Forensic Science International: Genetics 35 (2018) 141–148

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Massively parallel sequencing of 124 SNPs included in the precision ID identity panel in three East Asian minority ethnicities

Jing Liu[a,1], Zheng Wang[a,1], Guanglin He[a], Xueying Zhao[b], Mengge Wang[a], Tao Luo[c], Chengtao Li[d], Yiping Hou[a,*]

[a] Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, Chengdu 610041, China
[b] Shanghai Key Laboratory of Crime Scene Evidence, Shanghai Research Institute of Criminal Science and Technology, Shanghai 200083, China
[c] Laboratory of Infection and Immunity, West China Center of Medical Sciences, Sichuan University, Chengdu 610041, China
[d] Shanghai Key Laboratory of Forensic Medicine, Institute of Forensic Science, Ministry of Justice, Shanghai 200063, China

ARTICLE INFO

Keywords:
Massively parallel sequencing (MPS)
Precision ID identity panel
Genetic polymorphism
Forensic genetics

ABSTRACT

Massively parallel sequencing (MPS) technologies can sequence many targeted regions of multiple samples simultaneously and are gaining great interest in the forensic community. The Precision ID Identity Panel contains 90 autosomal SNPs and 34 upper Y-Clade SNPs, which was designed with small amplicons and optimized for forensic degraded or challenging samples. Here, 184 unrelated individuals from three East Asian minority ethnicities (Tibetan, Uygur and Hui) were analyzed using the Precision ID Identity Panel and the Ion PGM System. The sequencing performance and corresponding forensic statistical parameters of this MPS-SNP panel were investigated. The inter-population relationships and substructures among three investigated populations and 30 worldwide populations were further investigated using PCA, MDS, cladogram and STRUCTURE. No significant deviation from Hardy-Weinberg equilibrium (HWE) and Linkage Disequilibrium (LD) tests was observed across all 90 autosomal SNPs. The combined matching probability (CMP) for Tibetan, Uygur and Hui were 2.5880×10^{-33}, 1.7480×10^{-35} and 4.6326×10^{-34} respectively, and the combined power of exclusion (CPE) were 0.999999386152271, 0.999999607712827 and 0.999999696360182 respectively. For 34 Y-SNPs, only 16 haplogroups were obtained, but the haplogroup distributions differ among the three populations. Tibetans from the Sino-Tibetan population and Hui with multiple ethnicities with an admixture population have genetic affinity with East Asian populations, while Uygurs of a Eurasian admixture population have similar genetic components to the South Asian populations and are distributed between East Asian and European populations. The aforementioned results suggest that the Precision ID Identity Panel is informative and polymorphic in three investigated populations and could be used as an effective tool for human forensics.

1. Introduction

Single nucleotide polymorphisms (SNPs) are widespread in the human genome and usually appear as a bi-allelic feature [1]. SNPs can provide advantages over STRs, such as a lower mutation rate, shorter PCR amplicon length, and ancestry/phenotypic/lineage information, which plays an efficient role in specialized applications in human forensics [2–5]. Due to the limited number of alleles (typically two) for each SNP locus, more SNPs are required to reach an equivalent sufficient power of discrimination or random match probability [6]. However, using a capillary electrophoresis method, the mainstream technology for forensic DNA marker typing, the number of SNPs that can be multiplexed together is limited and often cannot provide adequate power of discrimination or random match probability [7,8]. In recent years, technological progress due to massively parallel sequencing (MPS) has attracted much interest among researchers, especially in the field of forensic science. MPS can sequence many targeted regions of multiple samples simultaneously at single-base resolution with high coverage [9–11]. The Precision ID Identity Panel (Thermo Fisher Scientific, MA, USA) contains 90 autosomal SNPs and 34 upper Y-Clade SNPs, which was designed with small amplicons and optimized for forensic degraded or challenging samples. This MPS-SNPs panel was validated using a well-designed study and demonstrated that this robust, efficient and informative assay has the potential for use as a tool

* Corresponding author at: Institute of Forensic Medicine, West China School of Basic Medical Sciences & Forensic Medicine, Sichuan University, 3-16 Renmin South Road, Chengdu 610041, China.
E-mail address: forensic@scu.edu.cn (Y. Hou).
[1] The author contributed equally to this work and should be considered co-first author.

https://doi.org/10.1016/j.fsigen.2018.05.002
Received 14 December 2017; Received in revised form 9 May 2018; Accepted 10 May 2018

for forensic individual identification and paternity testing [12]. Meanwhile, in several forensic laboratories, this SNP panel has been investigated in population groups to assess forensic characteristics and applications [12–15].

Tibetan, Uygur and Hui ethnic groups are the most representative populations for China in East Asia and are research hotspots of genetic polymorphism, population admixture and substructure analysis for forensic geneticists and anthropologists. Tibet is home to a typical high-altitude adaption population, which mainly resides in the Qinghai-Tibet Plateau (Tibet Autonomous Region, Sichuan province and Qinghai province) [16,17]. This population has its own languages and characteristics, and many individuals subscribe to Lamaism. The Uygur minority is chiefly distributed in the Xinjiang Uygur Autonomous Region, whose pragmatic language belongs to the Turkic sub-branch of the Altaic language family [18]. The Hui minority ethnicity is the most widespread population in China and originates from central and west Asian (such as Arabian and Persian populations) migrating Muslim populations. This group has a significant admixture with many ethnic groups (Mongolian, Han and Uyghur nationalities) [19].

In the present study, the forensic parameters and Y-haplogroup distributions of 124 SNPs were first investigated in three Chinese ethnic minorities. Furthermore, inter-population comparisons among the three investigated populations and 30 worldwide populations (Shenyang Han and Chongqing Han [12], Iraqi [13], Basque [14] and 26 populations from the 1000 Genomes Project [20]) based on 88 overlapped autosomal SNPs were made to investigate the genetic similarity and diversity among diverse populations.

2. Materials and methods

2.1. Sample preparation

Blood samples were collected with the approval of the Ethics Committee at the Institute of Forensic Medicine, Sichuan University. Peripheral blood samples were collected from unrelated individual donors after receiving written informed consent. Blood samples include 62 Tibetan individuals living in Chengdu city of Sichuan province, 64 Uygur individuals residing in Kumul city in Xinjiang Uyghur Autonomous Region and 58 Hui individuals located in Wuzhong city of Ningxia Hui Autonomous Region. All donors belonged to the self-declared indigenous people for the corresponding ethnicity groups or had lived in the collection areas for at least three generations.

Genomic DNA was extracted using the PureLink Genomic DNA Mini Kit (Thermo Fisher Scientific) following the manufacturer's instructions. Genomic DNA was quantified using a Quantifiler Human DNA Quantification kit (Thermo Fisher Scientific) on an Applied Biosystems 7500 Real-time PCR System (Thermo Fisher Scientific) according to the manufacturer's recommendations, and DNA samples were normalized to 1.0 ng/μL. All samples were stored at −20 °C until library amplification.

2.2. Library preparation

Libraries were prepared according to the Precision ID Identity Panel protocol (Revision C) and the Precision ID Library Kit (Thermo Fisher Scientific) with minor modifications.

To amplify targets, 1.0 μL of DNA (1.0 ng) was amplified in 20 μL of reaction volume containing 10 μL of Precision ID Identity Panel, 4 μL of 5 × HiFi Mix, and 5 μL of nuclease-free water. Thermal cycling was conducted using the ProFlex 3 × 32 PCR System (Thermo Fisher Scientific) with the following conditions: enzyme activation at 99 °C for 2 min, amplification for 21 cycles of 15 s at 99 °C and 4 min at 60 °C, and holding at 10 °C. Next, PCR amplicons were partially digested by adding 2 μL of FuPa Reagent and incubated for 10 min at 50 °C, 10 min at 55 °C, 20 min at 60 °C and held up to 1 h at 10 °C.

For the ligation of libraries with adaptor, 2 μL of DNA ligase, 4 μL of Switch solution, 0.5 μL of Ion P1 Adapter, 0.5 μL of Xpress Barcode X (X is chosen from the Ion Xpress Barcode Adapters Kit for different samples) and 1 μL of Nuclease-free Water were added to 22 μL of digested PCR reaction, incubated for 30 min at 22 °C, 10 min at 72 °C and held up to 1 h at 10 °C. After ligation, libraries were purified with 45 μL of Agencourt AMPure XP Reagent (Beckman Coulter, FL, USA) as the volume recommended by the manufacturer with minor modifications. Libraries were washed three times using freshly prepared 70% ethanol.

To assess the yield and normalize the libraries, 9 μL of diluted libraries (1:100 dilution) were quantified on an Applied Biosystems 7500 Real-time PCR System (Thermo Fisher Scientific) with the Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific) following the manufacturer's instructions. Each library was normalized to 8 pM, and the normalized libraries of a sequencing run were pooled in equal volumes for template preparation.

2.3. Template preparation and sequencing

A 25-μL sample of the pooled library was used to generate template-positive Ion Sphere Particles (ISPs) containing clonally amplified DNA. Emulsion PCR was conducted on the Ion OneTouch 2 System (Thermo Fisher Scientific) with the Ion PGM Hi-Q OT2 Kit (Thermo Fisher Scientific). Template-positive ISPs were enriched with the Ion OneTouch Enrichment System (Thermo Fisher Scientific) following the manufacturer's instructions.

Sequencing was performed on the Ion Torrent PGM instrument using an Ion PGM Hi-Q Sequencing Kit (Thermo Fisher Scientific) and an Ion 318 Chip v2 (Thermo Fisher Scientific). Sequencing primers and Control ISPs were added to the enriched, template-positive ISPs. After annealing, the sequencing primer and sequencing polymerase were added. A final volume of 30 μL was loaded onto the Ion 318 chips following the manufacturer's instructions. The average sample number loading in each Ion 318 Chip was 60, as described in previous validation study [12].

2.4. Sequencing data analysis

All sequencing data were analyzed using the Torrent Suite Software 5.2.2 (Thermo Fisher Scientific), and the HID SNP Genotyper Plugin (Thermo Fisher Scientific) was used for secondary analysis with the target regions file (PrecisionID_IdentityPanel_targets.bed) and the hotspot regions file (PrecisionID_ IdentityPanel_hotspots.bed) using the default analysis settings. All of the SNP genotypes were reviewed by two independent scientists. Four statistical parameters [21], including the locus balance (LB), locus strand balance (LSB), heterozygote balance (HB) and noise level (NL), were used to evaluate the performance of the Precision ID Identity Panel in our testing populations. LB was calculated as the coverage of a locus divided by the mean coverage of all loci per sample. LSB was measured from the forward strand reads divided by the total reads. HB was estimated as the number of reads per base divided by the other base in the order A, C, G and T. NL was determined using the ratio of coverage for non-alleles to the total coverage.

Allele frequencies and corresponding forensic parameters, including observed heterozygosity (Ho), expected heterozygosity (He), matching probability (MP), probability of exclusion (PE), polymorphism information content (PIC), typical paternity index (TPI) and the Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) tests, were analyzed using the STRAF online software (http://www.cmpg.iee.unibe.ch/services/shiny/index_eng.html) [22]. Y-haplogroup determination was confirmed according to the International Society of Genetic Genealogy (ISOGG) Y-DNA Haplogroup Tree 2017 Version 12.274 (https://isogg.org/tree/index.html).

The pairwise F_{st} values among our three studied populations and 30 worldwide populations were performed using STRAF. The details for relative populations and their abbreviations are listed in Supplementary