Forensic Science International: Genetics 13 (2014) 45–52

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Forensic application of SNP-based resequencing array for individual identification

Sohee Cho [a], Hyung Jin Yu [b], Jisung Han [b], Yoonsoo Kim [b], Jihyun Lee [a], Soong Deok Lee [a,c,*]

[a] Department of Forensic Medicine, Seoul National University, College of Medicine, Seoul, Republic of Korea
[b] DNA Link, Seoul, Republic of Korea
[c] Institute of Forensic Science, Seoul National University, College of Medicine, Seoul, Republic of Korea

ARTICLE INFO

Article history:
Received 21 March 2014
Received in revised form 2 June 2014
Accepted 1 July 2014

Keyword:
Single nucleotide polymorphism
DNA chip
Resequencing array
Human identification

ABSTRACT

In forensic field investigations using single nucleotide polymorphism (SNP) have been performed for various purposes. Based on the characteristics of SNP, it is essential to have a multi-amplification technology and a platform to analyze the amplified SNP markers accurately. Here, we have developed a platform based on the resequencing array of Affymetrix analyzing 169 SNP markers amplified via multiplex PCR and verified its forensic application. From the 1000 genomes database, the SNP markers were selected under the condition of less than 0.04 fixation index (Fst) and 0.3 linkage disequilibrium (LD) R^2 value, and 0.4–0.5 minor allele frequency (MAF). It was identified that more than 120 out of 169 SNPs were able to be typed with approximately 10 pg of DNA and artificially degraded samples in various tests. The DNA extracted from bones also showed a similar rate of success. The results indicated that our platform has a potential role to assist the current short tandem repeat (STR) method in analyzing harsh samples, such as bone or degraded DNA. With a possibility to expand the platform, it was expected to apply to various uses in different areas in the future.

1. Introduction

In recent years, Single Nucleotide Polymorphism (SNP) methodology has been paid attention for a potential role to assist the conventional Short Tandem Repeat (STR) practice in forensics [1–3]. SNPs have not only a lower mutation rate than STR and provide information on race and phenotypes, but also are suited for the automated qualitative allele calling system due to their biallelic characteristic. In addition, the amplicon size can be minutely designed, and the multiplexing technology allows for the high-throughput assay. The fact that a lot of forensic samples exist in small amounts or in degraded conditions at a crime scene or mega-disaster also makes SNP a fit method of analysis. Still, several drawbacks, such as a restricted number of SNPs that can be typed at once in forensic samples and a lack of appropriate statistical analysis methods, remain to be solved in order to expand the platform's usability. Nevertheless, ongoing studies of SNP among many research groups are expected to supplement the shortcomings and enhance the platform's overall quality [4,5].

The SNP research was not only limited within forensics, but also actively pursued in medicine and other areas of interest [6–8]. In order to identify SNP mutations, various methods were used depending on the specific purpose: (1) for direct discrimination [7,9], hybridization, primer extension, enzymatic cleavage, and ligation were used, and (2) for instrumental identification [10], SNaPshot, microarray, and mass spectrometry were often used. Although the choice of the techniques in carrying out a SNP research could be flexible, it is necessary to consider several factors such as the accuracy of SNP detection, number of SNPs detectable, possibility of automation, and the sensitivity in SNP recognition [11]. The amount of information obtainable from SNP is dependent on the number of SNPs, and it was reported that 40–60 SNPs were required to achieve a similar discrimination power of 13–15 STRs since SNPs are not as polymorphic as STRs [12–14]. While TaqMan and SNaPshot were presented as common methods usually for SNP analysis [8,15,16], it was not easy to overcome the aforementioned parameters due to their limited number of SNPs able to be typed at a time.

In present study, we tested the efficacy of AccuID Ver. 1.0 (hereafter referred to as AccuID), a SNP based microarray system

* Corresponding author at: Department of Forensic Medicine, Seoul National University College of Medicine, 28 Yongon-Dong, Chongno-Gu, Seoul 110-799, Republic of Korea. Tel.: +82 2 740 8359; fax: +82 2 764 8340.
E-mail address: sdlee@snu.ac.kr (S.D. Lee).

http://dx.doi.org/10.1016/j.fsigen.2014.07.005
1872-4973/

made by combining the Affymetrix Resequencing array technology (Santa Clara, CA, USA) and multiplex PCR technique (DNA Link, Inc., Seoul, Korea). Resequencing array was designed to sequence a portion of genome on the array having a high sensitivity and accuracy due to bidirectional sequencing [17], which was often used in germline mutation and cancer research. Using the array consisting of 169 SNP markers that were selected from 1000 genomes database, accuracy and sensitivity of the array and applicability of degraded samples were tested for forensic purpose. In terms of possibility of adding more SNP markers to the array, it could be designed to include more informative and meaningful SNP markers, such as lineage markers, phenotype predicting markers, and ancestral informative markers (AIMs), to increase the overall efficacy of the current array [18–20].

2. Materials and methods

2.1. Selection of SNP markers and primer design

Total of 169 autosomal SNP markers were selected from 1000 genomes database containing the data 1094 individuals. The chromosome distribution of markers, average amplicon size, minor allele frequency (MAF), and fixation index (Fst) are shown in Table 1. Prior to the selection the markers with high typing rate were sorted from the preliminary test. Less than 0.3 R^2 value of linkage disequilibrium (LD) was considered to minimize any connection between the markers. When calculating Fst, 1094 individual's data consisting of 14 different ethnic types were grouped into four ethnic categories: (1) African: American of African Ancestry in SW USA (61), Luhya in Webuye, Kenya (97) and Yoruba in Ibadan, Nigeria (88), (2) Asian: Han Chinese in Beijing, China (97), Han Chinese South, China (100) and Japanese in Tokyo, Japan (89), (3) European: Utah Residents (CEPH) with Northern and Western European ancestry (87), Finnish in Finland (93), British from England and Scotland, UK (89), Iberian populations in Spain (14) and Toscani in Italia (98), and (4) Latin American: Colombian in Medellin, Colombia (60), Mexican Ancestry in Los Angeles, USA (66) and Puerto Rican in Puerto Rico (55). Less than 0.04 (0.014 on average) Fst value was used when selecting markers in order to maximize the discrimination power without racial discrepancy. Primer was designed using Autoprimer.com's web tool (Beckman Coulter, Inc., Brea, CA, USA).

2.2. Samples preparation

Two human genomic DNA, NA12878 (a female of European Ancestry from Utah (CEU)), and NA10851 (a male of European Ancestry from Utah (CEU)) were obtained from the Coriell Institute for Medical Research (Camden, NJ, USA). 9947A and K562 control DNA (Promega, Madison, WI, USA) were also used to validate accuracy, sensitivity, and susceptibility to digestion in microarray platform. A sensitivity test was performed with diluted K562 DNA using six different DNA amounts: 5, 1, 0.5, 0.1, 0.05, and 0.01 ng. Two DNA samples from a femur bone, which had been kept underground for over 20 years, were extracted using a phenol/chloroform extraction method. Hair, swab and cigarette butt samples were collected from a volunteer with informed consent, and DNA extraction was performed using QIAamp DNA Investigator Kit (QIAGEN, Valencia, CA, USA) following the manufacturer's instructions. The DNA concentration was measured using NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA). To find out the typing rate of SNP in relation to the degree of degradation, an artificial degradation was induced by treating DNase I to the control DNA. Degraded DNA was prepared by following a modified method according to Nicklas and Buel [22]. Briefly, a 50 μl reaction mixture containing 1X reaction buffer, 1 μg DNA, and 1 U DNase I were prepared on ice, and 10 μl aliquots were isolated to serve as the "0" time point. The remaining mixture was divided into three separate tubes to serve as the 30 s, 90 s, and 15 min time point, and then incubated at 37 °C. To inactivate the enzyme at each time point, each tube was incubated at 65 °C for 10 min with 1 μl stop solution added.

2.3. Multiplex PCR

Multiplex amplification of DNA was performed in 50 μl reaction containing DL Premix (DNA Link, Inc., Seoul, Korea) and pooled 169 primer sets. For degraded DNA, such as bone DNA, maximum volume of DNA extracts were included in PCR reactions. Thermal cycling was performed with the following protocol:

Table 1
169 SNP markers included in the AccuID system with their amplicon size, MAF, and Fst in relation to each chromosome.

Chr	Number of markers	Minimum amplicon size	Maximum amplicon size	Average amplicon size	Average MAF	Average Fst
1	11	90	154	120	0.474	0.010
2	15	90	163	116	0.467	0.011
3	10	89	155	131	0.464	0.012
4	12	96	151	116	0.470	0.014
5	14	90	146	106	0.464	0.012
6	17	90	152	120	0.476	0.009
7	8	90	147	116	0.477	0.015
8	11	96	166	127	0.457	0.013
9	8	90	144	124	0.455	0.014
10	1	–	–	131	0.458	0.035
11	12	90	149	108	0.461	0.010
12	6	90	162	127	0.477	0.009
13	11	96	151	120	0.469	0.015
14	6	90	151	112	0.452	0.013
15	6	90	153	113	0.461	0.010
16	3	90	119	102	0.472	0.005
17	2	101	146	124	0.404	0.020
18	8	90	151	117	0.462	0.013
19	1	–	–	144	0.415	0.023
20	5	100	151	121	0.472	0.010
21	0	–	–	–	–	–
22	2	90	116	103	0.412	0.018
Total	169	89	166	118	0.458	0.014

MAF, minor allele frequency. Fst, fixation index