



# Systematic Evaluation of a Novel 6-dye Direct and Multiplex PCR-CE-Based InDel Typing System for Forensic Purposes

Haoliang Fan<sup>1,2</sup>, Yitong He<sup>1</sup>, Shuanglin Li<sup>1</sup>, Qiqian Xie<sup>1</sup>, Fenfen Wang<sup>3</sup>, Zhengming Du<sup>3</sup>, Yating Fang<sup>1</sup>, Pingming Qiu<sup>1,\*†</sup> and Bofeng Zhu<sup>1,4,5\*†</sup>

<sup>1</sup>Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou, China, <sup>2</sup>School of Basic Medicine and Life Science, Hainan Medical University, Haikou, China, <sup>3</sup>First Clinical Medical College, Hainan Medical University, Haikou, China, <sup>4</sup>Clinical Research Center of Shaanxi Province for Dental and Maxillofacial Diseases, College of Stomatology, Xi'an Jiaotong University, Xi'an, China, <sup>5</sup>Key Laboratory of Shaanxi Province for Craniofacial Precision Medicine Research, College of Stomatology, Xi'an Jiaotong University, Xi'an, China

## OPEN ACCESS

### Edited by:

Shaoqing Wen,  
Fudan University, China

### Reviewed by:

Mengge Wang,  
Sichuan University, China  
Atif Adnan,  
China Medical University, China

### \*Correspondence:

Pingming Qiu  
qiupmfy@126.com  
Bofeng Zhu  
zhubofeng7372@126.com

### †ORCID:

Pingming Qiu  
orcid.org/0000-0002-5579-1124  
Bofeng Zhu  
orcid.org/0000-0002-9038-2342

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

Received: 20 July 2021

Accepted: 29 October 2021

Published: 10 January 2022

### Citation:

Fan H, He Y, Li S, Xie Q, Wang F, Du Z,  
Fang Y, Qiu P and Zhu B (2022)  
Systematic Evaluation of a Novel 6-dye  
Direct and Multiplex PCR-CE-Based  
InDel Typing System for  
Forensic Purposes.  
Front. Genet. 12:744645.  
doi: 10.3389/fgene.2021.744645

Insertion/deletion (InDel) polymorphisms, combined desirable characteristics of both short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs), are considerable potential in the fields of forensic practices and population genetics. However, most commercial InDel kits designed based on non-Asians limited extensive forensic applications in East Asian (EAS) populations. Recently, a novel 6-dye direct and multiplex PCR-CE-based typing system was designed on the basis of genome-wide EAS population data, which could amplify 60 molecular genetic markers, consisting of 57 autosomal InDels (A-InDels), 2 Y-chromosomal InDels (Y-InDels), and Amelogenin in a single PCR reaction and detect by capillary electrophoresis, simultaneously. In the present study, the DNA profiles of 279 unrelated individuals from the Hainan Li group were generated by the novel typing system. In addition, we collected two A-InDel sets to evaluate the forensic performances of the novel system in the 1,000 Genomes Project (1KG) populations and Hainan Li group. For the Universal A-InDel set (UAIS, containing 44 A-InDels) the cumulative power of discrimination (CPD) ranged from 1–1.03 × 10<sup>-14</sup> to 1–1.27 × 10<sup>-18</sup>, and the cumulative power of exclusion (CPE) varied from 0.993634 to 0.999908 in the 1KG populations. For the East Asia-based A-InDel set (EAIS, containing 57 A-InDels) the CPD spanned from 1–1.32 × 10<sup>-23</sup> to 1–9.42 × 10<sup>-24</sup>, and the CPE ranged from 0.999965 to 0.999997. In the Hainan Li group, the average heterozygote (He) was 0.4666 (0.2366–0.5448), and the polymorphism information content (PIC) spanned from 0.2116 to 0.3750 (mean PIC: 0.3563 ± 0.0291). In total, the CPD and CPE of 57 A-InDels were 1–1.32 × 10<sup>-23</sup> and 0.999965, respectively. Consequently, the novel 6-dye direct and multiplex PCR-CE-based typing system could be considered as the reliable and robust tool for human identification and intercontinental population differentiation, and supplied additional information for kinship analysis in the 1KG populations and Hainan Li group.

**Keywords:** InDel, PCR-CE, East Asian population, Hainan Li group, 1000 Genomes Project, Human identification, Intercontinental population differentiation

## INTRODUCTION

Insertion/deletion (InDel) polymorphisms, the length polymorphisms resulting from the insertion or deletion of one or more nucleotides, are gradually becoming a type of alternative genetic markers for forensic purposes (Bus et al., 2016; Sun et al., 2016; Caputo et al., 2017; Sheng et al., 2018; Xie et al., 2018; Zhang et al., 2018; Sun et al., 2019; Zhang et al., 2019; Abel et al., 2020; Cui et al., 2020; Huang et al., 2020; Zhang et al., 2020). Low mutation rates ( $\sim 10^{-9}$ ) and no stutter/noise peaks are the overwhelming superiorities for InDels, which possess desirable properties of both short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) (Chakraborty et al., 1999; Weber et al., 2002; Bhangale et al., 2005; Mills et al., 2006; Pakstis et al., 2007; Mullaney et al., 2010; Pakstis et al., 2010; da Costa Francez et al., 2012; Kidd et al., 2012). In addition, the relatively small amplicon sizes of InDels enhance discrimination efficiencies in some dated or highly degraded samples from crime scenes when compared with STRs (Golenberg et al., 1996; Brinkmann et al., 1998; Jin et al., 2019). With the relatively uncomplicated chemical and operational approaches to detect the length variations in contrast to the determination methods of SNPs (Kwok 2002; Amoako et al., 2017; Matsuda 2017), the detection method by capillary electrophoresis (CE) for InDels could be extensively applied in distinct forensic scenarios (Zhao et al., 2018; Chen et al., 2019; Jin et al., 2019; Tao et al., 2019; Huang et al., 2020; Song et al., 2020; Zhang et al., 2020).

At present, the commercial and widely-used InDel kits present some issues, 1) they are not always suitable for East Asian (EAS) ancestry populations; 2) the insufficient utilization for the CE system; and 3) the time-consuming procedures for DNA extractions and/or purifications. The shortcomings for most InDel typing systems limit the promotion of forensic system effectiveness and the extension of forensic scenarios for EAS populations. Therefore, based on the underlying genome-wide data of the EAS populations from the 1,000 Genome Project (1KG) Phase 3 (Genomes Project et al., 2010; Genomes Project et al., 2015) and the engineering fundamental logic for the maximum utilization of multiplex PCR-CE system, a novel 6-dye direct and multiplex PCR-CE-based typing system with relatively short amplicons (<230 bp), consisting of 57 autosomal InDels (A-InDels), 2 Y-chromosomal InDels (Y-InDels), and Amelogenin, was studied to expand application scenarios for forensic purposes in EAS populations, especially for different Chinese populations. Moreover, forensic efficiencies and population genetic analyses of the direct and multiplex PCR-CE-based InDel typing system were further evaluated in 26 globally dispersed populations and the Hainan Li (HNL) group, which is a relatively isolated Chinese group revealed by the previous studies (Fan et al., 2018b; Fan et al. 2019b; Fan et al. 2021a; Wang et al., 2021).

## MATERIALS AND METHODS

### Sample and Data Collections

Bloodstain samples of 279 unrelated healthy Hainan Li individuals were collected after receiving their informed consents. The experiment was conducted in accordance with the guidelines of humane and ethical research of Xi'an

Jiaotong University and Southern Medical University, and warranted by the Ethics Committee of Xi'an Jiaotong University (No. 2019-1231). To evaluate the universal applicability of the novel 6-dye direct and multiplex PCR-CE-based typing system, we collected the population data of global 1KG populations from five continents.

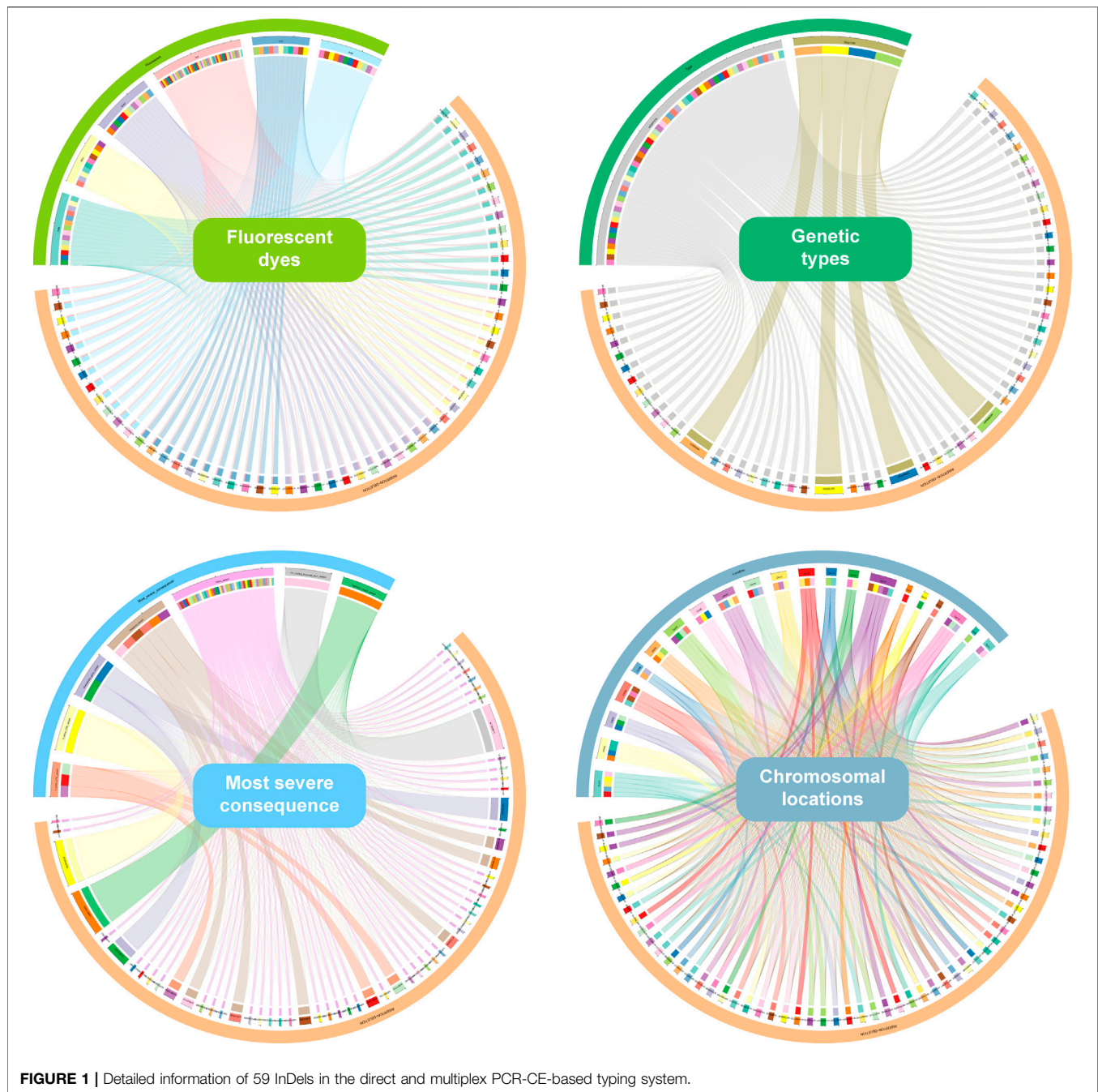
### Amplification and CE Detection

The amplification of the novel 6-dye direct and multiplex PCR-CE-based typing system was performed in a single multiplex PCR reaction (25  $\mu$ l in total) using 10  $\mu$ l of Reaction Mix (AGCU ScienTech Incorporation, Wuxi, Jiangsu, China), 1  $\mu$ l of U-Taq Enzyme (AGCU ScienTech Incorporation), 5  $\mu$ l of InDel 60 Primers (AGCU ScienTech Incorporation), and 9  $\mu$ l of  $\text{sdH}_2\text{O}$ . PCR cocktail was performed on the GeneAmp PCR System 9700 Thermal Cycler (Thermo Fisher Scientific, Waltham, MA, USA) based on the following parameters: initial denaturation at 95°C for 5 min; then 28 cycles of 94°C for 30 s, 60°C for 1 min, and 62°C for 1 min; and the final extension at 72°C for 10 min. Afterward, 1  $\mu$ l of PCR production was added to the cocktail of 0.5  $\mu$ l AGCU Marker SIZ-500 (AGCU ScienTech Incorporation) and 12  $\mu$ l of HiDi™ formamide (Thermo Fisher Scientific). The mixture was denatured at 95°C for 3 min and then immediately chilled on ice for 3 min. Finally, the product was detected on the 3500xL Genetic Analyzer (Thermo Fisher Scientific) using 36-cm capillary arrays (Thermo Fisher Scientific) with the POP-4® Polymer (Thermo Fisher Scientific). The CE parameters were as follows: 10 s injection at 2 kV, and electrophoresis at 15 kV for 1,400 s at 60°C. Genotyping data for each sample were determined by GeneMapper® ID-X software v1.6 (Thermo Fisher Scientific). Control DNA 9948 and deionized water were used as positive and negative controls, respectively.

### Statistical Analysis

Allele frequencies of InDels were calculated using SAS® 9.4 software (SAS Institute Inc., Cary, NC, USA). The forensic parameters, match probability (MP), power of discrimination (PD), polymorphism information content (PIC), power of exclusion (PE), typical paternity index (TPI), and heterozygote ( $H_e$ ), were conducted by PowerStats software (Promega, Madison, WI, USA). The Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) were evaluated by Arlequin v3.5 (Excoffier and Lischer 2010). The mean values and standard deviations of forensic relevant parameters were calculated by SAS® 9.4 software (SAS Institute Inc., Cary, NC, USA).

Population pairwise genetic distances (i.e.,  $F_{ST}$ ) and corresponding  $p$ -values between different populations were estimated by analysis of molecular variance (AMOVA) using raw data at Arlequin v3.5 (Excoffier and Lischer 2010). Genetic similarities and differences were further visualized by principal component analysis (PCA)

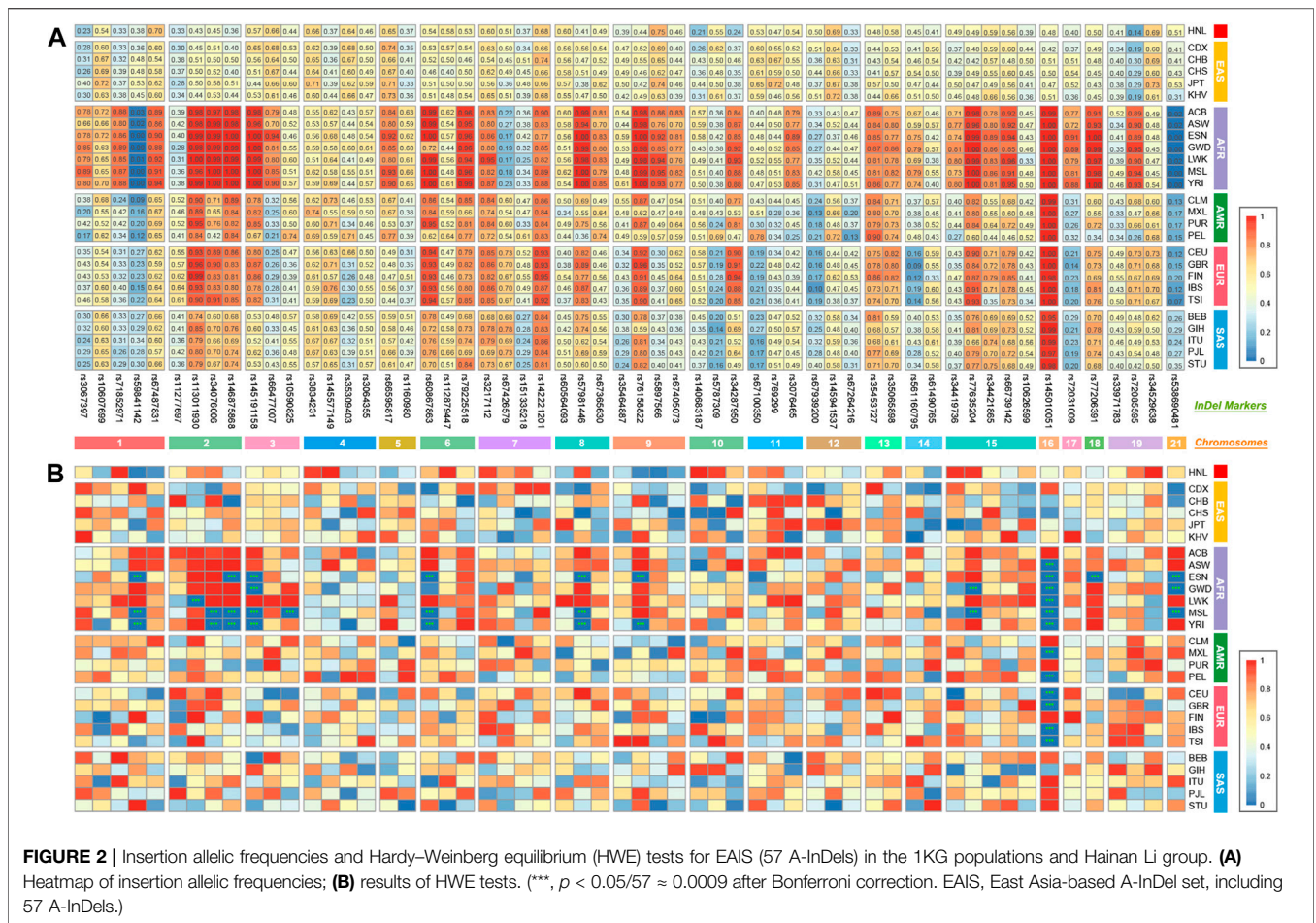


and multidimensional scaling plot (MDS) using *R* (<https://www.r-project.org/>) based on insertion allelic frequencies. Additionally, phylogenetic relationships among different populations were depicted in Molecular Evolutionary Genetics Analysis 7.0 (MEGA 7.0) software (Kumar et al., 2016) with neighbor-joining method (Saitou and Nei 1987) and visualized by the Interactive Tree of Life v5 (iTOL) (Letunic and Bork 2019). Other high-quality figures all used *R* to visualize.

## RESULTS AND DISCUSSION

### Details of the Novel Typing System and Distinct A-InDel Sets

**Supplementary Table S1** presents the detailed InDel information of the 6-dye direct and multiplex PCR-CE-based typing system. All genetic markers are autosomal and Y-chromosomal biallelic variations of InDels with the minimum allele frequency (MAF)  $\geq 0.25$  in most EAS populations (CHS and CHB in particular). A



total of six fluorescent dyes were applied to the direct typing system. The SIZ dye marked the internal reference item, and the HEX, LYN, SUM, PUR, and FAM dyes labeled 10, 11, 12, 13, and 14 InDels in accordance with the amplicon sizes, respectively (Figure 1 and Supplementary Table S1). Besides, the insertion/deletion fragment lengths ranged from 0 to 12 bp, and the amplicon sizes spanned from 76 to 226 bp, respectively. With the consideration of short amplicons for more forensic application scenarios, further validation and evaluation of forensic efficiencies, such as the ability to analyze dated/degraded samples, sensitivity, specificity, stability, and so on, should be performed in our subsequent study. In addition, the 59 InDels were distributed over 20 autosomes and Y chromosome. Of which, 45 InDels were located at the intron variant regions, 6 InDels were located at the intergenic variant regions, 3 InDels were located at the 3' UTR variant regions, 2 InDels were located at the downstream gene variant regions, 1 InDel was located at the 5' UTR variant region, 1 InDel was located at the regulatory region variant region, and 1 InDel was located at the non-coding transcript exon variant region (Figure 1). Physical distances of pairwise InDels located on the same chromosomes indicated that most pairwise InDels on the same chromosomes were more than 10 Mb apart except for the five pairs (rs3067397 and rs10607699,

~6.3 Mb; rs35464887 and rs76158822, ~4.1 Mb; rs76158822 and rs5897566, ~9.3 Mb; rs34419736 and rs77635204, ~8.8 Mb; and rs72085595 and rs34529638, ~5.9 Mb), but  $p$ -values (after Bonferroni corrections) of LD analyses (data not shown) were not statistically significant in the 1KG populations and Hainan Li group.

The direct and multiplex PCR-CE typing system (57 A-InDels, 2 Y-InDels, and Amelogenin) was studied based on the genome-wide data from the EAS populations. Thus, the East Asia-based A-InDel set (EAIS, including 57 A-InDels) of the typing system would be performing well for forensic purposes in the EAS populations. The results of HWE tests for 57 A-InDels in the 1KG populations and Hainan Li group are presented in Figure 2B and Supplementary Table S3. An overwhelming majority of A-InDel loci conformed to HWE in the 1KG populations and Hainan Li group after Bonferroni correction ( $0.05/57 = 0.0009$ ), while 13 A-InDels (rs59841142, rs113011930, rs34076006, rs146875868, rs145191158, rs10590825, rs60867863, rs57981446, rs76158822, rs77635204, rs145010051, rs77206391, and rs538690481) failed to pass the HWE tests, which mainly concentrated on African (AFR) ancestry populations. Therefore, we determined Universal A-InDel set (UAIS, including 44

**TABLE 1** | Comparisons of forensic relevant parameters in East Asia-based autosomal insertion/deletion set (EAIS) and universal autosomal insertion/deletion set (UAIS) for the 1,000 Genome Project (1KG) populations and Hainan Li group.

Population	n	Match probability (MP)		Power of discrimination (PD)		Polymorphism information content (PIC)		Power of exclusion (PE)		Typical paternity index (TPI)		Heterozygote (He)		
		EAIS	UAIS	EAIS	UAIS	EAIS	UAIS	EAIS	UAIS	EAIS	UAIS	EAIS	UAIS	
AFR	ACB	96	0.5699 ± 0.2096	0.4805 ± 0.1218	0.4301 ± 0.2096	0.5195 ± 0.1218	0.2478 ± 0.1275	0.3005 ± 0.0792	0.0940 ± 0.0753	0.1166 ± 0.0657	0.7722 ± 0.1898	0.8340 ± 0.1570	0.3098 ± 0.1773	0.3771 ± 0.1262
	ASW	61	0.5472 ± 0.2079	0.4549 ± 0.1049	0.4528 ± 0.2079	0.5451 ± 0.1049	0.2614 ± 0.1270	0.3167 ± 0.0703	0.1032 ± 0.0752	0.1294 ± 0.0624	0.7954 ± 0.1891	0.8657 ± 0.1472	0.3310 ± 0.1736	0.4042 ± 0.1092
	ESN	99	0.6112 ± 0.2368	0.5123 ± 0.1457	0.3888 ± 0.2368	0.4877 ± 0.1457	0.2256 ± 0.1429	0.2831 ± 0.0953	0.0916 ± 0.0849	0.1152 ± 0.0801	0.7616 ± 0.2169	0.8291 ± 0.1933	0.2885 ± 0.1987	0.3627 ± 0.1507
	GWD	113	0.5871 ± 0.2186	0.5005 ± 0.1283	0.4129 ± 0.2186	0.4995 ± 0.1283	0.2373 ± 0.1323	0.2875 ± 0.0855	0.0889 ± 0.0778	0.1093 ± 0.0726	0.7588 ± 0.1966	0.8166 ± 0.1728	0.2953 ± 0.1810	0.3598 ± 0.1339
	LWK	99	0.5935 ± 0.2277	0.4980 ± 0.1417	0.4065 ± 0.2277	0.5020 ± 0.1417	0.2346 ± 0.1378	0.2902 ± 0.0919	0.0926 ± 0.0793	0.1158 ± 0.0717	0.7662 ± 0.2019	0.8314 ± 0.1725	0.2988 ± 0.1891	0.3705 ± 0.1385
	MSL	85	0.6057 ± 0.2293	0.5060 ± 0.1408	0.3943 ± 0.2293	0.4940 ± 0.1408	0.2296 ± 0.1406	0.2859 ± 0.0938	0.0955 ± 0.0954	0.1140 ± 0.0813	0.7738 ± 0.2438	0.8272 ± 0.1943	0.2941 ± 0.1992	0.3620 ± 0.1458
YRI	108	0.6111 ± 0.2377	0.5123 ± 0.1503	0.3901 ± 0.2386	0.4893 ± 0.1511	0.2246 ± 0.1429	0.2816 ± 0.0972	0.0847 ± 0.0786	0.1055 ± 0.0733	0.7454 ± 0.2014	0.8060 ± 0.1767	0.2787 ± 0.1914	0.3486 ± 0.1444	
AMR	CLM	94	0.4545 ± 0.1186	0.4155 ± 0.0614	0.5455 ± 0.1186	0.5845 ± 0.0614	0.3194 ± 0.0765	0.3448 ± 0.0433	0.1340 ± 0.0685	0.1530 ± 0.0586	0.8753 ± 0.1638	0.9218 ± 0.1364	0.4062 ± 0.1238	0.4453 ± 0.0848
	MXL	64	0.4452 ± 0.1045	0.4201 ± 0.0591	0.5548 ± 0.1045	0.5799 ± 0.0591	0.3271 ± 0.0681	0.3450 ± 0.0405	0.1467 ± 0.0693	0.1622 ± 0.0634	0.9062 ± 0.1655	0.9439 ± 0.1493	0.4285 ± 0.1127	0.4570 ± 0.0859
	PEL	85	0.4616 ± 0.1216	0.4207 ± 0.0559	0.5384 ± 0.1216	0.5793 ± 0.0559	0.3151 ± 0.0772	0.3418 ± 0.0403	0.1313 ± 0.0657	0.1528 ± 0.0539	0.8685 ± 0.1580	0.9212 ± 0.1253	0.4022 ± 0.1250	0.4465 ± 0.0805
	PUR	104	0.4491 ± 0.1066	0.4309 ± 0.0705	0.5509 ± 0.1066	0.5691 ± 0.0705	0.3202 ± 0.0681	0.3309 ± 0.0485	0.1286 ± 0.0573	0.1331 ± 0.0524	0.8634 ± 0.1364	0.8756 ± 0.1217	0.4045 ± 0.1063	0.4174 ± 0.0845
EAS	CDX	93	0.3960 ± 0.0351	0.4011 ± 0.0341	0.6040 ± 0.0351	0.5989 ± 0.0341	0.3589 ± 0.0201	0.3559 ± 0.0216	0.1639 ± 0.0558	0.1628 ± 0.0566	0.9480 ± 0.1308	0.9453 ± 0.1323	0.4631 ± 0.0701	0.4611 ± 0.0721
	CHB	103	0.3890 ± 0.0305	0.3931 ± 0.0302	0.6110 ± 0.0305	0.6069 ± 0.0302	0.3641 ± 0.0145	0.3614 ± 0.0154	0.1720 ± 0.0451	0.1703 ± 0.0435	0.9663 ± 0.1059	0.9622 ± 0.1014	0.4769 ± 0.0520	0.4751 ± 0.0508
	CHS	105	0.4022 ± 0.0308	0.4027 ± 0.0322	0.5978 ± 0.0308	0.5973 ± 0.0322	0.3646 ± 0.0132	0.3621 ± 0.0140	0.1959 ± 0.0560	0.1883 ± 0.0571	1.0223 ± 0.1330	1.0049 ± 0.1351	0.5029 ± 0.0622	0.4939 ± 0.0640
	JPT	104	0.3948 ± 0.0289	0.3961 ± 0.0296	0.6052 ± 0.0289	0.6039 ± 0.0296	0.3607 ± 0.0174	0.3575 ± 0.0185	0.1700 ± 0.0507	0.1611 ± 0.0491	0.9618 ± 0.1182	0.9413 ± 0.1143	0.4728 ± 0.0598	0.4618 ± 0.0589
	KHV	99	0.4005 ± 0.0331	0.4050 ± 0.0344	0.5995 ± 0.0331	0.5950 ± 0.0344	0.3612 ± 0.0191	0.3584 ± 0.0202	0.1828 ± 0.0518	0.1805 ± 0.0562	0.9914 ± 0.1225	0.9867 ± 0.1333	0.4884 ± 0.0600	0.4846 ± 0.0650
	HNL	279	0.3986 ± 0.0409	0.4042 ± 0.0440	0.6014 ± 0.0409	0.5958 ± 0.0440	0.3563 ± 0.0291	0.3522 ± 0.0318	0.1639 ± 0.0372	0.1597 ± 0.0399	0.9463 ± 0.0856	0.9366 ± 0.0920	0.4666 ± 0.0560	0.4602 ± 0.0610
EUR	CEU	99	0.4903 ± 0.1374	0.4474 ± 0.0903	0.5097 ± 0.1374	0.5526 ± 0.0903	0.2957 ± 0.0881	0.3241 ± 0.0606	0.1183 ± 0.0722	0.1370 ± 0.0674	0.8375 ± 0.1746	0.8837 ± 0.1599	0.3754 ± 0.1365	0.4147 ± 0.1109
	GBR	91	0.4934 ± 0.1491	0.4454 ± 0.1003	0.5066 ± 0.1491	0.5546 ± 0.1003	0.2922 ± 0.0938	0.3222 ± 0.0674	0.1122 ± 0.0668	0.1309 ± 0.0599	0.8224 ± 0.1625	0.8693 ± 0.1413	0.3656 ± 0.1370	0.4082 ± 0.1045
	FIN	99	0.4897 ± 0.1430	0.4522 ± 0.1064	0.5103 ± 0.1430	0.5478 ± 0.1064	0.2986 ± 0.0920	0.3236 ± 0.0713	0.1256 ± 0.0795	0.1453 ± 0.0754	0.8543 ± 0.1924	0.9027 ± 0.1796	0.3830 ± 0.1460	0.4224 ± 0.1232
	IBS	107	0.4825 ± 0.1317	0.4381 ± 0.0824	0.5175 ± 0.1317	0.5619 ± 0.0824	0.2995 ± 0.0842	0.3288 ± 0.0559	0.1173 ± 0.0667	0.1384 ± 0.0590	0.8353 ± 0.1605	0.8870 ± 0.1384	0.3771 ± 0.1296	0.4212 ± 0.0990
	TSI	107	0.4859 ± 0.1398	0.4356 ± 0.0830	0.5141 ± 0.1398	0.5644 ± 0.0830	0.2971 ± 0.0878	0.3285 ± 0.0551	0.1157 ± 0.0630	0.1333 ± 0.0519	0.8312 ± 0.1530	0.8754 ± 0.1218	0.3748 ± 0.1310	0.4163 ± 0.0914
SAS	BEB	86	0.4245 ± 0.0704	0.4093 ± 0.0444	0.5755 ± 0.0704	0.5907 ± 0.0444	0.3394 ± 0.0472	0.3493 ± 0.0317	0.1520 ± 0.0556	0.1580 ± 0.0497	0.9194 ± 0.1306	0.9335 ± 0.1147	0.4447 ± 0.0837	0.4561 ± 0.0681
	GIH	103	0.4335 ± 0.0917	0.4158 ± 0.0528	0.5665 ± 0.0917	0.5842 ± 0.0528	0.3328 ± 0.0580	0.3449 ± 0.0372	0.1426 ± 0.0590	0.1522 ± 0.0565	0.8967 ± 0.1400	0.9201 ± 0.1313	0.4274 ± 0.1001	0.4451 ± 0.0822
	ITU	102	0.4295 ± 0.0862	0.4169 ± 0.0532	0.5705 ± 0.0862	0.5831 ± 0.0532	0.3345 ± 0.0563	0.3429 ± 0.0392	0.1417 ± 0.0568	0.1487 ± 0.0549	0.8948 ± 0.1337	0.9118 ± 0.1267	0.4274 ± 0.0953	0.4407 ± 0.0796
	PJL	96	0.4305 ± 0.0818	0.4146 ± 0.0513	0.5695 ± 0.0818	0.5854 ± 0.0513	0.3312 ± 0.0526	0.3421 ± 0.0362	0.1330 ± 0.0484	0.1418 ± 0.0463	0.8748 ± 0.1138	0.8959 ± 0.1066	0.4176 ± 0.0855	0.4335 ± 0.0714
	STU	102	0.4250 ± 0.0907	0.4124 ± 0.0626	0.5750 ± 0.0907	0.5876 ± 0.0626	0.3326 ± 0.0563	0.3405 ± 0.0408	0.1242 ± 0.0514	0.1270 ± 0.0531	0.8549 ± 0.1225	0.8621 ± 0.1244	0.4028 ± 0.0893	0.4086 ± 0.0808

**TABLE 2** | Comparisons of forensic system efficiencies in EAIS and UAIS for the 1KG populations and Hainan Li group.

Population		n	CMP		CPD		CPE	
			EAIS	UAIS	EAIS	UAIS	EAIS	UAIS
AFR	ACB	96	3.50E-16	2.82E-15	1-3.50E-16	1-2.82E-15	0.997060	0.996212
	ASW	61	3.28E-17	3.32E-16	1-3.28E-17	1-3.32E-16	0.998356	0.997997
	ESN	99	1.10E-14	3.25E-14	1-1.10E-14	1-3.25E-14	0.996773	0.996199
	GWD	113	1.74E-15	1.56E-14	1-1.74E-15	1-1.56E-14	0.996016	0.994716
	LWK	99	2.44E-15	1.03E-14	1-2.44E-15	1-1.03E-14	0.996858	0.996168
	MSL	85	8.00E-15	2.01E-14	1-8.00E-15	1-2.01E-14	0.997689	0.995993
	YRI	108	8.96E-15	2.55E-14	1-8.96E-15	1-2.55E-14	0.994830	0.993634
AMR	CLM	94	6.44E-21	1.08E-17	1-6.44E-21	1-1.08E-17	0.999771	0.999396
	MXL	64	2.80E-21	1.81E-17	1-2.80E-21	1-1.81E-17	0.999902	0.999636
	PEL	85	1.59E-20	2.00E-17	1-1.59E-20	1-2.00E-17	0.999721	0.999380
	PUR	104	4.54E-21	4.89E-17	1-4.54E-21	1-4.89E-17	0.999654	0.998282
EAS	CDX	93	9.42E-24	2.99E-18	1-9.42E-24	1-2.99E-18	0.999968	0.999637
	CHB	103	3.59E-24	1.27E-18	1-3.59E-24	1-1.27E-18	0.999981	0.999746
	CHS	105	2.40E-23	3.63E-18	1-2.40E-23	1-3.63E-18	0.999997	0.999908
	JPT	104	8.51E-24	1.78E-18	1-8.51E-24	1-1.78E-18	0.999978	0.999593
	KHV	99	1.85E-23	4.59E-18	1-1.85E-23	1-4.59E-18	0.999991	0.999859
	HNL	279	1.32E-23	3.93E-18	1-1.32E-23	1-3.93E-18	0.999965	0.999549
EUR	CEU	99	3.42E-19	1.96E-16	1-3.42E-19	1-1.96E-16	0.999371	0.998665
	GBR	91	3.69E-19	1.38E-16	1-3.69E-19	1-1.38E-16	0.999037	0.998123
	FIN	99	2.90E-19	2.52E-16	1-2.90E-19	1-2.52E-16	0.999627	0.999160
	IBS	107	1.54E-19	8.72E-17	1-1.54E-19	1-8.72E-17	0.999307	0.998716
	TSI	107	1.92E-19	6.68E-17	1-1.92E-19	1-6.68E-17	0.999220	0.998297
SAS	BEB	86	3.34E-22	6.67E-18	1-3.34E-22	1-6.67E-18	0.999927	0.999521
	GIH	103	8.19E-22	1.23E-17	1-8.19E-22	1-1.23E-17	0.999865	0.999368
	ITU	102	5.27E-22	1.39E-17	1-5.27E-22	1-1.39E-17	0.999855	0.999234
	PJL	96	6.30E-22	1.11E-17	1-6.29E-22	1-1.11E-17	0.999731	0.998879
	STU	102	2.46E-22	7.50E-18	1-2.46E-22	1-7.50E-18	0.999529	0.997665

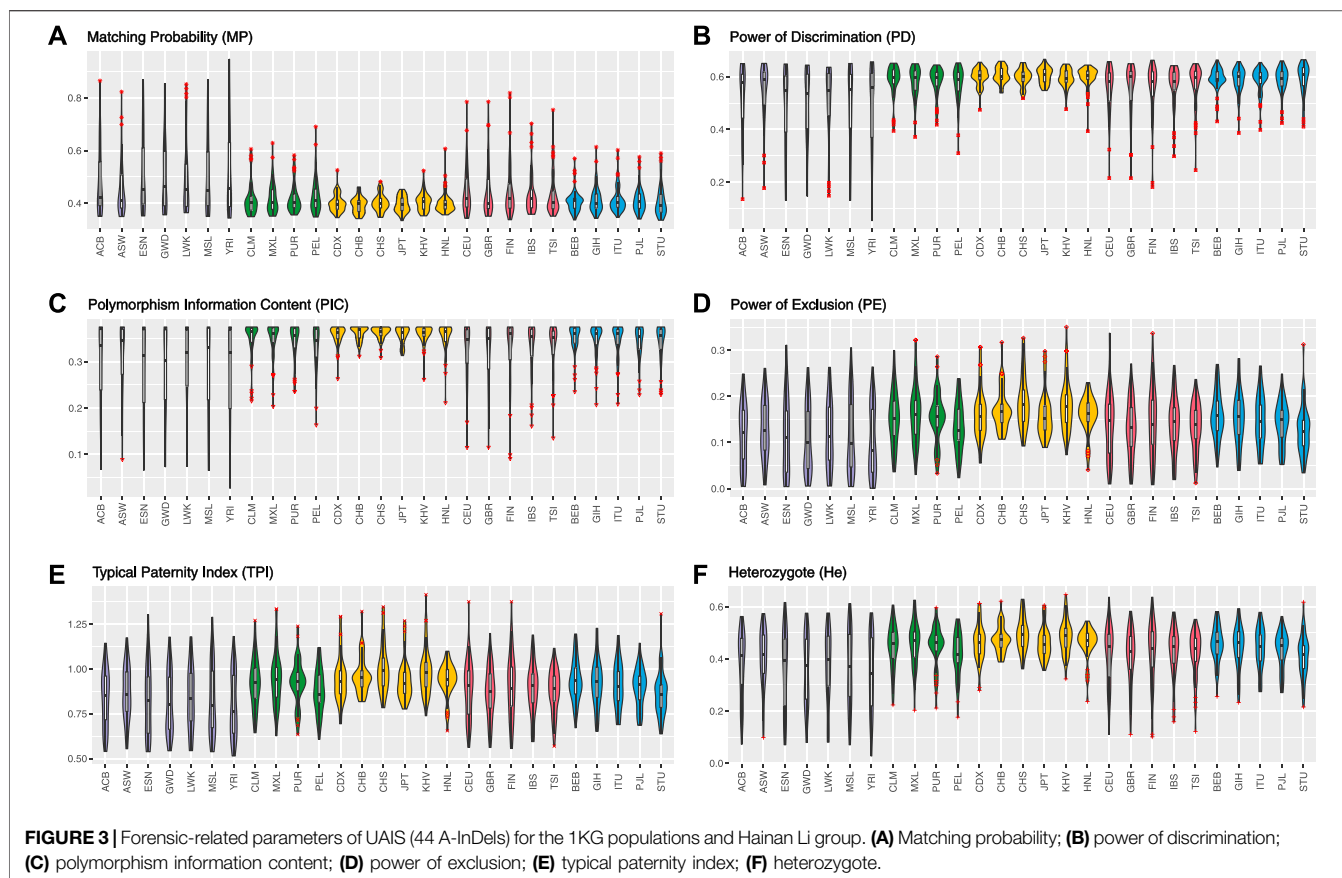
InDels which excluded the 13 A-InDels unconfirmed to HWE) to evaluate the forensic efficiencies for 26 universal 1KG populations and the Hainan Li group.

### Forensic Parameter Evaluations of UAIS (44 A-InDels) in the 1,000 Genomes Project Populations and Hainan Li Group

Allelic frequencies of all 57 A-InDel loci for the 1 KG populations and Hainan Li group are illustrated in **Figure 2A** and **Supplementary Table S4**. For UAIS, including a total of 44 A-InDels conformed by Hardy-Weinberg equilibrium and linkage equilibrium, the insertion allelic frequencies ranged from 0.0934 (GBR, rs561160795) to 0.9861 (YRI, rs79225518). The detailed forensic relevant parameters (MP, PD, PIC, PE, TPI, and He) of UAIS in the 1KG populations and Hainan Li group are calculated and summarized in **Supplementary Tables S5-S10**. In addition, the mean values and standard deviations for all forensic-related parameters in 26 different reference populations from the 1KG and Hainan Li group are shown in **Table 1**. The MP values (Mean  $\pm$  Standard Deviation) of UAIS ranged from  $0.3931 \pm 0.0302$  (CHB) to  $0.5123 \pm 0.1457$  (ESN) (mean MP:  $0.4386 \pm 0.0951$ ). The PIC values spanned from  $0.2816 \pm 0.0972$  (YRI) to  $0.3621 \pm 0.0140$  (CHS) (mean PIC:  $0.3297 \pm 0.0633$ ). The TPI values varied from  $0.8060 \pm 0.1767$  (YRI) to  $1.0049 \pm 0.1351$  (CHS) (mean TPI:  $0.8959 \pm 0.1523$ ).

The average PE value was  $0.1423 \pm 0.0643$  with a range from  $0.1055 \pm 0.0733$  (YRI) to  $0.1883 \pm 0.0571$  (CHS). For He, the mean value was  $0.4243 \pm 0.1071$  ranging from  $0.3486 \pm 0.1444$  (YRI) to  $0.4939 \pm 0.0640$  (CHS). What is more, the cumulative match probability (CMP), cumulative power of discrimination (CPD), and cumulative power of exclusion (CPE) values of UAIS for the 1KG populations and Hainan Li group are illustrated in **Table 2**. The CPD values ranged from  $1-1.03 \times 10^{-14}$  (LWK) to  $1-1.27 \times 10^{-18}$  (CHB), and the CPE varied from 0.993634 (YRI) to 0.999908 (CHS) in the 1KG populations and Hainan Li group.

In total, for UAIS which showed no evidence of deviation from HWE and LDs in both the 1KG populations and Hainan Li group, the forensic-related parameters were distributed relatively balanced (**Figure 3**), revealing that the UAIS had considerable potential in the field of forensic human identification for universal populations. The 44 A-InDels of UAIS possessed relatively reasonable genetic information ( $PIC > 0.25$ ) (Botstein et al., 1980), and the UAIS with the CMP range of  $1.27 \times 10^{-18}$  to  $1.03 \times 10^{-14}$  for 27 universal human populations satisfied the requirements for forensic human identification ( $10^{-15}$ - $10^{-14}$ ) (Pereira et al., 2009), which indicated that the UAIS could be considered as a powerful tool for human identification. Compared with the CPE provided by the common STR panels (Fan et al., 2019a; Fan et al., 2019b; Li et al., 2020), CPE for UAIS has outclassed 0.993634-0.999908. Therefore, the UAIS could supply additional information for the paternity tests.

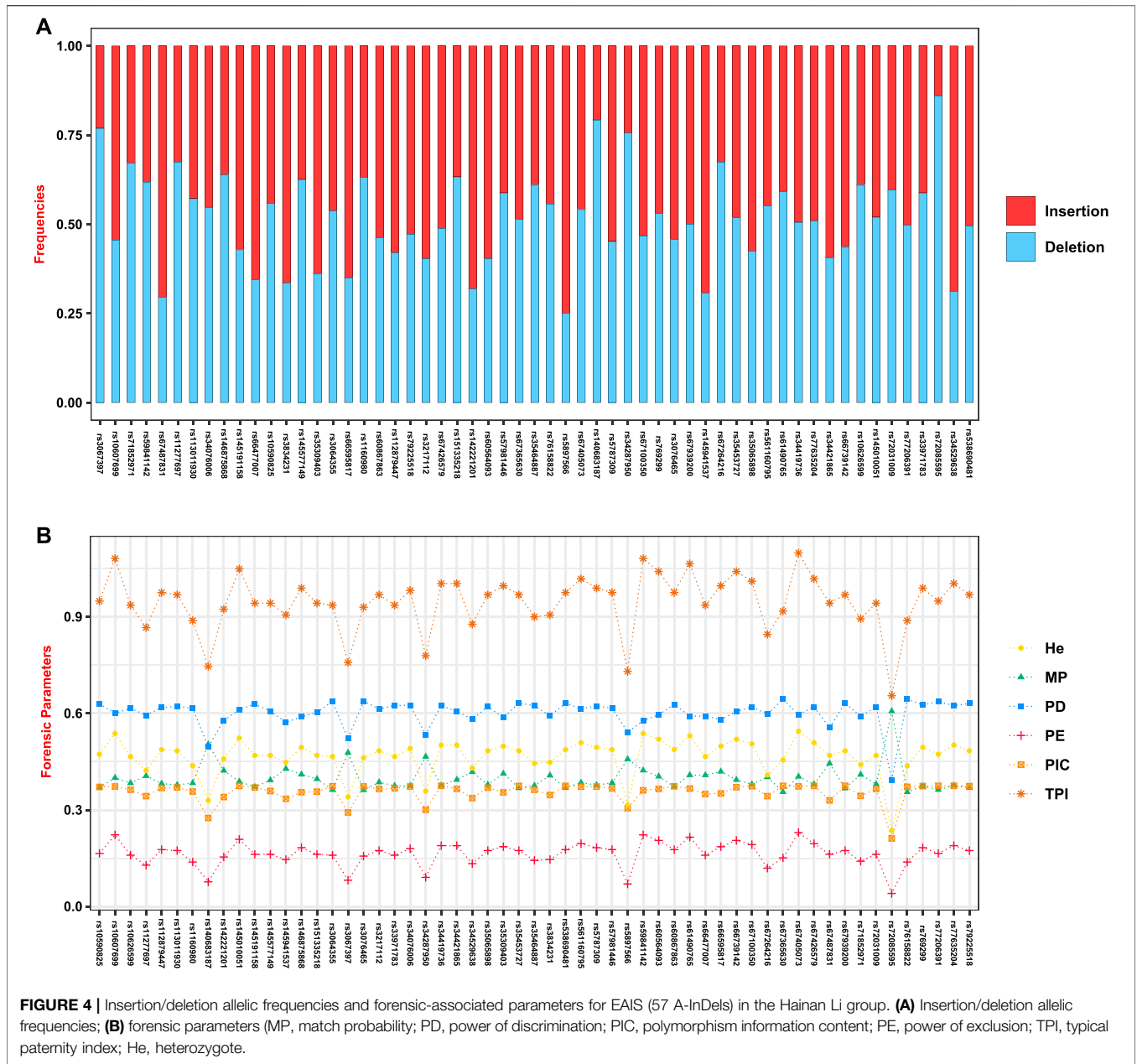


## Forensic Parameter Evaluations of EAIS (57 A-InDels) in the East Asian Populations and Hainan Li Group

The Hainan Li, inhabiting in the south of Hainan island, is a relatively isolated minority group in China, which is beneficial to clarify the exquisite population structure and develop specific genetic markers for subpopulations in the forensic genetic field (Liu et al., 2020b; Fan et al., 2021a). Hence, a total of 279 healthy Hainan Li individuals were collected for forensic evaluations of 57 A-InDel loci in EAIS. The insertion allelic frequencies of the Hainan Li group are demonstrated in **Figures 2A** and **4A** and **Supplementary Table S4**, which were distributed between 0.2079 (rs140683187) and 0.7491 (rs5897566), except for rs72085595 (0.1398). The forensic parameters (MP, PD, PIC, PE, TPI, and He) of the 57 A-InDels in the Hainan Li group are shown in **Supplementary Tables S5–S10** and **Figure 4B**. The MP values of the Hainan Li group ranged from 0.3557 (rs76158822) to 0.6069 (rs72085595) (mean MP:  $0.3986 \pm 0.0409$ ). The PIC values spanned from 0.2116 (rs72085595) to 0.3750 (rs67939200, rs34419736, rs77206391, and rs538690481) (mean PIC:  $0.3563 \pm 0.0291$ ). The TPI values varied from 0.6549 (rs72085595) to 1.0984 (rs67405073) (mean TPI:  $0.9463 \pm 0.0856$ ). The PE values ranged from 0.0405 (rs72085595) to 0.2298 (rs67405073) with an average of  $0.1639 \pm 0.0372$ . The He values varied from 0.2366

(rs72085595) to 0.5448 (rs67405073) (mean He:  $0.4666 \pm 0.0560$ ). Moreover, the CPE and CPD of the 57 A-InDels in the Hainan Li group was 0.999965 and  $1-1.32 \times 10^{-23}$ , which demonstrated that the EAIS have good performances for individual identification and paternity test in the Hainan Li group (**Table 2**). What is more, compared with the results of 47 A-InDels in 216 Hainan Li (Liu et al., 2020a) and 30 A-InDels in 207 Hainan Li (Liu et al., 2019), the majority of 57 A-InDels showed more balanced frequency distributions in the same population (**Figure 4A**). With the number of analyzed A-InDels increased, the CPD and CPE also increased, while the CMP decreased in the Hainan Li group (**Table 3**).

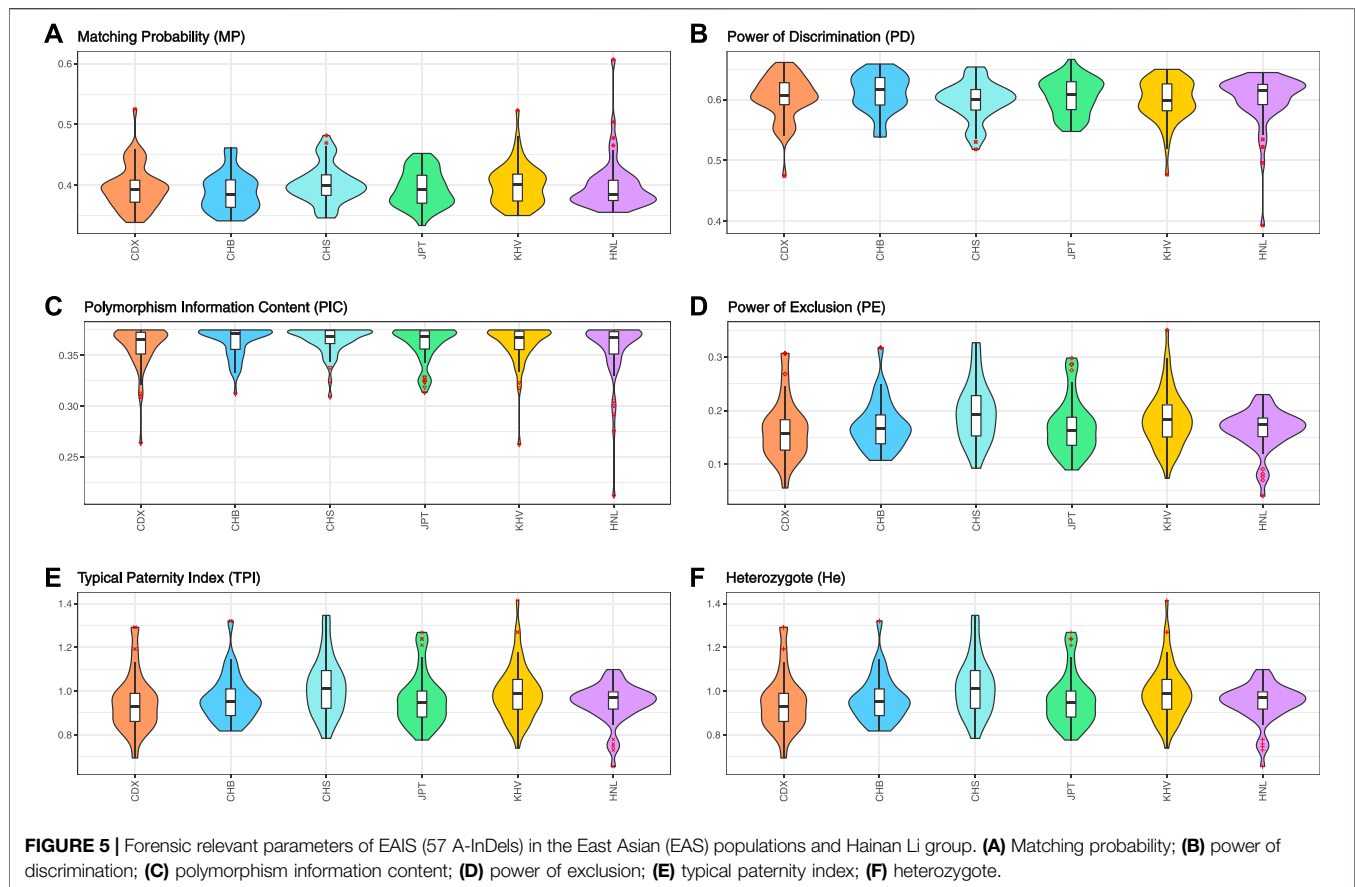
As shown in **Figure 5** and **Table 1**, the forensic-associated parameters of CDX, CHB, CHS, JPT, and KHV populations, and the Hainan Li group are illustrated. The MP values ranged from  $0.3890 \pm 0.0305$  (CHB) to  $0.4022 \pm 0.0308$  (CHS). The PIC values spanned from  $0.3563 \pm 0.0291$  (HNL) to  $0.3646 \pm 0.0132$  (CHS). The TPI values varied from  $0.9463 \pm 0.0856$  (HNL) to  $1.0223 \pm 0.1330$  (CHS). The PE values ranged from  $0.1639 \pm 0.0558$  (CDX) to  $0.1959 \pm 0.0560$  (CHS). The He values spanned from  $0.4631 \pm 0.0701$  (CDX) to  $0.5029 \pm 0.0622$  (CHS). In addition, the CPE varied from 0.999965 (HNL) to 0.999997 (CHS), and the CPD ranged from  $1-1.32 \times 10^{-23}$  (HNL) to  $1-9.42 \times 10^{-24}$  (CHB), respectively. The results revealed that the EAIS has sufficient system effectiveness for human identification and kinship analysis in the EAS populations.



**TABLE 3 |** Comparisons of forensic system efficiencies in different panels with distinct A-InDels for the Hainan Li group (N, number of A-InDel; n, number of population size).

Population	Panel	N	n	CMP	CPD	CPE
Hainan Li	Investigator DIPplex kit	30	207	2.92E-11	1-2.92E-11	0.986100
	AGCU InDel 50 kit	47	216	7.67E-18	1-7.67E-18	0.999283
	UAIS	44	279	3.93E-18	1-3.93E-18	0.999549
	EAIS	57	279	1.32E-23	1-1.32E-23	0.999965





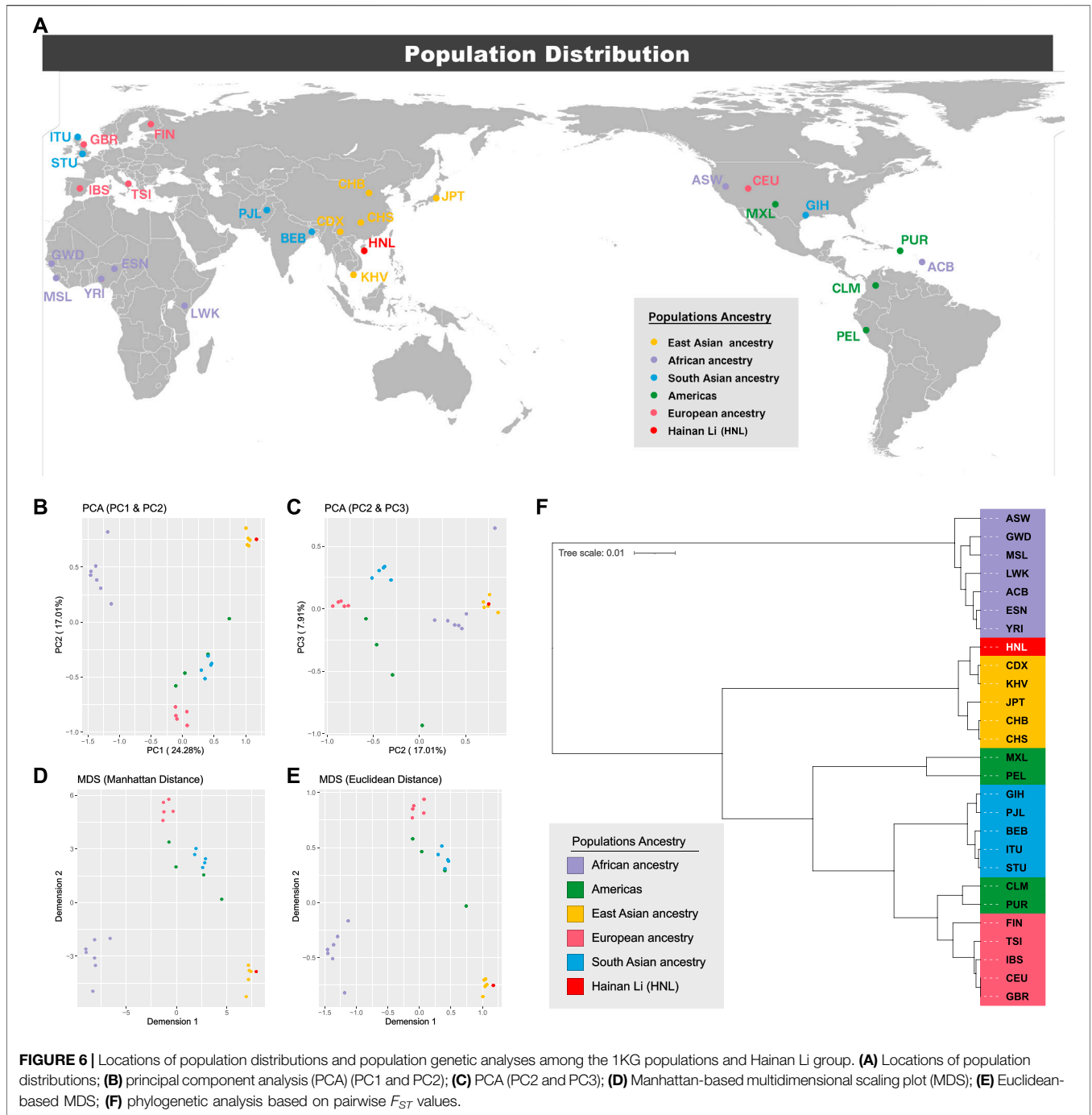
## Population Genetic Analyses Among the Hainan Li Group and 1,000 Genomes Project Populations

To illustrate the genetic landscapes among the 1KG populations and Hainan Li group, the dimensionality reduction analyses (PCA and MDS), which can accelerate the speed of algorithm execution, improve the performance of the analysis model, and reduce the complexity of data at the same time, were conducted based on insertion allelic frequencies of 44 A-InDels, which are illustrated in **Figure 2A** and **Supplementary Table S4**. As shown in **Figures 6B, C**, the first, second, and third components (PC1, PC2, and PC3) accounted for 24.28%, 17.01%, and 7.91% of the total variance observed within these populations, respectively. In the PCA diagrams (**Figures 6B, C**), populations from five different intercontinental ancestries clustered separately, the EAS populations and Hainan Li group clustered together on the upper right. While, the European populations located at the bottom, and the AFR populations distributed on the upper left. In addition, in order to make further confirmation about the genetic relationships between the Hainan Li group and populations from the 1KG conducted by PCA, the Manhattan and Euclidean distance-based MDS were conducted (**Figures 5D, E**), which also depicted the genetic relationships among the Hainan Li group and 1KG populations. The MDS results (**Figures 5D, E**) were in accordance with the genetic patterns of PCAs (**Figures**

**5B, C**). In brief, the dimensionality reduction analyses (PCA and MDS) made relatively clear distinctions, and the Hainan Li group had the close relationships with the EAS populations.

The results of pairwise  $F_{ST}$  and the corresponding  $p$ -values between the Hainan Li group and 26 worldwide populations from different continents are listed in **Supplementary Table S11**. The extreme values of  $F_{ST}$  were identified at CEU and GBR ( $F_{ST}=0.0001$ ,  $p < 0.0001$ ), and GWD and PEL ( $F_{ST}=0.2408$ ,  $p < 0.0001$ ). Phylogenetic relationships between the Hainan Li group and the other 26 reference populations are visualized in the neighbor-joining tree (**Figure 6F**). The EAS populations and Hainan Li group clustered together. For details, the Hainan Li got together with CDX ( $F_{ST}=0.0037$ ,  $p < 0.0001$ ) and KHV ( $F_{ST}=0.0048$ ,  $p < 0.0001$ ), and CHB and CHS clustered together with JPT in another inner branch. They all belong to Southeast Asia from the perspective of geography. The pairwise genetic distances indicated by  $F_{ST}$  values and the phylogenetic relationships based on neighbor-joining tree were consistent with the results of the abovementioned population genetic analyses (PCA and MDS), which manifested that the genetic distances of different populations were consistent with geographic scales in the present study to some degree.

In general, from the perspective of population genetic analyses, compared with paternal Y-STR genetic markers (Fan et al., 2018a; Fan et al., 2018c; Liu et al., 2020b; Ding et al., 2020; Fan et al., 2021b; Fan et al., 2021c; Luo et al., 2021), the novel 6-



dye direct and multiplex PCR-CE-based typing system also possessed the ability to differentiate intercontinental populations to a certain extent. The UAIS enabled to make the relatively clear distinctions among populations from five intercontinental ancestries, and the Hainan Li group had the close genetic relationships with EAS populations.

### CONCLUSION

In conclusion, the direct and multiplex PCR-CE-based typing system was studied based on genome-wide EAS population data, consisting of 57 A-InDels, 2 Y-InDels, and Amelogenin. We collected two A-InDel sets (EAIS and UAIS) according to the

numbers of A-InDels, which confirmed to HWE and evaluated the forensic system effectiveness for each set from the perspectives of EAS and global 1KG populations, respectively. For UAIS (44 A-InDels), the CPD ranged from  $1-1.03 \times 10^{-14}$  to  $1-1.27 \times 10^{-18}$ , and the CPE varied from 0.993634 to 0.999908. For EAIS (57 A-InDels), the ranges of CPD and CPE values were  $1-1.32 \times 10^{-23}$  to  $1-9.42 \times 10^{-24}$ , and 0.999965–0.999997, respectively. In addition, the CPD and CPE values of EAIS for the Hainan Li group were  $1-1.32 \times 10^{-23}$  and 0.999965, respectively. The population genetic analyses clarified the distinctions among the 1KG populations, and the Hainan Li group had close relationships with EAS populations. Consequently, the novel 6-dye direct and multiplex PCR-CE-based typing system should be considered as a reliable and robust tool for human identification and intercontinental population genetics, and supply additional information for kinship analysis in the 1KG populations and Hainan Li group.

## DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Ethics Committee of Xi'an Jiaotong University (No. 2019-1231). The patients/participants provided their written informed consent to participate in this study. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

## REFERENCES

- Abel, H. J., Larson, D. E., Larson, D. E., Regier, A. A., Chiang, C., Das, I., et al. (2020). Mapping and Characterization of Structural Variation in 17,795 Human Genomes. *Nature* 583, 83–89. doi:10.1038/s41586-020-2371-0
- Amoako, K. K., Thomas, M. C., Janzen, T. W., and Goji, N. (2017). Rapid SNP Detection and Genotyping of Bacterial Pathogens by Pyrosequencing. *Methods Mol. Biol.* 1492, 203–220. doi:10.1007/978-1-4939-6442-0\_15
- Bhangale, T. R., Rieder, M. J., Livingston, R. J., and Nickerson, D. A. (2005). Comprehensive Identification and Characterization of Diallelic Insertion-Deletion Polymorphisms in 330 Human Candidate Genes. *Hum. Mol. Genet.* 14, 59–69. doi:10.1093/hmg/ddi006
- Botstein, D., White, R. L., Skolnick, M., and Davis, R. W. (1980). Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms. *Am. J. Hum. Genet.* 32, 314–331.
- Brinkmann, B., Klintschar, M., Neuhuber, F., Hühne, J., and Rolf, B. (1998). Mutation Rate in Human Microsatellites: Influence of the Structure and Length of the Tandem Repeat. *Am. J. Hum. Genet.* 62, 1408–1415. doi:10.1086/301869
- Bus, M. M., Karas, O., and Allen, M. (2016). Multiplex Pyrosequencing of InDel Markers for Forensic DNA Analysis. *Electrophoresis* 37, 3039–3045. doi:10.1002/elps.201600255
- Caputo, M., Amador, M. A., Santos, S., and Corach, D. (2017). Potential Forensic Use of a 33 X-InDel Panel in the Argentinean Population. *Int. J. Leg. Med.* 131, 107–112. doi:10.1007/s00414-016-1399-z

## AUTHOR CONTRIBUTIONS

HF conceptualized the study, performed the formal analysis, wrote the original draft, reviewed and edited the manuscript, and provided the visualization of the study. FW and ZD procured the resources. HF and SL were in charge of the software for the study. YH, SL, YF, and QX performed the investigation. FW conducted the validation. ZD curated the data. PQ and BZ were in charge of the supervision and project administration. FW, ZD, and BZ acquired the funding. All authors have read and agreed to the published version of the manuscript.

## FUNDING

This study was supported by the National Undergraduate Innovation and Entrepreneurship Training Program (Nos. 201911810008 and 201911810023) and the Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (GDUPS, 2017).

## ACKNOWLEDGMENTS

We would like to thank all the volunteers who contributed the samples for this study.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.744645/full#supplementary-material>

- Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y., and Budowle, B. (1999). The Utility of Short Tandem Repeat Loci beyond Human Identification: Implications for Development of New DNA Typing Systems. *Electrophoresis* 20, 1682–1696. doi:10.1002/(sici)1522-2683(19990101)20:8<1682:aid-elps1682>3.0.co;2-z
- Chen, L., Du, W., Wu, W., Yu, A., Pan, X., Feng, P., et al. (2019). Developmental Validation of a Novel Six-Dye Typing System with 47 A-InDels and 2 Y-InDels. *Forensic Sci. Int. Genet.* 40, 64–73. doi:10.1016/j.fsigen.2019.02.009
- Cui, W., Jin, X., Guo, Y., Chen, C., Zhang, W., Kong, T., et al. (2020). Forensic Applicability of Autosomal Insertion/deletion Loci in Chinese Daur Ethnic Group and Genetic Affinity Evaluations between Daur Group and Reference Populations. *Leg. Med.* 47, 101741. doi:10.1016/j.legalmed.2020.101741
- da Costa Francez, P. A., Rodrigues, E. M. R., de Velasco, A. M., and dos Santos, S. E. B. (2012). Insertion-deletion Polymorphisms-Utilization on Forensic Analysis. *Int. J. Leg. Med.* 126, 491–496. doi:10.1007/s00414-011-0588-z
- Ding, J., Fan, H., Zhou, Y., Wang, Z., Wang, X., Song, X., et al. (2020). Genetic Polymorphisms and Phylogenetic Analyses of the Ü-Tsang Tibetan from Lhasa Based on 30 Slowly and Moderately Mutated Y-STR Loci. *Forensic Sci. Res.* 1–8. doi:10.1080/20961790.2020.1810882
- Excoffier, L., and Lischer, H. E. L. (2010). Arlequin Suite Ver 3.5: a New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Mol. Ecol. Resour.* 10, 564–567. doi:10.1111/j.1755-0998.2010.02847.x
- Fan, H., Du, Z., Wang, F., Wang, X., Wen, S.-Q., Wang, L., et al. (2021a). The Forensic Landscape and the Population Genetic Analyses of Hainan Li Based on Massively Parallel Sequencing DNA Profiling. *Int. J. Leg. Med.* 135, 1295–1317. doi:10.1007/s00414-021-02590-3

- Fan, H., Wang, X., Chen, H., Li, W., Wang, W., and Deng, J. (2019a). The Ong Be Language-Speaking Population in Hainan Island: Genetic Diversity, Phylogenetic Characteristics and Reflections on Ethnicity. *Mol. Biol. Rep.* 46, 4095–4103. doi:10.1007/s11033-019-04859-8
- Fan, H., Wang, X., Chen, H., Long, R., Liang, A., Li, W., et al. (2018a). The Evaluation of Forensic Characteristics and the Phylogenetic Analysis of the Ong Be Language-Speaking Population Based on Y-STR. *Forensic Sci. Int. Genet.* 37, e6–e11. doi:10.1016/j.fsigen.2018.09.008
- Fan, H., Wang, X., Chen, H., Zhang, X., Huang, P., Long, R., et al. (2018b). Population Analysis of 27 Y-Chromosomal STRs in the Li Ethnic Minority from Hainan Province, Southernmost China. *Forensic Sci. Int. Genet.* 34, e20–e22. doi:10.1016/j.fsigen.2018.01.007
- Fan, H., Wang, X., Ren, Z., He, G., Long, R., Liang, A., et al. (2019b). Population Data of 19 Autosomal STR Loci in the Li Population from Hainan Province in Southernmost China. *Int. J. Leg. Med.* 133, 429–431. doi:10.1007/s00414-018-1828-2
- Fan, H., Xie, Q., Li, Y., Wang, L., Wen, S.-Q., and Qiu, P. (2021b). Insights into Forensic Features and Genetic Structures of Guangdong Maoming Han Based on 27 Y-STRs. *Front. Genet.* 12, 690504. doi:10.3389/fgene.2021.690504
- Fan, H., Zeng, Y., Wu, W., Liu, H., Xu, Q., Du, W., et al. (2021c). The Y-STR Landscape of Coastal southeastern Han: Forensic Characteristics, Haplotype Analyses, Mutation Rates, and Population Genetics. *Electrophoresis* 42, 1578–1593. doi:10.1002/elps.202100037
- Fan, H., Zhang, X., Wang, X., Ren, Z., Li, W., Long, R., et al. (2018c). Genetic Analysis of 27 Y-STR Loci in Han Population from Hainan Province, Southernmost China. *Forensic Sci. Int. Genet.* 33, e9–e10. doi:10.1016/j.fsigen.2017.12.009
- Genomes Project, C., Abecasis, G. R., Altshuler, D., Auton, A., Brooks, L. D., Durbin, R. M., et al. (2010). A Map of Human Genome Variation from Population-Scale Sequencing. *Nature* 467, 1061–1073. doi:10.1038/nature09534
- Genomes Project, C., Auton, A., Brooks, L. D., Durbin, R. M., Garrison, E. P., Kang, H. M., et al. (2015). A Global Reference for Human Genetic Variation. *Nature* 526, 68–74. doi:10.1038/nature15393
- Golenberg, E. M., Bickel, A., and Weihs, P. (1996). Effect of Highly Fragmented DNA on PCR. *Nucleic Acids Res.* 24, 5026–5033. doi:10.1093/nar/24.24.5026
- Huang, Y., Liu, C., Xiao, C., Chen, X., Yi, S., and Huang, D. (2020). Development of a New 32-plex InDels Panel for Forensic Purpose. *Forensic Sci. Int. Genet.* 44, 102171. doi:10.1016/j.fsigen.2019.102171
- Jin, X. Y., Wei, Y. Y., Cui, W., Chen, C., Guo, Y. X., Zhang, W. Q., et al. (2019). Development of a Novel Multiplex Polymerase Chain Reaction System for Forensic Individual Identification Using Insertion/deletion Polymorphisms. *Electrophoresis* 40, 1691–1698. doi:10.1002/elps.201800412
- Kidd, K. K., Kidd, J. R., Speed, W. C., Fang, R., Furtado, M. R., Hyland, F. C. L., et al. (2012). Expanding Data and Resources for Forensic Use of SNPs in Individual Identification. *Forensic Sci. Int. Genet.* 6, 646–652. doi:10.1016/j.fsigen.2012.02.012
- Kumar, S., Stecher, G., and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Mol. Biol. Evol.* 33, 1870–1874. doi:10.1093/molbev/msw054
- Kwok, P.-Y. (2002). SNP Genotyping with Fluorescence Polarization Detection. *Hum. Mutat.* 19, 315–323. doi:10.1002/humu.10058
- Letunic, I., and Bork, P. (2019). Interactive Tree of Life (iTOL) V4: Recent Updates and New Developments. *Nucleic Acids Res.* 47, W256–W259. doi:10.1093/nar/gkz239
- Li, W., Wang, X., Wang, X., Wang, F., Du, Z., Fu, F., et al. (2020). Forensic Characteristics and Phylogenetic Analyses of One branch of Tai-Kadai Language-Speaking Hainan Hlai (Ha Hlai) via 23 Autosomal STRs Included in the Huaxia™ Platinum System. *Mol. Genet. Genomic Med.* 8, e1462. doi:10.1002/mgg3.1462
- Liu, J., Ye, Z., Wang, Z., Zhou, X., He, G., Mang, M., et al. (2019). Genetic Diversity and Phylogenetic Analysis of Chinese Han and Li Ethnic Populations from Hainan Island by 30 Autosomal Insertion/deletion Polymorphisms. *Forensic Sci. Res.* doi:10.1080/20961790.2019.1672933
- Liu, J., Du, W., Wang, M., Liu, C., Wang, S., He, G., et al. (2020a). Forensic Features, Genetic Diversity and Structure Analysis of Three Chinese Populations Using 47 Autosomal InDels. *Forensic Sci. Int. Genet.* 45, 102227. doi:10.1016/j.fsigen.2019.102227
- Liu, J., Wang, R., Shi, J., Cheng, X., Hao, T., Guo, J., et al. (2020b). The Construction and Application of a New 17-plex Y-STR System Using Universal Fluorescent PCR. *Int. J. Leg. Med.* 134, 2015–2027. doi:10.1007/s00414-020-02291-3
- Luo, C., Duan, L., Li, Y., Xie, Q., Wang, L., Ru, K., et al. (2021). Insights from Y-STRs: Forensic Characteristics, Genetic Affinities, and Linguistic Classifications of Guangdong Hakka and She Groups. *Front. Genet.* 12, 676917. doi:10.3389/fgene.2021.676917
- Matsuda, K. (2017). PCR-based Detection Methods for Single-Nucleotide Polymorphism or Mutation. *Adv. Clin. Chem.* 80, 45–72. doi:10.1016/bs.acc.2016.11.002
- Mills, R. E., Luttig, C. T., Larkins, C. E., Beauchamp, A., Tsui, C., Pittard, W. S., et al. (2006). An Initial Map of Insertion and Deletion (INDEL) Variation in the Human Genome. *Genome Res.* 16, 1182–1190. doi:10.1101/gr.4565806
- Mullaney, J. M., Mills, R. E., Pittard, W. S., and Devine, S. E. (2010). Small Insertions and Deletions (INDELs) in Human Genomes. *Hum. Mol. Genet.* 19, R131–R136. doi:10.1093/hmg/ddq400
- Pakstis, A. J., Speed, W. C., Fang, R., Hyland, F. C. L., Furtado, M. R., Kidd, J. R., et al. (2010). SNPs for a Universal Individual Identification Panel. *Hum. Genet.* 127, 315–324. doi:10.1007/s00439-009-0771-1
- Pakstis, A. J., Speed, W. C., Kidd, J. R., and Kidd, K. K. (2007). Candidate SNPs for a Universal Individual Identification Panel. *Hum. Genet.* 121, 305–317. doi:10.1007/s00439-007-0342-2
- Pereira, R., Phillips, C., Alves, C., Amorim, A., Carracedo, Á., and Gusmão, L. (2009). A New Multiplex for Human Identification Using Insertion/deletion Polymorphisms. *Electrophoresis* 30, 3682–3690. doi:10.1002/elps.200900274
- Saitou, N., and Nei, M. (1987). The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.* 4, 406–425. doi:10.1093/oxfordjournals.molbev.a040454
- Sheng, X., Bao, Y., Zhang, J. S., Li, M., Li, Y. N., Xu, Q. N., et al. (2018). Research Progress on InDel Genetic Marker in Forensic Science. *Fa Yi Xue Za Zhi* 34, 420–427. doi:10.12116/j.issn.1004-5619.2018.04.016
- Song, F., Lang, M., Li, L., Luo, H., and Hou, Y. (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. Genomic Med.* 8, e1224. doi:10.1002/mgg3.1224
- Sun, K., Ye, Y., Luo, T., and Hou, Y. (2016). Multi-InDel Analysis for Ancestry Inference of Sub-populations in China. *Sci. Rep.* 6, 39797. doi:10.1038/srep39797
- Sun, K., Yun, L., Zhang, C., Shao, C., Gao, T., Zhao, Z., et al. (2019). Evaluation of 12 Multi-InDel Markers for Forensic Ancestry Prediction in Asian Populations. *Forensic Sci. Int. Genet.* 43, 102155. doi:10.1016/j.fsigen.2019.102155
- Tao, R., Zhang, J., Sheng, X., Zhang, J., Yang, Z., Chen, C., et al. (2019). Development and Validation of a Multiplex Insertion/deletion Marker Panel, SifaInDel 45plex System. *Forensic Sci. Int. Genet.* 41, 128–136. doi:10.1016/j.fsigen.2019.04.008
- Wang, F., Du, Z., Han, B., Cao, S., Fu, F., Luo, Z., et al. (2021). Genetic Diversity, Forensic Characteristics and Phylogenetic Analysis of the Qiongzong Aborigines Residing in the Tropical Rainforests of Hainan Island via 19 Autosomal STRs. *Ann. Hum. Biol.* 48, 335–342. doi:10.1080/03014460.2021.1951352
- Weber, J. L., David, D., Heil, J., Fan, Y., Zhao, C., and Marth, G. (2002). Human Dialectic Insertion/deletion Polymorphisms. *Am. J. Hum. Genet.* 71, 854–862. doi:10.1086/342727
- Xie, T., Guo, Y., Chen, L., Fang, Y., Tai, Y., Zhou, Y., et al. (2018). A Set of Autosomal Multiple InDel Markers for Forensic Application and Population Genetic Analysis in the Chinese Xinjiang Hui Group. *Forensic Sci. Int. Genet.* 35, 1–8. doi:10.1016/j.fsigen.2018.03.007
- Zhang, H., He, G., Guo, J., Ren, Z., Zhang, H., Wang, Q., et al. (2019). Genetic Diversity, Structure and Forensic Characteristics of Hmong-Mien-speaking Miao Revealed by Autosomal Insertion/deletion Markers. *Mol. Genet. Genomics* 294, 1487–1498. doi:10.1007/s00438-019-01591-7
- Zhang, S., Zhu, Q., Chen, X., Zhao, Y., Zhao, X., Yang, Y., et al. (2018). Forensic Applicability of Multi-Allelic InDels with Mononucleotide Homopolymer Structures. *Electrophoresis* 39, 2136–2143. doi:10.1002/elps.201700468
- Zhang, W., Jin, X., Wang, Y., Kong, T., Cui, W., Chen, C., et al. (2020). Genetic Polymorphisms and Forensic Efficiencies of a Set of Novel Autosomal InDel

Markers in a Chinese Mongolian Group. *Biomed. Res. Int.* 2020, 3925189. doi:10.1155/2020/3925189

Zhao, X., Chen, X., Zhao, Y., Zhang, S., Gao, Z., Yang, Y., et al. (2018). Construction and Forensic Genetic Characterization of 11 Autosomal Haplotypes Consisting of 22 Tri-allelic Indels. *Forensic Sci. Int. Genet.* 34, 71–80. doi:10.1016/j.fsigen.2018.02.001

**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.

The handling editor and the reviewer MW declared a past coauthorship with the authors HF, FW, ZD, YF, PQ, and BZ at the time of review.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Fan, He, Li, Xie, Wang, Du, Fang, Qiu and Zhu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

## GLOSSARY

<b>InDel</b> insertion/deletion	<b>IBS</b> Iberian Populations in Spain
<b>STRs</b> short tandem repeats	<b>TSI</b> Toscani in Italia
<b>SNPs</b> single nucleotide polymorphisms	<b>BEB</b> Bengali in Bangladesh
<b>1KG</b> 1,000 Genomes Project	<b>GIH</b> Gujarati Indians in Houston, TX, United States
<b>AFR</b> African	<b>ITU</b> Indian Telugu in the United Kingdom
<b>AMR</b> Americas	<b>PJL</b> Punjabi in Lahore, Pakistan
<b>EAS</b> East Asian	<b>STU</b> Sri Lankan Tamil in the United Kingdom
<b>EUR</b> European	<b>AMOVA</b> analysis of molecular variance
<b>SAS</b> South Asian	$F_{ST}$ fixation index
<b>HNL</b> Hainan Li	<b>MAF</b> minimum allele frequency
<b>ESN</b> Esan in Nigeria	<b>MP</b> match probability
<b>GWD</b> Gambian in Western Division, Mandinka	<b>PD</b> power of discrimination
<b>LWK</b> , Luhya in Webuye, Kenya	<b>PIC</b> polymorphism information content
<b>MSL</b> Mende in Sierra Leone	<b>PE</b> power of exclusion
<b>YRI</b> Yoruba in Ibadan, Nigeria	<b>TPI</b> typical paternity index
<b>ACB</b> African Caribbean in Barbados	<b>He</b> heterozygote
<b>ASW</b> people with African Ancestry in Southwest United States	<b>CMP</b> cumulative match probability
<b>CLM</b> Colombians in Medellin, Colombia	<b>CPD</b> cumulative power of discrimination
<b>MXL</b> people with Mexican Ancestry in Los Angeles, CA, United States	<b>CPE</b> cumulative power of exclusion
<b>PEL</b> Peruvians in Lima, Peru	<b>HWE</b> Hardy-Weinberg equilibrium
<b>PUR</b> Puerto Ricans in Puerto Rico	<b>LD</b> linkage disequilibrium
<b>CDX</b> Chinese Dai in Xishuangbanna	<b>EAIS</b> East Asia-based A-InDel Set (57 A-InDels)
<b>CHB</b> Han Chinese in Beijing	<b>UAIS</b> Universal A-InDel Set (44 A-InDels)
<b>CHS</b> Southern Han Chinese	<b>PCA</b> principal component analysis
<b>JPT</b> Japanese in Tokyo	<b>MDS</b> multidimensional scaling plot
<b>KHV</b> Kinh in Ho Chi Minh City, Vietnam	<b>CE</b> capillary electrophoresis
<b>CEU</b> Utah residents (CEPH) with Northern and Western European ancestry	<b>MEGA</b> Molecular Evolutionary Genetics Analysis
<b>GBR</b> British in England and Scotland	<b>iTOL</b> Interactive Tree of Life
<b>FIN</b> Finnish in Finland	<b>SAS</b> <sup>®</sup> Statistical Analysis System
	<b>R</b> R project for statistical computing