International Congress Series 1261 (2004) 91–93

ELSEVIER

www.ics-elsevier.com

The mitochondrial DNA tree and forensic science

Martin Richards*

Department of Chemical and Biological Sciences, University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

Abstract. Studies of mitochondrial DNA (mtDNA) began with simple restriction assays, moved on to high-resolution restriction analysis and control-region sequencing, and now make use of complete mtDNA sequences. Forensic labs have, by and large, continued to rely on the use of control-region databases, but these databases often contain a disconcertingly large number of sequence errors. A phylogenetic approach to mtDNA data sets can help to uncover such errors, and it is argued that an understanding of both the phylogenetic and phylogeographic context of mtDNA lineages is a valuable tool from which the forensics community can benefit. © 2003 Published by Elsevier B.V.

Keywords: Mitochondrial DNA (mtDNA); Genealogy; Phylogenetic; Phylogeography

1. The mtDNA tree and phylogeography

Phylogenetic analysis is key to the evolutionary analysis of mtDNA. Tree-building methods perform poorly when faced with data sets from control-region sequencing, or even high-resolution RFLP analysis, because of the high levels of recurrent mutation. The introduction of the reduced-median (RM) algorithm for constructing phylogenetic networks has provided a simple means of phylogenetic analysis on such data sets, which performs well in simulation studies [1]. Network approaches portray ambiguities in the data, so that regions of uncertainty are picked out, and they present the phylogenetic estimate complete with uncertainties in a single diagram. The RM algorithm, furthermore, allows the more plausible pathways to be pinpointed using population-genetics criteria not normally employed in tree-building algorithms. By displaying the data in this compact form, networks are also extremely useful for exploratory data analysis. An example is phylogeographic analysis, which provides an additional encoding for the region of origin of each sample, showing how the lineages are distributed in space.

The picture that has emerged of the global mtDNA tree is one of clades of lineages, known as haplogroups, which are more or less geographically restricted at the continental level. The root of the tree lies within eastern and southern African lineages, and is about 200,000 years old [2]. If we trace our way through the tree towards the present, we reach

* Tel.: +44-1484-472428; fax +44-1484-472182.
E-mail address: m.b.richards@hud.ac.uk (M. Richards).

0531-5131/ © 2003 Published by Elsevier B.V.
doi:10.1016/S0531-5131(03)01704-7

the node at the base of haplogroup L3 which gave rise to both African and non-African descendants [3], including just two non-African branches, haplogroups M and N. Although not clearly distinguishable with either control-region sequencing or standard high-resolution RFLP analysis, M and N can be separated by typing a few additional positions in the coding region of the molecule [4]. The two non-African haplogroups are almost exactly the same age—about 65,000 years—and may well have dispersed from Africa together in a single, small group at about this time.

Haplogroups M and N themselves break down into smaller haplogroups that are more restricted geographically [5]. For example, south Asians mainly have lineages within one set of haplogroups within M and N, east Asians another set, and Australasians another. Most Europeans and Near Easterners, and many Central Asians, share a different set of haplogroups again, all within haplogroup N. Haplogroup M was perhaps lost as a result of founder effects as the early groups splintered and some made their way towards Europe.

2. Identifying errors in databases

Evidence is mounting that mtDNA databases are far from error-free [6]. These errors can arise in all sorts of ways: sample mix-up or cross-contamination in the lab; base misscoring due to sequencing artefacts; simple misreading of the output; and—perhaps most commonly—mistakes in transferring data into databases and tables. The impact of these errors on both evolutionary and forensic analyses may not be very great, but it seems risky to assume that it is negligible. In forensics, database errors would depress estimates of the population frequency of genuine mtDNA types, giving a match between a crime scene sample and a suspect more credibility than it deserved.

Fortunately, our knowledge of natural mtDNA variation makes it possible to identify many of these errors. The trick is to locate the new sequence to its approximate position in the mtDNA tree and to carry out a network analysis of it alongside closely related types from the database. To do this, if the typing is normally restricted to HVS-I and HVS-II, it may be necessary to type a few additional coding-region positions (since not all branches in the tree are resolved by control-region mutations—although HVS-I motifs are often sufficient). The new sequence will either match a known type, occur at a previously unsampled internal node in the tree, or generate a new tip with a private polymorphism. The latter are particular candidates for rechecking, since previously unseen variants may be the result of sequencing errors. Errors are particularly obvious when they affect a motif position for a particular mtDNA haplogroup. Bandelt et al. [7] have compiled a taxonomy of errors which can be checked quite systematically using network diagrams, assisted by a knowledge of the mutation rate of each position indicating bases likely to have undergone genuine recurrent mutation [8].

3. Phylogeography and ethnic affiliation

It may also be useful when carrying out forensic analyses to take into account the phylogeography of the mtDNAs one is interested in. The pitfalls of not doing so are illustrated by a recent BBC Television documentary, entitled ‘Motherland: A Genetic Journey’. This film showed several British African–Caribbeans undergoing mtDNA analysis and tracing their lineages back to a particular African ‘homeland’ before the

slave trade. However, there are major gaps in the African mtDNA database, and subsequent research has shown that one of the types, localised in the film to the small island of Bioko in tropical west-central Africa, also occurs as far afield as Mozambique, with close relatives in Tanzania. This surprisingly wide distribution is most likely a result of the Bantu dispersals out of west-central Africa 4000 years ago [3]. Similarly, knowledge of the phylogeography and inferred demographic history may usefully inform forensic analyses estimating ethnic affiliation, especially in the absence of databases with perfect geographic coverage.

Acknowledgements

Thanks to Vincent Macaulay and Han-Jürgen Bandelt for critically reading the manuscript.

References

[1] H.-J. Bandelt, et al., Median networks: speedy construction and greedy reduction, one simulation, and two case studies from human mtDNA, Mol. Phylogenet. Evol. 16 (2000) 8–28.
[2] D. Mishmar, et al., Natural selection shaped regional mtDNA variation in humans, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 171–176.
[3] A. Salas, et al., The making of the African mtDNA landscape, Am. J. Hum. Genet. 71 (2002) 1082–1111.
[4] L. Quintana-Murci, et al., Genetic evidence for an early exit of *Homo sapiens sapiens* from Africa through eastern Africa, Nat. Genet. 23 (1999) 437–441.
[5] M. Richards, V. Macaulay, Genetic data and the colonization of Europe: genealogies and founders, in: C. Renfrew, K. Boyle (Eds.), Archaeogenetics: DNA and the Population Prehistory of Europe, McDonald Institute for Archaeological Research, Cambridge, 2000, pp. 139–151.
[6] P. Forster, To err is human, Ann. Hum. Genet. 67 (2003) 2–4.
[7] H.-J. Bandelt, et al., Detecting errors in mtDNA data by phylogenetic analysis, Int. J. Leg. Med. 115 (2001) 64–69.
[8] H.-J. Bandelt, et al., The fingerprint of phantom mutations in mitochondrial DNA data, Am. J. Hum. Genet. 71 (2002) 1150–1160.