Legal Medicine 34 (2018) 41–47

Contents lists available at ScienceDirect

Legal Medicine

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

Genetic resolution of applied biosystems™ precision ID Ancestry panel for seven Asian populations

Ji Hyun Lee[a], Sohee Cho[b], Moon-Young Kim[a], Dong Hoon Shin[b,c], Allah Rakha[d], Vasant Shinde[e], Soong Deok Lee[a,b,*]

[a] Department of Forensic Medicine, Seoul National University College of Medicine, Seoul, South Korea
[b] Institute of Forensic Science, Seoul National University College of Medicine, Seoul, South Korea
[c] Department of Anatomy, Seoul National University College of Medicine, Seoul, South Korea
[d] Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan
[e] Department of Archaeology, Deccan University, Pune, India

ARTICLE INFO

Keywords:
Ancestry Informative Markers (AIMS)
SNP
Asians
Massively parallel sequencing
Population differentiation

ABSTRACT

Massively parallel sequencing (MPS) offers additional information in cases that lack reference samples for comparison. The HID-Ion AmpliSeq Ancestry Panel is a forensic multiplex platform consisting of 165 autosomal markers designed to provide biogeographic ancestry information. We analyzed seven different population groups from Asia to assess the accuracy and reliability of analysis, using this panel.

In this study, we have designated 750 unrelated Asians, from southern China (n = 99), Beijing (n = 100), Japan (n = 101), Korea (n = 100), Vietnam (n = 100), Nepal (n = 100), India (n = 51), and Pakistan (n = 99). The likelihood ratios of 750 Asians were calculated, using the Torrent Server and the HID SNP Genotyper Plugin Version 4.3.2.

The results reveal that all Northeast Asians (China, Japan, and Korea), and Vietnamese, were predicted as East Asians with the highest ethnicity likelihood values from reference data. Although the samples from Nepal, India and Pakistan (Southwest Asians), were predominantly predicted as South Asians, there were seven cases of which results revealed as Europeans, with similar likelihood patterns. Population structure analysis indicated that Southwest Asians have a genetic profile that is distinguishable from those of other Asian populations.

This panel was validated for potential usefulness in forensic analysis and the findings could be used as a basis for building additional Asian specific reference databases. Nevertheless, additional analyses comprising larger sample sizes will be necessary, especially Southeast Asians, to fully understand the Asian population structure, and to discriminate them with further details.

1. Introduction

The introduction of massively parallel sequencing (MPS) technology within the forensic laboratories has become a solution to simultaneous analyses of thousands of SNPs [1]. The development of ancestry informative marker (AIM) assays is valuable in forensic genetics, because AIMs can provide investigative leads in cases of unknown individuals. Studies using hundreds of markers suggest worldwide populations can be grouped based on allele frequency differences, closely corresponding to their continental distribution [2–5].

A commercial kit for the assay is available, the HID-Ion AmpliSeq Ancestry Panel (Thermo Fisher Scientific, Waltham, USA) that include 165 SNP loci previously selected by two laboratories [6–7]. The kit is optimized to work with 1 ng of genomic DNA, and the analysis program provided by the kit manufacturer assigns unknown individuals to proper population groups and calculates various statistical values such as likelihoods ratio. Random match probabilities for the observed genotypes can be calculated and ranked in the reference populations that the analysis program uses as database. This produces a list of ranked likelihoods of ancestry of the observed profiles among the reference populations. However, it is not known whether the kits are capable of accurately identifying the ancestry of individuals from populations not represented in the reference population [8].

The human population in Asia has a complex colonization history, and is characterized by migration events of mankind. Although there are limitations in gene flow for dividing the nation into the

* Corresponding author at: Department of Forensic Medicine, College of Medicine, Seoul National University, 103, Daehak-ro, Jongno-Gu, Seoul 03087, South Korea.
E-mail address: sdlee@snu.ac.kr (S.D. Lee).

https://doi.org/10.1016/j.legalmed.2018.08.007
Received 9 January 2018; Received in revised form 23 July 2018; Accepted 22 August 2018
Available online 23 August 2018
1344-6223/

geographical area, previous phylogeographic studies using uniparental DNA markers revealed a genetic divergence between northern and southern East Asian populations, most likely caused by southern and northern migration routes into East Asia [9–12]. In this study, we divided Asia into Northeast Asia (NEA), Southeast Asia (SEA), and Southwest Asia (SWA) for the purpose of obtaining additional investigation leads. We studied populations according to divided regions by using the HID-Ion AmpliSeq™ Ancestry Panel in 750 individuals representing Asian countries. Numerous autosomal SNPs have been identified, revealing allele frequency differences between East and Southeast Asian populations, and even between North and South Han Chinese [13–16]. The panel can now distinguish population clusters globally, but various Asian populations are not included in the HID-Ion Ampliseq database used to calculate population membership likelihoods. For this reason, we evaluated the performance of the Ancestry Panel and assessed the accuracy and applicability of the Genotyper plugin of the Torrent Server on 750 peoples in Asia.

2. Material and methods

2.1. DNA sample collection

DNA samples for Han Chinese South [CHS], Han Chinese in Beijing [CHB], Japanese in Tokyo [JPT], and Kinh in Ho Chi Minh City [KHV], were purchased from the Coriell Institute for Medical Research in Camden, New Jersey. Indian, Korean, Nepalese, and Pakistani samples were collected, and a complete list of these samples and geographic origins is presented in Table 1. Korean samples were collected directly from our laboratory, while other samples were collected with DNA extracts and FTA cards, under a material transfer agreement (MTA) with universities in each country. The FTA samples were extracted using the Maxwell CSC instrument (Promega Corporation, Madison, WI, USA), and reagents with a 1.2 mm^2 punch, following the manufacturer's protocols. Quantity of DNA was determined by the QuantStudio® 5 Real-Time PCR System (Thermo Fisher) using the Quantifiler Human DNA Quantification Kit (Thermo Fisher).

To check the quality of the sample, and to obtain additional information, autosomal STR and Y-STR typing were performed, using PowerPlex® fusion system and PowerPlex® Y23 System (Promega) and Mitochondrial DNA sequencing, following Rakha et al. [17]. All Y-STR haplotypes were submitted to Whit Atheys' Haplogroup Predictor, to obtain probabilities for the inferred haplogroups. The strategy was used to avoid redundant SNP-typing, though verification of the haplogroup with Y-SNPs was required [18]. Mitochondrial haplogroup assignment was processed with MitoTool (http://www.mitotool.org), using control region according to the PhyloTree Build 17 database [19]. The PCR products were analyzed using a 3130 Genetic Analyzer (Life Technologies), and the data was analyzed using GeneMapper® ID Software (Life Technologies). The collection of samples for this study was conducted in accordance with the guidelines of the Institutional Review Board (IRB) of Seoul National University Hospital, Korea, after approval by the IRB.

3. Library preparation, quantification and sequencing

Library construction was performed using the Precision ID Ancestry Panel and the Ion Precision ID Library Kit. Amplification of genomic DNA targets, partial digestion of primer sequences, ligation of barcode adapters to the amplicons and purification and quantification of the unamplified library were conducted according to the user guide [20]. Amplification of 1 ng input DNA was performed on the GeneAmp® PCR System 9700 (Thermo Fisher) and followed by the direct addition of a total of 2 μl FuPa reagent into each PCR product and then incubation. A total of 4 μl Switch solution, 2 μl diluted IonCode™ Barcode Adapters and 2 μl DNA ligase were consecutively added into each of 22 μl digested amplicon, followed by incubation for 30 min at 22 °C, 10 min at 72 °C and 1 h at 10 °C. Each unamplified library was purified using 1.5x Agencourt AMPure® XP reagent (Beckman Coulter, FL, USA) according to the user guides [20]. Purified unamplified libraries were quantified on the QuantStudio® 5 with the Ion Library TaqMan™ Quantitation Kit (Thermo Fisher) and then diluted to 30 pM. All barcoded libraries were pooled in equivolume. Template preparation consisting of emulsion PCR, enrichment of beads containing template, and chip loading was performed on the Ion Chef, and loaded sequencing chips were placed on the Ion S5xl System (Thermo Fisher Scientific), together with Ion 540™ Chip Kit (Thermo Fisher Scientific) following manufacturer's protocols [20]. We put 200 samples on one chip to make the average coverage over 1000 depths and the lowest coverage over 100 depths. Sequence analysis included signal processing, base-calling, and barcode de-convolution was conducted with the HID SNP Genotyper v4.3.2 Plugin with default settings.

3.1. Data analysis

The heterozygote balances were calculated, as the number of reads for one nucleotide, divided by the number of reads, for the other nucleotide in the order A, C, G, and T. Minor allele frequency, observed and expected heterozygosities, as well as the Linkage Disequilibrium (LD) among SNPs, were estimated using the HAPLOVIEW software [21], and figures were drawn with Sigmaplot 12.0 (Systat Software, Erkrath, Germany). The likelihood of observing these genotypes was calculated with the HID SNP Genotyper plugin of the Torrent Server. This server contains data from a total of 65 populations divided East Asia, Asia, Oceania, Siberia, North America, Europe, Africa, and South America into regional standards to calculate the population likelihood. There are 10 ethnic groups (Japanese-HapMap, Japanese, Koreans, Han-HapMap, Cambodians Khmer, Ami, Hakka, Taiwanese Han, Malaysians and Atayal) in East Asia and 14 ethnic groups (Lao Loum, Hazara, Khanty, Kachari, Keralite, Pashtun, Mohanna, Thoti, Negroid Makrani, Komi-Zyrian, Kuwaiti, Palestinian, Druze and Jews Yemenite) in Asia. The server uses this data to calculate the likelihood values and lists them in order from highest to lowest. This software also produces

Table 1
Details of populations analyzed.

Geographical region	Population sample description	Sample size (N)	Data source
Northeast Asia	Han Chinese South [CHS]	99	Coriell Institute
	Han Chinese in Beijing, China [CHB]	100	Coriell Institute
	Japanese in Tokyo, Japan [JPT]	101	Coriell Institute
	Korean from Seoul, Korea [KOR]	100	Seoul National Univ.
Southeast Asia	Kinh in Ho Chi Minh City, Vietnam [KHV]	100	Coriell Institute
Southwest Asia	Nepalese [NP]	100	Seoul National Univ.
	Indian [IN] (Residing at Deccan region)	51	Seoul National Univ.
	Pakistani in Pathan, Pakistan [PT]	49	Seoul National Univ.
	Punjabi in Lahore, Pakistan [PJ]	50	UHS Lahore
		750	