Int J Legal Med (2004) 118 : 1–4
DOI 10.1007/s00414-003-0394-3

ORIGINAL ARTICLE

I. Zupanič Pajnič · J. Balažic · R. Komel

Sequence polymorphism of the mitochondrial DNA control region in the Slovenian population

Received: 11 February 2003 / Accepted: 13 June 2003 / Published online: 8 October 2003

Abstract The forensic application of mitochondrial DNA (mtDNA) typing requires large and regionally well-defined databases. To expand the database for forensic identification purposes in Slovenia, the mtDNA control region sequences of the hypervariable regions HVI and HVII were determined in a population of 129 maternally unrelated Slovenians, using a fluorescent-based capillary electrophoresis sequencing method. A total of 111 different haplotypes resulting from 124 polymorphic positions (80 polymorphic positions in HVI and 44 in HVII) were found. Of these, 101 mtDNA types were unique, 6 haplotypes were shared by 2 individuals, 1 haplotype by 3 individuals, 2 haplotypes by 4 individuals, and the most common haplotype was found in 5 individuals. The most frequent haplotypes in the Slovenian population ,263(G), 315.1(C) and 263(G), 309.1(C), 315.1(C) are also the most common in other European populations. The data support the concept that these haplotypes may represent a common European mtDNA sequence types. The sequence poymorphisms were compared to the databases of west Austria and central Italy and the HVI and HVII sequence matching probabilities within and between populations were calculated. It is 1.1–4.5 times more likely to find a sequence match in a random pair of Slovenians than in a random Slovenian-Italian pair and in a random Slovenian-Austrian pair. The length heteroplasmy in the homopolymeric C-stretch regions located at nucleotide positions 16184–16193 in HVI and at positions 303–315 in HVII was observed in 17% and 8% of individuals, respectively. A statistical estimate of the results for this population showed the random match probability and the genetic diversity of 1.16% and 0.996, respectively.

I. Zupanič Pajnič (✉) · J. Balažic
Institute of Forensic Medicine, Faculty of Medicine,
Korytkova 2, 1000 Ljubljana, Slovenia
Fax: +386-1-5244974,
e-mail: izupanic@mf.uni-lj.si

R. Komel
Institute of Biochemistry, Faculty of Medicine,
Vrazov trg 2, 1000 Ljubljana, Slovenia

Electronic Supplementary Material Supplementary material is available in the online version of this article at http://dx.doi.org/10.1007/s00414-003-0394-3

Keywords Mitochondrial DNA · Sequencing · Heteroplasmy · Forensic identification · Population genetics · Slovenia

Introduction

Sequence analysis of the mitochondrial DNA control region is of central importance for forensic identity testing as well as for studies of human evolution. In order to use the mtDNA analysis in forensic casework we have performed mtDNA database by typing mtDNA HVI and HVII region in a sample of 129 Slovenians. The sequence polymorphisms were compared to the databases of west Austria and central Italy. The mtDNA sequence matching probabilities within and between European populations were calculated only for HVI region (Pfeiffer et al. 1999; Cali et al. 2001). Here we present the calculations of matching probabilities within and between the Slovenian, Austrian and Italian populations for both hypervariable segments (HVI and HVII) of the mtDNA D-loop region.

Materials and methods

DNA extraction

The population sample included 129 maternally unrelated people who were not preselected for ethnicity and is therefore representative of the current ethnical composition of Slovenia. DNA was extracted from bloodstains, oral swabs and hairs using the chelex extraction procedure (Walsh et al. 1991), followed by purification with Centricon 100 spin dialysis columns (Amicon). DNA from the EDTA blood samples was extracted using the QIAamp blood kit (Qiagen) following the manufacturer's recommendations.

MtDNA amplification

The two hypervariable regions of mtDNA were amplified by the polymerase chain reaction (PCR) performed in a Biometra UNO-

thermoblock. For bloodstains, oral swabs and EDTA blood samples the primers F15997/R16401 for HVI and F29/R408 for HVII (Parson et al. 1998) were used. The primers F15971/R16258, F16140/R16414 for HVI and F15/R274, F145/R389 for HVII (Holland et al. 1995) were used for amplification of mtDNA extracted from hairs. PCR was carried out in a 25 µl reaction mixture consisting of 2.5 µl GeneAmp 10X PCR Gold Buffer (1× 15 mM Tris-HCl, pH 8.0, 50 mM KCl) (Applied Biosystems), 1.5 µl 25 mM $MgCl_2$, 2 µl 2.5 mM dNTP mix, 2 µl 5 µM forward amplification primer, 2 µl 5 µM reverse amplification primer and 0.25 µl of AmpliTaq Gold DNA polymerase (Applied Biosystems). For amplification of mtDNA extracted from hairs 0.75 µl of AmpliTaq Gold DNA polymerase was used and approximately 3 ng of DNA was added to each reaction mix. The amplification of HVI and HVII regions with the primers F15997/R16401, F16140/R16414, F29/R408, F15/R274, and F145/R389 was carried out at 94°C for 11 min followed by 32 cycles of 94°C for 20 s, 56°C for 10 s and 72°C for 30 s. The PCR conditions for amplification of HVI with the primer F15971/R16258 was the same, except the annealing was done at 62°C for 20 s. Prior to sequencing the PCR products were purified using Centricon 100 spin dialysis columns (Amicon) following the manufacturer's recommendations.

MtDNA sequencing

Sequencing reactions were performed in a Biometra UNO-thermoblock in both orientations in order to verify the accuracy of base-calling. For mtDNA typing of bloodstains, oral swabs and EDTA blood samples the primers used for the sequencing of the PCR products were the same as for the amplification. The sequencing primers F15971/R16255, F16144/R16414 for HVI and F29/R270, F155/R381 for HVII (Holland et al. 1995) were used for cycle sequencing of amplicons obtained from mtDNA extracted from hairs. Sequencing reactions were carried out using 2–6 µl ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit with AmpliTaq FS DNA polymerase (Applied Biosystems), 6–10 µl purified PCR product as template, 2 µl 5 µM sequencing primer, and sterile distilled water to 20 µl for each sample. Both strands were sequenced with different amounts of the Ready Reaction kit. We used 2 µl of the Ready Reaction kit for forward sequencing reactions and 6 µl for reverse sequencing reactions, except for mtDNA typing of hairs where we used 6 µl for sequencing in both the forward and reverse directions. Sequencing was carried out at 96°C for 30 s, 50°C for 15 s and 60°C for 4 min, for 25 cycles. The removal of excess dye-deoxy terminators, primers and buffer was accomplished with MicroSpin G-50 columns (Amersham Pharmacia Biotech) following the manufacturer's recommendations, 12 µl of Template Suppression Reagent (Applied Biosystems) was added to 10 µl of purified sequencing product, heat-denatured and snap-cooled on ice. Automated DNA sequencing was carried out on an ABI Prism 310 Genetic Analyzer (Perkin Elmer) using the Performance Optimized Polymer POP 6 (Applied Biosystems). The denatured samples were electrokinetically injected for 30 s at 2.0 kV into a 47 cm capillary. The electrophoresis was run at 15.0 kV and 50°C for 36 min with POP6 polymer and the rapid sequencing module. Analysis of mitochondrial DNA sequencing data was performed on a Macintosh computer using the Sequencing Analysis Software, Version 3.0 (ABI). The sequences were aligned and compared with the Anderson sequence (Anderson et al. 1981) from 16030 to 16401 for HVI region and from 55 to 408 for HVII region. In cases of length heteroplasmy in the poly-C strand, the polymorphisms behind the C-stretch in the forward sequencing reaction and before the C-stretch in the reverse reaction were confirmed by repeating amplification and sequencing reactions.

The random match probability was defined according to Stoneking et al. (1991) and the genetic diversity was calculated according to Tajima (1989). The sequence poymorphisms were compared to the databases of west Austria (Parson et al. 1998) and central Italy (Tagliabracci et al. 2001) for nucleotide positions 16030–16394 in HVI and nucleotide positions 55–361 in HVII according to the Anderson sequence. The minimum diversity within the population (dw_{min}), the maximum matching probability within the population (mw_{max}), the minimum matching probability within the population (mw_{min}), and the minimum matching probability between two populations (mb_{min}) were calculated according to Brinkmann et al. (1999) and Pfeiffer et al. (1999). An upper (mw_{max}/mb_{min}) and a lower estimate (mw_{min}/mb_{min}) of how many more times it is likely to obtain a sequence match within a population than between two populations were defined.

Results and discussion

The sequences of hypervariable regions 1 and 2 were determined for 129 maternally unrelated Slovenians and 111 mtDNA types were observed. Of these, 101 mtDNA types were unique and 10 haplotypes were shared by more than one individual (Table S1). A total of 124 polymorphic positions were found in both regions. For the HVI 80 variable sites were detected. In this region 15.5% of the haplotypes presented were identical with the Anderson sequence which is similar to other European Caucasian databases, e.g. 12% for Germany (Pfeiffer et al. 1999), 14% for Switzerland (Dimo-Simonin et al. 2000) and Britain (Piercy et al. 1993), 16% for Italy (Tagliabracci et al.

Table 1 The nucleotide substitutions, deletions and insertions observed in the HVI and HVII region of the mitochondrial DNA in the Slovenian population

Mutation type	Number of position	Total number of mutations
Transitions		
Pu-Pu		
A-G	23	242
G-A	9	35
Py-Py		
C-T	39	150
T-C	40	243
Total	111	670
Transversions		
A-C	6	21
C-A	1	2
G-C	1	1
T-G	1	1
G-T	1	1
A-T	1	1
T-A	1	1
Total	12	28
Deletions		
–C	1	1
Total	1	1
Insertions		
+C	2	176
+2C	1	19
Total	3	195

Py pyrimidine.
Pu purine.
A adenine
T thymine.
G guanine.
C cytosine.

2001), 17% for Spain (Crespillo et al. 2000), 18% for France (Cali et al. 2001), and 20% for Austria (Parson et al. 1998). For the HVII region 44 variable sites were detected.

The distribution of nucleotide substitutions, deletions and insertions (Table 1) showed a large bias towards transitional changes with a transition:transversion ratio of 24:1. The majority of mutations in both regions were transitions (75%), insertions were observed in 22%, transversions in 3%, and deletions in 0.1%, which is in agreement with previously published data (Piercy et al. 1993; Tagliabracci et al. 2001; Crespillo et al. 2000; Dimo-Simonin et al. 2000). The majority of observed transitions were pyrimidines, with a pyrimidine to purine ratio of 1.4:1 (more pyrimidine than purine transitions were observed in HVI, while in HVII more purine transitions were found). The number of different positions where insertions and deletions were observed was 3 and 1, respectively. The HVII sequence only was involved in C insertions. The deletion was found in only one individual at position 16257 in HVI. The majority of polymorphic sites presented only 1 mutation type, however 2 different mutation events were observed at 3 out of 124 positions (16129, 16265, and 72). As observed in Germans by Pfeiffer et al. (1999) the appearance of some mutations is mechanistically linked. The 16182A to C transversion (7 out of 129 individuals) was always associated with the 16183A to C transversion (10 out of 129 individuals) and the 16183A to C transversion in turn was always associated with the 16189T to C transition if it created an uninterrupted C-stretch (28 out of 129 individuals). Compared to the Anderson reference sequence 1 A to G transition at nucleotide position 263 in 128 subjects were detected and 6 cytosine nucleotides instead of 5 at nucleotide positions 311–315 were observed in 127 out of 129 individuals.

An intact poly-cytosine tract between nucleotide positions 16184 and 16193 without a T at position 16189 resulting in length heteroplasmy, produced a characteristic blurred sequence in nucleotides beyond the tract (Bendall and Sykes 1995). In our study the T to C transition at position 16189 was found in 28 individuals. Where an additional transition from C to T takes place between the positions 16184–16193 no length heteroplasmy sequences were observed. In Slovenians 6 sequences with an additional transition from C to T were found (4 at position 16186 and 2 at position 16192). Of the 129 individuals sequenced, the 22 length heteroplasmy sequences were observed with a stretch of 10 or more Cs. Of these, in 3 sequences the transition at position 16189 was coupled with an A to C transversion at position 16183 generating 11 cytosine stretches, and in 7 sequences there was a double A to C transversion at positions 16182 and 16183 generating 12 cytosine stretches. In the remaining 12 length heteroplasmy sequences the stretch was 10 cytosines long. Our data showed that 17% of the individuals presented a mixture of different length variants for a heteroplasmic situation in HVI region. Similar results were observed in other European populations, e.g. 5% in Switzerland (Dimo-Simonin et al. 2000), 12% in Italians (Tagliabracci et al. 2001), 13% in Germans (Pfeiffer et al. 1999), 15% in Austrians (Parson et al. 1998), and 16% in Spanish (Crespillo et al. 2000). The length heteroplasmy in the homopolymeric C-stretch region located at positions 303–315 in the HVII region was observed in 10 individuals (8% of the samples). In Germany (Pfeiffer et al. 1999) in 12% of the population samples the length heteroplasmy at positions 303–315 in HVII was found, in Italy (Tagliabracci et al. 2001) in 17%, in Austria (Parson et al. 1998) in 19%, in Spain (Crespillo et al. 2000) in 41%, and in Switzerland (Dimo-Simonin et al. 2000) in 62%. Two individuals in our population sample showed at nucleotide position 310 in HVII region transition T to C, resulting in a stretch of 13 cytosine nucleotides, where the characteristic out-of-reading frame sequence beyond the C-stretch occurred. The same unusual transition was found in one mother-son pair in Italians (Tagliabracci et al. 2001). In our study only these two sequences lacked the C insertion at nucleotide positions 311–315.

Of the 129 sequences studied 14% were not unique but showed at least one counterpart with an identical haplotype. Table 2 shows the haplotypes observed in at least two individuals in our population. The most frequent haplotype was found in 5 individuals, 2 haplotypes were shared by 4 individuals, 1 haplotype by 3 individuals, and the remaining 6 non-unique sequence types were shared by 2 individuals each. The most common haplotype observed in our study was 263(G), 315.1(C) found in 3.9% of the population sample. The other two most frequent haplotypes in Slovenians were 263(G), 309.1(C), 315.1(C) and 16126(C), 16189(C), 16294(T), 16296(T), 73(G), 263(G), 309.1(C), 315.1(C) each found in 3.1% of the population sample. The results were compared with other European Caucasian populations. The haplotypes 263(G), 315.1(C) and 263(G), 309.1(C), 315.1(C) are also the most common sequences in Spanish (Crespillo et al. 2000) where the first one was found in approximately 5% and the sec-

Table 2 Haplotypes observed in the Slovenian population that were found at least in two individuals

Haplotypes	n
263(G), 315.1(C)	5
263(G), 309.1(C), 315.1(C)	4
16126(C), 16189(C), 16294(T), 16296(T), 73(G), 263(G), 309.1(C), 315.1(C)	4
16304(C), 263(G), 315.1(C)	3
16298(C), 72(C), 263(G), 309.1(C), 315.1(C)	2
16304(C), 16311(C), 263(G), 315.1(C)	2
16129(A), 16148(T), 16223(T), 16233(G), 16311(C), 16391(A), 73(G), 152(C), 199(C), 204(C), 250(C), 263(G), 315.1(C), 337(G)	2
16224(C), 16311(C), 73(G), 93(G), 146(C), 207(A), 263(G), 315.1(C)	2
16136(C), 16182(C), 16183(C), 16189(C), 16223(T), 16278(T), 16289(G), 73(G), 153(G), 195(C), 263(G), 315.1(C), 338(T)	2
16129(A), 16148(T), 16185(T), 16223(T), 16233(G), 16311(C), 16391(A), 73(G), 199(C), 204(C), 250(C), 263(G), 315.1(C), 337(G)	2

ond one in 4.2%. The haplotype 263(G), 315.1(C) is also the main haplotype in Great Britain (4%) (Piercy et al. 1993), Austrians (3%) (Parson et al. 1998), and Swiss (2.6%) (Dimo-Simonin et al. 2000) and the haplotype 263(G), 309.1(C), 315.1(C) is the most frequent in Italians (8.4%) (Tagliabracci et al. 2001) and in Germans (2.7%) (Pfeiffer et al. 1999). The results indicate that the Slovenian population shares the most frequent haplotypes with other European Caucasian populations. The data support the concept that the haplotypes 263(G), 315.1(C) and 263(G), 309.1(C), 315.1(C) may represent a common European mtDNA sequence types.

Among the parameters important for statistical evaluation of mtDNA typing in forensic casework, the random match probability and the genetic diversity were calculated. The probability of two randomly selected individuals from a population having identical mtDNA haplotype is 0.0116 (1.16%) and the genetic diversity is 0.996. In other European populations the random match probability ranges from 0.84% in Swiss (Dimo-Simonin et al. 2000) to 1.3% in Spanish (Crespillo et al. 2000), 1.9% in Italians (Tagliabracci et al. 2001), 2.6% in Britain (Piercy et al. 1993), 3.1% in Germans (Pfeiffer et al. 1999), and 4.4% in French (Cali et al. 2001). The genetic diversity range from 0.98 in Germans (Pfeiffer et al. 1999) to 0.99 in Spanish (Crespillo et al. 2000) and Italians (Tagliabracci et al. 2001), and 1.1 in Britain (Piercy et al. 1993).

The mtDNA HVI and HVII sequence matching probabilities within and between the Slovenian, Austrian and Italian populations are shown in Table 3. When both HVI and HVII are considered, the haplotype diversities (dw_{min}) are high within all three populations (in the Slovenians 98.84%, in the Austrians 99.55%, and in the Italians 98.13%). The maximum probability (db_{max}) of obtaining two different mtDNA haplotypes when sampling a Slovenian and an Austrian population is 99.74%, when sampling a Slovenian and an Italian 99.65%, and when sampling an Austrian and an Italian it is 99.69%. Table 3 shows that it is more likely to find matches within Slovenia than between Slovenia and the other two databases. In the maximum estimate it is about 4.5 times more likely to find a match between two Slovenians than between a Slovenian and an Austrian, and 3.3 times more likely than finding a Slovenia-Italy match. The lower estimates are 1.4 and 1.1, respectively. It is also more likely to draw an Italy-Italy match than an Italy-Slovenia and an Italy-Austria match (factors ranging from 1.9 to 6.0). For the Austria samples it is approximately equally likely to find an Austria-Austria sequence match for HVI and HVII than an Austria-Slovenia or an Austria-Italy sequence match (factors ranging from 0.8 to 1.7). A comparisons between the French, Sicilian and British mtDNA samples (Cali et al. 2001) and the northwest German, south German and Austrian mtDNA samples (Pfeiffer et al. 1999) showed that the sequence matching probabilities within populations differ by less than a factor of 2 from the matching probabilities between populations when only the HVI region is considered.

In conclusion, mtDNA sequences for HVI and HVII were determined for 129 individuals in order to use the mtDNA analysis in forensic identification purposes. This database enable the statistical evaluation of mtDNA typing results in our laboratory.

Table 3 MtDNA HVI and HVII sequence matching probabilities within and between Slovenia, Austria and Italy

	Slovenia (S)	Austria (A)	Italy (I)
n	129	101	83
dw_{min}	0.9884	0.9955	0.9813
mw_{max}	0.0116	0.0045	0.0187
mw_{min}	31/8256	13/5050	23/3403
mb_{min}	S-A: 0.0026 S-I: 0.0035	A-S: 0.0026 A-I: 0.0031	I-S: 0.0035 I-A: 0.0031
mw_{max}/mb_{min}	S-A: 4.5 S-I: 3.3	A-S: 1.7 A-I: 1.5	I-S: 5.3 I-A: 6.0
mw_{min}/mb_{min}	S-A: 1.4 S-I: 1.1	A-S: 1.0 A-I: 0.8	I-S: 1.9 I-A: 2.2

References

Anderson S, Bankier AT, Barrell BG et al. (1981) Sequence and organization of the human mitochondrial genome. Nature 290: 457–465

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

Brinkmann C, Forster P, Schürenkamp M, Horst J, Rolf B, Brinkmann B (1999) Human Y-chromosomal STR haplotypes in a Kurdish population sample. Int J Legal Med 112:181–183

Cali F, Le Roux MG, D'Anna R et al. (2001) MtDNA control region and RFLP data for Sicily and France. Int J Legal Med 114:229–231

Crespillo M, Loque A, 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

Dimo-Simonin N, 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

Holland MM, Fisher DL, Roby RK, Ruderman J, Bryson C, Weedn VW (1995) Mitochondrial DNA sequence analysis of human remains. Crime Lab Digest 22:109–115

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

Pfeiffer H, Brinkmann B, Hühne 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

Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic identification. Int J Legal Med 106:85–90

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

Tagliabracci A, Turchi C, Buscemi L, Sassaroli C (2001) Polymorphism of the mitochondrial DNA control region in Italians. Int J Legal Med 114:224–228

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

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