Forensic Science International: Genetics 6 (2012) 143–145

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

Current Next Generation Sequencing technology may not meet forensic standards

Hans-Jürgen Bandelt [a,b,*], Antonio Salas [a,b]

[a] Department of Mathematics, University of Hamburg, 20146 Hamburg, Germany
[b] Unidade de Xenética, Instituto de Medicina Legal, Facultade de Medicina, and Departamento de Anatomía Patolóxica e Ciencias Forenses, Facultade de Medicina, Universidade de Santiago de Compostela, 15782 Galicia, Spain

ARTICLE INFO

Article history:
Received 20 September 2010
Received in revised form 16 January 2011
Accepted 2 April 2011

Keywords:
mtDNA
Entire genome
Next-gen sequencing
Heteroplasmy
Genotyping errors

ABSTRACT

In a Nature paper of 2010, the concern was raised that intra-individual mtDNA variation may be more pronounced than previously believed, in that heteroplasmies are common and vary markedly from tissue to tissue. This claim taken at face value would have considerable impact on forensic casework. It turns out however that the employed technology detected the germ-line variation relative to the reference sequence only incompletely: on average at least five mutations were missed per sample, as an in silico reassessment of the data reveals. Before one can really set out to access to entire mtDNA genome data with relative ease for forensic purposes, one needs careful calibration studies under strict forensic conditions—or might have to wait for another generation.

1. Introduction

Intra-individual mtDNA variation is an issue of key importance in forensic casework [1–4]; and it was treated as such by the DNA commission almost a decade ago [5]. Recently, intra-individual variation has been re-assessed by He et al. [6] by using Next Generation Sequencing (NSG) techniques. According to the authors "Our new results clearly show that heteroplasmies affect the entire mitochondrial genome, are common in normal individuals and vary markedly from tissue to tissue... This suggests caution in excluding identity on the basis of a single or small number of mismatched alleles when the tissue in evidence (such as sperm) is not the same as the reference tissue of the suspect (such as blood or hair)". The quality of the results obtained by He et al. can be evaluated a posteriori by way of contrasting the mtDNA patterns observed by these authors with those predicted by the mtDNA phylogeny, in a similar way as executed in previous studies [3,7,8]. The current knowledge of the mtDNA phylogeny is now compiled in the PhyloTree project [9] which aims to reconstruct the worldwide mtDNA tree based on currently >8700 entire mtDNA genomes (Phylotree Build 11).

The present report examines the data quality of the He's et al. study by using in silico phylogenetic procedures. The results obtained by other authors on NGS of mtDNA variation [10] are also discussed. As demonstrated below, NGS is still far from meeting the high standards demanded in forensic routine casework.

2. Results

The mtDNA variation in 10 patients and two CEPH family cell lines of four individuals recorded in [6] can be read against the detailed mtDNA tree Build 11 offered by PhyloTree [9]. Allocation of mtDNA profiles from [6] to the respective haplogroups is easy and unambiguous in all cases, so that one can readily spot the missing variants in the mtDNAs that are expected by haplogroup status. The search results for the mtDNA variation of all 10 patients and the four CEPH cases are summarized in Table 1.

In detail, for example, the germ-line mutation profile for Patient 4 reported in Supplementary Table 6 of He et al. [6] is one of the most problematic ones. It is clear from more than half a dozen variants that the corresponding mtDNA lineage belongs to haplogroup T2. However, tracing the entire pathway from the rCRS (within haplogroup H2a2) to the ancestral T2 haplotype, it turns out that half of the expected mutational variants were not recorded in [6]. Proceeding within the T2 subtree one can see four further variants supporting the pathway T2 → T2a → T2a1 → T2a1b, but on the other hand, another two variants are then missing. To find more closely related sequences within haplogroup T2a1b we Google the variants from the patient's profile (as in [11,12]) not yet captured by T2a1b status one by one. In particular, by entering 'C12741T PhyloTree', we obtain a reference pointing to a T2a1b coding-region sequence (GenBank acc. no. EF657381.1) which possesses the expected variants (according to the haplotype profile provided by the corresponding

* Corresponding author at: Department of Mathematics, University of Hamburg, 20146 Hamburg, Germany.
E-mail address: bandelt@yahoo.com (H.-J. Bandelt).

1872-4973/$ – see front matter © 2011 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.fsigen.2011.04.004

Table 1
Deficiencies in the mtDNA sequences reported by He et al. [6].

Sample	HG[a]	Missed mutations (characterizing haplogroup)[b]	Comments
Patient 1	J1c3a1	15326 (H2a2), 2706 (H), 4216 (R2'JT), 3010[c] (J1), 13934 (J1c3)	Compare with GenBank acc. no. AY495211
Patient 2	J1b1a	15326 (H2a2), 2706 (H), 4216 (R2'JT), 3010[c] (J1), 16222 (J1b), 16261[c] (J1b), 5460[c] (J1b1)	
Patient 3	J1c or J1c3a	15326 (H2a2), either 2706 (H) or 13934 (J1c3), 4216 (R2'JT), 3010[c] (J1)	
Patient 4	T2a1b1	15326 (H2a2), 2706 (H), 4216 (R2'JT), 4917 (T), 8697 (T), 10463 (T), 15607 (T), 15928 (T), 16294[c] (T), 13965 (T2a), 13966 (T2a1b1)	Shares 12741 variant with GenBank acc. no. EF657381.1
Patient 5	N22	15326 (H2a2), 2706 (H), 16223[c] (R), 942 (N22), 16249[c] (N22)	
Patient 6	U5a1	15326 (H2a2), 2706 (H), 13617 (U5), 16270 (U5), 16256[c] (U5a)	Shares 12103 and 14893 variants with Dutch LHON 072U [14]
Patient 7	X2a2	15326 (H2a2), 2706 (H), 7028 (H), 16223[c] (R), 153 (X), 13966 (X), 16278[c] (X), 1719[c] (X2), 12397 (X2a'j), 8913 (X2a), 16213 (X2a)	Compare with GenBank acc. no. FJ168757
Patient 8	J1c3a	15326 (H2a2), 2706 (H), 4216 (R2'JT), 3010[c] (J1), 13934 (J1c3)	Compare with GenBank acc. no. AY495213
Patient 9	J1c	15326 (H2a2), 2706 (H), 4216 (R2'JT), 3010[c] (J1), 185[c] (J1c)	
Patient 10	H7c	15326 (H2a2), 4793 (H7)	
CEPH 45♀	T2b3	15326 (H2a2), 1438 (H2), 14905 (T)	
CEPH 45♂	H1	15326 (H2a2), 1438 (H2)	
CEPH1377♀	T1	15326 (H2a2), 1438 (H2), 14905 (T)	Shares 3867 variant with GenBank acc. no. EU369395
CEPH1377♂	K1b1a	15326 (H2a2), 1438 (H2), 152[c] (K1b1a)	Potential documentation error at site 15452

[a] HG = haplogroup according to mtDNA Build 11 from PhyloTree [9].
[b] Numbers indicate transitions relative to the rCRS [29]. Haplogroup status determined by the respective transition is in brackets. All variants were considered having allele frequencies >90% in Supplementary Tables 3, 4, and 6 (normal cells) from [6].
[c] Mutational hot spot [18].

PhyloTree entry). There exists yet another complete mtDNA sequence (not submitted to GenBank) which in fact belongs to this particular T2a1b branch (Patient 6 from Table 3 of [13]).

The mtDNA profile of Patient 7 does not fare better: it equally lacks 11 nucleotide variants, since Native North American haplogroup status X2a2 can be inferred with confidence. The mtDNA profile of Patient 5 is also somewhat unusual. The first guess is its allocation to haplogroup N because of the presence of the C12705T variant. Indeed, by searching site by site in PhyloTree, we get several hits for haplogroup status N22, although again a few mutations are missing. The mtDNA profile of Patient 6 can be well recognized as a typical European mtDNA lineage, pointing to haplogroup U5a1 (with characteristic sites 15218 and 16399), although the key HVS-I variants at sites 16256 and 16270 for haplogroup U5a are missing. Here the Google search for 'C12103A PhyloTree' as well as for 'A14893G PhyloTree' are each successful, leading to the particular U5a1 lineage of the Dutch LHON pedigree 072U [14]. The latter two mutations are thus characteristic for a new subhaplogroup of U5a1.

The two mtDNA profiles from "mother" and "father" of the CEPH 1377 family [6] reveal another problem, which may constitute a documentation error [15–17]. The transversion C15452A is a rare mutational event: it perfectly highlights haplogroup JT status, and therefore all J and T lineages should harbour this variant. The mutational scoring reported by Soares et al. [18] gave only one further hit for this mutation. In order to find this second hit, we queried the mtDB Website (http://www.genpat.uu.se/mtDB/) for the partial profile '4216T 15452A', which would be expected in any lineage outside haplogroup JT that had gained the variant C15452A independently. And indeed a single hit shows up, pointing to GenBank acc. no. AF382007, from the study by Maca-Meyer et al. [19]. All variants found in AF382007, except for 15452A are actually covered by EF660917, so that one cannot exclude the possibility that this variant entered the original sequence AF382007 by sample confusion affecting the sequence segment 15162–15720. In any case this variant is so rare that the occurrence in a K1b1a lineage is suspicious, especially as the other member of the CEPH 1377 family analyzed alongside belongs to haplogroup JT, where this variant naturally occurs.

Finally, one can also compare the putative heteroplasmic patterns reported in [6] with the ones reported by Irwin et al. [20] based on more than 5000 high-quality control region sequences collected by the EMPOP project. According to the latter authors, the top nine heteroplasmic sites in the control region are 16093, 152, 146, 204, 195, 16189, 150, 215, 16183; however, none of them, with the only exception of 16093, appears to be heteroplasmic in [6]. Standard Sanger sequencing can detect heteroplasmic status above the 10% level for the minimum allele, so the resolution of the technique cannot by itself explain the differences.

3. Discussion

Although we agree with the general conclusion formulated by He et al. regarding the existence of mtDNA intra-individual variation (which however is not new to forensic geneticists), the exact nature and amount of this variation remain to be elucidated. Thus, detailed phylogenetic reassessment of the data published by He et al. [6] clearly demonstrates the presence of several systematic oversights (involving site 15326 in the entire study, site 1438 in all CEPH cell lines, and sites 2706, 3010, and 4216 in patients' cells) and several oversights of a more erratic pattern. In total, the genomes reported in their study contain on average (at least) five errors each, as inferred from the current knowledge of the mtDNA phylogeny. In view of this error rate a forensic geneticist could safely conclude that the data presented by these authors do not meet the minimum quality standard for comparing mtDNA sequences from different tissues of the same individual.

Some optimism has been expressed among forensic geneticists concerning the prospect of NGS techniques [21]. The optimism is however somewhat dampened when confronted with the results obtained by He et al. and some more recent studies. The study by He et al. demonstrates that the analysis criteria to interpret their data were not appropriate and would need careful definition before applied to forensic samples. The results from Zaragoza et al. [10] using next-gen Roche 454 FLX sequencing and the conclusions about the merits of the new NGS approach are also misleading. These authors compared the sequencing results of the 454 sequencing methodology versus standard Sanger sequencing, observing 98% of concordance in variant detection. According to these authors, "for the five discordant variants detected by 454 (Table 2), retrospective review of the initial chromatograms and repeated Sanger sequencing confirmed the presence of each

variant". One should therefore wonder in the first place whether the initial Sanger sequencing was carried out using the appropriate standards instead of concluding that "...four false negative errors resulted from miscalled Sanger sequences, and the fifth Sanger "miss" to the lack of coverage [of 454]..." and that "these results support a potential role of next-generation sequencing in the discovery of novel mtDNA variants with heteroplasmy below the level reliably detected with Sanger sequencing". On the other hand, the 454 genotyping could not uncover the variation at homopolymeric tracks of the control region (which Sanger does), where inter-generation or tissue to tissue mtDNA differences commonly occur in healthy individuals.

Standard Sanger sequencing is the gold standard in mtDNA forensic genetics (and still so in other disciplines), although the technique is not exempt of problems [3,17] when applied without appropriate standards. Ultra-sequencing is emerging in biomedicine for the generation of a huge amount of data within the framework of very ambitious 'omic' projects (e.g. 1000 Genomes; http://www.1000genomes.org; [22]). Unfortunately, these NGS techniques cannot yet be applied to forensic casework for several reasons; for instance, the large amount and quality of DNA needed in these ultra-sequencing platforms is an obvious limitation for forensic casework. The use of whole genome amplification techniques could perhaps help to overcome this limitation, as successfully tested for high throughput mtDNA SNP genotyping [23], although multi-centric standardization exercises are mandatory before implementing these new techniques in real forensic casework. To date, there is no formal forensic assessment of the data quality generated by these NGS technologies, as it is commonly exercised with standard sequencing techniques [2,24–28]. Although the two studies examined here were carried out without the necessary forensic standards, the indirect quality assessment of their results as performed in the present article demonstrates that well-defined technical guidelines and interpretation rules need to be formulated, as it had been done in the case of Sanger sequencing, in order to allow for the production of high quality NGS data in forensics.

Acknowledgements

We thank Walther Parson for critical reading this article. This research received support from two grants from the Fundación de Investigación Médica Mutua Madrileña, and a grant from the Ministerio de Ciencia e Innovación (SAF2008-02971), given to AS.

References

[1] A. Salas, M.V. Lareu, Á. Carracedo, Heteroplasmy in mtDNA and the weight of evidence in forensic mtDNA analysis: a case report, Int. J. Legal Med. 114 (3) (2001) 186–190.
[2] A. Alonso, A. Salas, C. Albarrán, E. Arroyo, A. Castro, M. Crespillo, A.M. di Lonardo, M.V. Lareu, C.L. Cubria, M.L. Soto, J.A. Lorente, M.M. Semper, A. Palacio, M. Paredes, L. Pereira, A.P. Lezaun, J.P. Brito, A. Sala, M.C. Vide, M. Whittle, J.J. Yunis, J. Gómez, Results of the 1999–2000 collaborative exercise and proficiency testing program on mitochondrial DNA of the GEP-ISFG: an inter-laboratory study of the observed variability in the heteroplasmy level of hair from the same donor, Forensic Sci. Int. 125 (1) (2002) 1–7.
[3] A. Salas, H.-J. Bandelt, V. Macaulay, M.B. Richards, Phylogeographic investigations: the role of trees in forensic genetics, Forensic Sci. Int. 168 (1) (2007) 1–13.
[4] G.G. Paneto, J.A. Martins, L.V. Longo, G.A. Pereira, A. Freschi, V.L. Alvarenga, B. Chen, R.N. Oliveira, M.H. Hirata, R.M. Cicarelli, Heteroplasmy in hair: differences among hair and blood from the same individuals are still a matter of debate, Forensic Sci. Int. 173 (2–3) (2007) 117–121.
[5] Á. Carracedo, W. Bär, P. Lincoln, W. Mayr, N. Morling, B. Olaisen, P. Schneider, B. Budowle, B. Brinkmann, P. Gill, M. Holland, G. Tully, M. Wilson, DNA commission of the international society for forensic genetics: guidelines for mitochondrial DNA typing, Forensic Sci. Int. 110 (2) (2000) 79–85.
[6] Y. He, J. Wu, D.C. Dressman, C. Iacobuzio-Donahue, S.D. Markowitz, V.E. Velculescu, L.A. Diaz Jr., K.W. Kinzler, B. Vogelstein, N. Papadopoulos, Heteroplasmic mitochondrial DNA mutations in normal and tumour cells, Nature 464 (7288) (2010) 610–614.
[7] A. Salas, Y.-G. Yao, V. Macaulay, A. Vega, A. Carracedo, H.-J. Bandelt, A critical reassessment of the role of mitochondria in tumorigenesis, PLoS Med. 2 (11) (2005) e296.
[8] H.-J. Bandelt, A. Salas, Contamination and sample mix-up can best explain some patterns of mtDNA instabilities in buccal cells and oral squamous cell carcinoma, BMC Cancer 9 (1) (2009) 113.
[9] M. van Oven, M. Kayser, Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation, Hum. Mutat. 30 (2) (2009) E386–E394.
[10] M.V. Zaragoza, J. Fass, M. Diegoli, D. Lin, E. Arbustini, Mitochondrial DNA variant discovery and evaluation in human cardiomyopathies through next-generation sequencing, PLoS One 5 (8) (2010).
[11] H.-J. Bandelt, A. Salas, C.M. Bravi, What is a 'novel' mtDNA mutation—and does 'novelty' really matter? J. Hum. Genet. 51 (12) (2006) 1073–1082.
[12] H.-J. Bandelt, Y.-G. Yao, A. Salas, The search of 'novel' mtDNA mutations in hypertrophic cardiomyopathy: MITOMAPping as a risk factor, Int. J. Cardiol. 126 (3) (2008) 439–442.
[13] M.G. Shin, S. Kajigaya, B.C. Levin, N.S. Young, Mitochondrial DNA mutations in patients with myelodysplastic syndromes, Blood 101 (8) (2003) 3118–3125.
[14] N. Howell, C. Herrnstadt, C. Shults, D.A. Mackey, Low penetrance of the 14484 LHON mutation when it arises in a non-haplogroup J mtDNA background, Am. J. Med. Genet. A 119 (2) (2003) 147–151.
[15] H.-J. Bandelt, A. Salas, C.M. Bravi, Problems in FBI mtDNA database, Science 305 (5689) (2004) 1402–1404.
[16] H.-J. Bandelt, A. Salas, S. Lutz-Bonengel, Artificial recombination in forensic mtDNA population databases, Int. J. Legal Med. 118 (5) (2004) 267–273.
[17] A. Salas, Á. Carracedo, V. Macaulay, M. Richards, H.-J. Bandelt, A practical guide to mitochondrial DNA error prevention in clinical, forensic, and population genetics, Biochem. Biophys. Res. Commun. 335 (3) (2005) 891–899.
[18] P. Soares, L. Ermini, N. Thomson, M. Mormina, T. Rito, A. Röhl, A. Salas, S. Oppenheimer, V. Macaulay, M.B. Richards, Correcting for purifying selection: an improved human mitochondrial molecular clock, Am. J. Hum. Genet. 84 (6) (2009) 740–759.
[19] N. Maca-Meyer, A.M. González, J.M. Larruga, C. Flores, V.M. Cabrera, Major genomic mitochondrial lineages delineate early human expansions, BMC Genetics 2 (2001) 13.
[20] J.A. Irwin, J.L. Saunier, H. Niederstätter, K.M. Strouss, K.A. Sturk, T.M. Diegoli, A. Brandstätter, W. Parson, T.J. Parsons, Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples, J. Mol. Evol. 68 (5) (2009) 516–527.
[21] J.A. Irwin, W. Parson, M.D. Coble, R.S. Just, mtGenome reference population databases and the future of forensic mtDNA analysis, Forensic Sci. Int. Genet. (2010).
[22] R.M. Durbin, G.R. Abecasis, D.L. Altshuler, A. Auton, L.D. Brooks, R.A. Gibbs, M.E. Hurles, G.A. McVean, A map of human genome variation from population-scale sequencing, Nature 467 (7319) (2010) 1061–1073.
[23] M. Cerezo, V. Černý, Á. Carracedo, A. Salas, Applications of MALDI-TOF MS to large-scale human mtDNA population-based studies, Electrophoresis 30 (21) (2009) 3665–3673.
[24] M. Crespillo, M.R. Paredes, L. Prieto, M. Montesino, A. Salas, C. Albarrán, V. Álvarez-Iglesias, A. Amorin, G. Berniell-Lee, A. Brehm, J.C. Carril, D. Corach, N. Cuevas, A.M. Di Lonardo, C. Doutremepuich, R.M. Espinheira, M. Espinoza, F. Gómez, A. González, A. Hernández, M. Hidalgo, M. Jimenez, F.P. Leite, A.M. López, M. López-Soto, J.A. Lorente, S. Pagano, A.M. Palacio, J.J. Pestano, M.F. Pinheiro, E. Raimondi, M.M. Ramon, F. Tovar, L. Vidal-Rioja, M.C. Vide, M.R. Whittle, J.J. Yunis, J. Garcia-Hirschfel, Results of the 2003–2004 GEP-ISFG collaborative study on mitochondrial DNA: focus on the mtDNA profile of a mixed semen-saliva stain, Forensic Sci. Int. 160 (2–3) (2006) 157–167.
[25] M. Montesino, A. Salas, M. Crespillo, C. Albarrán, A. Alonso, V. Álvarez-Iglesias, J.A. Cano, M. Carvalho, D. Corach, C. Cruz, A. Di Lonardo, R. Espinheira, M.J. Farfán, S. Filippini, J. García-Hirschfeld, A. Hernández, G. Lima, C.M. López-Cubría, M. López-Soto, S. Pagano, M. Paredes, M.F. Pinheiro, A.M. Rodríguez-Monge, A. Sala, S. Sóñora, D.R. Sumita, M.C. Vide, M.R. Whittle, A. Zurita, L. Prieto, Analysis of body fluid mixtures by mtDNA sequencing: an inter-laboratory study of the GEP-ISFG working group, Forensic Sci. Int. 168 (1) (2007) 42–56.
[26] L. Prieto, A. Alonso, C. Alves, M. Crespillo, M. Montesino, A. Picornell, A. Brehm, J.L. Ramírez, M.R. Whittle, M.J. Anjos, I. Boschi, J. Buj, M. Cerezo, S. Cardoso, R. Cicarelli, D. Comas, D. Corach, C. Doutremepuich, R.M. Espinheira, I. Fernández-Fernández, S. Filippini, J. Garcia-Hirschfeld, A. González, B. Heinrichs, A. Hernández, F.P.N. Leite, R.P. Lizarazo, A.M. López-Parra, M. López-Soto, J.A. Lorente, B. Mechoso, I. Navarro, S. Pagano, J.J. Pestano, J. Puente, E. Raimondi, A. Rodríguez-Quesada, M.F. Terra-Pinheiro, L. Vidal-Rioja, C. Vullo, A. Salas, 2006 GEP-ISFG Q1 collaborative exercise on mtDNA. Reflections about interpretation, artefacts, and DNA mixtures, Forensic Sci. Int. Genet. 2 (2) (2008) 126–133.
[27] L. Prieto, M. Montesino, A. Salas, A. Alonso, C. Albarrán, S. Álvarez, M. Crespillo, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A.G. de la Vega, L. Gusmão, C.M. López, M. López-Soto, J.A. Lorente, M. Malaghini, C.A. Martínez, N.M. Modesti, A.M. Palacio, M. Paredes, S.D. Pena, A. Pérez-Lezaun, J.J. Pestano, J. Puente, A. Sala, M. Vide, M.R. Whittle, J.J. Yunis, J. Gómez, The 2000–2001 GEP-ISFG Collaborative Exercise on mtDNA: assessing the cause of unsuccessful mtDNA PCR amplification of hair shaft samples, Forensic Sci. Int. 134 (1) (2003) 46–53.
[28] A. Salas, L. Prieto, M. Montesino, C. Albarrán, E. Arroyo, M.R. Paredes-Herrera, A.M. Di Lonardo, C. Doutremepuich, I. Fernández-Fernández, A.G. de la Vega, C. Alves, C.M. López, M. López-Soto, J.A. Lorente, A. Picornell, R.M. Espinheira, A. Hernández, A.M. Palacio, M. Espinoza, J.J. Yunis, A. Pérez-Lezaun, J.J. Pestano, J.C. Carril, D. Corach, M.C. Vide, V. Álvarez-Iglesias, M.F. Pinheiro, M.R. Whittle, A. Brehm, J. Gómez, Mitochondrial DNA error prophylaxis: assessing the causes of errors in the GEP'02–03 proficiency testing trial, Forensic Sci. Int. 148 (2–3) (2005) 191–198.
[29] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat. Genet. 23 (1999) 147.