Science & Justice 59 (2019) 228–233

Contents lists available at ScienceDirect

Science & Justice

journal homepage: www.elsevier.com/locate/scijus

Mass spectrometry-based SNP genotyping as a potential tool for ancestry inference and human identification in Chinese Han and Uygur populations

Jiashuo Zhang[a,b], Jingyi Zhang[a,b], Ruiyang Tao[b,c], Zihao Yang[b,d], Suhua Zhang[b,*], Chengtao Li[a,b,*]

[a] Department of Forensic Science, Medical School of Soochow University, Suzhou 215123, China
[b] Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Academy of Forensic Sciences, Ministry of Justice, China, Shanghai 200063, China
[c] Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, China
[d] Department of Forensic Medicine, School of Basic Medical Science, Wenzhou Medical University, Wenzhou 325035, China

ARTICLE INFO

Keywords:
Ancestry informative SNP (AISNPs)
Ancestry inference
Forensic genetics
Matrix assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS)
Han population
Uygur population

ABSTRACT

Ancestry informative SNPs (AISNPs) are genetic variants that exhibit substantially different frequencies between populations from different geographical regions; thus, they can provide some valuable information regarding samples and be used in predicting an individual's ancestry origin. In this study, we selected the potentially best SNPs from our previous study with genome-wide high-density SNP data in mainland Chinese Uygur and Han populations and investigated the allele distribution patterns and genetic information of AISNPs with a mass spectrometry-based SNP genotyping panel. Mass spectrometry-based detection technology offers the opportunity to analyze forensic DNA samples and obtain SNP variants with accuracy and ease. The panel can distinguish and cluster Han and Uygur populations and is suitable for human identification and parentage testing in the two populations. Heatmap, PCA, and Structure analyses indicated that the ideal 64 AISNPs can collectively provide additional information on differences among populations from East Asia, South Asia, Europe and Africa. Additionally, the results proved that the Uygur population is the admixture of East Asia and Europe.

1. Introduction

Single nucleotide polymorphisms (SNPs) with low mutation rates, short amplicon length and high discrimination power have become effective markers for forensic application, such as human identification [1–8], ancestry inference [9,10] and prediction of externally visible characteristics, etc. [11–13]. Thus, various methods and panels for analyzing SNP markers have been explored. In 2012, Sequenom launched the iPLEX Sample ID Plus Panel, which amplifies 47 of the 52 SNPs in the SNPforID multiplex, amelogenin and two markers (XKRY-1 and HFSY-2) of the Y-chromosome. The iPLEX protocol involves a large multiplex PCR, a large multiplex SBE reaction, and detection of the SBE products by matrix assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS). In 2013, N. Morling et al. [6] evaluated the accuracy, reliability, sensitivity and forensic efficiency of the panel for human identification and sample tracking. These studies proved that the mass spectrometry-based method is an ideal choice for medium-resolution multiplex tool for parallel SNP genotyping because of the limited hands-on time of the iPLEX® assay, the low reagent costs and the automatic analysis of the mass spectra.

According to the Sixth National population Census in 2010 [14], Uygur is the fourth largest ethnic minority with approximately 10 million people who live primarily in the Xinjiang Autonomous Region. The Uygur population, who presents mixed anthropological features of Europeans and Asians, shows great difference with Chinese Han population in appearance, culture and habit. Jin et al. [15] demonstrated that large-scale SNPs show genetic difference between Uygur and Han populations. In a national project of multi-ethnic genetic background investigation which is supported by the National Key Research and Development Program of China, > 3000 individuals are sequenced with Genome-Wide Human SNP Array 6.0 (Thermo Fisher Scientific, MA, USA). The SNP Array features 1.8 million genetic markers, including 934,969 single nucleotide polymorphisms (SNPs) and > 946,000 probes for the detection of copy number variation. We observed some SNPs with highly differentiated allele frequencies between Han and Uygur Two essential criteria [16] were adopted to categorize a SNP as an ancestry-informative SNP (AISNP). These include (1) SNPs between the two populations are significant difference and (2) the distance between two contiguous SNPs is large enough to avoid strong linkage disequilibrium (LD). To estimate the genetic difference between

* Corresponding authors at: Shanghai Key Laboratory of Forensic Medicine, Shanghai Forensic Service Platform, Academy of Forensic Sciences, Ministry of Justice, China, Shanghai 200063, China.
E-mail addresses: zsh-daisy@163.com (S. Zhang), lichengtaohla@163.com (C. Li).

https://doi.org/10.1016/j.scijus.2019.01.006
Received 11 June 2018; Received in revised form 27 December 2018; Accepted 27 January 2019

Han and Uygur populations, we calculated unbiased estimates of F_{ST} according to Weir and Cockerham [17]. F_{ST} measures population differentiation or relatedness. We screened the 22 autosomes and dropped SNPs with little difference ($F_{ST} < 0.01$). For each pair of contiguous SNPs with an interval distance smaller than 500 Kb, we retained the one with a higher F_{ST} value. Finally, we ranked the markers that satisfied these criteria in descending order based on their F_{ST} values, and 78 SNPs were listed as candidates.

Considering the throughput, experiment time and costs, we try to design a multiplex panel for these targets with the MALDI-TOF MS platform. Concordance, accuracy and sensitivity of the novel custom panel were tested, as well as the performance for ancestry inference and human identification in Han and Uygur populations.

2. Materials and methods

2.1. Panel designing

MassARRAY Assay Designer (Sequenom, CA, USA) was adopted for multiprimer designing, resulting in three primer mix pools for 76 targeted markers (Supplementary Table S1). The three primer mix pools of W1, W2 and W3 aimed for 32, 29 and 15 targets, respectively. Two AISNPs (rs2594899 and rs6091928) were deleted from the final panel, as either the primer designing was difficult or the primer was not compatible with other primers. Pair of PCR primers and an extension primer were provided for each SNP. Primers were synthesized by GeneCore Biotech Inc. (Shanghai, China).

2.2. Mass spectrometry-based SNP genotyping (iPLEX assay)

The 76 targets were sequenced with the MALDI-TOF MS platform (Sequenom, CA, USA). The MassARRAY system is a scalable platform offering a suite of applications for quantitative and qualitative genomic analysis. The process combines the technologies of mass spectrometry, polymerase chain reaction (PCR), and primer extension to determine each allele.

The first step in this process is to perform PCR to amplify the region surrounding the target. The total 5 μL PCR reaction contained 1.3 μL H_2O, 0.5 μL 10× PCR buffer, 0.4 μL 25 mM $MgCl_2$, 0.1 μL 25 mM dNTP mix, 0.5 μL forward/reverse primer mix (1 μM final concentration), 0.2 μL PCR enzyme (5 U/μl) (Sequenom, CA, USA) and 2 μL sample DNA. The recommended DNA amount was 10 ng. The PCR was performed in a GeneAmp 9700 thermal cycler (Thermo Fisher Scientific, MA, USA) with the following conditions: 95 °C for 2 min followed by 45 cycles of 95 °C for 30 s, 56 °C for 30 s, and 72 °C for 1 min followed by 72 °C for 3 min.

The second step was treatment of the PCR reactions with shrimp alkaline phosphatase to dephosphorylate unincorporated nucleotides. The PCR products were treated with a cocktail of 1.53 μL H_2O, 0.17 μL Shrimp Alkaline Phosphatase (SAP) Buffer and 0.3 μL SAP enzyme (1.7 U/μL) (Sequenom, CA, USA) at 37 °C for 40 min followed by 85 °C for 5 min.

In the third step, a single base extension (SBE) reaction was performed using thermo sequenase to incorporate mass-modified dideoxy nucleotides. The extension reaction contained 7 μL of SAP-treated PCR products and 2 μL iPLEX mix (Sequenom, CA, USA). The iPLEX mix contained 0.2 μL iPLEX buffer, 0.2 μL iPLEX Termination mix, 0.94 μL extend primer mix (1.57 μM), 0.041 μL iPLEX enzyme (Sequenom, CA, USA), and 0.619 μL H_2O. The SBE reaction was performed with the following conditions: 94 °C for 30 s followed by 40 cycles of 94 °C for 5 s, 52 °C for 5 s, 80 °C for 5 s, 52 °C for 5 s, 80 °C for 5 s, 52 °C for 5 s, 80 °C for 5 s, 52 °C for 5 s, 80 °C for 5 s, 52 °C for 5 s, and 80 °C for 5 s followed by 72 °C for 3 min. A total of 41 μL molecular grade water and ion exchange resin (Sequenom, CA, USA) was added to each sample. Samples were rotated for approximately 15 min on a tube rotator and centrifuged at 4000 rpm for 5 min. This post-PCR primer extension reaction generated small DNA products that have a unique mass value based on each allele. These mass extension products were spotted on the SpectroCHIP array (Sequenom, CA, USA) using the RS1000 nano spotter (Sequenom, CA, USA), which consists of a matrix that assists in the ionization of the DNA products when exposed to high intensity laser. The time of flight of these ionized products depends on the mass of each allele that will be measured by the mass spectrometer. The results were visualized on the MassArray TyperAnalyzer version 4.0 (Sequenom, CA, USA).

2.3. Samples

Blood samples of 183 Chinese Eastern Han (CHE) and 100 Chinese Uygur (CHU) unrelated individuals were collected. A written informed consent was obtained from each participant, with the approval of the Ethics Committee of Academy Forensic Sciences, Ministry of Justice, China. DNA was extracted with QIAamp DNA Blood Mini Kit (QIAGEN, Hilden, Germany), according to the manufacturer's instruction. DNA concentration was determined by the Quantifiler Human DNA Quantification Kit (Thermo Fisher Scientific, MA, USA) with 7500 Real-time PCR System (Thermo Fisher Scientific, MA, USA). The concentrations of extracted DNA ranged from 16.38 to 47.34 ng/μL. All DNA samples were diluted to 5 ng/μL.

2.4. Sensitivity testing

For sensitivity testing of the custom panel, serial dilutions of a quantified control DNA 2800 M were tested in duplicate with quantities of 10 ng, 5 ng, 2.5 ng, 1.25 ng, 625 pg, 312.5 pg, 156.25 pg, 125 pg, 78.125 pg, 31.25 pg, 15.625 pg, and 7.8125 pg.

2.5. Concordance study

Genotypes of the control sample 2800 M (Promega, WI, USA) and 20 randomly selected samples at the included targets were validated with Sanger sequencing. Additionally, control DNA of 2800 M was used as positive sample during each batch of sequencing.

2.6. Case-type samples study

For case-type samples study, we tested vaginal blood, vaginal secretion, peripheral blood, hair and buccal swab of a female volunteer, and semen and saliva samples of a male volunteer.

2.7. Data analysis

The MassARRAY TYPER 4.0 software analyzed the results in real-time using a Gaussian mixture model for cluster analyses [6]. The allele genotypes with peak heights were exported as Plate Data File from TYPER 4.0 software. The heterozygote balance (H*b*) was calculated as the lower peak height dividing the higher peak height.

Genetic analyses, such as Hardy-Weinberg equilibrium (HWE), expected heterozygous (H*e*) and observed heterozygous (H*o*), were calculated with PLINK v 1.07 [18]. LD was tested using HaploView software. The detailed population genetic structure was performed using model-based clustering method implemented in STRUCTURE v.2.3.4 [19]. Each run used 100,000 estimation iterations for K = 2 to 10 after a 10,000 built-in length with 5 replicates. Graphics were generated using CLUMPP v1.1.2 [20]. STRUCTURE HARVESTER [21] was used to generate a table of mean likelihoods (L(K)) and variance per K value from STRUCTURE. Principal components analysis (PCA) was carried out using PLINK v 1.07, GCTA software and the “prcomp” function in the R v.3.4.0 package “stats”. Similarly, a heatmap was created based on population allele frequencies with R v.3.4.0 “pheatmap”. The neighbor-joining (NJ) tree was constructed using the “nj” function in the R package “ape” [22].

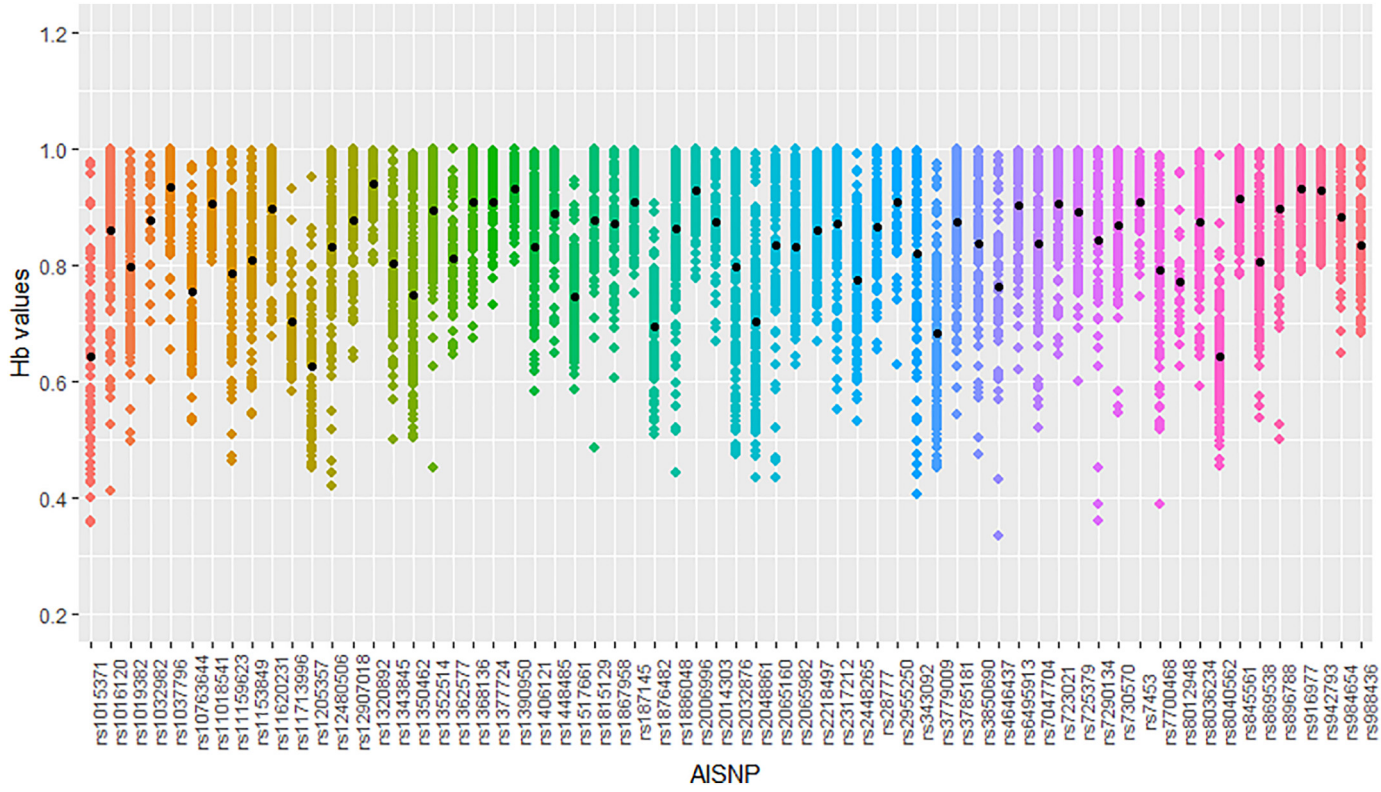

Fig. 1. Hb values of 7569 heterozygotes detected among 283 individuals at the 64 AISNPs. Hb values at each SNP are labeled in the same color. Black diamonds represent the average Hb values.

3. Results and discussion

3.1. Genotyping efficiency of the mass spectrometry-based SNP panel

183 CHE and 100 CHU individuals were double sequenced with the custom panel. Twelve AISNPs (rs11070629, rs11160371, rs1404402, rs2303798, rs2470102, rs279025, rs6713532, rs7163702, rs718092, rs728404, rs764138, rs963171) were detected with > 10% missing genotypes. Genotype dropout was not observed from the two detection wells on the SpectroCHIP Array at the remaining 64 AISNPs. Thus, the effective genotype calling rate of this panel was 84.21% (64/76). The following analyses were performed with the results at the 64 AISNPs from the first detection well.

A key factor in forensic SNP analysis is heterozygote balance, which is critical for reliable genotyping of heterozygotes. According to the results in reference [2], the H*b* for accurate heterozygotes calling should be above 60%. Fig. 1 shows the H*b* analysis of observed heterozygotes at the 64 AISNPs among the 283 tested samples. The average H*b* values ranged from 0.62 (rs1205357) to 0.94 (rs1320892). Regarding H*b* < 0.6, only 4% (348 out of 7569) of heterozygotes were observed, which may have been due to a sequencing effect.

3.2. Sensitivity results

According to the MALDI-TOF MS's manual instruction, 10 ng initial DNA was recommended for the first step of PCR preparation. However, DNA quantities obtained from forensic biological samples were frequently lower than the recommend quantities during routine work. To verify the detection limit of the custom panel, we performed sensitivity testing. Full genotypes were obtained at the 64 AISNPs with control DNA 2800 M at 10 ng and 5 ng in the 5 μL PCR reaction system (Fig. 2-A). The detection rate was calculated as: the number of observed SNP divided the total SNP here ($N = 64$). The average detection rate exceeded 90%, while the DNA quantities ranged from 2.5 ng to 156.25 pg. However, when the DNA quality dropped to 78.125 pg, the average detection rate of AISNPs was only 64%. Fig. 2-B illustrates the observed average H*b* values at different DNA amounts in the 5 μL PCR reaction system. H*b* values ranged from 0.715 (125 pg) to 0.810 (10 ng) with minor variation. The above data indicated that no < 156.25 pg DNA should be prepared for comparatively ideal results. In addition, even when there was a lower amount of DNA, the allelic balance of obtained heterozygotes was reliable.

3.3. Concordance results

The genotypes of 2800 M and 20 randomly selected samples at the 64 targets were consistent with Sanger sequencing results, and the genotypes of 2800 M were consistent during each batch of sequencing. Genotypes of control DNA 2800 M are listed in Supplementary Table S2.

3.4. Case-type samples study

The genotypes of different biological samples collected from the two volunteers were fully called at the 64 targets. The genotypes of samples of each volunteer were fully consistent.

3.5. Forensic application of the mass spectrometry-based SNP panel

3.5.1. Efficiency for human identification and parentage testing

Allele frequencies and corresponding parameters of the 64 SNPs in CHE and CHU are listed in Supplementary Table S3. No deviation from HWE was observed for any of the targets after Bonferroni's correction ($p = 0.05/64$) in CHE and CHU populations. In the CHE population, the Ho and He values ranged from 0.0109 (rs1032982) to 0.4993 (rs1037796) and 0.0109 (rs1032982) to 0.5738 (rs11620231), with mean values of 0.3467 and 0.3462, respectively; the DP values varied from 0.0216 (rs1032982) to 0.6566 (rs2317212); and the PE values

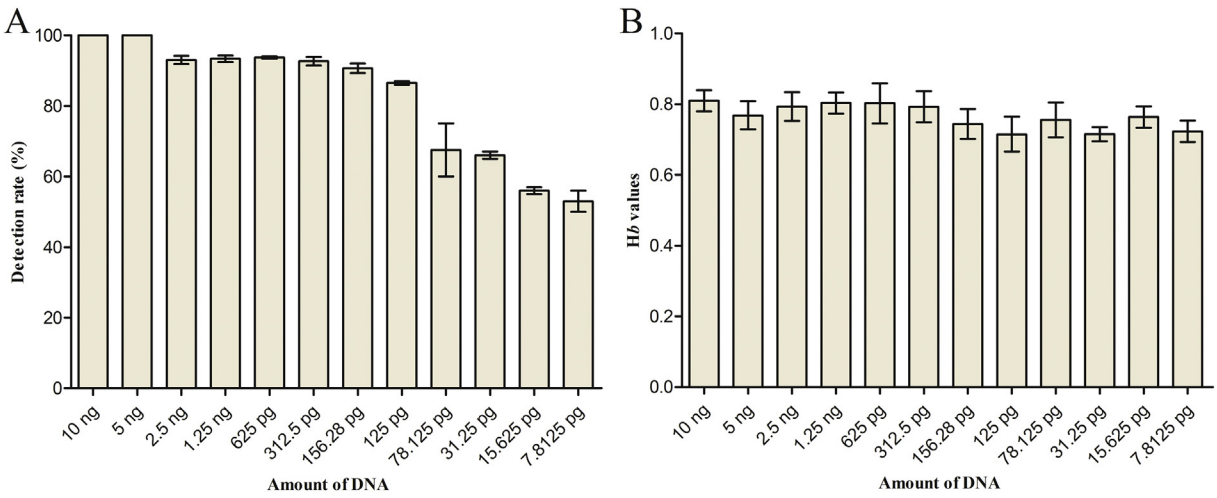

Fig. 2. Sensitivity testing of template DNA ranging from 10 ng to 7.8125 pg. (A) Average detection rate of AISNP was against DNA template mass. (B) Average Hb values were against DNA template mass. Error bars represent the plus and minus standard error in duplicate.

ranged from 0.0050 (rs1032982) to 0.2185 (rs1037796). In the CHU population, the Ho and He values ranged from 0.1040 (rs8012948) to 0.5000 (rs1517661) and 0.0700 (rs8012948) to 0.6289 (rs7047704), with mean values of 0.4214 and 0.4305, respectively; the DP values varied from 0.1666 (rs8012948) to 0.6249 (rs1016120); and the PE values ranged from 0.0506 (rs8012948) to 0.2187 (rs1517661 and rs1016120). The lowest Ho, He, DP and PE values were observed at rs1032982 locus in CHE, while the lowest value was found at rs8012948 in CHU population. Since LD was not identified between the pair of targets, the cumulative discrimination power (CDP) was 1–6.41E-20 in CHE and 1–3.03E-23 in CHU, while the cumulative exclusion power (CPE) was 0.999983345 in CHE and 0.999998407 in CHU. The results indicated that this panel would be a powerful tool for human identification and parentage testing in these two populations.

3.5.2. *Efficiency of ancestry inference between Han and Uygur*

Except for 100 CHU and 183 CHE individuals, 103 Beijing Han (CHB) and 103 Southern Han (CHS) from the 1000 Genomes Project (1KG) were referred here. In Fig. 3-A, the estimated cluster membership frequency is shown for each individual as a colored bar column, and K denotes the number of clusters specified. According to the table of mean likelihood (L(K)), as shown in Supplementary Table S4, the most likely and appropriate number of clusters is six. At K = 6, CHB, CHE and CHS are clustered into one group and emerge as a distinct cluster from CHU (Fig. 3-A). To obtain a more visual and intuitive exhibition, we carried out PCA based on the individual genotypes from CHE, CHS, CHB and CHU. In Fig. 3-B, the combined percentage of the variance of the data set explained by two PCA axes was 8.85% (4.66% + 4.19%). The PCA result indicated that individuals from Chinese Han clustered tightly, with exclusion of individuals from Uygur, which coincided with the results observed in the STRUCTURE.

3.5.3. *Efficiency of ancestry inference among populations*

For further evaluating the ancestry inference power of this panel, we referred the genotyping data at the 64 AISNPs from the other 22 populations of the 1KG. The sampling areas and sample size for each regional population are shown in Supplementary Table S5. We determined the allele frequency distribution characteristics in CHU, CHE and the 22 populations (Fig. 4-A). The heatmap matrix of 64 AISNPs provides more detail on the contribution of each SNP. SNPs that are very close to each other show a similar color pattern. The color pattern changes gradually from light yellow to dark green, indicating the lowest allele frequencies (0) to highest allele frequencies (1). The heatmap shows the relationships of the SNPs and the populations graphically in the marginal dendrograms. The heatmap also allows for determination of how these individual SNPs contribute to the differentiation of the specific populations analyzed. For example, in the left cluster, the 26 AISNPs showed high allele frequencies in African populations (ASW, ACB, LWK, YRI, ESN, GWD, MSL), while in the right cluster, most of the remaining 38 AISNPs showed relatively low allele frequencies. Fig. 4-B displays that the 24 populations were assigned into four distinct clusters in the highest likelihood run at K = 3. CHU has approximately 60% admixture from East Asian ancestry and 40% admixture from Europe ancestry. The populations of Gujarati Indian from Houston (GIH), Punjabi from Bangladesh (PJL), Bengali from Bangladesh (BEB), Sri Lankan Tamil from the UK (STU), and Indian Telugu from the UK (ITU), which belong to South Asia, displayed East Asian, European and African admixture. Supplementary Fig. S1-A shows that the first two factors account for 86.03% of the variation. Factor 1, which accounts for 44.78% of the variation, is defined at the extremes by populations from Africa and East Asia. Factor 2 (41.25% of the variation) separates Europe from East Asia and Africa, with South Asia lying between them. An unrooted tree (Supplementary Fig. S1-B) revealed the genetic relationships among CHE, CHU and the 22 previously published populations, separating them into four main branches.

4. Conclusion

Many AISNP panels have been published for ancestral inferences at various levels of resolution [9,10,23–29]. Most of the panels are able to correctly predict an individual's ancestry from Europe, Africa, America and Asia with high accuracy. In this study, taking advantage of our previous study with genome-wide high-density SNP data, we were able to screen a set of high quality AISNPs in distinguishing and clustering Han and Uygur populations. The mass spectrometry-based detection assay has proven to be a useful research tool. All reactions in the iPLEX protocol were performed in the same plate with little hands-on time, and the analyses were performed automatically by the TYPER 4.0 software. However, the call rates of some of the SNPs were not impressive. In particular, the rs11070629, rs11160371, rs1404402, rs2303798, rs2470102, rs279025, rs6713532, rs7163702, rs718092, rs728404, rs764138 and rs963171 loci should either be removed from the panel or the SNPs should be redesigned and optimized. The results at the remaining 64 SNPs showed that they yielded a relatively clean and robust finding between Han and Uygur populations as well as among four regions of East Asia, South Asia, Europe and Africa.

The Mass spectrometry-based SNP genotyping Panel required 10 ng

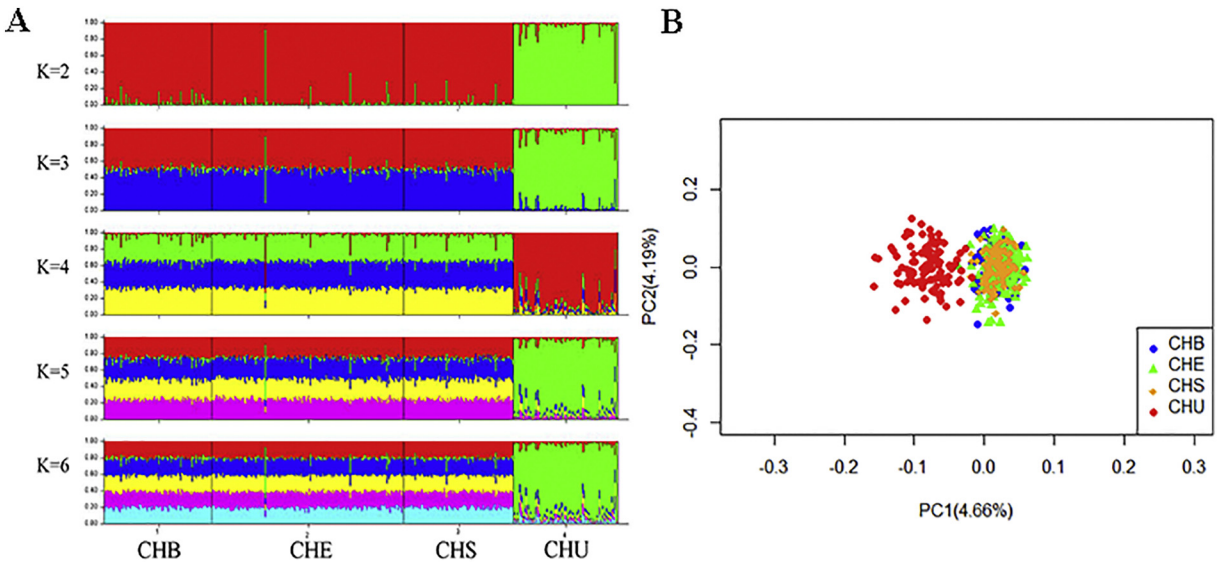

Fig. 3. (A) STRUCTURE analysis for CHB (n = 103), CHE (n = 183), CHS (n = 105) and CHU (n = 100). CHB and CHS from the 1000 Genomes Project (1KG). Each individual is represented by a vertical line partitioned into K colored segments, whose lengths are proportional to the estimated membership in each of the inferred K groups. Admixture and correlated allele frequency models were taken into account. (B) PCA plot of 491 individuals containing CHE, CHU, CHB and CHS analyzed with 64 AISNPs. Each individual is represented as a colored point according to the population of origin.

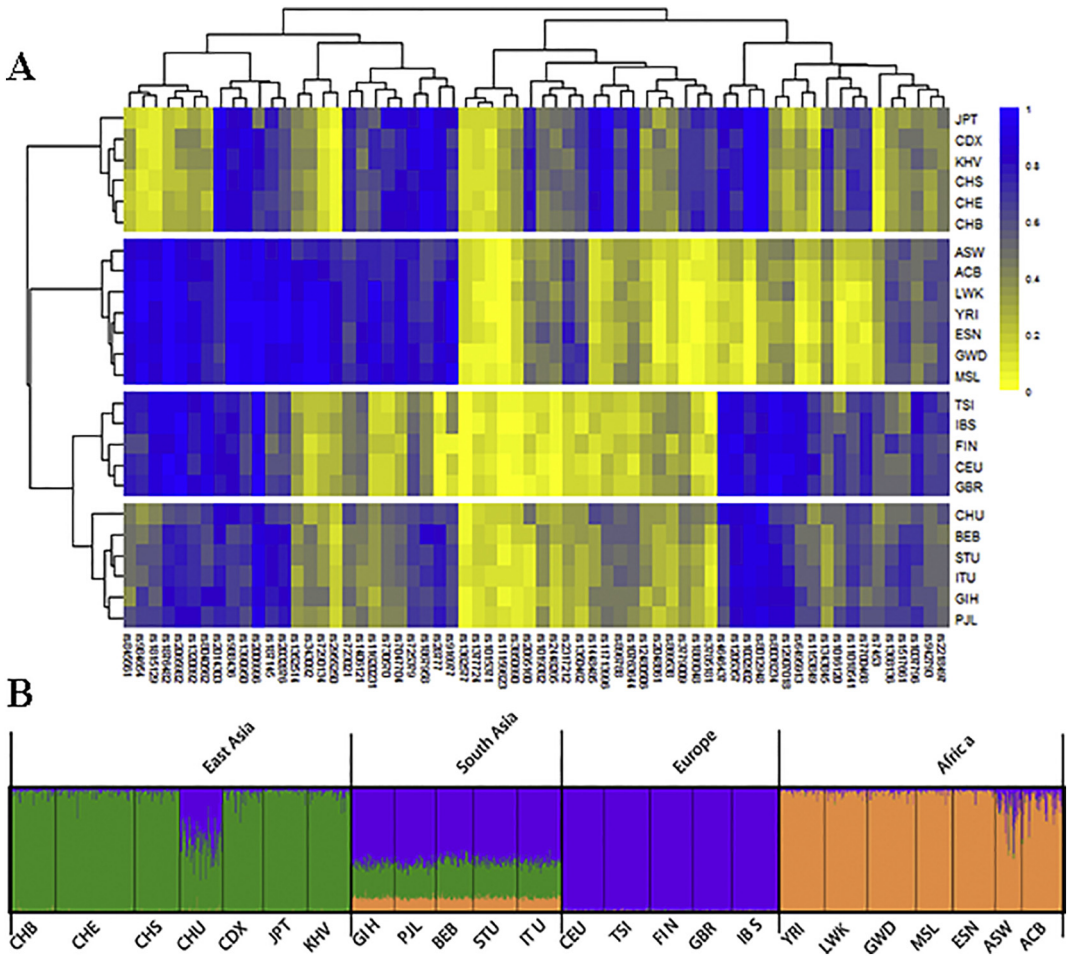

Fig. 4. (A) Heatmap representation of 64 AISNPs in 24 populations. Heatmap representation of the 24 population allele frequencies and 64 SNPs simultaneously. (B) STRUCTURE analysis of 24 populations based on genotypes at the 64 AISNPs. All populations were assigned to four distinct clusters in the highest likelihood run at K = 3.

for optimal performance, which is 50–100 times more than the amount of DNA typically used in the PCR of forensic genetic STR kits. Here, the custom panel required 5 ng for complete profiles. Allele drop-outs seemed to occur frequently when the DNA amount was lower than 78.125 pg. Therefore, the panel is not suitable for crime case investigations in its current form. However, it may be useful for complex parentage testing by providing additional genetic information, especially when the cost of the assay are taken into consideration.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scijus.2019.01.006.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgements

This study was supported by grants from the National Key R&D Program of China (grant numbers 2016YFC0800703), Standard Program of Shanghai Municipality (grant numbers 16DZ0501600, grant numbers 18DZ1200300), Public Interest Research Grant Programs of National Research Institutes (grant numbers GY2017D-2) and the National Natural Science Foundation (grant numbers 81625013, grant numbers 81772028). The funders had no role in study design, data analysis, publishing decisions, or manuscript preparation.

References

[1] S. Cho, H.J. Yu, J. Han, Y. Kim, J. Lee, S.D. Lee, Forensic application of SNP-based resequencing array for individual identification, Forensic Sci. Int. Genet. 13 (2014) 45–52.
[2] S. Zhang, Y. Bian, A. Chen, H. Zheng, Y. Gao, Y. Hou, C. Li, Developmental validation of a custom panel including 273 SNPs for forensic application using Ion Torrent PGM, Forensic Sci. Int. Genet. 27 (2017) 50–57.
[3] C. Borsting, S.L. Fordyce, J. Olofsson, H.S. Mogensen, N. Morling, Evaluation of the Ion Torrent HID SNP 169-plex: a SNP typing assay developed for human identification by second generation sequencing, Forensic Sci. Int. Genet. 12 (2014) 144–154.
[4] A. Sarkar, M.R. Nandineni, Development of a SNP-based panel for human identification for Indian populations, Forensic Sci. Int. Genet. 27 (2017) 58–66.
[5] H. Boonyarit, S. Mahasirimongkol, N. Chavalvechakul, M. Aoki, H. Amitani, N. Hosono, N. Kamatani, M. Kubo, P. Lertrit, Development of a SNP set for human identification: a set with high powers of discrimination which yields high genetic information from naturally degraded DNA samples in the Thai population, Forensic Sci. Int. Genet. 11 (2014) 166–173.
[6] P. Johansen, J.D. Andersen, C. Borsting, N. Morling, Evaluation of the iPLEX(R) sample ID plus panel designed for the Sequenom MassARRAY(R) system. A SNP typing assay developed for human identification and sample tracking based on the SNPforID panel, Forensic Sci. Int. Genet. 7 (2013) 482–487.
[7] F. Moreno, A. Freire-Aradas, C. Phillips, M. Fondevila, A. Carracedo, M.V. Lareu, SNP variation with latitude: analysis of the SNPforID 52-plex markers in north, mid-region and south Chilean populations, Forensic Sci. Int. Genet. 10 (2014) 12–16.
[8] Y. Ruiz, M.A. Chiurillo, L. Borjas, C. Phillips, M.V. Lareu, A. Carracedo, Analysis of the SNPforID 52-plex markers in four Native American populations from Venezuela, Forensic Sci. Int. Genet. 6 (2012) e142–e145.
[9] C. Phillips, A. Salas, J.J. Sanchez, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, M. Calaza, M.C. de Cal, D. Ballard, M.V. Lareu, A. Carracedo, S.N. Consortium, Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, Forensic Sci. Int. Genet. 1 (2007) 273–280.
[10] C. Phillips, A. Freire Aradas, A.K. Kriegel, M. Fondevila, O. Bulbul, C. Santos, F. Serrulla Rech, M.D. Perez Carceles, A. Carracedo, P.M. Schneider, M.V. Lareu, Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries, Forensic Sci. Int. Genet. 7 (2013) 359–366.
[11] Y. Ruiz, C. Phillips, A. Gomez-Tato, J. Alvarez-Dios, M. Casares de Cal, R. Cruz, O. Maronas, J. Sochtig, M. Fondevila, M.J. Rodriguez-Cid, A. Carracedo, M.V. Lareu, Further development of forensic eye color predictive tests, Forensic Sci. Int. Genet. 7 (2013) 28–40.
[12] K.B. Gettings, R. Lai, J.L. Johnson, M.A. Peck, J.A. Hart, H. Gordish-Dressman, M.S. Schanfield, D.S. Podini, A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. population, Forensic Sci. Int. Genet. 8 (2014) 101–108.
[13] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int. Genet. 7 (2013) 98–115.
[14] Communiqué of the national bureau of statistics of People's Republic of China on major figures of the 2010 population census(No.2)[J], China Popul. Today 2 (2011) 24–25.
[15] S. Xu, W. Huang, J. Qian, L. Jin, Analysis of genomic admixture in Uyghur and its implication in mapping strategy, Am. J. Hum. Genet. 82 (2008) 883–894.
[16] P. Qin, Z. Li, W. Jin, D. Lu, H. Lou, J. Shen, L. Jin, Y. Shi, S. Xu, A panel of ancestry informative markers to estimate and correct potential effects of population stratification in Han Chinese, Eur. J. Hum. Genet. 22 (2014) 248–253.
[17] B.S. Weir, C.C. Cockerham, Estimating F-statistics for the analysis of population structure, Evolution 38 (1984) 1358–1370.
[18] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, P.C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575.
[19] J.K. Pritchard, M. Stephens, P. Donnelly, Inference of population structure using multilocus genotype data, Genetics 155 (2000) 945–959.
[20] N.M. Kopelman, J. Mayzel, M. Jakobsson, N.A. Rosenberg, I. Mayrose, Clumpak: a program for identifying clustering modes and packaging population structure inferences across K, Mol. Ecol. Resour. 15 (2015) 1179–1191.
[21] D.A. Earl, B.M. vonHoldt, STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method, Conserv. Genet. Resour. 4 (2012) 359–361.
[22] X. Liu, F. Chen, Y. Niu, Y. Bian, S. Zhang, R. Zhu, C. Li, Population genetics of 30 insertion/deletion polymorphisms in Han Chinese population from Zhejiang Province, Forensic Sci. Int. Genet. 28 (2017) e33–e35.
[23] C.D. Royal, J. Novembre, S.M. Fullerton, D.B. Goldstein, J.C. Long, M.J. Bamshad, A.G. Clark, Inferring genetic ancestry: opportunities, challenges, and implications, Am. J. Hum. Genet. 86 (2010) 661–673.
[24] R. Santangelo, F. Gonzalez-Andrade, C. Borsting, A. Torroni, V. Pereira, N. Morling, Analysis of ancestry informative markers in three main ethnic groups from Ecuador supports a trihybrid origin of Ecuadorians, Forensic Sci. Int. Genet. 31 (2017) 29–33.
[25] C.M. Nievergelt, A.X. Maihofer, T. Shekhtman, O. Libiger, X. Wang, K.K. Kidd, J.R. Kidd, Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel, Investig. Genet. 4 (2013) 13.
[26] A.J. Pakstis, E. Haigh, L. Cherni, A.B.A. ElGaaied, A. Barton, B. Evsanaa, A. Togtokh, J. Brissenden, J. Roscoe, O. Bulbul, G. Filoglu, C. Gurkan, K.A. Meiklejohn, J.M. Robertson, C.X. Li, Y.L. Wei, H. Li, U. Soundararajan, H. Rajeevan, J.R. Kidd, K.K. Kidd, 52 additional reference population samples for the 55 AISNP panel, Forensic Sci. Int. Genet. 19 (2015) 269–271.
[27] R. Kosoy, R. Nassir, C. Tian, et al., Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America [J], Hum. Mutat. 30 (1) (2010) 69–78.
[28] C.X. Li, A.J. Pakstis, L. Jiang, Y.L. Wei, Q.F. Sun, H. Wu, O. Bulbul, P. Wang, L.L. Kang, J.R. Kidd, K.K. Kidd, A panel of 74 AISNPs: improved ancestry inference within Eastern Asia, Forensic Sci. Int. Genet. 23 (2016) 101–110.
[29] Y.L. Wei, L. Wei, L. Zhao, et al., A single-tube 27-plex SNP assay for estimating individual ancestry and admixture from three continents[J], Int. J. Legal Med. 130 (1) (2015) 1–11.