

---

## **mtDNAManager:**

### **An Updates on the Status of the Web-based Tool for Management and Quality Analysis of mtDNA Control Region Sequences**

*Hwan Young Lee, Woo Ick Yang, Kyoung-Jin Shin*

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



## **Outline**

---

### **➤ The first release of mtDNAManager in 2007**

Background

Aims

Implementation

### **➤ Recent updates**

Database expansion

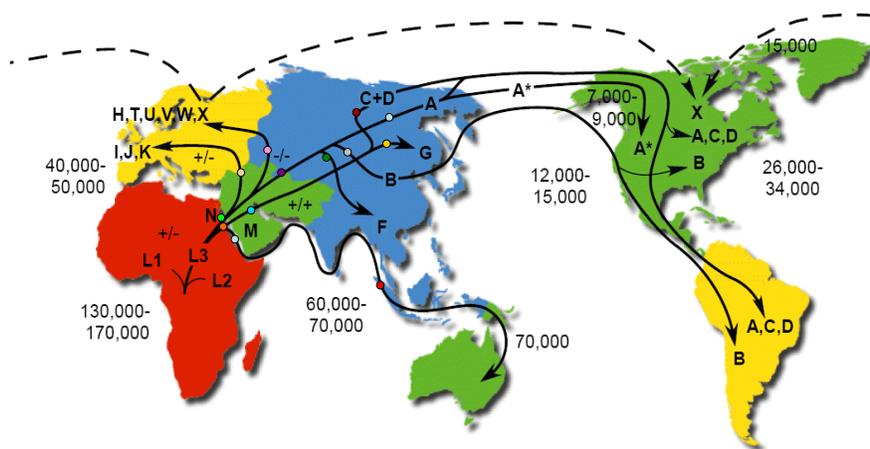
Refinement of control region mutation motifs



## Background

- Mitochondrial DNA (mtDNA) typing is more prone to human error than other forensic DNA analysis
- Errors were mainly due to misinterpretation of sequence raw data and due to the introduction of clerical errors during data transcription
- **Phylogenetic investigations** and **database screening** could have detected prevalent errors in published datasets (e.g., Bandelt et al. Science 2004, 305:1402)

## Human mtDNA migration



## How to avoid mtDNA sequence errors

---

- Attempts to **localize the sequence to a part of phylogeny (haplogroup)**. If the haplogroup motif is not fully represented, recheck the relevant positions in the sequence
- Have in mind the **relative mutability of sites**. **Be sensitive to rare mutations** on different sequence backgrounds in one batch of sequencing
- Look out for incongruence between parts of the sequences which have been obtained in different PCR or sequencing reactions (**artificial recombinants**)

*Bandelt et al. JLM (2001) 115:64-9*



## Need for software development

---

- Methodologies based on **mtDNA haplogroup determination** and **comparisons with existing mtDNA haplotypes** were proposed for preventing mtDNA errors
- Manual haplogroup estimation requires **a thorough understanding of the worldwide mtDNA phylogeny**
- Database screening for systematic error detection requires **high-quality databases** that are publicly available



## mtDNAManager's first release in 2007

**http://mtmanager.yonsei.ac.kr**

mtDNAManager - forensic mtDNA database

mtDNAManager

About  
Services  
Help  
Mutation motifs  
Recent updates  
Contact us

mtDNAManager: A forensic mitochondrial DNA database aimed at supporting data quality control and generating reliable frequency estimates

mtDNAManager provides a convenient web interface for analysis, query and storage for human mtDNA control region sequences. mtDNAManager is made free and open to all users and there is no login requirement. At the same time, mtDNAManager offers the option to store and match data with batch mode for registered users.

The aims of mtDNAManager is (1) to allow researchers to automatically estimate mtDNA haplogroups of their own mtDNA sequences, (2) to generate reliable frequency estimates from high quality mtDNA databases, and (3) to collect and integrate high quality mtDNA control region sequence data publicly available.

YONSEI UNIVERSITY

Copyright © 2007 Yonsei DNA Profiling Group  
Department of Forensic Medicine, Yonsei University College of Medicine  
250 Seongnam, Seodaemun-gu, Seoul 120-752, KOREA



## Worldwide access to mtDNAManager

Map Overlay

23 Aug 2007 - 12 Oct 2010



21,339 visits came from 86 countries/territories

Detail Level	Country/Territory	Visits	Pages/Visit	Avg. Time on Site	% New Visits	Bounce Rate
1.	South Korea	8,790	3.25	00:04:50	7.24%	29.40%
2.	United States	2,474	2.43	00:03:14	65.50%	52.22%
3.	Spain	1,752	2.09	00:02:48	18.21%	37.36%
4.	Germany	783	2.19	00:02:56	47.38%	53.61%
5.	China	472	2.41	00:02:58	62.08%	60.28%



## Worldwide access to mtDNAManager



## Aims of mtDNAManager

- To allow researchers to automatically estimate **the most-probable mtDNA haplogroups** of their mtDNA control region sequences
- To facilitate **database screening** with improved query tools
- To provide researchers with a **convenient interface** for managing and analyzing their own data in batch mode

## Design and content of mtDNAManager

- The mtDNAManager **interface** was designed to allow researchers to **easily query the database** and **immediately view the results** on a single page
- The mtDNAManager's first release contained 4839 mtDNA **control region sequences** from FBI and 593 Korean mtDNA control region sequences and **a set of bioinformatics tools** able to automatically characterize newly submitted data by estimating its most-probable mtDNA haplogroup based on more than 350 haplogroup-specific control region mutation motifs.



## The most-probable haplogroup estimation

- The phased designation of haplogroups (i.e. **expected haplogroup** and **estimated haplogroup**) suggests candidate sites that need reinvestigation by allowing the respective confirmation of the presence of **clear diagnostic mutations** and **accompanying mutations**.

Sample List							
Add	Edit	Delete	Import	Export	Print	N9a3: 16129-16223-16257A-16261-150	
Sample ID	Expected HG	Estimated HG	np 16024-16569	np 16024-16569	np 16024-16569	np 16024-16569	Comments
Demo-01	D4b1	D4b1	16223 16319 16362	16129 16223 16257A 16261	73 152 263 309.1C 315.1C	489 523d 524d	150?
Demo-02	N9a3						
Demo-03	A5b	A5b	16128 16223 16235 16290 16319 16519		73 235 263 315.1C	523d 524d	
Demo-04	D4a3	D4a3	16129 16223 16249 16266 16304 16362 16519		73 152 263 309.1C 315.1C	489	
Demo-05	F1b	F1b	16129R 16182C 16183C 16189 16232A 16249 ...		73 152 249d 263 315.1C	523d 524d	16129R
Demo-06	A4c	A4c	16223 16290 16319 16362		73 146 152 200 235 263 309.1C 315.1C	523d 524d	
Demo-07	F1b F1d	F1d	16158 16189 16232		73 249d 263 309.1C 315.1C	523d 524d	
Demo-08	D4/G	D4/G	16223 16260 16292 16319		73 146 263 309.1C 315.1C	489	16319 missed out?
Demo-09	A5a		16187 16223 16290 16519		73 146 195 235 263 309.1C 309.2C 31...	523d 524d	
Demo-10	D4j1	D4j1	16184 16223 16311 16362		73 263 315.1C	489	A5a: 16187-16223-16290-16319-235-523d-524d
Demo-11	M9a	M9a	16223 16234 16316 16362		73 263 315.1C	489	
Demo-12	B4c1a	B4c1a	16086 16183C 16189 16217 16311 16519		73 263 309.1C 315.1C	523d 524d	
Demo-13	M7b2	M7b2	16129 16189 16223 16297 16298		73 150 152 199 263 309.1C 315.1C	489 573.1C	
Demo-14	D4/G	D4/G	16223 16362		73 263 315.1C	489	
Demo-15	M7a	M7a	16209 16223		73 263 309.1C 315.1C	489 523d 524d	
Demo-16	F1a1	F1a1	16129 16162 16172 16304 16519		64 73 249d 263 309.1C 315.1C	523d 524d	
Demo-17	F2*	F2*			73 195 200 235 249d 263 309.1C 315.1C...	459d	
Demo-18	G3a	G3a	16223 16274 16362		73 143 152 204 263 315.1C	489	
Demo-19	A	A	16179 16223 16290 16319 16519		73 235 263 309.1C 315.1C	523d 524d	
Demo-20	M10b	M10b	16086 16223 16311		73 103 204 263 315.1C	489	

## Database search using query system

- A query system retrieves sequences that **include queried nucleotide polymorphisms** from a selected population or the entire population group of its open database.

**Target database**

**Include setting**

**Estimated mtDNA haplogroup affiliations using the bioinformatics resources of the mtDNAManager**

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
GRC CAU.0003	A4	A4	16223 16249 16290 16319 16362	73 152 236 263 309.1C 315.1C	5236 6249
USA CAU.0023	A2	A2	16111 16136 16172 16223 16290 16311 16319...	73 146 153 235 263 315.1C	
USA CAU.0378	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 315.1C	
USA CAU.0531	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 309.1C 315.1C	
USA CAU.0759	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 309.1C 315.1C	
USA CAU.0977	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 309.1C 315.1C	
USA CAU.1082	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 309.1C 315.1C	
USA CAU.1312	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 309.1C 315.1C	

## Database search using query system

- With the alternative setting of **match**, the mtDNAManager also searches sequences that match the queried sequence data from the database.

**Target database**

**Match setting**

**Frequency estimates = (x+2)/(n+2)**

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
USA CAU.0219	K1c	K1c	16224 16311	73 148 152 263 315.1C	
USA CAU.0341	K1c	K1c	16224 16311	73 148 152 263 315.1C	
USA CAU.0436	K1c	K1c	16224 16311	73 148 152 263 315.1C	
USA CAU.0531	K1c	K1c	16224 16311	73 148 152 263 315.1C	
USA CAU.0759	K1c	K1c	16224 16311	73 148 152 263 315.1C	
USA CAU.0977	K1c	K1c	16224 16311	73 148 152 263 315.1C	
USA CAU.1082	K1c	K1c	16224 16311	73 148 152 263 315.1C	
USA CAU.1312	K1c	K1c	16224 16311	73 148 152 263 315.1C	



mtDNAManager - forensic mtDNA database - Windows Internet Explorer

http://mtmanager.yonsei.ac.kr/search\_sample.php

mtDNAManager - forensic mtDNA database

Welcome, Colleague  
Logout Change Profile

**A query system**

Sample Match Query Home

Target Group  
 My mtDNA  
 Open DB: African

Name Description  
 Demo-1 Demo Sequences  
 Demo-2 Demo Sequences

Searching Option  
 Sequence: 16187 235  
 Region:  HV1  HV2  HV3  Control Region  
 Match  
 Ignore heteroplasmic insertions in the poly C-stretches  
 Maximum number of mismatched nucleotides: 0  
 Include the queried nucleotide polymorphism  
 Search Reset Frequency Estimates

Expected HG: A5a Estimated HG: H Comments:

Searching Result  
 No. of Matched Samples: 19 No. of Target Samples: 593 Match Probability: 0.0353  
 Export Print

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
K017	A5a	A5a	16187 16223 16290 16319	73 235 263 315.1C	523d 524d
K122	A5a	A5a	16187 16223 16290 16319	73 235 263 309.1C 315.1C	523d 524d
K173	A5a	A5a	16187 16189 16223 16290 16319	73 235 263 309.1C 315.1C	523d 524d
K184	A5a	A5a	16187 16223 16290 16296 16319	73 235 263 309.1C 315.1C	523d 524d
K191	A5a	A5a	16187 16223 16290 16319 16519	73 146 195 235 263 309.1C 309.2C 31...	523d 524d
K216	A5a	A5a	16187 16192 16223 16270 16290 16319	73 235 263 309.1C 315.1C	523d 524d
K222	A5a	A5a	16187 16223 16290 16319	73 235 263 309.1C 315.1C	523d 524d
K263	A5a	A5a	16086 16187 16223 16290 16319	73 235 263 309.1C 315.1C	523d 524d
K271	A5a	A5a	16187 16223 16290 16319 16519	73 195 235 263 309.1C 309.2C 315.1C	523d 524d
K308	A5a	A5a	16187 16223 16290 16319	73 235 263 315.1C	523d 524d
K382	A5a	A5a	16187 16223 16290 16319	73 235 263 315.1C	523d 524d
K386	A5a	A5a	16187 16215 16223 16290 16319	73 235 263 309.1C 315.1C	523d 524d

안료 인터넷 100%

mtDNAManager - a forensic mtDNA tool - Windows Internet Explorer

http://mtmanager.yonsei.ac.kr/index.php

mtDNAManager - a forensic mtDNA tool

E-mail: \_\_\_\_\_ Login  
 Password: \_\_\_\_\_ Register

mtDNAManager

Query Home

About  
 Services  
 Database  
 Statistics  
 Haplogroup  
 My mtDNA  
 Help  
 Query  
 Sample  
 Match  
 Mutation motifs  
 Recent updates  
 Contact us

mtDNAManager: a Web-based tool for the management and quality analysis of mitochondrial DNA control region sequences

mtDNAManager provides a convenient web interface to analyze, query and store human mtDNA control region sequences (*BMC Bioinformatics*, 2008;9(1):473). mtDNAManager is made free and open to all users and there is no login requirement. At the same time, mtDNAManager offers the option to store and match data with batch mode for registered users.

- [Access demo](#)

The aims of mtDNAManager are (1) to allow researchers to automatically estimate the most-probable mtDNA haplogroups of their mtDNA control region sequences, (2) to facilitate database screening with improved query tools and (3) to provide researchers with a convenient interface for managing and analysing their own data in batch mode.

YONSEI UNIVERSITY  
 Copyright (c) 2007 Yonsei DNA Profiling Group  
 Department of Forensic Medicine, Yonsei University, College of Medicine  
 250 Seongsanno, Seodaemun-gu, Seoul 120-752, KOREA

인터넷 100%

## Updates from the first release

- **The number of mtDNA control region sequences** in the mtDNAManager's open database has grown **from 5,432 to 9,294**, while the number of population groups has been increased to more than 20.
- **The number of control region mutation motifs** for the assignment of the most-probable mtDNA haplogroups has grown **from 350 to more than 630**.



## mtDNAManager's current open database

Metapopulation	Subpopulation	No. of sequences	References
African	African	1148	Forensic Sci Commun (2002) 4 Online
	Kenyan	100	Int J Legal Med (2004) 118:294-306
	U.S. African	248	Forensic Sci Int Genet (2009) 4(1):45-52
Caucasian	Caucasian	1655	Forensic Sci Commun (2002) 4 Online
	Armenian	273	Forensic Sci Int (2007) 166:164-175
	Darbazi	206	Forensic Sci Int Genet (2009) 4(4):87-88
	Finnish	200	Forensic Sci Int (2007) 172:171-178
	German	313	Forensic Sci Int (2007) 172:218-224 Int J Legal Med (2006) 120:310-314
West Eurasian	Greek	410	Int J Legal Med (2000) 122:87-89
	Hungarian	416	Int J Legal Med (2007) 121:377-383
	Macedonian	200	Forensic Sci Int Genet (2007) 1(1-4):4-9
	East Asian	753	Forensic Sci Commun (2002) 4 Online
East Asian	Japanese	211	Int J Legal Med (2003) 117:218-225
	Korean	593	Int J Legal Med (2006) 120:9-14
	Chinese (Hong Kong)	377	Forensic Sci Int Genet (2009) 3(4):119-125
	Malay	205	Legal Med (2007) 9: 33-37
	Vietnamese	187	Int J Legal Med (2008) 122:257-259
Oceanian	Oceanian	114	GenBank (Phylotree)
Admixed	Hispanic	686	Forensic Sci Commun (2002) 4 Online
	Debian	249	Forensic Sci Int Genet (2008) 2(1):9-10
	Egyptian	277	Forensic Sci Int Genet (2009) 3(3):97-103
	Northern African	120	Forensic Sci Int Genet (2009) 3(3):164-172
Total	U.S. Hispanic	253	Forensic Sci Int Genet (2008) 2(2):19-23 Forensic Sci Int Genet (2008) 2(3):45-48
	Venezuelan	100	Forensic Sci Int Genet (2008) 2(4):61-64
<b>Total</b>		<b>8284</b>	

20 more publications were added



# Expansion of mtDNAmanager's database

**DB at 1st release**

**Current DB**

Sample ID	Expected HG	Estimated HG	np 16024-16059	np 001-437
r1B4	K1b1	K1b1	16093 16224 16311 16319 16463 16519	73 151 152 198
r1E6	J1c0	J1c0	16089 16126 16201 16285 16319	73 185 228 283
r3D1	H5	H5	16304 16319	263 309 1C 309
mtG1	J1c0	J1c0	16089 16126 16193 16319 16519	73 152 263 295
m2E5	J1c0	J1c0	16089 16126 16201 162325 16285 16319	73 105 220 263

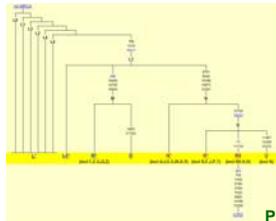


# Improved database search

Population	Group Name	Description	No. of Matched Samples	No. of Population Samples	Match Probability
African	African	FSI: Genet (2006), 2:481-91	348	1888	0.2383
Dutch and English	Dutch and English	FSI: Genet (2006), 2:481-91	19	249	0.0767
Hispanic	Hispanic	Forensic Sci Comm (2002), 4: (Online)	219	888	0.2457
Turkish and Moroccan	Turkish and Moroccan	FSI: Genet (2006), 2:481-91	1	128	0.0078
U.S. Hispanic	U.S. Hispanic	FSI: Genet (2006), 2:481-91; FSI: Genet (2006), 2:481-91	85	293	0.2902
Hispanic	Hispanic	FSI: Genet (2006), 2:481-91	34	168	0.2024
African	African	Forensic Sci Comm (2002), 4: (Online)	23	1148	0.0201
Hispanic	Hispanic	Int J Legal Med (2006), 118: 294-308	1	109	0.0091
U.S. African	U.S. African	FSI: Genet (2006), 2:481-91	3	248	0.0121
East Asian	East Asian	FSI: Genet (2006), 2:481-91	238	2328	0.1026
Chinese	Hong Kong	FSI: Genet (2006), 2:481-91	22	377	0.0583
East Asian	FSI: Asian	Forensic Sci Comm (2002), 4: (Online)	86	788	0.1091
Japanese	Japanese	Int J Legal Med (2006), 117: 219-228	24	211	0.1137
Korean	Korean/Chinese	Int J Legal Med (2006), 120: 5-14	85	593	0.1432
Indian	India	Int J Legal Med (2006), 120: 5-14	8	288	0.0278



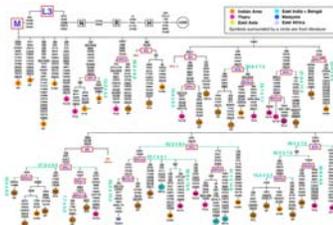
# Recent updates in phylogenetic trees



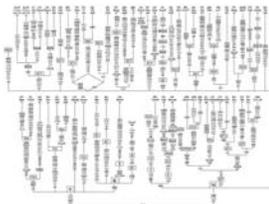
Phylotree.org



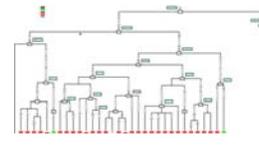
South East Asian: -M71, M72



South Asian: M31, M33, M51...



East Asian: M74, M75, M76, N10, N11...



African: L0-L5



YONSEI UNIVERSITY  
COLLEGE OF MEDICINE

miDNAManager - forensic mtDNA database - Windows Internet Explorer

http://mtmanager.yonsei.ac.kr/info.php

mtDNAManager - forensic

Control Region Mutation motif for more than 630 mtDNA haplogroups

mtDNAManager

Query Home

About

Services

- Haplogroup
- Database
- Statistics
- My mtDNA

Help

- Query
- Sample
- Match

Contact us

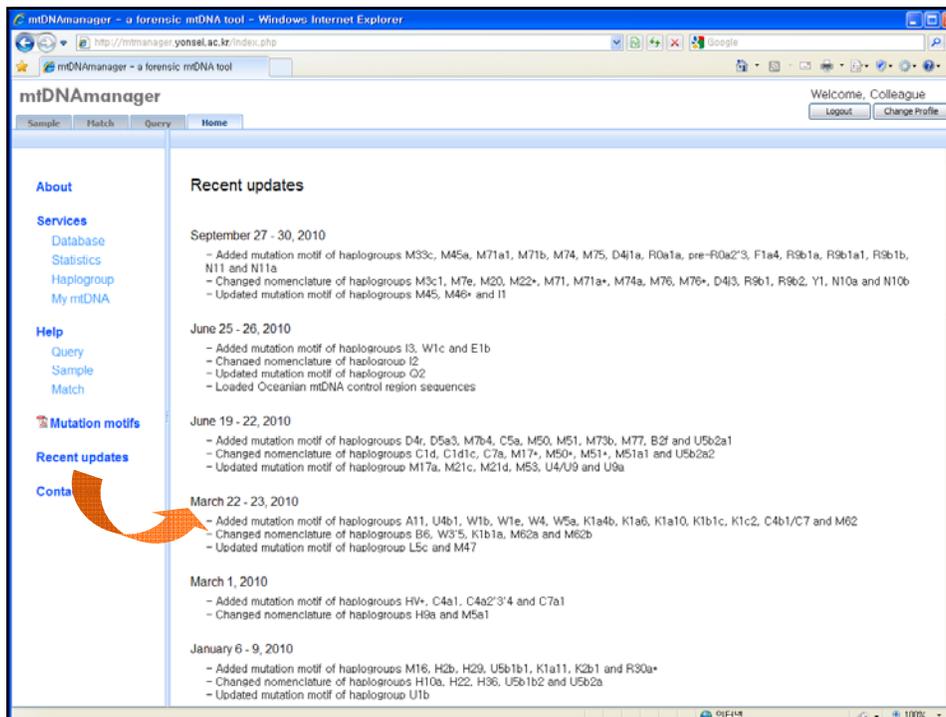
- Mutation motifs

mtDNA Haplogroup Specific Control Region Mutation Motifs

African Haplogroup	H1V1	H1V2	H1V3 etc.
L0	16129-16187-16189-16223-16230-16311	247	
L0a	16136-16148-16172-16187-16188G-16189-16231-16236-16311-16320	93-185-189-236-247	526A-526J
L0a1	16136-16148-16168-16172-16187-16188G-16189-16231-16236-16311-16320	93-185-189-(236)-247	526A-526J
L0a1a	16129-16148-16168-16172-16187-16188G-16189-16231-16236-16311-16320	93-185-189-200-236-247	526A-526J
L0a1b	16129-16148-16168-16172-16187-16188G-16189-16231-16236-16311-16320	93-95C-185-189-236-247	526A-526J
L0a2	16129-16148-16172-16187-16188G-16189-16231-16236-16311-16320	93-152-189-236-247	64, 325A-324A
L0f	16189-16187-16189-16223-(16230)-16278-16311-16322	(146)-189-247	2698B
L0k	16166C-16172-16187-16189-16214-16223-16230-16278-16291G-16311	140-189-195-247	
L0d	16129-(16187)-16189-16223-16230-16243-16311	247	
L1	16187-16189-16223-16278-16311	182-247	326A-526J
L1b	16136-16187-16189-16223-(16264)-16278-16291-16311	(182)-187F-(195)-247	357, 526A-526J
L1b1	16126-16187-16189-16223-16264-16270-16278-16291-16311	182-185T-195-247	357, 526A-526J
L1b1*	16136-16187-16189-16223-16264-16270-16278-16291-16311	182-185T-195-247	357, 526A-526J
L1c	16129-16187-16189-16223-16278-16311	152-(182)-186A-189C-247(see 247b)-316	526A-526J

Copyright (c) 2007 Yonsei DNA Profiling Group  
Department of Forensic Medicine, Yonsei University College of Medicine  
250 Seongsanno, Seodaemun-gu, Seoul 120-752, KOREA





## Concluding remarks

- With these improvements mtDNAManager will help in checking the quality of data and facilitate data comparisons from a phylogenetic perspective
- Continuous efforts are needed to collect and integrate high-quality mtDNA control region sequence data for various population groups in South East Asia and Oceania

---

*For comments, bug reports, suggestions for improvement, please contact us through the website (<http://mtmanager.yonsei.ac.kr>).*