Forensic Science International: Genetics 11 (2014) 166–173

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

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

Hathaichanoke Boonyarit [a,c,d], Surakameth Mahasirimongkol [b], Nuttama Chavalvechakul [c], Masayuki Aoki [d], Hanae Amitani [d], Naoya Hosono [d], Naoyuki Kamatani [d], Michiaki Kubo [d,*], Patcharee Lertrit [a,**]

[a] Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand
[b] Medical Genetic Section, National Institute of Health, Department of Medical Sciences, Ministry of Public Health, Nonthaburi 11000, Thailand
[c] Institute of Forensic Medicine, Police General Hospital, Royal Thai Police, Bangkok 10330, Thailand
[d] Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN Yokohama Institute, Kanagawa 230-0045, Japan

ARTICLE INFO

Article history:
Received 10 December 2013
Received in revised form 27 February 2014
Accepted 2 March 2014

Keywords:
Human identification
Degraded DNA
Short tandem repeat (STR)
Single nucleotide polymorphism (SNP)
Invader assay

ABSTRACT

This study describes the development of a SNP typing system for human identification in the Thai population, in particular for extremely degraded DNA samples. A highly informative SNP marker set for forensic identification was identified, and a multiplex PCR-based Invader assay was developed. Fifty-one highly informative autosomal SNP markers and three sex determination SNP markers were amplified in two multiplex PCR reactions and then detected using Invader assay reactions. The average PCR product size was 71 base pairs. The match probability of the 54-SNP marker set in 124 Thai individuals was 1.48×10^{-21}, higher than that of STR typing, suggesting that this 54-SNP marker set is beneficial for forensic identification in the Thai population. The selected SNP marker set was also evaluated in 90 artificially degraded samples, and in 128 naturally degraded DNA samples from real forensic casework which had shown no profiles or incomplete profiles when examined using a commercial STR typing system. A total of 56 degraded samples (44%) achieved the matching probability (PM) equivalent to STR gold standard analysis (successful genotyping of 44 SNP markers) for human identification. These data indicated that our novel 54-SNP marker set provides a very useful and valuable approach for forensic identification in the Thai population, especially in the case of highly to extremely degraded DNA.

In summary, we have developed a set of 54 Thai-specific SNPs for human identification which have higher discrimination power than STR genotyping. The PCRs for these 54 SNP markers were successfully combined into two multiplex reactions and detected with an Invader assay. This novel SNP genotyping system also yields high levels of genetic information from naturally degraded samples, even though there are much more difficult to recover than artificially degraded samples.

1. Introduction

The most common DNA typing system for forensic identification is a multiplex PCR system based on short tandem repeat (STR) loci [1,2]. A multiplex STR-based typing system for forensic identification is now commercially available and routinely used [3]. The STR-based typing system is considered a standard approach and has been proven to overcome most problems in forensic identification with an elevated discrimination power [4–6].

Nevertheless, there remains a problem with the STR-based system when it is applied to forensic identification of severely fragmented DNA, and such samples occur regularly in routine casework. Poor quality DNA samples produce partial STR profiles or no STR profiles when commercial STR typing kits (amplicon sizes of approximately 150–450 base pairs [7]) are used, resulting in a loss of genetic information [8]. To combat this problem, a mini-STR set has been developed, by placing amplification primers as

* Corresponding author at: Laboratory for Genotyping Development, RIKEN Center for Genomic Medicine, The Institute of Physical and Chemical Research, 1-7-22 Suehiro-cho, Tsurumi, Yokohama, Kanagawa 230-0045, Japan.
** Corresponding author at: Department of Biochemistry, Faculty of Medicine Siriraj Hospital, Mahidol University, 2 PrannokRoad, Bangkoknoi, Bangkok 10700, Thailand. Tel.: +66 2419 7581.
E-mail addresses: mkubo@src.riken.jp (M. Kubo), patcharee.ler@mahidol.ac.th, lertrito@yahoo.com, patlertrit@gmail.com (P. Lertrit).

http://dx.doi.org/10.1016/j.fsigen.2014.03.010
1872-4973/

close as possible to the core STR repeat, then generating amplicon sizes ranging from 70 to 283 bp [9], enabling extraction of the same information (genotype) as longer amplicons. However, 60% of the loci in the mini-STR kit overlap with commercial STR loci: hence in many highly degraded samples both the mini-STR and STR systems fail to yield adequate DNA profiles. Another genotyping system is needed for better identification of such samples.

One promising approach to obtain genetic information from highly degraded DNA is to use single nucleotide polymorphisms (SNPs) as markers [10–13]. Theoretically, SNP-based systems can recover greater information from highly degraded DNA than STRs, since they only require smaller amplicons. However, because each SNP only has a single polymorphic nucleotide, large numbers of individual SNPs are required to obtain reasonable discriminatory power and define a unique profile. There is an emerging consensus that sets of forty to sixty SNPs can be useful for forensic DNA testing, providing similar discriminatory power to the sets of 13–15 STR loci which are most commonly used today [14–16]. Many studies have investigated the use of SNPs in degraded DNA samples [17–21]. Unfortunately, many of the previously reported SNP marker sets require PCR amplicon sizes of more than 100 base pairs (typically amplicon sizes vary between 70 and 150 base pairs in the SNP sets). This limits their capacity for generating adequate data from severely degraded samples. Moreover, the discriminatory power of these previously reported SNPs have not been evaluated in the Thai population.

This study focused on the selection of SNP loci with high discrimination power in the Thai population, suitable for forensic DNA testing from routine case work and in severely degraded bone specimens. The aim of the study was to develop an informative, SNP-based marker system for forensic identification in the Thai population, especially for the analysis of highly degraded DNA samples.

2. Materials and methods

2.1. DNA sample preparation

The normal DNA samples in this study consisted of blood samples ($n = 91$) and buccal swab samples ($n = 33$). All samples were collected from unrelated Thai individuals who live in Thailand, and their pedigrees were traced back at least three generations. All of these 124 individuals gave informed consent. For the degraded DNA samples, bone samples ($n = 128$) from deceased individuals were obtained from the Institute of Forensic Medicine, Police General Hospital, Royal Thai Police, Bangkok, Thailand; with permission letter from legal representative of autopsy for human identification purpose. All of the decomposed human remains were submitted for routine forensic identification and were collected by forensic physicians. This study was approved by the Ethics Committee of Mahidol University, Faculty of Medicine Siriraj Hospital (no. *Si* 530/2009), and the study was conducted in accordance with the principles of the World Medical Association's Declaration of Helsinki. Some general information on the samples used in this study is shown in Supplementary Table 1. For comparison with the naturally degraded samples, SNPs were also genotyped in the 90 Asian HapMap artificially degraded DNA samples. These 90 Asian HapMap DNA samples come from 90 unrelated individuals in the Han Chinese (23 females and 22 males) and Japanese (22 females and 23 males) populations. All HapMap samples were purchased from the Coriell Institute for Medical Research (Camden, NJ, USA).

Genomic DNA extractions from buccal swab and bone samples were carried out using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) and the QIAamp DNA Micro Kit (QIAGEN, Hilden, Germany) respectively. Whole blood samples were extracted by a standard phenol-chloroform procedure. Intact DNA samples were quantified using a NanoDrop® ND-1000 Spectrophotometer (NanoDrop Technologies). Quantification of DNA from the decomposed tissue samples was performed using Picogreen (Molecular Probes, Eugene, OR). All of these procedures with kits were done according to the manufacturers' instructions.

2.2. SNP marker selection

SNP markers for human identification were selected according to the following criteria: (1) The SNPs selected were located outside the coding regions of each chromosome according to the NCBI dbSNP database (http://www.ncbi.nlm.nih.gov/snp/) [22]. SNPs were required to have minor allele frequency (MAF) greater than 40% in the Asian, European, African-American and Sub-Saharan African population samples in the NCBI database. (2) SNPs were selected to overlap with SNPs in the Affymetrix 100 K SNP array, 500 K SNP array, and Affymetrix™ Genome-Wide SNP array 6.0 K [23], in order to acquire more accurate genotype calls. (3) SNPs were required to have MAF of more than 0.4 in at least six of the eleven population groups in the HapMap phase III (Release 2) population database (http://hapmap.ncbi.nlm.nih.gov/) [24]. SNPs located in regions of copy number variation (CNV) were excluded. Linkage disequilibrium (LD) values from a Thai SNP study [25] were used to ensure that selected SNPs were unlinked. The median r^2-value was 0.004 (range 0–0.009). For gender determination SNP, three SNPs on the sex chromosomes were selected: one SNP in the amelogenin gene on the Y chromosome and two SNPs on the X and Y chromosomes reported by Hwan Young Lee et al. [26]. After applying these filtering criteria, 59 SNP markers were selected for further study (Supplementary Table 2).

2.3. Primer and probe design

PCR primers for selected SNPs were designed using Primer 3 software (http://www.bioinformatics.nl/cgi-bin/primer3/primer3_www.cgi) [27] to generate amplicon sizes less than 100 bp (Table 1). The average PCR amplicon size across this SNP marker set was 71 bp (range 51–98 bp). The primary probes and Invader oligonucleotides corresponding to each SNP allele were designed and synthesized according to the criteria of the Invader assay [28]. All PCR primers and HPLC-grade probes were synthesized by the Fasmac oligonucleotide service (Kanagawa, Japan). All reagents for the Invader assay were purchased from Third Wave Technologies, Inc. (Madison, WI) [29].

2.4. Multiplex PCR-based invader assay

A multiplex PCR reaction was performed in a reaction volume of 10 μl containing 0.1 μM of each primer, 1 mM of dNTPs, 4.5 mM of $MgCl_2$, 0.125 units of ExTaqHS (Takara Bio Inc, Shiga, Japan) and 2 μl (0.5 ng/μl) of DNA template. Amplification reactions were performed using a 96-well GeneAmp PCR system 9700 Thermal cycler (Applied Biosystems, Foster City, CA, USA) with the following conditions: 95 °C 1 min followed by 40 cycles of 95 °C 15 s, 58 °C 1 min, then the reaction mixtures were held at 4 °C until further use. Two sets of multiplex PCRs were set up in which 30 and 24 primers were randomly selected (Table 1). The Invader reaction was performed in a 384-well plate with a reaction volume of 4 μl containing 1 μl of multiplex PCR product (20-fold dilution), following the protocol recommended by Third Wave Technologies (Madison, WI). Each reaction volume of 3 μl contained 1X signal buffer (Third Wave Technologies), 1X FRET22/FRET7 mixture probes (Third Wave Technologies), 20 ng/μl cleavage enzyme (Cleavase 2.0, Third Wave Technologies) and 1X primary probe/Invader oligonucleotide cocktail. The Invader reaction was