ORIGINAL ARTICLE

The construction and application of a new 17-plex Y-STR system using universal fluorescent PCR

Jinding Liu[1] · Rongshuai Wang[2] · Jie Shi[1] · Xiaojuan Cheng[1] · Ting Hao[1] · Jiangling Guo[1] · Jiaqi Wang[1] · Zidong Liu[1] · Wenyan Li[1] · Haoliang Fan[3] · Keming Yun[1] · Jiangwei Yan[1] · Gengqian Zhang[1] 🔟

Abstract

Y-chromosomal short tandem repeat (Y-STR) polymorphisms are useful in forensic identification, population genetics, and human structures. However, the current Y-STR systems are limited in discriminating distant relatives in a family with a low discrimination power. Increasing the capacity of detecting Y chromosomal polymorphisms will drastically narrow down the matching number of genealogy populations or pedigrees. In this study, we developed a system containing 17 Y-STRs that are complementary to the current commercially available Y-STR kits. This system was constructed by multiplex PCR with expected sizes of 126–400 bp labeled by different fluorescence molecules (DYS715, DYS709, DYS716, DYS713, and DYS607 labeled by FAM; DYS718, DYS723, DYS708, and DYS714 labeled by JOE; DYS712, DYS717, DYS721, and DYS605 labeled by TAMRA; and DYS719, DYS726, DYS598, and DYS722 labeled by ROX). The system was extensively tested for sensitivity, male specificity, species specificity, mixture, population genetics, and mutation rates following the Scientific Working Group on DNA Analysis Methods (SWGDAM) guidelines. The genetic data were obtained from eight populations with a total of 1260 individuals. Our results showed that all the 17 Y-STRs are human- and male-specific and include only one copy of the Y-chromosome. The 17 Y-STR system detects 143 alleles and has a high discrimination power (0.996031746). Mutation rates were different among the 17 Y-STRs, ranging from 0.30 to 3.03%. In conclusion, our study provides a robust, sensitive, and cost-effective genotyping method for human identification, which will be beneficial for narrowing the search scope when applied to genealogy searching with the Y-STR DNA databank.

Keywords Y-STRs · Chinese population · Forensic DNA · Mutation rate · Human identification

---

Jinding Liu and Rongshuai Wang contributed equally to this work.

✉ Keming Yun
    yunkeming5142@163.com

✉ Jiangwei Yan
    yanjw@sxmu.edu.cn

✉ Gengqian Zhang
    gengqianzhang@sxmu.edu.cn

[1] School of Forensic Medicine, Shanxi Medical University, Jinzhong 030619, Shanxi, China

[2] Chongxin Judicial Expertise Center, Wuhan 430030, Hubei, China

[3] School of Basic Medicine and Life Science, Hainan Medical University, Haikou 571199, Hainan, China

Introduction

With the rapid development of DNA analysis technology, STR genotyping methods consisting of multiplex PCR with fluorescently labeled primers and capillary electrophoresis have been conventionally employed in the field of forensic medicine for individual identification and paternity testing [1, 2].

The Y chromosome acts as a unique tool for forensic investigations since it is inherited through the patrilineal line in a relatively conserved manner [3, 4]. Unlike autosomal STR markers, for which the typing of two samples must be compared to make a personal identification, typing of Y chromosome markers from stains from a crime scene could be helpful for inferring the potential perpetrator's origin if his familial DNA typing could be found in a DNA databank, i.e., to find a perpetrator or narrow the investigation scope by the "Sample to family" searching mode [5–7]. Y chromosome SNP (Y-

SNP) haplogroups are valuable for forensic applications of paternal biogeographical ancestry inference [8–10]. However, for many of the recently discovered and already phylogenetically mapped Y-SNPs, the population data are scarce in some populations. Y-chromosome microsatellites or short tandem repeats (Y-STRs) were first defined in Europe for forensic purposes and included nine Y-STR loci. Since then, more Y-STR loci have been added to the kits to increase the discrimination power for kinship analysis and human identification and for inferences on population history and evolution [11, 12]. A number of commercial Y-STR kits, such as the Yfiler™ Plus PCR Amplification kit (Applied Biosystems, Foster City, CA, USA), the PowerPlex Y23 System (Promega, Madison, WI, USA), STRtyper-27Y (Health Gene Technologies, China), and the Microreader™ 40Y ID System (Microread Genetics Incorporation, China), are available, and most incorporate 19–40 markers into the single multiplex systems, which have been validated with paternal genetic data in many populations and in forensic casework [13–15]. For the purpose of differentiating individuals (only males), Y-STR databases are established for either online public access or within criminal investigation laboratories, which are not available for public access. The US-Y-STR and Y-HRD are established for public access and are used to estimate the Y-STR haplotype frequency or to infer the ethnicity of the donor of a profile [16, 17]. The Y-STR databases established in crime laboratories are used to trace male lineages. The famous example of a Y-STR database is the Henan Provincial Y-STR database in China [5]. With this database, the potential profile of a DNA sample from a crime scene is searched by first screening families' Y-STR haplotypes and then subsequently investigating the identification of the individual in the Y haplotype-matched families. Dozens of cases have been solved in an expeditious manner by using this Y-STR database [5, 18].

Despite the robustness of the Y-STR markers in these commercial multiplex Y-STR systems and the capacity to discriminate two male individuals in most cases, the coincidence match probabilities are modest compared with a set of standard autosomal STR markers. Henan Province in China has established a large Y-STR (> 200,000 profiles) database for criminal investigators using either the Applied Biosystems Yfiler or Yfiler Plus PCR Amplification kits [5]. The limited number of Y-STRs has resulted in a few or even large numbers of false-positive matches with the enlargement of the databank. Therefore, improving the discrimination power by increasing the number of Y-STRs would be an effective way to reduce such false matches. Regarding of the strengths and weaknesses of Y-STR markers, more loci are still required to provide more information to assist forensic investigations, as also emphasized by Ge and Liu et al. [18].

In this study, we developed a new 17 Y-STR typing kit, which exceeds the current Y-STR system containing the trinucleotide loci DYS718 and DYS719; the tetranucleotide loci DYS715, DYS709, DYS713, DYS607, DYS708, DYS723, DYS712, DYS605, DYS726, and DYS722; and the pentanucleotide loci DYS716, DYS714, DYS717, DYS721, and DYS598 [19–21].

Currently, the analysis procedure is often performed by evaluating fluorescent fragments via capillary electrophoresis (CE) [22, 23]. Universal fluorescent PCR is a cost-effective genotyping method in which fluorescent fragments are obtained with non-labeled forward primers, non-labeled reverse primers, and fluorescent universal primers by a two-step amplification [24–26].

The combination of Y-STR and an economic fluorescence labeling technique can reduce the expenditures for forensic laboratories. We then used a new kind of genotyping method featuring FAM/JOE/TAMRA/ROX-labeled universal primers (M13 (−21) universal primer 5′-TGTAAAACGACGGC CAGT-3′) [25]. We validated this kit according to the recommendations on forensic analysis proposed by the International Society of Forensic Genetics [27, 28]. Following these guidelines, we defined the Y-STR and allele nomenclature and assessed the kit for PCR conditions, sensitivity, accuracy, species specificity, and the effects of DNA mixtures and mutation rates. Our results showed that the 17-plex Y STR typing system is a reproducible, accurate, sensitive, and economical tool for forensic identification.

Materials and methods

DNA samples

The following procedures were performed with the approval of the Ethics Committee of Shanxi Medical University. Written informed consent was obtained from all participants.

A total of 1600 blood samples were collected from 930 unrelated male samples, 10 unrelated female samples and 330 unrelated father-son pairs. The 1590 male samples were collected from 724 males (330 unrelated father-son pairs, which were confirmed by using autosomal STR analysis, and 64 unrelated males) in Taiyuan City (Shanxi Province), 162 unrelated males in Chongqing City, 154 unrelated males in Ulanqab City (Mongolia), 155 unrelated males in Sanmenxia City (Henan Province), 95 unrelated males in Foshan City (Guangdong Province), 113 unrelated males in Hainan Li, 63 unrelated males in Hainan Miao, and 124 unrelated males in Jingzhou City (Hubei Province) (see Figure S1). The 10 female samples were recruited from the Shanxi Han population.

DNA was extracted by using a QIAamp[R] DNA Investigator Kit (QIAGEN China (Shanghai)) and was quantified by using a Quantifiler[R] Human DNA Quantification Kit (Applied Biosystem, USA) on a Bio-Rad Real-time PCR

System (Bio-Rad Laboratories, USA). The quantification process was performed per the manufacturer's instructions.

The control DNAs 9948, 2800M, and 9947A were purchased from Promega (Promega, Madison, WI, USA).

The species specificity study evaluated the capacity for the system to avoid the detection of genetic information from nontargeted species. Nonhuman samples from chickens, cattle, fishes, pigs, rabbits, rats, sheep, and shrimps were obtained from Shanxi Medical University Animal Center. DNA was extracted using a QIAGEN DNA Tissue Mini Kit (Qiagen, USA) and was quantified by the standard OD260 method. For each species, 10 ng of DNA was amplified by the 17-plex Y-STR assay following the standard protocol. The human DNA sample 2800 M (4 ng) was used as a positive control.

Primer design

The loci were selected from our previously published articles [19–21]. Sequences for each locus were obtained from NCBI (https://www.ncbi.nlm.nih.gov/) using a standard nucleotide BLAST (Basic Local Alignment Search Tool) search. We selected 17 autosomal Y-STRs, which were classified into four sets (sets A, B, C, and D) based on the needs of multiplex amplification.

An 18-base universal M13 (− 21) sequence (TGTAAAAC GACGGCCAGT) was used as the standard sequence. The universal tail was added to the 5′ termini of the forward primers in each set. The primers were labeled with four fluorescent dyes (FAM, JOE, TAMRA, and ROX) according to the four sets. All primers were synthesized by Shanghai Sangon Biological Engineering Technology & Services Company, Shanghai, China.

PCR amplification

All individuals were analyzed in four independent multiplex PCRs (sets A, B, C, and D). Each PCR was performed by using 1 ng of DNA in a total volume of 15 μL, containing (0.008 μM–0.8 μM) forward primers with universal tails, reverse primers, 0.25 μM of the fluorescent universal M13 (− 21) primer and 1× PCR MasterMix (Mei5 Biotechnology (Beijing) Co, Ltd). The reaction conditions were as follows: 95 °C for 10 min; 25 cycles of 95 °C for 25 s, 56 °C for 25 s, and 72 °C for 25 s; 8 cycles of 95 °C for 25 s, 53 °C for 25 s, and 72 °C for 25 s; and a final extension at 72 °C for 60 min. PCR products (0.4 μL of each set) were added to 10 μL of HiDi™-Formamide (Applied Biosystems), containing 0.3 μL of the GeneScan™HD 500 ROX™ Size Standard (Applied Biosystems), and were detected by capillary electrophoresis (CE) using an ABI PRISM^R 3130 Genetic Analyzer (Applied Biosystems) with a 36-cm capillary and performance-optimized polymer (POP4). Data were analyzed using GeneMapper ID software (Applied Biosystems). Signal

intensities were represented by relative fluorescence units (RFU). The bins and panels for the multiplex system were programmed for genotyping.

Allelic ladders were created for the 17 loci by using TA cloning to represent the range of alleles observed in the Shanxi population [29, 30]. Alleles were amplified independently at each locus with monoclonal cells, and the products of each locus were then diluted and reanalyzed to produce a single allelic ladder for each locus. Finally, those single allelic ladders were mixed in appropriate proportions to create a "cocktail". Each reaction (total volume = 100 μL) contained 0.2 μM of each locus-specific primers, 2 μL of monoclonal cells, and the reagent from the multiplex PCR kit (Mei5 Biotechnology (Beijing) Co, Ltd) at 1× concentration. The reaction conditions were as follows: 95 °C for 10 min; 25 cycles of 95 °C 25 s, 58 °C 25 s, and 72 °C 25 s; 8 cycles of 95 °C 25 s, 53 °C 25 s, and 72 °C 25 s; and a final extension at 72 °C for 60 min.

Sensitivity

The sensitivity study estimated the capacity for the system to obtain the whole profile from a range of DNA quantities. It was performed with control DNA 9948 (Ori-Gene company, USA), and total DNA inputs were prepared in a serial dilution to 1 μL with the following template amounts: 12 ng, 4 ng, and 0.4 ng (3 ng, 1 ng, and 0.1 ng DNA in each reaction). All experiments were carried out according to the parameters provided above.

DNA mixtures

Male and female DNAs (formed by 2800 M and an anonymous female sample) were mixed in four combinations, which were previously diluted to 1:1, 1:50, 1:100, and 1:500. For each combination, the male DNA was maintained at 1 ng, and the amount of the female DNA was varied from 1 to 500 ng in each reaction. For male and male mixtures, a total of 2 ng of DNA was input in each reaction. The samples were prepared using an anonymous male sample and 2800 M human genomic DNA, with mixture ratios of 1:0, 1:1, 1:2, 1:4, and 1:9. Each mixture was tested in triplicate to reduce the accidental error and to ensure the accuracy of the result analysis.

Statistical analyses

The allele and haplotype frequencies were counted directly. Gene diversity (GD) was calculated as GD = $n(1-\Sigma P_a^2)/(n$-1), where n is the total number of samples and P_a is the relative frequency of the ath allele at the locus, respectively [31]. Haplotype diversity (HD) was calculated as HD = $n(1-\Sigma P_i^2)/$ $(n$-1), where P_i is the frequency of the ith haplotype and n indicates the total number of samples [32, 33]. The

discrimination capacity (DC) ratio was calculated as DC=$\frac{Nh}{Ns}$, where Nh represents the number of unique haplotypes and Ns represents the total number of individual samples [32]. Population Fst genetic distance and P value (P value was corrected by Sidak's correction for multiple testing ($P < 0.0018$, 28 pairs)) were obtained using Arlequin software to assess the genetic structure [34]. PCA based on the genotypes of the 17 Y-STRs in the eight populations was performed with SPSS 22.0. The subpopulation structure was examined via model-based clustering algorithms implemented in STRUCTURE 2.3.4, which was based on a Bayesian Markov Chain Monte Carlo algorithm. Analyses were performed from $K = 2$ to $K = 8$ using the no-admixture model and correlated allele frequencies to estimate the 17 Y-STRs. Structure Harvester was applied to estimate the optimal K value.

The mutations were counted directly, and the locus-specific mutation rate was calculated as the number of observed mutations divided by the number of father-son pairs. The 95% confidence intervals (CIs) were estimated from the binomial probability distribution available at http://statpages.info/confint.html [35].

Quality control

The above experiments were performed in the Forensic Genetics Laboratory of the School of Forensic Medicine of Shanxi Medical University, P.R. China, which is an accredited laboratory, in accordance to quality control measures. The DNAs 9947A and 9948 were used as negative and positive controls, respectively. All the alleles observed in this assay were validated by Sanger sequencing.

Results

Assessment of the system

The 17 loci were divided into four panels, and each panel was assigned a color dye. The allelic ladders and internal size standard in the 17-plex system are shown in Fig. 1. The set of designed primers act according to unified parameters, with a consistent temperature, to achieve a stable and balanced amplified sensitivity and efficient amplification in multiplex PCR processes. Primer-related information on each locus (sequence, melting temperature and distance from repeat motif, concentration, PCR product size, and dye label) is listed in Table 1. In all, we successfully established the 17-plex Y-STR system with universal fluorescence-labeled primers. The genotype results of 9948 and 2800 M are shown in Table 1.

Mixed samples consisting of DNA from two or more individuals are frequently encountered in many forensic cases. Thus, we assessed the typing system's capacity to analyze DNA mixtures. Mixtures of two individuals (male 2800 M and an anonymous female sample) were examined in various ratios (1:0, 1:1, 1:10, and 1:100) with 1 ng of template DNA in each set for male/female mixtures. Testing was performed in triplicate to ensure the accuracy of the results. All peak heights of the male 2800 M could be called for ratios of 1:1 and 1:100 (Fig. 2a). The two male samples (an anonymous male sample and 2800 M) were used to evaluate male/male mixtures with ratios of 1:0, 0:1, 1:1, 1:2, 1:4, and 1:9 for a total of 2 ng of template DNA in each set. The minor alleles dropped out at the 1:4 DNA (the anonymous male sample) ratio. The peak height ratios in the 1:4 and 1:9 mixtures were also 10% or less. We were unable to call alleles for some of the minor profiles at the DYS716 locus in the 1:4 ratio samples and at the DYS718, DYS712, and DYS605 loci in the 1:9 ratio samples (Fig. 2b). These results indicate that the 17 Y-STR system can be used to genotype DNA samples in mixtures with relatively small mixture ratios.

We assessed the sensitivity of the reaction by setting 50 RFU as the limit of detection and 60% as the stochastic threshold for peak balance using varying amounts of 9948 control DNA. We were able to obtain the full profile of sample 9948 when the DNA template provided was greater than 0.04 ng. For samples with more than 0.04 ng, the peak height increased with increasing amounts of DNA template in the reaction. Allelic drop-out and allelic imbalance occurred when the amount of template DNA was reduced to less than 0.04 ng. We thus used 4 ng of template DNA for our typing system (Figure S2).

Nonhuman genomic DNA samples from common animal species (chicken, cow, fish, pig, rabbit, rat, sheep, and shrimp) and female samples were amplified using the 17-plex Y-STR system. The results did not reveal the presence of any allele peaks within the genotyping range (Figure S3).

On this basis, it was concluded that the developed STR system was robust and unlikely to be affected by the presence of genetic material from these animal species and female samples.

Population data

We used the 17-plex Y-STR system to genotype 1260 unrelated Chinese males recruited from eight Chinese ethnic groups (Table S1). Table S2.1 shows the allele frequencies in eight populations. The allele frequencies ranged from 0.0025 to 0.9919. A total of 143 alleles were detected using this system. Table S2.2 shows the allele frequencies and parameters in 1260 individuals in China. A total of 1255 distinct haplotypes were generated; 1250 individuals had unique haplotype profiles, while the remaining 10 individuals exhibited the same haplotype as one other individual. Gene diversity (GD) ranged from 0.2939 (DYS726) to 0.9176 (DYS712). The haplotype diversity (HD) and discrimination capacity
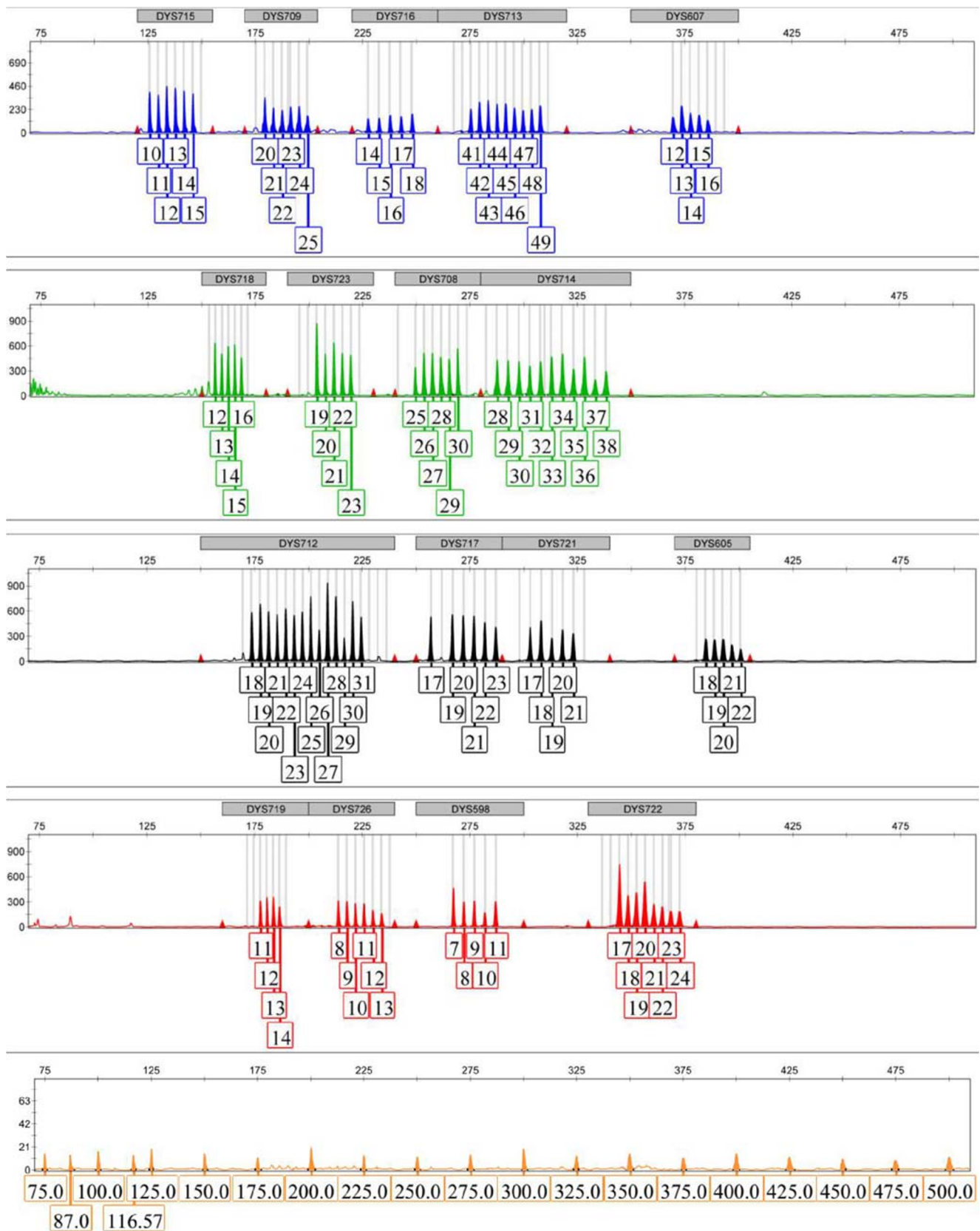
Fig. 1 Electropherogram of allelic ladders and internal size standard in the 17-plex system. The four dye panels for the allelic ladders correspond to (from top to bottom) FAM (blue), JOE (green), TAMRA (yellow), ROX (re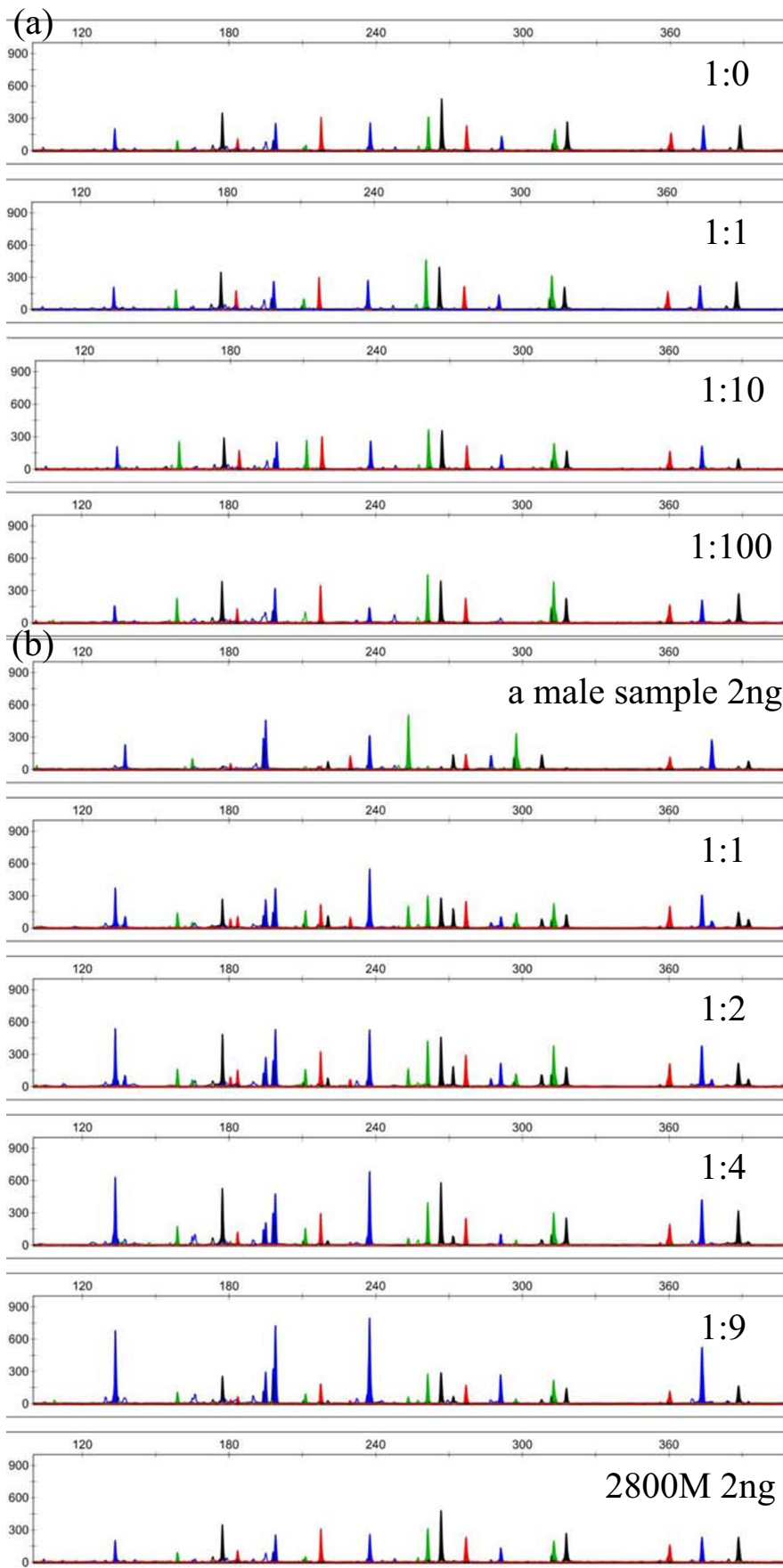d) dye-labeled peaks. The genotype is shown with the allele number displayed underneath each peak. The fifth panel shows the internal size standards labeled with Orange500 dye (a total of 20 fragments: 75, 87, 100, 116.5, 125, 130, 175, 200, 225, 250, 275, 300, 325, 350, 375, 400, 425, 450, 475, and 500 bp)

Table 1 General information on loci used in the new 17 Y-STR system

STR locus	Chromosomal location in GRCh38/hg38	NCBI accession numbers	Repeat motif	No. of alleles	Amplicon size (bp)	Allelic range	Primer sequences(5'-3')	Primer concentrations (μM)	9948	2800 M
SetA										
FAM-M13							TGTAAAACGACGGCCAGT	0.25		
DYS715	15,550,251–15,550,381	NG_001197.5	(TAGA)m	7	125–149	10,11,12,13,14,15	TGTAAAACGACGGCCAG TATGGTTGGAAGAA AGCATTGATGA / CATCCATCCATCCACATCT ATATCATCTTTA	0.04 / 0.04	14	12
DYS709	14,971,031–14,971,199	AC006989.3	(TTCT)4CTCT(TTCT) 2(CTTTTCT)2CTT(TTCT)m	8	175–199	20,21,22,22.3,23,24,25	TGTAAAACGACGGCC AGTTCTTTCCAATGA CCAAGACGTG / TGCAAATTGTTCACAT GTACCCT	0.008 / 0.075	20	25
DYS716	10,629,578–10,629,801	AC069323.5	(CACTC)m(CATTC)n	5	227–247	14,15,16,17,18	TGTAAAACGACGGCC AGTTAAATCAGA ATTCCTTTCCAATCCA / TCTGGGTTTCAGAG TGGGATAATT	0.04 / 0.1	15	16
DYS713	7,964,006–7,964,265	AC145782.1	(TCTT)mTC(TCTT)2(TCTG)1(TCTT) TTT(TCTT)ITC(TCTT)nITT (TCTT)5CCT(TCTT)IC(TCTT) T(TCTT)	12	267–311	41,41.2,42,43,44,45, 45.2,46,47,48,49	TGTAAAACGACGGC CAGTCCAGAAATA GATTTAITTCACGCTTG / CCTGGGTGACAGAC TCCATCTTAAA	0.08 / 0.8	43	45
DYS607	16,302,428–16,302,792	AC139189.1	(AAGG)m	7	369–393	12,13,14,15,16	TGTAAAACGACGCGCC AGTCATACAGCG TAATCACAGCTCAC / GTAATGATGCCTCC AGTAACCAA	0.04 / 0.1	15	12
SetB										
JOE-M13							TGTAAAACGACGG CCAGT	0.25		
DYS718	15,152,935–15,153,081	AC147710.3	(TTA)m	7	153–171	12,13,14,15,16	TGTAAAACGACGG CCAGTGGAGAAA AITTCAATGCAGTTACC ACACCAGCTTGGC ACATTTA	0.08 / 0.6	15	13
DYS723	15,170,017–15,170,206	AC011289.4	(GATA)2TAT(GATA) mGAT(GATA)1GAT(GATA)7	8	195–223	19,20,21,22,22.2,23	TGTAAAACGACGGC CAGTGACAGGT GGATGCATAAAITGG / TCTGGCATCTGTCT GCATAITT	0.08 / 0.1	19	21
DYS708	15,162,953–15,163,196	AC011289.4	(GATA)m (GACA)n	8	249–277	25,26,27,28,29,30	TGTAAAACGACGGC CAGTAGTGTAITC CGCCATGGTAGC / CTGCATTTTGGTA CCCCATA	0.08 / 0.2	28	28
DYS714	19,985,782–19,986,081	AC009233.3	(TTTTTC)m(TCTTTC)2 (TTTTC)2(TCTTC)2(TTTTC)2	13	282–342	28,29,30,31,32,32. 2,33,34,35,36,37,38	TGTAAAACGACGG CCAGTGCATCG AICTTTCTGGGAGC / GTGTGATGCTGACT TTGGGG	0.08 / 0.2	32	33
SetC										

Table 1 (continued)

STR locus	Chromosomal location in GRCh38/hg38	NCBI accession numbers	Repeat motif	No. of alleles	Amplicon size (bp)	Allelic range	Primer sequences(5'-3')	Primer concentrations (μM)	9948	2800 M
TAMRA-M13										
DYS712	13,446,527–13,446,695	AC146183.3	(AGAT)m(AGAC)n	18	169–237	18,19,20,21,22,23, 24,25,26,27,28,29,30,31	TGTAAAACGACGGCCAGT TGTAAAACGACGGC CAGTCAAGAAC AGCCTGGGTAACAGTG TATATGGTACAGCC CATGAACACTT	0.25 0.016 0.1	19	19
DYS717	15,201,262–15,201,511	AC010972.3	(TGTAT)2TAT(TGTAT) (TGTAC)m(TGTAT)n	7	257–287	17,19,20,21,22,23	TGTAAAACGACGGCC AGTGGCCGAGAG AATGGAATTGAT CCCGAACTTCAGCAC TATGAAATG	0.08 0.2	19	19
DYS721	15,163,835–15,164,125	AC147710.3	(AAGGG)mN10(AAGGG) 2 N7(AAGCA)6	7	298–328	17,18,19,20,21	TGTAAAACGACGGC CAGTGGGTGATA GAGGGAGGCTTCT CGGGCATGAGCTAT TGAGTC	0.016 0.1	19	20
DYS605	15,329,665–15,330,032	AC007007.3	(TATC)4CA(TATC) 3CG(TATC)m	6	380–400	18,19,20,21,22	TGTAAAACGACGGC CAGTAACATCTGG CTTACTTGTAGGTAG TCTTTGTCAGACA ATGATCTGTAA	0.08 0.2	20	18
SetD										
ROX-M13										
DYS719	14,944,954–14,945,115	AC142296.1	(ATA)m	7	171–189	11,12,13,14	TGTAAAACGACGGCCAGT TGTAAAACGACGG CCAGTTGACGAG TTAATGGGTGCAG GGAGAAAATTCAA TGCAGAT	0.25 0.016 0.1	12	13
DYS726	11,446,630–11,446,827	AC134879.3	(CTTC)3 N13 (CTTC) mN21(CTTT)3 T(CTTT)2	7	214–238	8,9,10,11,12,13	TGTAAAACGACGGCC AGTGGGTAAACC TCTGAAGACCATAC GAATGACAGACC AAGACTCTCTC	0.04 0.1	9	9
DYS598	8,544,962–8,545,208	AC016991.5	(AGAAC)m	5	267–287	7,8,9,10,11	TGTAAAACGACGG CCAGTCTTTATTA GGCAGGCAGTTTTG CCAGACAATGTA TGAGCAAGC	0.04 0.1	7	9
DYS722	15,166,237–15,166,580	AC011289.4	(GAAA)mAAGA(GAAA) 2A(GAAA)2GAG A(GAAA)2	11	336–376	17,18,19,20,21, 22,22.3,23,24	TGTAAAACGACGG CCAGTCCACTCA TCAGTGCTCAGCTA GCCAACCAGCAA TGTTGTC	0.04 0.1	21	20

Fig. 2 Representative electropherogram of DNA mixture. **a** Formed by 2800 M (1 ng DNA template in each set) and an anonymous female sample at the ratios 1:0, 1:1, 1:10, 1:100. **b** Mixtures of two male samples (an anonymous male sample and 2800 M) were examined in various ratios (1:0, 0:1, 1:1, 1:2, 1:4, and 1:9) with 2 ng total template DNA in each set

(DC) were 0.999993696 and 0.996031746, respectively (see Table S3).

STRUCTURE software (the program STRUCTURE implements a model-based clustering method for inferring population structure using genotyping data consisting of unlinked markers) runs for K = 2–8 showed an optimum K value (K = 5) estimated by Structure Harvester (http://taylor0.biology. ucla.edu/structureHarvester/). A bar plot of K = 5 analysis reveals that the Hainan Miao and Hainan Li samples can be primarily separated from other Chinese Han populations (Fig.

3a). The PCA demonstrated the population stratification in the tested samples (Fig. 3b).

The Fst and *P* values are indicators for evaluating the distribution differences of allele frequencies among populations. The Fst and *P* values for pairwise interpopulation comparisons were calculated based on allele frequencies of 17 Y-STRs by analysis of molecular variance (AMOVA) performed with ARLEQUIN version 3.5 software between eight populations (see Table S4). There were significant differences after Sidak's correction for multiple testing (*P* < 0.0018, 28 pairs) at DYS716, DYS713, DYS607, DYS718, DYS714, DYS717, DYS721, DYS605, DYS719, DYS726, and DYS598. At DYS716 and DYS713, the Chongqing population was different from the other seven populations. At the DYS607 locus, the Taiyuan (Shanxi) population could be differentiated. At DYS718, the Hainan Li population could be differentiated

Fig. 3 Component analysis results for 1260 Chinese samples from 8 population with the 17 selected Y-STR markers. **a** Bar plot of STRUCTURE run for K = 5. The bars with different colors represent different areas of samples. **b** Two-dimensional plot of PCA analysis. The first two principal components (PC1 vs. PC2) demonstrate the population stratification in the tested samples
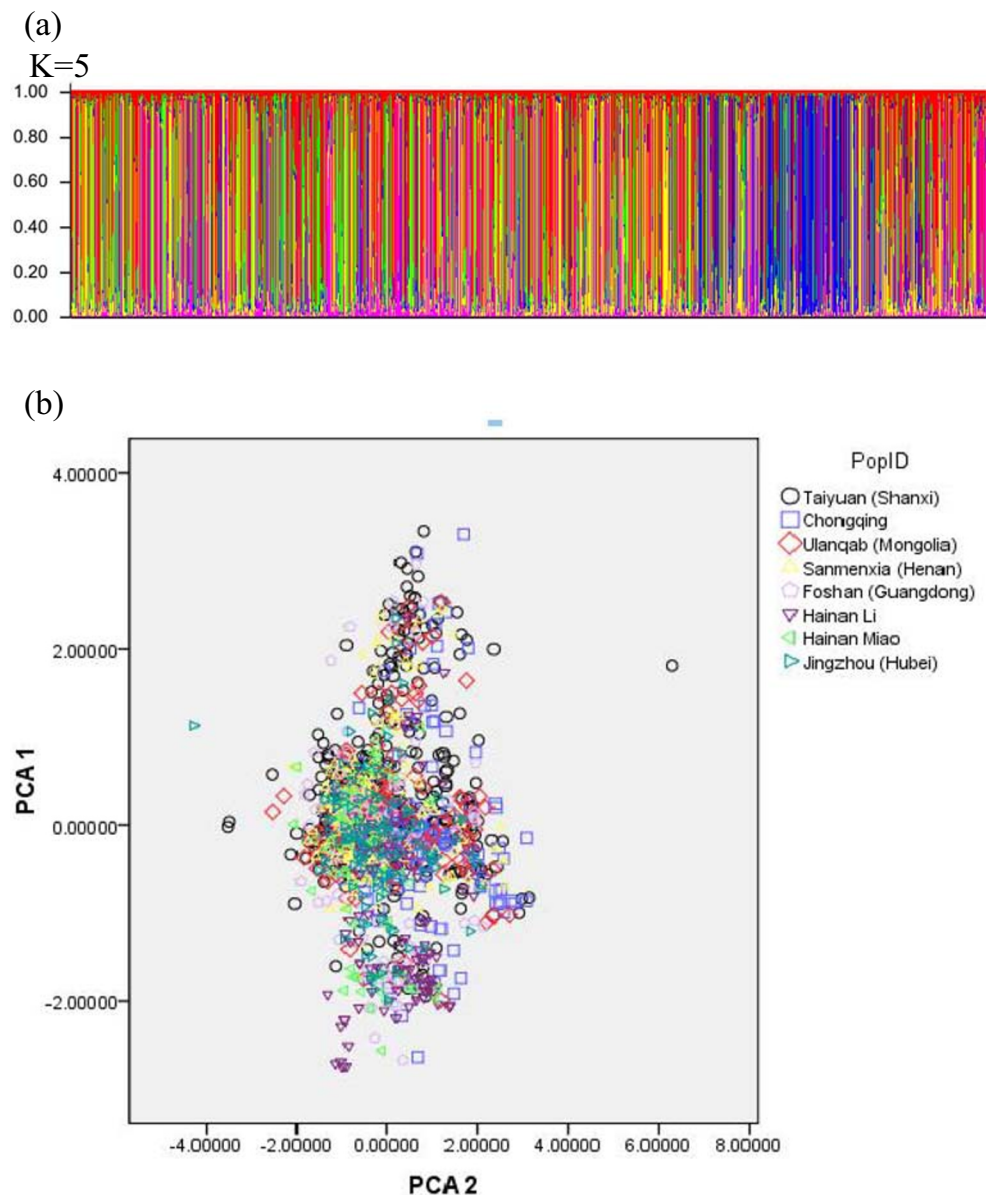
from the other seven populations. At the DYS714, DYS721, DYS605, DYS719, DYS726, and DYS598 loci, the Jingzhou (Hubei) population could be differentiated. At the DYS717 locus, the Hainan Li and Jingzhou (Hubei) populations could be differentiated from the other six populations.

A total of 330 allele transmission events from 330 father-son pairs were analyzed at the 17 Y-STRs. Mutations were observed in 38 father-son pairs in DYS715, DYS709, DYS716, DYS713, DYS607, DYS718, DYS714, DYS723, DYS712, DYS721, DYS719, DYS726, and DYS598 (see Table 2). The average mutation rate was 0.89%. The Y-STR DYS712 had the highest mutation rate of 3.03%. Highly polymorphic STRs had high mutation rates. However, no mutations were found in DYS708, DYS605, DYS717, and DYS722, which had relatively high gene diversity values. There were 38 mutations observed, including 16 repeat gains versus 22 repeat losses (1:1.375) and 35 one-step mutations versus three two-step mutations (11.667:1). A one-step mutation is shown in Figure S4.

Discussion

In this study, we developed a cost-effective multiplex 17 Y-STR genotyping method using four fluorescent universal primers. Multiplex PCRs using $\geqq 4$ ng of DNA produced intense signals for all 17 loci, all genotypes from eight different populations were fully concordant with STR profiles, and the mutation rate was assessed with 330 father-son pairs.

Standard DNA profiling using sets of well-selected, largely standardized, highly polymorphic autosomal STRs is very suitable for identifying the donor of a single-source crime scene trace, as long as the person's STR profile already exists in the DNA database. Currently, the STR profiles obtained from crime scene traces are compared with the stored profiles in the forensic DNA database to look for a match. However, this comparative autosomal STR profile matching for human identification is not successful for completely unknown perpetrators, whose STR profiles are not yet available. In addition, if the traces are from multiple-source stains, the autosomal STR profiles can be compromised because of the masking effect of the main contributor. As in sexual assault cases, for example, based on the excess of epithelial cells from the female major contributor, it is often difficult to single out the autosomal STR profile of the male contributor. This is where Y-chromosome STR profiling comes into play, as only the male perpetrator, but not the female victim, carries a Y chromosome.

However, when dealing with specific cases, additional loci are required. For cases to discriminate related males belonging to the same paternal lineage or to separate paternal lineages in populations with low Y-chromosome diversity, loci with relatively high mutation rates would be useful to improve the power of discrimination. For paternal lineage discrimination, loci with relatively low mutation rates would perform better. Therefore, more Y-STR loci should be studied.

In this study, we used a cost-effective multiplex genotyping method with four fluorescent universal primers. There are two different designs for labeling the PCR amplicons. One approach involves the addition of universal tails to only forward (or reverse) primers, while the other adds the tails to both forward and reverse primers. To reduce interactions between primers, we used the first strategy in this study. We selected four fluorescent labels with different colors (FAM, JOE,

Table 2 Mutation rates of 17 Y-STRs in a Han population of Shanxi, China

Y-STR locus	Allele transmission	Number of mutation	Number of gains	Number of losses	Number of two steps mutation	Mutation rate	Binominal 95% CIs
DYS715	330	1	1			0.30%	0.0001–0.0168
DYS709	330	1		1		0.30%	0.0001–0.0168
DYS716	330	4	2	2	1	1.21%	0.0033–0.0307
DYS713	330	2	1	1		0.61%	0.0007–0.0217
DYS607	330	4	1	3		1.21%	0.0033–0.0307
DYS718	330	2		2		0.61%	0.0007–0.0217
DYS714	330	5	3	2	1	1.52%	0.0049–0.0350
DYS723	330	1		1		0.30%	0.0001–0.0168
DYS712	330	10	4	6		3.03%	0.0146–0.0550
DYS721	330	1	1			0.30%	0.0001–0.0168
DYS719	330	3	1	2	1	0.91%	0.0019–0.0263
DYS726	330	1		1		0.30%	0.0001–0.0168
DYS598	330	3	2	1		0.91%	0.0019–0.0263
Total		38	16	22		0.89%	

TAMRA, and ROX), which are commonly used by the Health STRtyper series. This choice also enabled us to use the same capillary electrophoresis running module (Dye Set Any5dye) with the 36-cm capillary and POP4 polymer.

The usefulness of multiplex PCR systems has often been discussed based on interloci balances. By optimizing the concentrations of the fluorescent universal primers, our system can produce well-balanced signals of the 17 Y-STR loci. The concentrations of the reverse primers are generally 10 times higher than those of the forward primers when increasing the production of fluorescent fragments, which is higher than Masaru Asari et al. recommended in their research [26]. Optimal PCR conditions are required to acquire a robust detection system. To obtain well-balanced signals among the 17 loci, we evaluated the concentration ratios of the fluorescence-labeled primer, forward primers and reverse primers. We found that differential primer concentrations yielded much better results than those using even primer concentrations. The best signals were obtained when the ratios of the forward/reverse primer concentration ranged from 1:1 to 1:10. The PCR was performed with higher annealing temperatures for 25 cycles to amplify the target genome DNA and lower annealing temperatures for eight cycles to integrate the fluorescence to the PCR products. For commercial STR kits, the cost of fluorescence labeling cannot be neglected. Using the optimized PCR parameters, we were able to decrease the cost of fluorescence labeling to 25% for the following CE detection.

Previous studies of the 17 Y-STRs identified 90 alleles [19–21]. In this study, we detected 143 alleles in 8 populations. For all 17 Y-STRs, only the allele frequencies of DYS607 and DYS598 had been investigated in Cape Muslim, Belgium, and southeast China populations [36–38]. The number of alleles was the same in our population. We evaluated the mutation rates in all 17 loci in this study, while only DYS607 was evaluated in the Belgium population. The STRUCTURE result estimated the classification relationships among populations. A bar plot of K = 5 revealed that our 8 populations could not be separated into 8 clusters, which meant that our system could be used in all Chinese populations. PCA is a classical nonparametric linear dimensionality reduction technique that extracts the fundamental structure of a dataset without the need for any modeling. It can be used for inferring population clusters and assigning individuals to subpopulations. It is not possible to significantly distinguish all of the samples. Fst and P values were used to demonstrate the genetic differentiation. After Sikad's correction, the P values at 11 loci were less than 0.0018, suggesting that these markers had large differences in gene frequencies between the two populations.

STRs with mutation rates greater than 10^{-2} were termed rapidly mutating Y-STRs (RM Y-STRs) [39]. In our study, 38 mutations were observed in 13 loci: 35 mutations were one-step mutations, and 3 mutations were two-step mutations. Nine loci (DYS715, DYS709, DYS713, DYS718, DYS723,

DYS721, DYS719, DYS726, DYS598) showed low mutation rates ($< 10^{-2}$), and four loci (DYS716, DYS607, DYS714, and DYS712) could be regarded as RM Y-STRs, with the highest mutation rate of 3.03×10^{-2} at DYS712 and the lowest mutation rate of 1.21×10^{-2} at DYS716 and DYS607. The mutation rate at DYS607 in our population (1.21×10^{-2}) was higher than that in the Belgium population (2.54×10^{-3}) [37]. Among the 17 Y-STR markers, four loci (DYS708, DYS717, DYS605, and DYS722) with no mutations were observed, nine loci (DYS715, DYS709, DYS713, DYS718, DYS723, DYS721, DYS719, DYS726, and DYS598) had mutation rates lower than 1.00×10^{-2}, and the rest of the loci (DYS716, DYS607, DYS714, and DYS712) had mutation rates higher than 1.00×10^{-2} in our population. The 13 highly polymorphic Y-STR markers could be used to improve the discrimination capacity in populations with low genetic diversity. The newly discovered 4 Y-STR loci with high mutation rates can be used to differentiate different individuals from the same paternal lineage.

This multiplex assay remains a supplemental DNA tool which could provide additional Y-STR information. With the increment of 17-plex Y-STRs application in the future, we proposed to amplify in a single reaction and labeled the one of the two primers for each Y-STR at the 5′ end with dye to decrease the amount of input DNA.

Conclusion

With recent developments in forensic science, Y-STR analysis applied for forensic purposes has been continually improved. This article developed a new 17-plex Y-STR typing system for forensic genetic testing that incorporated the loci from of the current Yfiler Plus kit. Developmental validation studies that included PCR conditions as well as testing the cross-reactivity, sensitivity, anti-interference, and stability of the method and population data analysis have demonstrated it to be a sensitive, robust, and highly informative tool for use in forensic casework.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval The sampling was approved by the Ethics Committee of Shanxi Medical University, declaration of no objection no. 2019LL096.

Informed consent Blood samples were collected from volunteers with their written informed consent.

References

1. Bai X, Yao Y, Wang C, Li W, Wang Y, Song W, Mo X, Ma W, Ye J (2019) Development of a new 25plex STRs typing system for forensic application. Electrophoresis 40(12–13):1662–1676. https://doi.org/10.1002/elps.201900021

2. Fan H, Wang X, Ren Z, He G, Long R, Liang A, Song T, Deng J (2019) Population data of 19 autosomal STR loci in the Li population from Hainan Province in southernmost China. Int J Legal Med 133(2):429–431. https://doi.org/10.1007/s00414-018-1828-2

3. Tao R, Jin M, Ji G, Zhang J, Zhang J, Yang Z, Chen C, Zhang S, Li C (2019) Forensic characteristics of 36 Y-STR loci in a Changzhou Han population and genetic distance analysis among several Chinese populations. Forensic Sci Int Genet 40:e268–e270. https://doi.org/10.1016/j.fsigen.2019.03.011

4. Liu C, Han X, Min Y, Liu H, Xu Q, Yang X, Huang S, Chen Z, Liu C (2018) Genetic polymorphism analysis of 40 Y-chromosomal STR loci in seven populations from South China. Forensic Sci Int 291:109–114. https://doi.org/10.1016/j.forsciint.2018.08.003

5. Liu H, Li X, Mulero J, Carbonaro A, Short M, Ge J (2016) A convenient guideline to determine if two Y-STR profiles are from the same lineage. Electrophoresis 37(12):1659–1668. https://doi.org/10.1002/elps.201500566

6. Mo XT, Zhang J, Ma WH, Bai X, Li WS, Zhao XC, Ye J (2019) Developmental validation of the DNATyper()Y26 PCR amplification kit: an enhanced Y-STR multiplex for familial searching. Forensic Sci Int Genet 38:113–120. https://doi.org/10.1016/j.fsigen.2018.10.008

7. Zhang J, Mo X, Shang L, Jin X, Chen D, Zhu H, Zhang Y, Kang B, Li W, Ye J (2019) Genetic analysis of 29 Y-STR loci in Han population from Dongfang, Southern China. Int J Legal Med 133(4):1033–1035. https://doi.org/10.1007/s00414-018-1899-0

8. Lang M, Liu H, Song F, Qiao X, Ye Y, Ren H, Li J, Huang J, Xie M, Chen S, Song M, Zhang Y, Qian X, Yuan T, Wang Z, Liu Y, Wang M, Liu Y, Liu J, Hou Y (2019) Forensic characteristics and genetic analysis of both 27 Y-STRs and 143 Y-SNPs in Eastern Han Chinese population. Forensic Sci Int Genet 42:e13–e20. https://doi.org/10.1016/j.fsigen.2019.07.011

9. Ralf A, van Oven M, Montiel Gonzalez D, de Knijff P, van der Beek K, Wootton S, Lagace R, Kayser M (2019) Forensic Y-SNP analysis beyond SNaPshot: high-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using ion AmpliSeq and targeted massively parallel sequencing. Forensic Sci Int Genet 41:93–106. https://doi.org/10.1016/j.fsigen.2019.04.001

10. Song M, Wang Z, Zhang Y, Zhao C, Lang M, Xie M, Qian X, Wang M, Hou Y (2019) Forensic characteristics and phylogenetic analysis of both Y-STR and Y-SNP in the Li and Han ethnic groups from Hainan Island of China. Forensic Sci Int Genet 39:e14–e20. https://doi.org/10.1016/j.fsigen.2018.11.016

11. Cokic VP, Kecmanovic M, Zgonjanin Bosic D, Jakovski Z, Veljkovic A, Katic S, Keckarevic Markovic M, Keckarevic D (2019) A comprehensive mutation study in wide deep-rooted R1b Serbian pedigree: mutation rates and male relative differentiation capacity of 36 Y-STR markers. Forensic Sci Int Genet 41:137–144. https://doi.org/10.1016/j.fsigen.2019.04.007

12. Kayser M (2017) Forensic use of Y-chromosome DNA: a general overview. Hum Genet 136(5):621–635. https://doi.org/10.1007/s00439-017-1776-9

13. Olofsson JK, Mogensen HS, Buchard A, Borsting C, Morling N (2015) Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler(R) plus PCR amplification kit. Forensic Sci Int Genet 16:232–236. https://doi.org/10.1016/j.fsigen.2015.02.006

14. Thompson JM, Ewing MM, Frank WE, Pogemiller JJ, Nolde CA, Koehler DJ, Shaffer AM, Rabbach DR, Fulmer PM, Sprecher CJ, Storts DR (2013) Developmental validation of the PowerPlex(R) Y23 system: a single multiplex Y-STR analysis system for casework and database samples. Forensic Sci Int Genet 7(2):240–250. https://doi.org/10.1016/j.fsigen.2012.10.013

15. Liu YJ, Guo LH, Li J, Yue JT, Shi MS (2018) Genetic Polymorphisms of 27 Y-STR Loci in Dongxiang Population of Gansu Province. Fa Yi Xue Za Zhi 34(3):270–275. https://doi.org/10.12116/j.issn.1004-5619.2018.03.011

16. Willuweit S, Roewer L (2015) The new Y chromosome haplotype reference database. Forensic Sci Int Genet 15:43–48. https://doi.org/10.1016/j.fsigen.2014.11.024

17. Ge J, Budowle B, Planz JV, Eisenberg AJ, Ballantyne J, Chakraborty R (2010) US forensic Y-chromosome short tandem repeats database. Leg Med (Tokyo) 12(6):289–295. https://doi.org/10.1016/j.legalmed.2010.07.006

18. Ge J, Sun H, Li H, Liu C, Yan J, Budowle B (2014) Future directions of forensic DNA databases. Croat Med J 55(2):163–166. https://doi.org/10.3325/cmj.2014.55.163

19. Zhang GQ, Yang SY, Niu LL, Guo DW (2012) Structure and polymorphism of 16 novel Y-STRs in Chinese Han population. Genet Mol Res 11(4):4487–4500. https://doi.org/10.4238/2012.October.11.1

20. Zhang GQ, Wang Y, Zhang YX, Xu XL, Xing XP, Wang YY, Yun KM (2004) Study of polymorphism at new Y-STR DYS605 in a Chinese Han population of Shanxi. Yi Chuan 26(3):295–297

21. Zhang GQ, Yun KM, Zhang YX, Wang Y, Wang YY (2004) Allele frequencies for two new Y-chromosome STR loci DYS598 and DYS607 in Chinese Han population (Shanxi area). J Forensic Sci 49(3):630

22. Xu XM, Zheng JL, Lou Y, Wei XH, Wang BJ, Yao J (2019) Population genetics of 24 Y-STR loci in Chinese Han population from Jilin Province, Northeast China. Mol Genet Genomic Med: e984. https://doi.org/10.1002/mgg3.984

23. Wang Y, Dang Z, Zhang G, Li S, Liu Q, Li C, Hou X, Li H, Chen S, Cui W, Wang D, Kong X, Man D (2019) Genetic diversity and haplotype structure of 27 Y-STR loci in a Han population from Jining, Shandong province, eastern China. Forensic Sci Int Genet 42:e25–e26. https://doi.org/10.1016/j.fsigen.2019.06.008

24. Blacket MJ, Robin C, Good RT, Lee SF, Miller AD (2012) Universal primers for fluorescent labelling of PCR fragments–an efficient and cost-effective approach to genotyping by fluorescence. Mol Ecol Resour 12(3):456–463. https://doi.org/10.1111/j.1755-0998.2011.03104.x

25. Oka K, Asari M, Omura T, Yoshida M, Maseda C, Yajima D, Matsubara K, Shiono H, Matsuda M, Shimizu K (2014) Genotyping of 38 insertion/deletion polymorphisms for human identification using universal fluorescent PCR. Mol Cell Probes 28(1):13–18. https://doi.org/10.1016/j.mcp.2013.09.002

26. Asari M, Okuda K, Hoshina C, Omura T, Tasaki Y, Shiono H, Matsubara K, Shimizu K (2016) Multicolor-based discrimination of 21 short tandem repeats and amelogenin using four fluorescent universal primers. Anal Biochem 494:16–22. https://doi.org/10.1016/j.ab.2015.01.005

27. Gusmao L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, Morling N, Prinz M, Roewer L, Tyler-Smith C, Schneider PM, International Society of Forensic G (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. Int J Legal Med 120(4):191–200

28. 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 (2001) DNA Commission of the International Society of Forensic Genetics: recommendations on forensic

analysis using Y-chromosome STRs. Forensic Sci Int 124(1):5–10. https://doi.org/10.1016/s0379-0738(01)00498-4

29. Zhou MY, Gomez-Sanchez CE (2000) Universal TA cloning. Curr Issues Mol Biol 2(1):1–7

30. Yao S, Hart DJ, An Y (2016) Recent advances in universal TA cloning methods for use in function studies. Protein Eng Des Sel 29(11):551–556. https://doi.org/10.1093/protein/gzw047

31. Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci U S A 70(12):3321–3323. https://doi.org/10.1073/pnas.70.12.3321

32. Li M, Zhou W, Zhang Y, Huang L, Wang X, Wu J, Meng M, Wang H, Li C, Bian Y (2020) Development and validation of a novel 29-plex Y-STR typing system for forensic application. Forensic Sci Int Genet 44:102169. https://doi.org/10.1016/j.fsigen.2019.102169

33. Siegert S, Roewer L, Nothnagel M (2015) Shannon's equivocation for forensic Y-STR marker selection. Forensic Sci Int Genet 16:216–225. https://doi.org/10.1016/j.fsigen.2015.02.001

34. Zhu BF, Zhang YD, Liu WJ, Meng HT, Yuan GL, Lv Z, Dong N, Li Q, Yang CH, Zhang YH, Hou YL, Qian L, Fan SL, Xu P (2014) Genetic diversity and haplotype structure of 24 Y-chromosomal STR in Chinese Hui ethnic group and its genetic relationships with other populations. Electrophoresis 35(14):1993–2000. https://doi.org/10.1002/elps.201300574

35. Wu W, Ren W, Hao H, Nan H, He X, Liu Q, Lu D (2018) Mutation rates at 42 Y chromosomal short tandem repeats in Chinese Han population in Eastern China. Int J Legal Med 132(5):1317–1319. https://doi.org/10.1007/s00414-018-1784-x

36. Cloete K, Ehrenreich L, D'Amato ME, Leat N, Davison S, Benjeddou M (2010) Analysis of seventeen Y-chromosome STR loci in the Cape Muslim population of South Africa. Leg Med (Tokyo) 12(1):42–45. https://doi.org/10.1016/j.legalmed.2009.10.001

37. Claerhout S, Vandenbosch M, Nivelle K, Gruyters L, Peeters A, Larmuseau MHD, Decorte R (2018) Determining Y-STR mutation rates in deep-routing genealogies: identification of haplogroup differences. Forensic Sci Int Genet 34:1–10. https://doi.org/10.1016/j.fsigen.2018.01.005

38. Shi M, Bai R, Yu X, Lv J, Hu B (2009) Haplotype diversity of 22 Y-chromosomal STRs in a southeast China population sample (Chaoshan area). Forensic Sci Int Genet 3(2):e45–e47. https://doi.org/10.1016/j.fsigen.2008.05.009

39. Ay M, Serin A, Sevay H, Gurkan C, Canan H (2018) Genetic characterisation of 13 rapidly mutating Y-STR loci in 100 father and son pairs from South and East Turkey. Ann Hum Biol 45(6–8):506–515. https://doi.org/10.1080/03014460.2018.1559353