

# SCIENTIFIC REPORTS



OPEN

## Haplotypic polymorphisms and mutation rate estimates of 22 Y-chromosome STRs in the Northern Chinese Han father–son pairs

Yaran Yang<sup>1</sup>, Weini Wang<sup>2</sup>, Feng Cheng<sup>3</sup>, Man Chen<sup>1,4</sup>, Tong Chen<sup>1,4</sup>, Jing Zhao<sup>1,4</sup>, Chong Chen<sup>5</sup>, Yan Shi<sup>5</sup>, Chen Li<sup>6</sup>, Chuguang Chen<sup>6</sup>, Yacheng Liu<sup>5</sup> & Jiangwei Yan<sup>1,3,4</sup>

Y chromosome Short tandem repeats (Y-STRs) analysis has been widely used in forensic identification, kinship testing, and population evolution. An accurate understanding of haplotype and mutation rate will benefit these applications. In this work, we analyzed 1123 male samples from Northern Chinese Han population which including 578 DNA-confirmed father-son pairs at 22Y-STRs loci. A total of 537 haplotypes were observed and the overall haplotype diversity was calculated as  $1.0000 \pm 0.0001$ . Except that only two haplotypes were observed twice, all the rest of the 535 were unique. Furthermore, totally 47 mutations were observed during 13,872 paternal meiosis. The mutation rate for each locus estimates ranged from 0.0 to  $15.6 \times 10^{-3}$  with an average mutation rate  $3.4 \times 10^{-3}$  (95% CI  $2.5\text{--}4.5 \times 10^{-3}$ ). Among the 22 loci, DYS449, DYS389 II and DYS458 are the most prone to mutations. This study adds to the growing data on Y-STR haplotype diversity and mutation rates and could be very useful for population and forensic genetics.

Y chromosome Short tandem repeats (Y-STRs) are widely used in genetic epidemiology<sup>1</sup>, forensic genetics<sup>2</sup> and human migration<sup>3</sup> because of its paternal inheritance and human population structuring<sup>4</sup>. However, just the same as autosomal STR, Y-STRs also have high mutation rates<sup>5</sup>. Therefore, reliable estimates of mutation rates of Y-STRs are a prerequisite for the accurate application based on Y-STR analysis. Several studies on estimating Y-STR mutation rates had been reported, such as investigating the father–son pairs from confirmed paternity<sup>6</sup>, male individuals from deep-rooted pedigrees<sup>7</sup>, genotyping sperm cells<sup>8</sup>, and using Y-STR population data with known history<sup>9</sup>. Of these approaches, estimating Y-STR mutation rates through the direct observation of allelic transmission between father and son is the most accurate, as long as large numbers of meiosis could be investigated.

In this study, We determined the haplotypes and mutation rates for the 22 Y-STRs, DYS19, DYS385a/b, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS444, DYS447, DYS448, DYS449, DYS456, DYS458, DYS522, DYS527a/b, DYS635 and Y-GATA-H4 in 1123 Northern Chinese Han male individuals from 578 father–son pairs.

### Materials and Methods

**Samples and DNA extraction.** Blood samples were collected from 1123 healthy Northern Chinese Han male individuals. Among these individuals, there had 578 father–son pairs. All father–son pairs were confirmed by using autosomal STRs typing based on 39 autosomal STRs by using Microreader™ 21 ID and 23 SP system (Microread Genetics Incorporation, China), with a minimum paternity probability of 99.99%. All individuals signed the informed consent before participating in this study. Genomic DNA was extracted using Chelex resin method<sup>10</sup>. The quantity of DNA was quantified by Qubit® Quantitation System (Invitrogen, CA,

<sup>1</sup>CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, 100101, P.R. China. <sup>2</sup>Forensic Science Center of ShenZhen City, Guangdong, 518040, P.R. China. <sup>3</sup>Shanxi Medical University, Taiyuan, 030009, P.R. China. <sup>4</sup>University of Chinese Academy of Sciences, Beijing, 100049, P.R. China. <sup>5</sup>Beijing Tongda Shoucheng Institute of Forensic Science, Beijing, 100085, P.R. China. <sup>6</sup>Beijing Microread Genetics Co., Ltd, Beijing, 100044, P.R. China. Yaran Yang and Weini Wang contributed equally to this work. Correspondence and requests for materials should be addressed to J.Y. (email: [yanjw@big.ac.cn](mailto:yanjw@big.ac.cn))

No. of observed haplotypes	Combined Y-STRs			
	Minimal haplotype	Extended haplotype	Yfiler	This study
1(unique)	458	500	531	535
2	26	13	4	2
3	4	3	0	0
4	3	1	0	0
5	1	0	0	0
Number of haplotypes	492	517	535	537
Unique haplotypes	458	500	531	535
Fraction of unique haplotypes	93.09%	96.71%	99.25%	99.63%
Discrimination capacity	0.999537	0.999798	0.999963	0.999977
Haplotype diversity $\pm$ SD	0.9995 $\pm$ 0.0002	0.9998 $\pm$ 0.0001	1.0000 $\pm$ 0.0001	1.0000 $\pm$ 0.0001

**Table 1.** Haplotype diversity and forensic parameters for different combinations of Y-STRs in 539 unrelated Northern Chinese Han.

Locus	No. of meiosis	Mutation count	No. of mutations with repeat gains	No. of mutations with repeat losses	Mutation rate $\times 10^{-3}$	95% CI $\times 10^{-3}$
DYS19	578	1		1	1.7	0–9.6
DYS385a/b	1156	4	2	2	3.5	0.9–8.8
DYS389I	578	2	1	1	3.5	0.4–13.2
DYS389II	578	5	3	2	8.7	0.3–21.3
DYS390	578	3	2	1	5.2	1.1–16.0
DYS391	578	1	1		1.7	0–10.2
DYS392	578	1	1		1.7	0–10.2
DYS393	578	1	1		1.7	0–10.2
DYS437	578	0			0	0–6.4
DYS438	578	0			0	0–6.4
DYS439	578	4	4		6.9	1.9–17.6
DYS448	578	1		1	1.7	0–9.6
DYS456	578	1	1		1.7	0–10.2
DYS458	578	5	3	2	8.7	0.3–21.3
DYS635	578	1	1		1.7	0–10.2
Y_GATA_H4	578	1	1		1.7	0–10.2
DYS449	578	9	4	5	15.6	7.6–31.1
DYS447	578	0			0	0–6.4
DYS522	578	0			0	0–6.4
DYS388	578	0			0	0–6.4
DYS444	578	2		2	3.5	0.4–13.2
DYS527a/b	1156	5	1	4	4.3	1.4–10.1
Summary	13872	47	26	21	3.4	2.5–4.5

**Table 2.** The mutation and rates for the 22 Y-STR loci studied in Northern Chinese Han.

USA). All experiments of this study were carried out in accordance with the guidelines and regulations of the Ethical Committee of Beijing Institute of Genomics, Chinese Academy of Sciences (Protocol name: A study on the Haplotypic polymorphisms and mutation rate estimates of Y-chromosome STRs in father–son pairs. No. 2016033).

**Multiplex PCR amplification and genotyping.** The samples were amplified 22 Y-STR loci (DYS19, DYS385a/b, DYS388, DYS389I/II, DYS390, DYS391, DYS392, DYS393, DYS437, DYS438, DYS439, DYS444, DYS447, DYS448, DYS449, DYS456, DYS458, DYS522, DYS527a/b, DYS635 and Y-GATA-H4) using AGCU™ Y24 Plus amplification kit (AGCU ScienTech Incorporation, Wuxi, China) following manufacturer's recommendations. All the PCR amplification was proceeded respectively in an ABI PRISM® GeneAmp® 9700 thermal cycler. PCR products were detected on ABI PRISM® 3130xl Genetic Analyzer according to the manufacturer's recommendations. Electrophoretic result was analyzed using GeneMapper® ID-X software.

**Quality control.** A male DNA sample 9948 (Promega Corporation, WI, USA) and female DNA sample 9947 A (Promega Corporation) were used as reference and negative control for each batch of genotyping. All the experiments were carried out at the laboratory accredited by the China National Accreditation Service (CNAS)

Locus	Motif	Allele range	Short	Moderate	Long
DYS19	TAGA	10–19	1		
DYS385a/b	GAAA	7–28		4	
DYS389I	(TCTG)(TCTA)	9–17		2	
DYS389II	(TCTG)(TCTA)(TCTG)(TCTA)	24–34			5
DYS390	(TCTA)(TCTG)	17–28		3	
DYS391	TCTA	6–14	1		
DYS392	TAT	6–17	1		
DYS393	AGAT	9–17		1	
DYS437	TCTA	13–17			
DYS438	TTTTC	6–14			
DYS439	AGAT	9–14	1	3	
DYS448	AGAGAT	20–26			1
DYS456	AGAT	13–18		1	
DYS458	GAAA	13–20		2	3
DYS635	TSTA compound	17–27		1	
Y_GATA_H4	TAGA	8–13	1		
DYS449	TTTC	26–36			9
DYS447	TAAWA compound	22–29			
DYS522	GATA	8–17			
DYS388	ATT	10–18			
DYS444	TAGA	11–15	1	1	
DYS527a/b	(GAAA)(GGAA)	15–28			5
No. of mutations			6	18	23
No. of alleles			3200	7346	3326
Mutation rate $\times 10^{-3}$			1.9	2.5	6.9
95% CI $\times 10^{-3}$			0.7–4.1	1.5–3.9	4.4–10.4

**Table 3.** Mutation counts and rates by relative allele sizes (short, moderate, and long) for each locus. The allele sizes are generally categorized into 25%, 50%, and 25% quantiles for short, moderate and long allele sizes, respectively.

and strictly followed the recommendations on the analysis of Y-STRs by DNA Commission of the International Society of Forensic Genetics (ISFG)<sup>11</sup>.

**Statistical analysis.** Haplotype and allele frequencies were calculated by the gene counting method. Gene diversity (GD) for each locus was calculated using the formula:  $GD = [n(1 - \sum \pi_i^2)] / (n - 1)$ , where  $n$  is the number of alleles,  $\pi_i$  is the frequency of the  $i$ th allele. Discrimination capacity (DC) was determined as  $DC = N_{diff} / N$ , where  $N_{diff}$  and  $N$  was the number of different haplotypes and the sample size, respectively. Haplotype diversity (HD) and Standard error (SE) was calculated according to Nei's formula<sup>12</sup>. Mutation rates were calculated as the number of mutations divided by the number of Meiosis. Confidence intervals (CI) were estimated from the binominal standard deviation<sup>13</sup>. In mutation counting, there were two father–son pairs where one-step mutation seen for both DYS389I and DYS389II, for instance (13, 29)  $\rightarrow$  (14, 30). These were treated as one mutation instead of two because DYS389I is part of the sequence called DYS389II. According to the repeat numbers of alleles per locus, the alleles were categorized into short (25%), medium (50%) and long class (25%) as described by Ge *et al.*<sup>14</sup>, used to evaluate the relationship between allele size and corresponding mutation rate.

## Results and Discussion

**Allele frequencies and gene diversity.** Allele frequencies and gene diversity values for each locus are listed in Supplemental Table 1. A total of 190 alleles were detected at 22 Y-STR loci with the allele frequencies ranged from 0.0018 to 0.7676. The number of alleles at each locus ranged from 4 for Y\_GATA\_H4 to 15 for DYS447. At two multi-copy loci DYS385a/b and DYS527a/b, 71 allelic combinations with 16 separated alleles and 40 allelic combinations with 11 alleles were observed, respectively. Single-copy locus DYS449 and multi-copy locus DYS385a/b showed the highest GD values as 0.8883 and 0.9658, respectively. Except DYS391 and DYS438, the GD values for the other 20 loci were all above 0.5, which suggests high polymorphisms in the Northern Chinese Han population.

**Haplotype diversity.** The haplotype distribution in a sample of 539 unrelated Northern Chinese Han males for the 22 Y-STRs is shown in the Supplementary Table 2. Haplotype diversity and forensic parameter based on various combinations of Y-STRs, such as minimal haplotype, extended haplotype, Yfiler and this study are shown in Table 1. Within these unrelated 539 Northern Chinese Han individuals, a total of 537 haplotypes were observed at the 22-loci resolution and the haplotype diversity value of was 0.99998. 535 haplotypes (99.63%) were observed once and only 2 were observed twice (0.37%). 492 haplotypes were observed at the minimal haplotype STRs

Locus	This study				Ballantyne <i>et al.</i> <sup>25</sup>				Burgarella <i>et al.</i> <sup>26</sup>			
	No. of meiosis	Mutation count	Mutation rate $\times 10^{-3}$	95% CI $\times 10^{-3}$	No. of meiosis	Mutation count	Mutation rate $\times 10^{-3}$	95% CI $\times 10^{-3}$	No. of meiosis	Mutation count	Mutation rate $\times 10^{-3}$	95% CI $\times 10^{-3}$
DYS19	578	1	1.7	0–9.6	1756	7	4.4	2.0–8.2	14632	32	2.2	1.6–3.1
DYS385 a/b*	1156	4	3.5	0.9–8.8	1762	3	2.1	0.6–5.1	/	/	/	/
					1615	6	4.1	1.8–8.1	/	/	/	/
DYS389 I	578	2	3.5	0.4–13.2	1751	9	5.5	2.7–9.7	12651	32	2.5	1.8–3.6
DYS389 II	578	5	8.7	0.3–21.3	1743	6	3.8	1.6–7.5	/	/	/	/
DYS390	578	3	5.2	1.1–16.0	1758	2	1.5	0.4–4.1	14131	30	2.1	1.5–3.0
DYS391	578	1	1.7	0–10.2	1759	5	3.2	1.3–6.7	13995	38	2.7	1.8–3.7
DYS392	578	1	1.7	0–10.2	1728	1	9.7	0.1–3.2	13948	6	0.4	0.2–0.9
DYS393	578	1	1.7	0–10.2	1750	3	2.1	0.6–5	12576	13	1	0.6–1.8
DYS437	578	0	0	0–6.4	1760	2	1.5	0.4–4.1	9238	10	1.1	0.6–2
DYS438	578	0	0	0–6.4	1751	1	1	0.1–3.2	9339	4	0.4	0.2–1.1
DYS439	578	4	6.9	1.9–17.6	1736	6	3.8	1.6–7.5	9313	51	5.8	4.2–7.2
DYS448	578	1	1.7	0–9.6	1747	0	0	0.01–2.1	6655	11	1.7	0.9–3
DYS456	578	1	1.7	0–10.2	1757	8	4.9	2.4–9	6664	30	4.5	3.2–6.4
DYS458	578	5	8.7	0.3–21.3	1756	14	8.4	4.8–13.4	6684	46	6.9	5.2–9.2
DYS635	578	1	1.7	0–10.2	1732	6	3.9	1.6–7.6	7434	23	3.1	2.1–4.6
Y_GATA_H4	578	1	1.7	0–10.2	1755	5	3.2	1.3–6.6	7618	21	2.8	1.8–4.2
DYS449	578	9	15.6	7.6–31.1	1617	19	12.2	7.5–18.5	369	7	1.9	9.2–38.6
DYS447	578	0	0	0–6.4	1722	3	2.1	0.6–5.1	658	3	4.6	1.6–13.3
DYS522	578	0	0	0–6.4	1620	1	1	0.2–3.4	555	0	0	0–6.9
DYS388	578	0	0	0–6.4	1635	0	0	0.02–2.3	2394	1	0.4	0.02–2.4
DYS444	578	2	3.5	0.4–13.2	1775	9	5.5	2.7–9.7	80	0	0	0–45.8
DYS527 a/b	1156	5	4.3	1.4–10.1	/	/	/	/	/	/	/	/

**Table 4.** Comparison between the mutation rates of this study and other two published studies with large number of meiosis. \*The mutation rate for *DYS385 a/b* was calculated separately as *DYS385 a* and *DYS385 b* by Ballantyne's study. †The data were absent due to these loci were not included in the study.

resolution (*DYS19*, *DYS389I/II*, *DYS390*, *DYS391*, *DYS392*, *DYS393*, and *DYS385*). For the extended haplotype STRs (minimal haplotype STRs, *DYS438* and *DYS439*), 517 haplotypes were observed.

In the case of 17 Y-STRs (extended haplotype STRs, *DYS437*, *DYS448*, *DYS456*, *DYS458*, *DYS635*, and *GATA H4.1*) from AmpFlSTR® *Yfiler™* kit, 535 haplotypes were observed with the haplotype diversity value of 0.99996. From these, 531 haplotypes (99.25%) were observed once, 4 were observed twice (0.75%). Although the number of unique haplotype increased when additional Y-STR loci were combined, however, in this study, only 2 unique haplotype were increased with 5 loci were added compared with Y *filer*. This suggest that to achieve the goal for high haplotype resolution for Y-STR analysis, selecting appropriate loci, such as the Rapidly mutating Y-STRs<sup>15</sup>, should be considered.

**Variant alleles.** Thirty four copy number variants were detected in 1123 males. Variant alleles were confirmed by re-amplification and genotyping. Null alleles were observed at *DYS448* (6 father–son pairs), *DYS19* (1 father–son pair) and *DYS527a/b* (1 father–son pair). Primers were designed for larger PCR fragments of these 3 loci, but failed to produce amplicons in the test samples (data not shown). *DYS448* is located within the azoospermia factor c gene (*AZFc*) in the distal euchromatic part of the Y chromosome long arm. *AZFc* consists almost entirely of very long direct and inverted repeats. Therefore, it is prone to partial deletions or duplications by rearrangements<sup>16</sup>. The *DYS448* null allele has been reported by several studies<sup>17–20</sup>. The relatively high frequencies of the *DYS448* null allele in Asians suggest giving careful consideration to the use of *DYS448* for commercial genotyping and further database construction in Asians. Triplications were observed at *DYS527a/b* (8 father–son pairs) and *DYS385 a/b* (1 father–son pair). These variants are not rare in forensic casework and they should be interpreted carefully to exclude mixed profiles. These variants have been considered due to non-allelic, homologous recombination<sup>21</sup>.

**Mutation rates.** In this study, 578 meiosis from fathers to sons were observed, in which 47 mutations were found at all the studied loci except *DYS47*, *DYS438*, *DYS447*, *DYS522*, and *DYS388* (Table 2). There are no more than one locus mutations in the same father–son pair. Except one three-step mutation occurred at *DYS449* (32 → 29), all remaining mutations were single step, namely, 97.9% mutations were one step. This finding is consistent with the general notion that the majority of mutations comprise single step repeat gain or loss due to strand slippage during replication<sup>22</sup>. Among these 47 mutations, 26 mutations (i.e., 55.3%) gained repeats, and 21 mutations (i.e., 44.7%) lost repeats. Hence, the data herein support that mutations at these Y chromosome microsatellites do not have any contraction or expansion bias.

The average mutation rate across these 22 Y-STR loci was 0.0034 (95% confidence interval (CI), 0.0025–0.0045), which was close to the average mutation rates across 16 Y-STR markers of the Texas populations (i.e., 0.0021) by Ge *et al.*<sup>14</sup> and the South China Han population (i.e., 0.0023) by Weng *et al.*<sup>23</sup> The mutation rates of the 22 Y-STR loci ranged from 0.0000 (95% CI, 0.0000–0.0064) to 0.0156 (95% CI, 0.0076–0.0311). Mutation counts and rates by relative allele sizes (short, moderate, and long) for each locus is shown in Table 3. In the Northern Chinese Han population, the mutation rate of long alleles ( $6.9 \times 10^{-3}$ ) is significantly greater than short ( $1.9 \times 10^{-3}$ ) and moderate ( $2.5 \times 10^{-3}$ ) alleles. Therefore, the longer alleles are more likely to be mutated than short alleles, which is consistent with the previous studies<sup>14,23,24</sup>.

It is more accurate to estimate the Y-STR mutation rate is by testing a large number of meiosis from father-son pairs. Ballantyne *et al.*<sup>25</sup> provided Y-STR mutation rates for a large number of Y-STR markers in a reasonably large number of up to 2000 DNA-confirmed father-son pairs collected from the Germany and Poland. Burgella *et al.*<sup>26</sup> performed a meta-analysis to estimate the mutation rate for 110 Y-STRs combining population and father-son pair data. A comparison of our data to these published rates was shown in Table 4. The mutation rates for most of the shared loci were similar except DYS449, which was  $1.9 \times 10^{-3}$  reported by Burgarella and only approximately one eighth and one seventh of our and Ballantyne's study.

## Conclusion

In this study, we investigated the haplotype diversity and estimated mutation rates for 22 Y-STRs in 578 father-son pairs in a Northern Chinese Han population. We detected 537 distinct haplotypes in 539 male individuals, which indicating a high power to distinguish unrelated male individuals. Furthermore, totally 47 mutations were observed during 13,872 paternal meiosis. The mutation rate for each locus estimates ranged from 0.0 to  $15.6 \times 10^{-3}$  with an average mutation rate  $3.4 \times 10^{-3}$  (95% CI  $2.5$ – $4.5 \times 10^{-3}$ ). This study adds to the growing data on Y-STR haplotype diversity and mutation rates. It could be very useful for population and forensic genetics. However, to obtain precise knowledge of haplotype and mutation rate, more number of meiosis analyses involving more Y-STRs loci should be performed.

## References

1. Jobling, M. A. & Tylersmith, C. The human Y chromosome: an evolutionary marker comes of age. *Nature Reviews Genetics* **4**, 598 (2003).
2. Kayser, M. Uniparental markers in human identity testing including forensic DNA analysis. *Biotechniques* **43** (2007).
3. Underhill, P. A. & Kivisild, T. Use of y chromosome and mitochondrial DNA population structure in tracing human migrations. *Annual Review of Genetics* **41**, 539 (2007).
4. Karafet, T. M. *et al.* New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome research* **18**, 830 (2008).
5. Kayser, M. & Sajantila, A. Mutations at Y-STR loci: implications for paternity testing and forensic analysis. *Forensic science international* **118**, 116 (2001).
6. Kayser, M. *et al.* Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *American journal of human genetics* **66**, 1580 (2000).
7. Heyer, E., Puymirat, J., Dieltjes, P., Bakker, E. & De Knijff, P. Estimating Y Chromosome Specific Microsatellite Mutation Frequencies using Deep Rooting Pedigrees. *Human molecular genetics* **6**, 799–803 (1997).
8. Holtkemper, U., Rolf, B., Hohoff, C., Forster, P. & Brinkmann, B. Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. *Human molecular genetics* **10**, 629–633 (2001).
9. Zhivotovskiy, L. A. *et al.* The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time. *American journal of human genetics* **74**, 50–61 (2004).
10. Walsh, P. S., Metzger, D. A. & Higuchi, R. Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* **10**, 506–513 (1991).
11. Gusmão, L. *et al.* 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 science international* **120**, 191–200 (2006).
12. Nei, M. *Molecular evolutionary genetics*. 176–179 (Columbia University Press, 1987).
13. Clopper, C. J. & Pearson, E. S. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika* **26**, 404–413 (1934).
14. Ge, J. *et al.* Mutation rates at Y chromosome short tandem repeats in Texas populations. *Forensic Science International*. *Genetics* **3**, 179–184 (2009).
15. Ballantyne, K. N. *et al.* A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Science International Genetics* **6**, 208 (2012).
16. Kuroda-Kawaguchi, T. *et al.* The AZFc region of the Y chromosome features massive palindromes and uniform recurrent deletions in infertile men. *Nature genetics* **29**, 279 (2001).
17. Mizuno, N. *et al.* 16 Y chromosomal STR haplotypes in Japanese. *Forensic science international* **174**, 71–76 (2008).
18. Parkin, E. J. *et al.* Diversity of 26-locus Y-STR haplotypes in a Nepalese population sample: Isolation and drift in the Himalayas. *Forensic science international* **166**, 176–181 (2007).
19. Chang, Y. M., Perumal, R., Keat, P. Y. & Kuehn, D. L. Haplotype diversity of 16 Y-chromosomal STRs in three main ethnic populations (Malays, Chinese and Indians) in Malaysia. *Forensic science international* **167**, 70 (2007).
20. Yi, Y., Gao, J., Fan, G., Liao, L. & Hou, Y. Population genetics for 23 Y-STR loci in Tibetan in China and confirmation of DYS448 null allele. *Forensic Science International*. *Genetics* **16**, e7 (2015).
21. Jobling, M. A. Copy number variation on the human Y chromosome. *Cytogenetic & Genome Research* **123**, 253 (2008).
22. Budowle, B. *et al.* Twelve short tandem repeat loci Y chromosome haplotypes: genetic analysis on populations residing in North America. *Mutation research* **150**, 1–15 (2005).
23. Weng, W. *et al.* Mutation rates at 16 Y-chromosome STRs in the South China Han population. *International journal of legal medicine* **127**, 369–372 (2013).
24. Goedbloed, M. *et al.* Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFISTR® Yfiler® PCR amplification kit. *International journal of legal medicine* **123**, 471–482 (2009).
25. Ballantyne, K. N. *et al.* Mutability of Y- chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.* **87**, 341–353 (2010).
26. Burgarella, C. *et al.* Mutation rate estimates for 110 Y-chromosome STRs combining population and father-son pair data. *Eur. J. Hum. Genet.* **19**, 70–75 (2011).

## Acknowledgements

This project was supported by the National Natural Science Foundation of China (NSFC, No. 81330073) and the CAS Key Program (KGFZD-135-16-021).

## Author Contributions

J.Y. and Y.L. conceived and designed the experiments, Y.Y. and W.W. performed the experiment and wrote the main manuscript text, F.C., M.C. and T.C. analyzed data and the manuscript modification, C.C., Y.S. and C.C. performed the father–son pairs DNA confirmed experiment and data analysis, C.L. performed MDS analysis and provided the figure. All authors reviewed the manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-018-25362-3>.

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018