mtDNAmanager: a Web-based Tool for Mitochondrial DNA Haplogroup Estimation and Database Search

Hwan Young Lee · Woo Ick Yang · Kyoung-Jin Shin

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

Introduction

Since mitochondrial DNA (mtDNA) haplogroup is defined by differences in mtDNA sequence and is used to represent the major branch points on the phylogenetic tree, the haplogroup analysis of mtDNA has been widely utilized in various disciplines, including population, medical and forensic genetics. However, manual haplogroup estimation requires a thorough understanding of the worldwide mtDNA phylogeny, and a neighbourhood search to identify a subset of potentially closely related sequences requires high-quality databases that are publicly available. Therefore, we present a Web-based bioinformatics resource, mtDNAmanager (http://mtmanager.yonsei.ac.kr) that enables users to automatically estimate the most-probable mtDNA haplogroups according to control-region mutation motifs and to retrieve similar sequences from the database which includes over 9000 mtDNA control region sequences of worldwide populations.

Implementation

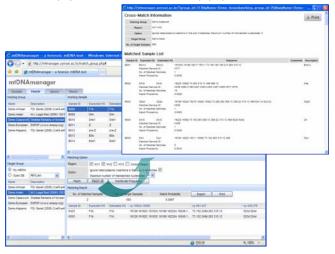### Database structure of mtDNAmanager

The mtDNAmanager Web interfaces were implemented using PHP and Asynchronous JavaScript and XML (AJAX). The mtDNAmanager system is optimized for Internet Explorer version 6.0 or later. mtDNAmanager has a multithreaded and multiuser SQL database management system designed and implemented using MySQL. The mtDNAmanager database currently comprises seven tables containing human mtDNA control-region sequences, data related to the samples and results of haplogroup estimation obtained by the mtDNA haplogroup-estimating resource that runs automatically .

Database of mtDNAmanager

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 / 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 / Forensic Sci Int Genet (2008) 2:e45-48
	Venezuelan	100	Forensic Sci Int Genet (2008) 2:e61-64
Total		**9180**	

Database search using the match system

The match system permits cross-matching of all sequence data between two selected groups as well as retrieval of matched sequences for a sample of the own database of the user or mtDNAmanager's open database. Clicking the "Match All" button will display cross-matched results in a new pop-up window.
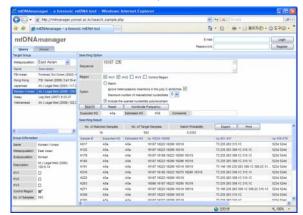

Estimation of the most-probable mtDNA haplogroup

The most-probable haplogroup of a given mtDNA sequence is estimated using a mathematical algorithm based on propositional logic via hierarchical verification of the presence or absence of haplogroup-specific diagnostic mutations. For that purpose, reliable control-region mutation motifs (strings of characteristic/diagnostic mutations shared by descent) for the assignment of more than 500 mtDNA haplogroups and subhaplogroups were identified based on well-characterized mtDNA phylogenies (see the list of mutation motifs at http://mtmanager.yonsei.ac.kr/help/MutationMotifs.pdf).
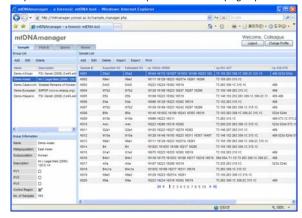
mtDNA haplogroup estimation and database search using the query system

The query system—using the include setting by default—retrieves sequences that include the queried nucleotide polymorphisms from a chosen population or the entire population group of its open database. The results are displayed on the same page that the query was entered, and while displaying retrieved sequences, mtDNAmanager shows frequency estimates for random matches from a selected group and automatically estimated haplogroup affiliations for both submitted data and retrieved sequences.

Data management using the sample system

The sample system allows users to manage and analyse large amounts of their own data in batch mode. Data are characterized whilst being imported by the automatically running haplogroup-estimating workflow, and accordingly, each sample is annotated with the most-probable mtDNA haplogroup.

Conclusion

- mtDNAmanager allows researchers to automatically estimate the most-probable mtDNA haplogroups of their mtDNA control region sequences, facilitates database screening with improved query tools, and provides researchers with a convenient interface for managing and analysing their own data in batch mode.
- mtDNAmanager will help in checking the quality of data and facilitate data comparisons from a phylogenetic perspective by displaying information—estimated haplogroup affiliations and nucleotide polymorphisms—of all sequences on a single page.
- mtDNAmanager could be very useful for population, medical and forensic studies that involve mtDNA analysis.

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