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

Department of Forensic Medicine, Yonsei University College of Medicine
Forensic DNA division, National Forensic Service

Mitochondrial DNA (mtDNA)

“16,569” bp

100s of copies per cell
Introduction

- Importance of mitochondrial hyper-variable (HV) regions and
coding region SNPs analysis in degraded samples identification
- Limitation of existing methods for mtDNA analysis using
degraded samples
- Next-generation sequencing (NGS) is expected to be useful
technique to analyze mtDNA effectively
- More simple library preparation method is needed to easily
access NGS analysis

Object

- Development of library preparation system for mtDNA sequencing
  - Construct a multiplex PCR system
    - Amplification of entire control region and targeted coding region SNPs
    - Small-sized amplicons (<250 bp)
    - Simultaneous amplification of coding region SNPs specific to global
      and East Asian haplogroups
  - Devise a simple library preparation method
- Validation of devised NGS system for mtDNA analysis
  - Test using artificially or naturally degraded DNA samples
Multiplex PCR design for mtDNA sequencing

- **Control regions** (6 amplicons from 2 multiplex PCRs)
  - Amplicon size range: 229 ~ 242 bp

- **Coding region SNPs** (22 amplicons from 2 multiplex PCRs)
  - Amplicon size range: 101 ~ 135 bp

Global mtDNA haplogroups
Multiplex PCR design for mtDNA sequencing

- Selected 32 coding region SNPs and haplogroups

Simplify library preparation

TruSeq Kit method – Ligation of adaptor

1. Fragment genomic DNA (PCR product)
2. End repairs
3. Beads purifications
4. Adenylate 3’ ends
5. PCR amplion + Adaptor ligation
6. Beads purifications
7. 2nd library enrichment PCR

Nextera indexing PCR method

1. PCR primer with adaptor sequence
2. 2nd Target specific PCR including Nextera adapter

Library

Beads purifications
Strategy for NGS library preparation

1st PCR; Specific to mtDNA sequence

- Template (mtDNA)
- Primers
  - Sequencing
  - Adaptor
  - Target to mtDNA
- 1st amplicon
  - mtDNA fragment + Sequencing primer + Adaptor

Strategy for NGS library preparation

2nd PCR; Specific to adaptor sequence

- Template (1st amplicon)
- Primers
  - Index
  - Clustering
  - Target to adaptor seq.
- 2nd amplicon
  - Sequencing-ready fragment
Materials and Methods

Workflow of library preparation and NGS run

1st PCR reaction; Midiplex I/II, mtSNP I/II

2nd PCR reaction; Adding indices

Bead purification

Library quantification

Library pooling

NGS (Illumina MiSeq)

1st PCR condition

<table>
<thead>
<tr>
<th>Reaction mixture</th>
<th>Midiplex I/II</th>
<th>mtSNP I/II</th>
</tr>
</thead>
<tbody>
<tr>
<td>10x STR buffer</td>
<td>2.0 ul</td>
<td>2.0 ul</td>
</tr>
<tr>
<td>Primers</td>
<td>Adjusted conc.</td>
<td>Adjusted conc.</td>
</tr>
<tr>
<td>Gold Taq polymerase</td>
<td>0.3 ul (1.5 U)</td>
<td>0.6 ul (3.0 U)</td>
</tr>
<tr>
<td>*Template</td>
<td>100 pg or 2 ul</td>
<td>100 pg or 2 ul</td>
</tr>
<tr>
<td>Fill up to with dH2O</td>
<td>20.0 ul</td>
<td>20.0 ul</td>
</tr>
</tbody>
</table>

*Template: 100 pg of degraded control DNA (2800M, 9947A and 9948) 2 ul of extracted DNA from bone

2nd PCR condition

<table>
<thead>
<tr>
<th>Reaction mixture</th>
<th>10x STR buffer</th>
<th>2.0 ul</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index 1</td>
<td>2.0 ul</td>
<td></td>
</tr>
<tr>
<td>Index 2</td>
<td>2.0 ul</td>
<td></td>
</tr>
<tr>
<td>Gold Taq polymerase</td>
<td>0.2 ul (1.0 U)</td>
<td></td>
</tr>
<tr>
<td><strong>Template</strong></td>
<td>2.0 ul</td>
<td></td>
</tr>
<tr>
<td>Fill up to with dH2O</td>
<td>20.0 ul</td>
<td></td>
</tr>
</tbody>
</table>

**Template: 1.0 ul of 1/100 diluent of 1st PCR I product + 1.0 ul of 1/100 diluent of 1st PCR II product

Thermal cycling

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>95 °C</td>
<td>11 min</td>
</tr>
<tr>
<td>94 °C</td>
<td>20 sec</td>
</tr>
<tr>
<td>56 °C</td>
<td>60 sec</td>
</tr>
<tr>
<td>72 °C</td>
<td>30 sec</td>
</tr>
<tr>
<td>72 °C</td>
<td>7 min</td>
</tr>
<tr>
<td>4 °C</td>
<td>forever</td>
</tr>
</tbody>
</table>

*Template: 1/100 diluent of 1st PCR I product + 1/100 diluent of 1st PCR II product
Materials and Methods

- **DNA samples**
  - Artificially degraded standard DNAs (2800M, 9947A and 9948)
  - 10 DNAs extracted from old skeletal remains

- **Generation of library by 2-step PCR**
  - Midiplex I/II
    (Average size: 372 bp)
  - mtSNP I/II
    (Average size: 257 bp)

Materials and Methods

- **Library quality control**
  - Library quantification using KAPA library quantification kit
  - Library QC using Bioanalyzer

- **Library pooling and NGS run**
  - 28 amplicons/sample
    ; 6 amplicons from Midiplex PCR + 22 amplicons from mtSNP PCR
  - 10 samples/run
  - NGS on Illumina platform
    ; MiSeq Reagent Nano Kit v2, 500 Cycles (2x250 bp)
Results

- The obtained data yield per sample

Read count distribution of 2800M control DNA

- Coverage for 22 amplicons for 32 coding region SNPs

- Coverage for 6 amplicons on mitochondrial control region
Read count distribution of a skeletal sample

- Coverage for 22 amplicons for 32 coding region SNPs
- Coverage for 6 amplicons on mitochondrial control region

Materials and Methods

- NGS data analysis
  - Miseq run: FASTQ
  - Bowtie 2; Align with rCRS
  - SAMtools; Convert to BAM
  - VarScan 2; Variant calling
  - IGV; Confirm calling results
  - Hg designation; Refer to PhyloTree Build 16
<table>
<thead>
<tr>
<th>Sample ID</th>
<th>Control Region</th>
<th>Coding Region</th>
<th>Haplogroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>2800M</td>
<td>16519C</td>
<td>8860G</td>
<td>H1c</td>
</tr>
<tr>
<td>9947A</td>
<td>16311C16519C</td>
<td>8448C8860G</td>
<td>H11</td>
</tr>
<tr>
<td>SR0022</td>
<td>1623T16362C16519C</td>
<td>4883T7028T8414T8860G9540C9824A</td>
<td>D4b2h</td>
</tr>
<tr>
<td>SR0023</td>
<td>1622T16213C16362C</td>
<td>4883T7028T8414T8860G9540C1169A</td>
<td>D4j</td>
</tr>
<tr>
<td>SR0026</td>
<td>16183C16189C16355G16527T</td>
<td>1005C6392C7028T8860G11719A</td>
<td>F2</td>
</tr>
<tr>
<td>SR0027</td>
<td>16223T16290T16319A16382C</td>
<td>4824G7028T8595G8795T8860G11719A</td>
<td>A15</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>SR7041</td>
<td>16187T16223T16290T16319A16519C</td>
<td>4824G7028T8795T8860G10670T11719A</td>
<td>A5a</td>
</tr>
</tbody>
</table>

### Summary

- **Construction of multiplex PCR system** to amplify entire control regions and targeted coding region SNPs of mitochondrial DNA
- **Development of simple library preparation system** for mtDNA NGS analysis
- **Demonstration validity of devised NGS system** for analysis of mitochondrial DNA sequences using degraded samples
Acknowledgement

This research was supported by a fund (NFS2014DNA06) from the Forensic Research Program of the National Forensic Service, Korea.