Forensic Science International: Genetics 30 (2017) 134–140

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Sequence-based diversity of 23 autosomal STR loci in Koreans investigated using an in-house massively parallel sequencing panel

Eun Hye Kim[a], Hwan Young Lee[a], So Yeun Kwon[a,b], Eun Young Lee[a], Woo Ick Yang[a], Kyoung-Jin Shin[a,b,*]

[a] Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea
[b] Brain Korea 21 PLUS Project for Medical Science, Yonsei University, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea

ARTICLE INFO

Article history:
Received 26 January 2017
Received in revised form 30 June 2017
Accepted 6 July 2017
Available online 9 July 2017

Keywords:
Massively parallel sequencing
Short tandem repeat
MiSeq
Sequence variation
Discrimination power
Koreans

ABSTRACT

As DNA databases continue to grow and international cooperation increases, forensic STR loci have expanded to increase the discriminatory power and inter-database compatibility. Current capillary electrophoresis (CE) and/or massively parallel sequencing (MPS)-based commercial STR analysis systems reflect such changing trends of expanding STR loci. Due to the general gains of larger multiplexing and the detection of sequence variation, the application of MPS technology to STR analysis has further improved discrimination and is expected to aid in mixture interpretation by increasing the effective number of alleles. However, high-throughput analysis has rarely been reported for forensic DNA databasing. In this study, we present the sequencing results from 250 Korean samples at 23 commonly used STR loci (D1S1656, TPOX, D2S441, D2S1338, D3S1358, FGA, CSF1PO, D5S818, D6S1043, D7S820, D8S1179, D10S1248, TH01, vWA, D12S391, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, and D22S1045) using an in-house assay designed for MPS. All amplicons in the multiplex exhibited a size range of 77 to 217 base pairs, and the barcoded library for the MPS run was easily prepared using a PCR-based library preparation method followed by sequencing on a MiSeq System (Illumina). We compared the STR genotyping results with those obtained using CE and scrutinized the sequence variations in both the targeted STR and flanking regions. MPS results of 23 autosomal STRs were 99.97% concordant with those of CE results. D12S391 and D21S11 exhibited, respectively, the highest number of alleles and genotypes by the MPS analysis. Single nucleotide polymorphisms and insertion and deletions (Indels) were observed in the flanking regions of D1S1656, D2S441, D5S818, D7S820, D13S317, D16S539, D21S11, and Penta D. Consequently, an MPS analysis of an expanded set of STRs, as demonstrated in the population statistics of a Korean population, will be of great practical use in forensic genetics.
© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Compared to the conventional short tandem repeat (STR) analysis with capillary electrophoresis (CE), massively parallel sequencing (MPS) of STRs has the advantage of producing a massive quantity of sequencing data with only a single reaction [1,2]. In addition, the reduced cost of sequencing through MPS has encouraged many forensic investigators to apply this technology to both casework and databasing. Moreover, several publications have emphasized the primary gains of using MPS related to an STR analysis (i.e., larger multiplexing and detection of sequence variation) [3–10]. Most of the published studies validate the new MPS system for forensic markers, including STRs [6–9], and have been performed on low-throughput samples and/or interrogated loci [10]. Therefore, a high-throughput approach is required to demonstrate the superior application of MPS to forensic DNA databasing in association with bioinformatics methods, which will enable the new database to maintain back compatibility with the CE-based method and existing forensic DNA databases.

Several commercial MPS assays, such as the PowerSeq Auto system (Promega, Madison, WI, USA) [6], ForenSeq™ DNA Signature Prep Kit (Illumina, San Diego, CA, USA) [7], and Precision ID GlobalFiler NGS STR Kit (Thermo Fisher, Waltham, MA, USA), are now available for STR analysis. These assays include 22, 27, and 29

* Corresponding author at: Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, Republic of Korea.
E-mail address: kjshin@yuhs.ac (K.-J. Shin).

http://dx.doi.org/10.1016/j.fsigen.2017.07.001
1872-4973/© 2017 Elsevier B.V. All rights reserved.

forensic autosomal STRs, respectively, and are all inclusive of the expanded US CODIS core loci and the 12 core European Standard Set loci. Such expanded STR loci are expected to increase the discriminatory power and inter-database compatibility, and are ready for study at the population level using an MPS method. To date, population genetic analysis has been performed on the MPS data obtained from 68 African Americans, 70 Caucasians, and 45 Hispanics using a prototype version of the PowerSeq Auto system [10]; however, there is a lack of MPS reports on the expanded STR loci of Asians.

In this study, we present the sequencing results from 250 Korean samples at 23 autosomal STR loci (D1S1656, TPOX, D2S441, D2S1338, D3S1358, FGA, CSF1PO, D5S818, D6S1043, D7S820, D8S1179, D10S1248, TH01, vWA, D12S391, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, and D22S1045) using an in-house multiplex assay designed for MPS. In addition to 22 commonly used STRs, our assay includes D6S1043, which is frequently studied in Asians, especially Koreans [11] and Chinese [12]. With the use of two bioinformatics pipelines and SNP annotation software, we identified sequence variations in both the repeat and flanking regions in association with allele frequencies and performed a population genetic analysis to provide forensic parameters for Koreans.

2. Materials and methods

2.1. DNA samples

A total of 250 DNA samples from our previous report [13] were used for this study. Among the 300 DNA samples extracted from buccal swabs using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany) around 2001, samples with small quantities and/or low concentrations (<1 ng/μL) through DNA quantification using a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) were excluded and the 250 selected Korean DNAs were normalized to 1.0 ng/μL for the following assay. This study was approved by the Institutional Review Board of Severance Hospital, Yonsei University in Seoul, Korea.

2.2. Multiplex PCR system for an MPS analysis of 23 autosomal STR loci

DNA samples were amplified with an in-house multiplex PCR system, including the same loci amplified in PowerPlex Fusion System, (Promega) and an additional STR locus, D6S1043: 23 autosomal STRs (D1S1656, TPOX, D2S441, D2S1338, D3S1358, FGA, CSF1PO, D5S818, D6S1043, D7S820, D8S1179, D10S1248, TH01, vWA, D12S391, D13S317, Penta E, D16S539, D18S51, D19S433, D21S11, Penta D, and D22S1045), DYS391, and amelogenin.

To construct the multiplex PCR system, the primer sequences of 17 autosomal STRs and amelogenin from our previous work [9] were utilized, some of which were redesigned to remove minor PCR interference and to increase the PCR yields. To add additional markers, the DNA sequence information of six autosomal STRs and a Y-STR (D1S1656, D2S441, D6S1043, D10S1248, D12S391, D22S1045, and DYS391) was collected from GenBank (http://www.ncbi.nlm.nih.gov/genbank) and the allele information was collected from STRBase (http://www.cstl.nist.gov/biotech/strbase). The PCR primers were designed using the Primer3 (http://frodo.wi.mit.edu/primer3/input.htm) program to ensure that the amplicon sizes could be as small as possible within a range of 75–250 bp, and none of the primer binding sites included mutations with a 1% or higher frequency on the dbSNP build 142 (http://www.ncbi.nlm.nih.gov/projects/SNP/) and the flanking region SNP information of forensic STRs reported by Gettings et al. [14]. For the following PCR-based MPS library preparation, the read sequences were attached at both 5′ ends of each designed primer as described by Lee et al. [12]. The primer information and amplicon sizes for the MPS analysis of the 23 autosomal STRs, a Y-STR, and amelogenin are listed in Supplementary Table S1.

2.3. PCR-based MPS library preparation

An MPS library was generated with the same PCR strategy as that used in previous work [15,16]. Briefly, the first round of PCR amplified targeted markers using a target-specific primer with read sequences. The second round of PCR added indices and platform-specific adapter sequences. The first round of PCR amplification was carried out using a 20 μL reaction volume containing 1 ng of template DNA, 5.0 U of AmpliTaq Gold® DNA Polymerase (Thermo Fisher Scientific), 2.0 μL of Gold ST*R 10 × Buffer (Promega), and an appropriate concentration (0.20–2.0 μM) of each primer (Supplementary Table S1). PCR was conducted on a Veriti® 96-Well Thermal Cycler (Applied Biosystems, Foster City, CA, USA) under the following conditions: 95 °C for 11 min; 29 cycles of 94 °C for 20 s, 59 °C for 1 min, and 72 °C for 30 s, followed by a final extension at 72 °C for 7 min with a 4 °C soak. The second round of PCR amplification was performed using a 20 μL reaction volume containing 1.0 μL of 100-fold diluted amplicons for the first round of PCR and with 2.0 U of AmpliTaq Gold® DNA Polymerase (Thermo Fisher Scientific), 2.0 μL of Gold ST*R 10 × Buffer (Promega) and 2.0 μL of paired-end index 1 (i7) and index 2 (i5) from Nextera XT v2 Index Kit (Illumina). The PCR was performed on a Veriti® 96-Well Thermal Cycler under the following conditions: 95 °C for 15 min, 15 cycles of 94 °C for 20 s, 59 °C for 30 s, and 72 °C for 30 s, and a final extension at 72 °C for 7 min with a 4 °C soak.

Next, the remaining primers and nucleotides were removed via PCR product purification with 1.2× Agencourt® AMPure® XP beads (Beckman Coulter, Indianapolis, IN, USA) according to the manufacturer's guidelines. The quality of the libraries was evaluated by examining the size distribution of the amplicons using an Agilent DNA 1000 Kit (Agilent Technologies, Santa Clara, CA, USA) on an Agilent 2100 Bioanalyzer (Agilent Technologies) (Supplementary Fig. S1). The quantity was measured using KAPA Library Quantification Kits (KAPA Biosystems, Wilmington, MA, USA) for Illumina® platform on an AB 7500 Real-Time PCR System (Applied Biosystems) according to the manufacturer's instruction.

2.4. MPS data generation and analysis

The final libraries were normalized and pooled to 10 nM with equal volumes. MPS was performed using a MiSeq Reagent Kit v2 (2 × 250 cycles) or a MiSeq Reagent Kit v3 (2 × 300 cycles) on the MiSeq System (Illumina). The pooled libraries were sequenced in two separate runs, and FASTQ files were generated for each sample. The raw FASTQ files were sorted as their index information. To obtain reliable results, two analytical methods were adopted in this study. First, the STRait Razor program was used to evaluate the locus-specific read count and size-based STR allele call from the FASTQ file of each sample using the modified config setting from STRait Razor v1.5 [17]. Second, sequence alignment was performed following the protocol presented by Kim et al. [9] with modified reference sequence data. Here, the information pertaining to the allele sequences was obtained from STRbase and Gettings et al. [14]. The flanking region sequences of 500–550 bp in length were acquired from the human reference genome GRCh37/hg19. Each FASTQ file was quality-trimmed with the Q30 value using the FASTX toolkit (http://hannonlab.cshl.edu/fastx_toolkit/), aligned with the custom-made reference sequences using the BWA program [18], and generated into a SAM file.

Download English Version:

https://daneshyari.com/en/article/6462720

Download Persian Version:

https://daneshyari.com/article/6462720

Daneshyari.com