Population genetic analysis of Shaanxi male Han Chinese population reveals genetic differentiation and homogenization of East Asians

Luyao Li | Xing Zou | Guanjun Zhang | Hongyan Wang | Yongdong Su | Mengge Wang | Guanglin He

1Department of Pathology, The First Affiliated Hospital of Xi’an Jiaotong University, Xi’an, Shaanxi, China
2Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, Chengdu, Sichuan, China
3Forensic Identification Center, Public Security Bureau of Tibet Tibetan Autonomous Region, Lhasa, Tibet Tibetan Autonomous Region, China

Correspondence
Mengge Wang and Guanglin He, Institute of Forensic Medicine, West China School of Basic Science and Forensic Medicine, Sichuan University, 3-16 Renmin South Road, Chengdu, Sichuan 610041, China.
Email: Menggewang2021@163.com (M. W.) and Guanglinhescu@163.com (G. H.)

Abstract

Background: Shaanxi province, located in the upper Yellow River, has been evidenced as the geographic origin of Chinese civilization, Sino-Tibetan-speaking language, and foxtail or broomcorn millet farmers via the linguistic phylogenetic spectrum, archeological documents, and genetic evidence. Nowadays, Han Chinese is the dominant population in this area. The formation process of modern Shaanxi Han population reconstructed via the ancient DNA is on the way, however, the patterns of genetic relationships of modern Shaanxi Han, allele frequency distributions of high mutated short tandem repeats (STRs) and corresponding forensic parameters are remained to be explored.

Methods: Here, we successfully genotyped 23 autosomal STRs in 630 unrelated Shaanxi male Han individuals using the recently updated Huaxia Platinum PCR amplification system. Forensic allele frequency and parameters of all autosomal STRs were assessed. And comprehensive population genetic structure was explored via various typical statistical technologies.

Results: Population genetic analysis based on the raw-genotype dataset among 15,803 Eurasian individuals and frequency datasets among 56 populations generally illustrated that linguistic stratification is significantly associated with the genetic substructure of the East Asian population. Principal component analysis, multidimensional scaling plots and phylogenetic tree further demonstrated that Shaanxi Han has a close genetic relationship with geographically close Shanxi Han, and showed that Han Chinese is a homogeneous population during the historic and recent admixture from the STR variations. Except for Sinitic-speaking populations, Shaanxi Han harbored more alleles sharing with Tibeto-Burman-speaking populations than with other reference populations. Focused on the allele frequency correlation and forensic parameters, all loci are in accordance with the minimum requirements of HWE and LD. The observed combined probability of discrimination of 8.2201E-28 and the
1 | INTRODUCTION

Forensic DNA profiling with sets of highly polymorphic short tandem repeat (STR) loci has become a pivotal niche in forensic investigations for nearly 30 years (Hagelberg, Gray, & Jeffreys, 1991; Kayser & de Knijff, 2011). STRs, also referred to as microsatellites, are DNA sequences containing a variable number of tandemly repeated short sequence motifs (2–6 bp) and are ubiquitously scattered throughout the eukaryotic genomes (Ellegren, 2004). STR profiling has played a key role in identifying perpetrators and missing persons, determining kinship and establishing national forensic DNA databases. During the past few decades, the increasing body of STR-based population data has replenished different national DNA databases and facilitated data sharing. To minimize adventitious matches and improve discriminating power, the officially recommended 13-CODIS (Combined DNA Index System) core loci were expanded to 20 STRs with the addition of 7 new loci (D1S1656, D2S1338, D2S441, D10S1248, D12S391, D19S433, and D22S1045) (Hares, 2015). For the sake of increasing the compatibility with expanded CODIS and the world's largest DNA database - Chinese National Database (CND), the Huaxia Platinum System (Thermo Fisher Scientific) covering all recommended loci in the expanded CODIS and the CND has been launched (Wang et al., 2016b, 2018). This system is a six-dye, 25-locus, multiplex assay that allows co-amplification and fluorescent detection of the 23 autosomal STRs (D1S1656, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D21S11, D22S1045, D6S1043, CSF1PO, FGA, TH01, TPOX, VWA, Penta D and Penta E) and Amelogenin as well as Y-InDel (rs2032678) for sex determination. However, previous Huaxia Platinum System-based studies have focused almost exclusively on ethnic groups (Liu, Wang, He, Wang, & Hou, 2019; Wang et al., 2016b, 2018).

The Han Chinese population, as the largest ethnic group in the world, is nonetheless underrepresented in forensic investigations to catalog the forensically genetic variants (Chen, Wu, Luo, et al., 2019; He, Wang, Liu, Hou, & Wang, 2018; He, Wang, Wang, Zou, et al., 2018). Due to its large population size, large-scale demographic migration and population expansion facilitated by ancient agriculture, genetic admixture with adjacent ethnic groups, and substantial genetic diversity among Han Chinese had been observed in previous studies (Chen, Wu, Luo, et al., 2019; Chiang, Mangul, Robles, & Sankararaman, 2018; Gao et al., 2019; Lang et al., 2019; Stoneking & Delfin, 2010). Previous whole-genome or uniparentally genetic studies (Gao et al., 2019; Lang et al., 2019; Li, Ye, et al., 2019; Liu et al., 2018) have shed light on a general South-North genetic divergence among Han Chinese. Genetic evidence based on low-coverage whole-genome sequencing of over ten thousand Han Chinese revealed an East-West cline (Chiang et al., 2018). Furthermore, archaeological, anthropological, lexical, and genetic findings have provided evidence that the Han Chinese could trace a common ancestry in the Yellow River basin of northern China (Blench, Sagart, & Sanchez-Mazas, 2005; Zhang, Yan, Pan, & Jin, 2019), and the population expansions and migrations of Han Chinese were driven by the development of the Yangshao and/or Majiayao Neolithic cultures. Our previous study (Chen, Wu, Luo, et al., 2019) has investigated the forensic features, genetic diversity and phylogenetic affinity of northern Han Chinese residing in Shanxi province on the basis of 23 autosomal STRs. Nevertheless, forensic characteristics and genetic makeup of Han Chinese living in Shaanxi Province are still underrepresented. Shaanxi province, lying in central China, stretching from the Qin Mountains and Shannan in the South to the Ordos Desert in the North and comprising the Wei Valley and much of the surrounding Loess Plateau, is considered one of the early cradles of Chinese civilization. Recent archeological plant documents of the earliest staple crop domestication further demonstrated that broomcorn and foxtail millet farmers originated from Shaanxi and surrounding regions (Leipe, Long, Sergusheva, Wagner, & Tarasov, 2019). Linguistic and mitochondrial evidence further supported that this region is the cradle of the formation of Tibeto-Burman and Sinitic-speaking populations (Li, Tian, et al., 2019; Zhang et al., 2019). The current capital of Shaanxi province—Xi'an, cumulative power of exclusion of 0.9999999995 in Shaanxi Han demonstrated that the studied STR loci are informative and polymorphic, and this system can be used as a powerful routine forensic tool in personal identification and parentage testing.

Conclusion: Both the geographical and linguistic divisions have shaped the genetic structure of modern East Asian. And more forensic reference data should be obtained for ethnically, culturally, geographically and linguistically different populations for better routine forensic practice and population genetic studies.

KEYWORDS
forensic science, genetic differentiation, Han Chinese, population genetics, short tandem repeats
is one of the four great ancient capitals of China and is the eastern terminus of the Silk Road. Hence, Shaanxi Province plays a significant role in the peopling of Neolithic populations and the dissection of genetic variations of Han Chinese settling in Shaanxi province is indispensable for uncovering the origin, migration, expansion, and admixture of the Han Chinese population.

2 | MATERIALS AND METHODS

2.1 | Sample preparation and DNA extraction

A batch of blood samples was collected from 630 healthy unrelated male Han Chinese individuals residing in Shaanxi province. All participants enrolled in the present study had signed the written informed consents and provided self-declared ethnicity information. This project was endorsed by the institutional review board of the First Affiliated Hospital of Xi’an Jiaotong University and carried out in accordance with the recommendations of the Declaration of Helsinki (Nicogossian, Kloiber, & Stabile, 2014). Human genomic DNA was isolated by applying the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer’s guidelines and the quantity of DNA template was estimated using the Nanodrop-2000c (Thermo Fisher Scientific).

2.2 | PCR amplification and profiling

All samples were typed using the Huaxia Platinum PCR amplification kit (Thermo Fisher Scientific) according to the manufacturer’s instructions. Multiplex amplification was performed on a ProFlex 96-well PCR System (Thermo Fisher Scientific) following the manufacturer’s protocol. The reaction mix for each sample was prepared in 25 μl volume containing 10 μl of the master mix, 10 μl of primer set, 1 μl of DNA template and 4 μl of deionized water. We employed the following thermal cycler conditions: pre-denaturation for 1 min at 95°C, followed by 26 cycles of 94°C for 3 s, 59°C for 16 s, 65°C for 29 s, then a final extension at 60°C for 5 min, and holding at 4°C. The PCR products were electrophoresed and detected on the Applied Biosystems 3500XL Genetic Analyzer (Thermo Fisher Scientific) using POP-4 polymer. The genotype profiles were obtained by comparing with the matching allelic ladder via GeneMapper ID-X v.1.4 (Thermo Fisher Scientific).

2.3 | Quality control

This study was conducted in an ISO 17025 accredited laboratory, which has also been accredited by the China National Accreditation Service for Conformity Assessment (CNAS). The experiment was carried out in strict accordance with the recommendations proposed by the International Society for Forensic Genetics (ISFG) (Schneider, 2007). Laboratory internal standards and manufacturer’s protocols were strictly abided to minimize errors. Negative control (ddH2O) and positive control (Control DNA 007) were genotyped for each batch of genotyping.

2.4 | Dataset composition

We first merged our 630 raw genotypes of 20 overlapping STRs among different commercial STR amplification kits with 15,173 genotypes from 19 Eurasian populations (six Turkic-speaking populations [Chen, Zou, Wang, Wang, & He, 2019; Chen, Zou, Wang, Gao, Su, et al., 2019; Jin et al., 2017; Liu et al., 2019]: Urumqi Uyghur, Hotan Uyghur, Kumul Uyghur1, Xinjiang Uyghur, Artux Uyghur, and Akto Kyrgyz; five Han Chinese populations [Chen, Zou, Wang, Gao, Su, et al., 2019; He, Wang, Liu, et al., 2018; He, Wang, Wang, Zou, et al., 2018; Liu et al., 2019; Wang et al., 2018]: Zhujiang Han, Shaxi Han, Chengdu Han, Wuzhong Hui, and Hainan Han; four Tibeto-Burman-speaking populations [Liu et al., 2019; Wang et al., 2018]: Liangshan Tibetan, Chengdu Tibetan, Tibet Tibetan, and Liangshan Yi; four western Eurasian populations [Alsafiah, Goodwin, Hadi, Alshaikhi, & Wepeba, 2017; Chen, Adnan, Rakha, et al., 2017; Ossowski et al., 2017; Sadam et al., 2015]: Quetta Hazara, Estonian, Poland, and Saudi Arabian). We referred to this dataset as the raw-genotype dataset. Subsequently, a dataset merging allele frequency distribution among Shaanxi Han population and other 55 worldwide populations (Almeida et al., 2015; Choi et al., 2017; Fuji et al., 2014; Gaviria et al., 2013; Guerreiro, Ribeiro, Porto, Carneiro de Sousa, & Dario, 2017; Hossain et al., 2016; Moyses et al., 2016; Ossowski et al., 2017; Park et al., 2013, 2016; Taylor, Bright, McGovern, Neville, & Grover, 2017; Wang et al., 2016a; Wu, Pei, Ran, & Song, 2017; Yang et al., 2018; Zhang, Xia, et al., 2016; Zhang, Yang, et al., 2016) was edited from the published literature (here referred as frequency dataset) based on the 20 overlapping STRs (CSF1PO, D10S1248, D12S391, D13S317, D16S539, D18S51, D19S433, D1S1656, D21S11, D22S1045, D2S1338, D2S441, D3S1358, D5S818, D7S820, D8S1179, FGA, TH01, TPOX, and vWA).

2.5 | Statistical analysis

We performed the exact test of Hardy-Weinberg equilibrium (HWE) in the Arlequin with the following parameter
settings: the number of steps in Markov Chn is 1,000,000 and the number of dememorization steps is 100,000. And we tested the Linkage disequilibrium between all pairs (23 autosomal STRs) of loci with the parameter settings: number of permutations: 10,000 and number of initial conditions of expectation-maximization (EM): 2 using Arlequin 3.5 (Excoffier & Lischer, 2010). The expected heterozygosity (Ho) and expected heterozygosity (He) were also calculated using the aforementioned parameters instrumented in Arlequin 3.5 (Excoffier & Lischer, 2010). We calculated forensic allele frequency and corresponding forensic parameters, including gene diversity (GD), polymorphism information content (PIC), matching probability (PM), discrimination power (PD), typical paternity index (TPI), power of paternity exclusion (PE), and p values of Hardy–Weinberg equilibrium using the STRAF (Gouy & Zieger, 2017). We calculated the pairwise Fst genetic distance among 20 populations included in the raw-genotype dataset using STRAF and calculated the pairwise Nei’s genetic distance among 56 global populations based on the frequency dataset using the Phylip software (Cummings, 2004). Principal component analysis (PCA) based on the allele frequency distribution among 56 populations were performed using the Multivariate Statistical Package (MVSP) software 3.22 (Kovach, 2007), and we subsequently pruned the populations out of Eurasian or East Asian to explore and zoom in the patterns of genetic relationship between eastern Eurasian or East Asian. Multidimensional scaling plots among worldwide populations or East Asians were performed using our in-house R-script. Phylogenetic relationships among worldwide or East Asian populations were reconstructed using Mega 7.0 (Kumar, Stecher, & Tamura, 2016). Model-based Structure analysis was carried out using STRUCTURE (Evanno, Regnaut, & Goudet, 2005).

3 | RESULTS AND DISCUSSION

3.1 | Allele frequency correlation and forensic parameters

We successfully genotyped 23 autosomal STRs and two sex-determination loci in 630 unrelated Han Chinese individuals residing in Shaanxi province located in the central plain of northern China using Huaxia Platinum amplification kit. No deviations from the linkage disequilibrium were observed after Bonferroni Correction (Table S1). Allele frequency and corresponding forensic parameters of 23 autosomal STRs are presented in Table 1. All 23 autosomal STRs are in line with the Hardy-Weinberg equilibrium. Here, a total of 271 alleles were identified in Shaanxi Han with corresponding allele frequency spanning from 0.0008 to 0.5143. TH01 harbored the smallest allele number (6), followed by TPOX (7), while Penta E possessed the largest allele number (20), followed by FGA (19). The GD values were identified ranging from 0.6436 (TH01) to 0.9170 (Penta E). GD and Ho ranged from 0.6436 to 0.9170 and 0.6238 to 0.9190, respectively. PIC values were observed spanning 0.5920 (TH01) to 0.9101 (Penta E), which is consistent with the observed minimum and maximum allele number. PM was observed ranging from 0.0146 to 0.1857 and PD spanned from 0.8143 to 0.9854. The PE values ranged from 0.3204 in the locus of TPOX and 0.8345 in the locus of Penta E. The observed individual forensic parameters of 23 autosomal STRs are suitable for choosing and applying these markers in individual identification and parentage testing, not for biogeographic ancestry inference of Shaanxi Han population. Combining the forensic effectiveness of all included loci, we identified that the combined probability of discrimination in Shaanxi Han Chinese is 8.2201E-28 and the cumulative power of exclusion in this studied population is 0.9999999995. In accordance with the observed patterns of forensic characteristics in Chinese Turkic-speaking, Tibetan-speaking and other Sinitic-speaking populations, all included loci (23 autosomal STRs and two sex-determinate loci) included in the Huaxia Platinum amplification kit are informative and polymorphic in Shaanxi Han Chinese population. This kit with more included STR loci than other STR kits such as the AmpF/STR Identifier PCR amplification kit (He, Su, et al., 2019) (15 STRs, Thermo Fisher Scientific) is more suitable for Chinese National Database construction and forensic routine personal identification and paternity discrimination.

3.2 | Population comparisons among Eurasian populations via raw-genotype dataset

To explore the similarities and differences in the genetic material of Shaanxi Han population and Eurasian reference populations, pairwise Fst genetic distances among 20 populations included in the raw-genotype dataset were calculated and presented in Table S2 and visualized in Figure 1a. Shaanxi Han population has a close genetic relationship with the geographically close Shaanxi Han population with the smallest pairwise Fst genetic distance (0.0002), followed by populations belonging to the same language groups (Sinitic-speaking Zhujiang Han: 0.0007, Chengdu Han: 0.0008 and Wuzhong Hui: 0.0009). The distant genetic relationship with Shaanxi Han in the raw-genotype dataset was identified with the western Eurasian population (Poland: 0.0163). Turkic-speaking populations have intermediate relationships with this studied Han Chinese population (average ± standard error: 0.0040 ± 0.0016). Patterns of genetic similarities and differences were then explored via MDS based on the top three dimensions and visualized in Figure 1b,c. Western Eurasian populations (Saudi Arabian, Poland and Estonian)
TABLE 1

Allele frequency distributions and corresponding forensic parameters of 23 autosomal short tandem repeats (STRs) included in the Huaxia Platinum kit in Shaanxi Han populations.

| Locus   | Allele Frequency | Forensic Parameter |
|---------|------------------|--------------------|
| CSF1PO  | 0.0008            | 0.0024             |
| D10S1248| 0.0000            | 0.0016             |
| D12S391 | 0.0008            | 0.0040             |
| D13S317 | 0.0000            | 0.0008             |
| D16S539 | 0.0000            | 0.0000             |
| D18S51  | 0.0000            | 0.0000             |
| D19S433 | 0.0000            | 0.0000             |
| D1S1656 | 0.0000            | 0.0000             |
| D21S11  | 0.0000            | 0.0000             |
| D22S1045| 0.0000            | 0.0000             |
| D2S1338 | 0.0000            | 0.0000             |
| D2S441  | 0.0000            | 0.0000             |
| D3S1358 | 0.0000            | 0.0000             |
| D5S818  | 0.0000            | 0.0000             |
| D6S1043 | 0.0000            | 0.0000             |
| D7S820  | 0.0000            | 0.0000             |
| D8S1179 | 0.0000            | 0.0000             |
| FGA     | 0.0000            | 0.0000             |
| PentaD  | 0.0000            | 0.0000             |
| PentaE  | 0.0000            | 0.0000             |
| TH01    | 0.0000            | 0.0000             |
| TPOX    | 0.0000            | 0.0000             |
| vWA     | 0.0000            | 0.0000             |

(Continues)
| Locus | CSF1PO | DIBS1248 | D12S391 | D13S317 | D16S539 | D18S51 | D19S433 | D1S1656 | D21S11 | D22S1045 | D2S1338 | D4S546 | D7S820 | D8S1179 | FGA | PentaD | PentaE | TH01 | TPOX | vWA |
|-------|--------|-----------|---------|---------|---------|--------|---------|---------|-------|---------|---------|---------|--------|---------|------|--------|--------|------|------|------|
| 22    | 0.0778 | 0.0175    | 0.0484  | 0.016   | 0.1571  | 0.0113 |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 22.2  |        |           |         |         |         |        |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 22.3  |        |           |         |         |         | 0.0071 |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 23    | 0.0365 | 0.003     | 0.2294  |         |         | 0.2102 | 0.0079  |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 23.2  |        |           |         |         |         | 0.119  |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 24    | 0.0135 | 0.0040    | 0.1802  |         |         | 0.1665 |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 24.2  |        |           |         |         |         | 0.040  |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 25    | 0.0111 | 0.0032    | 0.095   |         |         | 0.1135 | 0.0032  |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 25.2  |        |           |         |         |         | 0.063  |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 26    | 0.0024 | 0.0016    | 0.038   |         |         | 0.057  | 0.0024  |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 26.2  |        |           |         |         |         | 0.008  |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 27    |        |           |         | 0.0008  | 0.016  | 0.040  |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 27.2  |        |           |         | 0.0044  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 28    |        |           |         | 0.0087  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 28.2  |        |           |         | 0.2683  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 29    |        |           |         | 0.0040  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 29.2  |        |           |         | 0.0016  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 30    |        |           |         | 0.2881  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 30.2  |        |           |         | 0.0127  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 30.3  |        |           |         | 0.0234  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 31    |        |           |         | 0.1008  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 31.2  |        |           |         | 0.0698  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 31.3  |        |           |         | 0.0271  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 32    |        |           |         | 0.1079  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 32.2  |        |           |         | 0.0711  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 33    |        |           |         | 0.0500  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 34    |        |           |         | 0.0024  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |
| 34.2  |        |           |         | 0.0024  |         |         |         |         |       |         |         |         |        |        |      |        |        |      |      |      |

Abbreviations: He, expected heterozygosity; Ho, observed heterozygosity; PD, discrimination power; PE, power of paternity exclusion; pHWE, p values of Hardy–Weinberg equilibrium; PIC, polymorphism information content; PM, matching probability; TPI, typical paternity index.
were scattered than other patterns of eastern Eurasian populations. Here, we found that a close genetic affinity was identified between Hazara and Turkic-speaking populations, which is consistent with our recent findings that the Hazara population is mixture descendants of Mongolian and local central Asians via high-density genome-wide data and indel markers (He, Adnan, et al., 2019). Shaanxi Han was scattered in Figure 1b and located between Chengdu Tibetan and Liangshan Tibetan in Figure 1c. These observed patterns of genetic affinity may partially reflect the common origin of Sino-Tibetan-speaking populations in the Upper and middle Yellow River (including the studied Shaanxi province) approximately 5,900 years before the present (Zhang et al., 2019). It should be cautious that some artifacts can be made due to the low discrimination of STR markers in population substructure exploration. Thus, to provide more

**FIGURE 1** The genetic affinity between Shaanxi Han and 19 Eurasian reference populations based on the raw-genotype data. (a) Heatmap of the pairwise Fst genetic distance. (b and c) Genetic relationship patterns among 20 populations inferred from genetic distance matrix based on the top three dimensions. (d) Phylogenetic relationship reconstruction revealed the genetic affinity between Shaanxi Han and both Sinitic and Tibeto-Burman-speaking populations. (e) Model-based Structure results showed the individual and population ancestry proportion of Shaanxi Han and other reference populations.
FIGURE 2  Principal component analysis among 56 worldwide populations based on the allele frequency distribution. (a–d) Two-dimensional plots respectively reconstructed based on the random combination of the top four components.

FIGURE 3  Geographic position of Shaanxi Han and other included reference populations and the pairwise Nei's genetic distance between Shaanxi Han and the reference populations. Red color means the larger genetic distance between the studied population and the targeted reference and green color denotes the smaller genetic distance.
genetic evidence of the similarities and differences of genetic inheritance of these populations, we reconstructed the neighbor-joining tree in Figure 1d. Four genetic clusters can be identified in the phylogenetic relationship reconstruction result: Tibeto-Burman-speaking cluster, Sinitic-speaking cluster, Turkic-speaking cluster, and western Eurasian cluster. Here, we observed that Shaanxi Han was localized in the intermediate position between Tibeto-Burman-speaking populations and Sinitic-speaking populations.

Individual and population ancestry composition was dissected via model-based Structure analysis among 15,803 individuals (Figure 1e). At k = 2, all individuals were assigned two predefined ancestries: AntiqueWhite ancestry represented as western Eurasian ancestry and LightSkyBlue ancestry represented as eastern Eurasian ancestry. LightSkyBlue ancestry was maximized in Chengdu Tibetan (0.978) and AntiqueWhite ancestry was maximized in Poland (0.977). Turkic-speaking populations can be modeled as mixture of one population associated with European ancestry and one population-linked with east Asian (Xinjiang Uyghur (0.477; 0.523), Urumqi Uyghur (0.487; 0.513), Kumul Uyghur (0.508; 0.492), Hotan Uyghur (0.53; 0.47), Akto Kyrgyz (0.571; 0.429), Artux Uyghur (0.593; 0.407)). These patterns of European-Asian admixture were further supported our previous findings of the mixed formation of modern Turkic-speaking populations via ancestry-informative markers (He, Wang, Wang, Luo, et al., 2018) and the previous genome-wide survey of northern and southern Uyghurs via Xu, Huang, Qian, and Jin (2008). At k = 3, two predefined ancestries enriched in Han Chinese populations were observed (LightSkyBlue ancestry only enriched in Han Chinese populations and maximized in Zhujiang Han:0.495;
and ForestGreen ancestry enriched in all eastern Asians and maximized in Liangshan Tibetan: 0.903 and Chengdu Tibetan: 0.868). Here, we can define LightSkyBlue ancestry as Han-dominant ancestry and ForestGreen ancestry as Tibetan-dominant ancestry. The third AntiqueWhite ancestry is representative of European ancestry, which was maximized in Poland (0.871). We can model Shaanxi Han as a mixture of 0.752 Liangshan Tibetan-related ancestry, 0.231 Zhujiang Han-related ancestry and only 0.017 Poland-related ancestry. At k = 4 or 5, two new ancestries maximized in Saudi Arabian and Turkic populations were identified.

### 3.3 Comprehensive genetic relationship among worldwide populations via frequency-dataset

To further investigate the genetic homogeneity and heterozygosity between the Shaanxi Han population and more reference populations and dataset consisting of allele frequency distribution, we merged our allele frequency correlation with allele frequency data from 55 worldwide populations. We first carried out the principal component analysis among 56 populations based on 613 alleles of 20 autosomal STRs. The top ten components can extract 84.083% variations from the genetic variations of 56 worldwide populations. First to tenth component accounted for 32.489%, 15.532%, 10.029%, 7.900%, 6.400%, 3.544%, 2.509%, 2.064%, 1.840%, and 1.775%, respectively. Patterns of genetic relationship among 56 populations revealed by the top four components (65.951%) are showed in Figure 2. Generally, PC1 can differentiate East Asians from other populations and PC2 ~ PC4 mainly differentiate some small substructure within-continental populations. Due to some migrant reference populations from Africa, America and Oceania were included here, no obvious population cluster could be identified. We further removed these continental populations and focused on the genetic variations of East Asians and South Asians. 88.439% variations can be extracted via the top ten components (PC1 ~ PC10 can, respectively, account for 39.334%, 11.917%, 10.210%, 6.819%, 5.258%, 4.186%, 3.246%, 3.07%, 2.427%, and 1.969%). As shown in Figure S1, PC1 can separate East Asians and Turkic-speaking populations, and PC2 can separate South Asian Bangladeshi and Indian. PC3 and PC4 can separate Tibeto-Burman-speaking populations and Japonic&Koreanic populations from others, respectively. The genetic affinity between Shaanxi Han and Central Han, Shanxi Han, and Sichuan Han can be identified here. We finally excluded two South Asian populations (Bangladeshi and Indian) and carried another PCA analysis (Figure S2). A total of 88.637% variations can be revealed by the first ten components. Three clear clusters can be observed: Turkic cluster, Tibeto-Burman cluster, and others.

Subsequently, population genetic relationships were explored via pairwise genetic distance, multidimensional scaling plots and phylogenetic relationship reconstruction. Figure 3 and Table S3 showed the pairwise Nei’s genetic distances between Shaanxi Han and the other 55 worldwide reference populations. The smallest genetic distance was 0.0029 observing between Shaanxi Han and Shanxi Han, followed by 0.0031 in Central Han, 0.0059 in Guangdong Han. As expected, the largest genetic distance with Shaanxi Han was identified in South African populations (AmaZulu: 0.1794 and AmaXhosa: 0.2012), followed by the indigenous Oceanian Polynesian. Heatmap among 56 populations based on the pairwise Nei’s genetic distance matrix is shown in Figure 4. The largest distances (shown as green color) were identified between Polynesian and South Asian indigenous

![Multidimensional scaling plot result of populations included in the dataset2. (a) Patterns of genetic similarities and differences among all 56 worldwide populations. (b) Population genetic substructure is associated with linguistic stratification in East Asian](image-url)
population and the smallest genetic distances (shown as red color) were observed within-continental populations, especially in Han Chinese populations. Heatmap also clustered Shaanxi Han with Central Han, Southern Xiamen Han, and Northern Shanxi Han. Genetic clusters further explored using Multidimensional scaling plots among 56 worldwide populations (Figure 5a) and East Asian populations (Figure 5b). East Asians were localized in the left part and Turkic speakers were located in the intermediate position between East Asians and others from Europe, Africa, America and Oceania in the worldwide two-dimensional plots. In the East Asian two-dimensional plots, similar patterns of genetic relationships with PCA results were observed. Shanxi Han and Central Han Chinese populations clustered tightly with Shaanxi Han. We finally built the phylogenetic relationship on the basis of the Nei’s genetic distance matrix (Figure 6). Three main branches can be categorized: African and Oceanian indigenous branch, European and American branch, and East Asian Branch. We can find that populations with similar ethnic origins tend to be formed a clade. Linguistic stratification was significantly

**FIGURE 6** Neighbor-Joining phylogenetic tree. Similar color denotes the common geographic origin of continental populations or the same linguistic origin of East Asian populations.
associated with population genetic substructure in East Asian, significant examples for Sinitic, Tibeto-Burman and the Turkic language groups included here. Shaanxi Han was first clustered with southern Han Chinese populations (Sichuan Han and Central Han) and then formed a clade with Shanxi Han.

4 | CONCLUSION

We provided the first batch forensic dataset of 23 STRs included in the Huaxia Platinum PCR amplification kit from the Han population residing in Shaanxi, near the Loess Plateau which was thought of as the origin of Chinese civilization and Sino-Tibetan-speaking populations. Comprehensive population genetic analyses based on the raw-genotype dataset and frequency-dataset consistently provided new insights into the population substructure of East Asians: linguistic stratification was significantly associated with population genetic substructure. Pairwise genetic distance, PCA, MDS, heat map, neighbor-joining tree, as well as model-based individual and population ancestry composition dissection demonstrated that Shaanxi Han harbored a close genetic relationship with the geographically close Shanxi Han, followed by other Han Chinese and Tibeto-Burman-speaking populations. Significant genetic homogenization was identified in Han Chinese and genetic differentiation was observed among populations belonging to different language families. Allele frequency distribution, parameters focused on forensic effectiveness indicated that forensic markers included in the Huaxia Platinum kit are highly informative and polymorphic in Shaanxi Han populations and can be used as the routine forensic practice.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

AUTHOR CONTRIBUTIONS

GH and LL conceived the idea for the study. GZ, HW, XZ, and MW performed or supervised the laboratory work. GH, LL, GZ, HW, XZ, and MW analyzed the data. GH, MW wrote and edited the manuscript. We would like to thank Prof. Renata Jacewicz (Department of Forensic Genetics, Pomeranian Medical University in Szczecin, Poland), Prof. Hussain M. Alsafiah (Forensic Genetics Laboratory, General Administration of Criminal Evidences, Public Security, Ministry of Interior, Saudi Arabia), Prof. I. Zupanič Pajnič (Institute of Forensic Medicine, Faculty of Medicine, University of Ljubljana) for sharing the raw genotype data as our reference data.

ETHICS STATEMENT

This study was approved by the First Affiliated Hospital of Xi’an Jiaotong University. This study followed the recommendations of the World Medical Association Declaration of Helsinki.

ORCID

Luyao Li https://orcid.org/0000-0003-1397-5045
Guanglin He https://orcid.org/0000-0002-6614-5267

REFERENCES

Almeida, C., Ribeiro, T., Oliveira, A. R., Porto, M. J., Costa Santos, J., Dias, D., & Dario, P. (2015). Population data of the GlobalFiler(R) Express loci in South Portuguese population. Forensic Science International: Genetics, 19, 39–41. https://doi.org/10.1016/j.fsigen.2015.06.001
Alsafiah, H. M., Goodwin, W. H., Hadi, S., Alshaikhi, M. A., & Wepeba, P. P. (2017). Population genetic data for 21 autosomal STR loci for the Saudi Arabian population using the GlobalFiler(R) PCR amplification kit. Forensic Science International: Genetics, 31, e59–e61. https://doi.org/10.1016/j.fsigen.2017.09.014
Blench, R., Sagart, L., & Sanchez-Mazas, A. (Eds.). (2005). The people of East Asia: Putting together archaeology, linguistics and genetics. London: Routledge.
Chen, P., Adnan, A., Rakha, A., Wang, M., Zou, X., Mo, X., & He, G. (2019). Population background exploration and genetic distribution analysis of Pakistan Hazara via 23 autosomal STRs. Annals of Human Biology, 46(6), 514–518. https://doi.org/10.1080/03014460.2019.1673483
Chen, P., Wu, J., Luo, L. I., Gao, H., Wang, M., Zou, X., … He, G. (2019). Population genetic analysis of modern and ancient DNA variations yields new insights into the formation, genetic structure, and phylogenetic relationship of Northern Han Chinese. Frontiers in Genetics, 10, 1045. https://doi.org/10.3389/fgene.2019.01045
Chen, P., Zou, X., Wang, B., Wang, M., & He, G. (2019). Genetic admixture history and forensic characteristics of Turkic-speaking Kyrgyz population via 23 autosomal STRs. Annals of Human Biology, 46(6), 498–501. https://doi.org/10.1080/03014460.2019.1676918
Chen, P., Zou, X., Wang, M., Gao, B., Su, Y., & He, G. (2019). Forensic features and genetic structure of the Hotan Uyghur inferred from 27 forensic markers. Annals of Human Biology, 46(7-8), 589–600. https://doi.org/10.1080/03014460.2019.1687751
Chiang, C. W. K., Mangul, S., Robles, C., & Sankararaman, S. (2018). A comprehensive map of genetic variation in the World’s Largest Ethnic Group—Han Chinese. Molecular Biology and Evolution, 35(11), 2736–2750. https://doi.org/10.1093/molbev/msy170
Choi, E.-J., Park, K.-W., Lee, Y.-H., Nam, Y.-H., Suren, G., Ganbold, U., … Kim, W. (2017). Forensic and population genetic analyses of the GlobalFiler STR loci in the Mongolian population. Genes & Genomics, 39(4), 423–431. https://doi.org/10.1007/s13258-016-0511-6
Cummings, M. P. (2004). PHYLIP (Phylogeny Inference Package). John Wiley & Sons Inc.
Ellegren, H. (2004). Microsatellites: Simple sequences with complex evolution. Nature Reviews Genetics, 5(6), 435–445. https://doi.org/10.1038/nrg1348
Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A
simulation study. *Molecular Ecology, 14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Excoffier, L., & Lischer, H. E. (2010). Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources, 10*(3), 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fujii, K., Iwashima, Y., Kitayama, T., Nakahara, H., Mizuno, N., & Sekiguchi, K. (2014). Allele frequencies for 22 autosomal short tandem repeat loci obtained by PowerPlex Fusion in a sample of 1501 individuals from the Japanese population. *Legal Medicine, 16*(4), 234–237. https://doi.org/10.1016/j.legalmed.2014.03.007

Gao, Y., Zhang, C., Yuan, L., Ling, Y. C., Wang, X., Liu, C., … Xu, S. (2019). PGG.Han: The Han Chinese genome database and analysis platform. *Nucleic Acids Research*. 48(D1), D971–D976. https://doi.org/10.1093/nar/gkz829

Gaviria, A., Zambrano, A. K., Morejon, G., Galarza, J., Aguirre, V., Vela, M., … Burgos, G. (2013). Twenty two autosomal microsatellite data from Ecuador (Powerplex Fusion). *Forensic Science International: Genetics Supplement Series, 4*(1), e330–e333. https://doi.org/10.1016/j.fsigen.2013.10.169

Gouy, A., & Zieger, M. (2017). STRAF-A convenient online tool for STR data evaluation in forensic genetics. *Forensic Science International: Genetics, 30*, 148–151. https://doi.org/10.1016/j.fsigen.2017.07.007

Guerreiro, S., Ribeiro, T., Porto, M. J., Carneiro de Sousa, M. J., & Dario, P. (2017). Characterization of GlobalFiler loci in Angolan and Guinean populations inhabiting Southern Portugal. *International Journal of Legal Medicine, 131*(2), 365–368. https://doi.org/10.1007/s00414-016-1497-y

Hagelberg, E., Gray, I. C., & Jeffreys, A. J. (1991). Identification of the skeletal remains of a murder victim by DNA analysis. *Nature*, 352(6334), 427–429. https://doi.org/10.1038/352427a0

Hares, D. R. (2015). Selection and implementation of expanded CODIS core loci in the United States. *Forensic Science International: Genetics, 17*, 33–34. https://doi.org/10.1016/j.fsigen.2015.03.006

He, G., Adnan, A., Rakha, A., Yeh, H.-Y., Wang, M., Zou, X., … Wang, C.-C. (2019). A comprehensive exploration of the genetic legacy and forensic features of Afghanistan and Pakistan Mongoloid-descent Hazara. *Forensic Science International: Genetics, 42*, e1–e12. https://doi.org/10.1016/j.fsigen.2019.06.018

He, G., Su, Y., Zou, X., Wang, M., Liu, J., Wang, S., … Wang, Z. (2019). Allele frequencies of 15 autosomal STRs in Chinese Nakh and Yi populations. *International Journal of Legal Medicine, 133*(1), 105–108. https://doi.org/10.1007/s00414-018-1931-4

He, G., Wang, M., Liu, J., Hou, Y., & Wang, Z. (2018). Forensic features and phylogenetic analyses of Sichuan Han population via 23 autosomal STR loci included in the Huaxia Platinum System. *International Journal of Legal Medicine, 132*(4), 1079–1082. https://doi.org/10.1007/s00414-017-1679-2

He, G., Wang, Z., Wang, M., Luo, T., Liu, J., Zhou, Y., … Hou, Y. (2018). Forensic ancestry analysis in two Chinese minority populations using massively parallel sequencing of 165 ancestry-informative SNPs. *Electrophoresis, 39*(21), 2732–2742. https://doi.org/10.1002/elps.201800019

He, G., Wang, Z., Wang, M., Zou, X., Liu, J., Wang, S., & Hou, Y. (2018). Genetic variations and forensic characteristics of Han Chinese population residing in the Pearl River Delta revealed by 23 autosomal STRs. *Molecular Biology Reports, 45*(5), 1125–1133. https://doi.org/10.1007/s11033-018-4264-y

Hossain, T., Hasan, M., Mazumder, A. K., Momtaz, P., Sufian, A., Khandaker, J. A., & Akhteruzzaman, S. (2016). Genetic polymorphism studies on 22 autosomal STR loci of the PowerPlex Fusion System in Bangladeshi population. *Legal Medicine, 23*, 44–46. https://doi.org/10.1016/j.legalmed.2016.09.005

Jin, X., Wei, Y., Chen, J., Kong, T., Mu, Y., Guo, Y., … Zhu, B. (2017). Phylogenetic analysis and forensic genetic characterization of Chinese Uyghur group via autosomal multi STR markers. *Oncotarget, 8*(43), 73837–73845. https://doi.org/10.18632/oncotarget.17992

Kayser, M., & de Knijff, P. (2011). Improving human forensics through advances in genetics, genomics and molecular biology. *Nature Reviews Genetics*, 12*(3), 179–192. https://doi.org/10.1038/nrg2952

Kovach, W. L. (2007). MVSP-A MultiVariate Statistical Package for Windows, ver. 3.1. Pentreath, UK: Kovach Computing Services.

Kumar, S., Stecher, G., & Tamura, K. (2016). MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution, 33*(7), 1870–1874. https://doi.org/10.1093/molbev/msw054

Lang, M., Liu, H., Song, F., Qiao, X., Ye, Y. I., Ren, H. E., … Hou, Y. (2019). Forensic characteristics and genetic analysis of both 27 Y-STRs and 143 Y-SNPs in Eastern Han Chinese population. *Forensic Science International: Genetics, 42*, e13–e20. https://doi.org/10.1016/j.fsigen.2019.07.011

Leipe, C., Long, T., Serghesheva, E. A., Wagner, M., & Tarasov, P. E. (2019). Discontinuous spread of millet agriculture in eastern Asia and prehistoric population dynamics. *Science Advances, 5*(9), eaax6225. https://doi.org/10.1126/sciadv.aax6225

Li, Y.-C., Tian, J.-Y., Liu, F.-W., Yang, B.-Y., Gu, K.-S.-Y., Rahman, Z. U., … Kong, Q.-P. (2019). Neolithic millet farmers contributed to the permanent settlement of the Tibetan Plateau by adopting barley agriculture. *National Science Review, 6*(5), 1005–1013. https://doi.org/10.1093/nsr/nzw080

Li, Y.-C., Ye, W.-J., Jiang, C.-G., Zeng, Z., Tian, J.-Y., Yang, L.-Q., … Kong, Q.-P. (2019). River valleys shaped the maternal genetic landscape of Han Chinese. *Molecular Biology and Evolution, 36*(8), 1643–1652. https://doi.org/10.1093/molbev/msz072

Liu, J., Wang, Z., He, G., Wang, M., & Hou, Y. (2019). Genetic polymorphism and phylogenetic differentiation of the Huaxia Platinum System in three Chinese minority ethnicities. *Scientific Reports, 9*(1), 3371. https://doi.org/10.1038/s41598-019-39794-y

Liu, S., Huang, S., Chen, F., Zhao, L., Yuan, Y., Francis, S. S., … Xu, X. (2018). Genomic analyses from non-invasive prenatal testing reveal genetic associations, patterns of viral infections, and Chinese population history. *Cell, 175*(2), 347–359.e14. https://doi.org/10.1016/j.cell.2018.08.016

Moyes, C. B., Tsutsumida, W. M., Raimann, P. E., da Motta, C. H., Nogueira, T. L., Dos Santos, O. C., … Gusmao, L. (2017). Population data of the 21 autosomal STRs included in the GlobalFiler(R) kits in population samples from five Brazilian regions. *Forensic Science International: Genetics, 26*, e28–e30. https://doi.org/10.1016/j.fsigen.2016.10.017

Nicogossian, A., Kloiber, O., & Stabile, B. (2014). The Revised World Protection of human research subjects and empowering ethics review committees. *World Medical & Health Policy, 6*(1), 1–3. https://doi.org/10.1002/wmh3.79

Ossowski, A., Diepenbroek, M., Szargut, M., Zielinska, G., Jedrzejczyk, M., Berent, J., & Jacewicz, R. (2017). Population analysis and forensic evaluation of 21 autosomal loci included in GlobalFiler PCR Kit
in Poland. Forensic Science International: Genetics, 29, e38–e39. https://doi.org/10.1016/j.fsigen.2017.05.003

Park, H. C., Kim, K., Nam, Y., Park, J., Lee, J., Lee, H., … Lim, S. (2016). Population genetic study for 24 STR loci and Y indel (GlobalFiler PCR Amplification kit and PowerPlex(R) Fusion system) in 1000 Korean individuals. Legal Medicine, 21, 53–57. https://doi.org/10.1016/j.legalmed.2016.06.003

Park, J. H., Hong, S. B., Kim, J. Y., Chong, Y., Han, S., Jeon, C. H., & Ahn, H. J. (2013). Genetic variation of 23 autosomal STR loci in Korean population. Forensic Science International: Genetics, 7(3), e76–e77. https://doi.org/10.1016/j.fsigen.2012.10.005

Sadam, M., Tasa, G., Tiidla, A., Lang, A., Axelsson, E. P., & Pajnic, I. Z. (2015). Population data for 22 autosomal STR loci from Estonia. International Journal of Legal Medicine, 129(6), 1219–1220. https://doi.org/10.1007/s00414-014-1089-7

Schneider, P. M. (2007). Scientific standards for studies in forensic genetics. Forensic Science International, 165(2–3), 238–243. https://doi.org/10.1016/j.forsciint.2006.06.067

Stoneking, M., & Delfin, F. (2010). The human genetic history of East Asia: Weaving a complex tapestry. Current Biology, 20(4), R188–R193. https://doi.org/10.1016/j.cub.2009.11.052

Taylor, D., Bright, J. A., McGovern, C., Neville, S., & Grover, D. (2017). Allele frequency database for GlobalFiler STR loci in Australian and New Zealand populations. Forensic Science International: Genetics, 28, e38–e40. https://doi.org/10.1016/j.fsigen.2017.02.012

Wang, M., Wang, Z., He, G., Jia, Z., Liu, J., & Hou, Y. (2018). Genetic characteristics and phylogenetic analysis of three Chinese ethnic groups using the Huaxia Platinum System. Scientific Reports, 8(1), 2429. https://doi.org/10.1038/s41598-018-20871-7

Wang, Z., Zhou, D., Jia, Z., Li, L., Wu, W., Li, C., & Hou, Y. (2016a). Developmental validation of the Huaxia Platinum System and application in 3 main ethnic groups of China. Scientific Reports, 6, 31075. https://doi.org/10.1038/srep31075

Wu, L., Pei, B., Ran, P., & Song, X. (2017). Population genetic analysis of Xiamen Han population on 21 short tandem repeat loci. Legal Medicine, 26, 41–44. https://doi.org/10.1016/j.legalmed.2017.03.002

Xu, S., Huang, W., Qian, J., & Jin, L. (2008). Analysis of genomic admixture in Uygur and its implication in mapping strategy. American Journal of Human Genetics, 82(4), 883–894. https://doi.org/10.1016/j.ajhg.2008.01.017

Yang, L., Zhang, X., Zhao, L., Sun, Y., Li, J., Huang, R., … Nie, S. (2018). Population data of 23 autosomal STR loci in the Chinese Han population from Guangdong Province in southern China. International Journal of Legal Medicine, 132(1), 133–135. https://doi.org/10.1007/s00414-017-1588-4

Zhang, H., Xia, M., Qi, L., Dong, L., Song, S., Ma, T., … Li, S. (2016). Forensic and population genetic analysis of Xinjiang Uyghur population on 21 short tandem repeat loci of 6-dye GlobalFiler PCR Amplification kit. Forensic Science International: Genetics, 22, 22–24. https://doi.org/10.1016/j.fsigen.2016.01.005

Zhang, H., Yang, S., Guo, W., Ren, B. O., Pu, L., Ma, T., … Li, S. (2016). Population genetic analysis of the GlobalFiler STR loci in 748 individuals from the Kazakh population of Xinjiang in northwest China. International Journal of Legal Medicine, 130(5), 1187–1189. https://doi.org/10.1007/s00414-016-1319-2

Zhang, M., Yan, S., Pan, W., & Jin, L. (2019). Phylogenetic evidence for Sino-Tibetan origin in Northern China in the Late Neolithic. Nature, 569(7754), 112–115. https://doi.org/10.1038/s41586-019-1153-z

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Li L, Zou X, Zhang G, et al. Population genetic analysis of Shaanxi male Han Chinese population reveals genetic differentiation and homogenization of East Asians. Mol Genet Genomic Med. 2020;8:e1209. https://doi.org/10.1002/mgg3.1209