BMC Genomics

RESEARCH

Open Access

Exploring Y-chromosomal STRs and SNPs for forensic and genetic insights in the Jiangsu Han population

Min Li[1†], Han Zhang[2†], Ruiyang Tao[3], Anqi Chen[2], Peng Zhou[1], Chunjiang Yu[1], Yingnan Bian[3], Suhua Zhang[2], Chen Fang[1*] and Chengtao Li[2,3*]

Abstract

Y-chromosome short tandem repeats (Y-STRs) and single nucleotide polymorphisms (Y-SNPs) are valuable genetic markers used for individual identification, forensic applications, and the study of paternal lineage history. This study analyzed the genetic polymorphism and paternal genetic structure of the Han population in Jiangsu Province by examining 374 unrelated male individuals using 29 Y-STRs and 183 Y-SNPs. Forensic parameters were calculated, and the discriminatory power of five Y-STR systems (MHT, EXT, PPY12, Yfiler, and Y29) was compared. Genetic structure was assessed in the context of the Jiangsu Han and other Chinese populations. Results showed that the Y29 system had the highest discriminatory capacity, identifying 374 unique haplotypes with HD and DC values of 1. Seven major haplogroups (C, D, J, K, O, Q, R) and 83 terminal haplogroups were identified, with haplogroup O being the most predominant (approximately 85%). Subdivision of haplogroup O revealed that the Jiangsu Han population exhibits genetic characteristics of both Southern and Northern Han groups. Population genetic analyses further confirmed that the Jiangsu Han clustered closely with Southern Han populations. Genetic admixture results revealed that the Jiangsu Han population derives 89% of their ancestry from Southern Han populations, while retaining 11% Northern Han contributions. The study provides valuable insights into the genetic structure of the Jiangsu Han population, with significant implications for forensic genetics, anthropological research, and broader population genetics studies.

Keywords Han populations, Y-SNPs, Y-STRs, Genetic substructure, Forensic characteristics

Introduction

The human Y chromosome spans approximately 60 megabases (Mb), with its terminal regions comprising pseudoautosomal regions (PARs) that account for ~5% of its total bases [1]. These PARs facilitate homologous recombination with the X chromosome during meiosis. The remaining 95% constitutes the non-recombining region (NRY), which lacks pairing capability with the X chromosome [2]. Due to the absence of recombination, the NRY is inherited as a haplotype from father to son, theoretically preserving identical genetic information within the same paternal lineage in the absence of

†Min Li and Han Zhang contributed equally to this work.

*Correspondence:
Chen Fang
chenfangyulix@126.com
Chengtao Li
lichengtao@fudan.edu.cn
[1]Department of Forensic Medicine, School of Clinical and Basic Medical Sciences, Shandong First Medical University& Shandong Academy of Medical Sciences, Jinan 250117, China
[2]Institute of Forensic Science, Fudan University, Shanghai 200032, China
[3]Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Key Laboratory of Forensic Science, Academy of Forensic Sciences, Ministry of Justice, Shanghai 200063, China

mutations [3]. These characteristics make the Y chromosome valuable in anthropology, population genetics, molecular archaeology, and human evolution studies [4–6]. To date, over 65,000 genetic markers have been identified on the human Y chromosome [7]. These markers are broadly categorized into two classes: biallelic variants, such as single nucleotide polymorphisms (SNPs), and multiallelic loci, exemplified by short tandem repeats (STRs) [8–9]. Two primary genetic markers on the Y chromosome, Y-STRs and Y-SNPs, are widely used in forensic paternity testing, personal identification, and population structure analysis.

Y-STRs are highly polymorphic and can be readily detected using standard genetic analyzers, making them a preferred tool for forensic DNA analysis. They are instrumental in solving crimes, accurately determining paternity, and identifying kinship, even in the absence of maternal or other familial samples [10]. However, their relatively high mutation rates (3.95×10^{-4}-1.40×10^{-2} per locus per generation, https://yhrd.org/) can compl icate studies involving distant relatives or ancient DNA. In contrast, Y-SNPs have much lower mutation rates ($\sim3\times10^{-8}$ mutations/generation), offering stability over generations [11–12]. This stability underpins their role in constructing the Y-chromosome haplogroup phylogenetic tree, which classifies global male populations into hierarchical clades.

Although less polymorphic than Y-STRs, Y-SNPs complement them by providing additional insights, enhancing the accuracy and reliability of forensic analyses. Together, Y-STRs and Y-SNPs form a robust dual-marker system, enabling precision paternity testing, kinship analysis, and a deeper understanding of population genetic structures [13–14].

The Han ethnic group, the largest in China, exhibits notable genetic diversity influenced by geography, historical migrations, and cultural exchanges. While broad genetic patterns among Southern and Northern Han populations are relatively well-studied [15], subgroups such as those in Jiangsu Province remain underexplored. Jiangsu, a critical province in eastern China, serves as a cultural and geographical bridge between south and north regions. As a hub for population and cultural exchanges, studying the Han population in Jiangsu offers an opportunity to uncover the genetic characteristics of this transitional zone and gain deeper insights into the genetic diversity and distribution within the Han ethnicity.

Despite the importance of Jiangsu's Han population, previous studies have focused primarily on Y-STR markers, often analyzing an insufficient number of loci. While these studies have provided data on genetic polymorphisms and intergroup variations, they have not adequately addressed the distribution of Y-SNP haplogroups or the paternal genetic structure of the population. A more comprehensive approach that combines Y-STR and high-resolution Y-SNP data is essential to fill these gaps.

This study employs 29 Y-STR loci and a high-resolution haplogroup system comprising 183 Y-SNP markers to analyze genetic polymorphisms in the Han population of Jiangsu Province. The objectives are to elucidate the paternal genetic structure of this population, expand the genetic database of eastern China, and provide accurate data to advance forensic and human genetic research.

Materials and methods

Sample collection and DNA extraction

Blood samples were collected with the approval of the Ethics Committee of Academy of Forensic Science, Ministry of Justice, China (2014-3-3-5). Male participants were eligible if their ancestors had resided in Jiangsu Province for at least three generations. Peripheral blood was collected after receiving written informed consent. Finally, a total of 374 unrelated male individuals were collected. The geographic location of the Jiangsu Province is detailed in Supplementary Fig. S1. Genomic DNA was extracted using QIAamp DNA blood Mini Kit (Qiagen, Germany), following the manufacturer's instructions, and quantified with NanoDrop 2000c (Thermo Fisher Scientific). DNA samples were stored at -20 °C until PCR amplification.

Y-STR typing via capillary electrophoresis

The Goldeneye 29Y kit (Peoplespot Technology Ltd., Beijing, China), referred to as Y29, was used to generate Y-chromosome haplotypes for 29 Y-STR loci. This kit includes all loci from the Yfiler kit, supplemented with the four rapidly mutating Y-STR loci and eight loci with low-medium mutation rates. Specifically, the 29 loci analyzed were: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385, DYS437, DYS438, DYS439, DYS448, DYS456, DYS458, DYS635, Y-GATA-H4, DYS526a, DYS526b, DYS570, DYS576, DYS626, DYS388, DYS460, DYS481, DYS593, DYS596, DYS643, DYS645. PCR was performed on a GeneAmp 9700 thermal cycler (Thermo Fisher Scientific) in accordance with the manufacturer's recommendations. Separation and detection of PCR products were carried out with the Applied Biosystems 3130XL Genetic Analyser (Applied Biosystems, CA, USA). STR profiles were analysed with the ORG-500 Size Standard and allelic ladder (Peoplespot Technology Ltd., Beijing, China) and the set of bins and panels provided by the manufacturer using GeneMapper ID v3.2 software (Applied Biosystems, CA, USA).

Y-SNP typing via MALDI-TOF MS platform

Y-SNPs typing employed a self-designed panel of 183 markers, developed based on DNA mass spectrometry

(MALDI-TOF MS). The panel included 36 Y-SNPs from major world-wide haplogroups (A to R) and 147 additional markers specific to haplogroup O. Marker nomenclature followed the International Society of Genetic Genealogy (ISOGG2019) guidelines. Detailed information regarding the primers of 183 Y-SNP markers, the names and corresponding haplogroups of the 183 Y-SNPs, the chromosomal positions relative to human reference sequence hg19, dbSNP rs numbers, and mutation information is shown in the Supporting Information of our previous study [16].

Data analysis

Y-STR analytic methods

Allele and haplotype frequencies were estimated by the direct counting method. Haplotype diversity (HD) and gene diversity (GD) were calculated according to the formula of Nei and Tajima [17]: $HD/GD = n\,(1 - \sum Pi^2)/\,(n-1)$, where n is the total number of samples and Pi denotes the frequency of the i-th allele/haplotype. Match probability (MP) was computed as the sum of squared haplotype frequencies ($MP = \sum Pi^2$). Discrimination capacity (DC) was calculated as the ratio between the number of observed haplotypes and the number of total samples. The percentage of unique haplotype (PUH) was calculated by dividing the number of unique haplotypes by the total sample size.

The Y29 system encompasses all the STR loci present in the European minimal haplotype (MHT), extended haplotype (EXT), PowerPlex Y12 (PPY12), and Yfiler kits. In light of this comprehensive coverage, we partitioned the haplotypes generated by the Y29 system into five distinct haplotype systems, namely Y29 itself, MHT, EXT, PPY12, and Yfiler. Subsequently, we computed the relevant forensic efficiency parameters, specifically haplotype diversity (HD), discrimination capacity (DC), and the proportion of unique haplotypes (PUH), for each of these five systems separately.

Y-STR haplotypes from 24 published Chinese populations (comprising both Han Chinese and diverse ethnic minority groups) alongside 9 geographically distinct outgroups were curated as reference datasets [18–48]. Detailed group information was listed in Table S1. To characterize population genetic structure, pairwise Fst genetic distances were computed by Arlequin (version 3.5.2.2) software [49] and visualized via multidimensional scaling (MDS) analysis using IBM SPSS Statistics 23 software (IBM Corp., Armonk, NY, USA). Additionally, a neighbor-joining (NJ) phylogenetic tree [50] was inferred from these Fst matrices using MEGA X software (version 10.0.5), which incorporates bootstrapping (1,000 replicates) to assess branch robustness [51]. Samples carrying duplicated alleles (and triplicated alleles), null alleles, and microvariants were excluded during the Fst analysis. P-values were calculated at a significance level of 0.05 using 1,000 permutations.

Y-SNP analytic methods

Haplogroup frequencies were estimated by the direct counting method. The haplogroup diversity and DC value were calculated using the same method as Y-STR. A total of 36 representative reference populations were selected based on haplogroup frequency distributions derived from supplementary datasets encompassing both ancient and modern Chinese populations [15, 52–58]. Since the haplogroup classifications of each reference population were different, we pooled the haplogroup results of these reference populations together using the main haplogroups and calculated the haplogroup frequency. The reference populations were listed in Supplementary Table S2. Based on the results of haplogroup frequencies, a Reynold's distance matrix was obtained with Arlequin (version 3.5.2.2) software [49], and used as input for MDS analysis. MDS plots were constructed using IBM SPSS Statistics 23 software (IBM Corp., Armonk, NY, USA).

Y- STR and Y-SNP joint analytic methods

Median-joining (MJ) networks of Y-STR haplotypes (comprising DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393) for O-clade sub-haplogroups in the Jiangsu Han, Northern Han, and Southern Han populations [15] were constructed using NETWORK 10.2 software (Fluxus Technology, Suffolk, UK) [59]. Graphical representation of these networks was subsequently generated with Network Publisher software (Fluxus Technology, Suffolk, UK). Locus-specific weights were assigned based on empirical mutation rate estimates [57, 60].

An analysis of molecular variance (AMOVA) was conducted to assess the population structure between the Jiangsu Han, Northern Han, and Southern Han populations, employing the Arlequin software (version 3.5.2.2) [49]. Both Y-STRs and Y-SNPs were utilized in this analysis. The Y-STR dataset consisted of 17 loci, consistent with those included in the Yfiler kit. The Y-STR and Y-SNP data for the Northern and Southern Han populations were sourced from the literature [15]. The Southern Han group encompassed six populations, namely the Han populations from Guangdong, Guangxi, Fujian, Zhejiang, Jiangxi, and Hunan. The Northern Han group was composed of five populations, including the Han populations from Beijing, Shandong, Heilongjiang, Henan, and Shanxi.

Admixture analysis between the Jiangsu Han, Northern Han, and Southern Han populations was performed using the coalescent-based approach implemented in Admix 2.0 software [61–62]. This method infers genetic contributions from hypothetical parental source populations by

modeling STR haplotype frequencies as alleles at a single locus. Specifically, the Northern and Southern Han populations were designated as the primary parental groups for admixture estimation. The analysis incorporated seven Y-STR loci (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, and DYS393), selected based on their mutation rate parameters and forensic relevance [57].

Results

Y-STR haplotype analysis

In this study, 374 unique haplotypes were identified from 374 male samples (Supplementary Table S3), with a total of 206 alleles observed. Allele frequency distributions and gene diversity values for Y-STR loci were listed in Supplementary Table S4. The number of alleles at single-copy Y-STRs varied from 4 (DYS645, DYS393, DYS437 and DYS593) to 12 (DYS458). For the multi-copy DYS385a/b, 17 alleles and 58 haplotypes were identified.

Additionally, one null allele was observed at DYS438 and six intermediate alleles [DYS460 (7.2), DYS576 (18.3), DYS645 (7.1), DYS448 (19.2) and DYS458 (18.2/19.2)] were detected in seven samples. Of these, microvariants (except DYS645) were observed in YHRD. DYS645 (7.1) was reported for the first time.

The GD values exceeded 0.5 for 25 of the 29 loci, with the lowest being 0.0628 (DYS645) and the highest 0.9598 (DYS385). Key forensic parameters showed robust discriminative power: HD and DC values were both 1, while the MP value was 0.0027 (Supplementary Table S5).

Further analysis of forensic parameters across five Y-STR typing systems-MHT, EXT, PPY12, Yfiler, and Y29-highlighted significant differences in discriminative power as the number of Y-STR loci increased (Table 1). In the MHT system (9 Y-STRs), the DC was only 0.8021 and HD was 0.9983, including 300 different haplotypes. While in the Yfiler system (17 Y-STRs), the DC was 0.9786 and HD was 0.9999, including 366 different haplotypes. After the addition of 12 Y-STR markers, the DC and HD values were significantly increased to 1 in the Y29 system with 374 different haplotypes. The value of the PUH ranged from 66.84% (MHT) to 100% (Y29).

Table 1 The number of haplotypes and the forensic parameters estimated for different combinations of Y-STR loci ($N = 374$)

Time(s) a haplotype was observed	MHT	EXT	PPY12	Yfiler	Y29
1	250	288	290	358	374
2	37	26	28	8	-
3	9	7	5	-	-
4	1	-	-	-	-
6	2	1	1	-	-
7	1	1	1	-	-
Sample	374	374	374	374	374
No.	300	323	325	366	374
PUH	0.6684	0.7701	0.7754	0.9572	1
DC	0.8021	0.8636	0.8690	0.9786	1
HD	0.9983	0.9988	0.9989	0.9999	1

No.: Number of Haplotypes. PUH: Percentage of unique haplotypes. DC: discrimination capacity. HD: haplotype diversity. MHT: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385 a/b(9 Y-STR loci). EXT: MHT + DYS438 and DYS439(11 Y-STR loci). PPY12: EXT + DYS437 (12 Y-STR loci). Yfiler: PPY12 + DYS456, DYS458, DYS635, DYS448 and YGATA-H4(17 Y-STR loci). Goldeneye 29Y: Yfiler + DYS388, DYS460, DYS481, DYS526a, DYS526b, DYS570, DYS576, DYS593, DYS596, DYS626, DYS643 and DYS645(29 Y-STR loci)

Comparison of Jiangsu Han with other populations based on Y-STR

To better reveal the genetic relationship between Jiangsu Han and 24 other Chinese populations, pairwise genetic distance (Fst) and p values based on 17 Y-STRs (Yfiler kit) among 25 Chinese populations were summarized in Supplementary Table S6. The closest genetic distance was observed between the Jiangsu Han and Changzhou Han (Fst = 0.0033, $p = 0.2295$), while the furthest genetic distance was found with the Xinjiang Kazak (Fst = 0.550, $p = 0.4337$).

These genetic distances were visualized using a multidimensional scaling (MDS) plot (Fig. 1a) and a neighbor-joining (NJ) phylogenetic tree (Fig. 1b). The MDS plot revealed that the Jiangsu Han population showed no significant genetic differentiation from other Han subgroups across China but displayed clear distinctions when compared to certain ethnic minority groups. The Jiangsu Han clustered closely with the Southern Han groups in the central-right region of the MDS plot and maintained proximity to the Northern Han populations, indicating a transitional genetic structure. Ethnic minority groups were more dispersed around these clusters.

Similar patterns were reflected in the NJ phylogenetic tree, where the Jiangsu Han and other Han populations formed distinct branches separated from ethnic minorities, consistent with their geographic and genetic backgrounds. However, notable exceptions included the Hainan Li and Liangshan Yi groups, which appeared closer to the Han populations. This could be attributed to historical intermarriage between these groups and the Han population in recent decades, influencing their genetic composition.

Population comparisons of the Jiangsu Han with the other 9 populations from the outgroups were also conducted based on 17 Y-STRs (Supplementary Figure S2). The results revealed that Jiangsu Han had a close relationship with East Asian and Southeast Asian populations.

Y-SNP haplogroup distribution

A total of seven major haplogroups O (85.03%), C (9.09%), K (4.01%), D (0.53%), Q (0.8%), J (0.27%), and R (0.27%) were observed in the Jiangsu Han (Fig. 2). Within these, 83 unique terminal haplogroups were detected,

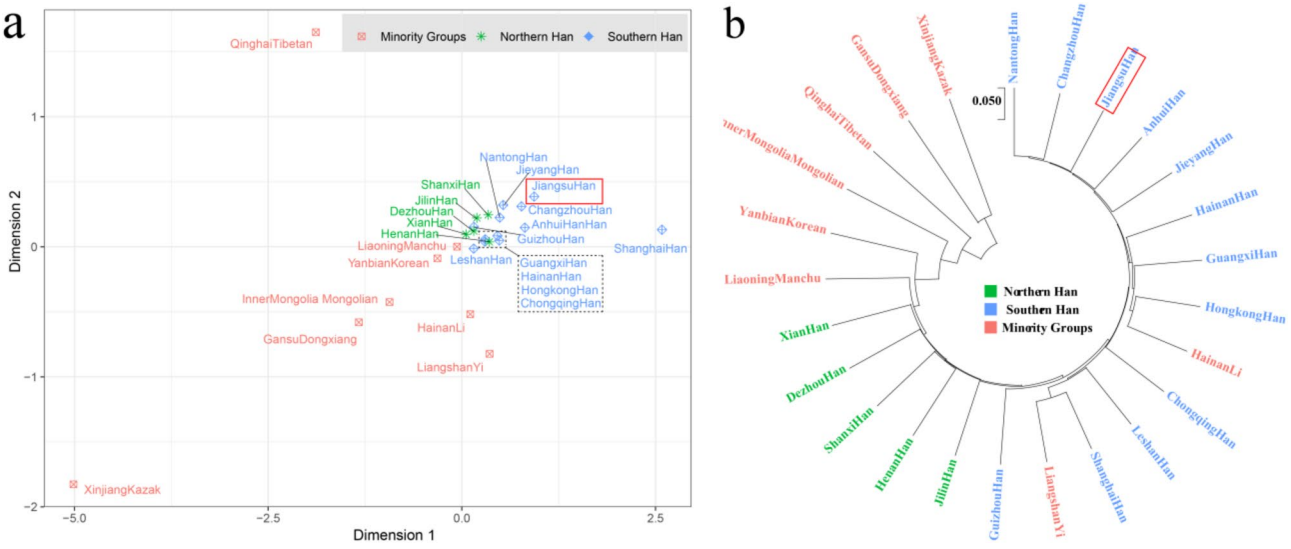

Fig. 1 Jiangsu Han and other reference populations based on Fst value of 17 Y-STRs. **(a)** MDS plot. **(b)** N-J phylogenetic tree. Jiangsu Han marked with a red box

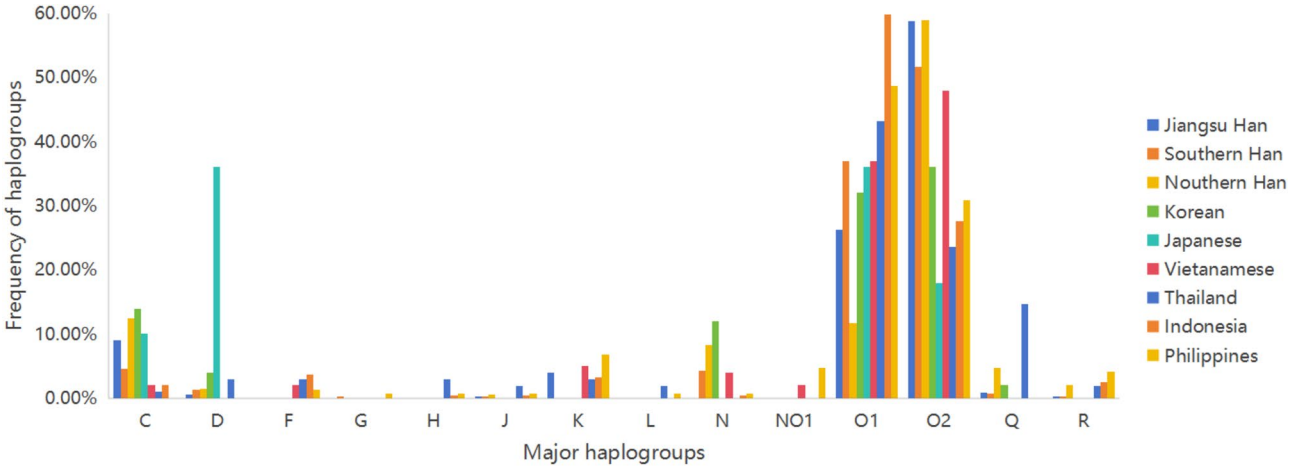

Fig. 2 The major haplogroup frequency distribution in Jiangsu Han and East/Southeast Asian populations

and their frequencies were listed in Supplementary Table S7. Forensic parameters revealed a HD of 0.9480 and a DC value of 0.2219 (Table S5).

To investigate the frequency distributions of haplogroups across Chinese populations and other East Asian and Southeast Asian populations, comparative analyses were conducted. Haplogroup frequency data for Southern Han, Northern Han [15], other East Asian nations (Japan and Korea) [56], and Southeast Asian populations (Thailand, Vietnam, Indonesia, and the Philippines) [57]were systematically retrieved from the literature. The outcomes of these inter-population comparisons are visually depicted in Fig. 2, enabling a clear assessment of haplogroup distribution patterns among the studied groups.

A comprehensive analysis of nine East Asian and Southeast Asian populations revealed the presence of a total of fourteen major haplogroups (C, D, F, G, H, J, K, L, N, NO1, O1, O2, Q, and R). Evidently, the O haplogroup emerged as the predominant lineage across all populations under investigation. The O haplogroup can be further classified into O1 and O2 sub-haplogroups. In Chinese Han populations, Korea and Vietnam, the frequency of the O2 sub-haplogroup exceeded that of the O1 sub-haplogroup. Conversely, in Japan and three Southeast Asian countries (Thailand, Indonesia, and the Philippines), the O1 sub-haplogroup was more prevalent than the O2 sub-haplogroup.

In the Jiangsu Han population, the frequency of the O1 haplogroup was 26.20%, which is consistent with the characteristics of Southern Han populations, where the O1 haplogroup typically accounts for over 25%. In contrast, in Northern Han populations, the frequency of the O1 haplogroup usually remains below 20%. The

frequency of the O2 haplogroup in the Jiangsu Han population was 58.82%, and it was further subdivided into four subgroups: O2a1 (17.11%), O2a2a (10.16%), O2a2b (30.75%), and O2b1a (0.8%).

In East Asian populations, excluding Japan, the C haplogroup was the second-most frequent haplogroup. In Japan, the D haplogroup held this position. In Southeast Asian countries, haplogroups such as F, K, and Q were among the second-largest haplogroups, yet their distribution was relatively scarce in East Asian populations. These findings contribute to an in-depth understanding of the genetic structure and diversity within East Asian and Southeast Asian populations, providing valuable insights into the complex genetic history and population dynamics of these regions.

Comparison of Jiangsu Han with other populations based on Y-SNP

To further elucidate the genetic relationships among diverse populations, an MDS plot was constructed using haplogroup frequencies from the Jiangsu Han and 36 reference populations. As depicted in Fig. 3, Chinese Han populations (excluding the Guangxi Han) clustered closely on the right side of the MDS plot. This concentrated distribution underscores substantial genetic similarities among Han groups, likely

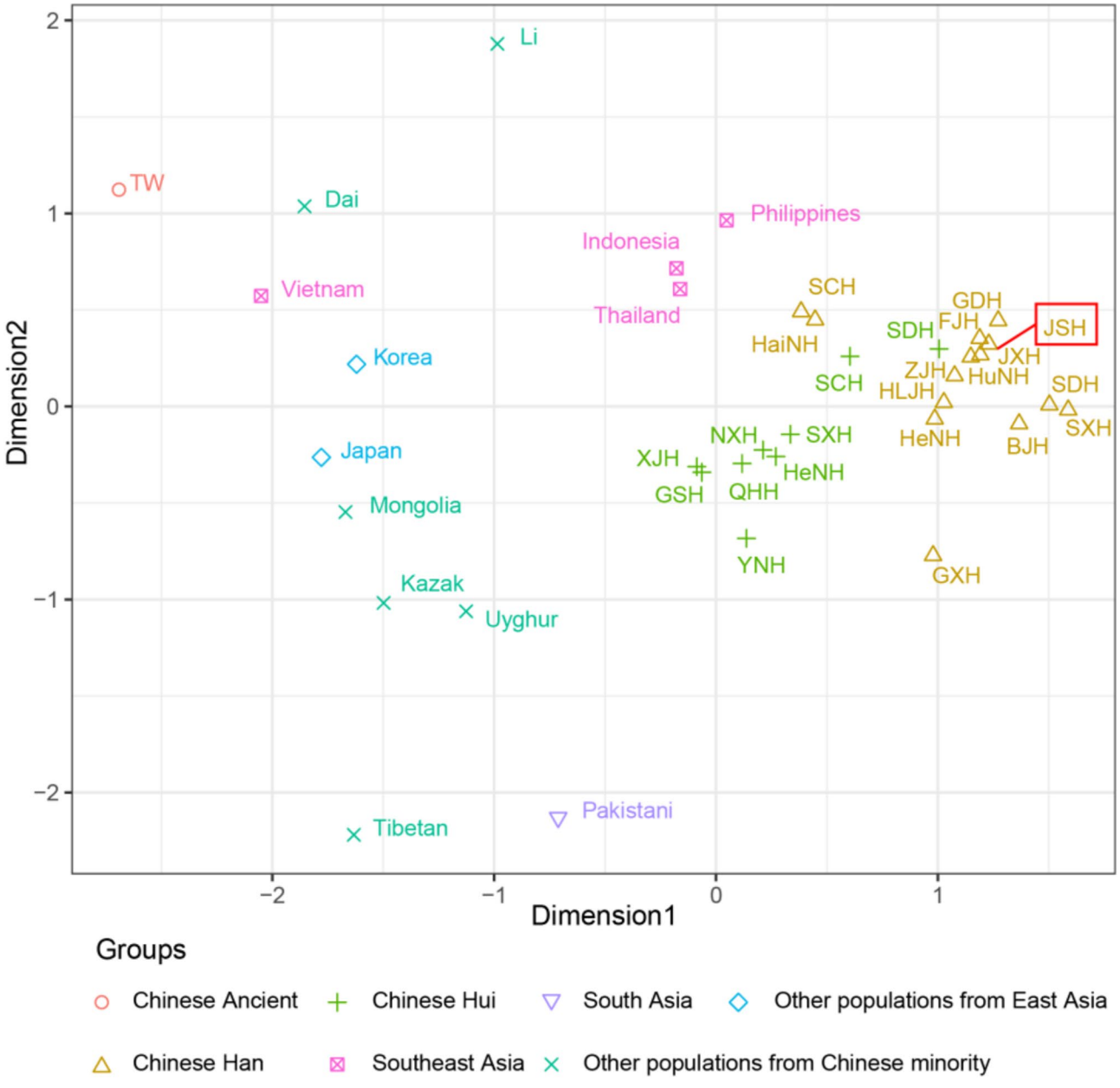

Fig. 3 MDS based on the frequencies of Y-SNP haplogroups. Jiangsu Han marked with a red box

stemming from shared ancestral origins or recent genetic exchanges. Furthermore, the Jiangsu Han population exhibited tight clustering with Southern Han groups (e.g., Guangdong, Fujian, Zhejiang, Hunan, and Jiangxi populations), exemplifying genetic proximity. In contrast, it demonstrated distinct separation from Chinese Hui populations (e.g., Yunnan Hui, Gansu Hui) and Chinese ethnic minority groups (e.g., Li, Dai). Additionally, the Jiangsu Han population was genetically distanced from Southeast Asian populations (Vietnam, Thailand), South Asian populations (Pakistani), and East Asian populations (Korea, Japan), highlighting its unique genetic positioning within the Han ethnic framework. We collected a dataset of the ancient Taiwan population (TW), dating from approximately 6000 BCE to 1000 CE. As visualized in the MDS plot, the ancient Taiwan population was distinctly positioned apart from most modern populations.

Network analysis

To investigate the genetic structure among Jiangsu Han, Northern and Southern Han in detail, we carried out a series of analyses. We used the median joining (MJ) method to create the STR haplotype network within the O sub-haplogroup. The median-joining network based on 7 Y-STRs for haplogroups O1, O2a1, O2a2a, and O2a2b revealed distinct clustering patterns among Jiangsu Han, Southern Han, and Northern Han populations (Fig. 4). As depicted in Fig. 4, for Jiangsu Han, within each haplogroup network, Jiangsu Han individuals were interspersed with Southern Han and Northern Han, indicating shared haplotype connections. In the O1 network, Jiangsu Han clustered centrally, suggesting a core contribution to the haplotype diversity. Similarly, in the O2a1, O2a2a, and O2a2b networks, Jiangsu Han nodes were distributed across various branches, with some forming prominent subclusters. The network topology showed a central star-like structure dominated by O2a2a, which suggested

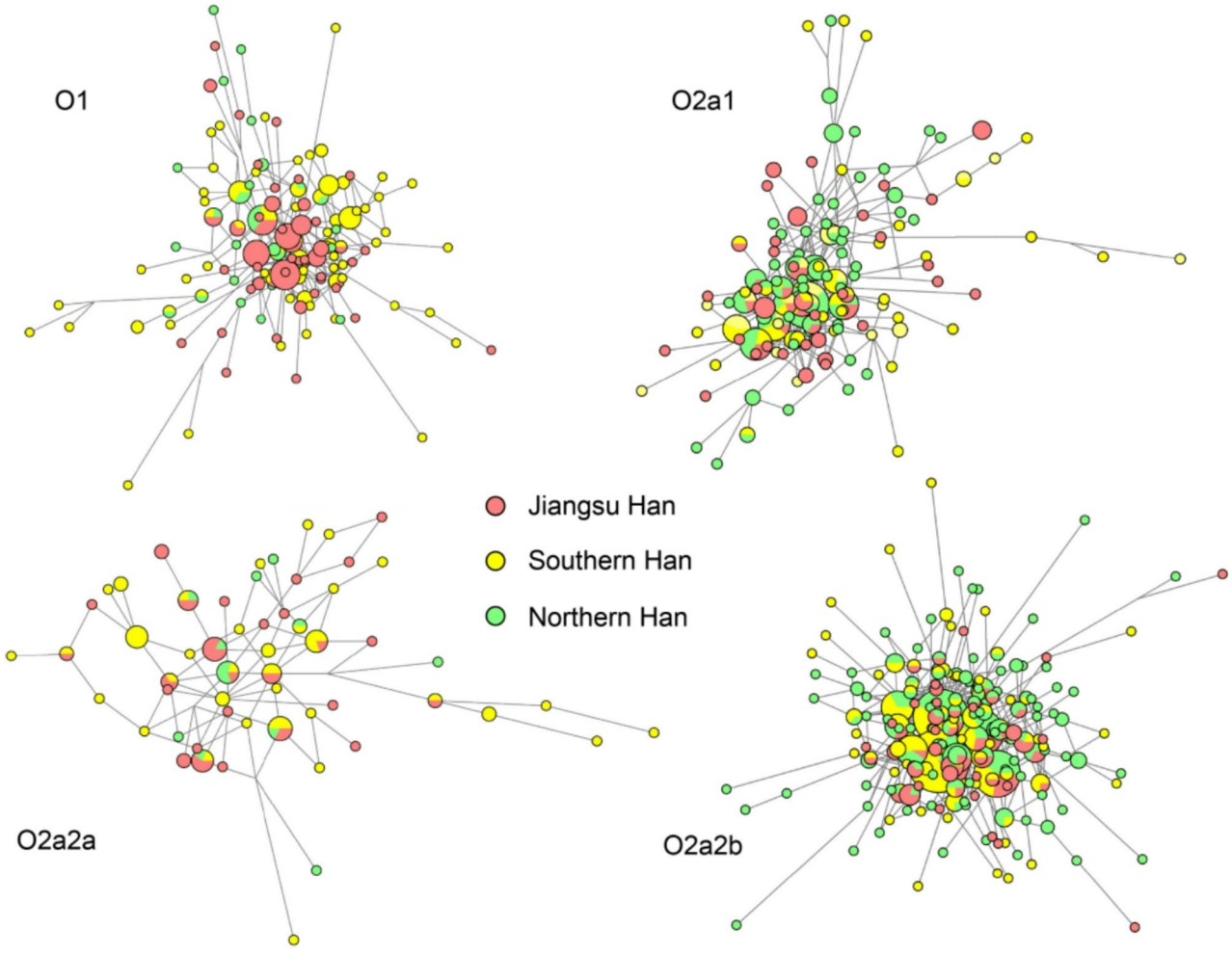

Fig. 4 Median-joining network based on 7 Y-STRs

Table 2 Analysis of molecular variance (AMOVA) using Y-SNPs and Y-STRs

	Classification criterion	Total variation (%)	Fixation indices	(*p* value)
Y-SNPs	Among groups	3.08	F_{CT}=0.03082	0.02737
	Among populations within groups	3.08	F_{SC}=0.03177	0.00000
	Within population	93.84	F_{ST}=0.06161	0.00000
Y-STRs	Among groups	-0.12	F_{CT}= -0.00119	-0.67644
	Among populations within groups	1.22	F_{SC}=0.01216	0.00000
	Within populations	98.90	F_{ST}=0.01099	0.00000

Geographic Partitioning (three groups). (1) Jiangsu Han. (2) Southern Han. (3) Northern Han. F_{SC}=variance among populations within groups; F_{ST}=variance within population; F_{CT}=variance among groups. AMOVA test for Y-STR was based on 17 STR loci

Table 3 Gene contribution estimates using 7 Y-STR

Putative parental groups	mγ	SD
Northern Han	0.11	0.19
Southern Han	0.89	0.19

mγ:admixture proportions; SD: standard error

a founder effect or recent population expansion. Subclusters of O1 and O2a2b displayed closer genetic proximity to Southern Han populations, while O2a1 exhibited partial overlap with Northern Han haplotypes. Notably, Jiangsu Han occupied an intermediate position between Southern and Northern Han in the network, indicating bidirectional genetic influences.

AMOVA and admixture

To estimate the ancestry proportion, AMOVA was conducted between the Jiangsu Han population and the groupings of Northern Han and Southern Han Chinese populations, as shown in Table 2. For Y-SNP data, as expected, the majority of variance (93.84%, $p < 0.001$) originated from within-population differences. Moderate variance (3.08%, $p < 0.001$) was observed within groups, while significant differences (3.08%, $p = 0.027$) were detected among groups. Regarding Y-STR data, the vast majority of variance (98.90%, $p < 0.001$) resided within populations. Slight variance (1.22%, $p < 0.001$) occurred within groups, and no significant differences were identified among groups. For both Y-SNPs and Y-STRs, intraspecific variation within populations served as the dominant contributor to genetic variance. Notably, Y-SNPs exhibit the capability to unravel differentiation patterns at both the intergroup level and the intragroup population level, thereby highlighting their superior sensitivity in dissecting interpopulation genetic structures.

Using Admix 2.0, the contributions of the two hypothetical parental groups "Southern Han" and "Northern Han" to the hybrid population (Jiangsu Han) were estimated, with results presented in Table 3. The findings reveal that Jiangsu Han derived approximately 11% of its ancestry from Northern Han and 89% from Southern Han, thus supporting the hypothesis that Jiangsu Han predominantly originated from Southern Han.

Discussion

In forensic genetics practice, Y-STRs and Y-SNPs serve as crucial genetic marker systems, demonstrating unique research value and application potential. Y-SNPs, characterized by extremely low mutation rates, effectively elucidate the paternal genetic history of populations by identifying haplogroup distribution patterns in specific geographic regions, providing molecular evidence for reconstructing ancient population migration routes. Conversely, Y-STRs exhibit significant advantages in studying genetic relationships among closely related populations and short-term evolutionary events due to their high mutation rates. This study systematically analyzed the genetic polymorphism characteristics and paternal genetic structure of the Jiangsu Han population through integration of 29 Y-STR loci and 183 Y-SNP markers. Among the five Y-STR systems, the Y29 system demonstrated the highest discriminatory power (DC = 1), with its exceptional haplotype resolution (HD = 1) proving ideal for forensic applications in the Jiangsu Han population.

The global Y-chromosomal phylogenetic tree reveals that East Asian populations primarily consist of four major haplogroups: C-M130, D-M174, N-M231, and O-M175 [63]. Notably, the O-M175-derived O1-F265 and O2-M122 haplogroup (formerly O3-M122) show high-frequency distribution in East Asian populations. The two principal subgroups of O1 haplogroup (O1a and O1b) predominantly distribute in Southern East Asia/Southeast Asia. Our findings align with previous studies [64], demonstrating higher O1 frequencies in Southeast Asian populations compared to O2, with O1 being more prevalent in Southern Chinese Han populations than their Northern counterparts.

The O2 haplogroup is dominant in East Asian populations and accounts for 58.82% of the Jiangsu Han population. O2 in Jiangsu Han is further classified into four main sub-haplogroups: O2a1, O2a2a, O2a2b, and O2b1a. According to the research records of Yan et al., O2a1 exhibits a relatively uniform distribution pattern in the Han Chinese population across the geographical regions of the east, north, and south [64]. O2a2b is also widely distributed in China, and its proportion is

much higher among the Han population in the Northern region than that in the Southern region (Northern region: 34.11% vs. Southern region: 21.54%) [65].

The C haplogroup ranks as the second most frequent in East Asian populations (excluding Japan) in our study and primarily distributes in Northern Mongolia, Korea, and Northeastern China. The D haplogroup shows a high frequency among the Japanese population. Additionally, it has a notable prevalence in the populations of Tibet and the Andaman Islands. Specifically, haplogroup D1a1 is extensively spread among the Tibetan, Sichuan Yi, Yunnan Yi (Tangut-Qiang), and Lolo populations, as noted in [66]. In contrast, D1a2 is most commonly found in Tibetans, as reported in [67]. The distribution of the N haplogroup in our study primarily occurs in Korea (12%) and Northern Chinese Han (8.31%), with previous research indicating that East Asian N1a1 and N1a2 sub-haplogroups mainly distribute in Mongolia and Northern China, occasionally appearing in Manchu, Xibe, and Ewenki populations [68].

Median-joining network analysis based on Y-STR data revealed Jiangsu Han occupying central nodal positions in O1 haplogroup networks, suggesting their potential role as a genetic radiation center. In O2a subhaplogroup networks, Jiangsu Han displayed multibranch divergent distributions, consistent with gene flow models between Northern and Southern Han populations, reflecting the complex historical processes of Central Plains population southward migration and Baiyue ethnic integration [69]. MDS analysis demonstrated significant genetic clustering between Jiangsu Han and Southern Han populations, with substantially smaller genetic distances compared to Northern Han and minority groups. AMOVA and admixture analysis clearly revealed Jiangsu Han's genetic admixture patterns from both Northern and Southern Han populations. Admixture analysis indicated 89% Southern Han contribution and 11% Northern Han component in Jiangsu Han's genetic composition, robustly confirming their close genetic affinity with Southern Han populations. This pattern likely results from historical large-scale southward migrations and sustained regional gene flow. While Northern Han contributions remain relatively minor, they indicate ongoing genetic exchanges potentially attributable to early Northern Han migrations or inter-ethnic interactions.

Conclusions

This study systematically characterizes the paternal genetic architecture of the Jiangsu Han population through integrated Y-STR and Y-SNP analysis, offering novel insights into their evolutionary history and forensic relevance. Genetic admixture results reveal Jiangsu Han population derive 89% of their ancestry from Southern Han populations, consistent with historical southward migrations, while retaining 11% Northern Han contributions, reflecting long-term regional gene flow. Median-joining network analysis further highlights Jiangsu Han as a central node in the O1 haplogroup radiation, underscoring their role in shaping Southern East Asian genetic diversity. Forensic applications are advanced by the Y29 multiplex system, which exhibits exceptional discriminative power (DC = 1) and haplotype diversity (HD = 1), providing a robust tool for individual identification in this population. These findings establish a critical foundation for constructing population-specific forensic databases and refining kinship inference models in East Asia. While this research enhances understanding of Jiangsu Han genetics, limitations include restricted sub-regional sampling and reliance on Y-chromosomal markers. Future studies integrating autosomal and mitochondrial data will further resolve fine-scale population structure. Collectively, this work contributes to the genetic characterization of Han subgroups, advancing both anthropological knowledge and forensic practice.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11634-6.

Supplementary Material 1
Supplementary Material 2
Supplementary Material 3
Supplementary Material 4
Supplementary Material 5
Supplementary Material 6
Supplementary Material 7
Supplementary Material 8
Supplementary Material 9

Author contributions

CF and CL proposed the initial idea. SZ and YB developed the methodology. ML carried out experiments and mainly wrote the manuscript. CF and CL supervised and administrated the project, HZ composed and drew figures. RT and AC investigated resulting data and validated results. PZ and CY made formal analysis. CF and ML acquired the funding. All authors read and approved the final manuscript.

Funding

This research was funded by Natural Science Foundation of Shandong Province(No. ZR2022QH337), the Visiting Research Fund for Teachers of Ordinary Undergraduate Universities in Shandong Province, National Natural Science Foundation of China (No. 82002006) and Qingchuang Talents Induction Program of Shandong Higher Education Institution (2022 Forensic medicine innovation team).

Data availability

Data is provided within the manuscript or supplementary information files.

Declarations

Ethics approval and consent to participate
The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Committee of Academy of Forensic Science, Ministry of Justice, China (2014-3-3-5). Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the samples to publish this paper.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 7 January 2025 / Accepted: 24 April 2025
Published online: 02 May 2025

References

1. Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature. 2003;423(6942):873–6.
2. Cantu ES, Marsh RD, Boecklen KE, Ostrer H. Selective protection of specific DNA sequences in the heterochromatin of C-banded human Y-chromosomes. Am J Hum Genet. 1988;43(6):948–53.
3. Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, CavalliSforza LL. Construction of human Y-chromosomal haplotypes using a new polymorphic A transition to G transition. Hum Mol Genet. 1994;3(12):2159–61.
4. Hammer MF, Zegura SL. The role of the Y chromosome in human evolutionary studies. Evol Anthropol. 1997;6(1):10.
5. Jobling MA, Pandya A, TylerSmith C. The Y chromosome in forensic analysis and paternity testing. Int J Legal Med. 1997;110(3):118–24.
6. Roewer L. Y chromosome STR typing in crime casework. Forensic Sci Med Pat. 2009;5(2):77–84.
7. Poznik GD, Xue Y, Mendez FL, Willems TF, Massaia A, Sayres MAW, Ayub Q, McCarthy SA, Narechania A, Kashin S, et al. Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. Nat Genet. 2016;48(6):593.
8. de Knijff P. Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome.am. J Hum Genet. 2000;67(5):1055–61.
9. Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, Zegura SL. Hierarchical patterns of global human Y-chromosome diversity. Mol Biol Evol. 2001;18(7):1189–203.
10. Kayser M. Forensic use of Y-chromosome DNA: a general overview. HUM GENET. 2017;136(5):621–35.
11. Pinto N, Gusmao L, Amorim A. Mutation and mutation rates at Y chromosome specific short tandem repeat polymorphisms (STRs): A reappraisal. Forensic Sci Int-Gen. 2014;9:20–4.
12. Xue YL, Wang QJ, Long Q, Ng BL, Swerdlow H, Burton J, Skuce C, Taylor R, Abdellah Z, Zhao YL, et al. Human Y chromosome Base-Substitution mutation rate measured by direct sequencing in a Deep-Rooting pedigree. Curr Biol. 2009;19(17):1453–7.
13. Hammer MF, Chamberlain VF, Kearney VF, Stover D, Zhang G, Karafet T, Walsh B, Redd AJ. Population structure of Y chromosome SNP haplogroups in the united States and forensic implications for constructing Y chromosome STR databases. Forensic Sci Int. 2006;164(1):45–55.
14. Cortellini V, Verzeletti A, Cerri N, Marino A, De Ferrari F. Y-chromosome polymorphisms and ethnic group - a combined STR and SNP approach in a population sample from Northern Italy. Croat Med J. 2013;54(3):279–85.
15. Lang M, Liu H, Song F, Qiao X, Ye Y, Ren H, Li J, Huang J, Xie M, Chen S, et al. Forensic characteristics and genetic analysis of both 27 Y-STRs and 143 Y-SNPs in Eastern Han Chinese population. Forensic Sci Int-Gen. 2019;42:E13–20.
16. Li M, Zhang Y, Luo L, Bian Y, Li C. Development and validation of a custom panel including 183 Y-SNPs for Chinese Y-chromosomal haplogroups dissection using a MALDI-TOF MS system. Electrophoresis. 2020;41(23):2047–54.
17. Nei M, Tajima F. DNA polymorphism detectable by restriction endonucleases. Genetics. 1981;97(1):145–63.
18. Fan H, Zhang X, Wang X, Ren Z, Li W, Long R, Liang A, Chen J, Song T, Qu Y, et al. Genetic analysis of 27 Y-STR loci in Han population from Hainan Province, southernmost China. Forensic Sci Int-Gen. 2018;33:E9–10.
19. Zhang J, Mo X, Zhang Y, Ding G, Wang X, Li W, Zhao X, Ye J. Genetic analysis of 26 Y-STR loci in Han population from Leshan, Southwest China. Forensic Sci Int-Gen. 2018;37:E15–6.
20. Liu C, Han X, Min Y, Liu H, Xu Q, Yang X, Huang S, Chen Z, Liu C. Genetic polymorphism analysis of 40 Y-chromosomal STR loci in seven populations from South China. Forensic Sci Int. 2018;291:109–14.
21. Ip SCY, Lin S, Lam T. Haplotype data of 27 Y-STR loci in Hong Kong Chinese. Forensic Sci Int-Gen. 2019;38:E14–5.
22. Tao R, Wang S, Zhang J, Zhang J, Yang Z, Zhang S, Li C. Genetic characterization of 27 Y-STR loci analyzed in the Nantong Han population residing along the Yangtze basin. Forensic Sci Int-Gen. 2019;39:E10–3.
23. Guan T, Song X, Xiao C, Sun H, Yang X, Liu C, Chen L. Analysis of 23 Y-STR loci in Chinese Jieyang Han population. Int J Legal Med. 2020;134(2):505–7.
24. Tao R, Jin M, Ji G, Zhang J, Zhang J, Yang Z, Chen C, Zhang S, Li C. Forensic characteristics of 36 Y-STR loci in a Changzhou Han population and genetic distance analysis among several Chinese populations. Forensic Sci Int-Gen. 2019;40:E268–70.
25. Zhou Y, Shao C, Li L, Zhang Y, Liu B, Yang Q, Tang Q, Li S, Xie J. Genetic analysis of 29 Y-STR loci in the Chinese Han population from Shanghai. Forensic Sci Int-Gen. 2018;32:E1–4.
26. Sun H, Su K, Fan C, Long F, Liu Y, Sun J, Mo X, Ge Y, Zhang L, Zhai L, et al. Y-STRs' genetic profiling of 1953 individuals from two Chinese Han populations (Guizhou and Shanxi). Forensic Sci Int-Gen. 2019;38:E8–10.
27. Wang C, Zhang J, Li X, Bai R, Shi M, Wang C. Haplotype analysis of 36 Y-STR loci in a Chinese Han population from Anhui Province, Eastern China. Int J Legal Med. 2020;134(6):2063–5.
28. Wang C, Li S, Hao X, Wang C, Bai R, Li H, Ma S. Forensic characteristics and phylogenetic analysis of the Chinese Han population from Chongqing municipality, Southwest China. LEGAL MED-TOKYO; 2021. p. 53.
29. Chen W, Feng Z, Jin W, Zhang Y, Wang Q, Li P, Li S, Wei L, Liu G. Genetic analysis of 17 Y-STR loci from 1026 individuals of Han populations in Jilin Province, Northeast China. Int J Legal Med. 2018;132(5):1309–11.
30. Huang Y, Guo L, Wang M, Zhang C, Kang L, Wang K, Ma Y, Jiao H, Li X, Sun H. Genetic analysis of 39 Y-STR loci in a Han population from Henan Province, central China. Int J Legal Med. 2019;133(1):95–7.
31. Zhang J, Tao R, Zhong J, Sun D, Qiao L, Shan S, Yang Z, Zhang J, Zhang S, Li C. Genetic polymorphisms of 27 Y-STR loci in the Dezhou Han population from Shandong Province, Eastern China. Forensic Sci Int-Gen. 2019;39:E26–8.
32. Zhao J, Jin X, Song Z, Song S, Cui W, Chen C, Shen C. Genetic distributions and diversity analyses of 23 Y-STR loci in Xi'an Han population. Legal Med - Tokyo. 2019; 41.
33. Cao S, Bai P, Zhu W, Chen D, Wang H, Jin B, Zhang L, Liang W. Genetic portrait of 27 Y-STR loci in the Tibetan ethnic population of the Qinghai Province of China. Forensic Sci Int-Gen. 2018;S1048572841.
34. Adnan A, Kasim K, Rakha A, Noor A, Cheema AS, Hadi S, Xing J. Population data of 23 Y STRs from Manchu population of Liaoning Province, Northeast China. Int J Legal Med. 2019;133(3):785–8.
35. Xuan J, Adnan A, Khan RA, Xing J, Yao J, Wang B. Population genetics of 19 Y-STR loci in Yanbian Korean samples from China. Ann Hum Genet. 2019;83(3):134–40.
36. Wang J, Wen S, Shi M, Liu Y, Zhang J, Bai R, Li H. Haplotype structure of 27 Yfiler plus loci in Chinese Dongxiang ethnic group and its genetic relationships with other populations. Forensic Sci Int-Gen. 2018;33:E13–6.
37. Fan H, Wang X, Chen H, Zhang X, Huang P, Long R, Liang A, Song T, Deng J. Population analysis of 27 Y-chromosomal STRs in the Li ethnic minority from Hainan Province, southernmost China. Forensic Sci Int-Gen. 2018;34:E20–2.
38. Wang C, Su M, Li Y, Chen L, Jin X, Wen S, Tan J, Shi M, Li H. Genetic polymorphisms of 27 Yfiler plus loci in the Daur and Mongolian ethnic minorities from Hulunbuir of inner Mongolia autonomous region, China. Forensic Sci Int-Gen. 2019;40:E252–5.
39. Chen D, Lv M, Huang Y, Hou Y, Yu T, Liang W. Population genetics of 27 Y-STRs for the Yi population from Liangshan Yi autonomous Prefecture, China. Int J Legal Med. 2021;135(2):441–2.
40. Li X, Zhang J, Li L, Zha L, Shi M, Ding M. Genetic polymorphism of 24 Y-STR loci in Altay Hui and Kazakh populations from Northwest China. Legal Med - Tokyo. 2020; 47.
41. Zhabagin M, Sarkytbayeva A, Tazhigulova I, Yerezhepov D, Li S, Akilzhanov R, Yeralinov A, Sabitov Z, Akilzhanova A. Development of the Kazakhstan

Y-chromosome haplotype reference database: analysis of 27 Y-STR in Kazakh population. Int J Legal Med. 2019;133(4):1029–32.
42. Adnan A, Rakha A, Kasim K, Noor A, Nazir S, Nadi S, Pang H. Genetic characterization of Y-chromosomal STRs in Hazara ethnic group of Pakistan and confirmation of DYS448 null allele. Int J Legal Med. 2019;133(3):789–93.
43. Khubrani YM, Wetton JH, Jobling MA. Extensive geographical and social structure in the paternal lineages of Saudi Arabia revealed by analysis of 27 Y-STRs. Forensic Sci Int-Gen. 2018;33:98–105.
44. Oh YN, Lee HY, Lee EY, Kim EH, Yang WI, Shin K. Haplotype and mutation analysis for newly suggested Y-STRs in Korean father-son pairs. Forensic Sci Int-Gen. 2015;15:64–8.
45. Li L, Xu Y, Luis JR, Alfonso-Sanchez MA, Zeng Z, Garcia-Bertrand R, Herrera RJ. Cebu, Thailand and Taiwanese aboriginal populations according to Y-STR loci. Gene. 2019;721S:100001.
46. Pickrahn I, Mueller E, Zahrer W, Dunkelmann B, Cemper-Kiesslich J, Kreindl G, Neuhuber F. Yfiler plus amplification kit validation and calculation of forensic parameters for two Austrian populations. Forensic Sci Int-Gen. 2016;21:90–4.
47. Ottaviani E, Vernarecci S, Asili P, Agostino A, Montagna P. Preliminary assessment of the prototype Yfiler plus kit in a population study of Northern Italian males. Int J Legal Med. 2015;129(4):729–30.
48. Iacovacci G, D'Atanasio E, Marini O, Coppa A, Sellitto D, Trombetta B, Berti A, Cruciani F. Forensic data and microvariant sequence characterization of 27 Y-STR loci analyzed in four Eastern African countries. Forensic Sci Int-Gen. 2017;27:123–31.
49. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and windows. Mol Ecol Resour. 2010;10(3):564–7.
50. Saitou N, Nei M. The Neighbor-Joining method-A new method for reconstructing phylogenetic trees. Mol Biol Evol. 1987;4(4):406–25.
51. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35(6):1547–9.
52. Gao T, Yun L, Zhou D, Lang M, Wang Z, Qian X, Liu J, Hou Y. Next-generation sequencing of 74 Y-SNPs to construct a concise consensus phylogeny tree for Chinese population. Forensic Sci Int-Gen SS. 2017;6:e96–8.
53. Song M, Wang Z, Zhang Y, Zhao C, Lang M, Xie M, Qian X, Wang M, Hou Y. 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-Gen. 2019;39:E14–20.
54. Li M, Zhou W, Zhang Y, Huang L, Wang X, Wu J, Meng M, Wang H, Li C, Bian Y. Development and validation of a novel 29-plex Y-STR typing system for forensic application. Forensic Sci Int-Gen. 2020; 44.
55. Xie M, Song F, Li J, Lang M, Luo H, Wang Z, Wu J, Li C, Tian C, Wang W, et al. Genetic substructure and forensic characteristics of Chinese Hui populations using 157 Y-SNPs and 27 Y-STRs. Forensic Sci Int-Gen. 2019;41:11–8.
56. Choi SS, Park KH, Nam DE, Kang TH, Chung KW. Y-chromosome haplogrouping for Asians using Y-SNP target sequencing. Forensic Sci Int-Gen SS. 2017;6:e235–7.
57. Trejaut JA, Poloni ES, Yen J, Lai Y, Loo J, Lee C, He C, Lin M. Taiwan Y-chromosomal DNA variation and its relationship with Island Southeast Asia. BMC Genet. 2014; 15.
58. Wang CC, Yeh HY, Popov AN, Zhang HQ, Matsumura H, Sirak K, Cheronet O, Kovalev A, Rohland N, Kim AM, et al. Genomic insights into the formation of human populations in East Asia. Nature. 2021;591(7850):413.
59. Bandelt HJ, Forster P, Rohl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999;16:37–48.
60. Dupuy B, Stenersen M, Egeland T, Olaisen B. Y-chromosomal microsatellite mutation rates: differences in mutation rate between and within loci. Hum Mutat. 2004;23:117–24.
61. Bertorelle G, Excoffier L. Inferring admixture proportions from molecular data. Mol Biol Evol. 1998;15:1298–311.
62. Dupanloup I, Bertorelle G. Inferring admixture proportions from molecular data: extension to any number of parental populations. Mol Biol Evol. 2001;18(4):672–5.
63. Zhong H, Shi H, Qi X, Duan Z, Tan P, Jin L, Su B, Ma RZ. Extended Y chromosome investigation suggests postglacial migrations of modern humans into East Asia via the Northern route. Mol Biol Evol. 2011;28(1):717–27.
64. Yan S, Wang C, Li H, Li S, Jin L. An updated tree of Y-chromosome haplogroup O and revised phylogenetic positions of mutations P164 and PK4. Eur J Hum Genet. 2011;19(9):1018.
65. Yin C, He Z, Wang Y, He X, Zhang X, Xia M, Zhai D, Chang K, Chen X, Chen X, et al. Improving the regional Y-STR haplotype resolution utilizing haplogroup-determining Y-SNPs and the application of machine learning in Y-SNP haplogroup prediction in a forensic Y-STR database. A pilot study on male Chinese Yunnan Zhaoyang Han population. Forensic Sci Int-Gen. 2022;57:102659.
66. Wen B, Xie XH, Gao S, Li H, Shi H, Song XF, Qian TZ, Xiao CJ, Jin JZ, Su B, et al. Analyses of genetic structure of Tibeto-Burman populations reveals sex-biased admixture in Southern Tibeto-Burmans. Am J Hum Genet. 2004;74(5):856–65.
67. Qi XB, Cui CY, Peng Y, Zhang XM, Yang ZH, Zhong H, Zhang H, Xiang K, Cao XY, Wang Y, et al. Genetic evidence of paleolithic colonization and neolithic expansion of modern humans on the Tibetan plateau. Mol Biol Evol. 2013;30(8):1761–78.
68. Rootsi S, Zhivotovsky LA, Baldovic M, Kayser M, Kutuev IA, Khusainova R, Bermisheva MA, Gubina M, Fedorova SA, Ilumae A, et al. A counter-clockwise Northern route of the Y-chromosome haplogroup N from Southeast Asia towards Europe. Eur J Hum Genet. 2007;15(2):204–11.
69. Ma H, Zhou Y, Wang R, Yan F, Chen H, Qiu L, Zhao J, Jin L, Wang C. Ancient genomes shed light on the long-term genetic stability in the central plain of China. Sci Bull. 2025;70(3):333–7.

Publisher's note