

# mtDNAprofiler

A web based Program for Nomenclature  
and Comparison of mtDNA Sequences

In Seok Yang  
Dept. of Forensic Medicine  
Yonsei University College of Medicine

## Human mitochondrial DNA (mtDNA)

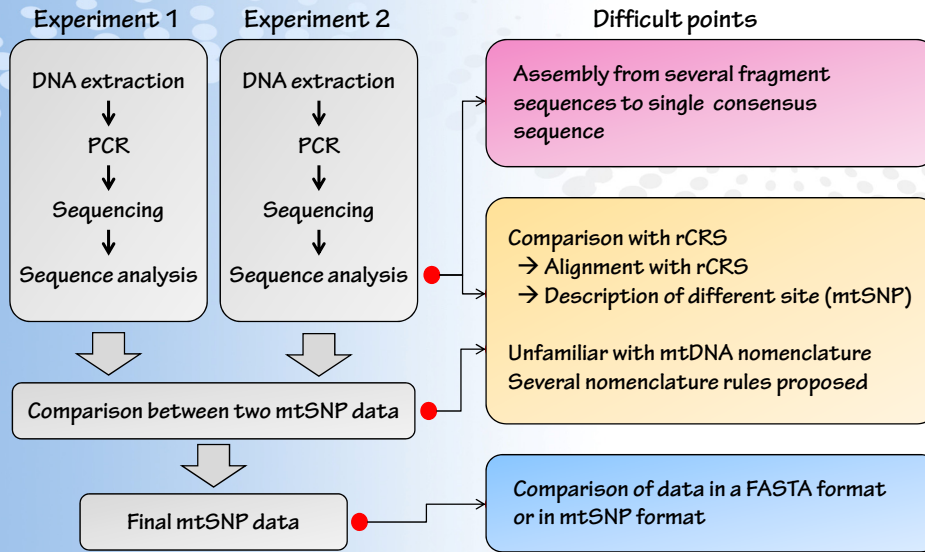
### ■ Characteristics

- 16,569 base pair, circular DNA
- Maternal inheritance, high mutation rate, high copy number, lack of recombination
- Certain mutations in mtDNA → diseases or clinical phenotypes

### ■ Applications

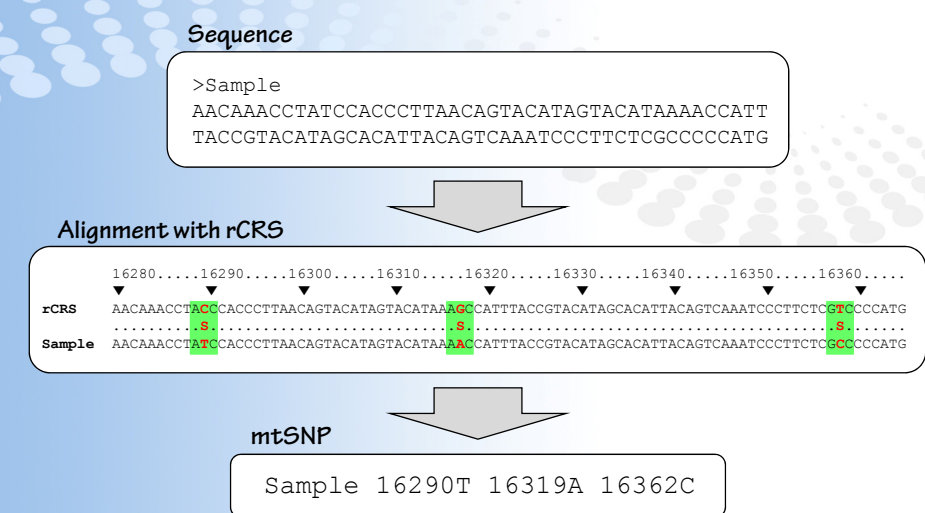
- Population genetics → genetic difference among populations
- Medical genetics → diseases related with mtDNA variations
- forensic genetics → personal identification

# Human mtDNA analysis from samples



Page = 3

## Nomenclature : mtDNA seq → mtSNP



Page = 4

# mtDNAProfiler (<http://mtprofiler.yonsei.ac.kr>)

## Four tools

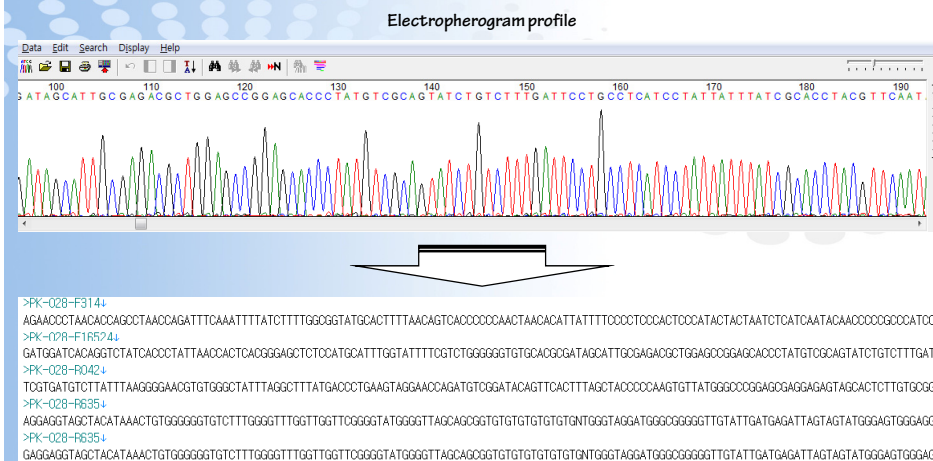
1. mtDNA nomenclature
2. mtDNA assembly
3. mtSNP conversion
4. mtSNP comparison

## Development

1. Web environment
2. Language : PHP
3. Operating system : Linux
4. Web-service program : Apache HTTP server

Page = 5

# Sample sequences → FASTA format



Sample sequences in FASTA format

Page = 6

# mtDNA nomenclature protocol

1. Alignment in mtDNAprofiler basically uses parsimonious Smith-Waterman algorithm and follows alignment rules which is selected by scientific working group on DNA analysis methods (SWGAM).
2. Its protocol is composed of two main rules as following: least number of differences and indels (insertions and deletions) detection and rearrangement.
3. Second rule is further divided into three sub-rules, which selects one rule according to the region.

## 1. Least number of differences

### 2. Indels detection and rearrangement

#### a. AC repeat region

Maintain repeat unit

#### b. Hypervariable 2 C-stretch region

(1) Prefer indels to substitutions

(2) 3'-most placement of indels

#### c. Non-AC repeat & non-hypervariable 2 C-stretch region

(1) Prefer substitutions to indels

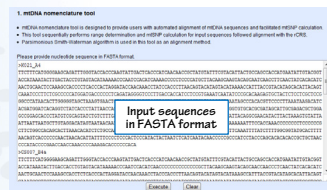
(2) 3'-most placement of indels

(3) Contiguous placement of indels

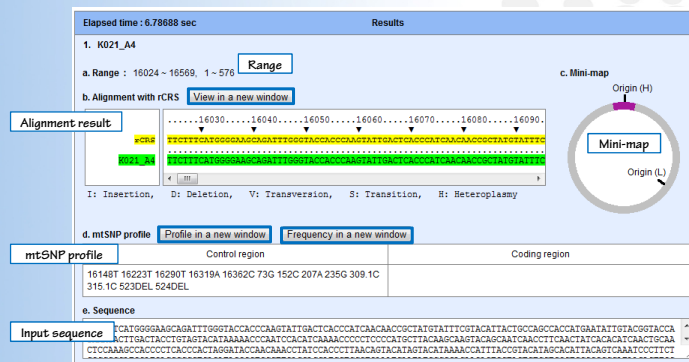
Page = 7

## 1. mtDNA nomenclature tool

mtDNA nomenclature tool is designed to provide users with mtSNP data through automated alignment of mtDNA sequences with rCRS.



Start page



Page = 8

Result page

Alignment results in a new window

mtSNP frequency

Num	mtSNP	Number of each mtSNP / Number of samples
1	73G	5/5
2	150T	1/5
3	152C	2/5
4	195C	1/5
5	199C	1/5
6	207A	1/5
7	235G	1/5
8	249DEL	1/5
9	263G	4/5
10	309.1C	4/5

mtSNP frequency in a new window

mtSNP profile

Region: All region Control region Coding region Download button

Select All Inverse Reset 5 samples Download mtSNP profile

Check	Sample	mtSNP list in all region
<input checked="" type="checkbox"/>	K021_A4	16148T 16223T 16290T 16319A 16362C 73G 152C 207A 235G 309.1C 315.1C 523DEL 524DEL
<input checked="" type="checkbox"/>	CS007_B4A	16182C 16183C 16189C 16194C 16195C 16217C 16261T 16519C 16524G 73G 263G 315.1C 523DEL 524DEL
<input checked="" type="checkbox"/>	CS016_C4A	16093C 16129A 16223T 16298C 16327T 16519C 73G 195C 249DEL 263G 309.1C 309.2C 315.1C 489C
<input checked="" type="checkbox"/>	DR0150_D4A	16129A 16223T 16362C 16519C 73G 152C 263G 309.1C 315.1C 489C
<input checked="" type="checkbox"/>	DR0143_M7B2	16093C 16129A 16183C 16189C 16223T 16297C 16298C 73G 150T 199C 263G 309.1C 309.2C 315.1C 489C

Select All Inverse Reset Download mtSNP profile

mtSNP profile in a new window

mtSNP profile in downloaded file

	A	B	C	D	E	F	G	H	I	J
1 Samples	SNP lists									
2 K021_A4	16148T	16223T	16290T	16319A	16362C	73G	152C	207A	235G	
3 CS007_B4A	16182C	16183C	16189C	16194C	16195C	16217C	16261T	16519C	16524G	
4 CS016_C4A	16093C	16129A	16223T	16298C	16327T	16519C	73G	195C	249DEL	
5 DR0150_D4A	16129A	16223T	16362C	16519C	73G	152C	263G	309.1C	315.1C	
6 DR0143_M7B2	16093C	16129A	16183C	16189C	16223T	16297C	16298C	73G	150T	

## 2. mtDNA assembly tool

The mtDNA assembly tool extracts a consensus sequence from several mtDNA fragment sequences obtained from multiple amplification reactions.

1. mtDNA assembly tool

- mtDNA assembly tool extracts a consensus sequence from several fragment sequences (cloning fragments) obtained from multiple amplification reactions.
- In particular, this tool is useful when range of the sequences is indicated in control region of mtDNA.
- Even if the sequences are included in the control region, it can be processed with the tool.
- This tool provides the selected sequences as the best and worst choices from assembled sequences as well as alignment results with rCRS and mtSNP calculation.

Please provide nucleotide sequence in FASTA format.

Input sequences in FASTA format

Start page

Page = 10

[Assembled sequences]

Download buttons Download mtSNP profile Download assembled seq

Elapsed time: 6.8774 sec Results

1. Overview

Assembly overview

2. Detail

a. Range: 15988 ~ 16569, 1 ~ 636 View in a new window

Assembly detail

b. Mini-map

3. mtSNP profile

mtSNP profile

Sample	Control region	Coding region
Best choice	16051G 16088C 16258A 16267T 16291T 16326G 16333T 16396C 73G 146C 234G 263G 309.1C 315.1C 519R 524.1A 524.2C 524.3A 524.4C	
Worst choice	16051G 16088C 16258A 16267T 16291T 16326G 16333T 16396C 61N 62N 73G 146C 159N 189N 234G 263G 287M 309.1C 315.1C 369R 369R 513N 524.1A 524.2C 524.3A 524.4C	589DEL

4. Alignment with rCRS

Alignment detail

5. Sequences

a. Best choice

b. Worst choice

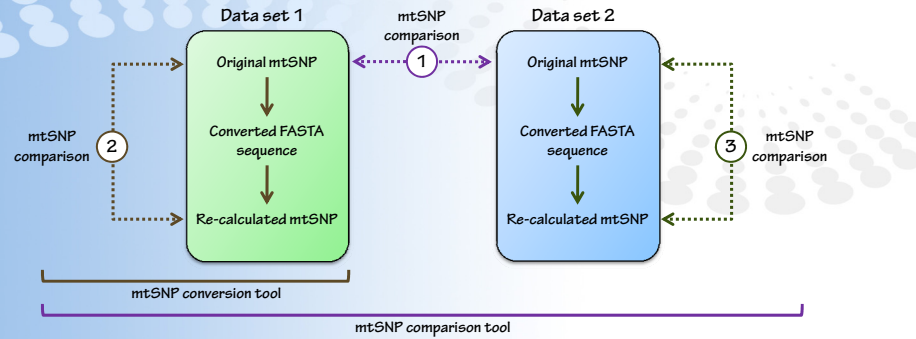
Two candidates of assembled sequences

Download mtSNP profile Download assembled seq

Result page

# Protocols of two mtSNP tools

## - mtSNP conversion and comparison tools



1. mtSNP conversion tool : mtSNP list → nucleotide sequence in FASTA format
2. mtSNP comparison tool : Comparison function between two mtSNP data gathered from independent two experiments
3. Both tools have validity checkup function for the mtSNP in comparison with original input mtSNP and re-calculated mtSNP.

Page ■ 11

## 3. mtSNP conversion tool

The screenshot shows the '1. mtSNP conversion tool' interface. It includes a 'Range selection' button and two output format options: 'mtDNAmanager format' and 'EMPOP format'. The interface also displays mtSNP data and control region information.

Start page – single data set input

The mtDNA assembly tool extracts a consensus sequence from several mtDNA fragment sequences obtained from multiple amplification reactions.

Page ■ 12

The screenshot shows the 'Results' page of the mtSNP conversion tool. It includes a 'Download button' and a 'Download converted FASTA seq' button. The results are organized into sections: 'Included mtSNPs 1', 'Included mtSNPs 2', and 'Excluded mtSNPs'. Each section shows alignment with rCRS and validity checkup of mtSNP. The results are also displayed in a table format.

Result page (Tab1 – alignment with rCRS)

← Tab2 –  
Converted  
FASTA  
sequences

← Tab3 –  
Validity  
checkup of  
mtSNP



# 4. mtSNP comparison tool

**1. mtSNP comparison tool**

- mtSNP comparison tool provides comparison results between two mtSNP data gathered from independent two experiments for the same sample.
- In context of mtSNP conversion tool, sequence range of input mtSNP data will be automatically determined in this tool.
- It will check the validity of mtSNP through the mtDNA reference sequence.

**Data set 1** **Data set 2**

**mtDNA manager format** **or** **EMPOP format**

**Start page - two data sets input**

mtSNP comparison tool provides comparison results between two mtSNP data gathered from independent two experiments for the same sample.

**2. Comparison of sequences converted from two mtSNP data sets**

**Sample 1.**

a. Comparison between two mtSNP data sets

- mtSNPs between two samples (PK-070 of data set 1 and PK-070 of data set 2) are not correct.

**Black : consensus mtSNP, Red : difference mtSNP**

**313.1C**  
**315.1C**

**b. Comparison of sequences converted from two mtSNP data sets**

Alignment with rCRS Validity checkup of mtSNP

mtSNP data set 1 mtSNP data set 2

a. PK-070

a. Range: 16100 - 16560

b. Included mtSNPs: PK-070 16223 16519 T>C 228 234 2496 243 313.1C 345 489 569

c. Alignment with rCRS

rs\_14206-16569

a. Range: 1 - 574

b. Included mtSNPs: PK-070 228 234 2496 243 313.1C 345 489 569

c. Alignment with rCRS

rs\_1-574

d. Excluded mtSNPs: None

**Result page (Tab1 - alignment with rCRS)**

**b. Comparison of sequences converted from two mtSNP data sets**

Alignment with rCRS Validity checkup of mtSNP

mtSNP data set 1 At least, one mtSNP is not correct.

Original PK-070 16111a 16223 16519 T>C 228 234 2496 243 313.1C 345 489 569

Re-calculated PK-070 16111a 16223 16519 T>C 228 234 2496 243 315.1C 345 489 569

mtSNP data set 2 All mtSNPs are correct.

Original PK-070 16111a 16223 16519 T>C 228 234 2496 243 315.1C 345 489 569

Re-calculated PK-070 16111a 16223 16519 T>C 228 234 2496 243 315.1C 345 489 569

**Tab2 - Validity checkup of mtSNP**

1. The description and analysis of mtDNA sequence variations compared to rCRS can be easily carried out using mtDNA nomenclature tool of mtDNAprofiler.
2. mtDNA assembly tool allows users to extract a consensus sequence from several fragment sequences of mtDNA.
3. mtSNP conversion and comparison tools help users to compare and manage the mtSNP data gathered by two independent experiments.
4. Therefore, mtDNAprofiler is a useful program to characterize and analyze mtDNA sequences for researcher dealing with mtDNA.



Thank you