known that SNPs are ancestry informative markers.[11] Across the 90 Han Chinese examined, only three individuals (3.33%) showed the allele 10C and 14 individuals (12.22%) showed allele 11C. On the contrary, 19 out of 52 individuals (36.53%) had allele 10C and only two individuals (3.84%) had allele 11C in the samples of ethnic minorities. Therefore, the alleles 10C or 11C will be helpful to differentiate the Han and non-Han Chinese populations.

Furthermore, the copy numbers of an allele with SNP C-type are significantly different between the samples from Han and non-Han Chinese populations. Based on the

### Table 1: The genotypes of DYF371 locus in Chinese population (n=142)

| n  | Genotype     | Frequency |
|----|--------------|-----------|
| 1  | 10T,13C      | 0.0493    |
| 2  | 10T,13C,14C  | 0.0070    |
| 3  | 10T,13C,15C  | 0.0141    |
| 4  | 10T,14C,15C  | 0.0070    |
| 5  | 10T,15C      | 0.0070    |
| 6  | 10T,15C,17C  | 0.0070    |
| 7  | 10T,11T,13C  | 0.0423    |
| 8  | 10T,11T,13C,14C | 0.0141 |
| 9  | 10T,11T,13C,15C | 0.0775 |
| 10 | 10T,11T,13C,16C | 0.0141 |
| 11 | 10T,11T,14C,14C | 0.0070 |
| 12 | 10T,11T,14C,15C | 0.1268 |
| 13 | 10T,11T,15C  | 0.0352    |
| 14 | 11T,10C,12C  | 0.0070    |
| 15 | 11T,10C,12C,13C | 0.0352 |
| 16 | 11T,10C,12C,14C | 0.0070 |
| 17 | 11T,10C,12C,15C | 0.0211 |
| 18 | 11T,10C,13C  | 0.0423    |
| 19 | 11T,10C,13C,14C | 0.0141 |
| 20 | 11T,10C,13C,15C | 0.0070 |
| 21 | 11T,10C,15C  | 0.0070    |
| 22 | 11T,11C      | 0.0704    |
| 23 | 11T,11C,12C,13C | 0.0352 |
| 24 | 11T,11C,13C,13C | 0.0070 |
| 25 | 11T,12C,13C  | 0.0141    |
| 26 | 11T,13C      | 0.0070    |
| 27 | 11T,13C,14C  | 0.0282    |
| 28 | 11T,13C,15C  | 0.2042    |
| 29 | 11T,13C,16C  | 0.0070    |
| 30 | 11T,14C      | 0.0070    |
| 31 | 11T,14C,15C  | 0.0070    |
| 32 | 11T,15C      | 0.0141    |
| 33 | 11T,9C,12C   | 0.0070    |
| 34 | 12T,10C,13C,15C | 0.0070 |
| 35 | 12T,11C,12C  | 0.0211    |
| 36 | 12T,13C,15C  | 0.0070    |
| 37 | 13T,10C,13C  | 0.0070    |

### Table 2: The distribution of three alleles with DYF371C-type in Chinese Han population and minorities

| n  | DYF371T        | DYF371C     | Population |
|----|----------------|-------------|------------|
| 1  | 11T            | 11C,12C,13C | 4          | Han        |
| 2  | 11T            | 11C,13C,13C | 1          | Han        |
| 3  | 11T            | 10C,12C,13C | 1          | Kazakh     |
| 4  | 11T            | 10C,13C,15C | 1          | Kazakh     |
| 5  | 11T            | 10C,12C,13C | 2          | Kazakh     |
| 6  | 11T            | 10C,13C,14C | 1          | Mongolian  |
| 7  | 11T            | 10C,12C,13C | 1          | Mongolian  |
| 8  | 11T            | 10C,13C,14C | 1          | Hui        |
| 9  | 11T            | 11C,12C,13C | 1          | Zhuang     |
| 10 | 11T            | 10C,12C,13C | 1          | Tibetan    |
| 11 | 11T            | 10C,12C,15C | 1          | Tibetan    |
| 12 | 11T            | 10C,12C,14C | 1          | Tibetan    |
| 13 | 12T            | 10C,13C,15C | 1          | Tibetan    |
| 14 | 11T            | 10C,12C,15C | 2          | Tibetan    |
electrophoretograms, only five males (5.55%) were determined unambiguously that they had three copies of alleles with SNP C-types in the Han samples. On the contrary, similar genotypes were observed clearly in 14 individuals (26.92%) from the samples of minorities [Table 2]. The genotype proportion comprised of three distinct copies of the allele with SNP C-types in Han sample was greatly lower than that in the sample of minorities [Table 2]. Therefore, the genotype pattern that consists of three distinguishing copies of the allele with SNP C-types will be very informative for Chinese bio-geographical ancestry analysis.

**Conclusion**

A multiplex PCR has been developed to analyze the SNP-STR of DYF371 locus. This PCR detected simultaneously the size of DYF371 and the SNPs in the flanking region. A total of 37 SNP-STR haplotypes were found in a sample of 142 individuals from Chinese. The haplotype diversity and discrimination capacity were 0.8321 and 0.2606, respectively. The result indicated that the SNP-STR-type 10C and genotypes had 11T and three copies of the C-type alleles with different repeat numbers and SNP C-type were more common in non-Han population than those in Han population in Chinese. Therefore, the SNP-STR of DYF371 is a very valuable marker with population-differentiated informative for inferring of Chinese bio-geographical ancestry.

**Financial support and sponsorship**

Nil.

**Conflicts of interest**

There are no conflicts of interest.

**References**

1. Trombetta B, Cruciani F. Y chromosome palindromes and gene conversion. Hum Genet 2017;136:605-19.
2. Tanaka H, Yao MC. Palindromic gene amplification-An evolutionarily conserved role for DNA inverted repeats in the genome. Nat Rev Cancer 2009;9:216-24. Epub 2009 Feb 12.
3. Trombetta B, Atanasio ED, Cruciani F. Patterns of inter-chromosomal gene conversion on the male-specific region of the human Y chromosome. Front Genet 2017;8:54.
4. Kayser M, Kittler R, Erler A, Hedman M, Lee AC, Mohyuddin A, et al. A comprehensive survey of human Y-chromosomal microsatellites. Am J Hum Genet 2004;74:1183-97.
5. Redd AJ, Agellon AB, Kearney VA, Contreras VA, Karafet T, Park H, et al. Forensic value of 14 novel STRs on the human Y chromosome. Forensic Sci Int 2002;130:97-111.
6. Wang X, Jiang L, Qian E, Long F, Cui W, Ji A, et al. Genetic polymorphisms and haplotype structure analysis of the Guizhou Gelao ethnic group based on 35 Y-STR loci. Legal Med 2020;43:101666.
7. Hanson EK, Ballantyne J. An ultra-high discrimination Y chromosome short tandem repeat multiplex DNA typing system. PLoS One 2007;2:e688.
8. Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, et al. Recurrent duplication and deletion polymorphisms on the long arm of the Y chromosome in normal males. Human molecular genetics 1996;5:1767-75.
9. Hanson EK, Ballantyne J. A highly discriminating 21 locus Y-STR "megaplex" system designed to augment the minimal haplotype loci for forensic casework. J Forensic Sci 2004;49:40-51.
10. Hanson EK, Ballantyne J. Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. Leg Med (Tokyo) 2006;8:110-20. Epub 2005 Dec 7.
11. Kayser M. Forensic use of Y-chromosome DNA: A general overview. Human Genetics 2017;136:621-35.
Supplementary Figure 1: Localization of DYF371 locus (C-type and T-type) on the palindrome P1 and P5 of the human Y chromosome

Supplementary Figure 2: The sequencing result of allele 10C of control DNA 9948. The single-nucleotide polymorphism site and repeat structure are indicated by the box