mtDNAmanager:

A Forensic Mitochondrial DNA Database Aimed at Supporting Data Quality Control and Generating Reliable Frequency Estimates

Hwan Young Lee[1], Injee Song[2], Eunho Ha[3], Sung-Bae Cho[2], Woo Ick Yang[1], Kyoung-Jin Shin[1]

(1) Department of Forensic Medicine, Yonsei University College of Medicine, Seoul, Korea
(2) Department of Computer Science, Yonsei University, Seoul, Korea
(3) Department of Information and Statistics, Yonsei University at Wonju, Gangwon-do, Korea

---

mtDNA errors on forensic work

- In crime-scene samples
  - **False exclusion**
- In surveys to enlarge the reference database
  - **Generation of new mtDNA types**
  - **Reduce the estimate of the frequency of real type**

➡ ***Distorted match probability making a match between a crime scene and a suspect more impressive than it really is***

mtDNA data quality control

Can we tell from a single entry of an mtDNA data table whether it is authentic?

Errors in mtDNA data can be detected

Science 3 September 2004:
Vol. 305. no. 5689, pp. 1402 - 1404
DOI: 10.1126/science.305.5689.1402b

LETTERS

Problems in FBI mtDNA Database

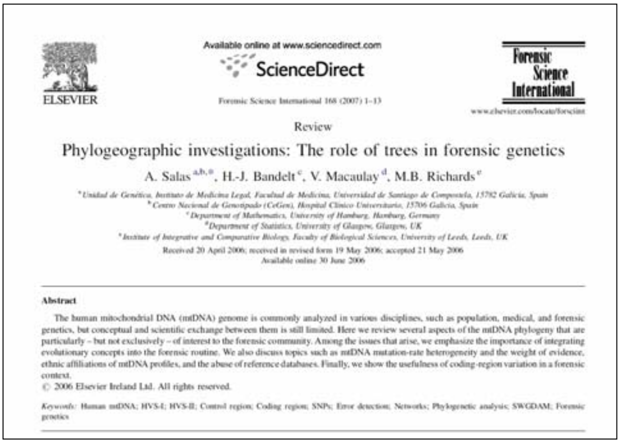
In his Editorial "Forensic science: oxymoron?" (5 Dec. 2003, p. 1625), D. Kennedy wonders why the U.S. National Institute of Justice has regularly resisted comprehensive evaluations of the science underlying forensic techniques. A possible answer can be found in the poor quality of the forensic human mitochondrial DNA (mtDNA) database used by the Federal Bureau of Investigation (FBI), which is included in the Scientific Working Group on DNA Analysis Methods (SWGDAM) database (1).

A thorough inspection of the original "African-American" database, which has been contributed by the FBI laboratory, reveals a number of major deficiencies. Among 1148 entries, each comprising two separately sequenced segments from the mtDNA control region, we detected as many as five artificial combinations of totally unrelated mtDNA segments stemming from different samples, which suggest fatal sample mix-up in the lab or during data transcription (2). The most striking hybrid (USA.AFR.000942) we found combined segment I from an African haplogroup (referred to as L1b) (3) with segment II from a Native American haplogroup (called C1) (4).

mtDNA database - USA.AFR.000942

HV1	HV2
16126-16187-16189-16223-16264 16270-16278-16293-16311-16519	73-249d-263-290d-291d 309.1C-315.1C-489
African haplogroup L1b	Asian haplogroup C1

How can we estimate mtDNA data quality



Available online at www.sciencedirect.com
ScienceDirect
Forensic Science International 168 (2007) 1–13
www.elsevier.com/locate/forsciint

Review

Phylogeographic investigations: The role of trees in forensic genetics

A. Salas[a,b,*], H.-J. Bandelt[c], V. Macaulay[d], M.B. Richards[e]

[a] Unidad de Genética, Instituto de Medicina Legal, Facultad de Medicina, Universidad de Santiago de Compostela, 15782 Galicia, Spain
[b] Centro Nacional de Genotipado (CeGen), Hospital Clínico Universitario, 15706 Galicia, Spain
[c] Department of Mathematics, University of Hamburg, Hamburg, Germany
[d] Department of Statistics, University of Glasgow, Glasgow, UK
[e] Institute of Integrative and Comparative Biology, Faculty of Biological Sciences, University of Leeds, Leeds, UK

Received 20 April 2006; received in revised form 19 May 2006; accepted 21 May 2006
Available online 30 June 2006

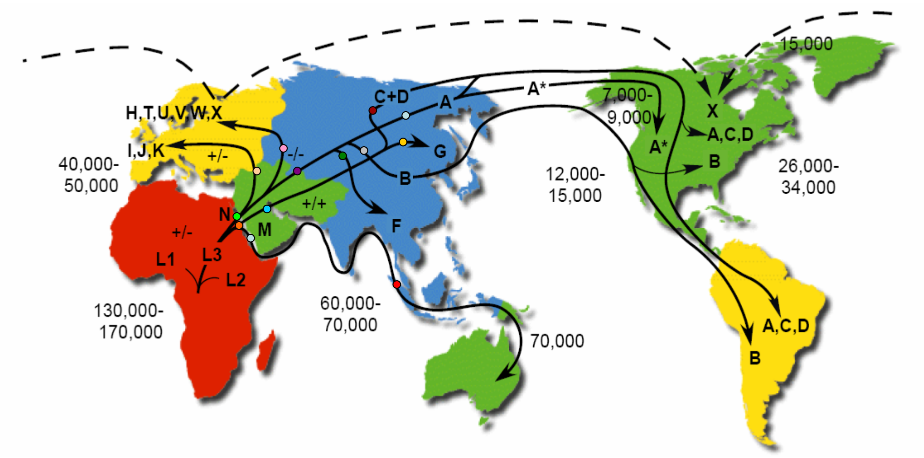
Abstract

The human mitochondrial DNA (mtDNA) genome is commonly analyzed in various disciplines, such as population, medical, and forensic genetics, but conceptual and scientific exchange between them is still limited. Here we review several aspects of the mtDNA phylogeny that are particularly – but not exclusively – of interest to the forensic community. Among the issues that arise, we emphasize the importance of integrating evolutionary concepts into the forensic routine. We also discuss topics such as mtDNA mutation-rate heterogeneity and the weight of evidence, ethnic affiliations of mtDNA profiles, and the abuse of reference databases. Finally, we show the usefulness of coding-region variation in a forensic context.

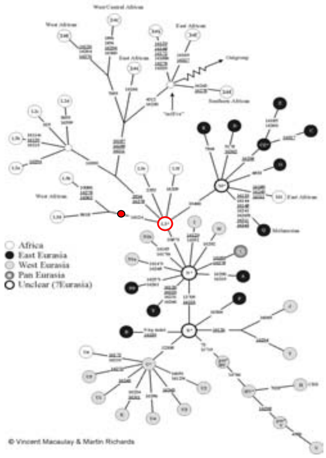
Keywords: Human mtDNA; HVS-I; HVS-II; Control region; Coding region; SNPs; Error detection; Networks; Phylogenetic analysis; SWGDAM; Forensic genetics

Phylogenetic analysis is the key tool

mtDNA evolves along a tree



A skeleton phylogeny of human mtDNA

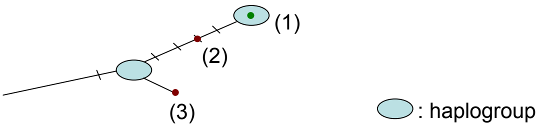


© Vincent Macaulay & Martin Richards

- **Haplotype** -- the mtDNA sequence variations within an individual (e.g. your HV1/HV2 type)
- **Haplogroup** (coined by Torroni et al) -- a group of related haplotypes. These form monophyletic clade on a phylogenetic tree. ***The member of a haplogroup carry a specific sequence motif as a consequence of sharing a common ancestor***

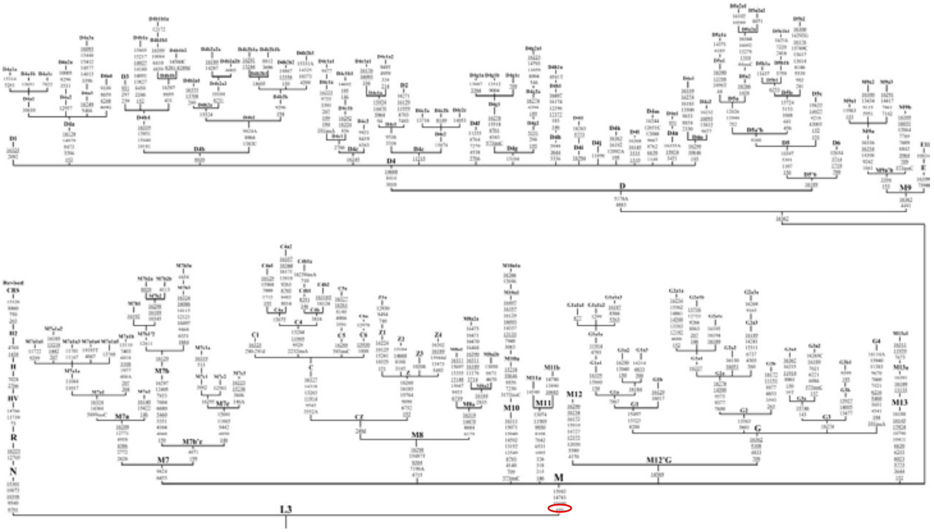
A single entry of an mtDNA data

- The sequence matches a type already found – Allocation of a sequence to **an appropriate haplogrouop (1)**
- It is located at a previously unsampled interior node of the tree (2)
- It creates a new tip of the tree by an unusual mutation (3)
- Haplogroup determination failure due to artificial recombination (4)



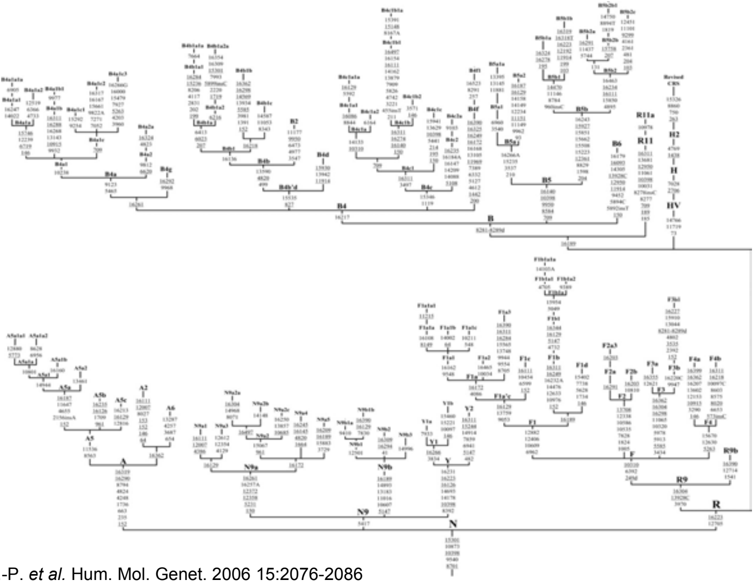
If artefacts occurred in the data set, then they are normally manifest in those seeming #2, #3 and #4

East Asian macrohaplogroup M



Kong, Q.-P. *et al.* Hum. Mol. Genet. 2006 15:2076-2086

East Asian macrohaplogroup N



Kong, Q.-P. *et al.* Hum. Mol. Genet. 2006 15:2076-2086

Self-help guide to avoid 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**)

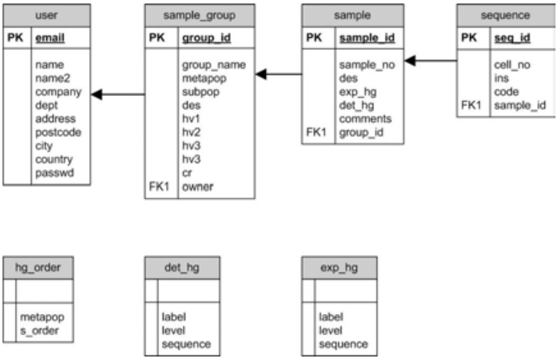
Need to establish a big reliable database with mtDNAs having appropriate haplogroup designation

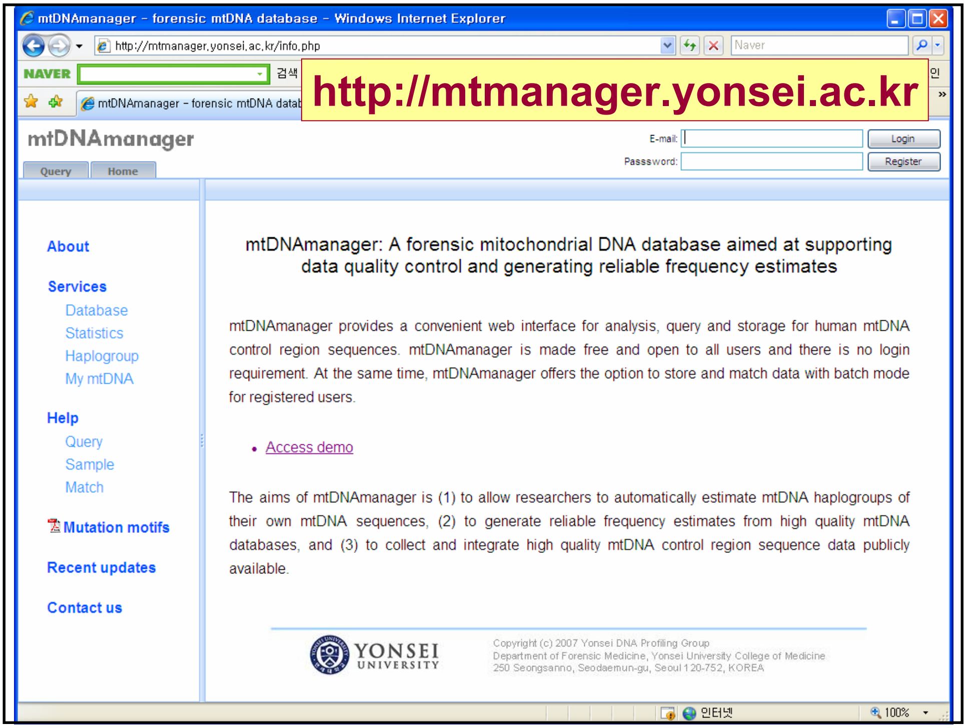
mtDNAmanager (http://mtmanager.yonsei.ac.kr)

- The goal of mtDNAmanager is to provide a Web-based forensic mitochondrial DNA bioinformatics resource for supporting **data quality control** and generating **reliable frequency estimates** using a new approach based on haplogroup estimation and data comparison with the contents of a given database.
- mtDNAmanager consists of previously reported **high quality mtDNA sequences**, and a set of **bioinformatics tools**, able to automatically characterize newly submitted data by **estimating its haplogroup according to the haplogroup-specific control region mutation motif**.

mtDNAmanager structure

- The database is of relational type, designed and implemented using the multithreaded, multi-user **SQL database management system (DBMS), MySQL**.
- Web interfaces have all been implemented using Asynchronous JavaScript and XML (AJAX) technique based on **Javascript and PHP**.
- The system is optimized with **internet explorer v. 6.0 or more**.





http://mtmanager.yonsei.ac.kr

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.

- Access demo

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.

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

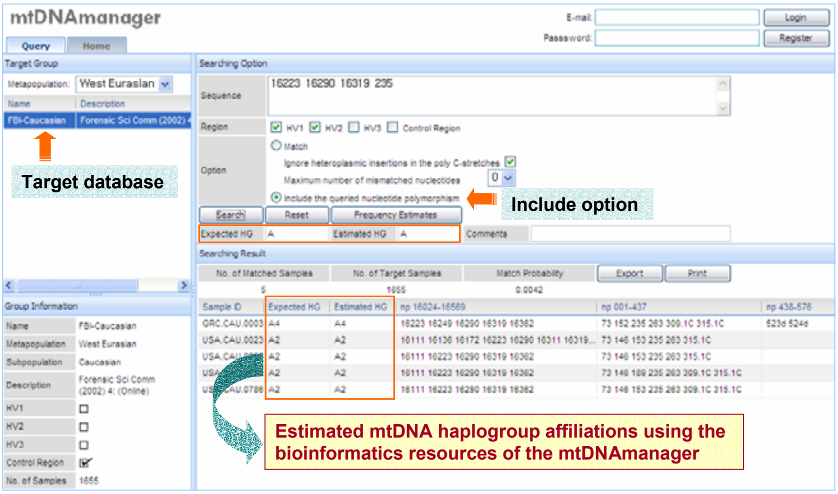
mtDNAmanager Database

- The mtDNAmanager's open database contains **6597 mtDNA** control region sequences and they are grouped in the following five subsets: **African (1248), West Eurasian (2857), East Asian (1557), Oceanian, and Admixed (935)**.

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
West Eurasian	Caucasian	1655	Forensic Sci Commun (2002) 4 *Online*
	Austrian	273	Forensic Sci Int (2007) 166:164-175
	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
	Hungarian	416	Int J Legal Med (2007) 121:377-383
East Asian	East Asian	753	Forensic Sci Commun (2002) 4 *Online*
	Japanese	211	Int J Legal Med (2003) 117:218-225
	Korean	593	Int J Legal Med (2006) 120:5-14
Oceanian	-	-	-
Admixed	Hispanic	686	Forensic Sci Commun (2002) 4 *Online*
	Dubaian	249	Forensic Sci Int: Genetics (2007) *in press*
Total		**6597**	

Database search

- By default but with an exchangeable setting, a **query system** retrieves sequences which **include the queried nucleotide polymorphisms** therein from a user-selected database.



Metapopulation: West Eurasian

Name	Description
FBI-Caucasian	Forensic Sci Comm (2002) 4

Target database

Sequence: 16223 16290 16319 235

Region: ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region

Option:
○ Match
Ignore heteroplasmic insertions in the poly C-stretches ☑
Maximum number of mismatched nucleotides 0
● Include the queried nucleotide polymorphism

Include option

Expected HG: A
Estimated HG: A
Comments:

No. of Matched Samples	No. of Target Samples	Match Probability
5	1655	0.0042

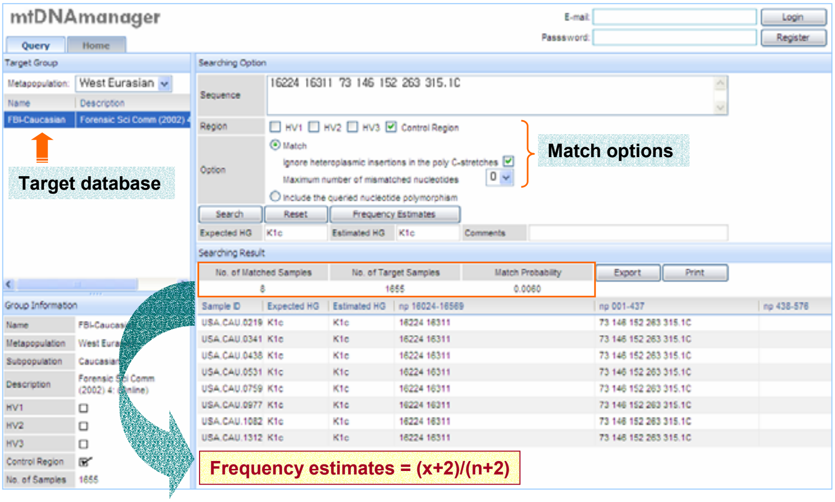
Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
GRC.CAU.0003	A4	A4	16223 16249 16290 16319 16362	73 152 235 263 309.1C 315.1C	523d 524d
USA.CAU.0023	A2	A2	16111 16136 16172 16223 16290 16311 16319...	73 146 153 235 263 315.1C	
USA.CAU[illegible]	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 315.1C	
USA[illegible]	A2	A2	16111 16223 16290 16319 16362	73 146 189 235 263 309.1C 315.1C	
US[illegible]CAU.0786	A2	A2	16111 16223 16290 16319 16362	73 146 153 235 263 309.1C 315.1C	

Group Information	
Name	FBI-Caucasian
Metapopulation	West Eurasian
Subpopulation	Caucasian
Description	Forensic Sci Comm (2002) 4: (Online)
HV1	☐
HV2	☐
HV3	☐
Control Region	☑
No. of Samples	1655

Estimated mtDNA haplogroup affiliations using the bioinformatics resources of the mtDNAmanager

Database search and statistics

- A query system of the mtDNAmanager also searches sequences which show **a match with the queried sequence data** from the database.



Target Group

Metapopulation: West Eurasian

Name	Description
FBI-Caucasian	Forensic Sci Comm (2002) 4

Searching Option

Sequence: 16224 16311 73 146 152 263 315.1C

Region: ☐ HV1 ☐ HV2 ☐ HV3 ☑ Control Region

Option: ● Match

Ignore heteroplasmic insertions in the poly C-stretches ☑

Maximum number of mismatched nucleotides: 0

○ Include the queried nucleotide polymorphism

Expected HG: K1c | Estimated HG: K1c | Comments:

Searching Result

No. of Matched Samples	No. of Target Samples	Match Probability
8	1655	0.0060

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
USA.CAU.0219	K1c	K1c	16224 16311	73 146 152 263 315.1C	
USA.CAU.0341	K1c	K1c	16224 16311	73 146 152 263 315.1C	
USA.CAU.0438	K1c	K1c	16224 16311	73 146 152 263 315.1C	
USA.CAU.0531	K1c	K1c	16224 16311	73 146 152 263 315.1C	
USA.CAU.0759	K1c	K1c	16224 16311	73 146 152 263 315.1C	
USA.CAU.0977	K1c	K1c	16224 16311	73 146 152 263 315.1C	
USA.CAU.1082	K1c	K1c	16224 16311	73 146 152 263 315.1C	
USA.CAU.1312	K1c	K1c	16224 16311	73 146 152 263 315.1C	

Group Information

Name	FBI-Caucasian
Metapopulation	West Eurasian
Subpopulation	Caucasian
Description	Forensic Sci Comm (2002) 4: (Online)
HV1	☐
HV2	☐
HV3	☐
Control Region	☑
No. of Samples	1655

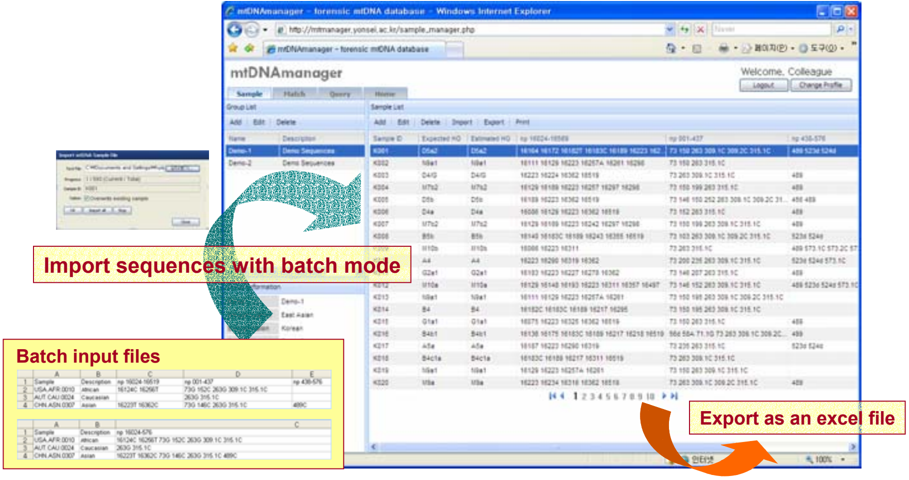
Target database

Match options

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

My mtDNA

- **Upon registration**, the mtDNAmanager provides web interfaces where **users can submit and store their own data with batch mode** and search sequences which show a match or include the queried nucleotide polymorphisms from their own database as well as the mtDNAmanager database.



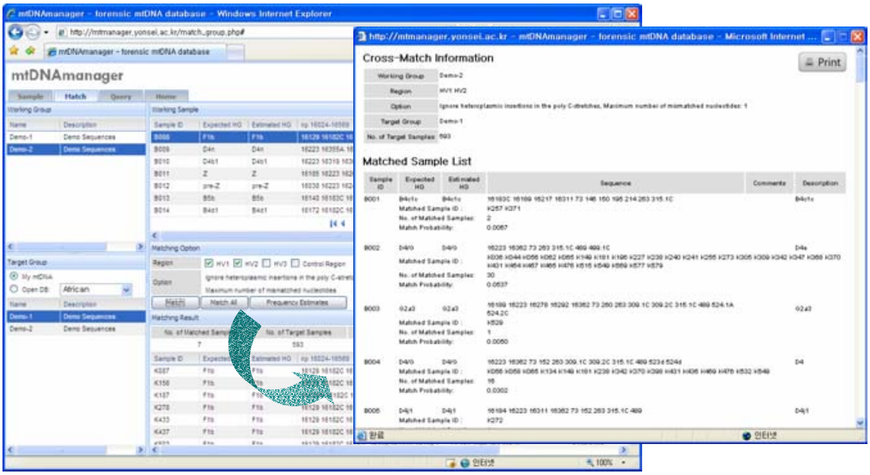
Import sequences with batch mode

Batch input files

Export as an excel file

Cross-match between data groups

- **Upon registration**, a match system permits **cross match of all sequence data between two selected groups** as well as retrieval of matched sequences for one of the data of a user-created data group from their own database or from the mtDNAmanager database.



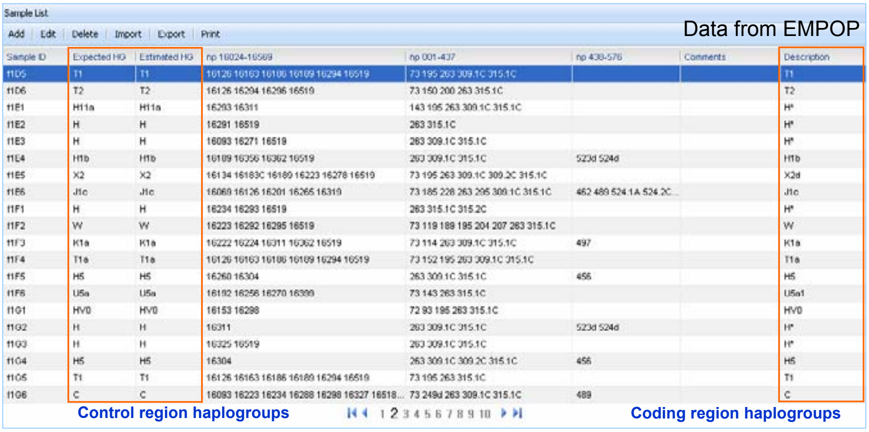
Cross-Match Information

Working Group	Demo-2
Region	HV1 HV2
Option	Ignore heteroplasmic insertions in the poly C-stretches, Maximum number of mismatched nucleotides: 1
Target Group	Demo-1
No. of Target Samples	593

Matched Sample List

---

mtDNA haplogroup estimation



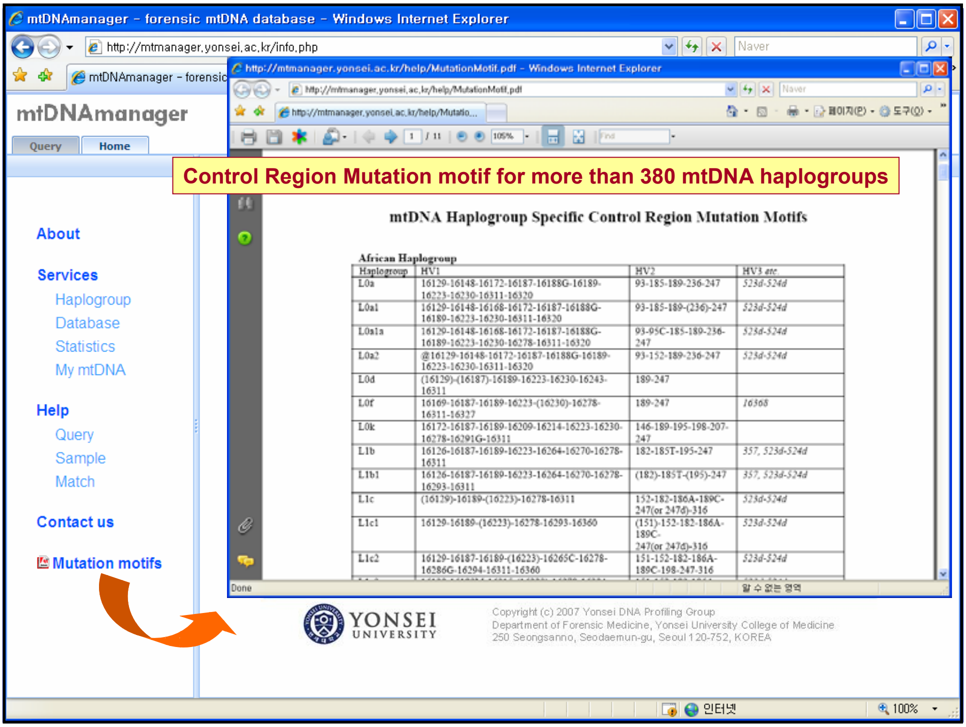
Data from EMPOP

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576	Comments	Description
f1D5	T1	T1	16126 16163 16186 16189 16294 16519	73 195 263 309.1C 315.1C			T1
f1D6	T2	T2	16126 16294 16296 16519	73 150 200 263 315.1C			T2
f1E1	H11a	H11a	16293 16311	143 195 263 309.1C 315.1C			H*
f1E2	H	H	16291 16519	263 315.1C			H*
f1E3	H	H	16093 16271 16519	263 309.1C 315.1C			H*
f1E4	H1b	H1b	16189 16356 16362 16519	263 309.1C 315.1C	523d 524d		H1b
f1E5	X2	X2	16134 16183C 16189 16223 16278 16519	73 195 263 309.1C 309.2C 315.1C			X2d
f1E6	J1c	J1c	16069 16126 16201 16265 16319	73 185 228 263 295 309.1C 315.1C	462 489 524.1A 524.2C...		J1c
f1F1	H	H	16234 16293 16519	263 315.1C 315.2C			H*
f1F2	W	W	16223 16292 16295 16519	73 119 189 195 204 207 263 315.1C			W
f1F3	K1a	K1a	16222 16224 16311 16362 16519	73 114 263 309.1C 315.1C	497		K1a
f1F4	T1a	T1a	16126 16163 16186 16189 16294 16519	73 152 195 263 309.1C 315.1C			T1a
f1F5	H5	H5	16260 16304	263 309.1C 315.1C	456		H5
f1F6	U5a	U5a	16192 16256 16270 16399	73 143 263 315.1C			U5a1
f1G1	HV0	HV0	16153 16298	72 93 195 263 315.1C			HV0
f1G2	H	H	16311	263 309.1C 315.1C	523d 524d		H*
f1G3	H	H	16325 16519	263 309.1C 315.1C			H*
f1G4	H5	H5	16304	263 309.1C 309.2C 315.1C	456		H5
f1G5	T1	T1	16126 16163 16186 16189 16294 16519	73 195 263 315.1C			T1
f1G6	C	C	16093 16223 16234 16288 16298 16327 16518...	73 249d 263 309.1C 315.1C	489		C

Control region haplogroups

Coding region haplogroups

Using the bioinformatics tools of mtDNAmanager, **more than 98% of mtDNAs** of high quality datasets can be allocated to the same mtDNA haplogroups as those confirmed with the coding region SNP information.



Control Region Mutation motif for more than 380 mtDNA haplogroups

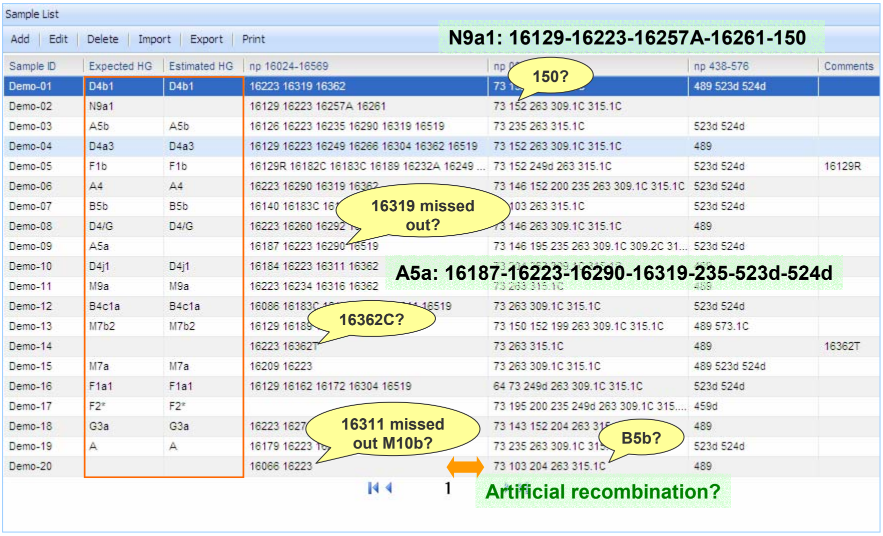
mtDNA Haplogroup Specific Control Region Mutation Motifs

African Haplogroup

Haplogroup	HV1	HV2	HV3 *etc.*
L0a	16129-16148-16172-16187-16188G-16189-16223-16230-16311-16320	93-185-189-236-247	*523d-524d*
L0a1	16129-16148-16168-16172-16187-16188G-16189-16223-16230-16311-16320	93-185-189-(236)-247	*523d-524d*
L0a1a	16129-16148-16168-16172-16187-16188G-16189-16223-16230-16278-16311-16320	93-95C-185-189-236-247	*523d-524d*
L0a2	@16129-16148-16172-16187-16188G-16189-16223-16230-16311-16320	93-152-189-236-247	*523d-524d*
L0d	(16129)-(16187)-16189-16223-16230-16243-16311	189-247	
L0f	16169-16187-16189-16223-(16230)-16278-16311-16327	189-247	*16368*
L0k	16172-16187-16189-16209-16214-16223-16230-16278-16291G-16311	146-189-195-198-207-247	
L1b	16126-16187-16189-16223-16264-16270-16278-16311	182-185T-195-247	*357, 523d-524d*
L1b1	16126-16187-16189-16223-16264-16270-16278-16293-16311	(182)-185T-(195)-247	*357, 523d-524d*
L1c	(16129)-16189-(16223)-16278-16311	152-182-186A-189C-247(or 247d)-316	*523d-524d*
L1c1	16129-16189-(16223)-16278-16293-16360	(151)-152-182-186A-189C-247(or 247d)-316	*523d-524d*
L1c2	16129-16187-16189-(16223)-16265C-16278-16286G-16294-16311-16360	151-152-182-186A-189C-198-247-316	*523d-524d*

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

mtDNA data quality control



Sample ID	Expected HG	Estimated HG	np 16024-16569	np 0	np 438-576	Comments
Demo-01	D4b1	D4b1	16223 16319 16362	73 1	489 523d 524d	
Demo-02	N9a1		16129 16223 16257A 16261	73 152 263 309.1C 315.1C		
Demo-03	A5b	A5b	16126 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	A4	A4	16223 16290 16319 16362	73 146 152 200 235 263 309.1C 315.1C	523d 524d	
Demo-07	B5b	B5b	16140 16183C 16	103 263 315.1C	523d 524d	
Demo-08	D4/G	D4/G	16223 16260 16292	73 146 263 309.1C 315.1C	489	
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	[illegible]	[illegible]	
Demo-11	M9a	M9a	16223 16234 16316 16362	73 263 315.1C	489	
Demo-12	B4c1a	B4c1a	16086 16183C [illegible] 16519	73 263 309.1C 315.1C	523d 524d	
Demo-13	M7b2	M7b2	16129 16189	73 150 152 199 263 309.1C 315.1C	489 573.1C	
Demo-14			16223 16362T	73 263 315.1C	489	16362T
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....	459d	
Demo-18	G3a	G3a	16223 1627	73 143 152 204 263 315	489	
Demo-19	A	A	16179 16223 16	73 235 263 309.1C 315	523d 524d	
Demo-20			16066 16223	73 103 204 263 315.1C	489	

N9a1: 16129-16223-16257A-16261-150

150?

16319 missed out?

A5a: 16187-16223-16290-16319-235-523d-524d

16362C?

16311 missed out M10b?

B5b?

Artificial recombination?

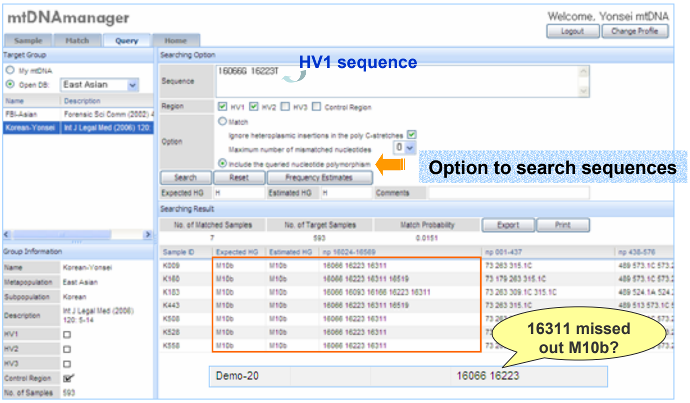
Detection of errors in mtDNA sequence

16311 missed out M10b?

B5b?

Demo-20			16066 16223	73 103 204 263 315.1C	489

Artificial recombination?



HV1 sequence

Option to search sequences

16311 missed out M10b?

Target Group

○ My mtDNA

● Open DB: East Asian

Name	Description
FBI-Asian	Forensic Sci Comm (2002) 4
Korean-Yonsei	Int J Legal Med (2006) 120:

Group Information

Name	Korean-Yonsei
Metapopulation	East Asian
Subpopulation	Korean
Description	Int J Legal Med (2006) 120: 5-14
HV1	☐
HV2	☐
HV3	☐
Control Region	☑
No. of Samples	593

Searching Option

Sequence: 16066G 16223T

Region: ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region

Option: ○ Match

Ignore heteroplasmic insertions in the poly C-stretches ☑

Maximum number of mismatched nucleotides 0

● Include the queried nucleotide polymorphism

Expected HG: H Estimated HG: H Comments:

Searching Result

No. of Matched Samples	No. of Target Samples	Match Probability
7	593	0.0151

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
K009	M10b	M10b	16066 16223 16311	73 263 315.1C	489 573.1C 573.2
K160	M10b	M10b	16066 16223 16311 16519	73 179 263 315.1C	489 573.1C 573.2
K183	M10b	M10b	16066 16093 16166 16223 16311	73 263 309.1C 315.1C	489 524.1A 524.2
K443	M10b	M10b	16066 16223 16311 16519	73 263 315.1C	489 513 573.1C 5
K508	M10b	M10b	16066 16223 16311	73 26	573.2
K528	M10b	M10b	16066 16223 16311	7	
K558	M10b	M10b	16066 16223 16311	73 2	573.2

Demo-20			16066 16223

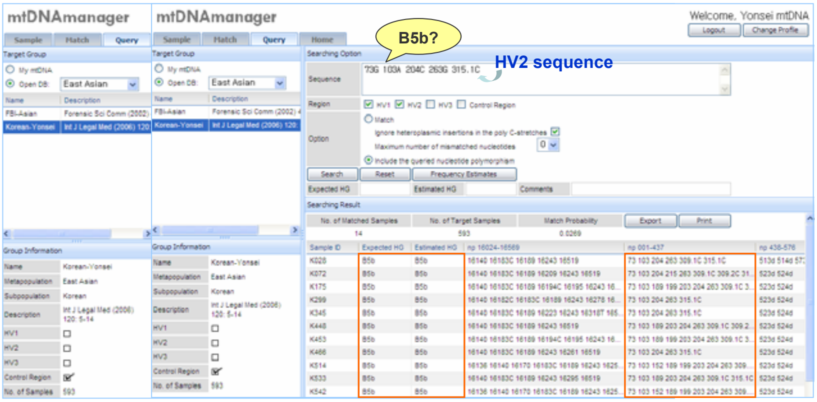
Detection of errors in mtDNA sequence

16311 missed out M10b?

B5b?

Demo-20			16066 16223	73 103 204 263 315.1C	489

Artificial recombination?



B5b?

HV2 sequence

Target Group

○ My mtDNA

● Open DB: East Asian

Name	Description
FBI-Asian	Forensic Sci Comm (2002) 4
Korean-Yonsei	Int J Legal Med (2006) 120:

Group Information

Name	Korean-Yonsei
Metapopulation	East Asian
Subpopulation	Korean
Description	Int J Legal Med (2006) 120: 5-14
HV1	☐
HV2	☐
HV3	☐
Control Region	☑
No. of Samples	593

Searching Option

Sequence: 73G 103A 204C 263G 315.1C

Region: ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region

Option: ○ Match

Ignore heteroplasmic insertions in the poly C-stretches ☑

Maximum number of mismatched nucleotides 0

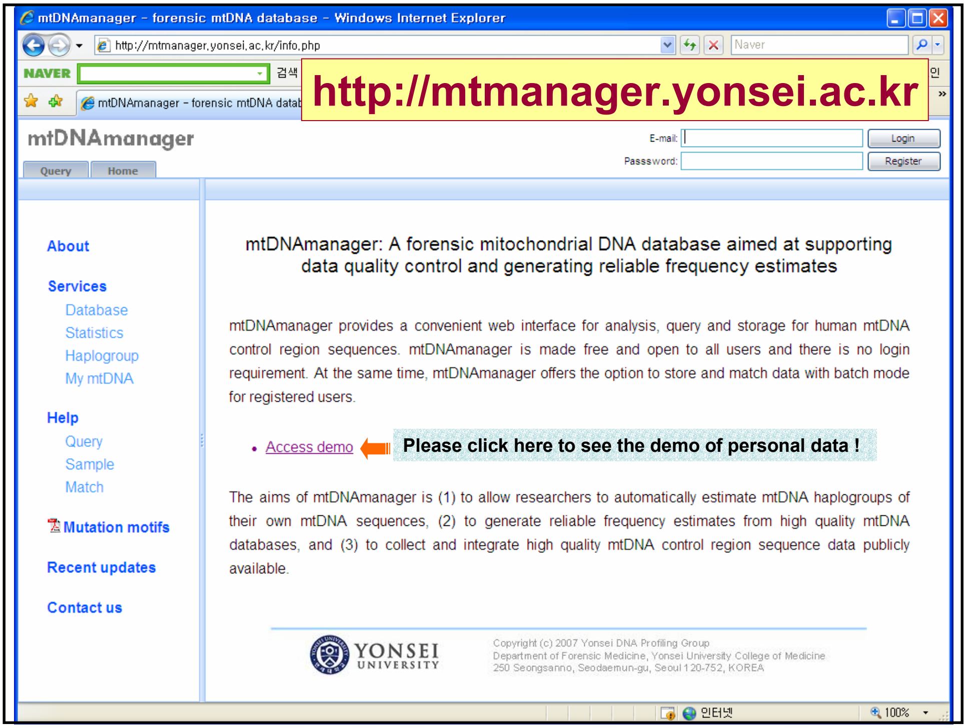
● Include the queried nucleotide polymorphism

Expected HG: Estimated HG: Comments:

Searching Result

No. of Matched Samples	No. of Target Samples	Match Probability
14	593	0.0269

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
K028	B5b	B5b	16140 16183C 16189 16243 16519	73 103 204 263 309.1C 315.1C	513d 514d 57
K072	B5b	B5b	16140 16183C 16189 16209 16243 16519	73 103 204 215 263 309.1C 309.2C 31...	523d 524d
K175	B5b	B5b	16140 16183C 16189 16194C 16195 16243 16...	73 103 189 199 203 204 263 309.1C 3...	523d 524d
K299	B5b	B5b	16140 16182C 16183C 16189 16243 16278 16...	73 103 204 263 315.1C	523d 524d
K345	B5b	B5b	16140 16183C 16189 16223 16243 16318T 165...	73 103 204 263 315.1C	523d 524d
K448	B5b	B5b	16140 16183C 16189 16243 16519	73 103 189 203 204 263 309.1C 309.2...	523d 524d
K453	B5b	B5b	16140 16183C 16189 16194C 16195 16243 16...	73 103 189 199 203 204 263 309.1C 3...	523d 524d
K466	B5b	B5b	16140 16183C 16189 16243 16261 16519	73 103 204 263 315.1C	523d 524d
K514	B5b	B5b	16136 16140 16170 16183C 16189 16243 1625...	73 103 152 189 199 203 204 263 309...	523d 524d
K533	B5b	B5b	16140 16183C 16189 16243 16295 16519	73 103 189 203 204 263 309.1C 315.1C	523d 524d
K542	B5b	B5b	16136 16140 16170 16183C 16189 16243 1625...	73 103 152 189 199 203 204 263 309...	523d 524d



http://mtmanager.yonsei.ac.kr

- **About**
- **Services**
  - Database
  - Statistics
  - Haplogroup
  - My mtDNA
- **Help**
  - Query
  - Sample
  - Match
- **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.

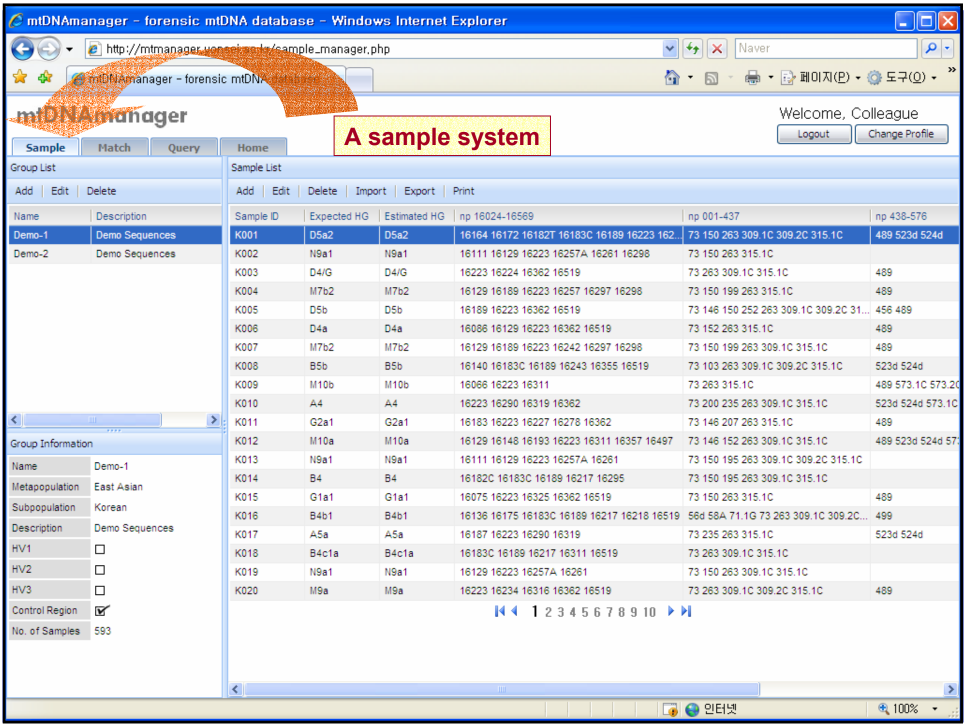
- Access demo

Please click here to see the demo of personal data !

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



A sample system

Welcome, Colleague

Group List

Add | Edit | Delete

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

Group Information

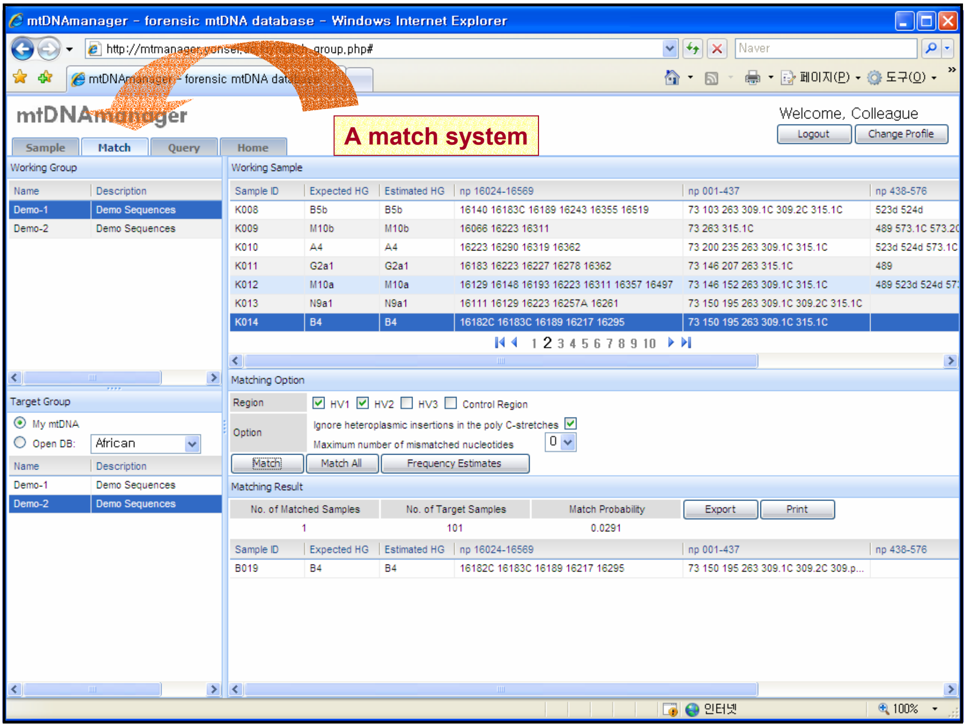
Name	Demo-1
Metapopulation	East Asian
Subpopulation	Korean
Description	Demo Sequences
HV1	☐
HV2	☐
HV3	☐
Control Region	☑
No. of Samples	593

Sample List

Add | Edit | Delete | Import | Export | Print

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
K001	D5a2	D5a2	16164 16172 16182T 16183C 16189 16223 162...	73 150 263 309.1C 309.2C 315.1C	489 523d 524d
K002	N9a1	N9a1	16111 16129 16223 16257A 16261 16298	73 150 263 315.1C	
K003	D4/G	D4/G	16223 16224 16362 16519	73 263 309.1C 315.1C	489
K004	M7b2	M7b2	16129 16189 16223 16257 16297 16298	73 150 199 263 315.1C	489
K005	D5b	D5b	16189 16223 16362 16519	73 146 150 252 263 309.1C 309.2C 31...	456 489
K006	D4a	D4a	16086 16129 16223 16362 16519	73 152 263 315.1C	489
K007	M7b2	M7b2	16129 16189 16223 16242 16297 16298	73 150 199 263 309.1C 315.1C	489
K008	B5b	B5b	16140 16183C 16189 16243 16355 16519	73 103 263 309.1C 309.2C 315.1C	523d 524d
K009	M10b	M10b	16066 16223 16311	73 263 315.1C	489 573.1C 573.2
K010	A4	A4	16223 16290 16319 16362	73 200 235 263 309.1C 315.1C	523d 524d 573.1C
K011	G2a1	G2a1	16183 16223 16227 16278 16362	73 146 207 263 315.1C	489
K012	M10a	M10a	16129 16148 16193 16223 16311 16357 16497	73 146 152 263 309.1C 315.1C	489 523d 524d 57
K013	N9a1	N9a1	16111 16129 16223 16257A 16261	73 150 195 263 309.1C 309.2C 315.1C	
K014	B4	B4	16182C 16183C 16189 16217 16295	73 150 195 263 309.1C 315.1C	
K015	G1a1	G1a1	16075 16223 16325 16362 16519	73 150 263 315.1C	489
K016	B4b1	B4b1	16136 16175 16183C 16189 16217 16218 16519	56d 58A 71.1G 73 263 309.1C 309.2C...	499
K017	A5a	A5a	16187 16223 16290 16319	73 235 263 315.1C	523d 524d
K018	B4c1a	B4c1a	16183C 16189 16217 16311 16519	73 263 309.1C 315.1C	
K019	N9a1	N9a1	16129 16223 16257A 16261	73 150 263 309.1C 315.1C	
K020	M9a	M9a	16223 16234 16316 16362 16519	73 263 309.1C 309.2C 315.1C	489

1 2 3 4 5 6 7 8 9 10


A match system

Working Group

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

Working Sample

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
K008	B5b	B5b	16140 16183C 16189 16243 16355 16519	73 103 263 309.1C 309.2C 315.1C	523d 524d
K009	M10b	M10b	16066 16223 16311	73 263 315.1C	489 573.1C 573.2C
K010	A4	A4	16223 16290 16319 16362	73 200 235 263 309.1C 315.1C	523d 524d 573.1C
K011	G2a1	G2a1	16183 16223 16227 16278 16362	73 146 207 263 315.1C	489
K012	M10a	M10a	16129 16148 16193 16223 16311 16357 16497	73 146 152 263 309.1C 315.1C	489 523d 524d 57
K013	N9a1	N9a1	16111 16129 16223 16257A 16261	73 150 195 263 309.1C 309.2C 315.1C	
K014	B4	B4	16182C 16183C 16189 16217 16295	73 150 195 263 309.1C 315.1C	

1 2 3 4 5 6 7 8 9 10

Target Group

● My mtDNA
○ Open DB: African

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

Matching Option

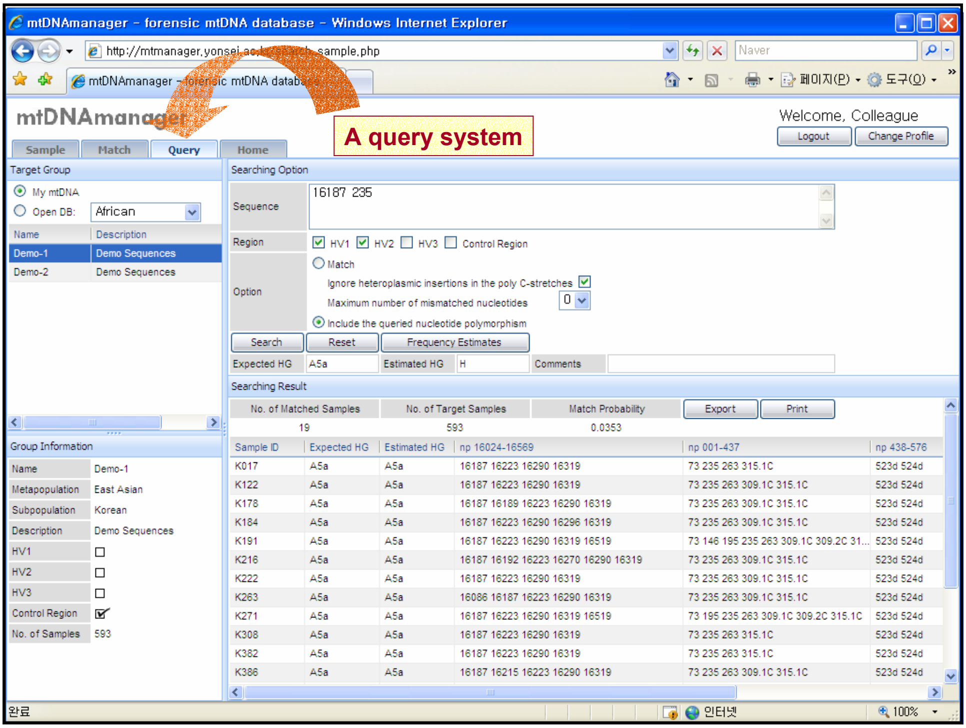
Region: ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region

Option: Ignore heteroplasmic insertions in the poly C-stretches ☑
Maximum number of mismatched nucleotides 0

Matching Result

No. of Matched Samples	No. of Target Samples	Match Probability
1	101	0.0291

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576
B019	B4	B4	16182C 16183C 16189 16217 16295	73 150 195 263 309.1C 309.2C 309.p...	


A query system

Target Group

● My mtDNA
○ Open DB: African

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

Group Information

Name	Demo-1
Metapopulation	East Asian
Subpopulation	Korean
Description	Demo Sequences
HV1	☐
HV2	☐
HV3	☐
Control Region	☑
No. of Samples	593

Searching Option

Sequence: 16187 235

Region: ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region

Option:
○ Match
Ignore heteroplasmic insertions in the poly C-stretches ☑
Maximum number of mismatched nucleotides 0
● Include the queried nucleotide polymorphism

Expected HG: A5a Estimated HG: H Comments:

Searching Result

No. of Matched Samples	No. of Target Samples	Match Probability
19	593	0.0353

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
K178	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

Killed in action recovery and identification

학도병 전사자 유해발굴 시작

경남 하동군 화개장터 일대… 출신학교 재학생 8명 동참

6·25 전사자 유해발굴 작업이 시작된 이후 처음으로 학도병 유해발굴이 이뤄진다.

국방부 유해발굴감식단(단장 박신한 육군대령)은 23일 "6·25 당시 학도병 전사자 20~30명이 집단 가매장된 것으로 추정되는 경남 하동군 화개면 탑리 화개장터 일대에서 유해발굴 작업을 24일부터 26일까지 벌인다"고 밝혔다.

그동안 국군의 6·25 전사자 유해발굴 도중 학도병의 것으로 추정되는 유품이 발견된 적은 있지만 학도병 유해라는 징후를 갖고 발굴에 착수한 것은 이번이 처음이다.

이 일대는 북한군이 파죽지세로 남하하던 1950년 7월 13일 전남 여수·순천 일대 고교생 180여 명이 당시 여수에 주둔하던 5사단 15연대에 자원 입대, 학도병으로는 처음으로 북한 정규군(6사단)과 전투를 벌인 곳이다.

16~18세의 어린 나이에 위기에 처한 나라를 구하기 위해 최소한의 군사훈련만 받고 군복도 입지 못한 채 전장에 투입된 이들은 전차·박격포 등으로 무장한 북한군에 소총으로 맞서다 30여 명이 전사했다.

학도병의 시신은 당시 마을 주민들에 의해 인근 지역에 가매장된 것으로 전해진다.

24일 오전 10시부터 하동군 화개면 탑리 화개면사무소 뒤편 야산에서 실시되는 유해발굴 작업에는 6·25 당시 학도병으로 참전한 여수고·여수공고·순천고·순천매산고 재학생 대표 8명이 선배들의 유해 발굴에 동참한다. 57년의 세월을 넘어 유명을 달리한 학도병 선배와 후배들 간의 만남이 이뤄지는 것이다.

유해발굴감식단은 이번에 발굴되는 유해에 대해 DNA 검사를 통해 유족 확인 절차를 거쳐 국립현충원 등에 안장할 계획이다.

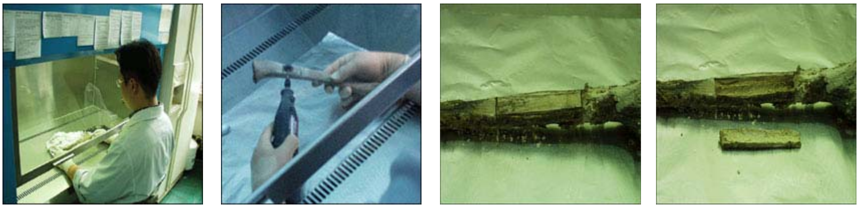
지난 1월 기존 육군본부 산하에 설치된 군 유해발굴 조직이 국방부 조직으로 확대·재창설된 유해발굴감식단은 3월 5일부터 올해 유해발굴 작업을 재개, 지난해 같은 기간보다 네 배나 많은 163구의 전사자 유해를 발굴했다.

유해발굴감식단은 또 24일부터 내달 20일까지 경남 창녕과 경기 연천지역에서 미국 유해발굴전문부대인 미 합동전쟁포로·실종자확인사령부(JPAC)와 함께 한미 합동 유해발굴을 실시하는 한편 터키군 전사자 유해발굴을 위한 탐사활동도 병행할 예정이다.

김가영 기자

Excavation on the finding spot

DNA extraction from skeletal remains

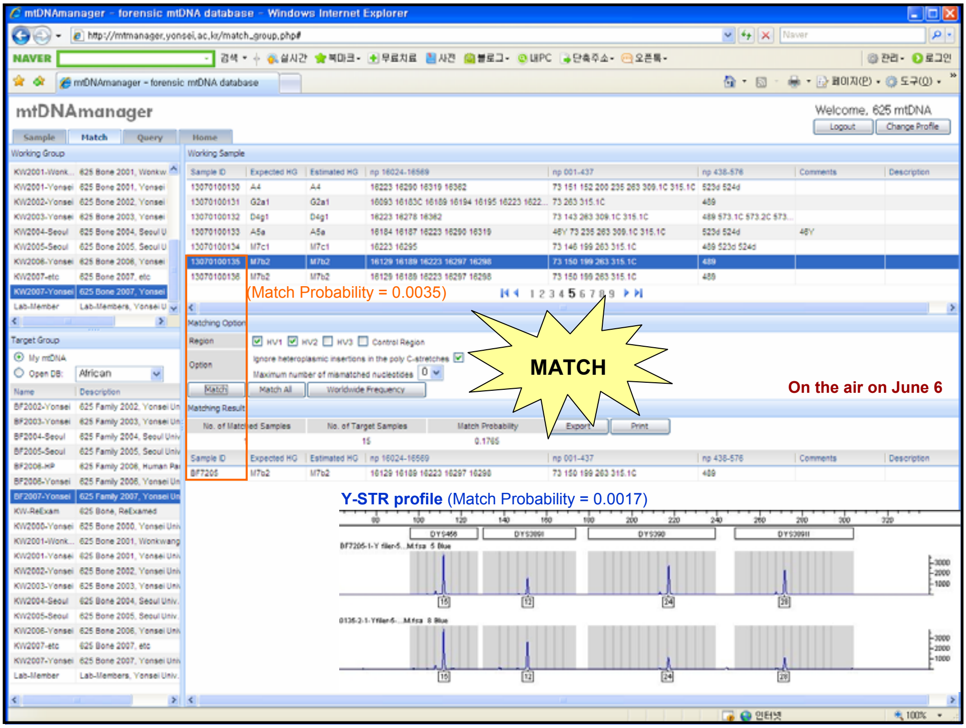


DNA extraction using Qiagen kit

Cryogenic grinding with Spex 6750 Freezer/Mill

한국법과학회
KOREAN SOCIETY OF FORENSIC SCIENCE

YONSEI UNIVERSITY
COLLEGE OF MEDICINE



mtDNAmanager

Welcome, 625 mtDNA

Working Group

KW2001-Wonk...	625 Bone 2001, Wonkw
KW2001-Yonsei	625 Bone 2001, Yonsei
KW2002-Yonsei	625 Bone 2002, Yonsei
KW2003-Yonsei	625 Bone 2003, Yonsei
KW2004-Seoul	625 Bone 2004, Seoul U
KW2005-Seoul	625 Bone 2005, Seoul U
KW2006-Yonsei	625 Bone 2006, Yonsei
KW2007-etc	625 Bone 2007, etc
KW2007-Yonsei	625 Bone 2007, Yonsei
Lab-Member	Lab-Members, Yonsei U

Target Group

● My mtDNA
○ Open DB: African

Name	Description
BF2002-Yonsei	625 Family 2002, Yonsei Un
BF2003-Yonsei	625 Family 2003, Yonsei Un
BF2004-Seoul	625 Family 2004, Seoul Univ
BF2005-Seoul	625 Family 2005, Seoul Univ
BF2006-HP	625 Family 2006, Human Pa
BF2006-Yonsei	625 Family 2006, Yonsei Un
BF2007-Yonsei	625 Family 2007, Yonsei Un
KW-ReExam	625 Bone, ReExamed
KW2000-Yonsei	625 Bone 2000, Yonsei Univ
KW2001-Wonk...	625 Bone 2001, Wonkwang
KW2001-Yonsei	625 Bone 2001, Yonsei Univ
KW2002-Yonsei	625 Bone 2002, Yonsei Univ
KW2003-Yonsei	625 Bone 2003, Yonsei Univ
KW2004-Seoul	625 Bone 2004, Seoul Univ.
KW2005-Seoul	625 Bone 2005, Seoul Univ.
KW2006-Yonsei	625 Bone 2006, Yonsei Univ
KW2007-etc	625 Bone 2007, etc
KW2007-Yonsei	625 Bone 2007, Yonsei Univ
Lab-Member	Lab-Members, Yonsei Univ.

Working Sample

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576	Comments	Description
13070100130	A4	A4	16223 16290 16319 16362	73 151 152 200 235 263 309.1C 315.1C	523d 524d		
13070100131	G2a1	G2a1	16093 16183C 16189 16194 16195 16223 1622...	73 263 315.1C	489		
13070100132	D4g1	D4g1	16223 16278 16362	73 143 263 309.1C 315.1C	489 573.1C 573.2C 573...		
13070100133	A5a	A5a	16184 16187 16223 16290 16319	48Y 73 235 263 309.1C 315.1C	523d 524d	48Y	
13070100134	M7c1	M7c1	16223 16295	73 146 199 263 315.1C	489 523d 524d		
13070100135	M7b2	M7b2	16129 16189 16223 16297 16298	73 150 199 263 315.1C	489		
13070100136	M7b2	M7b2	16129 16189 16223 16297 16298	73 150 199 263 315.1C	489		

(Match Probability = 0.0035)

Matching Option

Region: ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region

Option: Ignore heteroplasmic insertions in the poly C-stretches ☑

Maximum number of mismatched nucleotides 0

MATCH

On the air on June 6

Matching Result

No. of Matched Samples	No. of Target Samples	Match Probability
	15	0.1765

Sample ID	Expected HG	Estimated HG	np 16024-16569	np 001-437	np 438-576	Comments	Description
BF7205	M7b2	M7b2	16129 16189 16223 16297 16298	73 150 199 263 315.1C	489		

Y-STR profile (Match Probability = 0.0017)

Concluding remarks

- The mtDNAmanager is a forensic mitochondrial DNA database aimed at **supporting data quality control** and **generating reliable frequency estimates** using an approach based on haplogroup estimation and data comparison.
- The **mtDNAmanager** provides **a convenient Web interface** for analysis, query and storage of human mtDNA control region sequences.

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