Forensic Science International: Genetics 22 (2016) 1–7

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Massively parallel sequencing of 17 commonly used forensic autosomal STRs and amelogenin with small amplicons

Eun Hye Kim[a,b], Hwan Young Lee[b], In Seok Yang[b], Sang-Eun Jung[b], Woo Ick Yang[b], Kyoung-Jin Shin[a,b,*]

[a] Brain Korea 21 PLUS Project for Medical Science, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea
[b] Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea

ARTICLE INFO

Article history:
Received 6 March 2015
Received in revised form 9 December 2015
Accepted 5 January 2016
Available online 8 January 2016

Keywords:
Next-generation sequencing
Autosomal STR
Small-sized amplicon
Sequence variation
Mixture
Degraded DNA

ABSTRACT

The next-generation sequencing (NGS) method has been utilized to analyze short tandem repeat (STR) markers, which are routinely used for human identification purposes in the forensic field. Some researchers have demonstrated the successful application of the NGS system to STR typing, suggesting that NGS technology may be an alternative or additional method to overcome limitations of capillary electrophoresis (CE)-based STR profiling. However, there has been no available multiplex PCR system that is optimized for NGS analysis of forensic STR markers. Thus, we constructed a multiplex PCR system for the NGS analysis of 18 markers (13CODIS STRs, D2S1338, D19S433, Penta D, Penta E and amelogenin) by designing amplicons in the size range of 77–210 base pairs. Then, PCR products were generated from two single-sources, mixed samples and artificially degraded DNA samples using a multiplex PCR system, and were prepared for sequencing on the MiSeq system through construction of a subsequent barcoded library. By performing NGS and analyzing the data, we confirmed that the resultant STR genotypes were consistent with those of CE-based typing. Moreover, sequence variations were detected in targeted STR regions. Through the use of small-sized amplicons, the developed multiplex PCR system enables researchers to obtain successful STR profiles even from artificially degraded DNA as well as STR loci which are analyzed with large-sized amplicons in the CE-based commercial kits. In addition, successful profiles can be obtained from mixtures up to a 1:19 ratio. Consequently, the developed multiplex PCR system, which produces small size amplicons, can be successfully applied to STR NGS analysis of forensic casework samples such as mixtures and degraded DNA samples.

1. Introduction

Short tandem repeat (STR) analysis using capillary electrophoresis (CE) has been the gold standard method in forensic genetics [1–4]. The CE method can detect fluorescent dye-labeled fragments based on length and is very simple and easy. However, because the analyzable number of loci with overlapping size ranges is restricted to the number of different fluorescent dyes, it is sometimes difficult to simultaneously analyze many markers [5,6].

Next-generation sequencing (NGS) technology, which can produce massive parallel sequencing data, is now expected to become an alternative or additional method to overcome these limitations of CE-based forensic STR typing because NGS does not require size separation between amplicons and therefore enables the simultaneous analysis of a large number of markers [7]. Also, the application of NGS technology to STR analysis can provide information on sequence variation, which allows for the discrimination of same sized alleles [4,8–13]. Moreover, targeted sequencing data produces high depths of coverage [12], thereby enhancing the possibility of successful analysis of challenging forensic specimens such as mixtures and highly degraded DNA samples.

Previous studies have evaluated the potential of NGS for forensic STR analysis using several in-house multiplex PCR systems or commercial kits [5,6,10,12–15]. Scheible et al. [12] showed the potential of NGS technology in recovering genetic information from degraded and low DNA quantity samples such as skeletal remains using a series of multiplexes containing miniSTRs. Fordyce et al. [13] reported a 10 STR-plex with a sensitivity as low as 50 pg. However, these studies used multiplex PCR systems that often do not include the whole 13 Combined DNA Index System (CODIS) STRs [10,13] or have not undergone optimization for NGS analysis

* Corresponding author at: Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 120-752, South Korea. Fax: +82 2 362 0860.
E-mail address: kjshin@yuhs.ac (K.-J. Shin).

http://dx.doi.org/10.1016/j.fsigen.2016.01.001
1872-4973/

[14]. Therefore, there is a need to develop a multiplex PCR system that is optimized for NGS analysis of 13 CODIS STRs. Because amplicon size is the most significant factor for the read lengths of the NGS platform to cover the entire range of STRs for exact typing [16,17], we developed a multiplex PCR system with small sized amplicons for 18 markers including 13CODIS STR loci, four loci from the commonly used Identifiler or PowerPlex 16 system (D2S1338, D19S433, Penta D and Penta E) and amelogenin. Then, we evaluated the system with control DNA as single-source samples, mixtures and degraded DNA samples.

2. Materials and methods

2.1. DNA samples

One nanogram each of control male DNA 2800 M (Promega, Madison, WI, USA) and female DNA 9947A (Promega) were used as single-source samples. The concentrations of DNA samples were measured using the Quantifiler® Duo DNA Quantification Kit (Applied Biosystems, Foster City, CA, USA) following the manufacturer's instructions. Mixtures of two control DNAs were prepared in mixture ratios of 1:1, 1:3, 1:6, 1:9, 1:19 and 1:49 (male: female). The total amount of mixed DNA at each ratio was 1 ng.

An artificially degraded DNA sample was prepared by digesting 1.7 μg of K562 DNA (Promega) with 0.006 U of DNase I (New England Biolabs, Ipswich, MA, USA) at 37 °C for 15 min and incubating at 75 °C for 10 min. Enzyme digestions of the K562 DNA sample were duplicated, and DNA fragmentation was confirmed in agarose gel electrophoresis.

2.2. Construction of multiplex PCR system

The STR markers selected for a multiplex PCR were composed of 18 forensic markers that included 13CODIS STRs (D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, CSF1PO, FGA, TH01 and vWA), four loci from commonly used commercial kits (D2S1338, D19S433, Penta D and Penta E) and amelogenin. The allele information about the selected STR markers was collected from STRBase (http://www.cstl.nist.gov/biotech/strbase) and the DNA sequence information of target markers was collected from GenBank (http://www.ncbi.nlm.nih.gov/genbank). Primers for PCR amplification were designed using the Primer3 (http://frodo.wi.mit.edu/primer3/input.htm) program such that the amplicon sizes of the 18 targeted markers were less than 250 bp, and the primers did not bind to the region with a greater than 1% mutation rate based on the NCBI SNP information (http://www.ncbi.nlm.nih.gov/SNP/). To determine primer interference and PCR efficiency among designed candidate primers, PCR amplicons with fluorescent dye-labeled primers were temporarily designed and tested.

2.3. Multiplex PCR amplification and quantification

For single-sources and mixtures, the multiplex PCR reaction was carried out in a 25 μL reaction volume containing 1 ng of template DNA, 4.5 U of AmpliTaq Gold® DNA Polymerase (Applied Biosystems), 2.5 μL of Gold ST*R 10 × Buffer (Promega), and 0.351.0 μM of each primer (Table 1). PCR was conducted on a Veriti® 96-Well Thermal Cycler (Applied Biosystems) under the following conditions: 95 °C for 11 min; 33 cycles of 94 °C for 20 s, 59 °C for 90 s, and 72 °C for 60 s and a final extension at 60 °C for 45 min with a 4 °C soak. For an artificially degraded DNA sample, the amplification reaction was carried out under the same PCR conditions except that 200 pg of artificially degraded DNA and 6.0 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems) were used in 36 cycles of PCR amplification.

Table 1
PCR primer sequences and final concentrations of the developed multiplex PCR system for 18 forensic markers.

Locus	Primer Sequences (5′–3′)	Conc. (μM)
D2S1338	TGGAAACAGAAATGGCTTGG	0.8
	AGTTATTCAGTAAGTTAAAGGATTGC	0.8
D3S1358	CAGTCCAATCTGGGTGACAG	0.5
	ATCAACAGAGGCTTGCATGT	0.5
D5S818	TGATTTTCCTCTTTGGTATCCTT	0.6
	CAACATTTGTATCTTTATCTGTATCCT	0.6
D7S820	TGATAGAACACTTGTCATAGTTTAGAA	0.5
	CTCATTGACAGAATTGCACCA	0.5
D8S1179	TTTTTGTATTTCATGTGTACATTCGT	0.9
	GTAGATTATTTTCACTGTGGGGAA	0.9
D13S317	TCTAACGCCTATCTGTATTTACAA	0.8
	AGACAGAAAGATAGATAGATGATTGA	0.8
D16S539	AATACAGACAGACAGACAGGTG	0.5
	AGCATGTATCTATCATCCATCTCTG	0.5
D18S51	GTTGCTACTATTTCTTTTCTTTTTCTC	0.9
	CTGAGTGACAAATTGAGACCTTG	0.9
D19S433	GCAAAAAGCTATAATTGTACCAC	0.6
	AAAAATCTTCTCTCTTTCTTCCTCTC	0.6
D21S11	AATTCCCCAAGTGAATTGCC	0.7
	GGTAGATAGACTGGATAGATAGACGA	0.7
CSF1PO	ACTGCCTTCATAGATAGAAGAT	0.5
	GACCCTGTTCTAAGTACTTCCT	0.5
FGA	AAATAAAATTAGGCATATTTACAAGC	1.0
	GCCAGCAAAAAAGAAAGGAA	1.0
Penta D	GCAAGACACCATCTCAAGAAAG	1.0
	TGGTCATAACGATTTTTTTGAGA	1.0
Penta E	GGCGACTGAGCAAGACTCA	1.0
	TGGGTTATTAATTGAGAAAACTCCTT	1.0
TH01	GATTCCCATTGGCCTGTTC	0.5
	CAGGTCACAGGGAACACAGA	0.5
TPOX	CAGAACAGGCACTTAGGGAAC	0.4
	TCCTTGTCAGCGTTTATTTGC	0.4
vWA	GAATAATCAGTATGTGACTTGGATTG	1.0
	TGATAAATACATAGGATGGATGG	1.0
Amelogenin	CCTTTGAAGTGGTACCAGAGCAT	0.8
	GCATGCCTAATATTTTCAGGGAATAA	0.8

Then, a total of 25 μL of PCR product was purified with 10 μL of ExoSAP-IT® (USB, Cleveland, OH, USA) by incubating the sample at 37 °C for 45 min, followed by heat inactivation at 80 °C for 15 min. Then, PCR products were purified further using the Qiaquick PCR Purification Kit (Qiagen, Hilden, Germany). The quality of purified amplicons was evaluated by examining the size distribution using a Bioanalyzer 2100 with the Agilent High Sensitivity DNA Kit (Agilent Technologies, Palo Alto, CA, USA). In addition, purity was measured using the NanoDrop 1000 spectrophotometer (Thermo. Fisher Scientific, Waltham, MA, USA), and quantification was carried out using the Quant-iT™ PicoGreen® dsDNA Assay Kit (Invitrogen, Carlsbad, CA, USA) following the manufacturer's instructions.

2.4. Library preparation and NGS data generation

Subsequent barcoded libraries were prepared for sequencing on the MiSeq system with obtained amplicons. Library preparation from the individual DNA samples was performed using the TruSeq® Nano DNA Sample Preparation Kit (Illumina Inc., San Diego, CA, USA) with more than 200 ng of purified PCR products. All steps followed the manufacturer's recommendations, but the beads ratio for size selection was adjusted from 1:1.0 to 1:1.3 (DNA: beads) in bead purification steps based on a previous report [18]. Bioanalyzer was used to determine the size distribution as well as the absence of short fragments. The DNA concentration of the libraries was estimated using the KAPA Library Quantification Kit (KAPA Biosystems, Woburn, MA, USA) according to the manufacturer's instructions. The indexed DNA libraries were normalized