*Forensic Sciences Research*, 2024, Vol. 9, Issue 1
https://doi.org/10.1093/fsr/owad047
Advance access publication date 21 December 2023
**Research Article**

# Systematical explorations of forensic feature and population genetic diversity of the Chinese Mongolian group from northwest China *via* a self-constructed Multi-InDel panel

**Xuebing Chen[1], Hui Xu[1], Wei Cui[1], Ming Zhao[1], Bofeng Zhu[1,2,3,*]**

[1]Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou, China
[2]Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an, China
[3]Clinical Research Center of Shaanxi Province for Dental and Maxillofacial Diseases, College of Stomatology, Xi'an, China

*Corresponding author. E-mail: zhubofeng7372@126.com

## Abstract

This study aimed to investigate the genetic polymorphisms and population characteristics of Chinese Mongolian group from northwest China (NCM) through a self-developed panel including 43 autosomal insertion/deletion (A-InDel) polymorphism genetic markers. Herein, 288 unrelated healthy individuals from the NCM group were employed to obtain the genetic data of 43 A-InDels through multiplex PCR amplification and InDel genotyping using capillary electrophoresis platform. In addition, multiplex population genetic analyses were performed between the NCM group and 27 reference populations. There were no deviations at 43 loci from Hardy–Weinberg equilibrium in the NCM group. The observed heterozygosity (Ho) values ranged from 0.312 8 to 0.559 2, and the combined power of discrimination (CPD) and cumulative probability of exclusion (CPE) values in the NCM group were 0.999 999 999 999 999 998 77 and 0.999 814, respectively. The forensic parameter values indicated that this panel was polymorphic and informative in the NCM group and could be used as an effective tool for forensic personal identification. Furthermore, the results of pairwise genetic distances, principal component analysis, multidimensional scaling analysis, phylogenetic tree construction, and admixture analysis among the NCM group and 27 reference populations revealed that there were closer genetic relationships between the NCM group and East Asian populations, especially Chinese Hui group (CHH) from the northwest China, which is consistent with the geographical location. These present findings contributed to the ongoing genetic explorations and insights into the genetic architecture of the NCM group.

**Keywords:** forensic sciences; insertion/deletion; Chinese Mongolian group; individual identification; degraded sample; genetic polymorphism

## Introduction

The insertion/deletion (InDel) polymorphic genetic marker, which is widely distributed in the human genome, plays a crucial role in forensic and population genetic study. Since the initial research was reported by Weber et al. [1], InDel markers have generated significant attentions from forensic geneticists. InDel markers have several advantages over traditional short tandem repeat (STR) loci, making them the mutual complementation tool to the commonly used STRs nowadays [2, 3]. Recently, some studies have confirmed the indispensable role of InDel genetic markers in forensic genetics, particularly in population genetics and personal identification in degraded sample [4–6]. Personal identification is a crucial aspect of forensic practice, especially in that case involving degraded sample. However, commonly used STR genotyping of highly degraded sample may fail to provide a complete genotype profile due to the loss of large sizes of amplicons [7, 8]. Previously, this novel genotyping system consisting of 43 A-InDels and an Amelogenin gene locus, was found to be suitable for personal identification in Chinese Hui group (CHH) [9]. Besides, the amplicon sizes of this self-developed multiplex PCR panel, which includes 43 A-InDels, are less than 200 bp, making it ideal tool for obtaining complete genotyping profile of degraded sample. Nevertheless, this amplification system is still in its early stage in forensic genetics, and further population data are required before its widespread adoption.

The Mongolian group is an ancient nomadic people with unique language, characters and customs. According to the 2020 census, Chinese Mongolian group reached 6.29 million, which is one of the most populous ethnic groups in China (http://www.stats.gov.cn/tjsj/ndsj/2021/indexch.htm). Mongolian, which belongs to the Mongolian branch of the Altaic language family, is the national official language of Mongolia. Mongolians are renowned for their Mongol Empire, which was established by Genghis Khan in the 13th century. With the establishment of the Mongolian Empire and

**Received:** October 15, 2022. **Accepted:** October 16, 2023

the continuous expansion of its territory, it promoted the gene exchanges between Mongolian group and other nationalities, especially reflected in the gene structure of Eurasia continent [10, 11]. The polymorphism analyses of diverse genetic markers such as STR [12], single nucleotide polymorphism (SNP) [13] and mitochondrial DNA (mtDNA) [14] also largely affirm the intimate genetic associations between the Mongolians and the East Asians. An in-depth exploration of the genetic background and structure of the Mongolian group from northwest China (NCM) may help to better understand the background of Mongol group and their adjacent groups.

In the current study, genetic distributions and forensic efficiencies of all the A-InDels using the self-developed panel in the NCM group were further investigated. Moreover, *Nei's* $D_A$ distances, pairwise $F_{ST}$ values, principal component analysis (PCA), multidimensional scaling (MDS) analysis, phylogenetic tree construction, and genetic structure analysis based on the same 43 A-InDels were used to reveal the genetic differentiations and relationships among the NCM group and 27 worldwide comparison populations.

## Materials and methods

### Sample collections and comparison populations

Bloodstain samples were collected from 288 unrelated healthy donors from the Mongolian group residing in northwest China. Prior to this experiment, all donors gave written informed consents. As for the 3 626 individuals from 27 reference populations, 26 of them were from the Project Phase 3 database of the 1 000 Genomes Project and the remaining population was from our previously published research [9]. In addition, we combined the two sets of genotype data of Han Chinese in Beijing (CHB) group from the 1 000 Genomes Project Phase 3 and Beijing Han from our previously published work [15] to create a combined dataset of 404 Beijing Han individuals, which is named as CHB in the present study. Moreover, detailed information about the 28 populations is presented in Supplementary Table S1.

### PCR amplification and subsequent InDels genotyping

Genomic DNA was extracted by the Chelex-100 method. The GeneAmp PCR System 9700 Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA) was used to amplify the 43 A-InDels following the PCR protocol outlined in our previous study [9]. Subsequently, PCR amplification products were separated and detected through the ABI 3500xL Genetic Analyzer (Thermo Fisher Scientific). InDels genotyping was performed using GeneMapper ID-X software v1.5 (Thermo Fisher Scientific). The positive and negative controls during the experimental procedures were DNA 9947A, 9948, and deionized water, respectively.

### Forensic statistical analysis

The Hardy–Weinberg equilibrium (HWE) analyses of the 43 A-InDels in the NCM group were conducted by Genepop software v4.7.5 [16], while linkage disequilibrium (LD) analyses for pairwise InDels were performed using SNPAnalyzer v2.0 (Istech, Republic of Korea) software [17]. The allele frequencies of 43 A-InDels and values of forensic parameters containing the polymorphism information content (PIC), typical paternity index (TPI), match probability (MP), probability of exclusion (PE), expected heterozygosity (He), and observed heterozygosity (Ho) in the NCM group were calculated by the STRAF online program v1.0.5 [18]. Meanwhile, the cumulative power of discrimination (CPD) and combined probability of exclusion (CPE) values of the 43 A-InDels in the NCM group were carried out by the corresponding formula in the Excel. Moreover, locus-by-locus analysis of molecular variance (AMOVA) was performed on 28 populations, and the calculation of pairwise fixation index ($F_{ST}$) values was executed by Arlequin software v3.5 [19]. In addition, the $D_A$ distances between pairs of populations were calculated by the DISPAN program [20].

Phylogenetic trees were visualized and managed using the ITOL online tool (https://itol.embl.de). Heatmaps of the InDel allele frequencies, pairwise $D_A$ distances, and $F_{ST}$ values were drawn using Origin v2019 following the previous study [21]. Meanwhile, PCA plots, including the population-level map based on allele frequencies, and individual-level map derived from the 43 A-InDels genotyping data, along with MDS plot based on pairwise $F_{ST}$ values among the 28 populations were also drawn by *R* software v3.6.2. The genetic structure analysis was performed using the software STRUCTURE v2.3.4 [22]. An estimate of the optimum *K* value was obtained using the online tool STRUCTURE HARVESTER [23].

## Results

### HWE and LD analyses

In the present study, the LD tests were performed on pairwise loci from 43 A-InDels to evaluate the independence of every locus in the NCM group (Supplementary Figure S1). Pairwise $r^2$ values were less than 0.25 (Supplementary Table S2) in the NCM group, meaning that these loci could be used as independent genetic markers. Meantime, all A-InDel loci were found to comply with HWE by Bonferroni's correction ($P = 0.05/43 = 0.0012$), and the lowest *P* value of the HWE was 0.0322 (rs63064161) in the NCM group (Table 1).

### The allele frequency distributions and forensic parameter evaluations

The allele frequencies and corresponding forensic parameters of the 43 A-InDels in the NCM group are listed in Table 1. Results showed that the insertion allelic frequencies of 43 A-InDel loci ranged from 0.3611 (rs201844336) to 0.6319 (rs5821525) in the NCM group, the 30 loci of which ranged from 0.4 to 0.6. Besides, the values of PIC, He, and Ho in the NCM group ranged from 0.3550 (rs201844336) to 0.3750 (rs140025863), 0.4622 (rs201844336) to 0.5009 (rs140025863), and 0.4306 (rs5821525) to 0.5556 (rs63064161), respectively. The values of PE in the NCM group ranged from 0.1336 (rs5821525) to 0.2409 (rs63064161). Moreover, the CPD and CPE values of 43 A-InDels in the NCM group were 0.999 999 999 999 999 998 77 and 0.999 814, respectively. The above results indicated that this panel was informative and polymorphic and could be usedfor individual identification, as one of the effective complementary tools for the NCM group in forensic paternity testing. In addition, this studied panel including 43 A-InDels had higher CPE and CPD values in the NCM group when compared to the previously reported InDel systems (Supplementary Table S3), such as the 21-plex InDels [24], 30 InDels [25], 32 InDels [26], and 35 InDels [27] panels.

**Table 1.** Allelic frequencies and forensic parameters of 43 A-InDel loci in the Mongolian group from northwest China (NCM, $n = 288$)

| Loci | AF+ | PIC | MP | PE | TPI | He | Ho | *P*-HWE |
|---|---|---|---|---|---|---|---|---|
| rs142281120 | 0.4184 | 0.3683 | 0.3833 | 0.1785 | 0.9796 | 0.4875 | 0.4896 | 1.0000 |
| rs146880183 | 0.3628 | 0.3555 | 0.3883 | 0.1409 | 0.8944 | 0.4632 | 0.4410 | 0.4634 |
| rs3036240 | 0.3785 | 0.3598 | 0.4011 | 0.1815 | 0.9863 | 0.4713 | 0.4931 | 0.4585 |
| rs142159306 | 0.4653 | 0.3738 | 0.3740 | 0.1815 | 0.9863 | 0.4985 | 0.4931 | 0.9055 |
| rs35700881 | 0.4740 | 0.3743 | 0.3594 | 0.1562 | 0.9290 | 0.4995 | 0.4618 | 0.2316 |
| rs3830564 | 0.4115 | 0.3670 | 0.3738 | 0.1562 | 0.9290 | 0.4852 | 0.4618 | 0.4587 |
| rs3092307 | 0.4792 | 0.3746 | 0.3831 | 0.1999 | 1.0286 | 0.5000 | 0.5139 | 0.7325 |
| rs79287422 | 0.5538 | 0.3721 | 0.4109 | 0.2373 | 1.1163 | 0.4951 | 0.5521 | 0.0608 |
| rs5852131 | 0.3889 | 0.3623 | 0.3997 | 0.1875 | 1.0000 | 0.4761 | 0.5000 | 0.4485 |
| rs6144473 | 0.4201 | 0.3685 | 0.3844 | 0.1815 | 0.9863 | 0.4881 | 0.4931 | 0.9041 |
| rs33990282 | 0.6059 | 0.3635 | 0.3755 | 0.1458 | 0.9057 | 0.4784 | 0.4479 | 0.3230 |
| rs10537321 | 0.4913 | 0.3749 | 0.3882 | 0.2097 | 1.0511 | 0.5007 | 0.5243 | 0.4799 |
| rs201844336 | 0.3611 | 0.3550 | 0.4171 | 0.1936 | 1.0141 | 0.4622 | 0.5069 | 0.1220 |
| rs4019986 | 0.3924 | 0.3631 | 0.3826 | 0.1589 | 0.9351 | 0.4777 | 0.4653 | 0.7071 |
| rs5825145 | 0.4479 | 0.3723 | 0.3955 | 0.2130 | 1.0588 | 0.4954 | 0.5278 | 0.2802 |
| rs10541072 | 0.5139 | 0.3748 | 0.3826 | 0.1999 | 1.0286 | 0.5005 | 0.5139 | 0.7232 |
| rs10533337 | 0.6076 | 0.3631 | 0.3884 | 0.1699 | 0.9600 | 0.4777 | 0.4792 | 1.0000 |
| rs67941259 | 0.5694 | 0.3701 | 0.3846 | 0.1875 | 1.0000 | 0.4912 | 0.5000 | 0.8063 |
| rs10589141 | 0.4167 | 0.3680 | 0.3707 | 0.1536 | 0.9231 | 0.4870 | 0.4583 | 0.3346 |
| rs3993057 | 0.4063 | 0.3661 | 0.3892 | 0.1815 | 0.9863 | 0.4833 | 0.4931 | 0.8031 |
| rs5882232 | 0.3906 | 0.3628 | 0.4081 | 0.2031 | 1.0360 | 0.4769 | 0.5174 | 0.1724 |
| rs10555434 | 0.4896 | 0.3749 | 0.3521 | 0.1433 | 0.9000 | 0.5007 | 0.4444 | 0.0584 |
| rs63136060 | 0.5347 | 0.3738 | 0.3708 | 0.1756 | 0.9730 | 0.4985 | 0.4861 | 0.7213 |
| rs142623177 | 0.4722 | 0.3742 | 0.3638 | 0.1643 | 0.9474 | 0.4993 | 0.4722 | 0.4083 |
| rs55714089 | 0.3872 | 0.3619 | 0.3810 | 0.1510 | 0.9172 | 0.4754 | 0.4549 | 0.5263 |
| rs63064161 | 0.5625 | 0.3711 | 0.4152 | 0.2409 | 1.1250 | 0.4930 | 0.5556 | 0.0322 |
| rs35974596 | 0.5590 | 0.3715 | 0.3892 | 0.1999 | 1.0286 | 0.4939 | 0.5139 | 0.5578 |
| rs10540867 | 0.5208 | 0.3746 | 0.3950 | 0.2197 | 1.0746 | 0.5000 | 0.5347 | 0.2887 |
| rs140025863 | 0.4983 | 0.3750 | 0.3804 | 0.1968 | 1.0213 | 0.5009 | 0.5104 | 0.8178 |
| rs5822909 | 0.4670 | 0.3739 | 0.3721 | 0.1785 | 0.9796 | 0.4987 | 0.4896 | 0.8074 |
| rs10588341 | 0.4236 | 0.3691 | 0.3684 | 0.1536 | 0.9231 | 0.4892 | 0.4583 | 0.3235 |
| rs10573809 | 0.6076 | 0.3631 | 0.3799 | 0.1536 | 0.9231 | 0.4777 | 0.4583 | 0.5289 |
| rs16646 | 0.4913 | 0.3749 | 0.3769 | 0.1906 | 1.0070 | 0.5007 | 0.5035 | 1.0000 |
| rs147682692 | 0.3906 | 0.3628 | 0.3939 | 0.1785 | 0.9796 | 0.4769 | 0.4896 | 0.7054 |
| rs3830885 | 0.4809 | 0.3746 | 0.3775 | 0.1906 | 1.0070 | 0.5001 | 0.5035 | 1.0000 |
| rs10544053 | 0.3802 | 0.3602 | 0.3895 | 0.1616 | 0.9412 | 0.4721 | 0.4688 | 0.8993 |
| rs10555133 | 0.5226 | 0.3745 | 0.3678 | 0.1728 | 0.9664 | 0.4998 | 0.4826 | 0.6310 |
| rs142392113 | 0.5191 | 0.3746 | 0.3675 | 0.1728 | 0.9664 | 0.5001 | 0.4826 | 0.5541 |
| rs144537609 | 0.5104 | 0.3749 | 0.3686 | 0.1756 | 0.9730 | 0.5007 | 0.4861 | 0.6378 |
| rs5821525 | 0.6319 | 0.3570 | 0.3823 | 0.1336 | 0.8780 | 0.4660 | 0.4306 | 0.1993 |
| rs10584875 | 0.5451 | 0.3730 | 0.3941 | 0.2130 | 1.0588 | 0.4968 | 0.5278 | 0.3557 |
| rs5892949 | 0.4635 | 0.3737 | 0.3868 | 0.2031 | 1.0360 | 0.4982 | 0.5174 | 0.5581 |
| rs3043804 | 0.4497 | 0.3725 | 0.3784 | 0.1845 | 0.9931 | 0.4958 | 0.4965 | 1.0000 |

AF+: insertion allelic frequency; PIC: polymorphism information content; MP: match probability; PE: probability of exclusion; TPI: typical paternity index; Ho: observed heterozygosity; He: expected heterozygosity; *P*-HWE: *P* value of Hardy–Weinberg equilibrium test.

As shown in Supplementary Figure S2, results from the heatmap based on insertion allelic frequencies of 43 A-InDel loci showed different insertion allelic frequency distributions among 28 populations. In addition, all 28 populations were divided into two branches: one branch consisted of seven African populations clustering together, while the other populations gathered together in another branch. The similar distributions of allelic frequencies in those populations tend to gather together, even if they do not come from the same continent. In the East Asian populations, almost all the 43 A-InDel loci displayed middle insertion allelic frequencies. In addition, insertion allelic frequencies in the NCM group were between 0.36 and 0.63. Based on the allelic frequency distributions, the NCM group and East Asian subpopulations clustered closely. All values of insertion allelic frequencies at 43 A-InDels in the 28 populations are shown in Supplementary Table S4.

To better understand the genetic differentiations among the NCM group and 27 reference populations based on the 43 A-InDels, the locus-by-locus *P* values of AMOVA are listed in Supplementary Table S5. The NCM group had significant differences with East Asian populations on one locus (Japanese in Tokyo, JPT), three (CHH), four (Kinh in Ho Chi Minh City, Vietnam, KHV; and Southern Han Chinese, CHS), and five (Han Chinese in Beijing, CHB; and Chinese Dai in Xishuangbanna, CDX) loci, respectively. In addition, the NCM group had significant differences with European populations on 15 (Finnish in Finland, FIN), 18 (Toscani in Italy, TSI; Iberian populations in Spain, IBS; and Utah residents with Northern and Western European ancestry, CEU), and 19 (British in England and Scotland, GBR); with African populations on 17 (African Ancestry in Southwest US, ASW), 21 (Luhya in Webuye, Kenya, LWK), 23 (Gambian in Western Division, The Gambia, GWD), 24 (Esan in Nigeria,

ESN; and African Caribbean in Barbados, ACB), 25 (Mende in Sierra Leone, MSL), and 28 (Yoruba in Ibadan, Nigeria, YRI) loci, respectively.

### $F_{ST}$ and $D_A$ distance analyses

The values of pairwise $F_{ST}$ and $D_A$ distances among 28 populations are shown in Supplementary Tables S6 and S7. In comparisons with other populations, the NCM group exhibited the lowest $F_{ST}$ value with CHH (0.0039), followed by JPT (0.0047) and CHB (0.0066). Conversely, the highest $F_{ST}$ value was found between NCM and YRI (0.0903), followed by the ESN (0.0872) and MSL (0.0856). Meanwhile, the results of $D_A$ distances among the NCM group and 27 comparison populations and $F_{ST}$ values showed similar trends. The closest $D_A$ distance was observed between the NCM group and CHH (0.0013), followed by the JPT (0.0020) and CHB (0.0025); while the greatest $D_A$ distance was between the NCM group and YRI (0.0277), followed by the ESN (0.0269) and MSL (0.0269). Furthermore, two heatmaps were intuitively displayed through different colours, one was based on the values of pairwise $F_{ST}$ (Supplementary Figure S3A) and the other was on $D_A$ distances (Supplement Figure S3B). As displayed in the heatmaps, the NCM group exhibited greater $F_{ST}$ values and $D_A$ distances with populations from Africa, while the NCM group displayed smaller $F_{ST}$ values and $D_A$ distances with populations from East Asia.

### PCA and MDS analyses

PCA and MDS analyses were also performed to explore the genetic relationships among the NCM group and 27 comparison populations. Moreover, the results of allele frequency-based PCA and pairwise $F_{ST}$-based MDS are exhibited in Figure 1A and B, respectively, while results of genotype-based PCA are presented in Figure 1C and D. The first two principal components, PC1 and PC2, cumulatively contributed 53.64% of the total variation with PC1 and PC2 contributing 34.07% and 19.57% at the population level, respectively, as depicted in Figure 1A. Furthermore, the result of the MDS analysis based on pairwise $F_{ST}$ values is exhibited in Figure 1B. The similar population distribution pattern was also found in the MDS plot, further confirming the close relationships between the NCM group and East Asian populations involved in the present research. At the individual level, 9.85% of the total variation could be attributed to the first two principal components (PC1, 6.67%; PC2, 3.18%), as shown in Figure 1C. Subsequently, we conducted the PCA analysis among the populations from three continents (Africa, Europe, and East Asia) and the NCM group. Compared with the results shown in Figure 1D, more obvious boundaries were observed among the African, European, and East Asian populations. Populations from the same continent were displayed with the same colour. There were most overlapping dots between the NCM and the East Asian populations, suggesting the close relationships between the NCM group and East Asian subpopulations.

### Population genetic structure analyses and phylogenetic relationship reconstructions

In this study, we performed a STRUCTURE analysis of 28 populations based on the genotypic data of the 43 A-InDels. The result of online Harvest program analysis showed that the optimal $K$ value was 3 (Supplementary Figure S4). Figure 2 and Supplementary Figure S5 showed the STRUCTURE results intuitively among the 28 populations when $K = 2$–4. Noteworthy, the genetic structure of the NCM group was similar to those of East Asian populations, but differed from those of other continental populations when $K = 2$–4 at the individual level.

On the basis of the phylogenetic tree, we were able to determine the genetic relationships among the 28 populations. As displayed in Figure 3, there were five major clusters in the phylogenetic tree. Seven African populations clustered, which were marked in green, seven East Asian populations clustered, which were marked in red, five European populations clustered, which were marked in yellow, and five South Asian populations clustered and marked in blue. Moreover, the NCM group was situated in the East Asian cluster branch, particularly in proximity to the JPT and CHH groups.

## Discussion

In the present study, we evaluated the genetic features of 43 A-InDels in the NCM group to determine the potential forensic application of this multiplex PCR system for purposes of individual identification and paternity testing. The present results showed that there were no A-InDels deviated from HWE. Furthermore, all pairwise A-InDels were confirmed to linkage equilibrium, indicating that all loci were independent and could be adapted to the following population genetic andforensic application analyses. In a population, heterozygosity of a genetic marker is the proportion of heterozygotes among all genotypes. The high degree of heterozygosity suggests that the genetic marker holds great application value in forensic personal identification. In the NCM group, all 43 A-InDels showed heterozygosity values above 0.4, and 17 loci of them were greater than 0.5. As compared with the previous InDels systems used in Chinese Mongolian group from different regions (Supplementary Table S3), the studied system including 43 A-InDels had higher CPE and CPD values in the NCM group, which indicated that this studied panel improved personal identification ability and could be as a supplement tool for forensic paternity testing in the NCM group.

We then compared the NCM group with 27 reference populations based on the 43 A-InDels to gain more in-depth understanding of their genetic relationships. The insertion allele frequencies of 43 A-InDels were similar in the same intercontinental populations, with the exception of the American populations. The results of AMOVA, pairwise $F_{ST}$ values, and $D_A$ distances all indicated that the NCM group had the greatest genetic differentiations with African populations and the smallest differentiations with the East Asian populations, specifically with the CHH group. Throughout the history of China, the Hui and Mongolian groups have a long history of frequent and close social exchanges [28]. The Mongolian expeditions to the west and south promoted the migration and integration of ethnic groups. According to relevant historical records and published studies, a large number of Muslims from West and Central Asia, such as Turks, Persia, and Arabia, came to live and multiply in China, known as “Huihui”, which might lead to close genetic relationships between Mongolia and Hui groups [29–31]. It is worth noting that there are only 27 reference populations, of which only six are from East Asian populations, which are not representative of the whole East Asian populations. In the future, more populations need

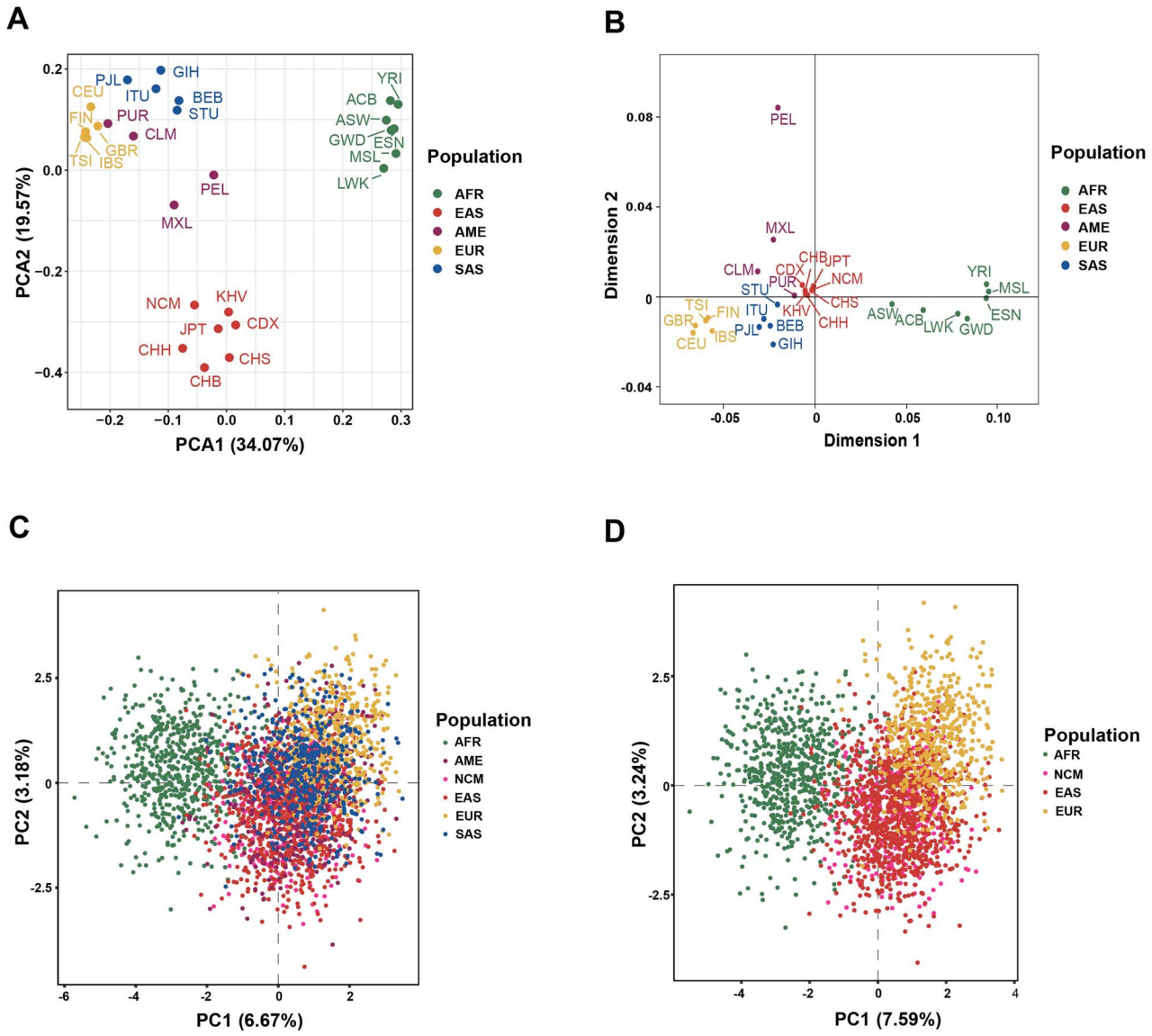


**Figure 1** The principal component analysis (PCA) and multidimensional scaling (MDS) for the studied NCM group and 27 reference populations. (A) PCA plot based on the allele frequency values and (B) MDS based on the pairwise $F_{ST}$ values among the NCM group and 27 reference populations at the population level. (C) PCA plot from 28 populations at the individual level. (D) PCA plot among the NCM group and populations from three continents (East Asia, Europe, and Africa) at the individual level. NCM: Mongolian group from northwest China; CHH: Chinese Hui group, China; CDX: Chinese Dai in Xishuangbanna, China; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese, China; KHV: Kinh in Ho Chi Minh City, Vietnam; JPT: Japanese in Tokyo, Japan; CEU: Utah residents with Northern and Western European ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian populations in Spain; TSI: Toscani in Italy; CLM: Colombian in Medellin, Colombia; MXL: Mexican Ancestry in Los Angeles, CA; PEL: Peruvian in Lima, Peru; PUR: Puerto Rican in Puerto Rico; PJL: Punjabi in Lahore, Pakistan; GIH: Gujarati Indian in Houston, TX; ITU: Indian Telugu in the UK; STU: Sri Lankan Tamil in the UK; BEB: Bengali in Bangladesh; ACB: African Caribbean in Barbados; ASW: African Ancestry in Southwest US; ESN: Esan in Nigeria; GWD: Gambian in Western Division, The Gambia; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria.

to be included, especially the East Asian sub-populations, to investigate their genetic relationships with the NCM group in detail.

The outcomes of phylogenetic relationship reconstruction showed that the NCM group closely clustered with East Asian populations, which was consistent with the results obtained from PCA and MDS analyses. The analysis of population genetic structure revealed that the proportions of ancestral compositions in the NCM group were similar to those in East Asian populations. Previously, several researchers have investigated the genetic polymorphisms of Mongolians in China based on different InDel panels. Huang et al. [26] found that the Mongolian population based on the 32-plex InDels panel showed mixed ancestral components related to East Asian and European populations, and Zhang et al. [27] reported that Chinese Mongolian group may have similar genetic structures and more closely related genetic relationships to the East Asian populations, which are consistent with our present study. In addition, the polymorphism analyses of other diverse genetic markers were also used to study the genetic characteristics of the Mongolian group. For example, the results of 22 A-STR [12], 23 Y-STR [32], and 60 mtDNA [14] loci primarily confirmed the most intimate genetic relationships between the Mongolian group and the East Asian populations as compared to Mongolian group and non-East Asian

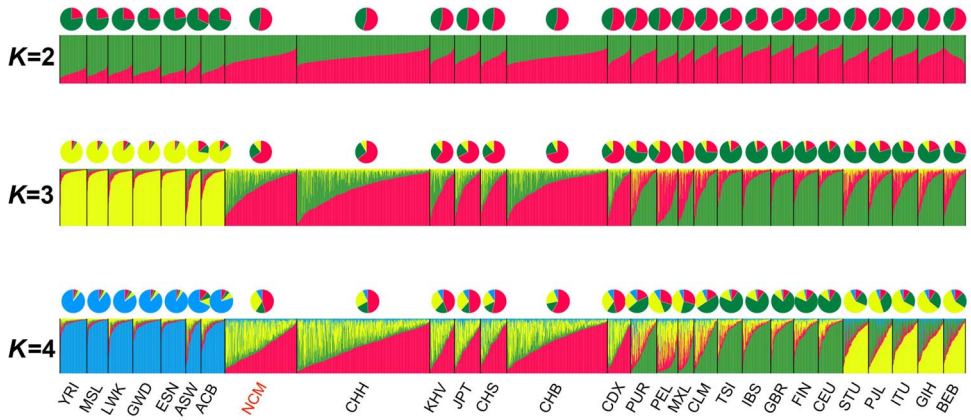


**Figure 2** The population genetic structure analysis among the NCM group and 27 reference populations when $K$ = 2–4. NCM: Mongolian group from northwest China; CHH: Chinese Hui group, China; CDX: Chinese Dai in Xishuangbanna, China; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese, China; KHV: Kinh in Ho Chi Minh City, Vietnam; JPT: Japanese in Tokyo, Japan; CEU: Utah residents with Northern and Western European ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian populations in Spain; TSI: Toscani in Italy; CLM: Colombian in Medellin, Colombia; MXL: Mexican Ancestry in Los Angeles, CA; PEL: Peruvian in Lima, Peru; PUR: Puerto Rican in Puerto Rico; PJL: Punjabi in Lahore, Pakistan; GIH: Gujarati Indian in Houston, TX; ITU: Indian Telugu in the UK; STU: Sri Lankan Tamil in the UK; BEB: Bengali in Bangladesh; ACB: African Caribbean in Barbados; ASW: African Ancestry in Southwest US; ESN: Esan in Nigeria; GWD: Gambian in Western Division, The Gambia; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria.

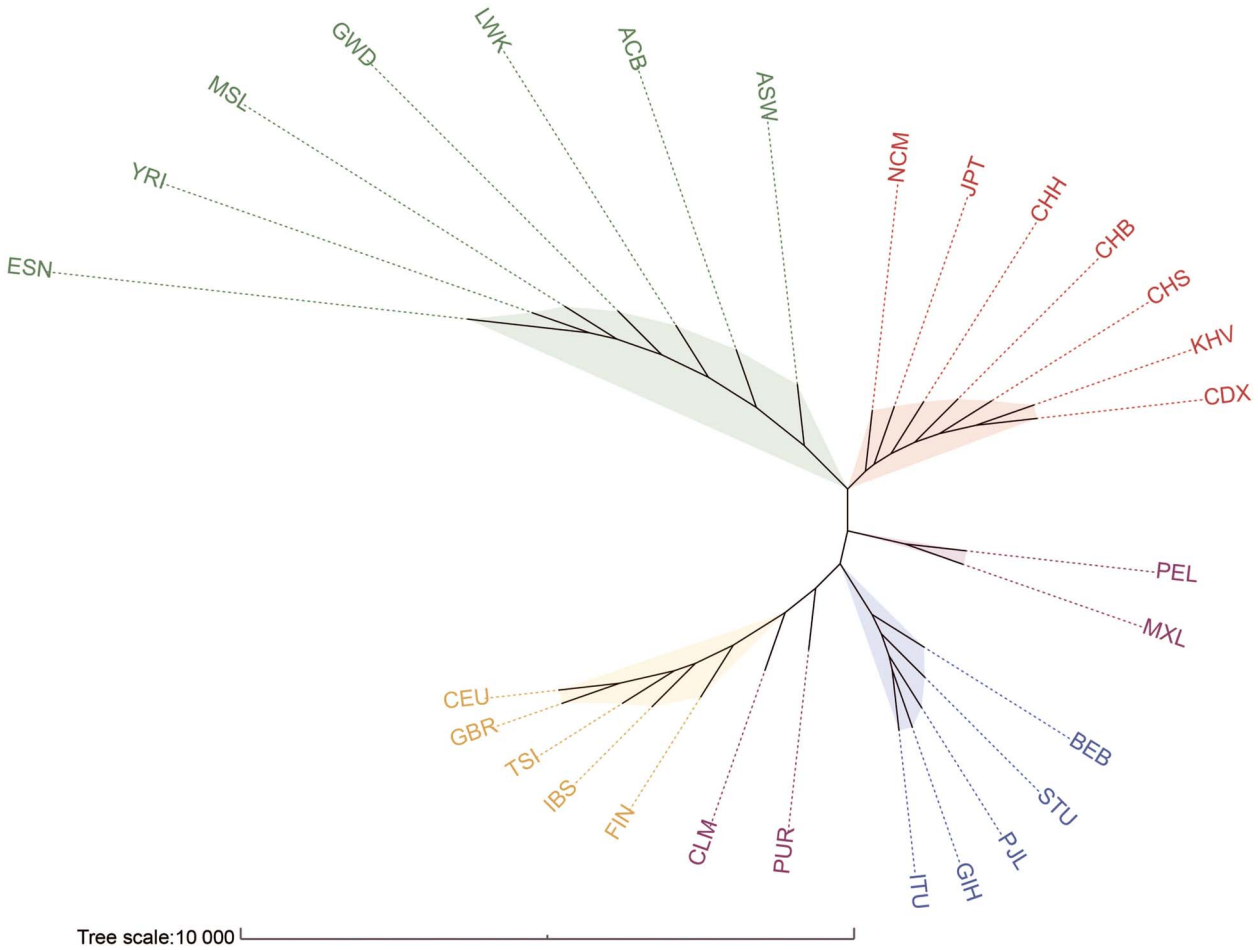


**Figure 3** The phylogenetic tree reconstruction based on the pairwise $D_A$ distances among the NCM group and 27 reference populations. NCM: Mongolian group from northwest China; CHH: Chinese Hui group, China; CDX: Chinese Dai in Xishuangbanna, China; CHB: Han Chinese in Beijing, China; CHS: Southern Han Chinese, China; KHV: Kinh in Ho Chi Minh City, Vietnam; JPT: Japanese in Tokyo, Japan; CEU: Utah residents with Northern and Western European ancestry; FIN: Finnish in Finland; GBR: British in England and Scotland; IBS: Iberian populations in Spain; TSI: Toscani in Italy; CLM: Colombian in Medellin, Colombia; MXL: Mexican Ancestry in Los Angeles, CA; PEL: Peruvian in Lima, Peru; PUR: Puerto Rican in Puerto Rico; PJL: Punjabi in Lahore, Pakistan; GIH: Gujarati Indian in Houston, TX; ITU: Indian Telugu in the UK; STU: Sri Lankan Tamil in the UK; BEB: Bengali in Bangladesh; ACB: African Caribbean in Barbados; ASW: African Ancestry in Southwest US; ESN: Esan in Nigeria; GWD: Gambian in Western Division, The Gambia; LWK: Luhya in Webuye, Kenya; MSL: Mende in Sierra Leone; YRI: Yoruba in Ibadan, Nigeria.

populations, which further bolsters the reliability of our research findings.

Herein, we utilized the self-developed panel to conduct a thorough assessment of the forensic effectiveness of this multiplex PCR amplification system in the NCM group and investigate the genetic relationships between the NCM group and 27 comparison populations. The results of forensic parameters based on 43 A-InDels demonstrated that this in-house panel holds great potential as a reliable tool for individual identification. The evaluation of the genetic relationship showed the NCM group was a close relationship with the CHH group. In summary, this study will not only provide a robust foundation for the application of InDels in forensic genetics, but also enrich the resources of the InDel database and promote more comprehensive understanding of the genetic architecture of the NCM group.

## Acknowledgements

The authors want to thank the volunteers in this research.

## Compliance with ethical standards

This study was performed in accordance with the principles of the Declaration of Helsinki and approved by the Ethics Committees of Southern Medical University, Guangzhou, China and Xi'an Jiaotong University, Xi'an, China (No. XJTULAC201). Written informed consent was obtained from all the participants.

## Disclosure statement

The authors declare that they have no conflict of interest.

## Funding

This work was supported by the National Natural Science Foundation of China [Grant No. 81772031].

## References

1. Weber JL, David D, Heil J, et al. Human diallelic insertion/deletion polymorphisms. Am J Hum Genet. 2002;71:854–862.
2. Wei YL, Qin CJ, Dong H, et al. A validation study of a multiplex INDEL assay for forensic use in four Chinese populations. Forensic Sci Int Genet. 2014;9:e22–e25.
3. Zhu B, Lan Q, Guo Y, et al. Population genetic diversity and clustering analysis for Chinese Dongxiang group with 30 autosomal InDel loci simultaneously analyzed. Front Genet. 2018;9:279.
4. Oldoni F, Castella V, Grosjean F, et al. Sensitive DIP-STR markers for the analysis of unbalanced mixtures from "touch" DNA samples. Forensic Sci Int Genet. 2017;28:111–117.
5. Guo Y, Shen C, Meng H, et al. Population differentiations and phylogenetic analysis of Tibet and Qinghai Tibetan groups based on 30 InDel loci. DNA Cell Biol. 2016;35:787–794.
6. Guo Y, Chen C, Jin X, et al. Autosomal DIPs for population genetic structure and differentiation analyses of Chinese Xinjiang Kyrgyz ethnic group. Sci Rep. 2018;8:11054.
7. Alaeddini R, Walsh SJ, Abbas A. Forensic implications of genetic analyses from degraded DNA—a review. Forensic Sci Int Genet. 2010;4:148–157.
8. Takahashi M, Kato Y, Mukoyama H, et al. Evaluation of five polymorphic microsatellite markers for typing DNA from decomposed human tissues—correlation between the size of the alleles and that of the template DNA. Forensic Sci Int. 1997;90:1–9.
9. Jin R, Cui W, Fang Y, et al. A novel panel of 43 insertion/deletion loci for human identifications of forensic degraded DNA samples: development and validation. Front Genet. 2021;12:610540.
10. Bai H, Guo X, Zhang D, et al. The genome of a Mongolian individual reveals the genetic imprints of Mongolians on modern human populations. Genome Biol Evol. 2014;6:3122–3136.
11. Zerjal T, Xue Y, Bertorelle G, et al. The genetic legacy of the Mongols. Am J Hum Genet. 2003;72:717–721.
12. Fang Y, Xie T, Lan Q, et al. Multiple genetic analyses to investigate the polymorphisms of Chinese Mongolian population with an efficient short tandem repeat panel. Croat Med J. 2019;60:191–200.
13. Wu R, Li R, Wang N, et al. Genetic polymorphism and population structure of Torghut Mongols and comparison with a Mongolian population 3000 kilometers away. Forensic Sci Int Genet. 2019;42:235–243.
14. Lan Q, Xie T, Jin X, et al. MtDNA polymorphism analyses in the Chinese Mongolian group: efficiency evaluation and further matrilineal genetic structure exploration. Mol Genet Genomic Med. 2019;7:e00934.
15. Zhao C, Yang J, Xu H, et al. Genetic diversity analysis of forty-three insertion/deletion loci for forensic individual identification in Han Chinese from Beijing based on a novel panel. J Zhejiang Univ Sci B. 2022;23:241–248.
16. Rousset F. Genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour. 2008;8:103–106.
17. Yoo J, Lee Y, Kim Y, et al. SNPAnalyzer 2.0: a web-based integrated workbench for linkage disequilibrium analysis and association analysis. BMC Bioinformatics. 2008;9:290.
18. Gouy A, Zieger M. STRAF—a convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet. 2017;30:148–151.
19. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010;10:564–567.
20. Nei M, Tajima F, Tateno Y. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. J Mol Evol. 1983;19:153–170.
21. Chen X, Nie S, Hu L, et al. Forensic efficacy evaluation and genetic structure exploration of the Yunnan Miao group by a multiplex InDel panel. Electrophoresis. 2022;43:1765–1773.
22. Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol. 2005;14:2611–2620.
23. Earl DA, Vonholdt BMJCGR. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. 2012;4:359–361.
24. Jiang Y, He W, Wang H, et al. Population genetic analysis of a 21-plex DIP panel in seven Chinese ethnic populations. Int J Leg Med. 2018;132:145–147.
25. Li CT, Zhang SH, Zhao SM. Genetic analysis of 30 InDel markers for forensic use in five different Chinese populations. Genet Mol Res. 2011;10:964–979.
26. Huang Y, Chen X, Liu C, et al. Genetic analysis of 32 InDels in four ethnic minorities from Chinese Xinjiang. PloS One. 2021;16:e0250206.
27. Zhang W, Jin X, Wang Y, et al. Genetic polymorphisms and forensic efficiencies of a set of novel autosomal InDel markers in a Chinese Mongolian group. Biomed Res Int. 2020;2020:3925189.
28. Wang CC, Lu Y, Kang L, et al. The massive assimilation of indigenous East Asian populations in the origin of Muslim Hui people inferred from paternal Y chromosome. Am J Phys Anthropol. 2019;169:341–347.

29. Chen C, Li Y, Tao R, et al. The genetic structure of Chinese Hui ethnic group revealed by complete mitochondrial genome analyses using massively parallel sequencing. Genes (Basel). 2020;11:1352.
30. Hong W, Chen S, Shao H, et al. HLA class I polymorphism in Mongolian and Hui ethnic groups from northern China. Hum Immunol. 2007;68:439–448.
31. Yao YG, Kong QP, Wang CY, et al. Different matrilineal contributions to genetic structure of ethnic groups in the Silk Road region in China. Mol Biol Evol. 2004;21:2265–2280.
32. Gao T, Yun L, Gao S, et al. Population genetics of 23 Y-STR loci in the Mongolian minority population in Inner Mongolia of China. Int J Leg Med. 2016;130:1509–1511.