Int J Legal Med (2010) 124:133–142
DOI 10.1007/s00414-009-0385-0

ORIGINAL ARTICLE

Evaluating length heteroplasmy in the human mitochondrial DNA control region

Lucy Forster · Peter Forster · Susan M. R. Gurney · Matthew Spencer · Christopher Huang · Arne Röhl · Bernd Brinkmann

Received: 28 July 2009 / Accepted: 16 October 2009 / Published online: 24 November 2009

Abstract We present allelic data for three known and one new C-tract in the human mitochondrial DNA (mtDNA) control region, and we measure intergenerational mutation rates at such C-tracts. In detail, in a sample of 1,172 mtDNA sequences, we demonstrate the existence of an instability threshold of eight consecutive cytosines, at and above which the phenomenon of length heteroplasmy arises. To determine mutation rates, we draw on mtDNA sequences in up to four generations of 248 pedigrees for families living in high or low-radiation environmental conditions. The high-radiation sample gives the most conservative (fastest) mutation rate likely to be encountered in any forensic context. We find that the C-tract mutation rate is up to 6% per generation, and we observe an excess of cytosine gains over losses. Case studies and guidelines for evaluating mtDNA heteroplasmy are provided.

Keywords C-stretch · D-loop · Displacement loop · Indel · Radioactivity · Mutagenesis · Kerala · Thorium · Monazite

L. Forster · P. Forster · A. Röhl
Institute of Legal Medicine, University of Münster,
Röntgenstrasse 23,
48149 Münster, Germany

P. Forster (✉)
Cambridge Society for the Application of Research,
Churchill College,
Cambridge CB3 0DS, UK
e-mail: pf223@cam.ac.uk

P. Forster · A. Röhl
Genetic Ancestor Ltd.,
PO Box 503, Cambridge CB1 0AN, UK

S. M. R. Gurney
Institute of Continuing Education, University of Cambridge,
Madingley Hall,
Cambridge CB23 8AQ, UK

M. Spencer
School of Environmental Sciences, University of Liverpool,
Liverpool L69 7ZB, UK

C. Huang
Physiological Laboratory, University of Cambridge,
Downing Street,
Cambridge CB2 3EG, UK

B. Brinkmann
Institute of Forensic Genetics,
Im Derdel 8,
48161 Münster, Germany

Introduction

In a high-profile case, the remains of the Russian Tsar and his family, shot in 1918, were recovered and subjected to forensic mitochondrial DNA (mtDNA) analysis for identification [10]. A comparison of Tsar Nicholas's mtDNA with that of the present Duke of Fife (who share a matrilineal ancestor in Louise of Hesse-Cassel) revealed that the dead Tsar and his living relative had similar mtDNA, except at a single nucleotide position, where the Tsar's bone sample had a mixture of matching T and mismatching C bases, a phenomenon called point heteroplasmy. This initially was a complication in the identification process, until the same heteroplasmic point mutation was discovered in the Tsar's brother, Georgij Romanov [13]. Thus, from a forensic point of view, the occurrence of heteroplasmy can be seen either as a nuisance which is best disregarded in the statistical analysis, or as valuable additional discriminatory information.

In this manuscript, we wish to focus on a related but more commonly encountered type of mitochondrial polymorphism, namely length heteroplasmy. This is, to our

knowledge, so far, the only type of polymorphism capable of genetically distinguishing between identical twins or triplets [17], or between a mother's mtDNA and her offspring's mtDNA [8].

The mtDNA control region in humans is approximately 1,122-bp long and comprises four regions with repetitive tracts of consecutive cytosines (Fig. 1). Such "C-tracts" are prone to length heteroplasmy; in other words, one individual can have various mtDNA lengths in his or her cells which differ by the number of C's at these C-tracts [3, 20]. Research on the cellular mechanisms and circumstances causing heteroplasmy is ongoing [8, 12, 14, 16, 19]. These C-heteroplasmies are not rare, but indeed are typical in the majority of casework.

At present, the usual practice both in forensics and in human evolutionary studies is to disregard heteroplasmy, because the mechanism of heteroplasmy is unclear, the intergenerational mutation rate at which it occurs is unknown, the notation of heteroplasmy can be complex [2] and even its presence difficult to detect if the researcher does not resort for example to bacterial cloning [15]. In order to exploit heteroplasmy for forensic identification and anthropological work, it would therefore be desirable to understand the mechanisms and mutation rates of human mtDNA length heteroplasmy. In this paper, we use a pedigree data set comprising 248 families to elucidate the mutation mechanism and the mutation rates of length heteroplasmy, and we present worked examples to assist the forensic practitioner.

Subjects and methods

The data we use here draws on our previously published mtDNA control region sequences (hypervariable regions I and II) from documented pedigrees comprising individuals living in high-radiation and low-radiation coastal areas of Kerala [7, 8]. We expect that the high-radiation sample gives the most conservative (fastest) mutation rate likely to be encountered in a forensic context. This high- and low-radiation "coastal" data set comprises a sample size of 1012 mtDNA individuals covering 802 mtDNA transmissions in 248 pedigrees. Furthermore, in a control population of 151 "inland" Keralese, we have sequenced the C-tract around np460 (numbering according to [1]). The extraction and sequencing protocols, and most of the raw data themselves, with the exception of the np460 sequencing results, are detailed in the electronic PhD thesis of Forster [8], available at the University of Münster at http://miami.uni-muenster.de/index.html. Furthermore, we selected nine sequences with relevant alleles: six mtDNA sequences from genealogical casework processed by Genetic Ancestor Ltd (Cambridge, United Kingdom), and three sequences from Bali and Java [9]. DNA extraction, PCR amplification, gel purification, bacterial cloning, and dye terminator sequencing was performed as described in Forster et al. [7] and Forster [8]. To separate true mutant families from any undiscovered adoptions, maternity testing was performed by autosomal nonaplex, and where statistically necessary to achieve a maternity probability of >98%, by the highly discriminative locus ACTBP2. A total of 26 of the 40 cases of mutant second C-tracts were confirmed to be true maternal relatives, and these and the remaining 14 cases were confirmed by their matching mtDNA control region sequence in the range nps15990–16569 and nps35–465, revealing that no adoptions were present. There is, incidentally, a high proportion of 16189C alleles in our sample due to the preponderance of mtDNA type U1a [7].

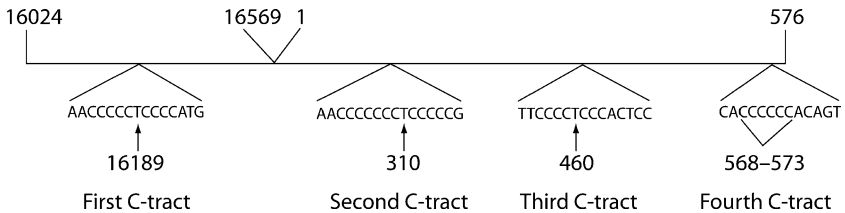

Fig. 1 Four human mitochondrial C-tracts susceptible to length heteroplasmy. Nucleotide positions (nps) are numbered according to Anderson et al. [1].The mtDNA control region ranges from nps16024–00576, and within this the displacement loop (D-loop) ranges from approx. nps16080–00191 [1], and the preceding RNA primer from nps00192–00407 [4, 6]. *First C-tract at nps16184–16193*: this region generally becomes a C-tract in individuals who show the mutation T16189C. *Second C-tract at nps303–309*: this long C-tract is prone to heteroplasmy even without the rare T310C mutation, while the short C-tract after np311 is stable and usually contains 6C's. An important exception is the woman sequenced for the Cambridge reference sequence [1], who has only 5C's at nps311–315, hence most other humans are denoted as having 315.1C. *Third C-tract at nps456–463*: similar to the first C-tract, heteroplasmy can be generated here by the mutation T460C. *Fourth C-tract at nps568–573*: normally containing six cytosines, heteroplasmy was observed here for tracts of 8–12 cytosines [20]

Results

Instability threshold of eight cytosines for mtDNA C-tracts

By visual inspection of each chromatogram in the mtDNA sequence from nps15990–16569 and nps35–465, we observe in our data set the presence of three cytosine-rich repetitive regions which have developed heteroplasmy in one or more individuals. Two of these repetitive regions are the well-known C-tracts at nps16184–16193 and nps303–309. In addition, we observe a novel heteroplasmic C-tract at nps456–463 (Fig. 1). We will refer to these as the first, second, and third C-tracts, respectively. A fourth heteroplasmic C-tract at nps568–573 has already been described by Torroni et al. [20].

Our overall impression was that shorter cytosine tracts are stable and homoplasmic, whereas longer C-tracts are always associated with length heteroplasmy. Examining this idea in detail, we observe, as have others, that for the first C-tract, longer C-tracts are created by the transition T16189C, affecting 206 out of 248 families, and we find that all 206 are heteroplasmic. In the second C-tract at nps303–309, we observe that 173 out of 248 families are heteroplasmic. The third C-tract at nps456–463 is rarely heteroplasmic, and we observe length heteroplasmy here in none of the 248 coastal families, and in only two of the 151 unrelated inland individuals, who both show the mutation T00460C, yielding a continuous C-tract. The rarity of this mutation (two of 399 unrelated individuals) probably explains why heteroplasmy at the third C-tract has not been reported before. The fourth C-tract at nps568–573 becomes heteroplasmic at eight and more consecutive cytosines [20].

To identify whether there is a general C-tract length threshold above which heteroplasmy occurs, we compiled the (dominant) C-tract lengths of each individual and scored the occurrence of heteroplasmy for the first, second, and third C-tracts from our data (Tables 1, 2, and 3).

We conclude that heteroplasmy is largely absent at a length of seven cytosines or less, and borderline heteroplasmy occurs at lengths of eight cytosines and more (Fig. 2). It appears that the instability threshold of 8C's applies to all regions of the mtDNA control region.

An important distinction between the first and the second C-tract with regards to length heteroplasmy is that HV1 length heteroplasmy affects the preceding A-tract. For anthropological studies, it is common practice to eliminate this A-tract information in any analysis. The implicit assumption probably is that the mutation rate of A16182C and A16183C is too fast between generations to be of reliable use. In the next section, we will look at such mutation rates in detail.

Mutation rates within the first C-tract

When screening heteroplasmic C-tracts for new mutations by comparing mothers and their daughters, any observed shift in the predominant allele length can be due to two quite different "mutation" mechanisms: "mutation by drift" within the cell's population of existing mtDNA alleles, enhanced by a bottleneck [16], as opposed to mutation by nucleotide insertion or deletion. For the first C-tract, we cannot distinguish these two underlying causes, and we simply record the observed shifts in the dominant allele between mothers and their children (Table 4) to arrive at a forensically applicable "observed" mutation rate.

In addition to the C-tract information shown in Table 4, we have inspected the pedigrees for changes in the short but variable A-tract (Fig. 1) immediately preceding the C-tract. The normal allele in most populations is AAAA, but in the wake of the T16189C mutation, this often shortens to AAA or AA. We have many pedigrees with these short alleles and find that AA never lengthens to AAA or to the common AAAA allele in our sample (n=440 transmissions in families with 16189C). In other words, a child having a longer A-tract than the putative mother is unlikely to be her biological child.

We can observe the shortening of the A-tract happen in one of our pedigrees, in the final stages. Among our sample of 107 pedigrees with 16189T, three have the mutation T16189C. In one of these (family 83 in [8]), both the AAAA-16189T minority allele and the AA-16189C majority allele are initially present in the grandmother. In the daughter, the AAAA-16189T minority allele finally disappears, leaving the pure AA-16189C allele. This case directly documents the frequently observed co-occurrence of the A-tract shortening with the T16189C mutation.

Mutation rates within the second C-tract

For the first C-tract, we stated it is generally not possible to distinguish "mutation by drift" within the cell's population of existing mtDNA alleles on the one hand, from mutation by nucleotide insertion or deletion on the other hand. However, in the second C-tract we can make this distinction in a few pedigrees. This is because in some mutant families, a nucleotide insertion or deletion in the second C-tract will push a genotype above or below the heteroplasmy threshold of eight cytosines, causing a striking change in the appearance of the chromatogram. In most forensic work, however, we have to be content to score whether or not there has been a shift in the predominant allele, irrespective of the underlying mechanism. In the following, we will discuss the pedigree in Fig. 3 as an example for macroscopic allele shift and also as an example for definite biochemical insertion/deletion

Table 1 First mtDNA C-tract (nps16182–16193): predominant allele length and status in 545 individuals

Length	High radiation[a]	Low radiation[a]	Bali[b]	Casework[c]	Status
2C	1	0			Homoplasmic
3C	5	4			Homoplasmic
4C	1	0			Homoplasmic
5C	85	20			Homoplasmic
6C	0	0			Not observed
7C	0	0	0	3[d]	Homoplasmic
8C	0	0	1	2	Heteroplasmic
9C	0	0	2	1	Heteroplasmic
10C	1	0			Heteroplasmic
11C	1	0			Heteroplasmic
12C	271	128			Heteroplasmic
13C	15	7			Heteroplasmic
Total	380	159	3	6	

[a] Representative pedigree sample from [7, 8]

[b] Selected alleles from [9]

[c] Selected alleles from Genetic Ancestor Ltd., Cambridge

[d] Those with 16186T–16189C

Table 2 Second mtDNA C-tract (nps303–309): predominant allele length and status in 250 individuals

Length	High radiation[a,b]	Low radiation[a]	Status
3C	0	1	Homoplasmic
4C	0	0	Not observed
5C	0	1	Homoplasmic
6C	8	0	Homoplasmic
7C	60	11	Homoplasmic, except for 4 individuals
8C	63	31	Heteroplasmic, except for 6 individuals
9C	48	24	Heteroplasmic
10C	3	0	Heteroplasmic
Total	182	68	

[a] Representative sample of unrelated individuals from [7, 8]

[b] Includes two adopted children not used for the pedigree study

Table 3 Third mtDNA C-tract (nps456–463): predominant allele length and status in 691 individuals

Length	High radiation[a]	Low radiation[a]	Inland[b]	Status
3C	32	16	0	Homoplasmic[c]
4C	150	68	149	Homoplasmic
5C	0	0	0	Not observed
6C	0	0	0	Not observed
7C	0	0	1	Homoplasmic[d]
8C	0	0	0	Not observed
9C	0	0	0	Not observed
10C	0	0	1	Heteroplasmic[d]
Total	380	160	151	

[a] Representative pedigree sample from [7, 8]

[b] Representative sample of unrelated individuals

[c] Those with 456T

[d] Has 460C

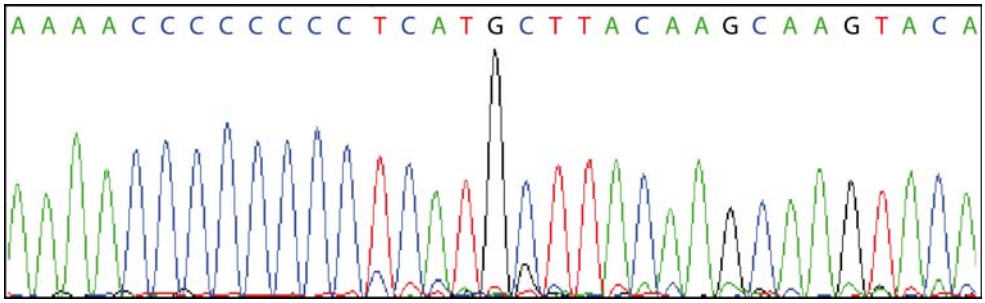

Fig. 2 Borderline heteroplasmy for allele length 8 in the first C-tract. Heteroplasmy is visible as small shoulder peaks to the right of each major peak after the last C in the C-tract. In this first C-tract, the same instability threshold at 8C's applies as for the second C-tract, but as can be seen in this figure and in our other cases (not shown), the resulting length heteroplasmy in the first C-tract is hardly noticeable for allele length 8

events, and we will quantify the two mutation rates in Table 5 as separate sections Introduction and Subjects and methods.

1. Observed "shifts in predominant allele" (potential mutations) are summarized in Table 5 and exemplified in Fig. 3. The mother (I.1) has predominantly 9C in her mtDNA, but also some mtDNA molecules with 10C and 11C (as can be seen by the multiple minor T peaks at np310). One daughter (II.1) also has predominantly 9C, but in her, the proportion of 10C has increased (the two T peaks are now nearly the same height). If the proportion of 10C had increased even more then 10C would have become the predominant molecule in the daughter (II.1) and, in forensic notation, the daughter would differ from the mother by a "mutation", even though an insertion or deletion has probably not taken place.

According to this forensic "mutational" notation, we observed in 600 high-radiation transmissions that mothers and offspring differed in 35 cases by their modal C-length in the C-tract between np303 and np309 (Appendix 2 in [8]). Since in some cases the offspring's modal C differed from the mother's by not only a single nucleotide position but by two positions or even four positions, the total number of shifted C-positions in the irradiated families is 42 modal C shifts in 600 transmissions. In 202 low-radiation transmissions, only five modal C-length differences between mother and offspring were observed (Table 5).

This is an approximately threefold higher rate in the irradiated families, which at first glance appears significantly higher ($p<0.01$). However, more intermediate generations of family members were sequenced in the high-radiation families than in the low-radiation families. This extended sequencing revealed two length changes (shortening in the daughter and then lengthening in the granddaughter) in radioactive family 159, which need to be subtracted before comparing the mutation rates in irradiated families and low-radiation families. The significance of the difference (40 changes in 600 transmissions versus five changes in 202 transmissions) then falls slightly below the 99% confidence level, but still well above the 95% confidence level. Biologically speaking, it is interesting that radiation not only increases hereditary point mutations as we have previously shown [7], but also increases changes in allele-length proportions in an individual.

2. Mutations affecting the heteroplasmy/homoplasmy status (definite mutations) are also listed in Table 5. In such cases, a nucleotide insertion or deletion mutation is clearly distinguishable from shifts in proportions of existing mtDNA types in the cell. This happens when a nucleotide insertion or deletion will push a genotype above or below the heteroplasmy threshold of eight cytosines, causing a striking change in the appearance of the chromatogram (Fig. 3). In our pedigree samples covering 802 transmissions, three such cases are available, all of these in the irradiated families (Table 5).

It may be argued that the presence of these three definite DNA mutations in the irradiated families versus zero definite DNA mutations in the low-radiation families further supports the hypothesis of a mutating effect of radiation. However, inspection of Table 2 reveals that the high- and the low-radiation families differ considerably in their allele profile, with on average longer alleles in the low-radiation sample. There is, therefore, more opportunity to observe definite mutations in alleles of the irradiated families.

Table 4 Mutation rates in first C-tract (nps16184–16193) based on shifts in predominant allele length in 440 transmissions

[a] Only families with 16189C included.

	High radiation[a]	Low radiation[a]
Families witnessing allele shifts	12 of 93 families	6 of 48 families
Children differing from their mother	12 in 314 transmissions	7 in 126 transmissions
Cytosines gained and lost after shift	12 gains, 0 losses	6 gains, 1 loss

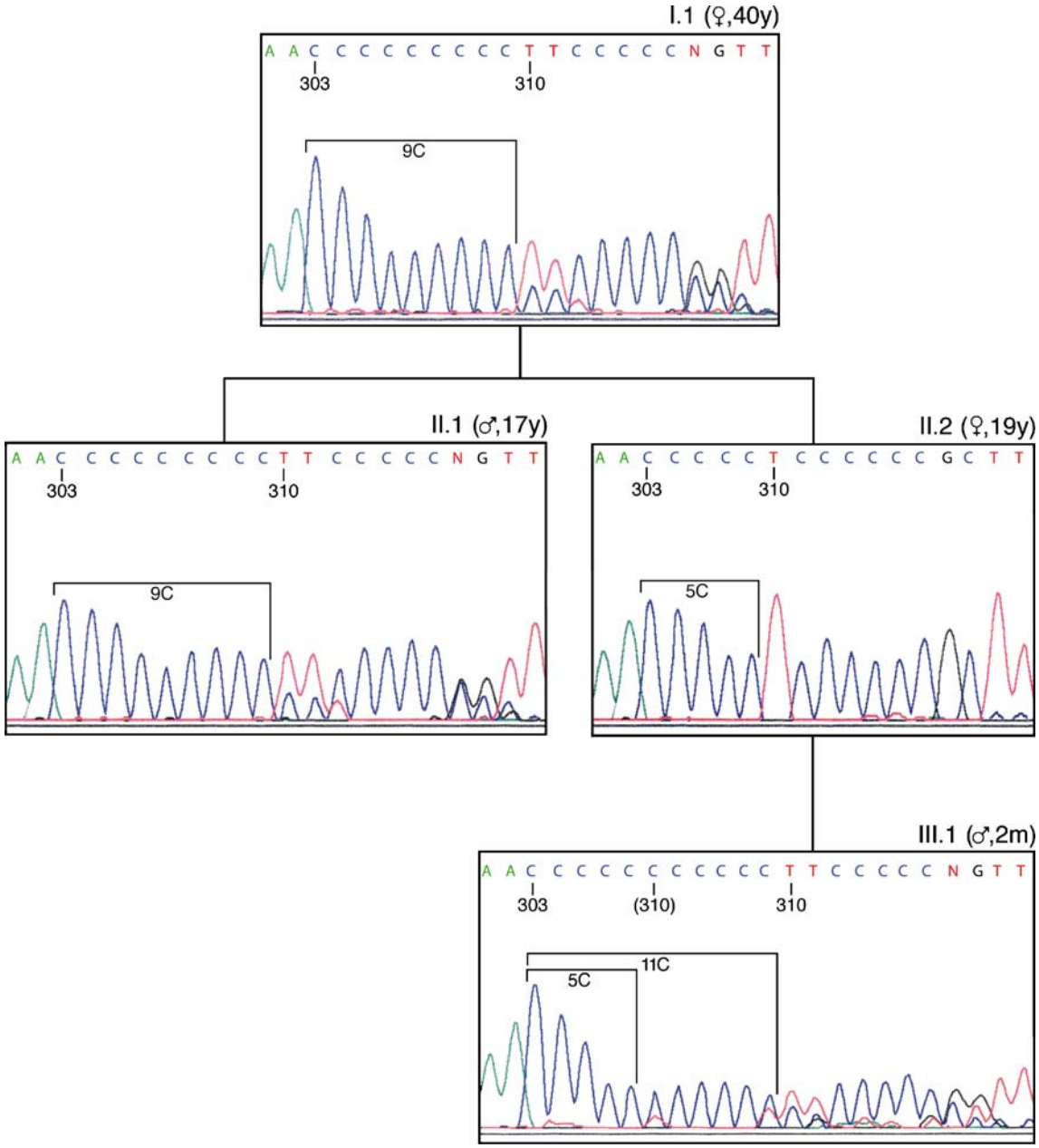

Fig. 3 Inheritance of normal and unusual length variation at nps303–309. In family 164, the mother (I.1) has various mtDNA types with 9, 10, and 11 cytosines. The simultaneous presence of several lengths causes the single T peak at np310 to appear as multiple virtual T peaks. In the mother, the highest T peak occurs after 9C, indicating that the 9C allele is predominant. In one daughter (II.1), the proportions of length variants change slightly (the second virtual T peak, indicating a tract of ten cytosines, is now nearly as high as the first T peak). The other daughter (II.2) shows a remarkable deletion of four cytosines in her saliva sample and no length heteroplasmy at all. The granddaughter (III.1) seems to have inherited the deletion as a minor variant, whereas most of her mitochondria have 11 cytosines

Point mutation rate versus length mutation rate

To compare qualitatively the in vivo mutation speed of insertions/deletions at nps303–309 with mtDNA point mutations, one individual (149.2) who was heteroplasmic at nps303–309 was measured by bacterial cloning (Appendix 7 in [8]). Four length variants were detected: 4% 7C, 75% 8C, 17% 9C, and 4% 10C. This same individual had three different mtDNA types due to new mutations at positions 144 and 152. No one-to-one correlation was seen between the three point mutation mtDNA types (144C–152T, 144T–152C, 144C–152C) and the four length variants. This strongly suggests that at nps303–309, insertions/deletions mutate faster than even the fastest point mutation np152 (Table 4 in [7]) within the mtDNA sections nps16093–16390 and nps35–315.

Table 5 Mutation rates in second C-tract (nps303–309) based on 802 transmissions

	High radiation	Low radiation
1. Shifts in predominant allele (potential mutations)		
Families witnessing allele shifts	24 of 180 families	5 of 68 families
Children differing from their mother	35 in 600 transmissions	5 in 202 transmissions
Differing by 1-nucleotide shifts	30 in 600 transmissions	5 in 202 transmissions
Differing by 2-nucleotide shifts	4 in 600 transmissions	0 in 202 transmissions
Differing by 3-nucleotide shifts	0 in 600 transmissions	0 in 202 transmissions
Differing by 4-nucleotide shifts	1 in 600 transmissions	0 in 202 transmissions
Total number of shifted C-positions	42 in 600 transmissions	5 in 202 transmissions
Cytosines gained and lost after shift	30 gains, 11 losses, 1 uncertain	4 gains, 1 loss
2. Mutations affecting homoplasmy (definite mutations)		
	3 in 600 transmissions Details: Family 83, homo→het Family 164[a], het→homo→het Family 171a, homo→het	0 in 202 transmissions

[a] See Fig. 3

Biological caveat for the observed mutation rates

Although forensically applicable, the C-tract mutation rate analyses have a shortcoming for those researchers interested in the biological basis of the observed pedigree mutations. This is the inability to distinguish between somatic and germline mutations. The ancestral state of a length allele (7C, 8C, …) cannot be inferred from an evolutionary mtDNA tree because our mutation rate data show that length alleles, unlike point mutations, have evolved back and forth too fast for an mtDNA tree to capture. And with few exceptions, the families in this study are too shallow to follow an ancestral allele through at least two generations (to confirm that it is the germline type) and then to follow the mutated allele another two generations (to confirm that the mutated allele is in the germline). Only mutant families 86 and 164 happen to fulfill these requirements. In family 86, we can confirm a germline length mutation, whereas in family 164 the complicated length mutations are probably at least partly explained by somatic mutation, pending further cloning experiments. This biological caveat does not directly affect forensic applications, but provides a useful point of departure for further basic biological research based on mtDNA pedigrees.

Case studies

First C-tract

We are fortunate to observe a possible case of nascent length heteroplasmy in one of our pedigrees (family 186), where the mother has 16189T and is homoplasmic, and her daughter has developed 16189C, leading to a C-tract of ten cytosines and concomitant length heteroplasmy (Fig. 4). It is alternatively possible that the mother already has 16189C as a barely detectable minority type in her cells, although the other family members in this pedigree do not show evidence for this.

Whether or not this mutation happened in the mother or in a preceding generation, this point mutation (precipitating a length mutation in this case) is one of 23 point mutations that we have observed in 802 transmissions [7] in the mitochondrial control region. The forensically relevant message here is that we observe all our new point mutations to be, without exception, heteroplasmic in the first generation, as is evident in Fig. 2, individual II.3 at np16189. We turn now to length mutation differences between a putative mother and a child in maternity casework.

Second C-tract

Looking back at the pedigree case depicted in Fig. 3, there are two counterintuitive features that could cause problems in casework interpretation.

One striking feature in the chromatogram is the loss of several cytosines in one generation. Daughter II.2 has only the allele 5C, clearly shorter than the shortest allele 9C detectable in the mother I.1. The loss of 4 cytosines at one stroke is remarkable and in fact unique in our sample of 802 pedigree transmissions, so we performed autosomal maternity testing to exclude the possibility of an adopted child. We find that the maternity probability between

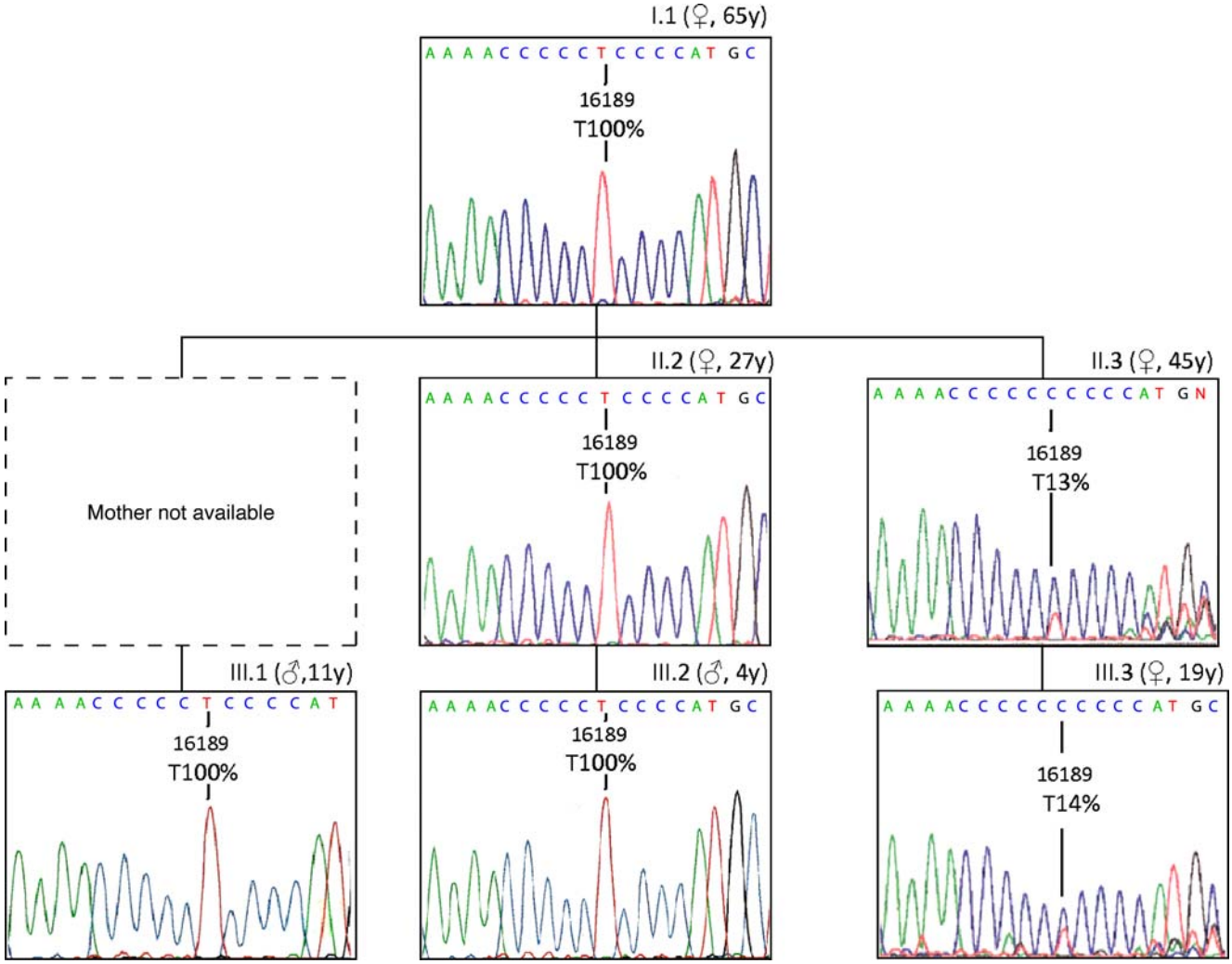

Fig. 4 Nascent length heteroplasmy in first C-tract due to mutation T16189C. Note that this mutation extends a stable tract of 5C's to an unstable tract of 10C's. Pedigree 186 from Forster [8]

grandmother and mother is 99.99%, and between the mother and daughter it is 99.38%, as determined by autosomal nonaplex typing.

Another striking feature in this case is that the reverse sequence at first glance suggests a discrepancy to the forward sequence, in that the reverse sequence does not show multiple thymine peaks. However, this apparent discrepancy can be rationalized as shown in Fig. 5.

Having thus interpreted the chromatograms, we can take the length mutation discrepancy between the mother and her child as an example for calculating maternity probabilities for cases where the maternity is in dispute and where there is an observed C-tract length difference between the putative mother and child. The first step is to determine a relevant comparative population. Circumstances indicate a local mother, and in a global database search [18] of 40,000 native mtDNA sequences, we find that the mtDNA type (in the range nps16051–16362 and 73–315) of the grandmother I.1 in family 164 is most commonly found in coastal southwestern India, where there are 20 exact matches out of a random sample of 248 unrelated individuals. Her daughter's mtDNA type, which differs by a loss of 4C's in the second C-tract, is observed once in the 248 local unrelated individuals, and is not observed elsewhere in the global database. Also, there are the following close matches with somewhat longer C-tracts, also restricted to this local population: 12 individuals differing from the child by three cytosines, and four individuals differing from the child by two cytosines. Assuming in the high-radiation area a mutation rate of one out of 1/600 transmissions for loss of four cytosines, one out of 1/600 for loss of three cytosines, and four out of 4/600 for a loss of two cytosines (Table 5), the probability to observe the child's sequence, given the putative mother is the true mother, is approximately 1/600= 0.17%. The probability to observe the child's sequence if the true mother is someone else is $(20/248) \times (1/600) +$

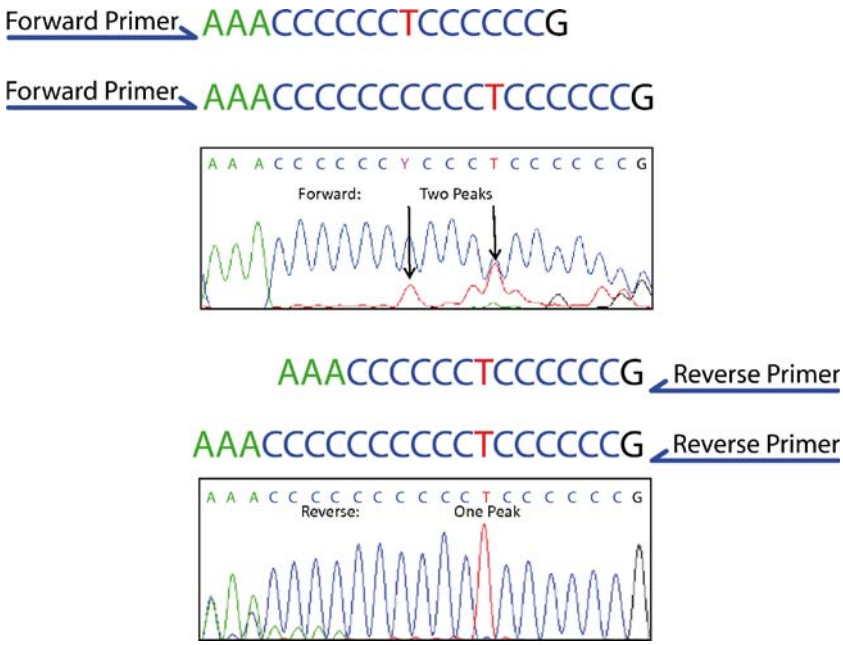

Fig. 5 Interpreting the apparent discrepancy between the forward and reverse sequences of heteroplasmy. This individual has heteroplasmy consisting of stable 6C's and variable 10C's, visible by virtual T peaks only in the forward sequence (*top*) and not in the reverse sequence (*bottom*). The explanation is that there is no second T, only invisibly superimposed C's. The situation is the same as for the unrelated case in Fig. 4, where individual III.1 appears to have contradictory chromatograms for forward (shown) and reverse (not shown) reactions

$(12/248) \times (1/600) + (4/248) \times (4/600) + (1/248) \times (1 - 35/600)$, approximately 0.41%. This calculation leaves aside, for the sake of simplicity, uncertainty in the frequencies and in the mutation rate. Hence, the mtDNA profile is about two to three times less likely to originate from the putative mother than from a randomly sampled woman represented by the database, which is hardly a conclusive result. So, in this particular case, an autosomal maternity test was additionally performed, which confirmed maternity with a probability of 99.99% [8], and therefore confirmed that the unusual 4C-mutation had indeed taken place. We note that this multiple deletion has immediately eliminated detectable heteroplasmy in the daughter, in contrast to the persistence of heteroplasmy in the 23 cases of point mutations we observed among the total of 802 transmissions.

Guidelines for handling C-tracts in the mtDNA control region

Based on the pedigree results presented here, we suggest the following guidelines for legal medical casework.

1. In maternity casework using mtDNA (where for example autosomal profiling is not possible), a point mutation difference between mother and child in the range nps15990–16569 and nps35–465 decreases the likelihood of maternity, but less so if point heteroplasmy is still visible. We argue this because in 802 transmissions, we have never observed a homoplasmic nucleotide substitution between mother and child, but we have observed 23 heteroplasmic nucleotide substitutions between mother and child.
2. If eight or more consecutive cytosines are present in the mtDNA control region, chromatograms should be inspected for heteroplasmy. In rare cases, seven cytosines can also be associated with heteroplasmy.
3. Forward and reverse sequencing may appear to contradict each other with regard to the presence of heteroplasmy. The apparent contradiction should be rationalized as exemplified in Fig. 5.
4. In heteroplasmic sequences, the dominant length variant should be determined by reference to a flanking peak in the chromatogram [11].
5. At nps16180–16183, an apparent lengthening of the alleles AA or AAA to the allele AAAA between a putative mother and her putative child decreases the likelihood of maternity, as we have never observed such lengthening in 802 transmissions.
6. In heteroplasmy of the first C-tract due to np16189C, the dominant length is 12 cytosines in 95% of cases, and we have observed occasional deviations from this length only in the youngest generation within our pedigrees. Pending further studies on selective or other distorting effects, we suggest length variation in the first C-tract is disregarded, as is current practice.
7. In contrast, the second C-tract (nps303–309) harbours much higher allelic diversity, and within our 802 transmissions we have observed both lengthening and shortening in this C-tract in up to 6% of transmissions in extreme natural radiation conditions. Allelic differences in maternity casework should take such mutation

rates into account, as exemplified in our second case study.

Discussion

In mitochondrial DNA, cytosine tracts above a certain length become unstable and then develop length heteroplasmy. The requisite length can be achieved in the second and fourth C-tract (Fig. 1) by insertion of one or more C's, or, as is typical in the first and third C-tract, by mutating an intervening T into a C, linking together two shorter C-tracts into one long tract. In our initial study [8], we found a stability threshold of eight cytosines in HV1 and HV2, independently confirmed by Irwin et al. [12]. Both we and they find occasional heteroplasmy at the second C-stretch with seven cytosines, and further pedigree research will show whether or not this is residual heteroplasmy from an ancestor who predominantly had eight or more cytosines before undergoing a deletion shortening the tract to predominantly seven cytosines. Based on the data presented so far, the instability threshold of eight cytosines appears to apply across the mtDNA control region for all four C-tracts. For avoidance of doubt, the instability threshold is not restricted to the displacement (D-) loop, as three of the C-tracts are outside the triple-stranded D-loop (Fig. 1). In other words, the instability threshold of eight cytosines does not appear linked to the presence or absence of the triple-stranded DNA. Finally, can the instability threshold for C-tracts be generalized to other tracts? Evidently not, as it has independently been reported [5, 12, 20] that the instability threshold for heteroplasmy in the CA-tract at nps514–523 is five repeats, representing a length of ten nucleotides.

Acknowledgments We thank Sabine Lutz-Bonengel (Freiburg) for bacterial cloning, Marianne Schürenkamp (Münster) for maternity tests, and two anonymous referees for improving the manuscript.

References

1. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJ, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–465
2. Bandelt HJ, Parson W (2008) Consistent treatment of length variants in the human mtDNA control region: a reappraisal. Int J Legal Med 122:11–21
3. Bendall KE, Sykes BC (1995) Length heteroplasmy in the first hypervariable segment of the human mtDNA control region. Am J Hum Genet 57:248–256
4. Chang DD, Clayton DA (1985) Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc Natl Acad Sci USA 82:351–355
5. Chung U, Lee HY, Yoo J-E, Park MJ, Shin K-J (2005) Mitochondrial DNA CA dinucleotide repeats in Koreans: the presence of length heteroplasmy. Int J Legal Med 119:50–53
6. Fish J, Raule N, Attardi G (2004) Discovery of a major D-loop replication origin reveals two modes of human mtDNA synthesis. Science 306:2098–2101
7. Forster L, Forster P, Lutz-Bonengel S, Willkomm H, Brinkmann B (2002) Natural radioactivity and human mitochondrial DNA mutations. Proc Natl Acad Sci USA 99:13950–13954
8. Forster L (2004) PhD thesis, University of Münster
9. Forster P, Vizuete-Forster M, Matsumura S, Blumbach P, Dewar R (2008) The genetic prehistory of Madagascar. In: Matsumura S, Forster P, Renfrew C (eds) Simulations, genetics and prehistory. McDonald Institute Press, University of Cambridge, UK
10. Gill P, Ivanov PL, Kimpton C, Piercy R, Benson N, Tully J, Evett I, Hagelberg E, Sullivan K (1994) Identification of the remains of the Romanov family by DNA analysis. Nature Genet 6:130–135
11. Hühne J, Pfeiffer H, Brinkmann B (1998) Heteroplasmic substitutions in the mitochondrial DNA control region in mother and child samples. Int J Legal Med 112:27–30
12. Irwin JA, Saunier JL, Niederstätter H, Strouss KM, Sturk KA, Diegoli TM, Brandstätter A, Parson W, Parsons TJ (2009) Investigation of heteroplasmy in the human mitochondrial DNA control region: a synthesis of observations from more than 5000 global population samples. J Mol Evol 68:516–527
13. Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ (1996) Mitochondrial DNA sequence heteroplasmy in the grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nature Genet 12:417–420
14. Lacan M, Thèves C, Amory S, Keyser C, Crubézy E, Salles J-P, Ludes B, Telmon N (2009) Detection of the A189G mtDNA heteroplasmic mutation in relation to age in modern and ancient bones. Int J Legal Med 123:161–167
15. Lutz-Bonengel S, Sänger T, Pollak S, Szibor R (2004) Different methods to determine length heteroplasmy within the mitochondrial control region. Int J Legal Med 118:274–281
16. Marchington DR, Macaulay V, Hartshorne GM, Barlow D, Poulton J (1998) Evidence from human oocytes for a genetic bottleneck in an mtDNA disease. Am J Hum Genet 63:769–775
17. Pfeiffer H, Lutz-Bonengel S, Pollak S, Fimmers R, Baur MP, Brinkmann B (2004) Mitochondrial DNA control region diversity in hairs and body fluids of monzygotic triplets. Int J Legal Med 118:71–74
18. Röhl A, Brinkmann B, Forster L, Forster P (2001) An annotated mtDNA database. Int J Legal Med 115:29–39
19. Santos C, Sierra B, Alvarez L, Ramos A, Fernandez E, Nogues R, Aluha MP (2008) Frequency and pattern of heteroplasmy in the control region of human mitochondrial DNA. J Mol Evol 67:191–200
20. Torroni A, Lott MT, Cabell MF, Chen Y-S, Lavergne L, Wallace DC (1994) mtDNA and the origin of Caucasians: identification of ancient Caucasian-specific haplogroups, one of which is prone to a recurrent somatic duplication, in the D-loop region. Am J Hum Genet 55:760–776