Human Genomics

RESEARCH

Open Access

Establishment and validation of a DIP panel for forensic ancestry inference and personal identification

Shuanglin Li[1,2], Shuyan Mei[3], Yanfang Liu[4], Wei Cui[1] and Bofeng Zhu[1,5*]

Abstract

Background Biallelic Deletion/Insertion polymorphisms (DIPs), known for their significant diversity across various populations, serve as valuable markers for forensic ancestry inference and personal identification. In this study, we utilized DIPs to provide a potentially powerful forensic examination tool specifically tailored for East Asian populations. Our focus on ancestry allows us to delve deeper into the genetic signatures that characterize this diverse group, offering enhanced resolution in forensic analyses.

Methods A total of 56 autosomal DIPs, 3 Y-chromosome DIPs, and the Amelogenin were selected to build the 60-panel. Population genetic parameters, principal component analysis (PCA), STRUCTURE analysis, and phylogenetic tree construction were employed to evaluate the capacity for ancestry inference. The verification guidelines recommended by the Scientific Working Group on DNA Analysis Methods were followed in the developmental validations of the 60-panel.

Results The PCA, STRUCTURE, and phylogenetic tree constructions were not only consistent with each other but also corroborated by previous research. The combined probability of discrimination and the cumulative probability of paternity exclusion values were 0.999999999999 and 0.9937, respectively. These values indicate that the 60-panel is not only a useful tool for personal identification testing within the East Asian population but also provides valuable biogeographic information. Furthermore, the validation study of the 60-panel, which included assessments of PCR conditions, sensitivity, species specificity, stability, mixture analysis, reproducibility, and case sample studies, as well as analysis of degraded samples, demonstrated that the panel is well-suited for forensic testing. The panel's performance was particularly notable in the analysis of degraded samples, showcasing its potential for use in challenging forensic cases.

Conclusion The newly developed 60-panel demonstrated robust performance in validation tests, yielding reliable genotypes even from poor-quality samples like degraded DNA. It offers valuable biogeographic insights and sufficient polymorphism for personal identification, which assisted forensic examinations in East Asian populations.

Keywords Forensic ancestry inference, Forensic personal identification, DIPs, Validation

*Correspondence:
Bofeng Zhu
zhubofeng7372@126.com
Full list of author information is available at the end of the article

Introduction

Deletion/Insertion Polymorphisms (DIPs), Single Nucleotide Polymorphisms (SNPs), and Short Tandem Repeats (STRs) are DNA markers that serve as essential tools in forensic ancestry inference and personal identification. In the field of forensic ancestry inference, population genetic methods are commonly employed to provide investigative evidence. These methods always utilize ancestry informative marker-DIPs (AIM-DIPs) and ancestry informative marker-SNPs (AIM-SNPs) to draw genetic characteristics. The examination of SNPs depends on the second-generation sequencing platform or SNaPShot technology, which has a high price and complex operation [1]. On the other hand, the DIPs depend on the second-generation sequencing platform or capillary electrophoresis platform (CE), which could overcome the limitations of the SNPs and have the advantages of Short Tandem Repeats (STRs). The mutation rates of DIPs are estimated on the order of 10^{-8}, ensuring their stability in the genome [2]. The DIPs also have no stutter peaks when examined by the PCR-CE platform, ensuring typing accuracy [3]. Considering the characteristics of the DIPs, they could be applied to various forensic specimens in forensic ancestry inference and individual identification.

East Asian populations are of high homogeneity. According to the "Southern coastal route hypothesis", the initial peopling of West Asia, India, Southeast Asia, and East Asia began between roughly 50,000–70,000 years ago. Their descendants later eventually colonized Eurasia [4, 5]. Recently, another theory, the "Northern" route hypothesis, has been proposed based on evidence of archaeology, molecular anthropology, and the geographic status of the Pleistocene. According to the hypothesis, later than the southern dispersal period, modern humans expanded eastward, passing through Central Asia and North Asia, and finally reached East Asia. They interbred with other archaic humans during the dispersal about 30,000–50,000 years ago [6]. After arriving in East Asia, modern humans continued their expansion and interbreeding in the long historical process. To better understand the population stratification and to perform ancestry assignment of East Asians precisely, several assays have been developed: A 39-AIM-InDel panel tested several Chinese groups and provided valuable information [7, 8]. A twelve Multi-InDel markers assay has proved effective in distinguishing Han and Tibetan people [9]; A 49AISNP panel can differentiate three ethnic groups (northern Han, Japanese, and Korean ethnic groups) in the north of East Asia [10]. In the present study, we aimed to explore the ancestry inference capacity of DIP markers and select candidates that could successfully assign populations from the north and south of East Asia.

Finally, a total of 56 autosomal DIPs, 3 Y-Chromosome DIPs, and an Amelogenin marker were selected to construct the forensic ancestry inference assay. The amplicons of the newly established 60-panel were limited to 200 bp to meet the need of typing various forensic specimens, especially the degraded ones. Then, the F_{ST} value, principal component analysis (PCA), STRUCTURE analysis, and phylogenetic tree construction were involved in the ancestry inference ability assessment. The forensic parameters, CPD, and CPE values were used for individual identification ability assessment. Finally, the verification guidelines recommended by the Scientific Working Group on DNA Analysis Methods (SWGDAM) were followed [11] to assess the PCR amplification performance and typing results by CE.

Materials and methods

Candidate selection

DIPs were selected from the 1000 Genomes Project database (http://grch37.ensembl.org) and the Nucleotide Polymorphism Database (https://www.ncbi.nlm.nih.gov/snp). Genomic data of five global populations and five East Asia subgroup populations were employed in the selection and evaluation process. Candidates were subject to the following seven criteria for inclusion:

a. Each marker must have a minimum allele frequency (MAF) of ≥ 0.1;
b. Markers must be bi-allelic ins, del, or delins;
c. The allele length variation of each indel from 1 to 20 bp;
d. Different markers must be located on different chromosomes or chromosomal arms, or be more than 5 Mb apart if on the same chromosomal arm, and they should be independently inherited;
e. The candidate markers should not deviate from Hardy–Weinberg equilibrium (HWE) in Japanese in Tokyo (JPT), Kihn in Ho Chi Minh city (KHV), Chinese Dai in Xishuangbanna (CDX), Han Chinese in Beijing (CHB), and Southern Han Chinese (CHS);
f. The candidates should exhibit allele frequency differences between populations: For East Asian, European, and African populations, the pairwise difference should not be less than 0.5; For South Asian, East Asian, and American populations, the pairwise difference should not be less than 0.3; For European, South Asian, and American populations, the pairwise difference should not less than 0.2; For the 5 subgroups of East Asian, the pairwise difference should not less than 0.2. A candidate locus will be included if it meets just one of these specified conditions;

g. The flanking sequence of the candidate markers should have no polynucleotides, indels, or genetic variations such as SNPs.

Candidates evaluation

Allele frequencies and five forensic statistical parameters in East Asia, including typical paternity index (TPI), power of exclusion (PE), polymorphic information content (PIC), match probability (MP), and power of discrimination (PD) were calculated by STRAF [12]. The combined probability of discrimination (CPD) was conducted based on the formulae: $CPD = 1-(1-DP_1)\ (1-DP_2)\ \ldots\ (1-DP_k)$, and the cumulative probability of paternity exclusion (CPE) was conducted based on the formulae: $CPE = 1-(1-PE_1)\ (1-PE_2)\ \ldots\ (1-PE_k)$. The HWE and the Linkage disequilibrium (LD) were conducted using the GENEPOP 4.0 [13].

To explore population affinities and genetic structures, pairwise F_{ST} was performed using STRAF, and DA distance was calculated using the DISPAN program. Principal component analysis (PCA) based on the 2504 individual genotypes was conducted using ggbiplot of *R* software. The genetic structure was dissected using STRUCTURE v2.3.4 with 10,000 Markov Chain Monte Carlo steps and 10,000 burn-in periods. *K* was set from 2 to 7. The ancestry components were plotted with Distruct 1.1 [14]. Molecular Evolutionary Genetics Analysis Version 7 (Mega 7) was used to discover the patterns of genetic affinities through phylogenetic tree reconstruction with the neighbor-joining method, minimum evolution method, and UPGMA method instrumented in the software [15].

Establishment and validation of the 60-panel

The newly established 6-dye system 60-panel enables to simultaneously type 56 A-DIPs, 3 Y-DIPs, and the Amelogenin loci. The DNA sequences were obtained from GenBank. Primer Premier 5.0 and AutoDimer software were applied in the primer design and evaluation. The ratio of Primer mix, reaction mix, Taq-polymerase, magnesium ion, and DNA Taq polymerase was adjusted to get optimal performance.

Samples

The PCR condition studies were conducted using 1 ng 9948. For the Reaction Mix and the Primer Mix volume test, 9948 was amplified with 0.5×, 0.75×, 1×(recommended), 1.25×, 1.5×Reaction Mix and Primer Mix, respectively; For the reaction volume, 5 µL, 10 µL, 15 µL, 20 µL, and 25 µL PCR premixtures was prepared with the ratios of components remain unchanged in test1, then, 8 µL, 9 µL, 10 µL (recommended), 11 µL, and 12 µL premixtures were prepared in test 2 by adjusting the amount of water; For the denaturation temperature test, 9948 was amplified with 89 °C, 90 °C, 91 °C, 92 °C, 93 °C, 94 °C (recommended), 95 °C, 96 °C, 97 °C, 98 °C, and 99 °C; For the annealing temperature test, 9948 was amplified with 55 °C, 56 °C, 57 °C, 58 °C, 59 °C, 60 °C (recommended), 61 °C, 62 °C, 63 °C, 64 °C, and 65 °C; For the cycling number test, 9948 was amplified with 21, 22, 23, 24, 25 (recommended), 26, and 27 cycles; For the final extension time test, 9948 was amplified with 5 min, 10 min, 15 min, 20 min, 25 min. Each cycle number was tested in triple.

The sensitivity study was conducted using 5 ng, 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, and 31.25 pg male genomic DNA 9948. Each concentration was tested three times.

The species specificity was tested using 10 ng genomic DNA extracted from samples of horse, cow, pig, sheep, dog, cat, chicken, duck, rabbit, mouse, fish, and Escherichia coli (JM109). Negative and positive control 9948 were contained in the study as well.

The stability study was tested using six common inhibitors (Ethanol, calcium ion, tannic acid, EDTA, humic acid, and hematin). The concentration of inhibitors was set as follows: 50%, 60%, 70%, 80%, and 90% Ethanol; 10 mM, 15 mM, 20 mM, 25 mM, and 35 mM calcium ion; 8 mg/mL, 10 mg/mL, 20 mg/mL, 30 mg/mL, and 40 mg/mL tannic acid; 3 mM, 4 mM, 5 mM, 8 mM, and 10 mM EDTA; 1 mol/L, 1.2 mol/L, 1.5 mol/L, 2 mol/L, and 2.5 mol/L humic acid; 4 mM, 5 mM, 6 mM, 7 mM, and 8 mM hematin.

The mixture study was conducted by mixing up 9947A and 9948 at the following ratios: 1:1, 1:4, 1:9, 1:19, 19:1, 9:1, and 4:1.

The reproducibility study employed 3 buccal-indicating FTA® cards and 3 blood FTA® cards, two operators conducted the experiments at the same time.

The case sample study was conducted using one blood on the dagger, two cigarette butts, two epithelial cell abrasions, one costal cartilage bone, one semen, one muscle tissue, and one blood filter. The 9 samples were extracted by ML- DNA Extraction Kit (Bokun Biotech, China).

The degradation study examined mock degraded samples and dated blood samples. The mock degraded samples were prepared by exposing 1 ng genomic DNA 9948 to WD-9430C Ultraviolet instrument (254 nm) for 0 h, 0.5 h, 1.0 h, 1.5 h, and 2.0 h at room temperature. The dated blood filter papers have been preserved for 7 years.

Amplification and capillary electrophoresis

The amplification was performed on the ABI GeneAmp® PCR System 9700 Thermal Cycler (Thermo Fisher Scientific, USA). The standard thermal cycling parameters

were set to initial incubation at 95 ℃ for 5 min; 2 cycles consisting of denaturation at 94 ℃ for 10 s and annealing at 63 ℃ for 90 s, followed by 25 cycles of 94 ℃ for 10 s and 60 ℃ for 90 s; final extension time was 15 min.

The CE was performed on Applied Biosystems®3500 xL Genetic Analyzer (Thermo Fisher Scientific, USA) with the run condition set as follows: injection for 15 s at 1.2 kV, electrophoresis for 24 s at 1.2 kV; run at 60 ℃ for 1210 s. The loading mixture was made of 8.75 μL of the Hi-Di™ Formamide (Thermo Fisher Scientific, USA) and 0.25 μL SIZE-500(S). The PCR product (1 μL) was added to the corresponding well on the CE plate. A standard run condition was set as follows: samples injection for 15 s at 1.2 kV; electrophoresis for 24 s at 1.2 kV; run at 60 ℃ and lasted for 1210 s. The raw data were subsequently analyzed by GeneMapper® ID-X (GMID-X) Software v1.2 (Thermo Fisher Scientific, USA).

Results

Selected DIPs for forensic ancestry inference

The final list of the selected 56 A-DIPs was given in Table 1. The 56 A-DIPs distributed in 20 chromosomes. The corresponding F_{ST} for AFR, AMR, (EUR, EAS, and SAS varied from 0.01 to 0.46, while the value varied from 0 to 0.37 for five EAS subgroups (CDX, CHS, CHB, JPT, and KHV) [16]. Moreover, 3 Y-chromosome DIPs (rs771783753, rs759551978, and rs2032678) and the Amelogenin loci were added in the final panel to assist examiners to determine the sex of samples. F_{ST} calculations, pairwise δ calculations, HWE test, LD test, forensic parameter calculations, PCA, STRUCTURE analysis, and phylogenetic tree constructions were applied in the candidate selection and evaluation process.

The δ value is the difference of the allele frequency at a specific locus, which can be used to evaluate the difference of a locus between different populations [17]. When the screening value was 0.5, 25 loci and 13 loci can be used to differentiate AFR-EAS and EUR-EAS, respectively. When the screening value was 0.3, 24 loci and 21 loci can be used to differentiate EAS-AMR and EAS-SAS, respectively. For subgroups of East Asia, when the δ value was set great than 0.2, a total of 11 loci met the criteria. Among them, rs35171885 showed a distinguishing ability to differentiate all five subgroups (Fig. 1).

The HWE and LD was tested in the CDX, CHS, CHB, JPT, and KHV groups, respectively. When the significance level of the p-value set at 0.05, deviations from HWE were observed at rs10555216 ($p=0.0005$), rs2067350 ($p=0.0358$), and rs66715534 ($p=0.0085$) in the CDX ethnic group; at rs35171885 ($p=0.0138$), and rs56038358 ($p=0.0214$) in the CHB group; at rs56038358 ($p=0.0185$) in the KHV group. After applying the Bonferroni correction ($p=0.0009$) [18], only rs10555216 was

Table 1 The information of 56 A-DIPs

Loci	dbSNP rs#	Build 37 nt position	F_{ST} for all groups	F_{ST} for EAS groups
B1	rs34477782	Chr3: 194506203–194506204	0.25	0.00
B2	rs4647655	Chr4: 185557899–185557909	0.30	0.01
B3	rs35323999	Chr2: 212541539–212541542	0.23	0.01
B4	rs66715534	Chr6: 32576385–32576392	0.01	0.06
B5	rs2067350	Chr5: 33968753–33968754	0.17	0.02
B6	rs141022607	Chr2: 187695430–187695432	0.21	0.00
B7	rs3831885	Chr5: 109133010–109133011	0.28	0.00
B8	rs33948716	Chr4: 123774811–123774815	0.39	0.00
B9	rs34891898	Chr2: 153227885–153227886	0.24	0.00
B10	rs3988320	Chr8: 18908781–18908782	0.20	0.00
B11	rs3032990	Chr10: 118830844–118830849	0.16	0.00
B12	rs4147539	Chr4: 100237480–100237487	0.10	0.01
B13	rs10555216	Chr8: 27207896–27207903	0.16	0.01
B14	rs3029066	Chr1: 6396827–6396828	0.44	0.04
G1	rs3045215	Chr1: 234740918	0.26	0.02
G2	rs10560368	Chr9: 38440395–38440396	0.16	0.00
G3	rs56038358	Chr2: 77314047–77314049	0.19	0.03
G4	rs76533131	Chr9: 34261853–34261854	0.23	0.00
G5	rs3028822	Chr5: 172575763–172575771	0.12	0.01
G7	rs11273905	Chr11: 46085383–46085393	0.17	0.02
G8	rs146391383	Chr7: 138207599–138207606	0.19	0.00
G10	rs16432	Chr20: 2128698–2128706	0.33	0.00
G11	rs34024438	Chr13: 92421048–92421049	0.24	0.00
G12	rs5891435	Chr8: 52749044–52749051	0.46	0.01
G13	rs10580743	Chr9: 139791625–139791632	0.08	0.00
G14	rs10538061	Chr10: 74572523–74572525	0.21	0.01
Y1	rs35434967	Chr2: 133216470–133216476	0.15	0.01
Y2	rs5860664	Chr4: 102157449–102157450	0.21	0.04
Y4	rs34618829	Chr6: 121477884–121477887	0.37	0.00
Y5	rs34383321	Chr11: 10383323–10383324	0.17	0.00
Y6	rs66471060	Chr4: 118990299–118990302	0.36	0.02
Y7	rs67853165	Chr15: 27990090	0.14	0.00
Y8	rs57406754	Chr9: 133617278–133617281	0.23	0.00
Y9	rs68108181	Chr18: 77344953–77344955	0.34	0.01
Y10	rs35171885	Chr14: 106091744	0.20	0.37
Y11	rs67185986	Chr3: 129519507–129519510	0.36	0.03
R1	rs3840222	Chr3: 4356252–4356265	0.40	0.01
R2	rs148177611	Chr12: 111390455–111390460	0.16	0.04
R3	rs34530228	Chr7: 134761552–134761553	0.16	0.01
R4	rs71997819	Chr11: 60567526–60567530	0.26	0.00
R5	rs10534050	Chr15: 99343470–99343471	0.17	0.02
R6	rs10569275	Chr10: 11336961–11336962	0.31	0.00
R7	rs10533245	Chr17: 70347822–70347826	0.28	0.00
R8	rs61044006	Chr13: 46370078–46370079	0.22	0.01
R9	rs3842715	Chr22: 36708050–36708057	0.25	0.00
R10	rs5868949	Chr5: 78364989–78364990	0.25	0.02
R11	rs3835409	Chr1: 29633492–29633502	0.18	0.00

Table 1 (continued)

Loci	dbSNP rs#	Build 37 nt position	F_{ST} for all groups	F_{ST} for EAS groups
P1	rs2307840	Chr1: 36099089–36099093	0.38	0.00
P2	rs138324	Chr22: 41210550–41210552	0.16	0.01
P3	rs5840834	Chr20: 19190654–19190655	0.25	0.00
P4	rs35794676	Chr14: 92786986–92786990	0.11	0.04
P5	rs3044252	Chr20: 55044840–55044845	0.19	0.00
P7	rs5896844	Chr9: 19564552–19564563	0.27	0.01
P8	rs71960622	Chr9: 125966268–125966286	0.27	0.04
P9	rs3840794	Chr12: 1724156–1724174	0.15	0.01
P10	rs3033760	Chr3: 173557661–173557671	0.16	0.02

found to be out of HWE in the CDX group. No deviations were found in other four East Asian groups (Supplementary Table s1). Although the rs10555216 locus did not adhere to the HWE criteria within the CDX group, its value in the CHB, CHS, JPT, and KHV populations were 0.4010, 0.4416, 0.9494, and 0.0508, respectively. These values suggest a generally good performance. Further testing in additional populations, including the Han Chinese from Guizhou, Shaanxi Han [19], and the Manchu group from Inner Mongolia [20], also showed no evidence of HWE imbalance. Consequently, we have opted to retain this locus for continued analysis. The LD was carried out by Genepop and the results showed that they are independently inherited (Supplementary Table s2).

The forensic parameters were calculated based on the genotypes of 504 East Asian individuals (Supplementary Table s3). The CPE and the CPD values were 0.9937 and 0.999 999 999 999, respectively, similar to the Investigator® DIPplex kit (Qiagen, Hilden, Germany) [21], suggesting the 60-panel could also provide a powerful supplement for personal identification testing.

(A)

δ	AFR AMR	AFR EAS	AFR EUR	AFR SAS	AMR EAS	AMR EUR	AMR SAS	EAS EUR	EAS SAS	EUR SAS
rs10533245	0.22	0.61	0.45	0.55	0.39	0.23	0.33	0.16	0.06	0.10
rs10534050	0.44	0.43	0.37	0.21	0.01	0.07	0.23	0.06	0.22	0.16
rs10538061	0.00	0.47	0.13	0.10	0.47	0.13	0.10	0.60	0.37	0.23
rs10555216	0.15	0.44	0.00	0.08	0.30	0.14	0.06	0.44	0.36	0.08
rs10560368	0.30	0.49	0.07	0.17	0.19	0.23	0.14	0.42	0.33	0.10
rs10569275	0.48	0.59	0.49	0.55	0.11	0.02	0.07	0.09	0.04	0.06
rs10580743	0.30	0.00	0.10	0.06	0.30	0.20	0.24	0.10	0.06	0.04
rs11273905	0.14	0.11	0.38	0.28	0.25	0.24	0.14	0.49	0.39	0.10
rs138324	0.32	0.26	0.12	0.11	0.58	0.20	0.21	0.38	0.37	0.01
rs141022607	0.37	0.60	0.36	0.32	0.23	0.02	0.05	0.25	0.28	0.03
rs146391383	0.19	0.49	0.01	0.21	0.30	0.18	0.03	0.48	0.28	0.21
rs148177611	0.00	0.16	0.00	0.00	0.16	0.00	0.00	0.16	0.16	0.00
rs16432	0.23	0.68	0.03	0.31	0.45	0.26	0.08	0.71	0.37	0.34
rs2067350	0.08	0.34	0.15	0.17	0.42	0.07	0.25	0.50	0.17	0.33
rs2307840	0.36	0.04	0.68	0.45	0.40	0.33	0.09	0.72	0.49	0.23
rs3028822	0.11	0.15	0.31	0.01	0.26	0.20	0.10	0.45	0.16	0.29
rs3029066	0.67	0.72	0.68	0.64	0.05	0.01	0.03	0.03	0.08	0.05
rs3032990	0.27	0.05	0.43	0.37	0.22	0.16	0.09	0.38	0.32	0.06
rs3033760	0.05	0.41	0.09	0.12	0.46	0.04	0.17	0.50	0.29	0.21
rs3044252	0.11	0.49	0.07	0.06	0.39	0.18	0.05	0.56	0.44	0.12
rs3045215	0.32	0.60	0.10	0.42	0.29	0.22	0.10	0.51	0.19	0.32
rs33948716	0.54	0.77	0.59	0.56	0.23	0.05	0.02	0.18	0.21	0.03
rs34024438	0.22	0.53	0.31	0.56	0.31	0.09	0.34	0.22	0.04	0.26
rs34383321	0.14	0.48	0.04	0.09	0.35	0.17	0.05	0.52	0.39	0.13
rs34477782	0.28	0.20	0.58	0.41	0.08	0.30	0.12	0.38	0.20	0.18
rs34530228	0.20	0.47	0.23	0.00	0.27	0.02	0.20	0.25	0.47	0.22
rs34618829	0.35	0.54	0.69	0.67	0.18	0.34	0.31	0.16	0.13	0.03
rs34891898	0.05	0.50	0.15	0.12	0.55	0.10	0.17	0.64	0.38	0.26
rs35171885	0.18	0.50	0.26	0.23	0.32	0.08	0.05	0.24	0.27	0.03
rs35323999	0.16	0.63	0.17	0.27	0.47	0.01	0.11	0.46	0.36	0.10
rs35434967	0.13	0.14	0.35	0.05	0.28	0.22	0.09	0.49	0.19	0.31
rs35794676	0.16	0.31	0.37	0.32	0.15	0.21	0.16	0.06	0.01	0.05
rs3831885	0.19	0.59	0.09	0.33	0.39	0.10	0.14	0.50	0.25	0.25
rs3835409	0.21	0.53	0.13	0.33	0.32	0.07	0.13	0.39	0.19	0.20
rs3840222	0.58	0.77	0.52	0.64	0.19	0.06	0.06	0.25	0.13	0.12
rs3840794	0.04	0.37	0.16	0.03	0.33	0.20	0.00	0.53	0.33	0.20
rs3842715	0.17	0.57	0.17	0.40	0.40	0.00	0.24	0.40	0.16	0.24
rs3988320	0.29	0.03	0.44	0.39	0.26	0.15	0.11	0.41	0.36	0.05
rs4147539	0.25	0.02	0.01	0.03	0.23	0.24	0.28	0.00	0.05	0.05
rs4647655	0.21	0.61	0.00	0.20	0.40	0.20	0.00	0.60	0.40	0.20
rs56038358	0.40	0.13	0.51	0.37	0.27	0.11	0.03	0.38	0.24	0.14
rs57406754	0.38	0.22	0.61	0.42	0.17	0.23	0.04	0.40	0.21	0.19
rs5840834	0.22	0.57	0.03	0.08	0.35	0.19	0.14	0.55	0.50	0.05
rs5860664	0.29	0.11	0.57	0.33	0.18	0.28	0.03	0.46	0.22	0.24
rs5868949	0.50	0.59	0.45	0.41	0.10	0.05	0.09	0.14	0.18	0.04
rs5891435	0.68	0.54	0.79	0.73	0.14	0.11	0.05	0.25	0.19	0.06
rs5896844	0.35	0.52	0.31	0.46	0.17	0.04	0.12	0.21	0.05	0.16
rs61044006	0.34	0.60	0.16	0.27	0.27	0.18	0.07	0.45	0.34	0.11
rs66471060	0.65	0.66	0.59	0.60	0.01	0.07	0.05	0.07	0.06	0.01
rs66715534	0.00	0.03	0.06	0.09	0.03	0.06	0.09	0.03	0.06	0.04
rs67185986	0.54	0.27	0.70	0.65	0.28	0.15	0.10	0.43	0.38	0.05
rs67853165	0.22	0.18	0.24	0.05	0.40	0.01	0.27	0.41	0.12	0.29
rs68108181	0.40	0.74	0.50	0.58	0.34	0.10	0.17	0.24	0.16	0.08
rs71960622	0.24	0.04	0.49	0.47	0.28	0.25	0.23	0.54	0.52	0.02
rs71997819	0.29	0.62	0.26	0.53	0.33	0.04	0.24	0.37	0.09	0.27
rs76533131	0.13	0.57	0.02	0.23	0.44	0.15	0.10	0.59	0.34	0.25

(B)

δ	CDX CHB	CDX CHS	CDX JPT	CDX KHV	CHB CHS	CHB JPT	CHB KHV	CHS JPT	CHS KHV	JPT KHV
rs10533245	0.11	0.06	0.08	0.07	0.06	0.03	0.04	0.02	0.02	0.00
rs10534050	0.04	0.03	0.17	0.01	0.01	0.13	0.05	0.14	0.04	0.18
rs10538061	0.01	0.02	0.05	0.01	0.01	0.04	0.00	0.02	0.01	0.04
rs10555216	0.06	0.03	0.12	0.01	0.09	0.05	0.07	0.15	0.02	0.12
rs10560368	0.03	0.06	0.06	0.00	0.03	0.03	0.02	0.00	0.05	0.05
rs10569275	0.03	0.04	0.04	0.03	0.01	0.02	0.01	0.00	0.01	0.01
rs10580743	0.00	0.01	0.05	0.01	0.01	0.05	0.02	0.04	0.00	0.03
rs11273905	0.06	0.00	0.16	0.01	0.06	0.10	0.07	0.16	0.01	0.17
rs138324	0.04	0.07	0.11	0.06	0.04	0.07	0.02	0.04	0.02	0.05
rs141022607	0.01	0.00	0.07	0.03	0.01	0.06	0.03	0.07	0.04	0.03
rs146391383	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.01	0.01	0.01
rs148177611	0.12	0.22	0.12	0.09	0.11	0.00	0.02	0.11	0.13	0.02
rs16432	0.00	0.02	0.03	0.00	0.02	0.03	0.00	0.04	0.02	0.03
rs2067350	0.01	0.06	0.14	0.00	0.05	0.15	0.01	0.20	0.05	0.14
rs2307840	0.08	0.04	0.03	0.01	0.03	0.05	0.09	0.01	0.05	0.04
rs3028822	0.09	0.03	0.01	0.10	0.07	0.10	0.01	0.04	0.08	0.11
rs3029066	0.09	0.09	0.20	0.05	0.00	0.11	0.04	0.12	0.04	0.15
rs3032990	0.06	0.02	0.03	0.02	0.08	0.03	0.04	0.05	0.04	0.02
rs3033760	0.00	0.03	0.08	0.09	0.03	0.08	0.09	0.11	0.06	0.16
rs3044252	0.03	0.04	0.01	0.01	0.07	0.03	0.04	0.05	0.03	0.02
rs3045215	0.15	0.01	0.12	0.05	0.14	0.03	0.10	0.11	0.04	0.07
rs33948716	0.03	0.03	0.01	0.01	0.00	0.02	0.04	0.02	0.04	0.02
rs34024438	0.07	0.02	0.08	0.08	0.05	0.01	0.01	0.06	0.06	0.00
rs34383321	0.01	0.00	0.03	0.03	0.01	0.03	0.04	0.02	0.03	0.00
rs34477782	0.01	0.01	0.06	0.06	0.01	0.05	0.05	0.05	0.05	0.00
rs34530228	0.03	0.01	0.09	0.02	0.04	0.06	0.05	0.10	0.01	0.11
rs34618829	0.06	0.05	0.02	0.03	0.01	0.08	0.03	0.07	0.02	0.05
rs34891898	0.04	0.02	0.04	0.01	0.03	0.01	0.06	0.02	0.03	0.05
rs35171885	0.57	0.34	0.81	0.20	0.23	0.24	0.38	0.47	0.15	0.62
rs35323999	0.02	0.02	0.05	0.03	0.00	0.03	0.05	0.04	0.05	0.09
rs35434967	0.08	0.06	0.05	0.01	0.03	0.03	0.09	0.01	0.07	0.06
rs35794676	0.03	0.01	0.21	0.02	0.03	0.24	0.01	0.21	0.03	0.24
rs3831885	0.02	0.02	0.06	0.06	0.00	0.04	0.08	0.03	0.08	0.12
rs3835409	0.02	0.01	0.07	0.08	0.01	0.05	0.06	0.06	0.07	0.01
rs3840222	0.02	0.03	0.03	0.05	0.04	0.02	0.03	0.06	0.08	0.02
rs3840794	0.03	0.01	0.11	0.01	0.04	0.08	0.02	0.12	0.02	0.10
rs3842715	0.07	0.01	0.04	0.07	0.06	0.03	0.00	0.03	0.06	0.03
rs3988320	0.04	0.00	0.01	0.03	0.04	0.03	0.07	0.01	0.03	0.04
rs4147539	0.06	0.01	0.01	0.02	0.07	0.05	0.07	0.02	0.01	0.02
rs4647655	0.14	0.13	0.15	0.08	0.01	0.01	0.06	0.02	0.05	0.07
rs56038358	0.09	0.13	0.24	0.05	0.04	0.15	0.04	0.11	0.08	0.19
rs57406754	0.05	0.00	0.07	0.03	0.05	0.03	0.01	0.08	0.04	0.04
rs5840834	0.02	0.03	0.00	0.03	0.05	0.03	0.00	0.03	0.05	0.03
rs5860664	0.15	0.10	0.20	0.01	0.04	0.05	0.15	0.10	0.11	0.21
rs5868949	0.09	0.10	0.21	0.06	0.01	0.12	0.03	0.11	0.04	0.15
rs5891435	0.13	0.10	0.17	0.08	0.03	0.04	0.05	0.07	0.02	0.09
rs5896844	0.00	0.01	0.00	0.02	0.01	0.00	0.02	0.01	0.01	0.02
rs61044006	0.03	0.00	0.04	0.01	0.03	0.07	0.04	0.04	0.01	0.03
rs66471060	0.11	0.08	0.20	0.09	0.03	0.09	0.02	0.12	0.01	0.11
rs66715534	0.24	0.22	0.15	0.28	0.02	0.09	0.05	0.07	0.06	0.13
rs67185986	0.23	0.12	0.22	0.10	0.11	0.01	0.14	0.10	0.02	0.13
rs67853165	0.02	0.01	0.01	0.04	0.01	0.02	0.02	0.02	0.03	0.05
rs68108181	0.08	0.11	0.13	0.11	0.04	0.05	0.03	0.01	0.01	0.02
rs71960622	0.13	0.08	0.27	0.06	0.05	0.14	0.07	0.19	0.02	0.21
rs71997819	0.06	0.08	0.01	0.01	0.03	0.04	0.06	0.07	0.09	0.02
rs76533131	0.07	0.05	0.04	0.02	0.02	0.03	0.05	0.01	0.02	0.02

Legend
>0.5
0.3-0.5
0.2-0.3
0.1-0.2
<0.1

Fig. 1 The δ value of 56 A-DIPs: **A** The δ value between different continent populations; **B** The δ value between 5 subgroups of East Asian

PCA is a dimensionality reduction technique and groups samples together based on their similarity. Geographically close populations often mix genetically and this leaves a clear signal in PCA plots. In the present study, the individuals of the 26 worldwide populations were employed in the test, and a point in the PCA plot represented one person. The proportions of the top two principal components PC1 and PC2 were shown in the PCA plot. At the population level, whether it's intercontinental groups or intra-East Asian groups, the application of the 60-panel can achieve a high degree of discrimination (Fig. 2A, B). When the analysis included 661 African individuals, 504 East Asian individuals, and 503 European individuals, the PCA results demonstrated that Africans, East Asians, and Europeans formed well-defined clusters. Notably, the first two principal components accounted for 41% of the total genetic variation (Fig. 2C). However, when the analysis was expanded to include 347 American individuals and 489 South Asian individuals, some overlap was observed among the European (EUR), South Asian (SAS), and American (AMR) data points. This overlap is likely due to the lower resolution of these groups relative to the more distinct clusters formed by the three main continental groups (Fig. 2D). Further analysis focusing solely on the EUR, SAS, and AMR individuals revealed that they indeed formed three separate clusters. In this case, the first two principal components explained a smaller proportion of the variation, at 13.3%. The PCA results also indicated that the genetic background of AMR individuals is complex in terms of their dispersion (Fig. 2E). For the East Asian populations, the CDX individuals and JPT individuals were analyzed and a small number of individuals have overlapped (Fig. 2F). Then, we merged the CHS and the CHB individuals into

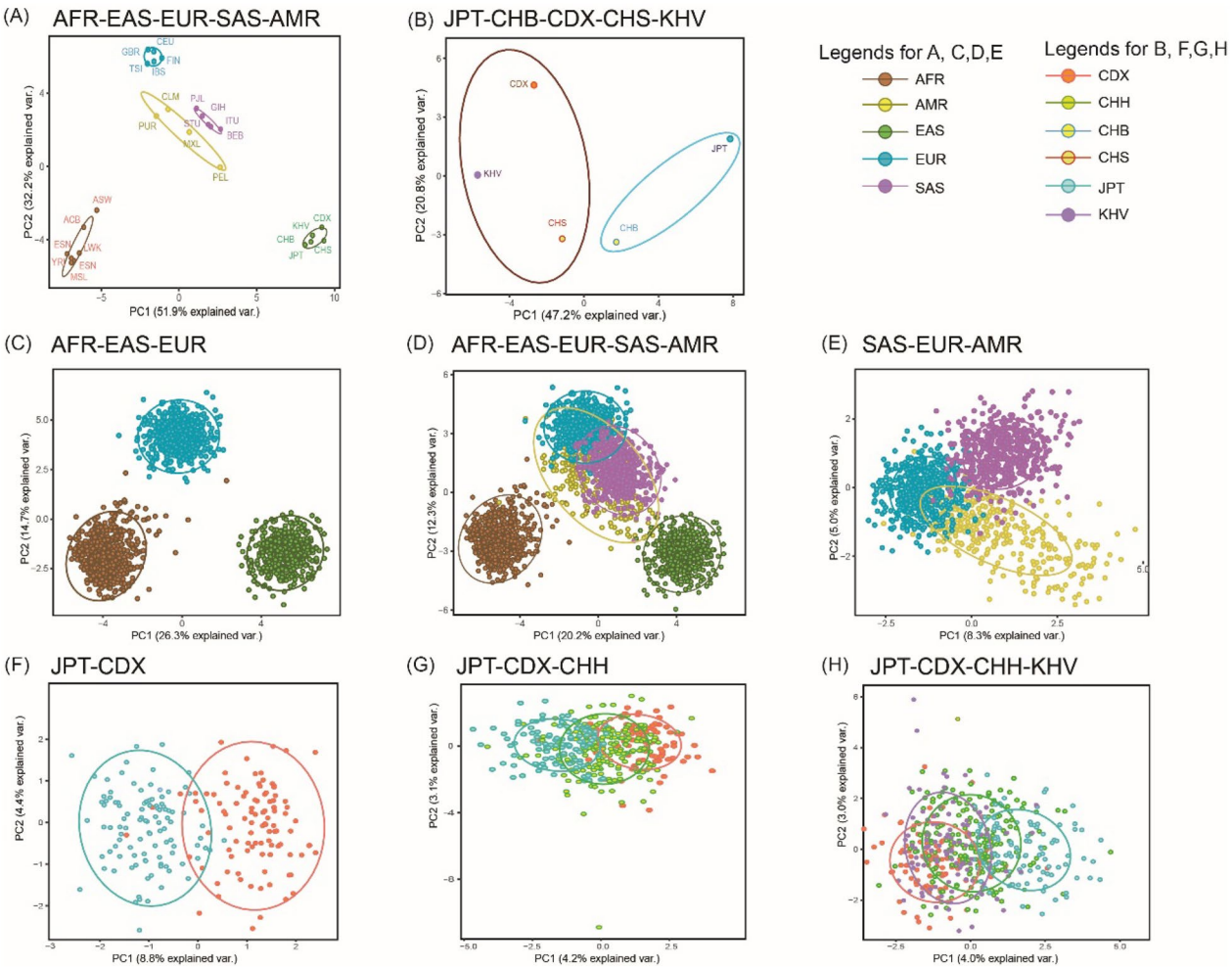

Fig. 2 The PCA plot revealed by the frequencies of 56 A-DIPs: **A** PCA Plot of 26 Populations from the 1000 Genomes Project; **B** PCA plot of 5 East Asian populations; The PCA plot revealed by the genotype of 56 A-DIPs: **C** PCA plot of AFR, EAS, and EUR populations; **D** PCA plot of AFR, EAS, EUR, AMR and SAS populations; **E** PCA plot of EUR, SAS, and AMR populations; **F** PCA plot of JPT, and CDX; **G** PCA plot of JPT, CHH, and CDX populations; **H** PCA plot of JPT, CDX, CHH, and KHV populations

the CHH group in the analysis. As expected, the CHH group was located between the CDX group and the JPT group, and had a certain degree of overlap with the two groups (Fig. 2G). When KHV individuals participated in the analysis, they partly coincided with the CHH groups, partly with CDX groups (Fig. 2H). For the 3 Chinese groups, the plots showed that the CHB, CHS, and CDX individuals harbor subtle differences along a North-to-South gradient (Supplementary Fig. s1).

STRUCTURE analysis used Bayesian iterative algorithm to infer the genetic structure. First, 26 worldwide populations were analyzed with the putative "ancestral populations" from 2 to 7. For $K=2$, AFR and other continental populations were separated, the results can be interpreted by the important demographic event of out-of-Africa bottleneck; For $K=3$, continental groups of AFR, EUR, and EAS had been separated; For $K=5$, all five continental groups have been separated. The Peruvian in Lima (PEL) and Mexican Ancestry in Los Angeles (MXL) from AMR showed AMR specific component, while Puerto Rican in Puerto Rico (PUR) and Columbians in Medellin (CLM) showed more EUR related component (Fig. 3A). Then, 5 East Asian subgroups were analyzed with the putative "ancestral populations" from 2 to 7. For all values, the JPT group can be distinguished from other 4 groups. For $K=5$, the dominant genetic component of JPT group was labelled blue, representing the Ancestral North EAS component; the dominant genetic component of CDX was labelled brown, representing the Ancestral South EAS component. The CHB, CHS and KHV groups harbored both components and a north–south cline was visible (Fig. 3B, Supplementary Fig. s2).

The neighbor-joining method, minimum evolution method, and UPGMA method were employed to reconstruct the phylogenetic tree (Fig. 3C, Supplementary Fig. s3). The results of the three methods confirmed each other and were consistent with the known morphology. The four main clusters in the phylogenetic trees were the East Asian, South Asian, European, and African clusters. Four American populations fell into two groups. The PUR and CLM were close to the EUR while the MXL and the PEL were close to the SAS.

Validation of the 60-panel

The PCR condition studies

Firstly, we tested the PCR conditions with each single locus. The initial typing results showed that non-specific peaks were detected at rs3045215, rs146391383, and rs34383321 (Supplementary Fig. s4). By adjusting the primers and the corresponding annealing temperature, the problems were solved. In the subsequent multiplex amplification of the 60 markers, a satisfactory profile was obtained by repeatedly adjusting the PCR conditions the ratio of the PCR components (Supplementary Fig. s5).

The allele dropout, peak height, and heterozygote balance (HB) were the three main parameters when we evaluated the typing results of the PCR condition studies.

For the Reaction Mix and the Primer Mix volume test, allele dropouts were observed at 0.5×Reaction Mix (G8, R3, and R9), 0.5×Primer Mix (B10), and 1.5×Primer Mix (B13, G8, and R11). When the 9948 was amplified with 0.75×, 1×, 1.25×, and 1.5×Reaction Mix, the average peak heights were 3387 RFUs, 5114 RFUs, 6042 RFUs, and 5944 RFUs, respectively. When the 9948 was amplified with 0.75×, 1×, 1.25×Primer Mix, the average peak heights were3387 RFUs, 5114 RFUs, and 6328 RFUs. The qualified peak height and peak height ratio (PHR) of heterozygous (>0.7) were obtained at 1×Reaction Mix and the Primer Mix volume (Supplementary Fig. s6).

For the reaction volume, full profiles can be obtained under all conditions. With the increase of volume, the average peak height decreased gradually, but the equilibrium between loci increased gradually (Supplementary Fig. s7).

For the denaturation temperature test, allele dropouts were observed at 89 °C, 90 °C, and 99 °C. The average peak heights were 3808 RFUs, 5391 RFUs, 6453 RFUs, 5114 RFUs, 2734 RFUs, 6290 RFUs, 4631 RFUs, and 4045 RFUs when it was from 91 to 98°C. Then, we calculated the PHR of heterozygous. The PHR value of all loci was higher than 0.7 when it was 94 °C. For the annealing temperature test, allele dropouts were observed at 64 °C and 65 °C. The average peak heights were 2359 RFUs, 2277 RFUs, 3311 RFUs, 5351 RFUs, 4475 RFUs, 5114 RFUs (recommended), 3108 RFUs, 3312 RFUs, and 2537 RFUs when it was from 55 °C to 63 °C and the PHR showed best performance in the Box plot when it was 60 °C (Supplementary Fig. s8).

(See figure on next page.)

Fig. 3 STRUCTURE analysis using 56 A-DIPs: **A** Stacked chart of 26 worldwide populations come from AFR, AMR, EAS, EUR and SAS. The dominant ancestral component of AFR was labeled with yellow, and it was pink, red, and brown for EAS, EUR, and SAS, respectively. For American populations, the dominant ancestral component of MXL and PEL was labeled with blue, while it was blue and red for PUR and CLM; **B** Stacked chart of five subpopulations of EAS and a north–south cline was visible. The Ancestral North EAS component was labeled with blue while the Ancestral South EAS component was labeled with brown; **C** The NJ-tree of 26 worldwide populations

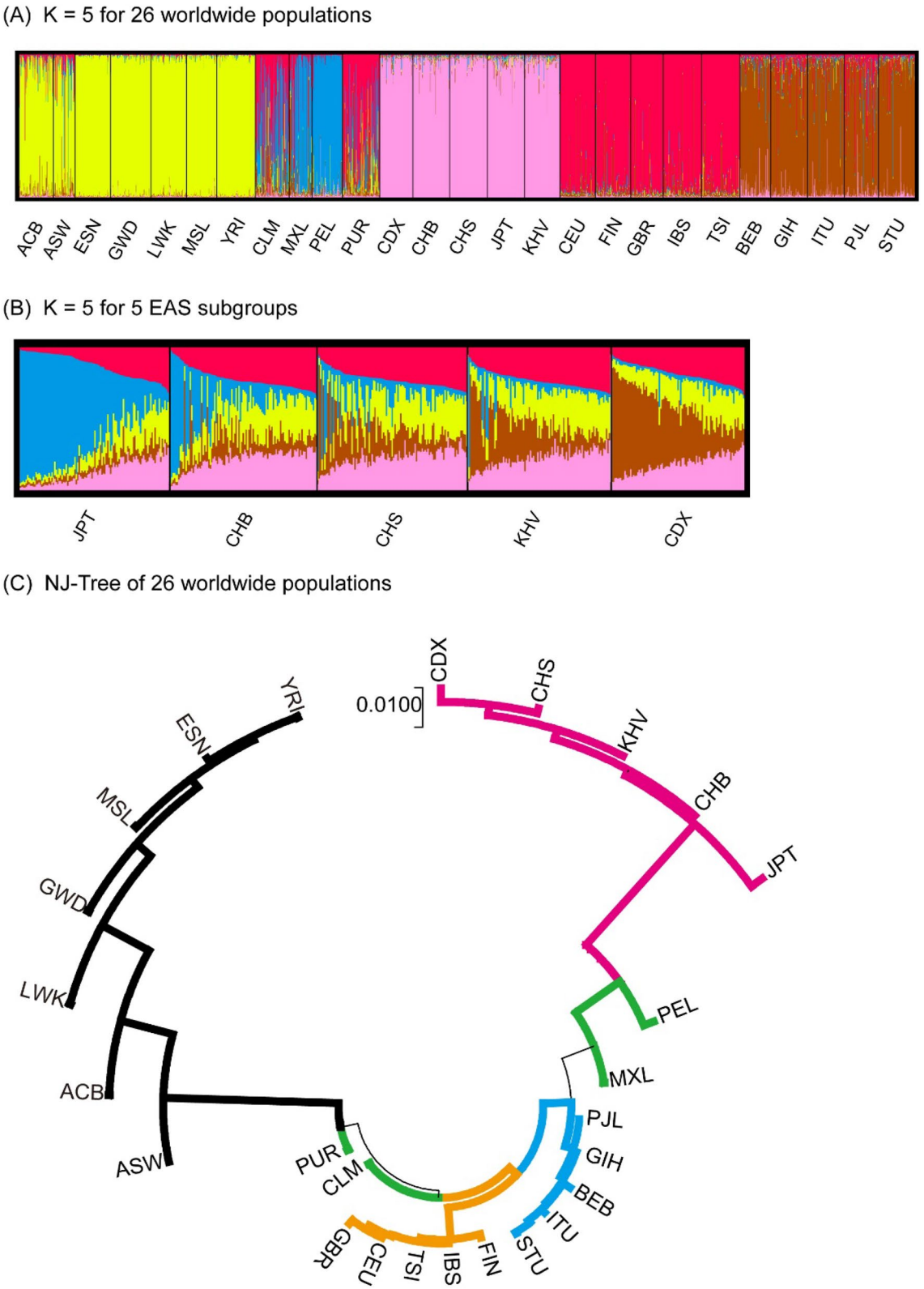

Fig. 3 (See legend on previous page.)

For the cycling number test, allele dropouts were observed at 20 cycles. Then, peak heights increased and HB advanced from 21 to 25 cycles. At 26 cycles, the peak heights increased but the HB decreased. At 27 cycles a large area of saturation was observed due to the overloading DNA quality (Fig. 4).

In the final extension time test, loci R4, B7, G11, Y11, B12, G13, and G14 had minus-A peaks for the 5 min-hold, the situation improved for the 10 min-hold, normal peak morphologies were obtained when the final extension time was more than 15 min (Supplementary Fig. s9).

Developmental validation of 60-panel

Besides the PCR-based studies, the developmental validation process also included the sensitivity study, species specificity study, the stability study, the mixture study, the reproducibility study, and the Case-type sample study.

For the sensitivity study, allele dropouts were observed when 31.25 pg template was amplified. Full profiles could be obtained from 62.5 pg to 5 ng. The fluorescent signal increased with the increase of the template amount. When 5 ng templated was amplified, large area of pull-up peaks and fluorescent saturation appeared; When 2 ng templated was amplified, most peak heights were between 2300 and 12,000 RFUs, with several pull-up peaks; When the amount of DNA was 1 ng, most peak heights were between 2300 and 12,000 RFUs; When the amount of DNA was 500 pg, most peak heights were between 501 and 2300 RFUs (Fig. 5A). The average peak heights were 28,349 RFUs, 9,416 RFUs, 5,163 RFUs, 1,825 RFUs, 978 RFUs, 614 RFUs, and 106 RFUs for template DNA amounts of 5 ng, 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, and 62.5 pg, respectively (Fig. 5B). Given that the peak heights were exceptionally high when the DNA amount exceeded 2 ng, and to provide a clearer representation, Fig. 5C focuses on displaying the average peak heights for DNA amounts of 1 ng or less. It's important to note that the data point for 31.25 pg was omitted from Fig. 5C due to locus dropout occurring at this concentration.

For the species specificity study, negative results were detected for all nine non-human samples as expected (Supplementary Fig. s10).

For the stability study, full profiles were yielded at 50% of Ethanol, 15 mM of calcium ion, 8 mg/mL of tannic acid, 3 mM of EDTA, 1 mol/L of humic acid, and 5 mM of hematin. The call rates decreased as the concentration of inhibitors increased, and the large amplicons dropped out first. When the concentration reached 90% of Ethanol, 35 mM of calcium ion, 40 mg/mL of tannic acid, 10 mM of EDTA, 2.5 mol/L of humic acid, and 8 of mM hematin, almost all alleles were lost (Supplementary Fig. s11).

For the mixture study, the allele number of a full profile was 86, 74, and 100 for 9948, 9947A, and the mixture, respectively. When the ratio was 1:1, 1:2, 1:4, 4:1, and 2:1, the call rates of the minor contributors were 100%. The allele dropouts of minor contributor appeared at the ratio of 1:9, 1:19, 9:1, and 19:1 (9948:9947A) (Table 2). It suggested that reliable profiles for analysis can be yielded when the DNA amount was greater than 200 pg.

For the reproducibility study, two operators conducted 6 samples separately, the typing results were concordant with each other. No significant differences in the peak heights, HB, or peak morphology were observed.

For the case-type sample study, nine samples from crime scenes were examined. All of them rendered full profiles and normal peak morphologies. The results showed that the 60-panel could be used to examine different forensic specimens (Supplementary Fig. s12).

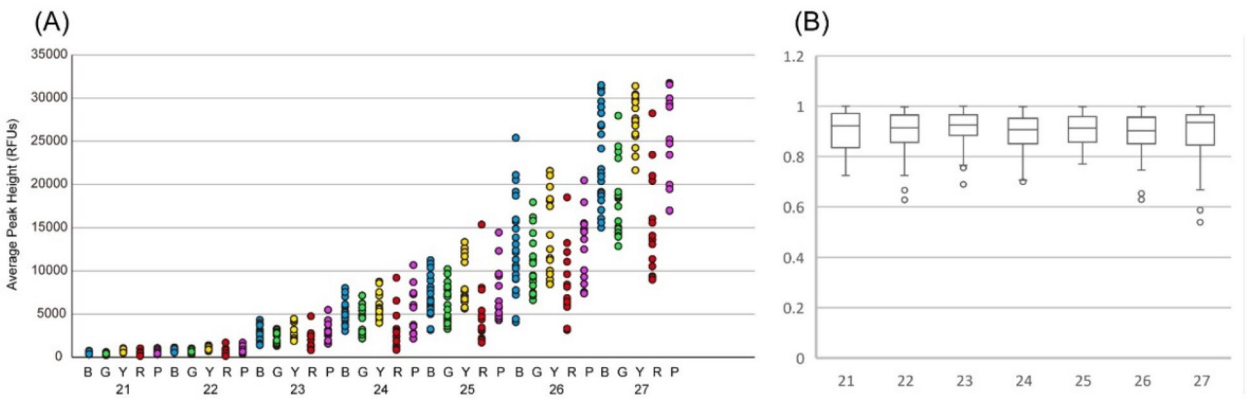

Fig. 4 The peak heights and heterozygote balances (HB) when the cycling number was from 21 to 27 cycles: **A** The dot plot showed the peak height of each locus. "B", "G", "Y", "R", and "P" represented loci labeled with FAM, JOE, TAMRA, ROX, and AEXA594, respectively; **B** The box chart showed the HB value for each cycling number

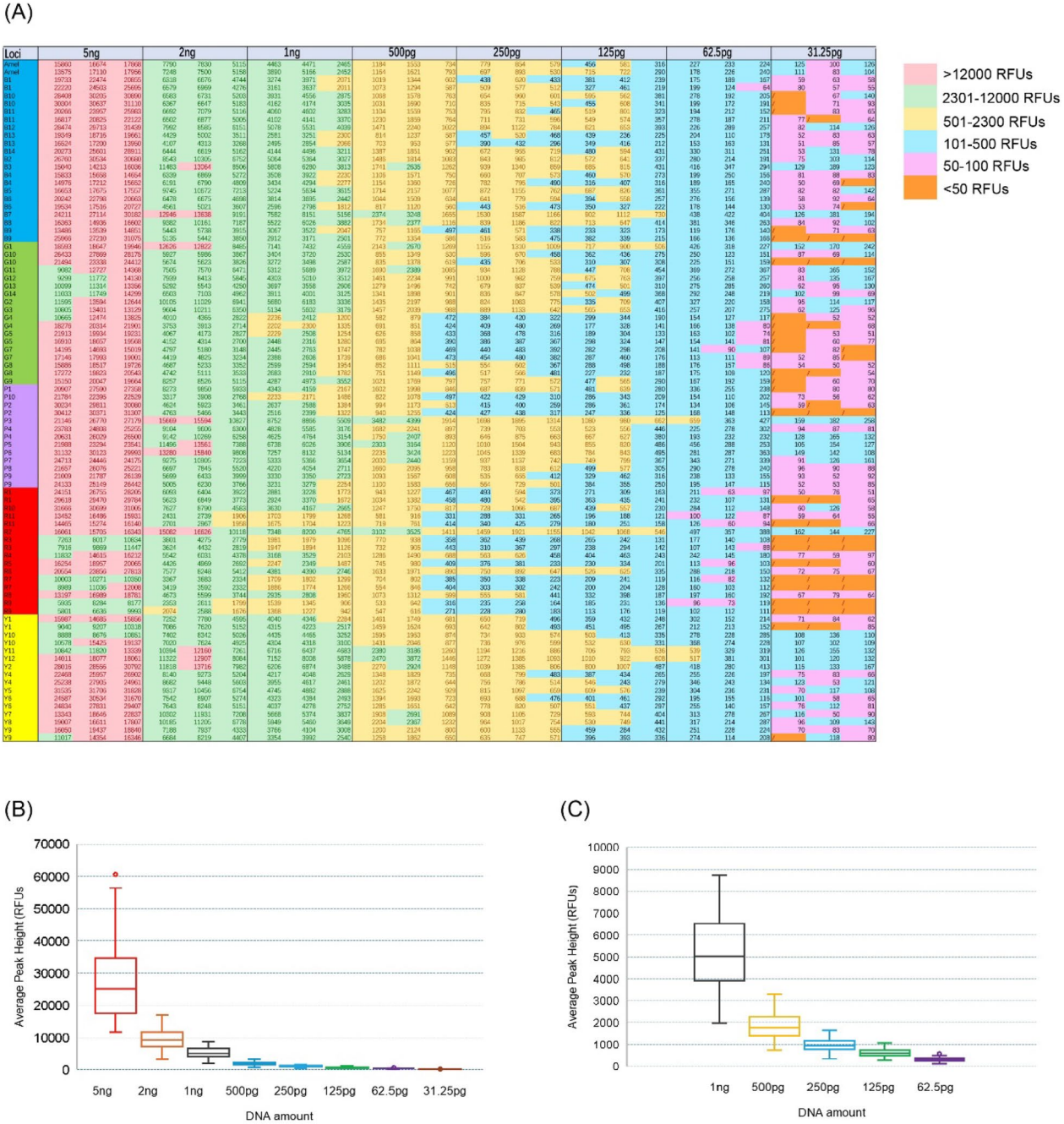

Fig. 5 **A** The peak heights for male control DNA 9948 when the following template amounts were amplified: 5 ng, 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, and 31.25 pg. The peak height ranges were highlighted as follows: < 50 RFUs (orange), 50–100 RFUs (purple), 101–500 RFUs (blue), 501–2300 RFUs (yellow), 2301–12,000 RFUs (green) and > 12,000 RFUs (pink); **B** The average peak heights when the following template amounts were amplified: 5 ng, 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, and 31.25 pg; **C** The average peak heights when the following template amounts were amplified: 1 ng, 500 pg, 250 pg, 125 pg, and 62.5 pg

For the degradation sample study, we examined a series of samples exposed to UV light and some blood spots kept for seven years. Theoretically, the degradation effect would increase as time goes on, and the results were as expected. For samples exposed to UV light for 0.5 h, the profile showed a degradation trend, and several drop-outs were observed. When the time extended to 2 h, about 50% of alleles were lost (Supplementary Fig. s13).

Table 2 The results of mixture studies using 60 panel

Ratio	Allele number			The allele call rate of the minor contributor (%)
	Test1	Test 2	Test 3	
9948:9947A = 1:1	100	100	100	100
9948:9947A = 1:2	100	100	100	100
9948:9947A = 1:4	100	100	100	100
9948:9947A = 1:9	94	96	93	93
9948:9947A = 1:19	87	92	88	83
9948:9947A = 2:1	100	100	100	100
9948:9947A = 4:1	100	100	100	100
9948:9947A = 9:1	94	94	90	90
9948:9947A = 19:1	87	89	89	84

For blood spots kept for seven years, which had a typical degradation pattern confirmed previously by a STR system: SureID®Y35 kit (Health gene Tech, China), all alleles were successfully yielded using the 60-panel (Supplementary Fig. s14).

Discussions

The newly established 60-panel consisted of the Amelogenin, 3 Y-chromosome DIPs, and 56 Autosomal DIPs. The Amelogenin and the 3 Y-chromosome DIPs were used in sex determination. The 56 Autosomal DIPs included were selected to distinguish the north and south of the East Asian populations on the premise of correctly discriminating the five major global populations (AFR, EAS, SAS, AMR, and EUR). The PCA results showed that AFR, EAS, and EUR formed distinct clusters, so did the EUR, SAS, and AMR. But when the five populations were analyzed together, some points of EUR, SAS, and AMR overlapped, reflecting the indelible mark left on South Asia and America by colonialism and their low resolution compared to the high accuracy of the three main continent groups. The STRUCTURE analysis and phylogenetic tree results corresponded to the PCA results.

For the East Asian populations, CHB and JPT were quite close compared with the other three EAS groups. Genealogical research had indicated similar genetic profiles of a less than 1% total variation among Han Chinese, Japanese and Korean [22]. Japanese bore genetic signatures from three ancient populations: hunter-gatherers called the Jomon, farmers called Yayoi, and migrations of the Kofun period. About 300 AD, people from China and Korea came to Japan and contributed approximately 71% genetic ancestry [23]. The genome-wide results were concordant with that revealed by the newly established ancestry inference panel. For the CHB and CHS, the differences between them are much smaller than that between Chinese Han and Japanese. The three large migratory waves and continuous small-scale migration of northern China inhabitants escaping to southern China made northern Han the primary contributors to the gene pool of the southern Han [24]. Given the historical genetic admixture of various ethnic groups with the Han population, the CHB was genetically close to the ethnic groups living in north China while the CHS was close to the ethnic groups living in the south. Thus, the CHS was also genetically close related to the CDX and the KHV. The Kinh is one of the main groups of Vietnam and interacted extensively with Hmong and Tai-Kadai groups. The possible origin included Southern China [25]. The high frequency of subclades of O-M175 among Tai people indicated a common ancestor of other southern Chinese [26].

The genetic evidence was concordant with the PCA plot, STRUCTURE results, and phylogenetic tree revealed by the 60-panel, proving the newly established panel was applicable in ancestral inference.

Forensic ancestry inference provides valuable clues when the sources of the specimens are unknown. For missing person cases, large-scale disasters, and armed conflicts, ancestry information could help practitioners quickly sift through the abundant information during the investigation and help handle the remains according to their cultural beliefs [27, 28]. Specimens of missing person cases often involve old bones or severely damaged tissues, while specimens of large-scale disasters and armed conflicts often involve corrupted, disfigured, and broken corpses. These challenging samples often have varying degrees of degradation, leading to failures of large fragments during the PCR. As known to all, the smaller the amplicon size, the more complete the profile. To meet the need of typing compromised forensic samples, the size of the amplicons of the 60-panel was limited to 200 bp considering three aspects: firstly, the nucleosome is essential for the maintenance of genome integrity and proper chromatin organization. Markers located in nucleosome core particles are prone to be successfully genotyped in ancient or degraded samples [29]. Structurally, the nucleosome subunit of chromatin consists of the core particle and the linker DNA. The core particle of approximately 146 bp DNA is very stable, and the stretches of linker DNA consisting of about 80 bp of DNA are unbound [30]. In total, the nucleosome contains about 200 bp of DNA, which can resist a certain degree of degradation. Secondly, the length of ancient human DNA fragments, the poorly preserved, and degraded samples are usually between 100 and 200 bp [31, 32]. Lastly, the Ion Torrent and Illumina systems have average read lengths of about 200 bp [33]. The amplicon size of the 60-panel is comparable to NGS systems and surpasses traditional STR systems.

In forensic practice, mixture samples are common, particularly in sexual assault cases or scenarios with multiple suspects. When DNA from multiple individuals is present in a sample, our panel's ability to distinguish and attribute genetic markers to each contributor becomes invaluable. The reliable generation of profiles from such mixtures, as evidenced with DNA amounts exceeding 200 pg, underscores the utility of our 60-DIP panel for forensic applications. This is especially true for cases involving degraded or trace samples. Consequently, this study not only confirms the panel's effectiveness but also establishes a benchmark for the minimum DNA requirements necessary for reliable analysis. Our panel's performance in validation studies, including assessments of simulated casework samples, degradation, inhibition, and mixed sources, has been commendable. These studies reinforce the suitability of our panel for forensic analysis, ensuring reliable and effective results in real-world applications.

Our newly established panel comprises 56 loci, selected with a priority on their potential for population discrimination and their level of polymorphism. The inclusion of loci with higher polymorphism significantly aids in personal identification—a fundamental aspect of forensic science with extensive applications. The Investigator® DIPplex Kit from Qiagen [34], which was the first to utilize DIPs for personal identification, includes 30 autosomal DIP markers. Our panel showed similar CPD and CPE values compared with the Investigator® DIPplex Kit, despite most of our loci not exhibiting the same high degree of polymorphism. However, by expanding the number of loci, we have managed to attain comparable effectiveness. The validation studies have confirmed that our 60-loci panel is well-suited for forensic testing, fulfilling the necessary criteria for such applications.

Conclusions

In the present study, we utilized the DIPs for forensic ancestry inference and personal identification in East Asian. A total of 56 autosomal DIPs, 3 Y-DIPs, and the Amelogenin were selected to build the 60-panel. The PCA, STRUCTURE, and phylogenetic tree construction results were consistent with each other, and concordant with the previous findings. Besides, the CPD value of the panel reached 0.999 999 999 999, suggesting the 60-panel not only provides some degree of biogeographic information but also a powerful supplement for personal identification in the East Asian population. Subsequent validation studies confirmed that the 60-DIP panel satisfies the stringent criteria for forensic examinations.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40246-025-00727-8.

Additional file 1.

Additional file 2.

Acknowledgements
We would thank Xingkai Zheng for his help in data collection and processing. We would also thank Xiaoye Jin and Yuxin Guo for their help in Marker selection.

Author contributions
Shuanglin Li: Conceptualization; Formal analysis; Software; Visualization; Writing—original draft; Shuyan Mei: Data curation; Methodology; Validation; Yanfang Liu: Methodology; review &editing; Wei Cui: Writing—review &editing; Bofeng Zhu: Conceptualization; Funding acquisition; Investigation; Project administration; Resources; Supervision;

Funding
This research was funded by National Natural Science Foundation of China (grant number 81930055) and Guangdong Province Universities and Colleges Pearl River Scholar Funded Scheme (2017).

Availability of data and materials
No datasets were generated or analysed during the current study.

Declarations

Competing interests
The authors declare no competing interests.

Author details
[1]Guangzhou Key Laboratory of Forensic Multi-Omics for Precision Identification, School of Forensic Medicine, Southern Medical University, Guangzhou 510515, China. [2]School of Basic Medical Sciences, Shenzhen University Medical School, Shenzhen University, Shenzhen 518055, Guangdong, China. [3]School of Basic Medicine and Forensic Medicine, Henan University of Science and Technology, Luoyang 471000, China. [4]School of Nursing, Guangdong Medical University, Dongguan 523808, China. [5]Microbiome Medicine Center Department of Laboratory Medicine, Zhujiang Hospital, Southern Medical University, Guangzhou 510515, China.

Received: 28 May 2024 Accepted: 12 February 2025
Published online: 21 March 2025

References

1. Daniel R, Santos C, Phillips C, Fondevila M, van Oorschot RA, Carracedo A, et al. A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Sci Int Genet. 2015;14:50–60.
2. Pranckoniene L, Jakaitieno A, Ambrozaityte L, Kavaliauskiene I, Kucinskas V. Insights Into de novo mutation variation in lithuanian exome. Front Genet. 2018. https://doi.org/10.3389/fgene.2018.00315.
3. Du W, Peng Z, Feng C, Zhu B, Wang B, Wang Y, et al. Forensic efficiency and genetic variation of 30 InDels in Vietnamese and Nigerian populations. Oncotarget. 2017;8(51):88934–40.
4. Haber M, Jones AL, Connell BA, Asan AE, Yang H, et al. A rare deep-rooting D0 African Y-chromosomal haplogroup and Its implications for the expansion of modern humans out of Africa. Genetics. 2019;212(4):1421–8.
5. Karmin M, Saag L, Vicente M, Wilson Sayres MA, Jarve M, Talas UG, et al. A recent bottleneck of Y chromosome diversity coincides with a global change in culture. Genome Res. 2015;25(4):459–66.

6. Li F, Vanwezer N, Boivin N, Gao X, Ott F, Petraglia M, et al. Heading north: late Pleistocene environments and human dispersals in central and eastern Asia. PLoS ONE. 2019;14(5):e0216433.
7. Zhang X, Shen C, Jin X, Guo Y, Xie T, Zhu B. Developmental validations of a self-developed 39 AIM-InDel panel and its forensic efficiency evaluations in the Shaanxi Han population. Int J Legal Med. 2021;135(4):1359–67.
8. Cui W, Nie S, Fang Y, Chen M, Zhao M, Lan Q, et al. Insights into AIM-InDel diversities in Yunnan Miao and Hani ethnic groups of China for forensic and population genetic purposes. Hereditas. 2022;159(1):22.
9. Sun K, Yun L, Zhang C, Shao C, Gao T, Zhao Z, et al. Evaluation of 12 Multi-InDel markers for forensic ancestry prediction in Asian populations. Forensic Sci Int Genet. 2019;43:102155.
10. Guo XY, Sun CC, Xue SY, Zhao H, Jiang L, Li CX. 49AISNP: a study on the ancestry inference of the three ethnic groups in the north of East Asia. Yi Chuan. 2021;43(9):880–9.
11. Scientific Working Group on DNA Analysis Methods (SWGDAM) Validation guidelines for DNA analysis methods. 2013. (http://swgdam.org/SWGDAM_Validation_Guidelines_APPROVED_Dec_2012.pdf (accessed 19.02.13))
12. Gouy A, Zieger M. STRAF-A convenient online tool for STR data evaluation in forensic genetics. Forensic Sci Int Genet. 2017;30:148–51.
13. Rousset F. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour. 2008;8(1):103–6.
14. Porras-Hurtado L, Ruiz Y, Santos C, Phillips C, Carracedo A, Lareu MV. An overview of STRUCTURE: applications, parameter settings, and supporting software. Front Genet. 2013;4:98.
15. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016;33(7):1870–4.
16. Buckleton J, Curran J, Goudet J, Taylor D, Thiery A, Weir BS. Population-specific FST values for forensic STR markers: a worldwide survey. Forensic Sci Int Genet. 2016;23:91–100.
17. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet. 2003;73(6):1402–22.
18. Sedgwick P. STATISTICAL QUESTION Multiple significance tests: the Bonferroni correction. Brit Med J. 2012;344:e509.
19. Lan Q, Li SL, Cai MM, Lei FZ, Shen CM, Zhu BF. A self-developed AIM-InDel panel designed for degraded DNA analysis: forensic application characterization and genetic landscape investigation in the Han Chinese population. Genomics. 2023;115(3):110620.
20. Liu L, Li SL, Cui W, Fang YT, Mei SY, Chen M, et al. Ancestry analysis using a self-developed 56 AIM-InDel loci and machine learning methods. For Sci Int. 2024;361:112065.
21. Larue BL, Ge JY, King JL, Budowle B. A validation study of the Qiagen Investigator DIPplex (R) kit; an INDEL-based assay for human identification. Int J Legal Med. 2012;126(4):533–40.
22. Wang Y, Lu D, Chung YJ, Xu S. Genetic structure, divergence and admixture of Han Chinese. Jpn Korean Popul Hered. 2018;155:19.
23. Cooke NP, Mattiangeli V, Cassidy LM, Okazaki K, Stokes CA, Onbe S, et al. Ancient genomics reveals tripartite origins of Japanese populations. Sci Adv. 2021;7(38):eabh2419.
24. Wen B, Li H, Lu D, Song X, Zhang F, He Y, et al. Genetic evidence supports demic diffusion of Han culture. Nature. 2004;431(7006):302–5.
25. Liu D, Duong NT, Ton ND, Van Phong N, Pakendorf B, Van Hai N, et al. Extensive ethnolinguistic diversity in Vietnam reflects multiple sources of genetic diversity. Mol Biol Evol. 2020;37(9):2503–19.
26. Srithawong S, Srikummool M, Pittayaporn P, Ghirotto S, Chantawannakul P, Sun J, et al. Genetic and linguistic correlation of the Kra-Dai-speaking groups in Thailand. J Hum Genet. 2015;60(7):371–80.
27. Cho S, Kim MY, Lee JH, Lee HY, Lee SD. Large-scale identification of human bone remains via SNP microarray analysis with reference SNP database. Forensic Sci Int Genet. 2020;47:102293.
28. Marsico FL, Vigeland MD, Egeland T, Pinero MH. Making decisions in missing person identification cases with low statistical power. Forensic Sci Int Genet. 2021;54:102519.
29. Dong CN, Yang YD, Li SJ, Yang YR, Zhang XJ, Fang XD, et al. Whole genome nucleosome sequencing identifies novel types of forensic markers in degraded DNA samples. Sci Rep. 2016;6:26101.
30. Bargaje R, Alam MP, Patowary A, Sarkar M, Ali T, Gupta S, et al. Proximity of H2A.Z containing nucleosome to the transcription start site influences gene expression levels in the mammalian liver and brain. Nucleic Acids Res. 2012;40(18):8965–78.
31. Der Sarkissian C, Allentoft ME, Avila-Arcos MC, Barnett R, Campos PF, Cappellini E, et al. Ancient genomics. Philos Trans R Soc Lond B Biol Sci. 2015;370(1660):20130387.
32. Maciejewska A, Jakubowska J, Pawowski R. Whole genome amplification of degraded and nondegraded DNA for forensic purposes. Int J Legal Med. 2013;127(2):309–19.
33. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
34. Neuvonen AM, Palo JU, Hedman M, Sajantila A. Discrimination power of Investigator DIPplex loci in Finnish and Somali populations. Forensic Sci Int Genet. 2012;6(4):e99–102. https://doi.org/10.1016/j.fsigen.2011.09.005.

Publisher's Note