



Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/fsigss

mtDNAmap: Geographic representation of mtDNA Haplogroups

Haroon ur Rasheed^a, Muhammad Jawad^b, Shahid Nazir^b, Saadia Noreen^b, Allah Rakha^{b,*}^a COMSATS Institute of Information Technology, Sahiwal, Pakistan^b Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan

ARTICLE INFO

Keywords:

mtDNA
Phylogeography
mtDNAmap
Haplogroup

ABSTRACT

Mitochondrial DNA is inherited maternally and is thought to be evolved stepwise from one population to another population in the history of mankind. Haplogroup for any mtDNA provides us a solution for the logical classification of the mitochondrial DNA based on established phylogenetic principles. There is a huge amount of scattered mtDNA sequence data from different global and regional populations. It demands a professional platform for representation of data to draw meaningful and simple-to-understand information about mtDNA distribution. Here, mtDNAmap provides geographical representation of mtDNA haplogroups' frequencies in various populations all over the world according to their present day reported locations. It is a haplogroup frequency database of different populations calculated from the published data using their reported valid mtDNA sequences. Publicly available MtDNA sequences, processed through mtDNAprofiler for SNP determinations based on revised Cambridge Reference Sequence and followed by Haplogrep 2.0 for the determination of the haplogroups on the basis of most updated Phylotree version-17, are graphically represented on the dynamic map in the form of frequencies. mtDNAmap provides the open access to the whole or part of published high-quality curated data. The tool is not only useful for researchers from forensic and anthropology backgrounds but also in general public.

1. Introduction

Human Mitochondria have several thousand copies of specific deoxyribonucleic acid known as mitochondrial DNA (mtDNA). Human mtDNA is a 16,569 bp double-stranded circular molecule that contains approximately 1121-bp non-coding region referred to as the control region or D-loop region, which harbors a series of regulatory segments for transcription and replication and remaining mtDNA encodes 37 genes, including 12S rRNA, 16S rRNA, 22 tRNA and 13 proteins that are critical for oxidative phosphorylation (OXPHOS) [1]. Being the maternally inherent and absence of recombination, the molecule conserved the same pattern of common ancestry and is geographically localized. A relatively higher rate of mutation in mtDNA (nearly 482 polymorphisms) [2] and along with other prominent properties like maternal inheritance and no recombination makes this molecule a vital element in the studies of population genetics, forensics and medical genetics [3]. mtDNA is thought to be evolved stepwise from one population to another population in the history of mankind and the use of mtDNA genealogies and relevant geographic information helps in understanding the genetic structure of the sampled population [4]. mtDNA mutations over time had classified the human population into several discrete haplogroups. The major haplogroups arose 40,000–150,000

years before present (YBP) and have defined different human populations as they migrated out of Africa and populated the globe [5]. The recent advancements in the study of mtDNA have equipped us with a complete automated system for the study of classification of the specific mtDNA in the form of haplogroups [6]. Haplogroup for any mtDNA provides us a solution for logical classification of mitochondrial DNA based on established phylogenetic principles. There is a huge amount of scattered mtDNA sequence data from different global and regional populations. It demands a professional platform for representation of data to draw meaningful and simple-to-understand information about mtDNA distribution. Here, mtDNAmap provides geographical representation of mtDNA haplogroups' frequencies in various populations all over the world according to their present day reported locations. It is a haplogroup frequency database of different populations calculated from the published data using their reported valid mtDNA sequences. Haplogroups are calculated by using HaploGerp 2.0 [6] with latest Phylotree [7], the *de facto* standard mtDNA tree.

2. Database architecture and management

mtDNAmap is currently implemented using MySQL, an open-source relational database management system (RDBMS) with the help of

* Corresponding author at: Department of Forensic Sciences, University of Health Sciences, Lahore, Pakistan.
E-mail address: dnatypist@gmail.com (A. Rakha).

<http://dx.doi.org/10.1016/j.fsigss.2017.09.208>

Received 10 September 2017; Received in revised form 22 September 2017; Accepted 25 September 2017
1875-1768/ © 2017 Published by Elsevier Ireland Ltd.

phpMyAdmin. phpMyAdmin is the web development suite and hosting server that provides a very user friendly interface for a customized database. The database stores the data of haplogroup frequency along with the information on geographical coordinates of sampling location, population, country, the range of analyzed sequence and source reference. In order to perform entry of the data into the database, a PHP based data entry system was developed.

3. User interface

The user interface for using the database to fetch the required information is developed by combining the powers of HTML, CSS, PHP, JavaScript and jQuery. A query interface was developed, which has been equipped with jQuery UI functionality for the auto completion. When the user enters a query, the database is searched for the given haplogroup and all occurrences are loaded after processed by leaflet and mapbox functions for interactive geo-maps. Leaflet JavaScript is free, open-source library for the development of the interactive map system. It has a power for mobile-friendly interactive maps. The JavaScript library, Mapbox, an open source mapping platform was also used for custom designed maps. The Mapbox provides a great opportunity to generate heatmaps. There are two optional displays and a user can switch between heatmap and a detailed map.

4. Data collection

The data are being collected from the publications having information on mtDNA sequences/variations of different populations. The sequences are retrieved in "FASTA" format from NCBI GenBank [8]. For the determination of SNPs (Single Nucleotide Polymorphisms) of each sequence, mtProfiler is used, i.e., a server providing facility of simultaneous alignment of mtDNA with revised Cambridge Reference Sequence (rCRS) for the determination of mtSNPs using the mtDNA nomenclature tool packaged online [9]. Finally, the mtSNPs list with sequence identifier is given as input to HaploGrep 2.0 [6] for the calculation of Haplogroups of given sequences. Latest versions of HaploGrep and PhyloTree 17 [7] are used for the estimation of Haplogroups. The published literature is being continuously searched for updating the database.

5. mtDNAmapp access

The mtDNAmapp database is available to the general public through the World Wide Web (<http://www.dnageography.com/mtDNAmapp.php>). mtDNAmapp provides the open access to the whole or part of published high-quality data. The tool is not only useful for researchers from forensic and anthropology backgrounds but also in general public.

The query box had been equipped with the feature of fast suggestions for indexed haplogroups. It provides the power of pattern matching search to provide fast, efficient and nearest matching haplogroup.

After the submission of a query, the data is loaded from the database. As the information fields are filled with relevant geo-coordinates, an interactive map building system activates to dynamically show the results on world map. Moreover, frequency calculations for each searched haplogroup are made and display on relevant geo-location of the

sampled populations. The system has been equipped with different color markers and texts for convenient understanding of results. The popup window having details regarding the population name, country, sequence range and reference is made available for detailed information. For the better understanding of data, we have another view to show results in the form of heatmap.

Remarks Mitomap [2] provides established phylogenetic tree illustrating evolution of haplogroups in different regions of the world. mtDNAmapp shows frequencies of all catalogued haplogroups geographically distributed around the globe. Although inspirations came from the Haplogroup Browser tool available through EMPOP [10] portal, whereas major limitation remains to be number of sequences submitted to the database. The important feature of mtDNAmapp is the utilization of all available sequence ranges from quality publications. Cosmetically, here a gradient color scheme for interactive visual display and popup window for each location provide convenient-to-use experience to all kinds of users.

Conflict of interest

The authors declare no conflict of interest.

Role of the funding source

This research was financially supported and funded by University of Health Sciences, Lahore, Pakistan research fund.

Acknowledgements

We acknowledge University of Health Sciences, Lahore for providing computational resources for this work.

References

- [1] S. Anderson, M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger, I.G. Young, Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome, *J. Mol. Biol.* 156 (1982) 683–717, [http://dx.doi.org/10.1016/0022-2836\(82\)90137-1](http://dx.doi.org/10.1016/0022-2836(82)90137-1).
- [2] A.M. Kogelnik, M.T. Lott, M.D. Brown, S.B. Navathe, D.C. Wallace, MITOMAP: a human mitochondrial genome database, *Nucleic Acids Res.* 24 (1996) 177–179.
- [3] L. Fan, Y.G. Yao, MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations, *Mitochondrion* 11 (2011) 351–356, <http://dx.doi.org/10.1016/j.mito.2010.09.013>.
- [4] C. Hillis, *Molecular Systematics*, Sinauer Associates, 1996, 1996.
- [5] J.B. Stewart, P.F. Chinnery, The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease, *Nat. Rev. Genet.* 16 (2015) 530–542, <http://dx.doi.org/10.1038/nrg3966>.
- [6] H. Weissensteiner, D. Pacher, A. Kloss-Brandstätter, L. Forer, G. Specht, H.-J. Bandelt, F. Kronenberg, A. Salas, S. Schönherr, HaploGrep 2 mitochondrial haplogroup classification in the era of high-throughput sequencing, *Nucleic Acids Res.* 44 (2016) W58–W63, <http://dx.doi.org/10.1093/nar/gkw233>.
- [7] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, *Hum. Mutat.* 30 (2009) 386–394, <http://dx.doi.org/10.1002/humu.20921>.
- [8] D.A. Benson, M. Cavanaugh, K. Clark, I. Karsch-Mizrachi, D.J. Lipman, J. Ostell, E.W. Sayers, GenBank, *Nucleic Acids Res.* 41 (2013) 36–42, <http://dx.doi.org/10.1093/nar/gks1195>.
- [9] I.S. Yang, H.Y. Lee, W.I. Yang, K.-J. Shin, mtDNAProfiler: a web application for the nomenclature and comparison of human mitochondrial DNA sequences, *J. Forensic Sci.* 58 (2013) 972–980, <http://dx.doi.org/10.1111/1556-4029.12139>.
- [10] W. Parson, A. Dür, EMPOP-A forensic mtDNA database, *forensic sci, Int. Genet.* 1 (2007) 88–92, <http://dx.doi.org/10.1016/j.fsigen.2007.01.018>.