

T. Vanecek · F. Vorel · M. Sip

## Mitochondrial DNA D-loop hypervariable regions: Czech population data

Received: 25 March 2003 / Accepted: 29 September 2003 / Published online: 31 October 2003

© Springer-Verlag 2003

**Abstract** In order to identify polymorphic sites and to find out their frequencies and the frequency of haplotypes, the complete D-loop of mitochondrial DNA (mtDNA) from 93 unrelated Czech Caucasians was sequenced. Sequence comparison showed that 85 haplotypes were found and of these 78 were unique, 6 were observed twice and 1 was observed three times. Genetic diversity (GD) was estimated at 0.999 and the probability of two randomly selected sequences matching (random match probability, RMP) at 1.2%. Additionally these calculations were carried out for hypervariable regions 1, 2 (HV1, HV2), for the area between HV1 and HV2 and for the area of the hypervariable region HV3. The average number of nucleotide differences (ANND) was established to be 10.2 for the complete D-loop. The majority of sequence variations were substitutions, particularly transitions. Deletions were found only in the region where HV3 is situated and insertions in the same place and in poly-C tracts between positions 303 and 315 in HV2. A high degree of length heteroplasmy was found especially in the regions of poly-C tracts between positions 16184 and 16193 in HV1 and between positions 303 and 315 in HV2. Position heteroplasms were found in two cases.

**Electronic Supplementary Material** Supplementary material is available in the online version of this article at <http://dx.doi.org/10.1007/s00414-003-0407-2>

**Keywords** Mitochondrial DNA · Population study · Czech Republic

### Introduction

The analysis of human mtDNA has become a powerful tool for genetic characterization of forensic biological specimens. Several features of the mitochondrial genome, such as a partial protection against nuclease activity due to a circular molecule or the presence of thousands of copies per cell [1, 2] make it relatively convenient for the genetic analysis of highly degraded material where DNA typing with nuclear markers would not be successful, such as bones, teeth, hair shafts, buried material, faeces etc. [3, 4, 5, 6, 7, 8, 9, 10]. Additionally, the substitution rate of mtDNA (especially in the non-coding region) is markedly higher compared to most nuclear genes. This is probably caused by mtDNA polymerase mistakes, deficiencies in repair activities, high oxidant levels and lack of recombination [11, 12]. All these mechanisms increase the genetic variability of mtDNA and therefore increase the chance of individual identification. Another advantage is also the maternal inheritance [13, 14], which allows intergeneration comparison to determine maternal family relationships even in the case of several missing generations [15].

The non-coding highly variable region of human mtDNA called the D-loop is used for the purpose of forensic genetics. The D-loop of human mtDNA, which is approximately 1150 bp long, is situated between the mitochondrial tRNA<sup>Pro</sup> and tRNA<sup>Phe</sup> genes. It contains two sequence variable regions called HV1 (position 16024–16365) and HV2 (position 73–340), and one sequence and length variable region HV3 (position 438–574) with CA dinucleotide repeats. Furthermore, the origin of replication of the H strand (OriH), from which 7S DNA is synthesized

T. Vanecek  
Sikl's Department of Pathology, Charles University Hospital,  
Pilsen, Czech Republic

T. Vanecek  
Faculty of Biological Science, University of South Bohemia,  
Ceske Budejovice, Czech Republic

F. Vorel  
Department of Forensic Medicine, University Hospital,  
Ceske Budejovice, Czech Republic

M. Sip (✉)  
Department of Medical Physics and Biophysics,  
Faculty of Health and Social Studies University  
of South Bohemia,  
Dukelska 3, 37001 Ceske Budejovice, Czech Republic  
Tel.: +420-60-1273498, Fax: +420-38-7438389,  
e-mail: sip@zsf.jcu.cz

during the first stage of replication, is situated here [16, 17, 18, 19, 20, 21].

At present, the frequencies of mtDNA types in a population are quantified by counting the number of occurrences in the sequence database. Therefore, the availability of large sequence databases is necessary. In the last decade many population studies dealing with mtDNA D-loop polymorphisms have been published [22, 23, 24, 25, 26, 27, 28]. The Czech population data of the D-loop region presented in this study is another small piece in the big puzzle of the world human mitochondrial population data.

## Materials and methods

### DNA isolation

The total DNA from 93 randomly chosen unrelated Caucasians from the Czech population was isolated from blood samples by chelex extraction [29] or with a QIAamp DNA Blood Mini kit (QIAGEN, Hilden, Germany) according to manufacturer's protocol.

### Polymerase chain reaction (PCR)

A total of 10 ng of DNA was amplified using the primers (Generi Biotech, Hradec Kralove, Czech Republic) displayed in Table 1, which anneal to the area of the tRNA<sup>Pro</sup> and tRNA<sup>Phe</sup> genes and border the D-loop region. The PCR was carried out in a total volume of 50 µl consisting of 1×PCR buffer (containing 1.5 mM MgCl<sub>2</sub>), 200 µM of each dNTP, 0.1 µM of each primer and 2 U Taq polymerase (TAKARA, Shiga, Japan).

Amplification was performed on a Progene thermocycler (Technique, Cambridge, UK) using the following conditions: denaturation at 95°C for 2 min and then 30 cycles with denaturation at 95°C for 1 min, annealing at 55°C for 1 min and extension at 72°C for 1 min. The program was finished with elongation at 72°C for 7 min.

Successfully amplified PCR products were purified with a QIAquick spin PCR purification kit (QIAGEN, Hilden, Germany).

### Sequencing and electrophoresis

Sequencing reactions were carried out using the Big Dye Terminator Sequencing kit (PE/Applied Biosystems, Foster City, CA) on a Progene thermocycler (Technique, Cambridge, UK) with 50–200 ng

of each PCR product and 10 pmol of one of the sequencing primers (Table 2) (Generi Biotech, Hradec Kralove, Czech Republic) was added to the reaction for sequencing the complete D-loop region. The reaction conditions were as follows: denaturation at 96°C for 30 s and then 25 cycles with denaturation at 96°C for 30 s, annealing at 50°C for 15 s and extension at 60°C for 4 min.

The samples were precipitated and purified twice with isopropanol, resolved in template suppression reagent (PE/Applied Biosystems, Foster City, CA), denatured and run on an automated sequencer ABI Prism 3100 Avant (PE/Applied Biosystems, Foster City, CA) at a constant voltage of 12.2 kV for 180 min.

### Data analysis

Analysis of sequences was carried out using ABI Prism sequence software (PE/Applied Biosystems, Foster City, CA) and sequences were aligned using the BioEdit program [30]. The genetic characteristics of the D-loop were calculated according to Tajima [31] and Stoneking et al. [32]. Statistical analysis of the D-loop region was performed using the DNA SP v3 program [33].

## Results and discussion

All sequences were manually checked in ABI Prism sequence software (PE/Applied Biosystems) and whole D-loop consensus sequences were created. Sequences were aligned by the BioEdit program [30] and compared with the reference sequence [16] (comparison table see ESM). For better utilization of all obtained data, several new terms were established in our paper. The region spanned by 7S DNA during replication of the H-strand, situated between positions 16366 and 72, was named 7S DNA spanned region (7S-SP) and the region between positions 341 and 574 containing HV3 was called HV3 extended region (HV3ex). All calculations were then made with regions HV1, 7S-SP, HV2 and HV3ex.

After the comparison of sequences of the complete D-loop region, 85 haplotypes (91.4%) were found. From these, 78 sequences (83.9%) were observed only once in the Czech population, 6 were observed twice (12.9%) and 1 was observed 3 times (3.2%). These results are slightly different from those given in the paper of Lutz et al. [20], where 95% had different haplotypes and 93% were unique sequences.

The most abundant haplotype (16519C, 263G, 315.1C) in comparison to Anderson et al. [16] was observed in 3.2% of samples, which is also the typical haplotype found by Lutz et al. [20]. When we focused only on the regions HV1 and HV2, type 263G, 315.1C was observed in 4.3% of cases and is also the typical haplotype found in other Caucasian populations (US Caucasians 4.3% [22], Spanish 5% [23], Swiss 2.6% [24], Austrians 3% [25], Germans 2% [34], British 4.5% [26], French 4% [27]).

Based on the observed frequencies of mitochondrial D-loop haplotypes and from data from each single area of this region, i.e. HV1, 7S-SP, HV2, HV3ex and HV1 together with HV2, genetic diversity [31] and random match probability [32] were counted (Table 3). The results for the complete D-loop and for HV1 together with HV2 regions are in a good accordance with values obtained by other authors for large populations [20, 22, 35], while the data for

**Table 1** Mt DNA amplification primers

Primer	Sequence
L15996	5'-CCA CCA TTA GCA CCC AAA GC-3'
H583	5'-GCT TTG AGG AGG TAA GCT AC-3'

**Table 2** Mt DNA sequencing primers

Primers	Sequence
L15996	5'-CCA CCA TTA GCA CCC AAA GC-3'
L27	5'-CAG GTC TAT CAC CCT ATT AAC-3'
H7	5'-GTG GTT AAT AGG GTG ATA GAC-3'
H409	5'-ACT GTT AAA AGT GCA TAC GCC-3'
H583	5'-GCT TTG AGG AGG TAA GCT AC-3'
H16401	5'-TGA TTT CAC GGA GGA TGG TG-3'

**Table 3** Genetic diversity, random match probability and average number of nucleotide differences in HV1, 7S-SP, HV2, HV3ex, HV1 and HV2 together<sup>a</sup> and D-loop region of mtDNA of 93 Czech Caucasians

	HV1	7S-SP	HV2	HV3ex	HV1HV2 <sup>a</sup>	D-loop
Genetic diversity	0.984	0.711	0.974	0.659	0.996	0.999
Random match probability	2.7%	29.6%	3.6%	34.8%	1.49%	1.2%

**Table 4** Average number of nucleotide differences in HV1, 7S-SP, HV2, HV3ex, HV1 and HV2 together <sup>a</sup> and D-loop region of mtDNA of 93 Czech Caucasians

	HV1	7S-SP	HV2	HV3ex	HV1HV2 <sup>a</sup>	D-loop
Average number of nucleotide differences	5.29	1.08	2.97	0.83	8.26	10.16

**Table 5** Observed nucleotide substitution, insertion, deletion and heteroplasmic events found in HV1, 7S-SP, HV2, HV3ex, HV1 and HV2 together <sup>a</sup> and D-loop region of mtDNA of 93 Czech Caucasians in comparison to reference Anderson et al. sequence [16]

Nucleotide	HV1	7S-SP	HV2	HV3ex	HV1HV2 <sup>a</sup>	D-loop
A>G	29	6	149	3	178	187
G>A	12	5	23	3	35	42
T>C	106	50	62	17	168	234
C>T	114	7	22	16	136	158
A>T	4	0	1	0	5	5
A>C	7	0	0	0	7	7
G>T	0	0	0	0	0	0
G>C	0	1	1	1	1	3
C>A	1	0	0	0	1	1
C>G	0	1	1	1	1	3
T>A	0	1	0	0	0	1
T>G	0	1	1	0	1	2
I Insertion.	D	0	0	19	0	19
D Deletion.	I	0	0	138	24	162
LH Length heteroplasmy.	LH	8	0	21	0	30
PH Position heteroplasmy.	PH	2	0	0	0	2
NPS Number of polymorphic sites (including gaps).	NPS	76	22	42	27	167

small isolated populations exhibited lower values of genetic diversity and higher values of random match probability [36]. Surprisingly, data of the HV3ex region exhibited higher RMP and lower GD than one would expect, i.e. they showed lower variability than other authors observed even if they used only HV3 region in their calculations [37].

The degree of polymorphism within the Czech population, i.e. the average number of nucleotide differences between individuals for each region and complete D-loop region as shown in Table 4 are not markedly different from other European Caucasian studies such as the French, British, German etc. [22].

Table 5 displays the number of differences in the complete D-loop and separately in HV1, 7S-SP, HV2, HV3ex regions of the Czech population data set in comparison with the Anderson et al. reference sequence [16]. Substitution, deletion and insertion mutations were found in the whole D-loop region at 167 positions. Substitutions were observed at 151 positions with a total of 645 differences. From these, 96.6% were transitions, mainly T to C and C to T (63.3%), which is in accordance with other studies even if they analyzed only HV1 and HV2 regions [22, 23,

24, 25, 26, 27]. Transversions were very rare, they were randomly spread over the whole region and there was a slight predominance of A to C and A to T mutations (Table 5 and ESM).

Most of the substitutions were distributed randomly but among them there were several sites which exhibit, as in other population studies [22, 23, 24, 25, 26, 27], a higher percentage of changes in comparison to the reference sequence [16]. These are particularly the already mentioned sites 16519C, which was observed in 48.3% of samples, 263G observed in 92.4 %, 16126C observed in 24.7%, 73G observed in 59.1% and 195C observed in 21.5%. All exhibited more than 20% change in comparison to the reference Anderson et al. sequence [16] (for more details see ESM).

Deletions were found at three positions and were only found in the HV3ex region and particularly in the CA repeats where a dinucleotide deletion at positions 522 and 523 was found in 9 cases. Additionally, the deletion of a C at position 527 was observed in 1 sample. (Table 5 and ESM).

Both types of insertion of nucleotides, i.e. heteroplasmic and homoplasmic, were found in 11 positions. The

frequency of the homoplasmic mutation type is relatively high and occurs in the region of the CA repeats in HV3ex. There was 1 CA dinucleotide added at positions 523.1 and 523.2 in 6 cases and 2 CA dinucleotides were added at positions 523.1–523.4 in 2 cases. Another insertion of a T at position 455.1 and a C at position 455.2 was observed in 1 sample (Table 5 and ESM). However, the main area where the insertion mutations occurred is situated in the poly-C tract called the C-stretch between positions 303–315 in HV2, where a C at position 315.1 was added in 94.6% and at position 309.1 in 53.8% of all samples (ESM). Similar results of a C insertion in this region were also observed in other populations [25, 34]. This type of variability is relatively often connected with another interesting phenomenon called length heteroplasmy, in which populations of mitochondrial haplotypes differing in length due to the insertion of nucleotide are found. Length heteroplasmy patterns between positions 303 and 315, similar to the data published by Parson et al. [25] were observed in 21 (22.5%) cases (Table 5 and ESM). Another length heteroplasmy can occur in the poly-C tract in HV1 due to a transition of T to C at position 16189 and addition of C between positions 16184 and 16193 as observed in 8 (8.6%) of our samples in agreement with other authors, i.e. in all where an observed change at position 16189 together with no substitution at position 16186 occurred [25, 38] (Table 5 and ESM).

Besides length heteroplasmy, two position heteroplasmy in HV1 at position 16155A/T in one sample and at position 16120A/C in a second sample were found (Table 5 and ESM).

Extreme care should be taken in determination of heteroplasmy, as for example different PCR strategies or sequencing artifacts can lead to different results [39]. In these cases new DNA extraction and reading both the forward and reverse reactions of the sequence should be considered. Alternatively, an error detection method covering not only errors leading to false determination of heteroplasmy but also of other types of errors in mtDNA data, based on phylogenetic analysis [40] could be used to minimize possible errors.

In conclusion, all these results and values, i.e. a sufficiently high genetic diversity, a sufficiently small RMP and a relatively high intrapopulation diversity, indicate that our data are relatively well suited for application to forensic casework and contribute to a better definition of continental and subcontinental distributions of mtDNA types [41]. Thus, it can be successfully applied especially to cases in which highly decomposed specimens are analyzed and typing with nuclear markers fails.

**Acknowledgment** The project was supported by the grant No. 350/2000 of the Ministry of Education of the Czech Republic.

## References

1. Robin ED, Wong R (1988) Mitochondrial DNA molecules and virtual number of mitochondria per cell in mammalian cells. *J Cell Physiol* 136:507
2. Nass MM (1966) The circularity of mitochondrial DNA. *Proc Natl Acad Sci U S A* 56:1215–1222
3. Allen M, Engstrom AS, Meyers S et al. (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *J Forensic Sci* 43:453–464
4. Bender K, Schneider PM, Rittner C (2000) Application of mtDNA sequence analysis in forensic casework for the identification of human remains. *Forensic Sci Int* 113:103–107
5. Goodwin W, Linacre A, Vazezis P (1999) The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims. *Electrophoresis* 20:1707–1711
6. Ginther C, Issel-Tarver L, King MC (1992) Identifying individuals by sequencing mitochondrial DNA from teeth. *Nat Genet* 2:135–138
7. Holland MM, Fisher DL, Mitchell LG, Rodriquez WC, Canik JJ, Merrill CR, Weedn VW (1993) Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *J Forensic Sci* 38:542–553
8. Hopwood AJ, Mannucci A, Sullivan KM (1996) DNA typing from human faeces. *Int J Legal Med* 108:237–243
9. Sullivan KM, Hoggood R, Gill P (1992) Identification of human remains by amplification and automated sequencing of mitochondrial DNA. *Int J Legal Med* 105:83–86
10. Szibor R, Michael M, Plate I, Krause D (2000) Efficiency of forensic mtDNA analysis. Case examples demonstrating the identification of traces. *Forensic Sci Int* 113:71–78
11. Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci U S A* 76:1967–1971
12. Cann RL, Brown WM, Wilson AC (1984) Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* 106:479–499
13. Giles RE, Blanc H, Cann HM, Wallace DC (1980) Maternal inheritance of human mitochondrial DNA. *Proc Natl Acad Sci U S A* 77:6715–6719
14. Hutchinson CA 3rd, Newbold JE, Potter SS, Edgell MH (1974) Maternal inheritance of mammalian mitochondrial DNA. *Nature* 251:536–538
15. Gill P, Ivanov PL, Kimpton C et al. (1994) Identification of the remains of the Romanov family by DNA analysis. *Nat Genet* 6:130–135
16. Anderson S, Bankier AT, Barrell BG et al. (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465
17. Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics* 103:287–312
18. Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. *Nature* 325:31–36
19. Greenberg BD, Newbold JE, Sugino A (1983) Intraspecific nucleotide sequence variability surrounding the origin of replication in human mitochondrial DNA. *Gene* 21:33–49
20. Lutz S, Weisser HJ, Heizmann J, Pollak S (1998) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany. *Int J Legal Med* 111:67–77
21. Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507
22. Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, Monson KL (1999) Mitochondrial DNA regions HVI and HVII population data. *Forensic Sci Int* 103:23–35
23. Crespiello M, Luque JA, Paredes M, Fernandez R, Ramirez E, Valverde JL (2000) Mitochondrial DNA sequences for 118 individuals from northeastern Spain. *Int J Legal Med* 114:130–132
24. Dimo-Simonin A, Grange F, Taroni F, Brandt-Casadevall C, Mangin P (2000) Forensic evaluation of mtDNA in a population from south west Switzerland. *Int J Legal Med* 113:89–97
25. Parson W, Parsons TJ, Scheithauer R, Holland MM (1998) Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of mtDNA sequence analysis to a forensic case. *Int J Legal Med* 111:124–132

26. Piercy R, Sullivan KM, Benson A, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. *Int J Legal Med* 106:85–90
27. Rousselet F, Mangin P (1998) Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework. *Int J Legal Med* 111:292–298
28. Wittig H, Augustin C, Baasner A et al. (2000) Mitochondrial DNA in the Central European population. Human identification with the help of the forensic mt-DNA D-loop-base database. *Forensic Sci Int* 113:113–118
29. Walsh PS, Metzger DA, Higuchi R (1991) Chelex 100 as a medium for simple extraction of DNA for PCR-based typing from forensic material. *Biotechniques* 10:506–513
30. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
31. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
32. Stoneking M, Hedgecock D, Higuchi RG, Vigilant L, Erlich HA (1991) Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am J Hum Genet* 48:370–382
33. Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15:174–175
34. Pfeiffer H, Brinkmann B, Huhne J et al. (1999) Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med* 112:291–298
35. Imaizumi K, Parsons TJ, Yoshino M, Holland MM (2002) A new database of mitochondrial DNA hypervariable regions I and II sequences from 162 Japanese individuals. *Int J Legal Med* 116:68–73
36. Budowle B, Allard MW, Fisher CL et al. (2002) HVI and HVII mitochondrial DNA data in Apaches and Navajos. *Int J Legal Med* 116:212–215
37. Chen MH, Lee HM, Tzen CY (2002) Polymorphism and heteroplasmy of mitochondrial DNA in the D-loop region in Taiwanese. *J Formos Med Assoc* 101:268–276
38. 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
39. Brandstätter A, Parson W (2003) Mitochondrial DNA heteroplasmy of artefacts – a matter of the amplification strategy? *Int J Legal Med* 117:180–184
40. Bandelt HJ, Lahermo P, Richards M, Macaulay V (2001) Detecting errors in mtDNA data by phylogenetic analysis. *Int J Legal Med* 115:64–69
41. Forster P, Cali F, Röhl A et al. (2002) Continental and subcontinental distributions of mtDNA control regions types. *Int J Legal Med* 116:99–108