Forensic Science International: Genetics 22 (2016) 37–43

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Massively parallel sequencing of the entire control region and targeted coding region SNPs of degraded mtDNA using a simplified library preparation method

Eun Young Lee[a], Hwan Young Lee[a], Se Yoon Oh[a], Sang-Eun Jung[a], In Seok Yang[a], Yang-Han Lee[b], Woo Ick Yang[a], Kyoung-Jin Shin[a,*]

[a] Department of Forensic Medicine, Yonsei University College of Medicine, 50-1 Yonsei-ro, Seodaemun-gu, Seoul 03722, South Korea
[b] Forensic DNA division, National Forensic Service, 10 Ipchun-ro, Wonju-si, Gangwon-do 26460, South Korea

ARTICLE INFO

Article history:
Received 11 June 2015
Received in revised form 12 January 2016
Accepted 20 January 2016
Available online 22 January 2016

Keywords:
Mitochondrial DNA
Next-generation sequencing
Library preparation
Haplogroup

ABSTRACT

The application of next-generation sequencing (NGS) to forensic genetics is being explored by an increasing number of laboratories because of the potential of high-throughput sequencing for recovering genetic information from multiple markers and multiple individuals in a single run. A cumbersome and technically challenging library construction process is required for NGS. In this study, we propose a simplified library preparation method for mitochondrial DNA (mtDNA) analysis that involves two rounds of PCR amplification. In the first-round of multiplex PCR, six fragments covering the entire mtDNA control region and 22 fragments covering interspersed single nucleotide polymorphisms (SNPs) in the coding region that can be used to determine global haplogroups and East Asian haplogroups were amplified using template-specific primers with read sequences. In the following step, indices and platform-specific sequences for the MiSeq® system (Illumina) were added by PCR. The barcoded library produced using this simplified workflow was successfully sequenced on the MiSeq system using the MiSeq Reagent Nano Kit v2. A total of 0.4 GB of sequences, 80.6% with base quality of >Q30, were obtained from 12 degraded DNA samples and mapped to the revised Cambridge Reference Sequence (rCRS). A relatively even read count was obtained for all amplicons, with an average coverage of 5200 × and a less than three-fold read count difference between amplicons per sample. Control region sequences were successfully determined, and all samples were assigned to the relevant haplogroups. In addition, enhanced discrimination was observed by adding coding region SNPs to the control region in *in silico* analysis. Because the developed multiplex PCR system amplifies small-sized amplicons (<250 bp), NGS analysis using the library preparation method described here allows mtDNA analysis using highly degraded DNA samples.

1. Introduction

Mitochondrial DNA (mtDNA) is a useful forensic marker because of its unique features such as the accumulation of mutations over time, which allows individuals to be differentiated; the absence of recombination; its maternal inheritance; and its presence in multiple copies per cell, which enables the genetic analysis of even highly degraded DNA samples [1–3]. Human mtDNA is 16,569 bp in length. The non-coding mtDNA control region contains the origin of replication and transcription regulatory elements, while the coding region comprises several genes that encode proteins critical to mitochondrial metabolism [4]. The control region shows elevated sequence variation due to its high mutation rate, particularly in the so-called hypervariable regions (HV1, HV2, and HV3). Accurate determination of the sequences of these polymorphic regions is a significant goal in discrimination analysis because it allows individual haplotypes or haplogroups to be defined [5,6]. Moreover, single nucleotide polymorphisms (SNPs) in the coding region can supplement discriminatory evidence in cases where there is a lack of control region sequence information [7,8]. Although Sanger sequencing can be used to obtain this information, it is a labor-intensive, expensive, and time-consuming method [9].

Next-generation sequencing (NGS) has increased in popularity in recent years due to rapid technological advances. NGS can produce a large number of sequences and can detect minor

* Corresponding author. Fax: +82 2 362 0860.
E-mail address: kjshin@yuhs.ac (K.-J. Shin).

http://dx.doi.org/10.1016/j.fsigen.2016.01.014
1872-4973/

variations due to the high coverage given a sequencing library [10,11]. In general, the library for NGS is prepared in several steps, including DNA fragment of interest generation, adaptor ligation, fragment size-selection and purification, library quantification, and pooling. Adaptor ligation is essential to obtain adequate sequencing results. The library has to contain several functional sequences such as read sequences for sequencing, short oligomers for barcoding in multiplex reactions, and sequences suitable for hybridization to an NGS platform [11,12]. Therefore, the library-building process is time consuming and technically demanding, and there is significant sample loss.

Here we propose a simplified library preparation method based on two PCR amplifications that is applicable to the Illumina platform. We targeted the entire mtDNA control region and several coding regions containing SNPs that can be used to designate global and East Asian haplogroups [7,13]. This PCR system can be used to sequence mtDNA from highly degraded DNA samples. To test the applicability of the devised system for mtDNA analysis, we obtained and analyzed NGS results from highly degraded DNA samples.

2. Materials and methods

2.1. DNA samples

Control DNAs, 2800M and 9947A (Promega Corp., Madison, MI, USA), were artificially degraded to 150–250-bp-sized fragments using a Covaris S2 Focused-ultrasonicator (Covaris, Inc., Woburn, MA, USA). The degraded fragments were electrophoresed using a high-sensitivity DNA chip (Agilent Technologies, Inc., Santa Clara, CA, USA) on an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.) (Supplementary material Fig. S1). Ten naturally degraded DNAs were obtained from 50-year-old skeletal remains, as described in a previous study [14]. The mtDNA control region of these samples had been amplified to obtain five PCR products sized below 300 bp and were analyzed by Sanger sequencing in both the forward and reverse directions without ambiguous interpretation. The primer information and PCR conditions for Sanger sequencing are presented on the website (http://forensic.yonsei.ac.kr/protocol/mtDNA-midi-mini.pdf).

2.2. Design of multiplex PCR systems to amplify mtDNA fragments

We designed two major multiplex PCR systems: one named “Midiplex PCR” to amplify the hyper-variable regions and the other named “mtSNP PCR” to amplify the 22 coding region fragments. The Midiplex PCR system consisted of two separate PCR sets, Midiplex I and II, each of which generated three amplicons (Fig. 1). The mtSNP PCR system also consisted of two multiplex PCR sets; the 22 targeted amplicons contain 32 SNPs that can be used to designate major global and East Asian haplogroups (Fig. 2). PCR primers were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/input.htm). Amplicons were designed to be shorter than 250 bp in Midiplex PCR and shorter than 135 bp in mtSNP PCR, facilitating amplification even from degraded samples.

2.3. Two-step PCR amplification and library preparation

We conducted two PCR amplifications to generate a library using primers with a modification referring to the sequence information from the Nextera® sample preparation kit (Illumina, Inc., San Diego, CA, USA) (Fig. 3). The first PCR amplification targeted the mtDNA itself. Primer sequences included mtDNA-specific sequences and read sequences as follows:

F_Rd1; 5′-TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG-(mtDNA-specific sequences)-3′

R_Rd2; 5′-GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG-(mtDNA-specific sequences)-3′

Multiplex PCR reactions were performed in a final volume of 20 μL that contained 100 pg of degraded control DNA or 2.0 μL of DNA extracted from skeletal remains, 2.0 μL of Gold ST*R 10 × buffer (Promega), 1.5 U of AmpliTaq Gold® DNA polymerase (Applied Biosystems, Foster City, CA, USA) for Midiplex PCR or 3.0 U AmpliTaq Gold® DNA polymerase for mtSNP PCR, and primer mix (Supplementary material Table S1). The reaction mixture was subjected to an initial denaturation step at 95 °C for 11 min, followed by 25 cycles of amplification consisting of denaturation at 94 °C for 20 s, annealing at 56 °C for 60 s, and extension at 72 °C for 30 s. This was followed by a final extension step at 72 °C for 7 min.

A second PCR was performed to ligate indices and platform-specific sequences. Primer sequences were as follows:

F_i5; 5′-AATGATACGGCGACCACCGAGATCTACAC[i5]TCGTCGGCAGCGTC-3′

R_i7; 5′-CAAGCAGAAGACGGCATACGAGAT[i7]GTCTCGTGGGCTCGG-3′

In the second PCR, 1.0-μL aliquots of 100-fold diluted PCR products from PCRs I and II were used as templates in 20-μL reactions containing 2.0 μL of Gold ST*R 10 × buffer, 2.0 μL each of i5 and i7 adapters from the Nextera® Index Kit (Illumina, Inc.), and 1.0 U of AmpliTaq Gold® DNA polymerase. Midiplex and mtSNP PCRs were performed in separate tubes. Each reaction mixture was subjected to an initial denaturation step at 95 °C for 15 min, followed by 15 cycles of amplification consisting of denaturation at 94 °C for 20 s, annealing at 59 °C for 30 s, extension at 72 °C for 30 s, and a final extension step at 72 °C for 7 min.

Following PCR cleanup with 1.0 × Agencourt® AMPure® XP beads (Beckman Coulter, Inc., Indianapolis, IN, USA), libraries were quantified using KAPA library quantification kits (KAPA Biosystems, Inc., Wilmington, MA, USA), and fragment sizes were assessed using an Agilent 2100 Bioanalyzer. Considering the copy numbers of fragments generated from each multiplex PCR system, the libraries were mixed at a ratio of 1:3 of Midiplex PCR product:

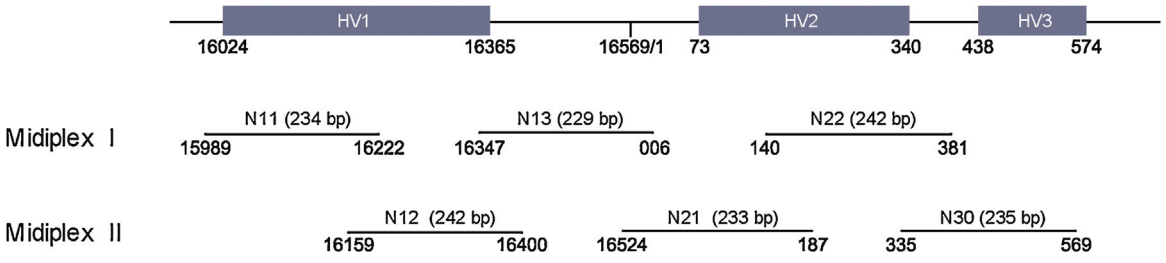

Fig. 1. Scheme depicting the Midiplex PCR sets I and II designed to amplify the control region. Six fragments (N11, N13, N22, N12, N21, and N30) were generated from two multiplex PCRs; the size of each fragment is indicated in parentheses. Numbers indicate the starting nucleotide position (np) and ending np of each fragment.