# Massively Parallel Sequencing of Mitochondrial DNA Control Region and SNPs for Global Haplogroup Determination

**Eun Young Lee · Hwan Young Lee · Sang Eun Jung · In Seok Yang · Kyoung-Jin Shin**

*Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea*

## Introduction

Because of the potential of high-throughput sequencing for recovering genetic information from multiple markers and multiple individuals in a single run, forensic genetic application of next-generation sequencing (NGS) technology is being explored by an increasing number of laboratories. For NGS, a cumbersome and technically challenging library construction process is required. Here, we propose a simplified library preparation method applicable to the Illumina platform for mitochondrial DNA (mtDNA) sequence analysis. A sequencing library for NGS is completed through two rounds of PCR amplification. Moreover, we analyzed sequence variations in control region and coding region SNPs from NGS results of two artificially degraded DNAs and ten naturally degraded DNAs.

## Materials and Methods

### DNA samples

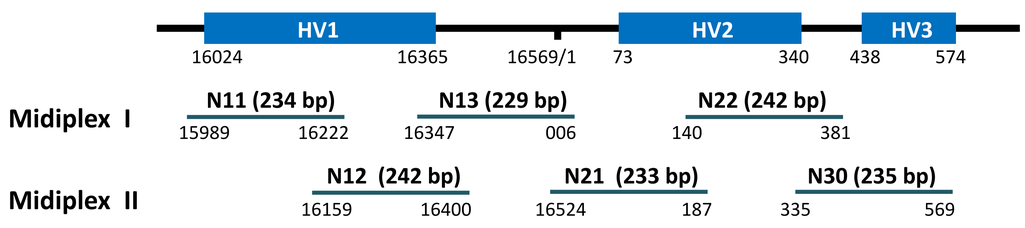
Control DNAs, 2800M and 9947A (Promega Corp., Madison, MI, USA), were artificially degraded using a Covaris S2 Focused-ultrasonicator (Covaris, Inc., Woburn, MA, USA) and the fragment size ranged 150 - 250 bp. Ten naturally degraded DNAs were obtained from 50-year-old skeletal remains.

### Design of multiplex PCR systems to amplify mtDNA fragments

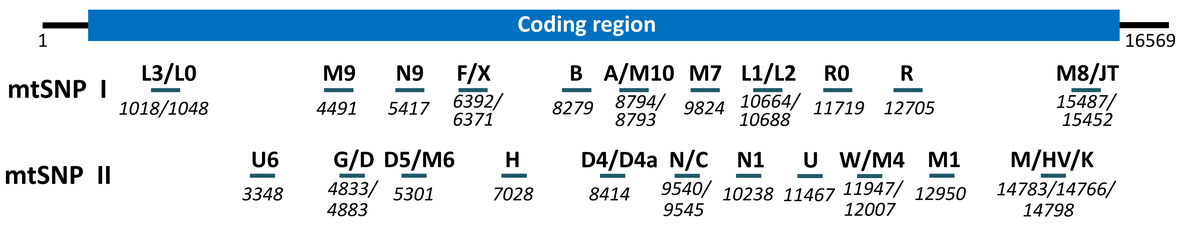
We have designed two groups of multiplex PCR system: one named “Midiplex PCR” to amplify control region and the other named “mtSNP PCR” to amplify 22 coding region fragments. Midiplex PCR consists of two separate PCR sets, Midiplex I and II, and each of these sets generates three amplicons. The mtSNP PCR is also divided into two multiplex PCR sets; the 22 target amplicons contain 32 SNPs that can be used to designate major global haplogroups and East Asian-specific haplogroups. Finally, primers were designed to have less than 250 bp of amplicon size.

**Fig. 1. Scheme of multiplex PCR systems to amplify the control region (a) and 32 SNPs in the coding region (b)**

(a) Six fragments (N11, N13, N22, N12, N21 and N30) are generated from two multiplex PCRs




(b) Each multiplex PCR amplify 11 fragments to target SNPs at the indicated nucleotide positions to make haplogroup determinations




### Library preparation by two rounds of PCR and MiSeq sequencing

The procedure of library preparation is summarized in Fig.2. First PCR reaction used 100 pg of degraded control DNA or 2.0 µl of skeletal DNA as template and amplified by 25 cycles. The 1:100 diluted first amplicons were used as template in the second PCR reaction and amplification repeated by 15 cycles. Following PCR cleanup with 1.0X SPRI beads, libraries were quantified using KAPA library quantification kits (KAPA Biosystems, Inc., Wilmington, MA, USA). The pooled library were analyzed on the MiSeq® System using a MiSeq Reagent Nano Kit v2 (2x250 cycles) (Illumina, Inc., San Diego, CA, USA).

**Fig. 2. Overview of library preparation by two rounds of PCR amplification**




The first PCR primer has mtDNA-specific sequences (blue) and read sequences (orange). In the second PCR, the primers bind to read sequences of the first amplicon and add indices (green) for multiplexing and platform-specific sequences (black). After two rounds of PCR amplifications, the fragments can sequenced using MiSeq® System.

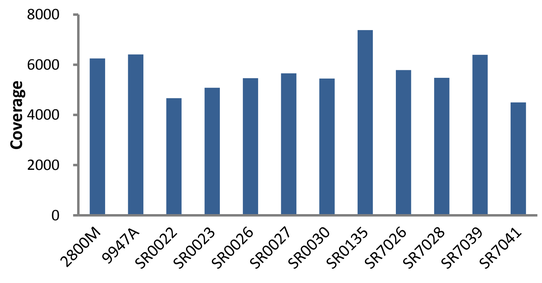
### NGS data analysis

MiSeq (FASTQ) → Bowtie 2 (Align with rCRS) → SAMtools (Convert to BAM) → Variant calling → Haplogroup designation

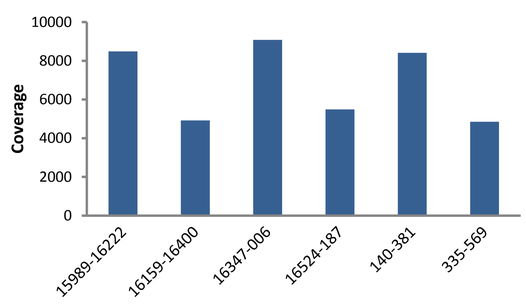
## Results

**Fig. 3. Read count distribution of 12 samples or 2800M control DNA**

(a) Average coverage observed in the 28 amplicons of 12 samples




(b) Coverage of the six amplicons targeting the control region of 2800M control DNA




(c) Coverage of the 22 amplicons targeting the coding region of 2800M control DNA

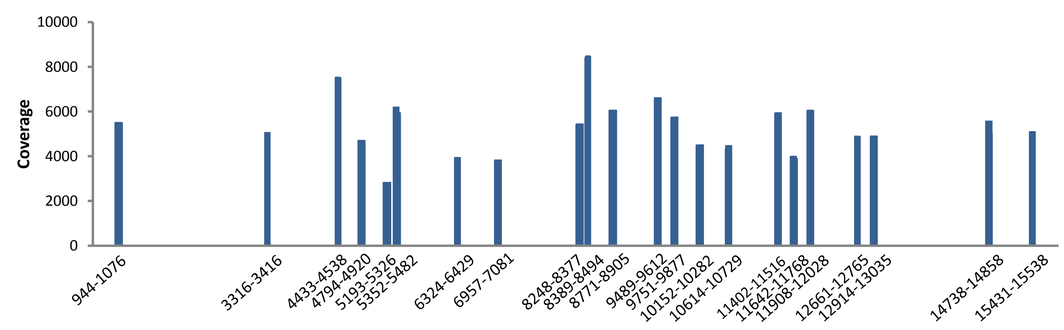



**Table 1. Observed mutation motifs and assigned haplogroups of the 12 samples**

| Sample ID | Control Region | Coding Region | Haplogroup |
|---|---|---|---|
| **2800M** | 16519C<br>152C 263G 315.1C 477C | 8860G | H1c |
| **9947A** | 16311C 16519C<br>93G 195C 214G 263G 309.1C<br>309.2C 315.1C | 8448C 8860G | H11 |
| **SR0022** | 16223T 16362C 16519C<br>73G 150T 194T 205A 263G 315.1C<br>489C 523d 524d | 4883T 7028T 8414T 8860G 9540C<br>9824A 11719A 12705T 14766T<br>14783C | D4b2b |
| **SR0023** | 16223T 16231C 16362C<br>73G 263G 315.1C 489C | 4883T 7028T 8414T 8860G 9540C<br>11696A 11719A 12705T 14766T<br>14783C | D4j |
| **SR0026** | 16183C 16189C 16335G 16527T<br>73G 152C 249d 263G 309.1C<br>309.2C 315.1C | 1005C 6392C 7028T 8860G<br>11719A 14766T | F2 |
| **SR0027** | 16223T 16290T 16319A 16362C<br>73G 152C 207A 235G 260A 309.1C<br>309.2C 315.1C 523d 524d | 4824G 7028T 8459G 8794T<br>8860G 11719A 12705T 14766T | A15 |
| **SR0030** | 16126C 16231C 16311C<br>73G 263G 309.1C 315.1C 338T<br>482C 523d 524d | 5417A 7028T 8860G 11719A<br>12705T 14766T 16126C | Y2b |
| **SR0135** | 16129A 16189C 16223T 16297C 16298C<br>73G 150T 199C 263G 315.1C 489C | 5460A 7028T 8860G 9540C 9824C<br>11719A 12705T 14766T 14783C | M7b1a1a1 |
| **SR7026** | 16129A 16140C 16187T<br>16189C 16266G 16519C<br>73G 93G 210G 263G 309.1C<br>315.1C 523d 524d | 3396C 7028T 8281-8289d 8860G<br>11719A 14766T | B5a2a1 |
| **SR7028** | 16223T 16245T 16362C<br>73G 152C 191.1A 194T 199C 207A<br>263G 309.1C 315.1C 489C | 3391A 4883T 7028T 8414T 8860G<br>9540C 11719A 12705T 14766T<br>14783C | D4c1a |
| **SR7039** | 16223T 16297C 16301T<br>73G 150T 199C 204C 263G<br>309.1C 315.1C 489C | 5460A 7028T 8860G 9540C 9824C<br>11719A 12705T 14766T 14783C | M7b1a1b |
| **SR7041** | 16187T 16223T 16290T 16319A 16519C<br>73G 235G 263G 309.1C 315.1C<br>523d 524d | 4824G 7028T 8794T 8860G<br>10670T 11719A 12705T 14766T | A5a |

## Conclusions

- We devised simplified library preparation method that is completed through two rounds of PCR amplification and applied to mtDNA analysis of entire control region and coding region fragments using Illumina platform.
- The 336 mitochondrial amplicons from 12 samples were sequenced showing average coverage of 5,200× and relatively even read depth was obtained between 28 amplicons within a sample.
- The sequence variations of the entire control region and 32 SNPs in the coding region were determined and relevant haplogroups were assigned for all the samples.
- Because PCR system was designed to generate small fragment size (less than 250 bp), sequence analysis of degraded DNA was performed successfully.

## Acknowledgement

- This research was supported by a fund (No. NFS2014DNA06) from the Forensic Research Program of the National Forensic Service, Korea.