Int J Legal Med (2008) 122:341–345
DOI 10.1007/s00414-007-0191-5

SHORT COMMUNICATION

Specificity of mtDNA-directed PCR—influence of NUclear MTDNA insertion (NUMT) contamination in routine samples and techniques

Ana Goios · Lourdes Prieto · António Amorim · Luísa Pereira

Received: 9 May 2007 / Accepted: 8 August 2007 / Published online: 14 September 2007

Abstract Nuclear mitochondrial insertions (NUMTs) are sequences homologous to mtDNA, which are present throughout the human nuclear genome. The possibility that these sequences may be accidentally amplified in reactions directed to mtDNA has been raised and evaluated by different groups and by different means. Despite that, data is still missing on the specificity of PCRs in routine procedures in what concerns contamination with nuclear mtDNA insertions (NUMTs). In this work, we performed PCR sequencing reactions with primers directed either to mitochondrial or to NUMT DNA with different annealing temperatures and in different tissues. We observed that (a) contamination with NUMTs depends on the sample and tissue, and (b) employing routine techniques, there is no risk of co-amplification. Only when mtDNA is almost completely removed from the samples does the number of NUMT copies exceed mitochondrial sequences, i.e., only in samples with virtually no mtDNA, such as those resulting from preferential semen lysis, is there a risk of accidental amplification of NUMTs. We suggest that to evaluate a possible co-amplification of NUMT DNA, it is more relevant to take into account sample processing and original tissue of the samples, and consequently the relative proportions of NUMT and mtDNA, rather than the presence of NUMTs by itself, irrespectively of its proportion.

Keywords mtDNA · NUMT · Contamination · Routine · PCR

Electronic supplementary material The online version of this article (doi: 10.1007/s00414-007-0191-5) contains supplementary material, which is available to authorized users.

A. Goios (✉) · A. Amorim · L. Pereira
Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP),
R. Dr. Roberto Frias, s/n,
4200-465 Porto, Portugal
e-mail: aalmeida@ipatimup.pt

A. Goios · A. Amorim
Faculdade de Ciências da Universidade do Porto,
Porto, Portugal

L. Prieto
Comisaría General de Policía Científica, DNA Laboratory,
Madrid, Spain

L. Pereira
Medical Faculty, University of Porto,
Porto, Portugal

Introduction

Mitochondrial DNA is a circular molecule that is present in cells in a high copy number, is maternally transmitted and encodes for 13 proteins of the respiratory chain. The particular characteristics of this molecule are the source of its use in a variety of fields, from forensic [1, 2] and population genetics [3, 4] to clinical studies [5–7]. Analyses of mtDNA can be performed through different techniques, such as RFLP, SNaPshot or sequencing, and most of these techniques are based on a prior PCR that requires the use of specific primers for the mtDNA molecule.

Primers designed for amplifications of mtDNA can potentially anneal with sequences in the nuclear genome that present a high homology to mtDNA—nuclear mtDNA insertions or NUMTs. In fact, NUMTs have already been mistaken as heteroplasmic positions, as in the case of the report of an association of mutations in *CO1* and *CO2* genes with the development of Alzheimer's disease [8], which were later shown using ρ^0 cells (lacking mtDNA), to be an artefact resulting from an accidental amplification of

pseudogenes [9, 10]. In other cases, NUMTs have been sporadically co-amplified in tissues with particular features, as shown by Bravi et al. [11] in ancient DNA from an Egyptian mummy that was believed to present an unusual mitochondrial control region sequence, and in a case of a supposed tissue-specific mosaicism for mtDNA in spermatozoa [12]. As it had been suggested that mtDNA anomalies in sperm could play a role in infertility [13], results like those obtained by Thangaraj et al. [12] are usually accepted with interest, and only when the association is demystified [14], do the artefacts come to light.

These examples show that the importance of NUMTs and other artefacts in mtDNA typing encompass a variety of fields from association studies to forensics and population genetics. At the same time, they suggest that a possible contamination occurs only in particular cases, in specific tissues or conditions of the samples.

NUMTs were suggested to have been integrated into the nuclear genome by a mechanism of non-homologous recombination [15, 16] and, because they evolve more slowly than mtDNA, they have been used as outgroup in some studies [17, 18]. Several studies based on in silico analyses have listed the NUMTs that are present throughout the human genome [16, 18, 19]. It was estimated that ≈250 NUMTs with similarities higher than 68% exist throughout the different human chromosomes [18] and, while most of them are unique, some exist in a number of paralogous copies that arose from duplication events. Parr et al. [20], in a study performed with ρ^0 cells, suggested that among the factors that determine whether a NUMT will co-amplify with mtDNA are the region of the mtDNA targeted by the PCR and the number of copies of the NUMT. Other works [21, 22] have addressed this subject with advice or suggestions to avoid NUMT contamination. However, as valuable as studies using BLAST searches or ρ^0 cells can be, there is still missing information on how specific PCRs are with routinely used procedures and samples, and how different tissues behave in the presence of primers directed to mtDNA or to NUMTs.

To understand if there is a real risk of co-amplification of NUMTs with routine samples and techniques, we performed a comparison between sequencing results in four different tissues obtained with two primer pairs, one specifically designed for NUMT and another for mtDNA.

Materials and methods

We selected the longest (≈6,000 bps) and most homologous (97% homology) NUMT (accession no. AL359496.30) from the list by Mishmar et al. [18], representing a region that corresponds to bases 3,914–9,755 in the mtDNA Cambridge Reference Sequence (CRS). We designed two sets of primers, specific either to mtDNA (MtF and MtR) or NUMT (NUMTF and NUMTR), so that both pairs would encompass essentially the same region. The specificity of each primer pair was determined by two different bases in both forward and reverse directions (ESM Fig. 1). We also designed one set of primers (M18F and M18R) that anneals in both NUMT and mtDNA. The amplified fragment was identified as mtDNA or NUMT by two variant positions (diagnostic positions [DP]). Although mtDNA polymorphisms have been reported for these positions [23, 24], they are quite infrequent. In any case, the use of a second diagnostic position allowed us to overcome any possible misclassification of the sequences.

Nuclear DNA was quantified by means of real-time PCR (Quantifiler, AB Applied Biosystems, Foster City, CA, USA). NUMTs or mtDNA amplifications were carried out with a regular *taq* DNA polymerase (Fermentas International, Burlington, Canada) by using the following conditions: 95°C for 2 min followed by 35 cycles as follows: 95°C for 30 s, annealing for 30 s and 72 °C for 1 min, and a final 10 min extension at 72°C. We tested different annealing temperatures, ranging from 53°C to 62°C. The amount of nuclear DNA in the PCR was always higher than 5 ng for a total reaction volume of 25 μl. A negative control was always included in the PCR.

PCR products were purified using Microspin S-300 HR columns (Amersham Biosciences, Uppsala, Sweden), according to the manufacturer's specifications. Sequence reactions were carried out using the Big-Dye Terminator Cycle Sequencing Ready Reaction kit (AB Applied Biosystems). Samples were further purified with Sephadex G-50 DNA Grade f (Amersham Biosciences) and run in an automatic sequencer ABI 3100 (AB Applied Biosystems).

All PCR sequencing analyses were performed in samples from different tissues (blood, cheek swabs and hair with root) from six individuals belonging to different haplogroups, after extraction with Chelex 100 Resin (BioRad, Hercules, CA), and in samples of semen preferential lyses from three different individuals. In an effective preferential lysis protocol, it is expected that nuclear DNA from non-spermatic cells and all mtDNAs remain in the first fraction, while the second fraction retains only nuclear DNA (nDNA) from spermatozoa [25]. These samples provided valuable information on how tissues with different relative proportions of mitochondrial and nuclear DNA behave in the presence of the different primers.

Results and discussion

We have tested a fragment of the mtDNA coding region for the accidental amplification of NUMTs. Although the control region is more frequently analysed in forensic

Table 1 Summary of the results obtained from the amplification with different primers from different tissues and individuals

Samples		Annealing temperatures					
		NUMT primers			MT primers		
		53°C	60°C	62°C	53°C	60°C	62°C
Blood	1	Mt	Mt+Numt	–	Mt	–	Mt
	2	Mt	–	Numt	Mt	–	Mt
	3	Mt	Numt+Mt	Numt	Mt	–	Mt
	4	Mt	Numt	Numt	Mt	–	–
	5	Mt	–	Numt	Mt	–	Mt
	6	Mt	–	Numt	Mt	–	Mt
	7	Mt	–	Numt	Mt	–	Mt
Hair	1	Mt	Mt	–	Mt	–	Mt
	2	Mt	–	Numt	Mt	–	Mt
	3	Mt	Mt	Mt+Numt	Mt	–	Mt
	4	Mt	–	–	Mt	–	Mt
	5	Mt	Numt	–	Mt	–	Mt
	6	Mt	Mt+Numt	Numt	Mt	–	Mt
	7	Mt	–	–	Mt	–	–
Swabs	1	Mt	Mt+Numt	–	Mt	–	Mt
	2	Mt	Mt+Numt	Numt	Mt	–	Mt
	3	Mt	Mt+Numt	Numt	Mt	–	Mt
	4	Mt	Numt	Numt	Mt	–	Mt
	5	Mt	Mt+Numt	Numt	Mt	–	Mt
	6	Mt	–	Numt	Mt	–	Mt
	7	Mt	Mt+Numt	Numt	Mt	–	–
Semen first lysis	8	Mt	Numt	Numt	Mt	Mt	Mt
	9	Mt	–	–	Mt	Mt	Mt
	10	Mt	Numt	Numt	Mt	Mt	Mt
Semen second lysis	8	Numt	Numt	Numt	Mt+Numt	Mt	Mt
	9	Numt	Numt	Numt	Mt+Numt	Mt	Mt
	10	Numt	Numt	Numt	Mt+Numt	Mt	Mt

"–" correspond to sequences not performed or not obtained.

genetics, in silico analyses [21] had already excluded NUMTs that encompass this region as potential sources of contamination, for being too divergent from the primer annealing sites. Moreover, the coding region has been more and more targeted in medical, forensic and population genetics.

Three kinds of results were obtained from the sequencing experiment (ESM Fig. 2): (a) pure mtDNA sequences, (b) a mixture of NUMT and mtDNA and (c) pure NUMT DNA. Except for the second semen fraction, all amplifications performed using a primer annealing temperature of 53°C yielded pure mtDNA (Table 1), either with mtDNA or NUMT-directed primers. This led us to assume that the annealing temperature did not permit enough specificity, and therefore we increased it to 60°C, giving rise to a mixture of NUMT and mtDNA with the NUMT-directed primers. Only with an annealing temperature of 62°C were we able to obtain clean NUMT DNA sequences with NUMT-directed primers. Amplifications with mtDNA-directed primers did not result in unspecific sequences at any temperature.

It was quite clear that the tested tissues did not all behave in the same manner. While blood and cheek swabs presented relatively uniform results in different individuals, hair samples presented a more diverse pattern with mtDNA sequences still resulting from amplifications with NUMT-directed primers at the highest temperatures.

These experiments show that with routine techniques for mtDNA amplification, there is no risk of contamination with NUMT DNA; furthermore, they show that it is more probable that mtDNA is co-amplified when trying to obtain NUMT DNA than the reverse contamination. This fact results from a high number of copies of mtDNA molecules relatively to nuclear DNA in the cells; a hypothesis that had already been raised by others [26]. This would also explain the differences among tissues; and for example, why hair samples present such a high diversity.

This hypothesis is also supported by the results obtained from the two semen fractions. In an effective preferential lysis protocol, it is expected that non-spermatic nuclei and all mtDNAs are obtained in the first fraction, and the only DNA that remains in the second fraction is nuclear DNA

from spermatic cells [25], which results in an absence of mtDNA in the second fraction and a high proportion of mtDNA in the first fraction. In this case, we observed that the lysis was not complete: control region (HVRI) mtDNA was amplified from the second fraction, although only when high amounts of nuclear DNA were used in the PCR. Despite this presence of mtDNA in the second fraction, pure NUMT sequences were observed in amplifications with M18F and M18R primers, which anneal with both kinds of sequences (not shown) and in all amplifications with NUMT-directed primers. In addition, mixed NUMT and mtDNA sequences were obtained when using primers for mtDNA at 53°C. This was, in fact, the only situation where an accidental amplification of nuclear DNA was observed. The differences in the proportions of nuclear and mitochondrial DNA may, therefore, be responsible for the two different results: (a) a higher concentration of mtDNA in the first fraction makes it more difficult to amplify the NUMT, even with primers specifically designed for the nuclear pseudogene, while (b) in the second fraction, a high proportion of nuclear DNA enhances the amplification of the NUMT, even when mtDNA-directed primers are used.

Based on these results, we cannot exclude the hypothesis raised by Parr et al. [20] after experiments with ρ^0 cells in which they suggested that NUMTs with multiple copies throughout the genome are more prone to be accidentally amplified, although a very large number of copies would be needed to exceed the mtDNA copy number.

We therefore conclude that there is no significant risk of NUMT contamination in analyses with routine techniques. However, it is advisable to maximise primer specificity and to take into account the kind of samples used, the original tissue, and how DNA is extracted with each protocol.

Acknowledgements We wish to thank the donors of the samples used in this study. We are also very grateful to Marta Montesino for the technical support. This work was partially supported by Fundação para a Ciência e a Tecnologia through a research grant to A.G. (SFRH/BD/16518/2004) and by "Programa Operacional Ciência, e Inovação 2010" (POCI 2010), VI Programa-Quadro (2002–2006).

References

1. Bandelt HJ, Salas A, Lutz-Bonengel S (2004) Artificial recombination in forensic mtDNA population databases. Int J Legal Med 118:267–273
2. Pereira L, Richards M, Goios A et al (2006) Evaluating the forensic informativeness of mtDNA haplogroup H sub-typing on a Eurasian scale. Forensic Sci Int 159:43–50
3. Pereira L, Richards M, Goios A et al (2005) High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Res 15:19–24
4. Torroni A, Achilli A, Macaulay V, Richards M, Bandelt HJ (2006) Harvesting the fruit of the human mtDNA tree. Trends Genet 22:339–345
5. Wallace DC (1999) Mitochondrial diseases in man and mouse. Science 283:1482–1488
6. Herrnstadt C, Howell N (2004) An evolutionary perspective on pathogenic mtDNA mutations: haplogroup associations of clinical disorders. Mitochondrion 4:791–798
7. Goios A, Nogueira C, Pereira C, Vilarinho L, Amorim A, Pereira L (2005) mtDNA single macrodeletions associated with myopathies: absence of haplogroup-related increased risk. J Inherit Metab Dis 28:769–778
8. Davis RE, Miller S, Herrnstadt C et al (1997) Mutations in mitochondrial cytochrome c oxidase genes segregate with late-onset Alzheimer disease. Proc Natl Acad Sci U S A 94:4526–4531
9. Wallace DC, Stugard C, Murdock D, Schurr T, Brown MD (1997) Ancient mtDNA sequences in the human nuclear genome: a potential source of errors in identifying pathogenic mutations. Proc Natl Acad Sci U S A 94:14900–14905
10. Hirano M, Shtilbans A, Mayeux R, Davidson MM, DiMauro S, Knowles JA, Schon EA (1997) Apparent mtDNA heteroplasmy in Alzheimer's disease patients and in normals due to PCR amplification of nucleus-embedded mtDNA pseudogenes. Proc Natl Acad Sci U S A 94:14894–14899
11. Bravi CM, Parson W, Bandelt HJ (2006) Numts revisited. In: Bandelt HJ, Macaulay V, Richards M (eds) Human mitochondrial DNA and the evolution of *Homo sapiens*. Springer, Berlin, pp 31–46
12. Thangaraj K, Joshi MB, Reddy AG, Rasalkar AA, Singh L (2003) Sperm mitochondrial mutations as a cause of low sperm motility. J Androl 24:388–392
13. May-Panloup P, Chrétien MF, Savagner F, Vasseur C, Jean M, Malthiery Y, Reynier P (2003) Increased sperm mitochondrial DNA content in male infertility. Hum Reprod 18:550–556
14. Pereira L, Gonçalves J, Goios A, Rocha T, Amorim A (2005) Increased sperm mitochondrial DNA content in male infertility. Int J Androl 28:241–247
15. Roth DB, Porter TN, Wilson JH (1985) Mechanisms of nonhomologous recombination in mammalian cells. Mol Cell Biol 5:2599–2697
16. Woischnik M, Moraes CT (2002) Pattern of organization of human mitochondrial pseudogenes in the nuclear genome. Genome Res 12:885–893
17. Zischler H, Geisert H, von Haeseler A, Paabo S (1995) A nuclear 'fossil' of the mitochondrial D-loop and the origin of modern humans. Nature 378:489–492
18. Mishmar D, Ruiz-Pesini E, Brandon M, Wallace DC (2004) Mitochondrial DNA-like sequences in the nucleus (NUMTs): insights into our African origins and the mechanism of foreign DNA integration. Hum Mutat 23:125–133
19. Ricchetti M, Tekaia F, Dujon B (2004) Continued colonization of the human genome by mitochondrial DNA. PLoS Biol 2:E273
20. Parr RL, Maki J, Reguly B et al (2006) The pseudo-mitochondrial genome influences mistakes in heteroplasmy interpretation. BMC Genomics 7:185–197
21. Goios A, Amorim A, Pereira L (2006) Mitochondrial DNA pseudogenes in the nuclear genome as possible sources of contamination. International Congress Series 1288. Elsevier, Amsterdam, pp 697–699
22. Bensasson D, Zhang D, Hartl DL, Hewitt GM (2001) Mitochondrial pseudogenes: evolution's misplaced witnesses. Trends Ecol Evol 16:314–321

23. Behar DM, Metspalu E, Kivisild T et al (2006) The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. Am J Hum Genet 78:487–497
24. Macaulay V, Hill C, Achilli A et al (2005) Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. Science 308:1034–1036
25. Montesino M, Salas A, Crespillo M et al (2006) Analysis of body fluid mixtures by mtDNA sequencing: An inter-laboratory study of the GEP-ISFG working group. Forensic Sci Int 168:42–56
26. Álvarez-Iglesias V, Jaime JC, Carracedo A, Salas A (2007) Coding region mitochondrial DNA SNPs: targeting east Asian and Native American haplogroups. Forensic Sci Int Genetics 1:44–55