



## Correspondence

<https://doi.org/10.1631/jzus.B2100507>



# Genetic diversity analysis of forty-three insertion/deletion loci for forensic individual identification in Han Chinese from Beijing based on a novel panel

Congying ZHAO<sup>1</sup>, Jinlong YANG<sup>2</sup>, Hui XU<sup>1</sup>, Shuyan MEI<sup>1</sup>, Yating FANG<sup>1</sup>, Qiong LAN<sup>1</sup>, Yajun DENG<sup>2</sup>✉, Bofeng ZHU<sup>1,3</sup>✉

<sup>1</sup>Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou 510515, China

<sup>2</sup>Beijing Zhongzheng DNA Evidence Institute, Beijing 101318, China

<sup>3</sup>Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, Xi'an 710004, China

Due to the virtues of no stutter peaks, low rates of mutation, and short amplicon sizes, insertion/deletion (InDel) polymorphism is an indispensable tool for analyzing degraded DNA samples from crime scenes for human identifications (Wang et al., 2021). Herein, a self-developed panel of 43 InDel loci constructed previously by our group was utilized to evaluate the genetic diversities and explore the genetic background of the Han Chinese from Beijing (HCB) including 301 random healthy individuals. The lengths of amplicons at 43 InDel loci in this panel ranged from 87 to 199 bp, which indicated that the panel could be used as an effective tool to utilize highly degraded DNA samples for human identity testing. The loci in this panel were validated and performed well for forensic degraded DNA samples (Jin et al., 2021). The combined discrimination power (PD) and combined probability of exclusion (PE) values in this panel indicated that the 43 InDel loci could be used as the candidate markers in personal identification and parentage testing of HCB. In addition, population genetic relationships between the HCB and 26 reference populations from five continents based on 19 overlapped InDel loci were displayed by constructing a phylogenetic tree, principal

component analysis (PCA), and population genetic structure analysis. The results illustrated that the HCB had closer genetic relationships with the Han populations from Chinese different regions.

At present, a series of commercial panels are applied in forensic investigations, such as the Investigator<sup>®</sup> DIPplex kit (30 autosomal InDels and Amelogenin) and AGCU InDel 50 kit (47 autosomal InDels, 2 Y-chromosomal InDels, and Amelogenin) (Wang et al., 2016; Du et al., 2017). However, some InDel loci, such as locus HLD118 in the Investigator<sup>®</sup> DIPplex kit, have shown poor polymorphic distributions because these loci were primarily designed for Europeans rather than East Asians, restricting the forensic applications of such kit in Chinese population (Shen et al., 2016). Herein, to satisfy the demands of highly polymorphic loci in the populations from China, we co-amplified the 43 InDel loci by a multiplex amplification system and verify their forensic efficacy in HCB.

In the past decade, the perspectives of research on the Han population mostly stemmed from historians, archaeologists, and anthropologists. The historians generally believe that the history of Han population could be traced to the Huaxia tribe in the 21st century BC. Subsequently, this tribe gradually formed the Han population through integration with other ethnic groups until the early Han Dynasty (Zhao et al., 2015). In the field of forensic genetics, a variety of studies based on different genetic markers, for

✉ Bofeng ZHU, [zhubofeng7372@126.com](mailto:zhubofeng7372@126.com)

Yajun DENG, [18601285369@163.com](mailto:18601285369@163.com)

✉ Bofeng ZHU, <https://orcid.org/0000-0002-9038-2342>

Yajun DENG, <https://orcid.org/0000-0003-4938-9646>

Received June 11, 2021; Revision accepted Dec. 10, 2021;  
Crosschecked Feb. 23, 2022

© Zhejiang University Press 2022

example, autosomal short tandem repeats (STRs), X-chromosomal STRs, Y-chromosomal STRs, single nucleotide polymorphisms (SNPs), and InDels (Chen et al., 2021; Jia et al., 2021), have deepened scholars' understanding of the genetic characteristics of the Han population. With the reconstruction of old urban areas and the acceleration of rural urbanization, the geographical boundaries were broken and gene exchanges occurred between different groups of people. However, up to now data on the allelic frequencies of HCB are still lacking.

Therefore, this study investigated the genetic polymorphisms of HCB by a self-developed panel consisting of 43 autosomal InDels. A total of 26 reference populations were simultaneously screened for population genetic analyses based on a public database (Table S1). Moreover, the genetic relationships between HCB and reference populations were further clarified by population genetic analyses.

After Bonferroni's correction, no loci were observed to deviate from the Hardy-Weinberg equilibrium ( $P > 0.05/43 = 0.0012$ ), and the results for linkage disequilibrium indicated that all pairs of loci were in the linkage equilibrium ( $P > 0.05/903 = 0.00006$ ) in HCB (Tables S2 and S3), which indicated that the selected samples were representative and all of the InDel loci were independent of each other.

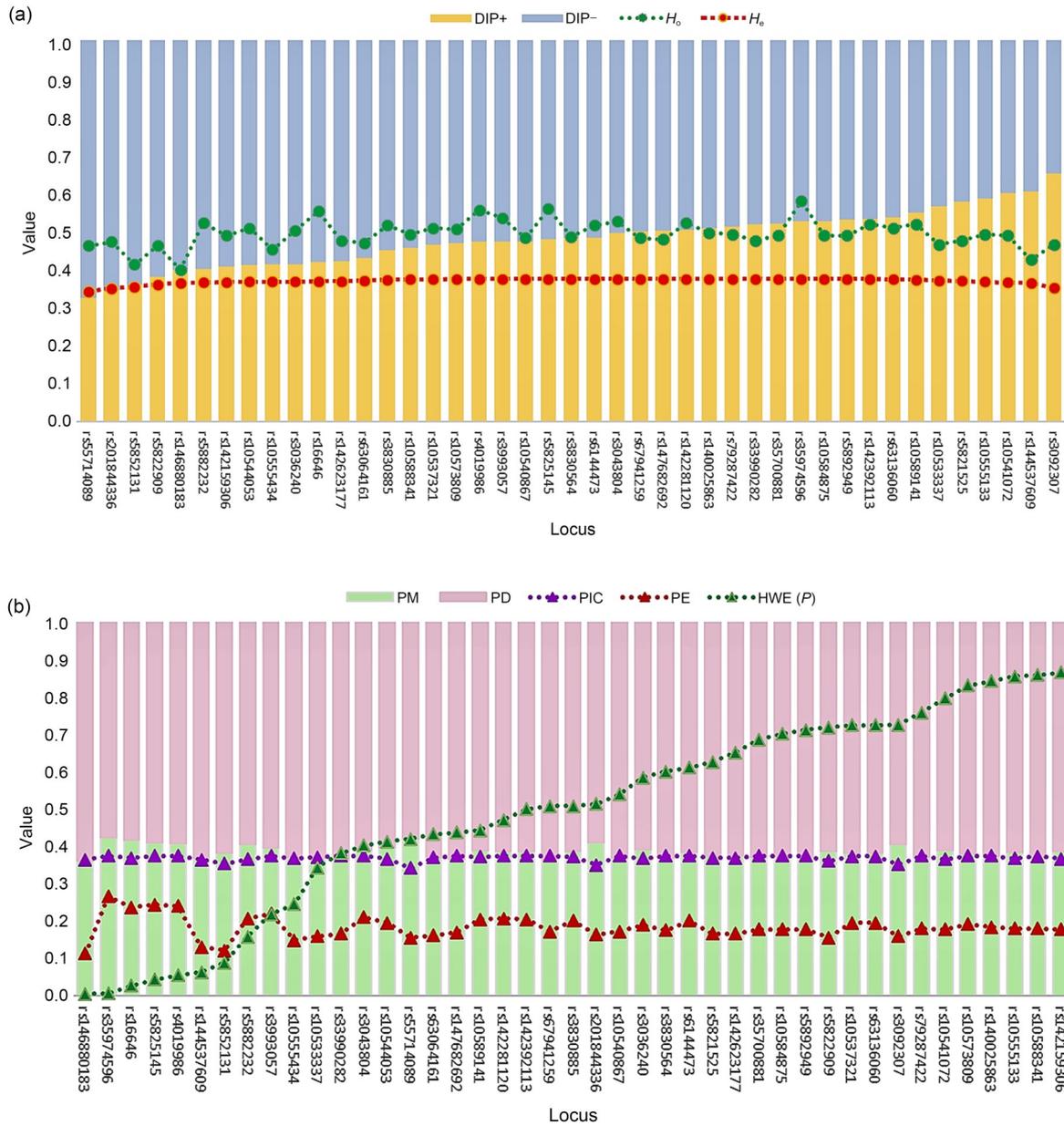
The calculations of insertion allelic frequencies (DIP+) in 43 InDel loci fluctuated within 0.3239 (rs55714089)–0.6512 (rs3092307) in HCB (Fig. 1a and Table S3). The mean value of DIP+ was 0.4803, and the DIP+ of most loci ranged from 0.4 to 0.6, indicating that these loci had relatively balanced frequency distributions. Besides, the observed heterozygosity ( $H_o$ ) and expected heterozygosity ( $H_e$ ) values spanned from 0.3987 (rs146880183) to 0.5781 (rs35974596), and 0.4387 (rs55714089) to 0.5008 (rs142281120, rs67941259, rs147682692, and rs3043804), respectively. The scopes of match probability (PM), PD, polymorphism information content (PIC), and PE at 43 loci in the HCB ranged from 0.3603 (rs146880183) to 0.4244 (rs35974596), 0.5756 (rs35974596) to 0.6397 (rs146880183), 0.3421 (rs55714089) to 0.3750 (rs142281120, rs67941259, rs140025863, rs147682692, and rs3043804), and 0.1131 (rs146880183) to 0.2654 (rs35974596), respectively (Fig. 1b and Table S3).

Furthermore, we calculated the combined PD and combined PE values, which were  $1-3.17 \times 10^{-18}$  and

0.999869, respectively. The values of DIP+ of 27 populations at 43 loci were shown in Fig. 2 and Table S4. The 27 populations were divided into five continents, and the cluster analyses on the vertical axis indicated that most populations from the same continent were clustered together. The HCB firstly clustered with Chinese Beijing Han (CHB) and Chinese Southern Han (CHS) populations, and then clustered with the remaining East Asia populations.

The pairwise fixation index ( $F_{ST}$ ) values based on the 43 InDel loci were calculated between different intercontinental populations (Table S5). An  $F_{ST}$  heatmap of the 43 InDel loci between intercontinental populations was performed to intuitively show the genetic differentiations between pairwise intercontinental populations by *R* software v.4.0.5 (Fig. 3). The loci including rs142392113, rs10589141, rs10541072, rs16646, rs5825145, rs67941259, rs10537321, rs10540867, rs147682692, rs3830885, rs5821525, rs33990282, and rs4019986 were observed to show relatively high differentiations ( $F_{ST} > 0.2$ ) between Africa populations (AFR) and non-Africa populations (non-AFR), implying that these InDel loci might distinguish AFR from non-AFR. Next, rs3092307 (East Asia populations (EAS)-Europe populations (EUR),  $F_{ST} > 0.15$ ), rs5852131 (America populations (AMR)-South Asia populations (SAS),  $F_{ST} > 0.1$ ), rs6144473 (AMR-SAS,  $F_{ST} > 0.1$ ), rs55714089 (EAS-AMR and EAS-SAS,  $F_{ST} > 0.1$ ), rs5822909 (EAS-SAS,  $F_{ST} > 0.1$ ), and rs144537609 (EAS-AMR,  $F_{ST} > 0.1$ ) loci could also be used to the distinguish these continental populations. Finally, the suitable 19 InDel loci of the above-mentioned 43 loci were selected to further dissect the genetic relationships between HCB and 26 reference populations.

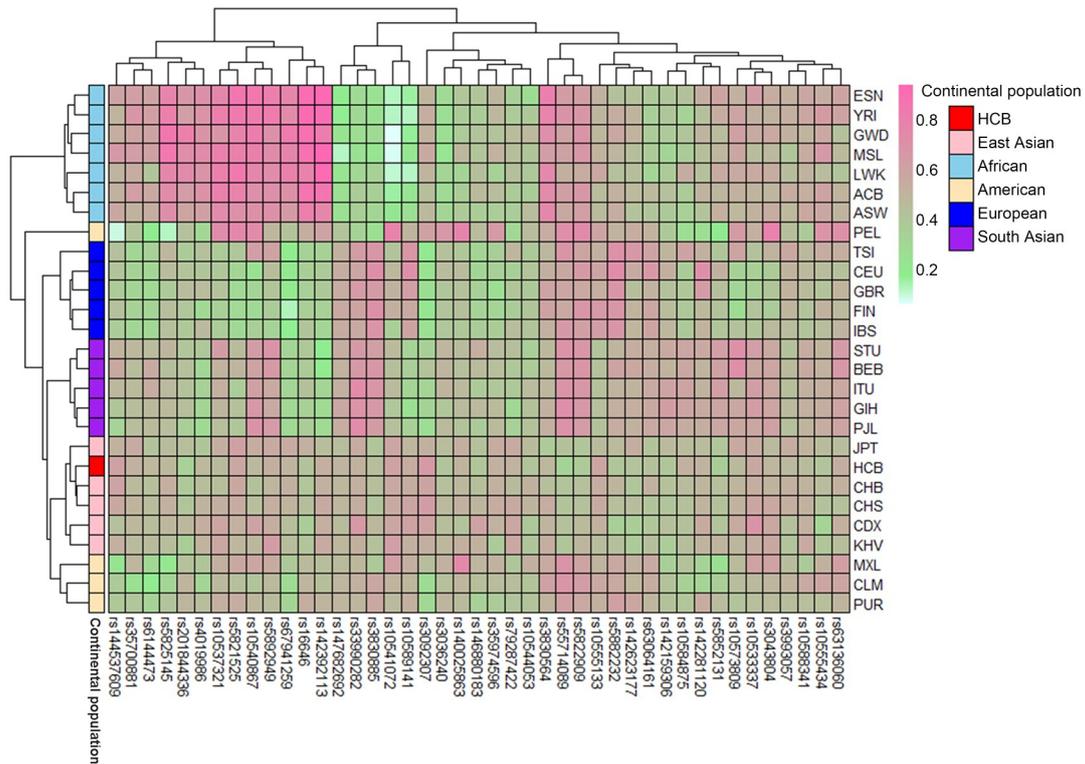
In order to verify the ancestry inference efficiencies of the selected 19 InDel loci, we used the Snipper online website (<http://mathgene.usc.es/snipper>) to conduct PCA on the individual level based on three continental (East Asia, Europe, and Africa) populations. These continental populations predominantly corresponded to three parts; most individuals of HCB overlapped with the cluster of East Asian individuals (Fig. S1). Then, PCA on the population level was conducted based on five continental (East Asia, South Asia, America, Europe, and Africa) populations. The first three principal components (PC1, PC2, and PC3) were 59.4%, 16.5%, and 13.0%, respectively, which could explain 88.9% variances (Fig. 4). Seven populations



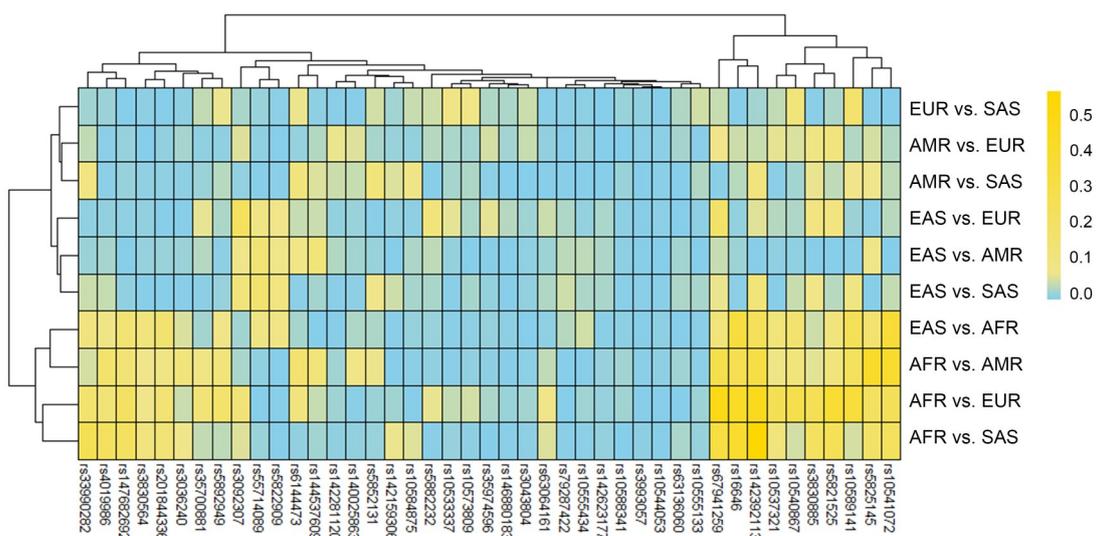
**Fig. 1** Diagrams of DIP+, DIP-, and forensic parameters in Han Chinese from Beijing based on 43 InDel loci. (a) DIP+, DIP-,  $H_o$ , and  $H_e$  values; (b) PM, PD, PIC, PE, and HWE ( $P$ ) values. DIP+, insertion allelic frequencies; DIP-, deletion allelic frequencies; InDel, insertion/deletion;  $H_o$ , observed heterozygosity;  $H_e$ , expected heterozygosity; PM, match probability; PD, discrimination power; PIC, polymorphism information content; PE, probability of exclusion; HWE ( $P$ ),  $P$  values of Hardy-Weinberg equilibrium.

from Africa (left side of plot) and five populations from Europe (right side of plot) could be distinguished in PC1 (Fig. 4a). Five populations from South Asia and six populations from East Asia could be divided into two independent clusters (top and bottom of the plots) in PC2 (Fig. 4a). The remaining four populations from America could be separated in PC3 (Fig. 4b).

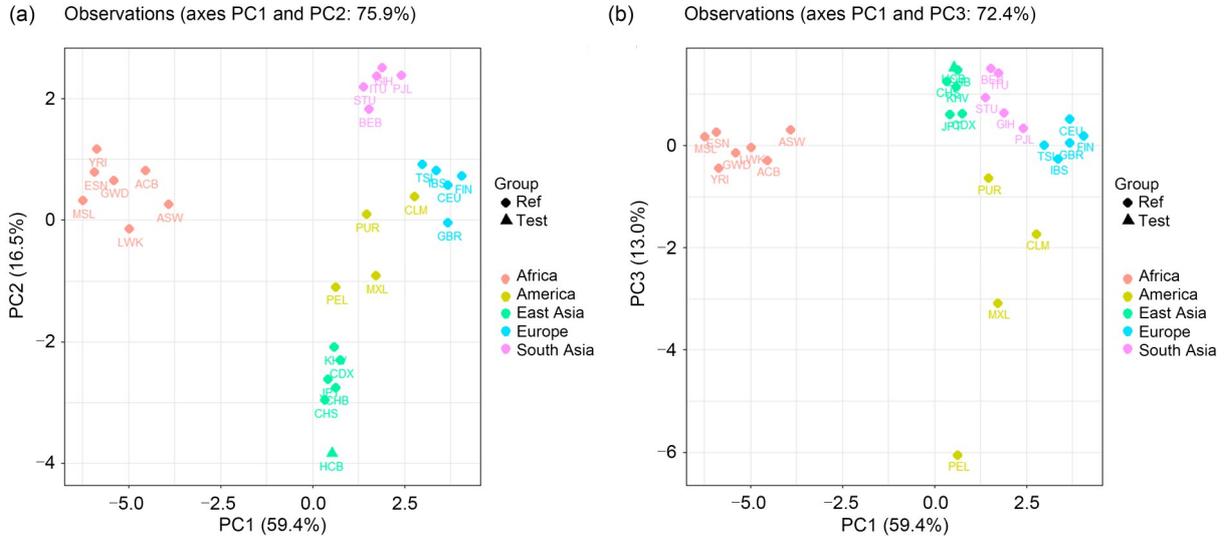
Next, we calculated the Nei's genetic distance ( $D_A$ ) and  $F_{ST}$  values between the HCB and 26 reference populations based on the 19 shared InDel loci (Tables S6 and S7), and constructed the two heatmaps of Nei's  $D_A$  and  $F_{ST}$  values (Fig. 5). Each grid represented a value of pairwise populations, and the depths of different colors represented the sizes of values. The first column of Nei's  $D_A$  values exhibited the genetic



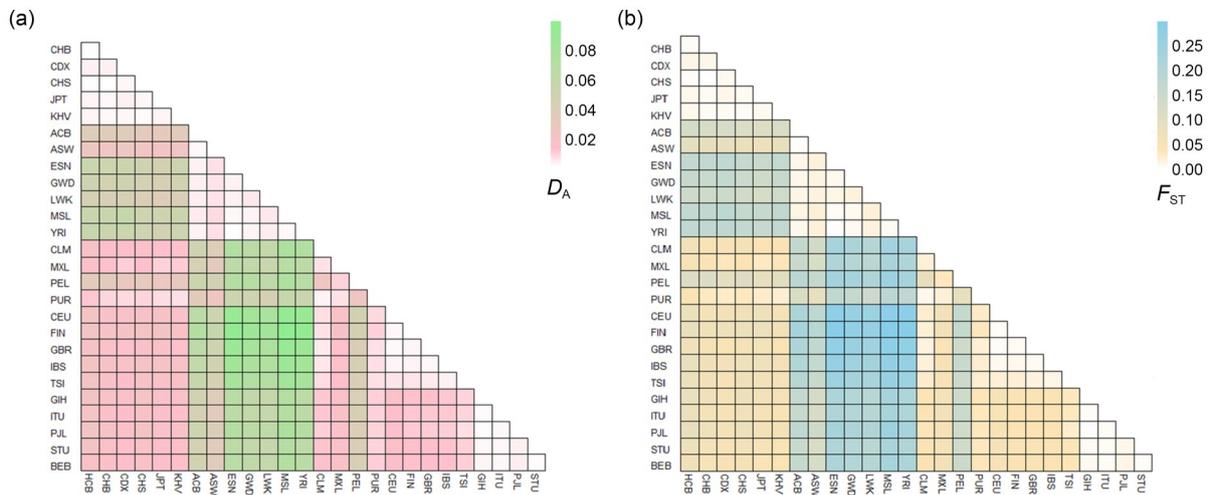
**Fig. 2** Heatmap of insertion allelic frequencies among HCB and 26 reference populations based on the 43 InDel loci. ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest USA; BEB, Bengali in Bangladesh; CDX, Chinese Dai in Xishuangbanna; CEU, Utah residents with Northern and Western European ancestry; CHB, Chinese Beijing Han; CHS, Chinese Southern Han; CLM, Colombian in Medellin; ESN, Esan in Nigeria; FIN, Finnish in Finland; GBR, British in England and Scotland; GIH, Gujarati Indian in Houston; GWD, Gambian in Western Division; HCB, Han Chinese from Beijing; IBS, Iberian populations in Spain; ITU, Indian Telugu in the UK; JPT, Japanese in Tokyo; KHV, Kinh in Ho Chi Minh City; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; MXL, Mexican Ancestry in Los Angeles; PEL, Peruvian in Lima; PJI, Punjabi in Lahore; PUR, Puerto Rican in Puerto Rico; STU, Sri Lankan Tamil in the UK; TSI, Toscani in Italy; YRI, Yoruba in Ibadan; InDel, insertion/deletion.



**Fig. 3** Heatmap of fixation index ( $F_{ST}$ ) values in pairwise continental populations from five continents based on the 43 InDel loci. InDel, insertion/deletion; EUR, Europe populations; SAS, South Asia populations; AMR, America populations; EAS, East Asia populations; AFR, Africa populations.



**Fig. 4** Principal component analyses of HCB and 26 reference populations performed by *R* software v.4.0.5 based on the 19 overlapped InDel loci. (a) PC1 and PC2 levels; (b) PC1 and PC3 levels. PC, principal component; Ref, reference; ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest USA; BEB, Bengali in Bangladesh; CDX, Chinese Dai in Xishuangbanna; CEU, Utah residents with Northern and Western European ancestry; CHB, Chinese Beijing Han; CHS, Chinese Southern Han; CLM, Colombian in Medellin; ESN, Esan in Nigeria; FIN, Finnish in Finland; GBR, British in England and Scotland; GIH, Gujarati Indian in Houston; GWD, Gambian in Western Division; HCB, Han Chinese from Beijing; IBS, Iberian populations in Spain; ITU, Indian Telugu in the UK; JPT, Japanese in Tokyo; KHV, Kinh in Ho Chi Minh City; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; MXL, Mexican Ancestry in Los Angeles; PEL, Peruvian in Lima; PJI, Punjabi in Lahore; PUR, Puerto Rican in Puerto Rico; STU, Sri Lankan Tamil in the UK; TSI, Toscani in Italy; YRI, Yoruba in Ibadan.

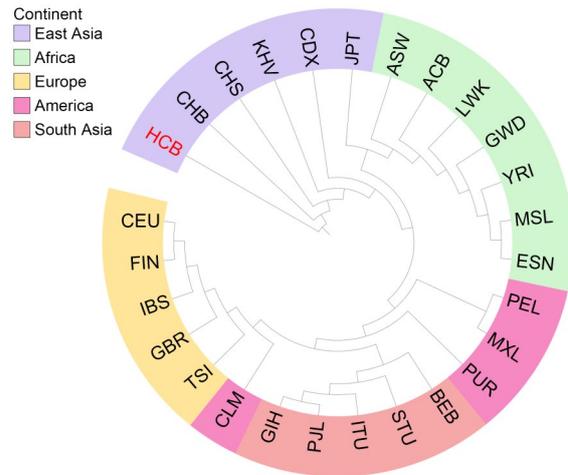


**Fig. 5** Heatmaps of the Nei's genetic distance ( $D_A$ ) values (a) and fixation index ( $F_{ST}$ ) genetic differentiation values (b) among HCB and 26 reference populations based on the 19 overlapped InDel loci. ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest USA; BEB, Bengali in Bangladesh; CDX, Chinese Dai in Xishuangbanna; CEU, Utah residents with Northern and Western European ancestry; CHB, Chinese Beijing Han; CHS, Chinese Southern Han; CLM, Colombian in Medellin; ESN, Esan in Nigeria; FIN, Finnish in Finland; GBR, British in England and Scotland; GIH, Gujarati Indian in Houston; GWD, Gambian in Western Division; HCB, Han Chinese from Beijing; IBS, Iberian populations in Spain; ITU, Indian Telugu in the UK; JPT, Japanese in Tokyo; KHV, Kinh in Ho Chi Minh City; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; MXL, Mexican Ancestry in Los Angeles; PEL, Peruvian in Lima; PJI, Punjabi in Lahore; PUR, Puerto Rican in Puerto Rico; STU, Sri Lankan Tamil in the UK; TSI, Toscani in Italy; YRI, Yoruba in Ibadan; InDel, insertion/deletion.

distances between the HCB and reference populations (Fig. 5a). Small  $D_A$  values could be observed between the HCB and five populations from East Asia including CHB, Chinese Dai in Xishuangbanna (CDX), CHS, Japanese in Tokyo (JPT), and Kinh in Ho Chi Minh City (KHV). The Nei's  $D_A$  values between the HCB and the populations from East Asia, Africa, America, Europe, and South Asia ranged from 0.0009 (CHB and CHS) to 0.0033 (CDX), 0.0269 (African Ancestry in Southwest USA (ASW)) to 0.0555 (Mende in Sierra Leone (MSL)), 0.0112 (Puerto Rican in Puerto Rico (PUR)) to 0.0339 (Peruvian in Lima (PEL)), 0.0181 (British in England and Scotland (GBR)) to 0.0232 (Finnish in Finland (FIN)), and 0.0178 (Bengali in Bangladesh (BEB)) to 0.0233 (Punjabi in Lahore (PJL)), respectively. The trend of  $F_{ST}$  values was consistent with the Nei's  $D_A$  values (Fig. 5b). The  $F_{ST}$  values between the HCB and the populations from East Asia, Africa, America, Europe, and South Asia ranged from 0.0002 (CHS) to 0.0098 (CDX), 0.0915 (ASW) to 0.1704 (Yoruba in Ibadan (YRI)), 0.0400 (PUR) to 0.1096 (PEL), 0.0639 (GBR) to 0.0797 (FIN), and 0.0628 (BEB) to 0.0827 (PJL), respectively. The intuitive charts of  $D_A$  and  $F_{ST}$  genetic differentiations which compared the studied HCB with the reference populations were displayed in Fig. S2.

We then constructed a phylogenetic tree based on Nei's  $D_A$  values to describe the genetic relationships among different populations (Fig. 6). The 27 populations involved in this study were divided into two main dendrimers: one branch consisted of seven populations from Africa, and the other branch included the remaining populations from South Asia, Europe, America, and East Asia. The HCB firstly grouped with the CHB population and then with other populations from East Asia in the same sub-branch.

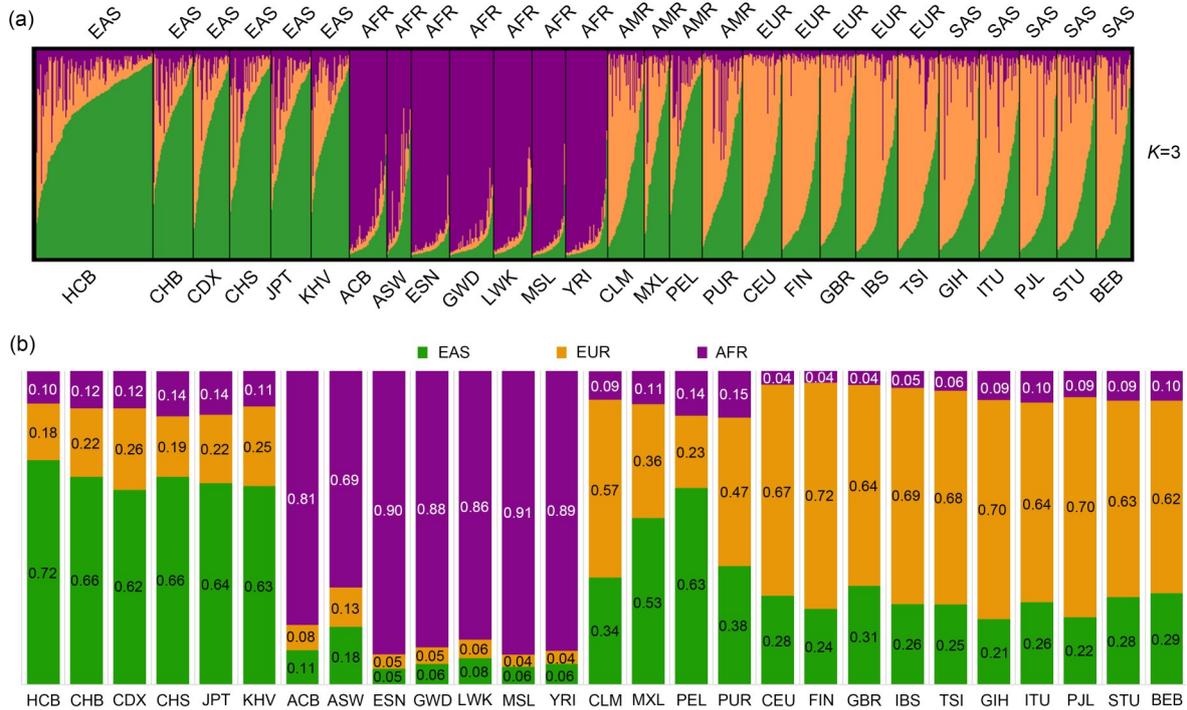
To uncover the genetic components of the HCB, we also performed population genetic structure analyses of 2805 individuals and 27 populations at individual and population levels, respectively.  $K$  value is used to infer ancestral components of population based on Bayesian algorithm. The optimal  $K$  value calculated by the Structure Harvester online website (<http://taylor0.biology.ucla.edu/structureHarvester>) was 3 (Fig. 7a). Different populations were separated by black bold lines, of which each thin line represented an individual component. The genetic structure analyses indicated that the 19 overlapped InDel loci could be



**Fig. 6** Phylogenetic tree among 27 populations constructed by the Interactive Tree of Life (iTOL) online website (<https://itol.embl.de>) based on 19 overlapped InDel loci. ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest USA; BEB, Bengali in Bangladesh; CDX, Chinese Dai in Xishuangbanna; CEU, Utah residents with Northern and Western European ancestry; CHB, Chinese Beijing Han; CHS, Chinese Southern Han; CLM, Colombian in Medellin; ESN, Esan in Nigeria; FIN, Finnish in Finland; GBR, British in England and Scotland; GIH, Gujarati Indian in Houston; GWD, Gambian in Western Division; HCB, Han Chinese from Beijing; IBS, Iberian populations in Spain; ITU, Indian Telugu in the UK; JPT, Japanese in Tokyo; KHV, Kinh in Ho Chi Minh City; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; MXL, Mexican Ancestry in Los Angeles; PEL, Peruvian in Lima; PJL, Punjabi in Lahore; PUR, Puerto Rican in Puerto Rico; STU, Sri Lankan Tamil in the UK; TSI, Toscani in Italy; YRI, Yoruba in Ibadan; InDel, insertion/deletion.

used as potential genetic markers to distinguish African, European, and East Asian populations. The HCB mainly belonged to the East Asian ancestral component, and the proportions of East Asian, European, and African ancestral components in HCB were 0.725, 0.178, and 0.097, respectively. Higher proportion from the East Asian ancestral component could be observed between HCB and CHB, CHS populations (Fig. 7b), and the similar result was previously reported by Song et al. (2020) based on 47 InDel loci.

Additionally, we performed the analysis of molecular variance (AMOVA) at 19 overlapped InDel loci between the HCB and reference populations. The values of locus-by-locus  $F_{ST}$  and significance ( $P < 0.05/19 = 0.0026$ ) were calculated by Arlequin software v. 3.5 (<http://cmpg.unibe.ch/software/arlequin35>) (Table S8). Significant differences could be observed between the HCB and JPT populations at 1 locus, CDX and KHV



**Fig. 7** Population genetic structure analyses conducted by STRUCTURE software v.2.3.4 when  $K=3$  at the levels of 2805 individuals (a) and 27 populations (b) based on the 19 overlapped loci. Different colors represent different ancestral components. ACB, African Caribbean in Barbados; ASW, African Ancestry in Southwest USA; BEB, Bengali in Bangladesh; CDX, Chinese Dai in Xishuangbanna; CEU, Utah residents with Northern and Western European ancestry; CHB, Chinese Beijing Han; CHS, Chinese Southern Han; CLM, Colombian in Medellin; ESN, Esan in Nigeria; FIN, Finnish in Finland; GBR, British in England and Scotland; GIH, Gujarati Indian in Houston; GWD, Gambian in Western Division; HCB, Han Chinese from Beijing; IBS, Iberian populations in Spain; ITU, Indian Telugu in the UK; JPT, Japanese in Tokyo; KHV, Kinh in Ho Chi Minh City; LWK, Luhya in Webuye, Kenya; MSL, Mende in Sierra Leone; and YRI, Yoruba in Ibadan.

populations at 2 loci, Mexican Ancestry in Los Angeles (MXL) population at 6 loci, PUR population at 7 loci, Colombian in Medellin (CLM) population at 9 loci, GBR and Iberian populations in Spain (IBS) populations at 10 loci, BEB population at 11 loci, Utah residents with Northern and Western European ancestry (CEU), FIN, Toscani in Italy (TSI), and Sri Lankan Tamil in the UK (STU) populations at 12 loci, ASW and Indian Telugu in the UK (ITU) populations at 13 loci, PEL, Gujarati Indian in Houston (GIH), and PJJ populations at 14 loci, African Caribbean in Barbados (ACB) population at 15 loci, Luhya in Webuye, Kenya (LWK) and MSL populations at 16 loci, and Esan in Nigeria (ESN), Gambian in Western Division (GWD), and YRI populations at 17 loci. No significant differentiations were observed between HCB and CHB, CHS populations.

In the present study, we successfully evaluated the forensic application efficiency of a self-developed

panel including 43 InDel loci in HCB, and investigated the genetic relationships between the HCB and 26 reference populations on basis of the 19 overlapped InDel loci. The results indicated high genetic polymorphisms of these loci in EAS, especially in Chinese Han population. Therefore, these results indicated that the panel could be a promising tool for human identification and parentage testing in the HCB. In addition, it could be applied as a complementary tool in the exploration of ancestral information inference in this population. Populations from different continents such as East Asia, Europe, and Africa could be distinguished from each other using the panel; however, it was not yet suitable for the differentiations of populations from the same continent. The genetic distance analyses showed that the studied HCB shared relatively closer relationships with CHB, CHS populations from the 1000 Genomes Project phase III.

## Materials and methods

Detailed methods are provided in the electronic supplementary materials of this paper.

## Acknowledgments

This study was supported by the National Natural Science Foundation of China (No. 81373248).

## Author contributions

Bofeng ZHU constructed the multiple amplification system, conceived and designed the experiment, and revised the manuscript. Congying ZHAO performed the majority of the experiment, analyzed the data, and wrote the manuscript. Jinlong YANG, Hui XU, Shuyan MEI, Yating FANG, and Qiong LAN performed part of the experiment and reviewed the manuscript. Yajun DENG collected the samples, performed part of the experiment, and revised the manuscript. All authors have read and approved the final manuscript, and therefore, have full access to all the data in the study and take responsibility for the integrity and security of the data.

## Compliance with ethics guidelines

Congying ZHAO, Jinlong YANG, Hui XU, Shuyan MEI, Yating FANG, Qiong LAN, Yajun DENG, and Bofeng ZHU declare that they have no conflict of interest.

The Ethics Committee of the Xi'an Jiaotong University Health Science Center and Southern Medical University approved all the processes including the sample collections, experimental design, and so on during this research (Approval No. XJTULAC201). We certify that all procedures in this investigation complied with the Helsinki Declaration of 1975, as revised in 2008 (5). Additional informed consent was obtained from all participants for which identifying information is included in this article.

## References

Chen QF, Kang KL, Song JJ, et al., 2021. Allelic diversity and forensic estimations of the Beijing Hans: comparative data on sequence-based and length-based STRs. *Forensic*

*Sci Int Genet*, 51:102424.

<https://doi.org/10.1016/j.fsigen.2020.102424>

Du WA, Peng ZY, Feng CL, et al., 2017. Forensic efficiency and genetic variation of 30 InDels in Vietnamese and Nigerian populations. *Oncotarget*, 8(51):88934-88940. <https://doi.org/10.18632/oncotarget.21494>

Jia J, Liu X, Fan QW, et al., 2021. Development and validation of a multiplex 19 X-chromosomal short tandem repeats typing system for forensic purposes. *Sci Rep*, 11:609. <https://doi.org/10.1038/s41598-020-80414-x>

Jin R, Cui W, Fang YT, et al., 2021. A novel panel of 43 insertion/deletion loci for human identifications of forensic degraded DNA samples: development and validation. *Front Genet*, 12:610540. <https://doi.org/10.3389/fgene.2021.610540>

Shen CM, Zhu BF, Yao TH, et al., 2016. A 30-InDel assay for genetic variation and population structure analysis of Chinese Tujia group. *Sci Rep*, 6:36842. <https://doi.org/10.1038/srep36842>

Song F, Lang M, Li LY, et al., 2020. Forensic features and genetic background exploration of a new 47-autosomal InDel panel in five representative Han populations residing in Northern China. *Mol Genet Genom Med*, 8(5):e1224. <https://doi.org/10.1002/mgg3.1224>

Wang L, Lv ML, Zaumsegel D, et al., 2016. A comparative study of insertion/deletion polymorphisms applied among Southwest, South and Northwest Chinese populations using Investigator® DIPplex. *Forensic Sci Int Genet*, 21:10-14. <https://doi.org/10.1016/j.fsigen.2015.08.005>

Wang MG, He GL, Gao S, et al., 2021. Molecular genetic survey and forensic characterization of Chinese Mongolians via the 47 autosomal insertion/deletion marker. *Genomics*, 113(4):2199-2210. <https://doi.org/10.1016/j.ygeno.2021.05.010>

Zhao YB, Zhang Y, Zhang QC, et al., 2015. Ancient DNA reveals that the genetic structure of the northern Han Chinese was shaped prior to 3,000 years ago. *PLoS ONE*, 10(5):e0125676. <https://doi.org/10.1371/journal.pone.0125676>

## Supplementary information

Tables S1–S8; Figs. S1 and S2; Materials and methods