# mtDNAmanager:

## Updates for generating high-quality mitochondrial DNA data from a phylogenetic perspective

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

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

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

---

# Outline

- The first release of mtDNAmanager in 2007
  - Background
  - Aims
  - Implementation
- Recent updates
  - Database expansion
  - Refinement of control region mutation motifs

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

## 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)

---

## 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**)

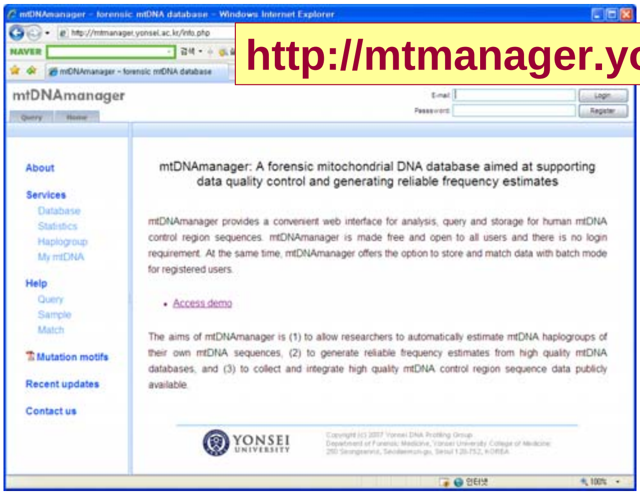
*Bandelt et al. IJLM (2001) 115:64-9*

# 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



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.

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

# Worldwide access to mtDNAmanager

지도 중첩

2007. 8. 23. - 2009. 11. 22.

75개 국가/지역을(를) 통해 13,296회의 방문이 이루어졌습니다.

| 방문수 | 페이지수/방문수 | 평균 사이트 방문 시간 | 신규 방문수 비율(%) | 반송률 |
|---|---|---|---|---|
| 13,296 | 2.93 | 00:03:49 | 28.79% | 41.42% |

| | 국가/지역 | 방문수 | 페이지수/방문수 | 평균 사이트 방문 시간 | 신규 방문수 비율(%) | 반송률 |
|---|---|---|---|---|---|---|
| 1. | South Korea | 5,947 | 3.41 | 00:04:07 | 7.08% | 29.88% |
| 2. | United States | 1,314 | 2.67 | 00:03:16 | 68.72% | 53.81% |
| 3. | Spain | 951 | 2.29 | 00:03:41 | 20.72% | 47.00% |
| 4. | Germany | 520 | 2.11 | 00:02:49 | 47.88% | 58.46% |
| 5. | China | 443 | 2.62 | 00:03:12 | 44.92% | 47.40% |

Since its first release in 2007, the mtDNAmanager has been accessed by users from 75 nations more than 13,000 times

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

# Monthly access to mtDNAmanager

대시보드

2009. 10. 24. - 2009. 11. 23.

사이트 용도

823 방문수 Number of access

2,518 페이지뷰

3.06 페이지수/방문수

40.34% 반송률

00:04:27 평균 사이트 방문 시간

33.78% 신규 방문수 비율(%) Rate of newcomers

: Revisit 66.22%

방문자 개요

351 방문자 Number of visitors

리포트 보기

지도 중첩

리포트 보기

트래픽 소스 개요

리포트 보기

콘텐츠 개요

| 페이지 | 페이지뷰 | 페이지뷰 비율(%) |
|---|---|---|
| /sample_manager.php | 744 | 29.55% |
| / | 663 | 26.33% |
| /index.php | 481 | 19.10% |
| /search_sample.php | 441 | 17.51% |
| /match_group.php | 157 | 6.24% |

리포트 보기

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

# 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

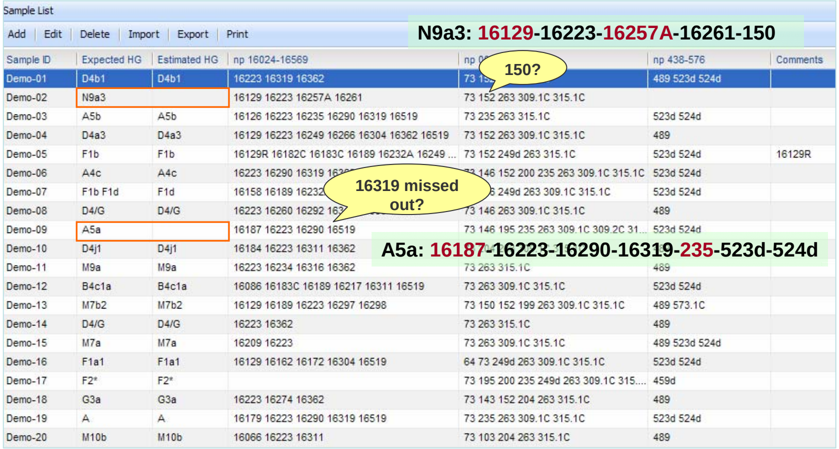
---

# 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.



**N9a3: 16129-16223-16257A-16261-150**

150?

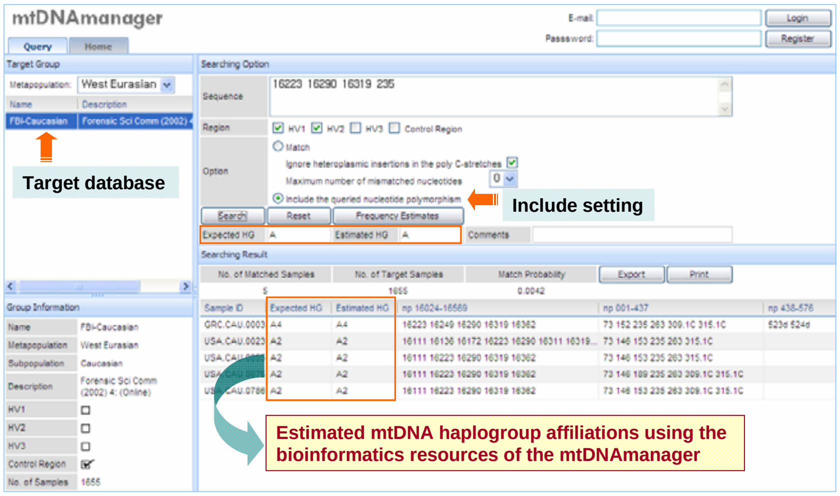
16319 missed out?

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

| Sample ID | Expected HG | Estimated HG | np 16024-16569 | np 0 | np 438-576 | Comments |
|---|---|---|---|---|---|---|
| Demo-01 | D4b1 | D4b1 | 16223 16319 16362 | 73 15 | 489 523d 524d | |
| Demo-02 | N9a3 | | 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 | A4c | A4c | 16223 16290 16319 16 | 73 146 152 200 235 263 309.1C 315.1C | 523d 524d | |
| Demo-07 | F1b F1d | F1d | 16158 16189 16232 | 6 249d 263 309.1C 315.1C | 523d 524d | |
| Demo-08 | D4/G | D4/G | 16223 16260 16292 163 | 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 | | | |
| 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.... | 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 | 16066 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 database or the entire population group of its open 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 setting**

Expected HG: A
Estimated HG: A
Comments:

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

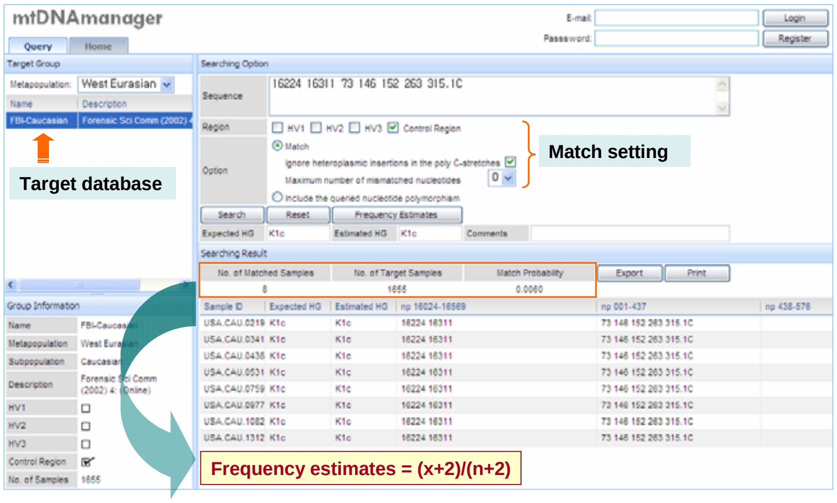
| Sample ID | Expected HG | Estimated HG | np 16024-16569 | np 001-437 | np 438-576 |
|---|---|---|---|---|---|
| GRC.CAU.0003 | A4 | A4 | 16223 16249 16290 16319 16362 | 73 152 235 263 309.1C 315.1C | 523d 524d |
| USA.CAU.0023 | A2 | A2 | 16111 16136 16172 16223 16290 16311 16319... | 73 146 153 235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

| | |
|---|---|
| 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 using query system

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



Target Group

Metapopulation: West Eurasian

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

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 |

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

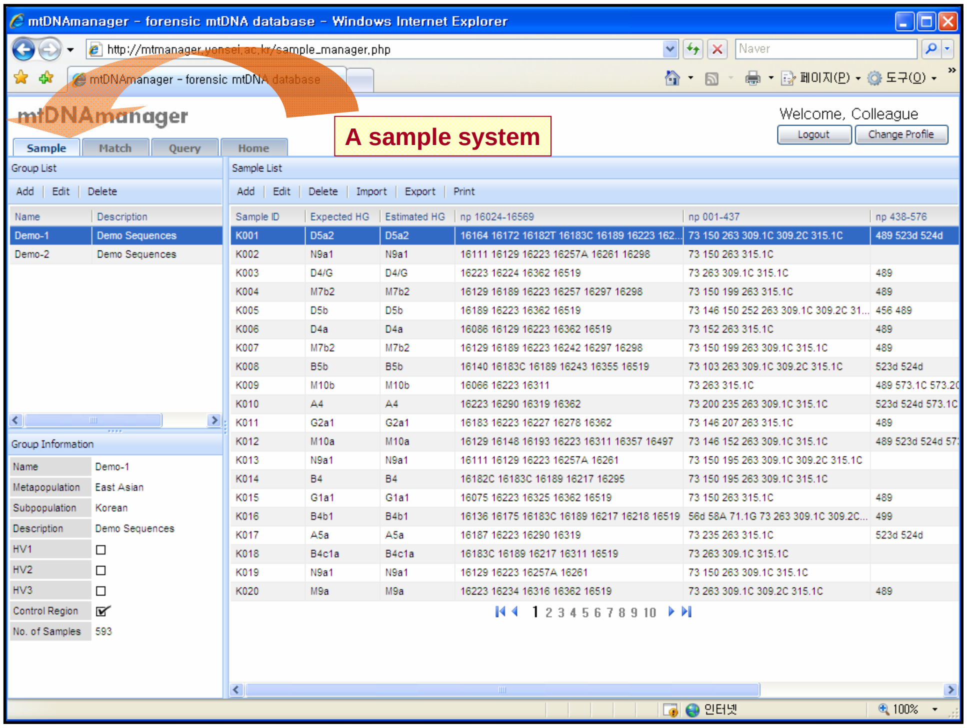
| Expected HG | K1c | Estimated HG | K1c | Comments | |
|---|---|---|---|---|---|

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 | |

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



Group List

Add | Edit | Delete

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

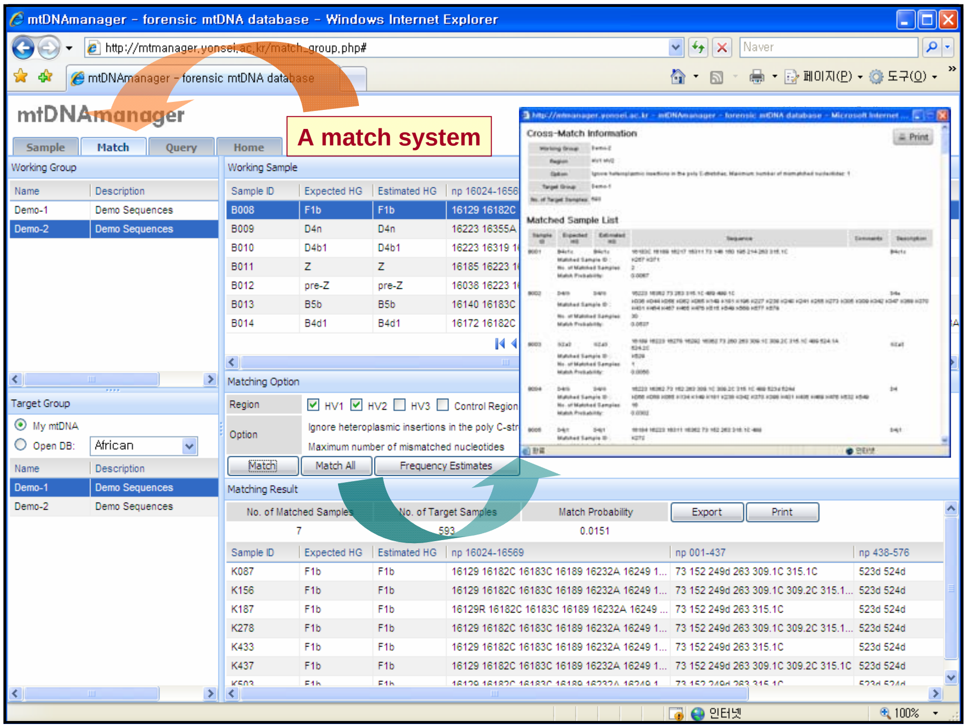
Group Information

| | |
|---|---|
| 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 |



**A match system**

Working Group

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

Target Group

● My mtDNA
○ Open DB: African

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

Working Sample

| Sample ID | Expected HG | Estimated HG | np 16024-1656 |
|---|---|---|---|
| B008 | F1b | F1b | 16129 16182C |
| B009 | D4n | D4n | 16223 16355A |
| B010 | D4b1 | D4b1 | 16223 16319 1 |
| B011 | Z | Z | 16185 16223 1 |
| B012 | pre-Z | pre-Z | 16038 16223 1 |
| B013 | B5b | B5b | 16140 16183C |
| B014 | B4d1 | B4d1 | 16172 16182C |

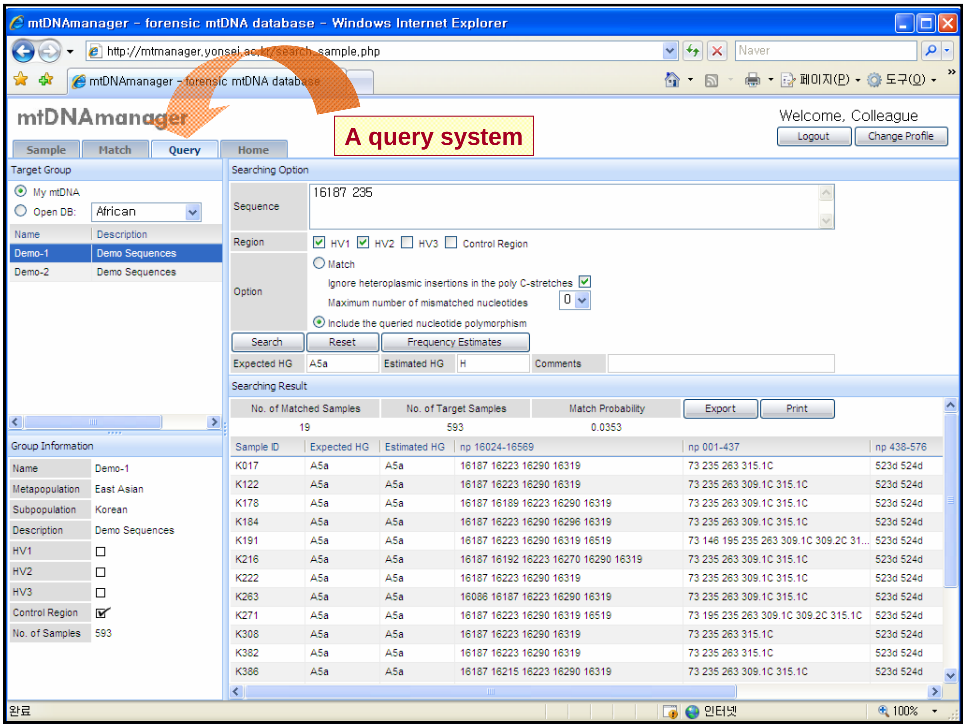
Matching Option

| | |
|---|---|
| Region | ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region |
| Option | Ignore heteroplasmic insertions in the poly C-str<br>Maximum number of mismatched nucleotides |

Matching 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 |
|---|---|---|---|---|---|
| K087 | F1b | F1b | 16129 16182C 16183C 16189 16232A 16249 1... | 73 152 249d 263 309.1C 315.1C | 523d 524d |
| K156 | F1b | F1b | 16129 16182C 16183C 16189 16232A 16249 1... | 73 152 249d 263 309.1C 309.2C 315.1... | 523d 524d |
| K187 | F1b | F1b | 16129R 16182C 16183C 16189 16232A 16249 ... | 73 152 249d 263 315.1C | 523d 524d |
| K278 | F1b | F1b | 16129 16182C 16183C 16189 16232A 16249 1... | 73 152 249d 263 309.1C 309.2C 315.1... | 523d 524d |
| K433 | F1b | F1b | 16129 16182C 16183C 16189 16232A 16249 1... | 73 152 249d 263 315.1C | 523d 524d |
| K437 | F1b | F1b | 16129 16182C 16183C 16189 16232A 16249 1... | 73 152 249d 263 309.1C 309.2C 315.1C | 523d 524d |



**A query system**

Welcome, Colleague

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

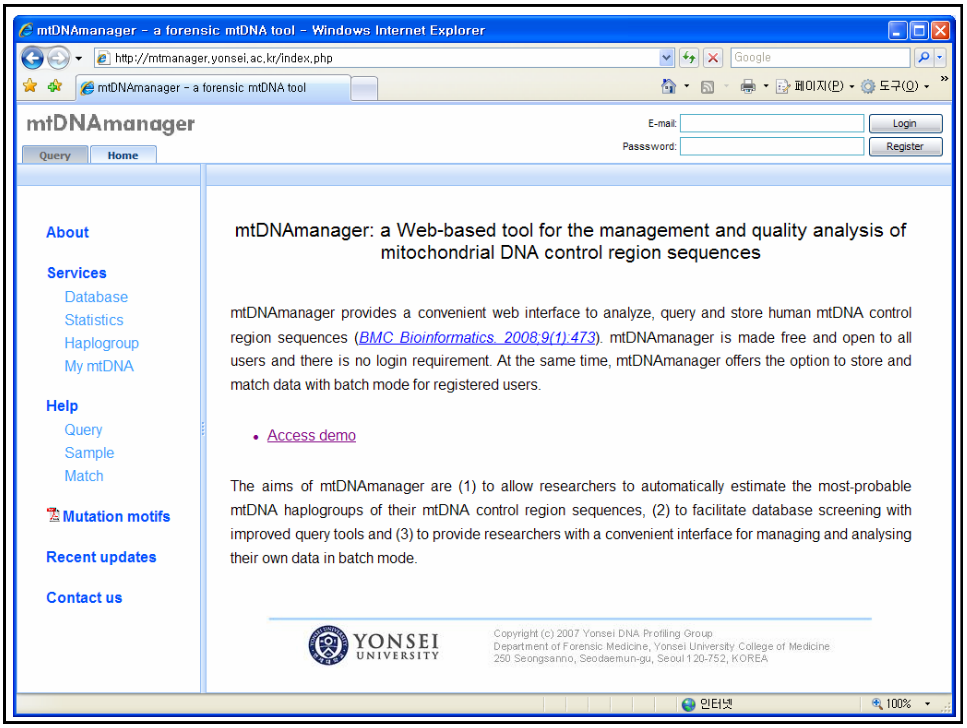
| | |
|---|---|
| Sequence | 16187 235 |
| Region | ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region |
| Option | ○ Match<br>Ignore heteroplasmic insertions in the poly C-stretches ☑<br>Maximum number of mismatched nucleotides 0<br>● Include the queried nucleotide polymorphism |

| 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 |



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.

Department of Forensic Medicine, Yonsei University College of Medicine
250 Seongsanno, Seodaemun-gu, Seoul 120-752, KOREA

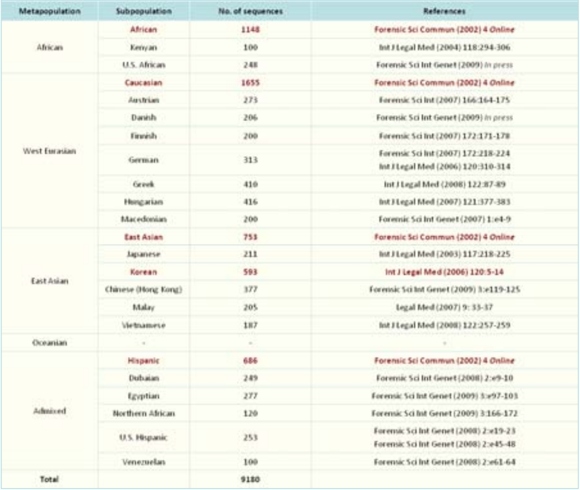
## Updates from the first release

- **The number of mtDNA control region sequences** in the mtDNAmanager's open database has grown **from 5,432 to 9,180**, 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 590**.

YONSEI UNIVERSITY
COLLEGE OF MEDICINE

# mtDNAmanager’s current open database

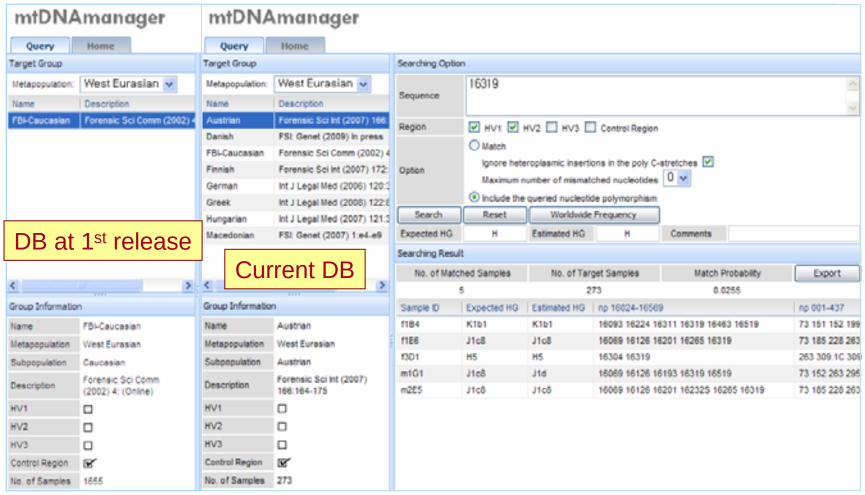


| Metapopulation | Subpopulation | No. of sequences | References |
|---|---|---|---|
| African | **African** | **1148** | **Forensic Sci Commun (2002) 4 *Online*** |
| | Kenyan | 100 | Int J Legal Med (2004) 118:294-306 |
| | U.S. African | 248 | Forensic Sci Int Genet (2009) *In press* |
| West Eurasian | **Caucasian** | **1655** | **Forensic Sci Commun (2002) 4 *Online*** |
| | Austrian | 273 | Forensic Sci Int (2007) 166:164-175 |
| | Danish | 206 | Forensic Sci Int Genet (2009) *In press* |
| | Finnish | 200 | Forensic Sci Int (2007) 172:171-178 |
| | German | 313 | Forensic Sci Int (2007) 172:218-224<br>Int J Legal Med (2006) 120:310-314 |
| | Greek | 410 | Int J Legal Med (2008) 122:87-89 |
| | Hungarian | 416 | Int J Legal Med (2007) 121:377-383 |
| | Macedonian | 200 | Forensic Sci Int Genet (2007) 1:e4-9 |
| 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** |
| | Chinese (Hong Kong) | 377 | Forensic Sci Int Genet (2009) 3:e119-125 |
| | Malay | 205 | Legal Med (2007) 9: 33-37 |
| | Vietnamese | 187 | Int J Legal Med (2008) 122:257-259 |
| Oceanian | - | - | - |
| Admixed | **Hispanic** | **686** | **Forensic Sci Commun (2002) 4 *Online*** |
| | Dubaian | 249 | Forensic Sci Int Genet (2008) 2:e9-10 |
| | Egyptian | 277 | Forensic Sci Int Genet (2009) 3:e97-103 |
| | Northern African | 120 | Forensic Sci Int Genet (2009) 3:166-172 |
| | U.S. Hispanic | 253 | Forensic Sci Int Genet (2008) 2:e19-23<br>Forensic Sci Int Genet (2008) 2:e45-48 |
| | Venezuelan | 100 | Forensic Sci Int Genet (2008) 2:e61-64 |
| **Total** | | **9180** | |

**20 more publications were added**

---

# Expansion of mtDNAmanager’s database



Metapopulation: West Eurasian

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

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

DB at 1st release

Metapopulation: West Eurasian

| Name | Description |
|---|---|
| Austrian | Forensic Sci Int (2007) 166: |
| Danish | FSI: Genet (2009) In press |
| FBI-Caucasian | Forensic Sci Comm (2002) 4 |
| Finnish | Forensic Sci Int (2007) 172: |
| German | Int J Legal Med (2006) 120: |
| Greek | Int J Legal Med (2008) 122: |
| Hungarian | Int J Legal Med (2007) 121: |
| Macedonian | FSI: Genet (2007) 1:e4-e9 |

Current DB

| | |
|---|---|
| Name | Austrian |
| Metapopulation | West Eurasian |
| Subpopulation | Austrian |
| Description | Forensic Sci Int (2007) 166:164-175 |
| HV1 | ☐ |
| HV2 | ☐ |
| HV3 | ☐ |
| Control Region | ☑ |
| No. of Samples | 273 |

Sequence: 16319

Region: ☑ HV1 ☑ HV2 ☐ HV3 ☐ Control Region

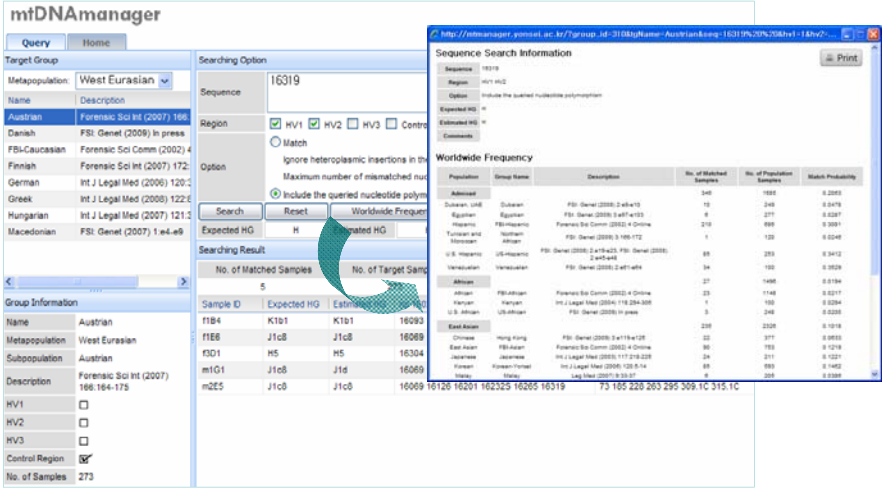
Option:
○ Match
Ignore heteroplasmic insertions in the poly C-stretches ☑
Maximum number of mismatched nucleotides 0
● Include the queried nucleotide polymorphism

Expected HG: H
Estimated HG: H
Comments:

| No. of Matched Samples | No. of Target Samples | Match Probability |
|---|---|---|
| 5 | 273 | 0.0255 |

| Sample ID | Expected HG | Estimated HG | np 16024-16569 | np 001-437 |
|---|---|---|---|---|
| f1B4 | K1b1 | K1b1 | 16093 16224 16311 16319 16463 16519 | 73 151 152 199 |
| f1E6 | J1c8 | J1c8 | 16069 16126 16201 16265 16319 | 73 185 228 263 |
| f3D1 | H5 | H5 | 16304 16319 | 263 309.1C 309 |
| m1G1 | J1c8 | J1d | 16069 16126 16193 16319 16519 | 73 152 263 295 |
| m2E5 | J1c8 | J1c8 | 16069 16126 16201 16232S 16265 16319 | 73 185 228 263 |

# Improved database search



**mtDNAmanager**

**Target Group**

Metapopulation: West Eurasian

| Name | Description |
|---|---|
| Austrian | Forensic Sci Int (2007) 166: |
| Danish | FSI: Genet (2009) In press |
| FBI-Caucasian | Forensic Sci Comm (2002) 4 |
| Finnish | Forensic Sci Int (2007) 172: |
| German | Int J Legal Med (2006) 120: |
| Greek | Int J Legal Med (2008) 122: |
| Hungarian | Int J Legal Med (2007) 121: |
| Macedonian | FSI: Genet (2007) 1:e4-e9 |

**Group Information**

| | |
|---|---|
| Name | Austrian |
| Metapopulation | West Eurasian |
| Subpopulation | Austrian |
| Description | Forensic Sci Int (2007) 166:164-175 |
| HV1 | ☐ |
| HV2 | ☐ |
| HV3 | ☐ |
| Control Region | ☑ |
| No. of Samples | 273 |

**Searching Option**

| | |
|---|---|
| Sequence | 16319 |
| Region | ☑ HV1 ☑ HV2 ☐ HV3 ☐ Contro |
| Option | ○ Match<br>Ignore heteroplasmic insertions in the<br>Maximum number of mismatched nuc<br>● Include the queried nucleotide polym |

| Expected HG | H | Estimated HG | |
|---|---|---|---|

**Searching Result**

| No. of Matched Samples | No. of Target Samp |
|---|---|
| 5 | 273 |

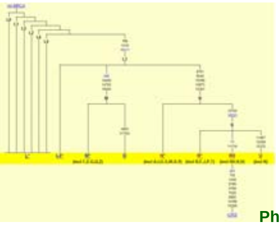
| Sample ID | Expected HG | Estimated HG | no 160 | |
|---|---|---|---|---|
| f1B4 | K1b1 | K1b1 | 16093 | |
| f1E6 | J1c8 | J1c8 | 16069 | |
| f3D1 | H5 | H5 | 16304 | |
| m1G1 | J1c8 | J1d | 16069 | |
| m2E5 | J1c8 | J1c8 | 16069 16126 16201 162325 16265 16319 | 73 185 228 263 295 309.1C 315.1C |

http://mtmanager.yonsei.ac.kr/?group_id=310&tgName=Austrian&seq=16319%20%20&hv1=1&hv2=...

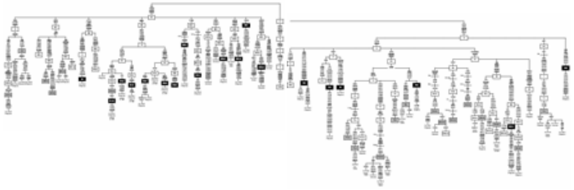
**Sequence Search Information**

**Worldwide Frequency**

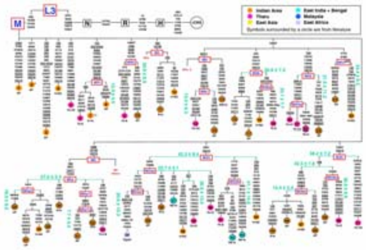
# Recent updates in phylogenetic trees



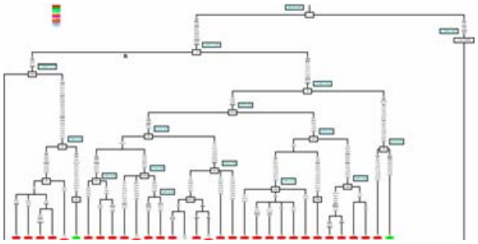
PhyloTree.org



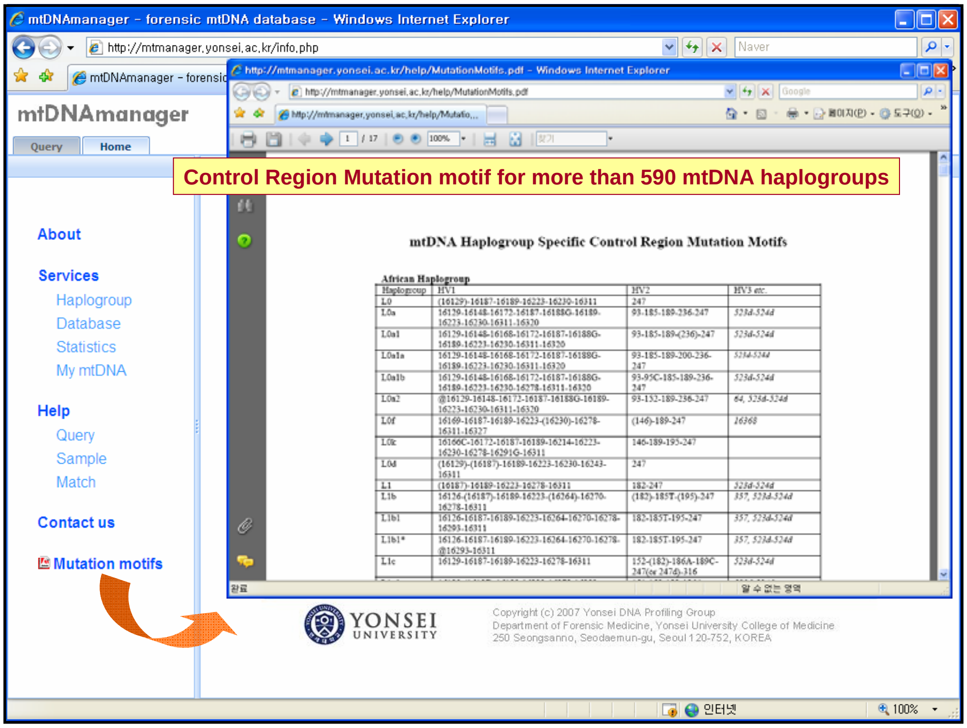
South East Asian: ~M71, M72



South Asian: M31, M33, M51…



African: L0~L5



Control Region Mutation motif for more than 590 mtDNA haplogroups

**mtDNA Haplogroup Specific Control Region Mutation Motifs**

**African Haplogroup**

| Haplogroup | HV1 | HV2 | HV3 *etc.* |
|---|---|---|---|
| L0 | (16129)-16187-16189-16223-16230-16311 | 247 | |
| 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-16189-16223-16230-16311-16320 | 93-185-189-200-236-247 | *523d-524d* |
| L0a1b | 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 | *64, 523d-524d* |
| L0f | 16169-16187-16189-16223-(16230)-16278-16311-16327 | (146)-189-247 | *16368* |
| L0k | 16166C-16172-16187-16189-16214-16223-16230-16278-16291G-16311 | 146-189-195-247 | |
| L0d | (16129)-(16187)-16189-16223-16230-16243-16311 | 247 | |
| L1 | (16187)-16189-16223-16278-16311 | 182-247 | *523d-524d* |
| 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* |
| L1b1* | 16126-16187-16189-16223-16264-16270-16278-@16293-16311 | 182-185T-195-247 | *357, 523d-524d* |
| L1c | 16129-16187-16189-16223-16278-16311 | 152-(182)-186A-189C-247(or 247d)-316 | *523d-524d* |

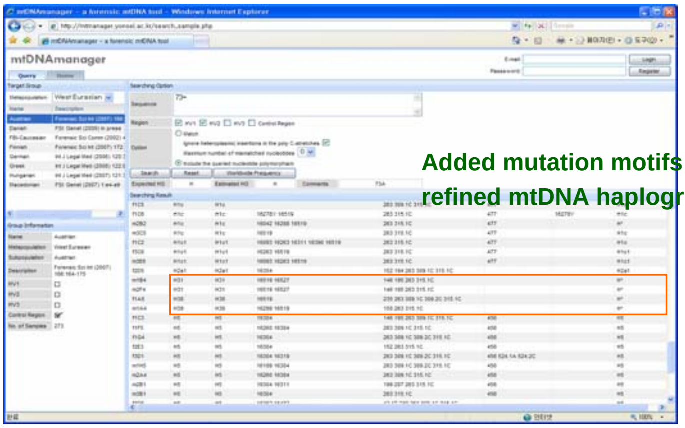
YONSEI UNIVERSITY

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

---

# Improved mtDNA haplogroup estimation

- **Refinement of mtDNA phylogeny** with more diagnostic mutations would facilitate the detection of more errors in mtDNA sequence since it is based on mutation motifs



**Added mutation motifs enables refined mtDNA haplogroup estimation**

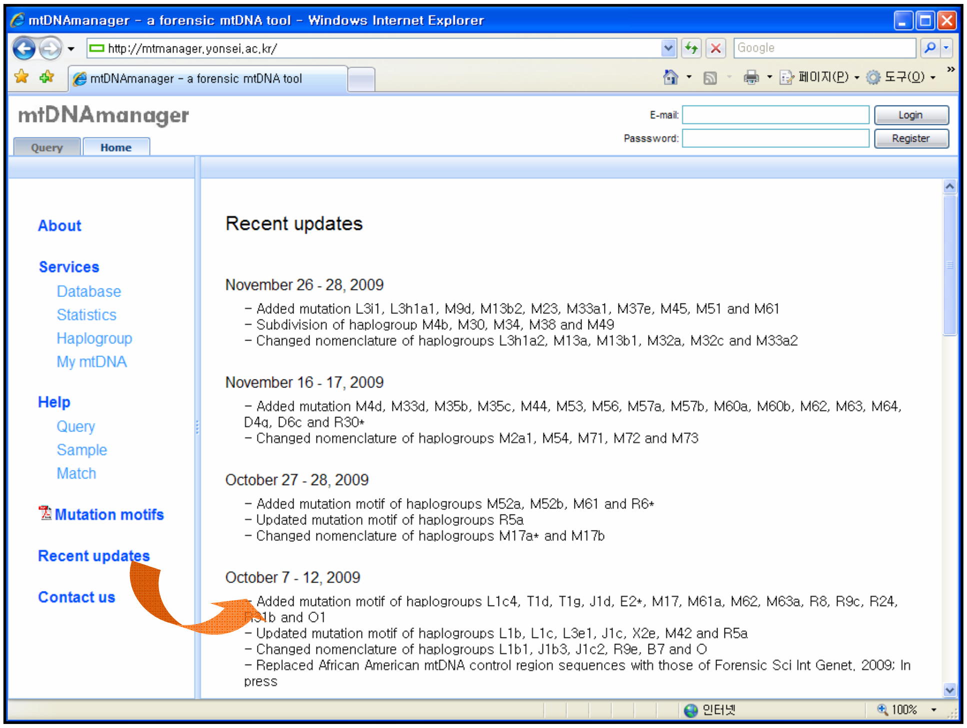
YONSEI UNIVERSITY
COLLEGE OF MEDICINE

# Updates in mtDNAmanager

- **Refinement of mtDNA phylogeny** with more diagnostic mutations would provide better algorithms for automatic estimation of the most-probable mtDNA haplogroups in diverse population groups, and facilitate the detection of more errors in mtDNA sequence data by suggesting more candidate sites for reinvestigation

- **A neighbourhood search for sequences in the expanded open database** would facilitate pinpointing errors through extensive data comparison using the expanded subset of the total database

YONSEI UNIVERSITY
COLLEGE OF MEDICINE



Recent updates

November 26 - 28, 2009

- Added mutation L3i1, L3h1a1, M9d, M13b2, M23, M33a1, M37e, M45, M51 and M61
- Subdivision of haplogroup M4b, M30, M34, M38 and M49
- Changed nomenclature of haplogroups L3h1a2, M13a, M13b1, M32a, M32c and M33a2

November 16 - 17, 2009

- Added mutation M4d, M33d, M35b, M35c, M44, M53, M56, M57a, M57b, M60a, M60b, M62, M63, M64, D4q, D6c and R30*
- Changed nomenclature of haplogroups M2a1, M54, M71, M72 and M73

October 27 - 28, 2009

- Added mutation motif of haplogroups M52a, M52b, M61 and R6*
- Updated mutation motif of haplogroups R5a
- Changed nomenclature of haplogroups M17a* and M17b

October 7 - 12, 2009

- Added mutation motif of haplogroups L1c4, T1d, T1g, J1d, E2*, M17, M61a, M62, M63a, R8, R9c, R24, R31b and O1
- Updated mutation motif of haplogroups L1b, L1c, L3e1, J1c, X2e, M42 and R5a
- Changed nomenclature of haplogroups L1b1, J1b3, J1c2, R9e, B7 and O
- Replaced African American mtDNA control region sequences with those of Forensic Sci Int Genet, 2009; In press

## 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).*