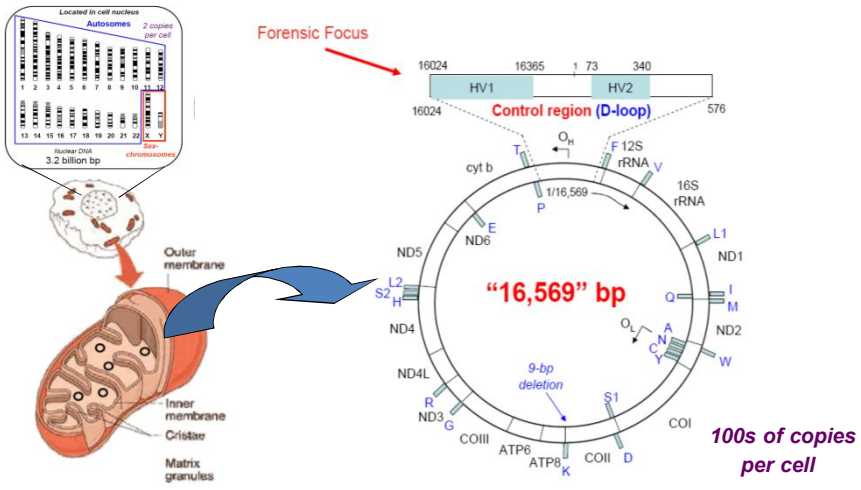
Massively parallel sequencing of the entire control region and targeted coding region SNPs of degraded mtDNA using a simplified library preparation method

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)



Introduction

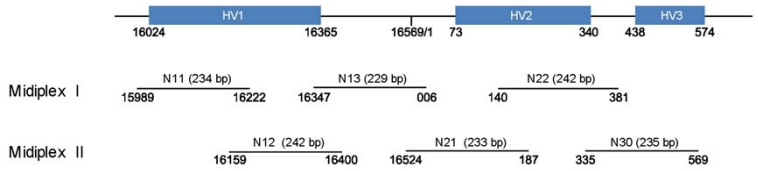
- Importance of mitochondrial hyper-variable (HV) regions and coding region SNPs analysis in degraded samples identification
- Limitation of existing methods for mtDNA analysis using degraded samples
- Next-generation sequencing (NGS) is expected to be useful technique to analyze mtDNA effectively
-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

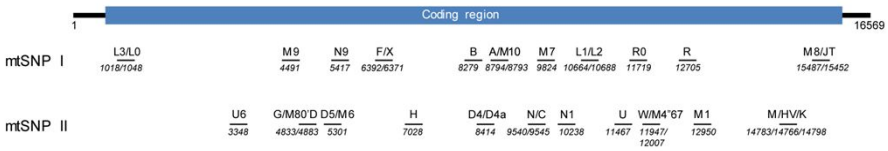
Multiplex PCR design for mtDNA sequencing

➢ Control regions (6 amplicons from 2 multiplex PCRs)



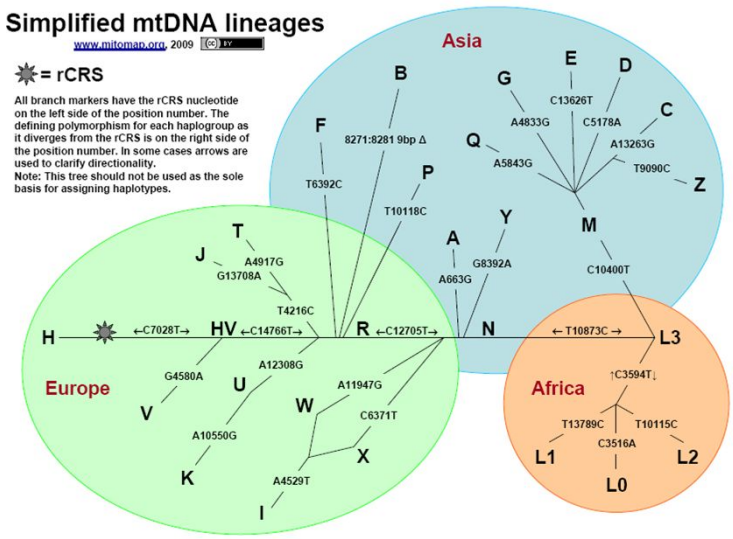
⇒ Amplicon size range; 229 ~ 242 bp

➢ Coding region SNPs (22 amplicons from 2 multiplex PCRs)



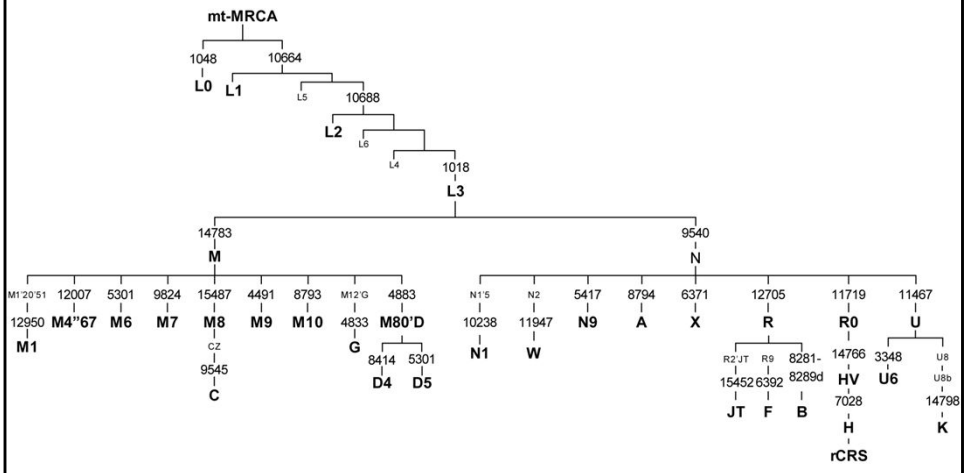
⇒ Amplicon size range; 101 ~ 135 bp

Global mtDNA haplogroups

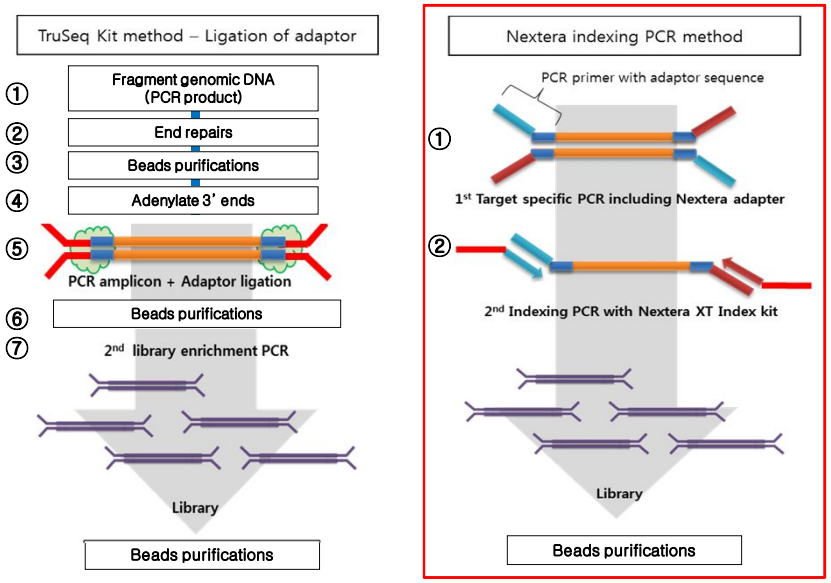


Multiplex PCR design for mtDNA sequencing

➢ Selected 32 coding region SNPs and haplogroups

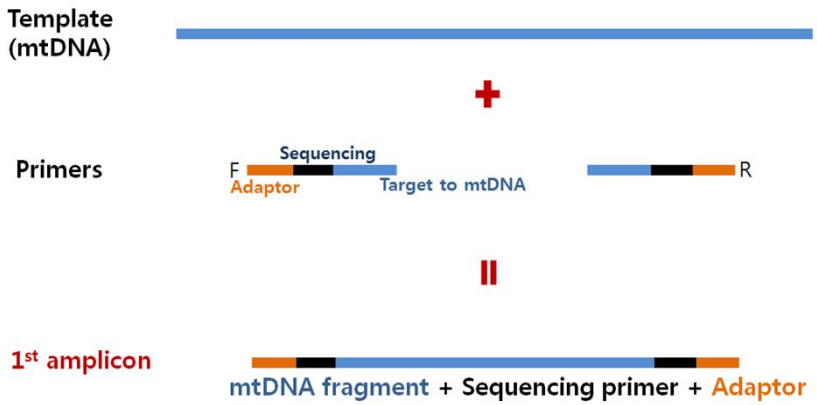


Simplify library preparation



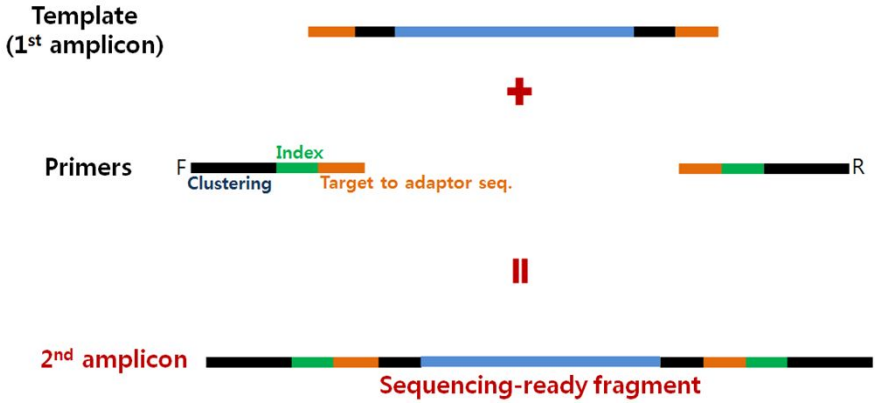
Strategy for NGS library preparation

➢ 1st PCR; Specific to mtDNA sequence



Strategy for NGS library preparation

➢ 2nd PCR; Specific to adaptor sequence



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)

Materials and Methods

➢ 1st PCR condition

Reaction mixture	Midiplex I/II	mtSNP I/II
10x STR buffer	2.0 ul	2.0 ul
Primers	Adjusted conc.	Adjusted conc.
Gold *Taq* polymerase	0.3 ul (1.5 U)	0.6 ul (3.0 U)
*Template	100 pg or 2 ul	100 pg or 2 ul
Fill up to with dH_2O	20.0 ul	20.0 ul

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

Thermal cycling

95 ℃	11 min	
94 ℃	20 sec	
56 ℃	60 sec	X 25 cycles
72 ℃	30 sec	
72 ℃	7 min	
4 ℃	forever	

➢ 2nd PCR condition

Reaction mixture	
10x STR buffer	2.0 ul
Index 1	2.0 ul
Index 2	2.0 ul
Gold *Taq* polymerase	0.2 ul (1.0 U)
**Template	2.0 ul
Fill up to with dH_2O	20.0 ul

Thermal cycling

95 ℃	15 min	
94 ℃	20 sec	
59 ℃	30 sec	X 15 cycles
72 ℃	30 sec	
72 ℃	7 min	
4 ℃	forever	

**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

Materials and Methods

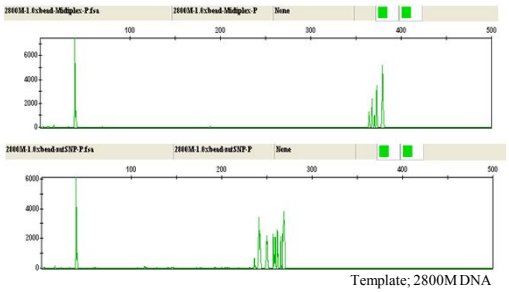
➢ DNA samples

- 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)

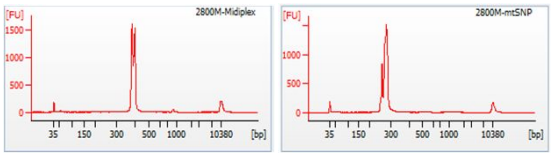


Template; 2800M DNA

Materials and Methods

➢ Library quality control

- Library quantification using using KAPA library quantification kit
- Library QC using Bioanalyzer



➢ Library pooling and NGS run

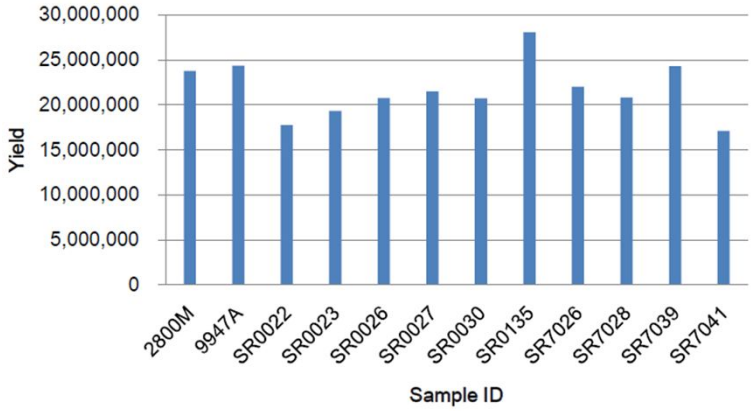
- 28 amplicons/sample

  ; 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)

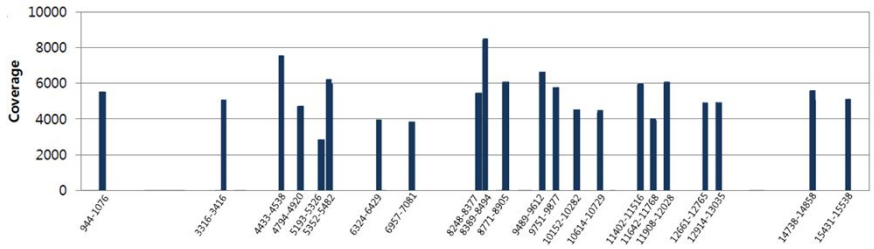
Results

➢ The obtained data yield per sample

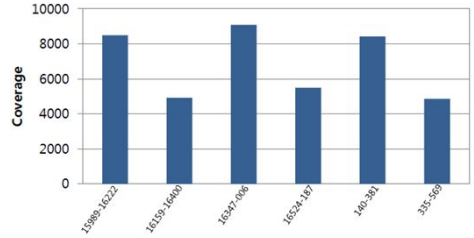


Read count distribution of 2800M control DNA

➢ Coverage for 22 amplicons for 32 coding region SNPs

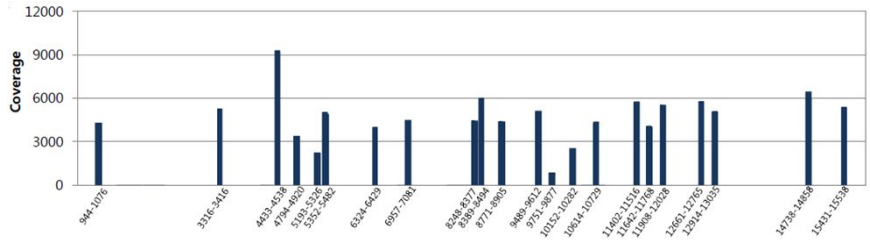


➢ Coverage for 6 amplicons on mitochondrial control region

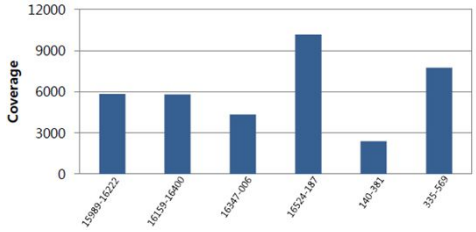


Read count distribution of a skeletal sample

- 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

mtDNA sequence and haplogroup

Sample ID	Control Region	Coding Region	Haplogroup
2800M	16519C 152C 263G 315.1C 477C	8860G	H1c
9947A	16311C 16519C 93G 195C 214G 263G 309.1C 309.2C 315.1C	8448C 8860G	H11
SR0022	16223T 16362C 16519C 73G 150T 194T 205A 263G 315.1C 489C 523d 524d	4883T 7028T 8414T 8860G 9540C 9824A 11719A 12705T 14766T 14783C	D4b2b
SR0023	16223T 16231C 16362C 73G 263G 315.1C 489C	4883T 7028T 8414T 8860G 9540C 11696A 11719A 12705T 14766T 14783C	D4j
SR0026	16183C 16189C 16335G 16527T 73G 152C 249d 263G 309.1C 309.2C 315.1C	1005C 6392C 7028T 8860G 11719A 14766T	F2
SR0027	16223T 16290T 16319A 16362C 73G 152C 207A 235G 260A 309.1C 309.2C 315.1C 523d 524d	4824G 7028T 8459G 8794T 8860G 11719A 12705T 14766T	A15
...	...	...	...
SR7041	16187T 16223T 16290T 16319A 16519C 73G 235G 263G 309.1C 315.1C 523d 524d	4824G 7028T 8794T 8860G 10670T 11719A 12705T 14766T	A5a

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.