Journal Pre-proof

Title: Developmental Validation of Y-SNP Pedigree Tagging System: a panel via quick ARMS PCR

Caiyong Yin, Yijie Ren, Atif Adnan, Junzhe Tian, Kejian Guo, Mingying Xia, Ziwei He, Dian Zhai, Xueyun Chen, Lei Wang, Xin Li, Xingjun Qin, Shilin Li, Li Jin

PII: S1872-4973(20)30042-9

DOI: https://doi.org/10.1016/j.fsigen.2020.102271

Reference: FSIGEN 102271

To appear in: *Forensic Science International: Genetics*

Received Date: 17 June 2019

Revised Date: 1 March 2020

Accepted Date: 2 March 2020

Please cite this article as: Yin C, Yijie R, Atif A, Junzhe T, Kejian G, Mingying X, Ziwei H, Dian Z, Xueyun C, Lei W, Xin L, Xingjun Q, Li S, Li J, Title: Developmental Validation of Y-SNP Pedigree Tagging System: a panel via quick ARMS PCR, *Forensic Science International: Genetics* (2020), doi: https://doi.org/10.1016/j.fsigen.2020.102271

Title: Developmental Validation of Y-SNP Pedigree Tagging System: a panel via quick ARMS PCR

Caiyong Yin [a,b,c#] *, Yijie Ren [a,b,c#], Atif Adnan [d#], Junzhe Tian [a,b,c#], Kejian Guo [a,e], Mingying Xia [a], Ziwei He [a,b,c], Dian Zhai [f], Xueyun Chen [f], Lei Wang [g], Xin Li [g], Xingjun Qin [g]*, Shilin Li [a,b,c] *, Li Jin [a,b,c]

[a] MOE Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200438, China

[b] Human Phenome Institute, Fudan University, Shanghai 200438, China

[c] Fudan University Taizhou Institute of Health Sciences, Taizhou 225300, Jiangsu, China

[d] Department of Human Anatomy, School of Basic Medicine, China Medical University, Shenyang 110122, China

[e] Public Security Bureau of Zibo City, Zibo 255000, Shandong, China

[f] Criminal Investigation Department of Yunnan Province, Kunming 650021, Yunnan, China

[g] Department of Oral & Maxillofacial-Head & Neck Oncology, Shanghai Ninth People's Hospital, College of Stomatology, National Clinical Research Center for Oral Diseases, Shanghai Key Laboratory of Stomatology & Shanghai Research Institute of Stomatology, Shanghai Jiao Tong University School of Medicine, Shanghai 200011, China

[#] These authors contributed equally

* Send correspondence to

Shilin Li,

MOE Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200438, China

E-mail: lishilin@fudan.edu.cn

Xingjun Qin,

Shanghai Ninth People's Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai Key Laboratory of Stomatology, Shanghai 200011, China

E-mail: qinxj1989@sina.com

Caiyong Yin,

MOE Key Laboratory of Contemporary Anthropology, Department of Anthropology and Human Genetics, School of Life Sciences, Fudan University, Shanghai 200438, China

E-mail: cyyin18@fudan.edu.cn

Highlights:

- To resolve the complicated relationships within close relatives or among unrelated individuals from different geographic areas, special importance was given to Y Chromosomal SNPs;
- Based on capillary electrophoresis, a novel Y-SNP Pedigree Tagging System typed via rapid ARMS-PCR was proposed, which can precisely predict 11 haplogroups and 13 sub-haplogroups selected from the ISOGG Y-DNA tree;
- The results of this multiplex system were consistent within male relative pairs and showed the different distribution patterns of males branching from various Y-DNA haplogroups;
- The introduction of Y-SNPs into forensic practices may help define the potential ancestry of near-matching Y-STR haplotypes, and hopefully enhance the male pedigree discriminability.

Abstract

The Y chromosomal short tandem repeats (Y-STRs) have been used widely to establish paternal relatedness and examine sub-structures in different geographical regions. However, the applications of Y-STRs showed their limitations when it comes to resolving the complicated relationships within close relatives or among unrelated individuals from different geographic areas. Here, we overcome these limitations by introducing a new strategy for Y-SNP multiplex typing using rapid ARMS (amplification-refractory mutation system) PCR. Newly developed Y-SNP Pedigree Tagging System is able to profile 24 Y-SNPs in a single reaction while the whole process takes 4-5 hours. The panel precisely defines the 11 haplogroups (E-M96, D-JST021355, N-M231, C-M130, O-P186, I-M170, IJ-M429, K-M9, QR-M45, G-M201, and IJK-M522) and 13 sub-haplogroups (D1a1a1-N1, D1a2a-P47, C2-M217, N1a1-M46, O1a-M119, O1b-M268, O1b2-M176, O2-M122, O2a1-KL1, O2a2-P201, O2a2b-P164, O2a2a1a2-M7 and O2a2b1a1-M117). This system could contribute to providing the haplogroup affiliation of unknown

pedigree and resolving the sub-structures of East Asian populations. In this study, the multiplex system was validated for: ability to detect degraded DNA, sensitivity, species specificity, reproducibility/repeatability, stability, performance in different scenarios, mixture studies, PCR amplification conditions, and population surveys. The Y-SNP information showed a consistent pattern within 40 father-son or brother-brother pairs. The results of this multiplex system showed the different distribution patterns of male donors from two Chinese Han populations. In this study, we try to discriminate the suspect's pedigree on the level of Y-SNP haplogroups. These results show that Y-SNP Pedigree Tagging System is a robust and reliable amplification kit which can be used for male haplogroup determination.

Key Words

Developmental validation; Amplification-refractory mutation system; Y chromosome database; Pedigree discrimination; Population substructures

1. Introduction

In population genetics, the nonrecombining region of the human Y chromosome (NRY) has attracted much attention with respect to its unique inheritance characteristics [1]. Y chromosomal single nucleotide polymorphisms (Y-SNPs), which have relatively low mutation rates (approximately 1×10^{-9} substitution/generation [2]), are usually utilized to trace back the human origin and diverse dispersal routes. Meanwhile, Y chromosomal short tandem repeats (Y-STRs) mutate more rapidly than Y-SNPs (generally 3.78×10^{-4} to 7.44×10^{-2} mutation/generation [3]), and are used in evolutionary studies and genealogical analysis to date historical divergent events [4-6].

In the forensic field, Y-STRs are widely used in criminal casework for characterizing male contributions to mixed male-female biological materials; particularly in sexual assault cases [7], and in paternity casework involving male offspring, especially in deficiency paternity cases where the putative father is unavailable and replaced by one of his male relatives [8]. Detailed knowledge on mutation rates and mutation properties of the applied Y-STRs is required for the interpretation of near-matches in both types of forensic applications of Y-STR haplotypes [9]. At present, the Chinese male database is all based on Y-STR haplotypes to characterize male pedigree, and to discriminate closely or distantly related male lineages. When near-matches are observed in forensic casework - i.e., similar but not identical Y-STR haplotypes in the crime scene trace versus known suspect reference sample, or in the putative father (or his relative in deficiency cases) versus male offspring of whom paternity is to be established, the question arises if the allelic differences can be explained by mutations, or not. In criminal casework, explaining observed differences in the Y-STR haplotype profile by mutation(s) supports conclusions that an unknown relative of the known suspect may be the sample donor. In paternity casework, explaining observed Y-STR haplotype differences by mutation(s) allows

conclusions that a relative of the tested putative father may be the true biological father. Knowledge about mutability of forensically used Y-chromosomal STRs (Y-STRs) is relevant for interpreting near-matches of Y-STR profiles in criminal as well as paternity casework. Moreover, knowledge about the molecular factors influencing Y-STR mutability is limited and consensus on Y-SNP and Y-STR mutation rates remains to be determined [2, 10].

Y chromosomal haplogroups are defined by Y-SNPs and could explain the pedigree source of the near-matching Y-STR haplotypes. Additionally, consistency or explicit relationships of mutation events between Y-STR haplotypes and Y-SNP haplogroups would also provide compelling clues for male pedigree discrimination. Until now, several attempts have been made to bring Y-SNP genotyping into forensic practice. The advent of next-generation sequencing technology has made it possible to type hundreds of SNPs in a single batch. Apparently, many studies have attempted to sequence Y-SNPs using massively parallel sequencing (MPS) for forensic applications [11-13], while others developed new Y-SNP panels, which give more strength to SNaPshot technology for detecting degraded DNA [14] or shedding light on the genetic substructures [15].

According to the phylogenetic Y tree constructed by the International Society of Genetic Genealogy (ISOGG), all modern male individuals could be classified into 20 monophyletic or paraphyletic groups (*aka* Y Haplogroup A to T alphabetically), and their sub-haplogroups (https://isogg.org/tree/HaplogroupTr2019.html). The 1000 Genome Project describes the Y haplogroup distribution of East Asian men in general. Although, with significantly diverse distribution patterns, Chinese Han (*n*=98), Chinese Dai (*n*=50), Korean (*n*=46) and Japanese (*n*=56) males mainly belonged to the D, C, F, G, N, O and R haplogroups (https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/). Among these, the proportion of haplogroup O in Chinese (90.5%) is marginally higher than Koreans (87%) and Japanese (53.6%). In an earlier study, the investigators revealed that haplogroup O covers over 80.9% of Chinese Han men [16] while approximately 69,280 Y chromosomal SNPs or insertions and deletions (InDels) have been detected and listed in the 2019 Y tree by ISOGG (version: 14.177, date: 8 October 2019) by researchers using the whole genome sequencing method [17]. The dominant Y haplogroup in East Asia, O, C, D and N covers approximately 8,000 sub-haplogroups. The number of sub-haplogroups within haplogroup O indicated by the ISOGG Y tree has increased to 642.

The amplification-refractory mutation system (ARMS) is one kind of simple method for detecting single base changes or small deletions [18]. The multiplex ARMS is an ideal alternative, since Y-SNPs can be amplified in a single reaction system and detected by capillary electrophoresis (CE) technology. According to previous studies on East Asian male [6, 19-21], only four haplogroups (C, D, N and O) contribute >93% of the Y chromosomal genetic make-up. Haplogroup E, G, H, I, J, L, Q, R and T, contribute <7%. Y chromosomal markers can determine

these haplogroups or sub-haplogroups to shed light on population structures as well as for increased pedigree discrimination resolution for East Asian populations [22]. These selected markers are specific to 11 haplogroups (E-M96, D-JST021355, N-M231, C-M130, O-P186, I-M170, IJ-M429, K-M9, QR-M45, G-M201 and IJK-M522) and 13 sub-haplogroups (D1a1a1-N1, D1a2a-P47, C2-M217, N1a1-M46, O1a-M119, O1b-M268, O1b2-M176, O2-M122, O2a1-KL1, O2a2-P201, O2a2b-P164, O2a2a1a2-M7 and O2a2b1a1-M117). In East Asian males, the major haplogroups N (K1a1a) and O (K1a1b) are both sub-branches of the old lineage haplogroup K (47,000-50,000 years ago). It has been reported that basal K (not carrying any derived allele downstream) is remarkably rare and has not been fully explored [23]. According to the report released by ISOGG, the basal K haplogroup can be found at very low frequencies in many continents. Hence, marker M9 for haplogroup K was chosen for this panel, in order to (i) study its distribution, (ii) screen its sub-haplogroups, and (iii) apply it forensically for East Asian males. In Zhong *et al.* [24], the comparison between their results (*n*=3,826) and from the literatures (in total, *n*=4,933) [25-27] revealed that H-M69 (0.24%), L-M20 (0.24%) and T-M70 (0.04%) have low frequencies in East Asian males. Consequently, haplogroups L, H and T were not included in the initial ARMS-PCR panel.

In this current study, we have developed and validated the Y-SNP Pedigree Tagging System. This is a robust multiplex PCR system, using 5 dye technology, and was tested with degraded DNA, sensitivity, species specificity, reproducibility/repeatability, stability, performance in complex forensic practice, mixture studies, PCR amplification conditions, and population studies. This system has two main advantages: (i) the whole process takes only 4-5 hours, and (ii) it has similar readable electrophoretogram to different STR kits, which makes it more suitable for building forensic applications such as pedigree dataset reconstructing. The developmental validation was implemented strictly according to scientific standards by FBI QAS (2009/2011) [28], the International Society of Forensic Genetics (ISFG [29]) and the Scientific Working Group on DNA Analysis Methods guidelines [30].

2. Methods and Materials

The study was approved by the ethical review board of Fudan University Shanghai, People's Republic of China (code: BE1806; date: March 3rd, 2018) and in accordance with the standards of the Declaration of Helsinki.

2.1 Panel design and primer preparation

The Y-SNPs included in the 2019 ISOGG Y-DNA Haplogroup Tree (version 14.62 by 7 April 2019) were screened as part of the panel. SNPs that determine either Y haplogroups or sub-haplogroups were viewed as candidates. The screening criteria were set as follows: (i) Y-SNPs which determine major haplogroups and O sub-haplogroups with different distribution weights

in Chinese populations; and (ii) loci which could be integrated into one single reaction system *via* ARMS PCR. A total of 25 Y-SNPs were selected as candidates. However, F-M89 was removed due to the failure in primer design. Referring to the Hg38 human genome, Primer 5.0 software (PREMIER Biosoft International, Corina Way, Palo Alto, CA, USA) was employed to design primers that specifically match to these 24 Y-SNPs.

Some bases of these primer sequences were modified and optimized with the purpose of successful amplification. FAM, HEX, TAMRA and ROX dyes were used to label these primers, while the 5th dye (orange) was utilized to label the internal size standard. The design of panel detailing Y-SNP-related amplificons is presented in Fig.1, while the information of Y-SNPs and primers is summarized in Table S1. This primer set was the basis of the Microreader™ Y-SNP 5× Primer Mix within the Y-SNP Pedigree Tagging System.

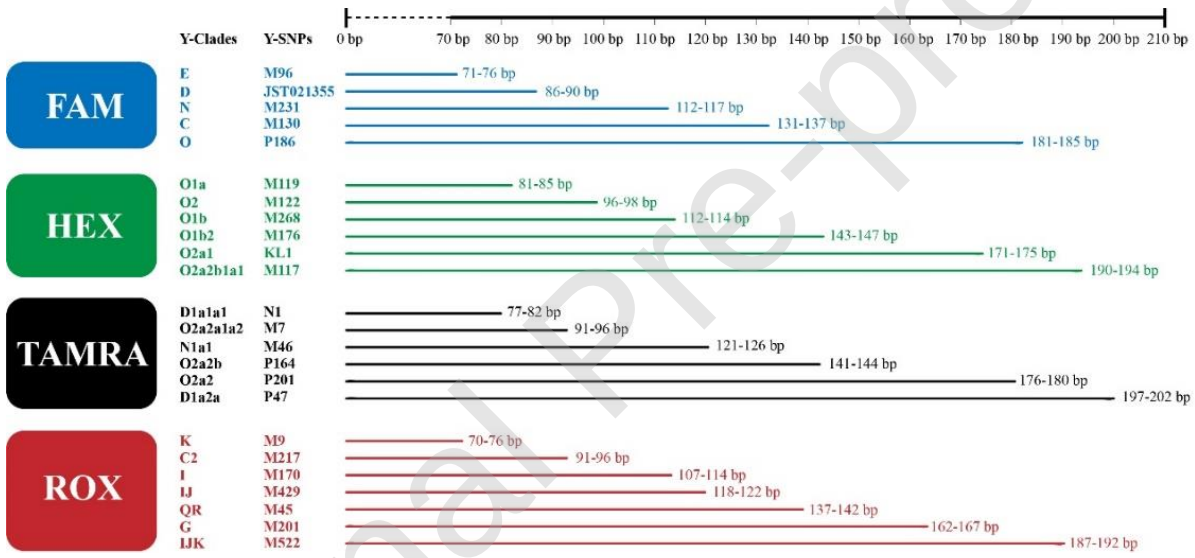

Fig.1 Configuration of the Y-SNP Pedigree Tagging System Different colors indicate the four-fluorescence labeling described in Table S1. Each horizontal line represents the amplicon length of the MT or WT allele at 24 Y-SNP loci (WT: wild type, *aka* ancestral allele; MT: mutation type, *aka* derived allele).

2.2 Optimization of PCR amplification conditions

In order to optimize the PCR conditions, emphasis was on the details about the concentrations of Microreader™ 2.5× Master Mix I and Microreader™ Y-SNP 5× Primer Mix (Suzhou Microread Genetics, Suzhou, Jiangsu, China), PCR instruments, the sizes of reaction volume, different cycles, annealing temperature, and the temperature and time of the final extension step (the number of replicates = 3). Control DNA 9948 was used.

The concentration of the added reagents varied at 50%, 75%, 100%, 125% and 150%. The 100% referred to be an ideal concentration of various reagents. The various components included in the base 100% concentration of Master Mix I are: potassium chloride (KCL, 50 mM), magnesium chloride ($MgCl_2$, 2 mM), tris hydrochloride (Tris-HCL, 35 mM), bovine serum albumin (BSA, 0.3 mg/ml), deoxyribonucleotide triphosphate (dNTP, 0.2 mM), ammonium sulfate ($(NH_4)_2SO_4$, 2.5 mM), dimethyl sulfoxide (DMSO, 6% in volume), betaine (100 mM), and *Thermus aquaticus* (Taq) DNA polymerase (1 U). While, the primer concentrations in the base 100% concentration of Primer Mix are summarized in Table S1.

Four thermocyclers (Veriti™ Thermal Cycler, GeneAmp® PCR system 9700, 2720 Thermal Cycler, and ProFlex™ PCR System (Applied Biosystems, CA, USA)) were used for the DNA amplification with different reaction volumes (5, 7, 10 and 25 µL). Other conditions were determined by altering the number of cycles (27-29), annealing temperature (57-61 °C), extension temperature (62, 65, 68, 72 and 74 °C), and extension time (30, 45 and 60 minutes). Additionally, cross-variations of annealing temperature (58, 59 and 60 °C) and extension temperature (68, 72 and 74 °C) were done to further optimize the system.

2.3 Multiplex amplification and electrophoresis analysis

Amplification was performed in a GeneAmp®PCR System 9700 thermal cycler (Life Technologies, CA, USA) on a gold-plated silver block. After optimization, the final reaction system was performed in a 25 µL volume containing 10 µL Reaction Mix of Microreader™ 2.5× Master Mix I, 10 µL of Microreader™ Y-SNP 5× Primer Mix, and 2 ng of template DNA under the following PCR conditions: pre-incubation at 95 °C for 5 minutes, 28 cycles of denaturation, annealing and extension step at 94 °C for 30 seconds, 59 °C for 60 seconds, and 72 °C for 60 seconds, followed by a final extension step at 60 °C for 30 minutes, and a final soak at 4 °C.

A customized allelic ladder was developed for the Y-SNP Pedigree Tagging System. The PCR products were detected and separated by capillary electrophoresis using Applied Biosystems 3500xl Genetic Analyzer (Life Technologies, CA, USA) with POP-4™ polymer (Life Technologies) and dye set of E5. The PCR product (1 µL) or allelic ladder (1 µL) was mixed with a mixture of 8.5 µL of Hi-Di formamide (Applied Biosystems) and 0.5 µL of QD550 internal size standard (Suzhou Microread Genetics, Suzhou, Jiangsu, China). Final mixture (PCR product or Allelic ladder, Hi-Di formamide and ISS) was denatured at 95 °C and chilled on ice for 3 minutes. Samples were injected at 3 kV for 10 s and electrophoresed at 15 kV for 1500 s at a run temperature of 60 °C. Data were analyzed by GeneMapper™ ID-X1.5 Software (Applied Biosystems). All peaks were set with an analytical threshold of 50 relative fluorescence units (RFU).

2.4 Validation studies

To evaluate the accuracy and precision of the SNPs genotyping, three replicates were performed for each sample. In order to make sure that genotyping is accurate, sequencing experiments were performed. As shown in Table S1, the PCR products were not long enough to be directly sequenced by Sanger sequencing, so a total of 48 sequences, which matched the amplicons of different alleles, were synthesized and constructed in the plasmids (Table S2). The Sanger sequencing results demonstrated that these constructed plasmid sequences successfully bind to the PCR products which shows that the typing of 24 Y-SNPs through this kit was correct (data not shown).

2.4.1 Ability to detect degraded DNA

DNA control 9948 (Marligen Biosciences, Ijamsville, MD, USA) was damaged to different degrees with ultraviolet light of 1300 Series A2 Biological Safety Cabinet (Thermo Fisher Scientific, Waltham, MA, USA). The UV degradation was performed for 15 to 60 minutes with an interval of 15 minutes, and both the Microreader™ 21 (Direct) ID System (Suzhou Microread Genetics, Suzhou, Jiangsu, China) and Y-SNP Pedigree Tagging System were used to count the success allele calling after DNA profiling. The ratio of loci detected was calculated by dividing the number of successful typing alleles by 24.

2.4.2 Sensitivity

For the sensitivity study, a serial concentration of control DNA 9948 (1.000000, 0.500000, 0.250000, 0.125000, 0.062500, 0.031250 and 0.015625 ng) were amplified in triplicate. Control DNA 9948 (10 ng/μL) was serially diluted with distilled water. Amplification processes were performed according to the manufacturer's instructions.

2.4.3 Species specificity

Whole genomic DNA from *Sus scrofa* (pig), *Bos taurus* (cattle), *Ovis aries* (sheep), *Gallus gallus* (chicken), *Anas platyrhynchos* (duck), *Rattus norvegicus* (brown rat), *Mus musculus* (mouse), *Columba livia domestica* (pigeon), *Canis lupus familiaris* (dog), *Equus caballus* (horse) and one non-human primate species – *Macaca fascicularis* (crab-eating monkey) was extracted by magnetic particles method [31]. 2 ng of DNA was used for each reaction volume for further amplification according to above-mentioned PCR conditions.

2.4.4 Reproducibility/Repeatability

After amplification using the Y-SNP Pedigree Tagging System, the amplicons of 9948 were detected through ABI3130XL, ABI3500, and ABI3730XL (Applied Biosystems, Waltham, MA, USA).

Blood or saliva samples from 598 unrelated healthy male individuals were collected on Flinders Technology Associates (FTA) cards (Whatman International Ltd., Maidstone, UK). In order to check the genotyping reproducibility/repeatability, these samples were independently genotyped at different forensic biology laboratories: (i) the MOE Key Laboratory of Contemporary Anthropology of the School of Life Sciences, Fudan University, (ii) the Complex laboratory 1 of Suzhou Microread Genetics, and (iii) the DNA Laboratory of Public Security Bureau of Zibo, Shandong province.

2.4.5 Stability

In order to test the stability of this robust Y-SNP Pedigree Tagging System, common forensic inhibitors were mixed with control DNA 9948 for the subsequent PCR experiments. The concentrations of three inhibitors were diluted, as follows: indigotine (0, 4, 8, and 12 mM), humic acid (0, 50, 100, 150 and 200 ng/μL), and hematin (0, 100, 200, 300 and 400 μM) by Sigma-Aldrich, Saint Louis, MO, USA.

2.4.6 Performance in different scenarios

Keeping the operative demands in mind, we have used different biological samples to check the performance of this kit. The biological sources of these samples were as follows: blood stain, saliva stain, semen stain, hair follicle, and human tissue (costicartilage). These five kinds of biological samples were obtained from 15 respective individuals, with three samples per biological source. No lineage relationship was observed among these sample donors. The blood and saliva stains were deposited on FTA cards, while the semen stains were stored on a cotton swab. The DNA was extracted from human tissue using the magnetic particles method [31] for further analysis, while the other four kinds of samples were directly amplified.

The Y-SNP panel was also tested using blood stains on FTA cards provided by 40 father-son or brother-brother pairs from 36 families to observe the consistency of Y-SNP through gamete transmission. The relation consanguinity of these father-son or brother-brother pairs has been validated by both pedigree investigation and paternity testing (autosomal STRs, data not shown).

2.4.7 Mixture studies

Artificial male-female and male-male mixture samples were prepared using control DNA 9947A, 9948 (Marligen Biosciences, Ijamsville, MD, USA) and private Sample_A (male). The mixtures of male 9948 to female 9947A were set as 1.0000, 0.5000, 0.2500, 0.1250, 0.0625 and 0.0000 ng of control DNA 9948 per 50 ng of 9947A (ratios: 1:50, 1:100, 1:200, 1:400, 1:800 and 0:50). For male-male mixture studies, the DNA concentrations of commercial DNA 9948 (Haplogroup QR-M45) and private Sample_A (Haplogroup O1b-M268) were set as 2 ng/μL. The ratios of the male-male mixtures were 1:9, 1:3, 1:1, 3:1 and 9:1.

2.4.8 Population surveys and statistical analysis

Initially, MT allelic frequencies were retrieved by referring to the 1000 Genomes Browser (phase 3, website: https://www.ncbi.nlm.nih.gov/variation/tools/1000genomes/) [32, 33]). The distributions of the allelic frequencies of 21 Y-SNP loci in the global scale and all five East Asian populations (CDX: Chinese Dai in Xishuanbanna, CHB: Han Chinese in Beijing, CHS: Southern Chinese Han, JPT: Japanese in Tokyo, and KHV: Kinh in Ho Chi Minh City) were obtained. The remaining three Y-SNPs (D1a1a1-N1, D1a2a-P47 and O2a2a1a2-M7) were unavailable in the 1000 Genomes (1KG) database, but allelic frequencies of these SNPs were obtained from other public databases or literatures [34-36].

To study the forensic genetic characteristics of the Y-SNP Pedigree Tagging System, a population study was performed by genotyping the 598 unrelated healthy male individuals which were living there for at least three generations. Among these, 288 Wujiang Han samples were from Jiangsu province (saliva samples on FTA cards) and 310 Zibo Han samples were from Shandong (blood samples on FTA cards). An informed consent was obtained from each participant. Optimized reaction conditions were applied to genotype all alleles at the 24 Y-SNPs of 598 samples. The MT allelic frequencies were calculated by direct counting. Furthermore, the assignment of an individual to the Y-DNA tree was determined on the basis of the profiles of the 24 Y-SNPs. Then, the proportion of males assigned to the Y-DNA haplogroups was computed in term of percentage.

2.5 Quality control

The validation of Y-SNP Pedigree Tagging System was conducted according to the recommendations of FBI QAS (2009/2011) [28], ISFG [29] and SWGDAM [30]. Nomenclature of the Y-SNP was adopted in accordance with ISOGG, and the corresponding custom Y tree was illustrated.

3. Results

3.1 PCR conditions

The optimization steps of PCR protocol are illustrated via bar chart in Fig.2.

3.1.1 Concentrations of Master Mix I

The concentration of Master Mix I was adjusted to balance the allele calling. The 100% referred to the ideal concentration of Master Mix I. With the exception of 50% or less than that amount of Master Mix I, all gave full profile for control DNA 9948. The peak heights of 100% and 125% of Master Mix I are shown in Fig.S1. The peak height balance was best using the 100% Master Mix I. Thus, the optimal concentration of Master Mix I was set at 100%.

3.1.2 Concentrations of 5× Primer Mix

The results (Fig.S2) showed the peak variations at various concentrations (50%, 75%, 100%, 125% and 150%). We have observed ideal peak heights at 100% concentration. Hence, 100% of 5× Primer Mix is recommended for completely genotyping these 24 Y chromosomal loci.

3.1.3 PCR instruments

Different laboratories have used different thermal cyclers, so four different thermal cyclers were used. The results indicated that the four thermal cyclers performed well in the amplification reactions, with the peak height analysis threshold setting at 50 RFU. The variety in peak heights and peak balance may be correlated to the block ramp rate of these different thermal cyclers (2-6 °C per second) and the inevitable minuscule sample loading errors (Fig.S3).

3.1.4 Cycles

At 27-29 cycles, the amounts of PCR products were in the detection range of genetic analyzer (Fig.S4). In order to balance the peak height and increase the time efficiency, we recommend 28 cycles are suitable for the amplification of this Y-SNPs panel.

3.1.5 Annealing temperatures

The optimum annealing temperature was observed at 59 °C with ideal peak balance and electrophoretogram integrity. Lower temperature would lead to an excessively high peak height, while higher temperatures would cause allele drop-outs (Fig.S5).

3.1.6 Extension temperatures and durations

At an extension temperature of 74 °C, the peak heights started to decrease when compared with 68 °C or lower temperatures. At 72 °C, all normal/ideal peaks were observed (Fig.S6). Moreover, in order to ensure the +A additions and maintain the time efficiency, 30 minutes of the final extension duration was enough for the extension reactions (Fig.S7).

3.1.7 Cross-variations of annealing and extension temperatures

Cross-variations were conducted to further optimize the annealing and extension temperatures for the PCR conditions. Three candidate temperatures (annealing temperatures: 58, 59 and 60 °C; extension temperatures: 68, 72 and 74 °C) were selected for each condition. The nine trials revealed that the peak heights and balance were best at 59-72 °C of annealing-extension temperatures (Fig.S8), and either high annealing (> 59 °C) or extension (> 72 °C) temperature would be likely to decrease the amplification efficiency.

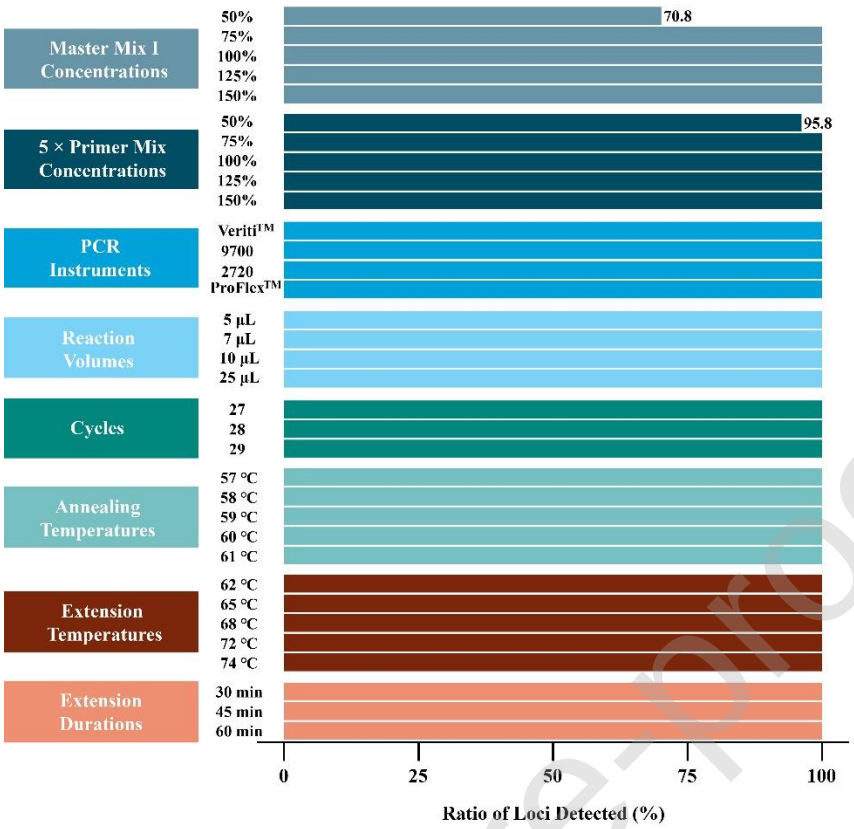

Fig.2 Various amplification conditions tested for the optimized PCR process The amplification conditions included the Master Mix I concentrations, 5× Primer Mix concentrations, PCR instruments, sizes of the reaction volume, cycles, annealing temperatures, and temperatures and time of the final extension step (the number of replicates = 3). The lengths of rectangles indicate the ratios of loci detected.

3.2 High-level performance in detecting degraded DNA

The control DNA 9948 was treated with UV light and then genotyped using Y-SNPs and autosomal STR kits (Fig.3 and Fig.S9). The drop-out of the SNP allele was initially observed at 45 minutes, while that of the STR allele at 30 minutes. Additionally, it was also found that amplicons specific to M117 and amplification fragments greater than 275 base pairs were undetectable in some cases.

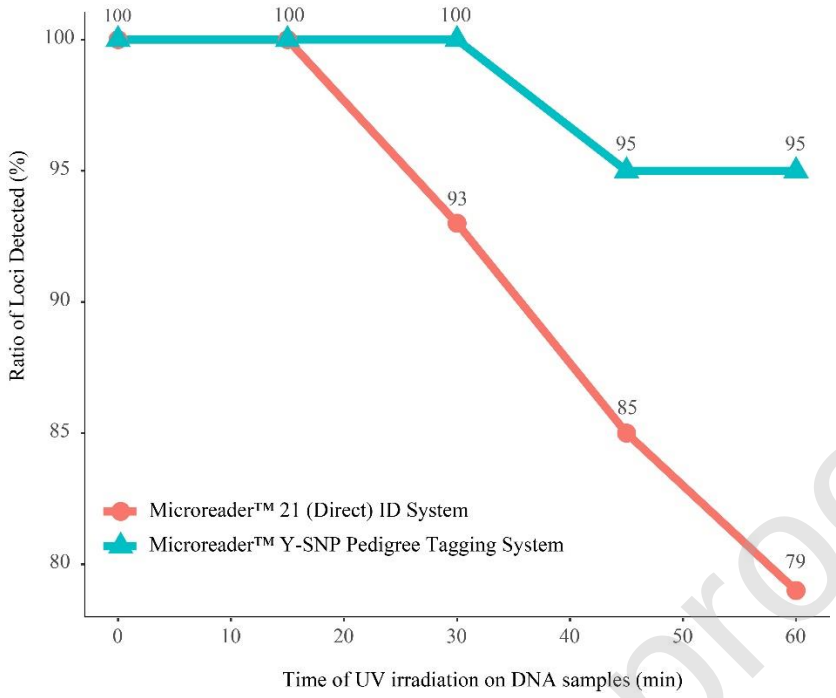

Fig.3 The influence of different UV irradiation durations on Y-SNP typing The control DNA was treated by UV irradiation for 0, 15, 30, 45 and 60 minutes, respectively (the number of replicates = 3). The blue triangles and orange circles indicate the ratios of loci detected (%) using the Y-SNP Pedigree Tagging System and Microreader™ 21 (Direct) ID system.

3.3 Sensitivity

The results indicated that 31.25 *pg* or a lower amount of DNA template was not completely amplified (Figs.4 and S10). When 62.5 *pg* amount of DNA template was used, complete amplification was obtained which shows that the Y-SNP panel exhibited high sensitivity.

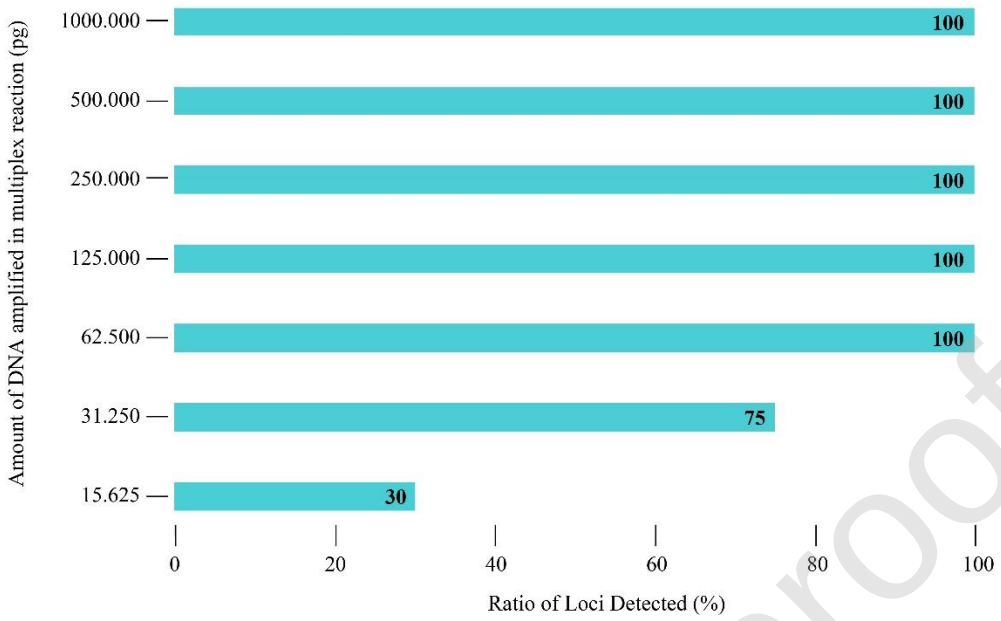

Fig.4 Sensitivity measurement of the Y-SNP Pedigree Tagging System The 7 different amounts of control DNA amplified in the reaction volume were 1000.000, 500.000, 250.000, 125.000, 62.500, 31.250 and 15.625 *pg* (the number of replicates = 3). The blue rectangles illustrate the ratios of loci detected.

3.4 Specificity

Male and female samples from non-primate species (pig, cattle, sheep, chicken, duck, brown rat, mouse, pigeon, dog and horse) and the male sample from non-human primate species (crab-eating monkey) were used. However, we did not observe any amplicon (male, Fig.S11A; female, Fig.S11B) which shows that the set of 24 Y-SNP primers are highly specific to *Homo sapiens* (Human), which could be utilized for practical forensic applications.

3.5 Reproducibility/Repeatability

The control DNA 9948 was amplified utilizing the Y-SNP Pedigree Tagging System in three different reactions, and amplification products were detected by three different genetic analyzers (ABI3130XL, ABI3500, and ABI3730XL). Results showed the consistency in peaks (allele call) on three different platforms of genetic analyzers (Fig.S12).

Subsequently, the matching Y-SNP profiles of 598 male individuals were obtained from three forensic biology laboratories, which revealed that Y-SNP Pedigree Tagging System is both reproducible and repeatable.

3.6 Stability

DNA from forensic casework samples often contain substances which can inhibit PCR reaction. The most common inhibitors encountered are from elements intrinsic to the sample itself (humic acid from soil and hematin from blood [37, 38]). Humic acid inhibits amplification by binding template DNA, while hematin inhibits amplification by curbing Taq polymerase [39]. Our results indicated that indigotine (0, 4, 8 and 12 mM) had no negative effects (Fig.S13A). Full profiles were observed with humic acid (≤ 50 ng/μL) and hematin (≤ 200 μM) (Fig.S13 B, C). The peak height decreased with increasing inhibitor concentration, and some peaks dropped when humic acid was ≥ 100 ng/μL and hematin ≥ 300 μM.

3.7 Applications in forensic practice

Different sources of human biological materials such as blood stain, saliva stain, semen stain, hair follicle, and human tissue (costicartilage) were collected and typed with the Y-SNP Pedigree Tagging System. Only the DNA of human tissue was extracted while other samples were amplified directly using the DNA extraction-free method. The complete profile of each sample showed that this kit is capable of detecting samples from various sources (Fig.S14 A-E, respectively).

In addition, the Y-SNP profiling of 40 father-son or brother-brother pairs were classified into Y tree branches in Table S3. We did not observe any mutation within the pedigrees, since it is extremely unlikely for new or reverse mutation events to occur at these haplogroup-determining Y-SNP loci. Therefore, the existing Y-DNA haplogroup tree could be employed to tag each male pedigree.

3.8 Mixture studies

Mixtures of DNA appear frequently in forensic casework, especially in sexual assault cases [40]. Template DNA mixed by 9948 and 9947A at various ratios (1:50, 1:100, 1:200, 1:400, 1:800, and 0:50) was tested in triplicate. The peak height of 9948 decreased with input reduction (Fig.S15). For the male-male mixture, the electrophoretic patterns demonstrated that all peaks of MT/WT alleles could be detected in the mixture ratios (9948 : Sample_A) of 1:9, 1:3, 1:1 and 3:1 (ratio of loci detected = 100%). Only in 9:1 was the drop-out of the WT allele observed at locus M45 of Sample_A (Fig.S16) and ratio of loci detected was 95.83%. In Fig.S16, the red box indicated the drop-out observations when compared with other electrophoretograms.

3.9 Population data

On the basis of ISOGG 2019 Y-DNA haplogroup Tree, we have demonstrated a customized Y haplogroup tree which is specific to the selected 24 Y-SNP loci (Fig.5). Previous studies and their MT allelic frequencies showed highly polymorphic nature among these 21 SNPs in six populations [32, 33]. According to the Y-DNA tree trunk phylogeny, the K haplogroup was formed by K1 (including L and T) and K2 sub-branches. The deep branches of K2 are K2a (including N and O), K2b (including Q and R), K2c, K2d and K2e. For East Asian males, major haplogroups N (K1a1a) and O (K1a1b) were both sub-branches of the old lineage haplogroup K (47,000-50,000 years ago) [23]. Thus, in a broader sense, K with MT allele at M9 covers QR, N and O. In Chinese male population with K haplogroup ancestry, either from the K sub-branches or basal K, composes the major structure (>90%). As Fig.5 showed, substructures also existed in the three Chinese populations. In particular, 13% of CHS individuals were from the N haplogroup, but no samples were observed in CDX or CHB, and *vice versa* in the Chinese O haplogroup. Moreover, distinct differences among the O sub-haplogroups were revealed, such as the similar O1b proportion in two Han populations, which was much higher than that of CDX, and the relatively low O2a1 proportion of CHS (15%), which was nearly half of CHB (33%).

In previous studies, the allelic frequencies were reported for the remaining 3 Y-SNPs: D1a1a1-N1 (rs3212291), D1a2a-P47 (rs868302452) and O2a2a1a2-M7 (rs3898). In the genome Aggregation Database (gnomAD) [34], the MT allelic frequency of rs3212291 is 0.0002 (all populations, *n*=13) and 0.0025 (East Asian populations, *n*=11). For D1a2a-P47, Wang *et al.* [35] found that in Qiangic Populations living in the Western Sichuan Corridor, the average frequency of the rs868302452 MT allele is 0.1969 (*n*=25). Furthermore, Gayden *et al.* [36] characterized that 18.6% of Tibetans belonged to sub-haplogroup D1a2a-P47. The data submitted to the dbSNP by the Stanford Genome Center indicated that the MT allelic frequency of rs3898 is 0.0014 (global scale, *n*=1).

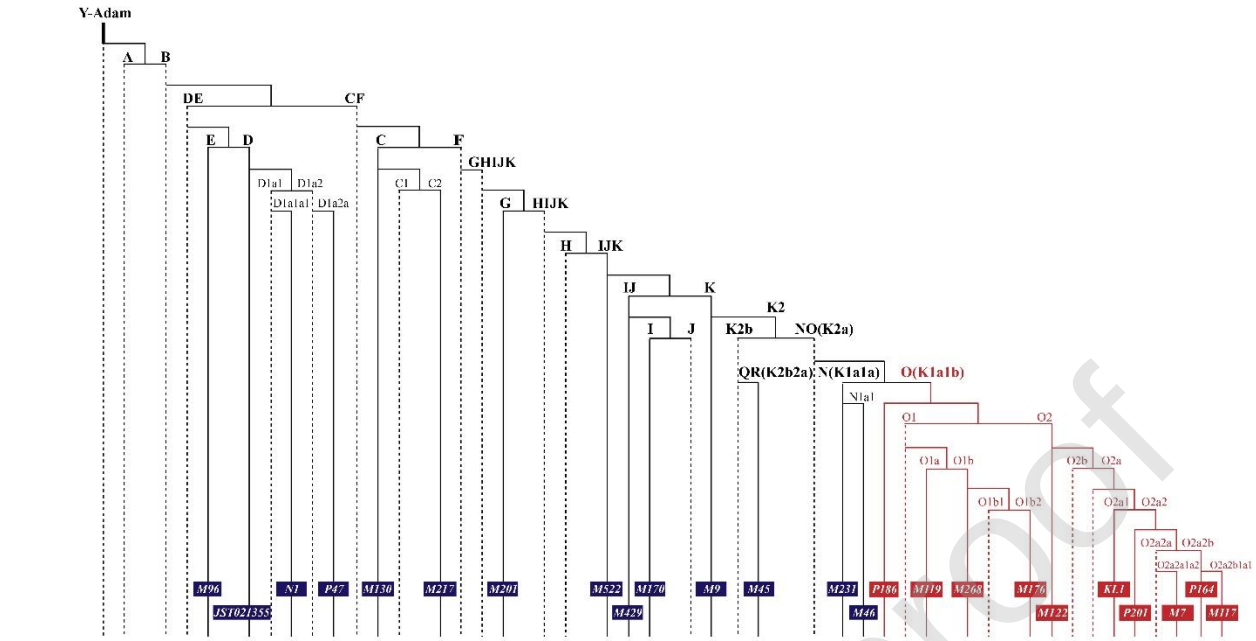

Fig.5 Distribution of the selected 24 Y-SNPs and determination of the Y-DNA haplogroups in the form of a custom ISOGG 2019 Y tree The targeted 24 Y-SNPs are indicated in solid lines, while the dotted lines indicate the haplogroups that were not covered. Y-SNPs that belong to the non-O haplogroups are depicted in blue color, while those from the O haplogroup are depicted in red color. Since three Y-SNPs are not included in the 1000 Genome Project phase III, the MT allele frequencies of the remaining 21 loci from the global population and all five Eastern Asian populations were revealed (CDX: Chinese Dai in Xishuanbanna, CHB: Han Chinese in Beijing, CHS: Southern Chinese Han, JPT: Japanese in Tokyo, and KHV: Kinh in Ho Chi Minh City).

The initial reference data revealed a high level of heterozygosity among the contributing Y-DNA haplogroups in various Chinese populations. In this study, two Chinese Han populations were selected to conduct population comparison. The MT allelic frequencies of the 24 Y-SNP loci and their percentage assigned to the 16 Y-DNA haplogroups are summarized in Fig.6.

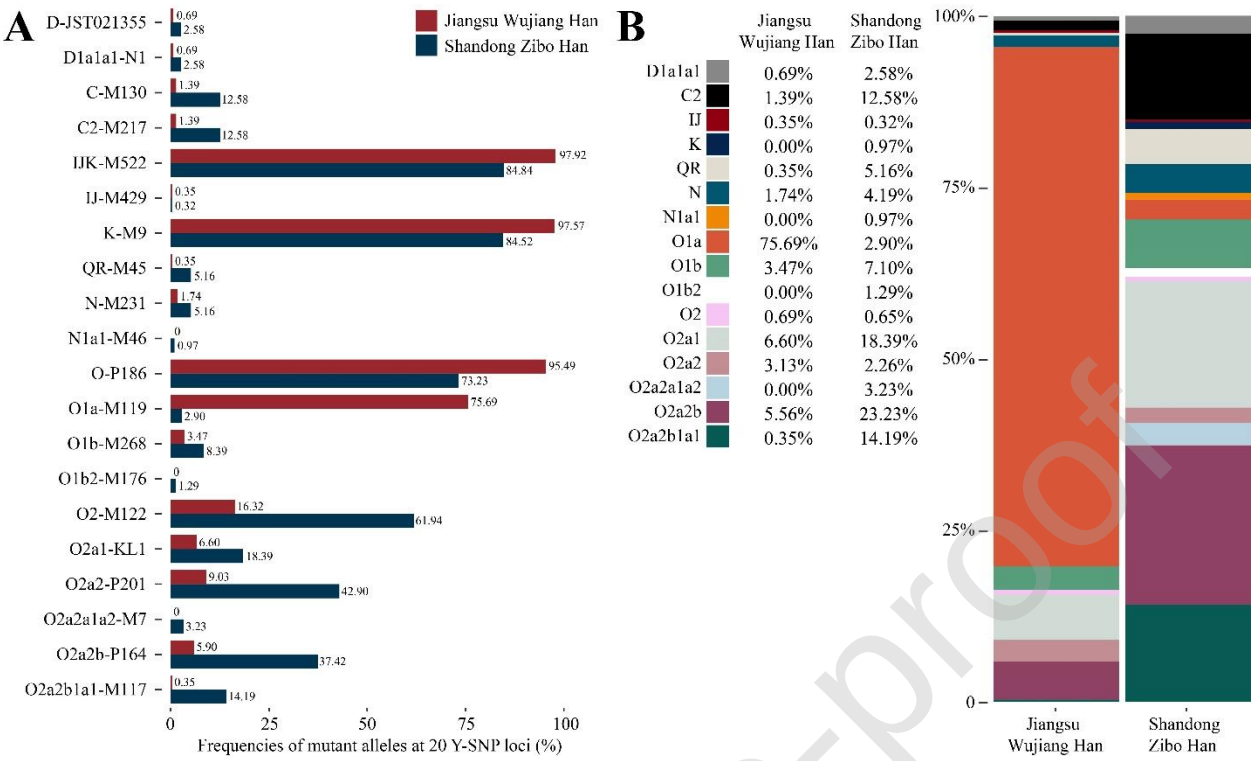

Fig.6 The MT allele frequencies and sub-haplogroup proportions investigated in the Jiangsu Wujiang Han (n=288) and Shandong Zibo (n=310) populations (A). The frequencies of MT alleles at the selected 24 Y-SNP loci were compared between the Jiangsu Wujiang Han (red) and Shandong Zibo Han populations (blue). (B) All 598 male individuals were assigned to 16 haplogroups or sub-haplogroups, as indicated in different colors. Each male individual would be labeled for only one haplogroup. Furthermore, those who have been assigned to relatively high-resolution Y-DNA haplogroups (such as O2a2) were excluded from the upstream haplogroups (such as O2 or O). This distribution assisted in observing the proportion distributions of individuals who could not be assigned to deeper branches.

The allelic frequency differences revealed the different outcomes of multi-stage population migration and mixture. No MT alleles were detected in E-M96, D1a2a-P47, G-M201 and I-M170 loci, so these haplogroups are not included in Fig.6A. Mainly, more male individuals from the Wujiang Han population belonged to the O haplogroup, when compared to Zibo Han population. On the other hand, in the Wujiang Han population, low-frequency mutations were observed in the other three dominating haplogroups (D, C and N), which makes up to 0.69%, 0.69%, and 1.74% while 2.58%, 12.58% and 5.16% in Zibo Han. In addition, 95.49% of Wujiang Han and 73.23% of Zibo Han belongs to O haplogroup. Among these, the O sub-haplogroups exhibited a significantly dichotomous distribution. Most of the Wujiang Han individuals of O haplogroup were categorized into O1 (82.90%) including 79.27% of O1a and

3.63% of O1b. In the Zibo Han population, 84.58% of O haplogroup individuals were from the O2 sub-haplogroup (Fig.6A). Depending on the Y-SNP profile of these 598 samples, they belong to 16 Y chromosomal clans which is shown in Fig.6B, indicating the structure heterozygosity of the primary Y-DNA haplogroup and high-resolution of O sub-haplogroups in the Wujiang and Zibo Han populations.

4. Discussion

The Y-DNA haplogroup polymorphisms provided insight into the long history of human migration and admixture. The mutation rates of Y-SNPs were lower than Y-STRs (2×10^{-8} *vs.* 5.9×10^{-3} mutation per marker per generation[3]). According to the Y-DNA tree, the identical Y-SNPs of male descents from the same ancestor would be defined by determining the Y-DNA variations among the Y haplogroups.

In forensic genetics, the pedigree discrimination is still a challenge [41]. Different working groups put much emphasis on STRs development and database construction [42]. To address the important issue related to pedigree discrimination, ARMS-PCR technology, which has been previously utilized for other forensic applications [43-45], allows the amplified products to be detected using capillary electrophoresis and whole process will be completed within 4-5 hours. The method utilized for male pedigree tagging has the prospect of being used in primary laboratories where high-throughput sequencing technology is unavailable.

The markers which are part of the existing Chinese Y chromosomal database are all Y-STRs, and its main aim was to exclude male suspects from involvement in crime, as indicated by Kayser *et al* [46]. In particular, emphasis was on rapidly mutating Y-STR loci to discriminate between male relatives or male with near-matching Y-STR haplotypes. Promisingly, in forensic practice, the Y-SNP haplogroups combined with the Y-STR haplotypes would provide more useful investigative leads. When handling forensic cases, the search in the existing database would generate a list of targeted individuals carrying the same or near-matching Y-STR haplotypes for further investigation. Previous scientific research has revealed the possibilities of Y-STR haplotype similarities among Y-SNP haplogroups [47]. Given the huge population base of China, Chinese investigators may give much efforts in ruling out possible suspects or victims. The males who branched from other Y-SNP sub-haplogroups would be precluded, and the investigative work would become less exhausting.

In Y-SNP Pedigree Tagging System panel, half SNPs are from the upstream or downstream of the O haplogroup, which makes this panel powerful for detection of O haplogroup along with its classification in Chinese population. Moreover, a small segment covers high resolution Y-SNP loci (D1a1a1-N1, N1a1-M46, O1a-M119, O1b-M268, O1b2-M176, O2a1-KL1, O2a2-P201,

O2a2a1a2-M7, O2a2b-P164 and O2a2b1a1-M117). According to new ISOGG Y haplogroup tree, only four main haplogroups (C, D, N and O) are further subdivided into 1,474 sub-haplogroups in the Chinese population. With the advancement in technology and more in-depth knowledge of genetic genealogy, this number will be increased. In the currently proposed panel, haplogroup L, H and T were not included because these are almost absent in East Asian populations. The main aim of Y-SNP Pedigree Tagging System panel is to address major Y-SNP haplogroup and some predominant sub-haplogroups in Chinese population.

Overall, in forensic identification, we mostly deal with unknown samples/biological materials, therefore it would be good for each sample to be initially typed with this panel (sensitivity=62.5 *pg*). Suppose that customized high-resolution Y-SNP panels under different major Y-SNP haplogroups are successfully developed, the initial typing results will give us a lead toward precise selection of panel for unraveling the possible affinity of the forensic sample to the targeted pedigree. This kind of typing strategy would protect the sample quantity from being wasted. Hopefully, the effectiveness of searching potential suspects within pedigree alternatives would be significantly facilitated with Y-SNP Pedigree Tagging System and quality of database development will also be enhanced.

5. Conclusion

In this current study, the investigators have proposed a new Y-SNP panel "Y-SNP Pedigree Tagging System", which has the ability to type 24 Y-SNPs in one multiplex system. After the validation work, we have tested its ability to detect degraded DNA, sensitivity, species specificity, reproducibility/repeatability, stability, performance in different scenarios, mixture analysis, PCR conditions and population study. This kit was optimized and found to be time– and cost–saving. Furthermore, the Y-SNPs information showed a consistent pattern within each pedigree of 40 father-son or brother-brother pairs, and different distribution patterns of males branching from various Y-DNA haplogroups. We also shed light on pedigree discrimination in the level of Y-SNP haplogroups, and lay a solid foundation for further Y database development using the newly proposed Y-SNP typing strategy. In the future, more emphasis should be given on database development, and customized panels that focus on the low-frequency Y-SNP loci which define major haplogroups and downstream high-resolution Y-SNPs under different major haplogroups.

Conflicts of Interest

The authors declared there are no conflicts of interest, and they are not employees of any commercial company and receive no compensation from any commercial company.

Acknowledgements

The investigators thank all the donors for providing biological samples. The monkey skin tissue is donated by Dr. Junqing Chen (State Key Laboratory of Reproductive Medicine, Department of Histology and Embryology, Nanjing Medical University, Nanjing, Jiangsu, China). This work was supported partly by the Science and Technology Committee of Shanghai Municipality [grant number 2017SHZDZX01] and the Institute of Forensic Science, Ministry of Public Security, People's Republic of China [grant numbers 2015FGKFKT01, 2017JB004].

References

[1] P.A. Underhill, T. Kivisild, Use of Y Chromosome and Mitochondrial DNA Population Structure in Tracing Human Migrations, 41(1) (2007) 539-564.
[2] O. Balanovsky, Toward a consensus on SNP and STR mutation rates on the human Y-chromosome, Human genetics 136(5) (2017) 575-590.
[3] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. von Wurmb-Schwark, P. de Knijff, D. Labuda, H. Vezina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, Am J Hum Genet 87(3) (2010) 341-53.
[4] L.A. Zhivotovsky, P.A. Underhill, C. Cinnioglu, M. Kayser, B. Morar, T. Kivisild, R. Scozzari, F. Cruciani, G. Destro-Bisol, G. Spedini, G.K. Chambers, R.J. Herrera, K.K. Yong, D. Gresham, I. Tournev, M.W. Feldman, L. Kalaydjieva, The effective mutation rate at Y chromosome short tandem repeats, with application to human population-divergence time, Am J Hum Genet 74(1) (2004) 50-61.
[5] S. Yan, C.-C. Wang, H.-X. Zheng, W. Wang, Z.-D. Qin, L.-H. Wei, Y. Wang, X.-D. Pan, W.-Q. Fu, Y.-G. He, L.-J. Xiong, W.-F. Jin, S.-L. Li, Y. An, H. Li, L. Jin, Y chromosomes of 40% Chinese descend from three Neolithic super-grandfathers, PloS one 9(8) (2014) e105691-e105691.
[6] C.C. Wang, H. Li, Inferring human history in East Asia from Y chromosomes, Investig Genet 4(1) (2013) 11.
[7] M. Prinz, K. Boll, H. Baum, B. Shaler, Multiplexing of Y chromosome specific STRs and performance for mixed samples, Forensic Sci Int 85(3) (1997) 209-18.
[8] K.N. Ballantyne, M. Kayser, Additional Y-STRs in Forensics: Why, Which, and When, Forensic Sci Rev 24(1) (2012) 63-78.
[9] M. Kayser, C. Krüger, M. Nagy, G. Geserick, P. de Knijff, L. Roewer, Y-chromosomal DNA-analysis in paternity testing: experiences and recommendations, in: B. Olaisen, B. Brinkmann, P.J. Lincoln (Eds.), Prog. Forensic Genet. Elsevier Amst 7 (1998) 494-496.
[10] A. Adnan, A. Ralf, A. Rakha, N. Kousouri, M. Kayser, Improving empirical evidence on differentiating closely related men with RM Y-STRs: A comprehensive pedigree study from Pakistan, Forensic Sci Int Genet 25 (2016) 45-51.
[11] M. Wang, Z. Wang, G. He, J. Liu, S. Wang, X. Qian, M. Lang, J. Li, M. Xie, C. Li, Y. Hou, Developmental validation of a custom panel including 165 Y-SNPs for Chinese Y-chromosomal haplogroups dissection using the ion S5 XL system, Forensic Sci Int Genet 38 (2019) 70-76.

[12] X. Qian, J. Hou, Z. Wang, Y. Ye, M. Lang, T. Gao, J. Liu, Y. Hou, Next Generation Sequencing Plus (NGS+) with Y-chromosomal Markers for Forensic Pedigree Searches, Sci Rep 7(1) (2017) 11324.
[13] R. Li, C. Zhang, H. Li, R. Wu, H. Li, Z. Tang, C. Zhen, J. Ge, D. Peng, Y. Wang, H. Chen, H. Sun, SNP typing using the HID-Ion AmpliSeq Identity Panel in a southern Chinese population, Int J Legal Med 132(4) (2018) 997-1006.
[14] M.J. Park, H.Y. Lee, N.Y. Kim, E.Y. Lee, W.I. Yang, K.J. Shin, Y-SNP miniplexes for East Asian Y-chromosomal haplogroup determination in degraded DNA, Forensic Sci Int Genet 7(1) (2013) 75-81.
[15] M. Xie, F. Song, J. Li, M. Lang, H. Luo, Z. Wang, J. Wu, C. Li, C. Tian, W. Wang, H. Ma, Z. Song, Y. Fan, Y. Hou, Genetic substructure and forensic characteristics of Chinese Hui populations using 157 Y-SNPs and 27 Y-STRs, Forensic Sci Int Genet 41 (2019) 11-18.
[16] S. Yan, C.C. Wang, H. Li, S.L. Li, L. Jin, An updated tree of Y-chromosome Haplogroup O and revised phylogenetic positions of mutations P164 and PK4, Eur J Hum Genet 19(9) (2011) 1013-5.
[17] P. Hallast, C. Batini, D. Zadik, P. Maisano Delser, J.H. Wetton, E. Arroyo-Pardo, G.L. Cavalleri, P. de Knijff, G. Destro Bisol, B.M. Dupuy, H.A. Eriksen, L.B. Jorde, T.E. King, M.H. Larmuseau, A. Lopez de Munain, A.M. Lopez-Parra, A. Loutradis, J. Milasin, A. Novelletto, H. Pamjav, A. Sajantila, W. Schempp, M. Sears, A. Tolun, C. Tyler-Smith, A. Van Geystelen, S. Watkins, B. Winney, M.A. Jobling, The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades, Mol Biol Evol 32(3) (2015) 661-73.
[18] S. Little, Amplification-refractory mutation system (ARMS) analysis of point mutations, Curr Protoc Hum Genet Chapter 9 (2001) Unit 9.8.
[19] L. Jin, B. Su, Natives or immigrants: modern human origin in east Asia, Nat Rev Genet 1(2) (2000) 126-33.
[20] H. Shi, H. Zhong, Y. Peng, Y.L. Dong, X.B. Qi, F. Zhang, L.F. Liu, S.J. Tan, R.Z. Ma, C.J. Xiao, R.S. Wells, L. Jin, B. Su, Y chromosome evidence of earliest modern human settlement in East Asia and multiple origins of Tibetan and Japanese populations, BMC Biol 6 (2008) 45.
[21] B. Su, J. Xiao, P. Underhill, R. Deka, W. Zhang, J. Akey, W. Huang, D. Shen, D. Lu, J. Luo, J. Chu, J. Tan, P. Shen, R. Davis, L. Cavalli-Sforza, R. Chakraborty, M. Xiong, R. Du, P. Oefner, Z. Chen, L. Jin, Y-Chromosome evidence for a northward migration of modern humans into Eastern Asia during the last Ice Age, Am J Hum Genet 65(6) (1999) 1718-24.
[22] M.H. Larmuseau, A. Van Geystelen, M. Kayser, M. van Oven, R. Decorte, Towards a consensus Y-chromosomal phylogeny and Y-SNP set in forensics in the next-generation sequencing era, Forensic Sci Int Genet 15 (2015) 39-42.
[23] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res 18(5) (2008) 830-8.
[24] H. Zhong, H. Shi, X.B. Qi, Z.Y. Duan, P.P. Tan, L. Jin, B. Su, R.Z. Ma, Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the northern route, Mol Biol Evol 28(1) (2011) 717-27.
[25] S. Sengupta, L.A. Zhivotovsky, R. King, S.Q. Mehdi, C.A. Edmonds, C.E. Chow, A.A. Lin, M. Mitra, S.K. Sil, A. Ramesh, M.V. Usha Rani, C.M. Thakur, L.L. Cavalli-Sforza, P.P. Majumder, P.A. Underhill, Polarity and temporality of high-resolution y-chromosome distributions in India identify both indigenous and exogenous expansions and reveal minor genetic influence of Central Asian pastoralists, Am J Hum Genet 78(2) (2006) 202-21.
[26] M.F. Hammer, T.M. Karafet, H. Park, K. Omoto, S. Harihara, M. Stoneking, S. Horai, Dual origins of the Japanese: common ground for hunter-gatherer and farmer Y chromosomes, J Hum Genet 51(1) (2006) 47-58.
[27] I. Nonaka, K. Minaguchi, N. Takezaki, Y-chromosomal binary haplogroups in the Japanese population and their relationship to 16 Y-STR polymorphisms, Ann Hum Genet 71(Pt 4) (2007) 480-95.

[28] Quality Assurance Standards (QAS) for DNA Databasing Laboratories, (2009) Available at http://www. fbi.gov/about-us/lab/codis/qas_databasalabs.
[29] P.M. Schneider, Scientific standards for studies in forensic genetics, Forensic Sci Int 165(2-3) (2007) 238-43.
[30] SWGDAM, Validation Guidelines for DNA Analysis Methods. <http://swgdam.org/SWGDAM_Validation_Guidelines_APPROVED_Dec_2012.pdf,>, 2010).
[31] S. Berensmeier, Magnetic particles for the separation and purification of nucleic acids, Appl Microbiol Biotechnol 73(3) (2006) 495-504.
[32] A. Auton, L.D. Brooks, R.M. Durbin, E.P. Garrison, H.M. Kang, J.O. Korbel, J.L. Marchini, S. McCarthy, G.A. McVean, G.R. Abecasis, A global reference for human genetic variation, Nature 526(7571) (2015) 68-74.
[33] P.H. Sudmant, T. Rausch, E.J. Gardner, R.E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, K. Ye, G. Jun, M.H. Fritz, M.K. Konkel, A. Malhotra, A.M. Stutz, X. Shi, F.P. Casale, J. Chen, F. Hormozdiari, G. Dayama, K. Chen, M. Malig, M.J.P. Chaisson, K. Walter, S. Meiers, S. Kashin, E. Garrison, A. Auton, H.Y.K. Lam, X.J. Mu, C. Alkan, D. Antaki, T. Bae, E. Cerveira, P. Chines, Z. Chong, L. Clarke, E. Dal, L. Ding, S. Emery, X. Fan, M. Gujral, F. Kahveci, J.M. Kidd, Y. Kong, E.W. Lameijer, S. McCarthy, P. Flicek, R.A. Gibbs, G. Marth, C.E. Mason, A. Menelaou, D.M. Muzny, B.J. Nelson, A. Noor, N.F. Parrish, M. Pendleton, A. Quitadamo, B. Raeder, E.E. Schadt, M. Romanovitch, A. Schlattl, R. Sebra, A.A. Shabalin, A. Untergasser, J.A. Walker, M. Wang, F. Yu, C. Zhang, J. Zhang, X. Zheng-Bradley, W. Zhou, T. Zichner, J. Sebat, M.A. Batzer, S.A. McCarroll, R.E. Mills, M.B. Gerstein, A. Bashir, O. Stegle, S.E. Devine, C. Lee, E.E. Eichler, J.O. Korbel, An integrated map of structural variation in 2,504 human genomes, Nature 526(7571) (2015) 75-81.
[34] M. Lek, K.J. Karczewski, E.V. Minikel, K.E. Samocha, E. Banks, T. Fennell, A.H. O'Donnell-Luria, J.S. Ware, A.J. Hill, B.B. Cummings, T. Tukiainen, D.P. Birnbaum, J.A. Kosmicki, L.E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D.N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M.I. Kurki, A.L. Moonshine, P. Natarajan, L. Orozco, G.M. Peloso, R. Poplin, M.A. Rivas, V. Ruano-Rubio, S.A. Rose, D.M. Ruderfer, K. Shakir, P.D. Stenson, C. Stevens, B.P. Thomas, G. Tiao, M.T. Tusie-Luna, B. Weisburd, H.H. Won, D. Yu, D.M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J.C. Florez, S.B. Gabriel, G. Getz, S.J. Glatt, C.M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M.I. McCarthy, D. McGovern, R. McPherson, B.M. Neale, A. Palotie, S.M. Purcell, D. Saleheen, J.M. Scharf, P. Sklar, P.F. Sullivan, J. Tuomilehto, M.T. Tsuang, H.C. Watkins, J.G. Wilson, M.J. Daly, D.G. MacArthur, C. Exome Aggregation, Analysis of protein-coding genetic variation in 60,706 humans, Nature 536(7616) (2016) 285-91.
[35] C.C. Wang, L.X. Wang, R. Shrestha, M. Zhang, X.Y. Huang, K. Hu, L. Jin, H. Li, Genetic structure of Qiangic populations residing in the western Sichuan corridor, PLoS One 9(8) (2014) e103772.
[36] T. Gayden, A.M. Cadenas, M. Regueiro, N.B. Singh, L.A. Zhivotovsky, P.A. Underhill, L.L. Cavalli-Sforza, R.J. Herrera, The Himalayas as a directional barrier to gene flow, Am J Hum Genet 80(5) (2007) 884-94.
[37] K. Bender, M.J. Farfan, P.M. Schneider, Preparation of degraded human DNA under controlled conditions, Forensic Sci Int 139(2-3) (2004) 135-40.
[38] T. Geng, R.A. Mathies, Minimizing inhibition of PCR-STR typing using digital agarose droplet microfluidics, Forensic Sci Int Genet 14 (2015) 203-9.
[39] K.L. Opel, D. Chung, B.R. McCord, A study of PCR inhibition mechanisms using real time PCR, J Forensic Sci 55(1) (2010) 25-33.
[40] J. Purps, M. Geppert, M. Nagy, L. Roewer, Validation of a combined autosomal/Y-chromosomal STR approach for analyzing typical biological stains in sexual-assault cases, Forensic Sci Int Genet 19 (2015) 238-242.

[41] P. Villaescusa, L. Palencia-Madrid, M.A. Campaner, J. Jauregui-Rada, M. Guerra-Rodriguez, A.M. Rocandio, M.M. de Pancorbo, Effective resolution of the Y chromosome sublineages of the Iberian haplogroup R1b-DF27 with forensic purposes, Int J Legal Med 133(1) (2019) 17-23.
[42] M. Nothnagel, G. Fan, F. Guo, Y. He, Y. Hou, S. Hu, J. Huang, X. Jiang, W. Kim, K. Kim, C. Li, H. Li, L. Li, S. Li, Z. Li, W. Liang, C. Liu, D. Lu, H. Luo, S. Nie, M. Shi, H. Sun, J. Tang, L. Wang, C.C. Wang, D. Wang, S.Q. Wen, H. Wu, W. Wu, J. Xing, J. Yan, S. Yan, H. Yao, Y. Ye, L. Yun, Z. Zeng, L. Zha, S. Zhang, X. Zheng, S. Willuweit, L. Roewer, Revisiting the male genetic landscape of China: a multi-center study of almost 38,000 Y-STR haplotypes, Human genetics 136(5) (2017) 485-497.
[43] L. Wang, W. He, J. Mao, H. Wang, B. Jin, H.B. Luo, W.B. Liang, L. Zhang, Development of a SNP-STRs multiplex for forensic identification, Forensic Science International: Genetics Supplement Series 5 (2015) e598-e600.
[44] J.H. Wetton, C.S. Tsang, C.A. Roney, A.C. Spriggs, An extremely sensitive species-specific ARMs PCR test for the presence of tiger bone DNA, Forensic Sci Int 140(1) (2004) 139-45.
[45] T. Wei, F. Liao, Y. Wang, C. Pan, C. Xiao, D. Huang, A novel multiplex assay of SNP-STR markers for forensic purpose, PLoS One 13(7) (2018) e0200700.
[46] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Human genetics 136(5) (2017) 621-635.
[47] C.C. Wang, L.X. Wang, R. Shrestha, S. Wen, M. Zhang, X. Tong, L. Jin, H. Li, Convergence of Y Chromosome STR Haplotypes from Different SNP Haplogroups Compromises Accuracy of Haplogroup Prediction, J Genet Genomics 42(7) (2015) 403-7.