

H. Pfeiffer · B. Brinkmann · J. Hühne · B. Rolf  
A. A. Morris · R. Steighner · M. M. Holland · P. Forster

## Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography

Received: 14 January 1999 / Accepted: 15 February 1999

**Abstract** Mitochondrial DNA control region sequences were determined in 109 unrelated German Caucasoid individuals from north west Germany for both hypervariable regions 1 (HV1) and 2 (HV2) and 100 polymorphic nucleotide positions (nps) were found, 63 in HV1 and 37 in HV2. A total of 100 different mtDNA lineages was revealed, of which 7 were shared by 2 individuals and 1 by 3 individuals. The probability of drawing a HV1 sequence match within the north west Germans or within published sets of south Germans and west Austrians is similar (within a factor of 2) to drawing a sequence match between any two of these three population samples. Furthermore, HV1 sequences of 700 male inhabitants of one village in Lower Saxony were generated and these showed a nearly linear increase of the number of different haplotypes with increasing number of individuals, demonstrating that the commonly used haplotype diversity measure (Nei 1987) for population samples tends to underestimate mtDNA diversity in the actual population.

**Key words** Mitochondrial DNA · Hypervariable regions · Population study · Match probability

### Introduction

The analysis of mitochondrial DNA is of great importance in forensic case work when only degraded DNA is available, because unlike nuclear DNA, mtDNA is present in

hundreds to thousands of copies per cell (Robin and Wong 1988). Mitochondrial DNA has been shown to evolve rapidly and mutate 5–10 times faster than nuclear DNA (Brown et al. 1979). The resulting sequence variation is concentrated within the non-coding control region consisting of the three hypervariable segments HV1, HV2 and HV3 (Aquadro and Greenberg 1983; Vigilant et al. 1991; Lutz et al. 1998, 1999). Since recombination is absent in mtDNA inheritance, the application of mtDNA analysis in forensic case work requires large population databases for estimating the probability of identity by chance (Allen et al. 1998). Another caution is the rapid evolutionary drift of mtDNA (due to its effective population size of only  $\frac{1}{4}$  compared to nuclear loci), which leads to population-specific lineage groups (e.g. Torroni et al. 1998). The question then arises whether it is appropriate to pool population samples from different regions.

The first part of our study presents a mtDNA database of 109 unrelated German Caucasoids living in north west Germany for both hypervariable regions (HV1 and HV2) of the mtDNA D-loop region. Mitochondrial DNA is generally used as a last resort when stain DNA is limited, thus it frequently occurs that only HV1, which has the highest density of polymorphic positions, is chosen. For this reason we discuss HV1 and HV1+HV2 separately. In order to investigate whether regional differences in the mtDNA sequences of German-speaking people are relevant for the probability of finding sequence matches by chance, mtDNA sequences from north west Germans were compared to a south German (Lutz et al. 1998, 1999) and a west Austrian population sample (Parson et al. 1998). Furthermore, sequence data for HV1 only were generated for 700 males living in a village and its surroundings in Lower Saxony, and the correlation between the number of haplotypes and the population number size was explored in order to obtain a realistic picture of the resolving power of mtDNA in a local context.

H. Pfeiffer (✉) · B. Brinkmann · J. Hühne · B. Rolf · P. Forster  
Institute of Legal Medicine, University of Münster,  
Von-Esmarch-Straße 62, D-48149 Münster, Germany  
Fax: +49-251-83-55-158

R. Steighner · M. M. Holland  
Armed Forces DNA Identification Laboratory,  
Armed Forces Institute of Pathology, 1413 Research Blvd.,  
Rockville, MD, USA 20850D3125

A. A. Morris  
Earswick Drive, Scarborough, Ontario, M1 E 1C6, Canada

## Materials and methods

### DNA extraction

EDTA blood samples from 109 unrelated German Caucasoids living in North Rhine Westphalia and 700 saliva samples from males living in one village area in Lower Saxony were collected. The latter 700 samples were neither preselected for ethnicity, nor for maternal unrelatedness and therefore represent the current genetic composition of the village. DNA from the EDTA blood samples was extracted using the QIAamp Blood Kit (Qiagen). DNA extraction from the saliva samples was performed with Chelex 100 (Walsh et al. 1991).

### MtDNA amplification and sequencing

The amplification for both hypervariable regions was performed in a Perkin Elmer 9600 thermal cycler. The following primer pairs were used:

HV1: F15971 (5' TTA ACT CCA CCA TTA GCA CC)  
 R16410 (5' GAG GAT GGT GGT CAA GGG AC)  
 HV2: F15 (5' CAC CCT ATT AAC CAC TCA CG)  
 R448 (5' TGA GAT TAG TAG TAT GGG AG)  
 (Holland et al. 1995).

The PCR master mix for a 25 µl reaction consisted of: 2.5 µl 10X Perkin-Elmer PCR Buffer (1X 10 mM Tris-HCl, pH 8.3, 50 mM KCl and 1.5 mM MgCl<sub>2</sub>), 2 µl 2.5 mM dNTP Mix, 1 µl 10 µM forward amplification primer, 1 µl 10 µM reverse amplification primer and 0.25 µl of AmpliTaq DNA polymerase. Of the DNA extract 2.5 µl (approximately 2.5 ng) was added to each reaction mix. The PCR conditions for both HV1 and HV2 were: 94°C for 30 s followed by 32 cycles 94°C for 20 s, 56°C for 10 s, 72°C for 30 s, followed by 5°C soak. Of the PCR product 5 µl was added to 1 µl 6 × agarose gel loading buffer, loaded on a 2% agarose gel, electrophoresis was carried out at constant voltage (90 V) for approximately 30 min and visualised on a transilluminator. A sizing ladder was included on each gel (1 µl of 123 base pair ladder, 4 µl dest. H<sub>2</sub>O and 1 µl of 6 × loading buffer). The gels were photographed with a Polaroid camera fitted with an orange filter. Prior to sequencing the PCR product was purified using Centricon 100 spin dialysis columns (Amicon). Cycle sequencing was performed using 8.0 µl ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit with AmpliTaq FS DNA polymerase, 1 µl 10 µM sequencing primer, 1–11.0 µl purified PCR product and sterile distilled water qs to 20 µl for each sample. The sequencing primers were as follows (Holland et al. 1995):

HV1: F15989 5' CCC AAA GCT AAG ATT CTA AT  
 R16410 5' GAG GAT GGT GGT CAA GGG AC  
 HV2: F29 5' CTC ACG GGA GCT CTC CAT GC  
 R381 5' GCT GGT GTT AGG GTT CTT TG

A Perkin Elmer 9600 thermal cycler was used under the following conditions: 25 cycles at: 96°C for 15 s, 50°C for 5 s and 60°C for 2 min. After sequencing the samples were passed through Centriflex gel filtration cartridges (Advanced Genetic Technologies) and dried in a vacuum centrifuge. Automated DNA sequencing was performed on an AB 373 DNA Sequencer and 4 µl loading buffer (50 mM EDTA/deionised formamide 1:5) was added to the dry samples. The samples were vortexed, denatured at 96°C for 2 min and loaded on a polyacrylamide gel. Electrophoresis was run at 18–21 mA, 980–1600 V and a temperature approaching 40°C for 12 h. Analysis of mitochondrial DNA sequencing data was performed on a Macintosh computer using the Sequence Navigator software (Version 1.0.1 ABI). HV1 was analysed between nps16024 and 16365, HV2 between nps73 and 340 according to the Anderson sequence. The sequence polymorphisms were compared to the databases of west Austria (Parson et al. 1998) and south Germany (Lutz et al. 1998, 1999) for nps16024–16365 and nps73–340 according to the Anderson sequence (Anderson et al. 1981).

The diversity measures  $dw_{\min}$  (minimum diversity within the population, also termed  $h$  by Nei 1987),  $mw_{\min}$  (minimum matching probability within the population),  $mw_{\max}$  (maximum matching probability within the population), and  $mb_{\min}$  (minimum matching probability between two populations) were calculated as in Brinkmann et al. (1999). The ratio  $mw_{\max}/mb_{\min}$  gives an upper estimate how many more times it is probable to find a match within a population than between two populations and the ratio  $mw_{\min}/mb_{\min}$  gives a lower estimate.

## Results and discussion

The sequences of HV1 and HV2 were determined for 109 individuals. For these individuals 63 polymorphic positions were found in HV1 and 37 in HV2 which resulted in 100 different haplotypes (Table 1). The most frequent haplotype (263 G, 309.1 C, 315.1 C) was found in 3 individuals, the remaining 7 non-unique sequence types were shared by only 2 individuals each (Table 1).

Compared to the reference sequence (Anderson et al. 1981), one A to G transition at np263 and six cytosine nucleotides instead of five at nps311–315 were detected in all individuals (Table 1). These differences are typical for nearly all individuals in other populations (Piercy et al. 1993; Lutz et al. 1998; Parson et al. 1998; Pfeiffer et al. 1998; Rousselet and Mangin 1998). A transition from T to C at np16189 occurred in 14 cases (13%) and was connected with length heteroplasmy in the homopolymeric tract between nps16184 and 16193, which produced a characteristic blurred sequence in nucleotides beyond the tract (Bendall and Sykes 1995). Length polymorphism in the homopolymeric tract of HV2 at nps303–315 was found in 13 individuals (12%).

Lutz et al. (1998) found in 200 south German Caucasoids 88 variable nucleotide positions in HV1 and 65 variable positions in HV2. A revised version of the Lutz et al. (1998) data was used for our calculations, (cf. Lutz et al. 1999). Parson et al. (1998) investigated the sequences of 101 west Austrian Caucasoids and described 74 sequence polymorphisms between nps16023 and 16365 (HV1) and 46 between nps73 and 340. To investigate whether it is legitimate to pool the Austrian, south German and north west German databases for forensic purposes, upper and lower bounds for matching probabilities within and between regions were calculated for HV1 data (nps16024–16365) (Table 2).

Table 2 shows that it is approximately as likely to find a Münster-Münster match as a Münster-Innsbruck or a Münster-Freiburg match (factors ranging from 0.8 to 1.3). The same is true for the Freiburg samples. In contrast, it is slightly more likely to find matches within Innsbruck than between Innsbruck and the other two databases, although the factor is small (ranging from 1.3 to 1.9), and larger samples ( $n > 200$  for German subpopulations) and shuffling/resampling techniques are required to test the stability of these values. In any case, a 1.5 to twofold difference in matching probabilities between the databases will not be relevant in most forensic cases. The resolution of HV1 in Caucasoids typically ranges from 1 match in every 4 in-



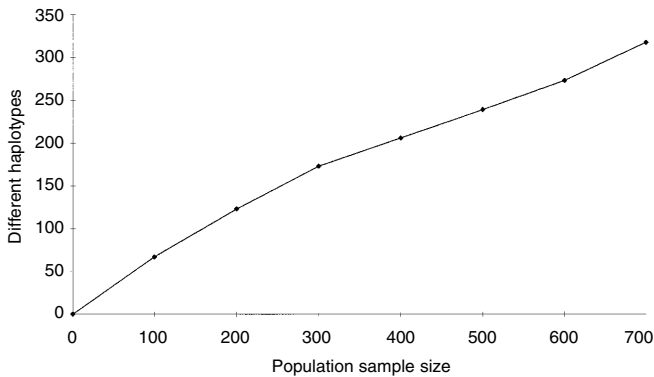




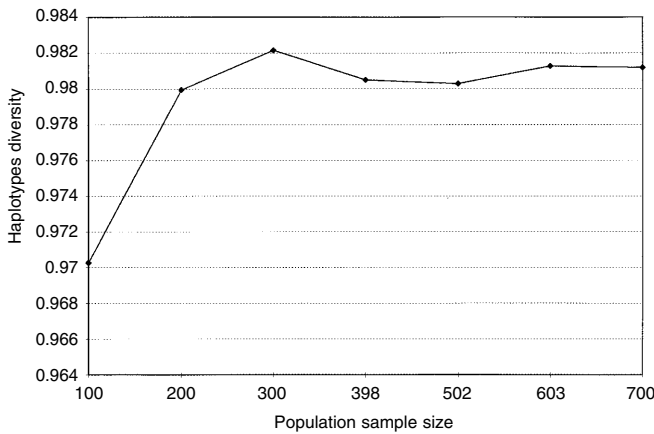


**Table 2** MtDNA HV1 sequence matching probabilities within and between the Münster, Freiburg, and Innsbruck samples

	Münster (M)	Innsbruck (I)	Freiburg (F)
<i>n</i>	109	101	200
$dw_{\min}$ (= <i>h</i> , Nei 1987)	0.9688	0.9460	0.9690
$mw_{\max} = 1 - dw_{\min}$	0.0312	0.0540	0.0310
$mw_{\min}$	131/5886	222/5050	519/19900
$mb_{\min}$	M-I: 0.0291 M-F: 0.0234	I-F: 0.0341 I-M: 0.0291	F-M: 0.0234 F-I: 0.0341
$mw_{\max}/mb_{\min}$	M/I: 1.1 M/F: 1.3	I/F: 1.6 I/M: 1.9	F/M: 1.3 F/I: 0.9
$mw_{\min}/mb_{\min}$	M/I: 0.8 M/F: 1.0	I/F: 1.3 I/M: 1.5	F/M: 1.1 F/I: 0.8



**Fig. 1** Correlation between population sample size and number of different mtDNA haplotypes in HV1



**Fig. 2** Correlation between population sample size and haplotype diversity (Nei 1987) in HV1. The samples are randomly selected from the database, the value for the sample size 100 is the mean value of three randomly selected population sub-samples of 100

that very few matches are found either within or between regions, making a detailed consideration as performed above for HV1 unnecessary.

In an independent study, 700 males from one village area of Lower Saxony were typed in HV1 only (sequence data can be obtained from the authors), where 137 variable positions resulted in 317 different haplotypes. As in the previous set of 109 Germans, the 16182A to C transversion (15

out of 809 individuals) was always associated with the np16183 A to C transversion (33 out of 809 individuals) and the np16183A to C transversion in turn was always associated with the 16189T to C transition if it created an uninterrupted C-stretch (105 out of 809 individuals). It thus appears that the appearance of these three mutations is mechanistically linked. Surprisingly, with increasing number of individuals the number of different haplotypes increases nearly linearly (Fig. 1) instead of approaching a level of saturation. Some of these village inhabitants of course are maternally related and an even higher number of different haplotypes would be expected if unrelated individuals were investigated. Of the 700 individuals 225 have unique mtDNA HV1 sequences.

It seems from Fig. 2 that a sample size of at least 200 is needed to obtain a realistic picture of the average matching probability for HV1 sequences in the actual population. We conclude that the forensic resolution of mtDNA HV1 is better than its reputation and will typically be forensically useful even in a single German village.

**Acknowledgements** We would like thank Kirsten Waterkamp and Karin Möller for their excellent technical assistance in the laboratory, as well as Hans-Jürgen Bandelt for statistical advice. This work was supported by the Deutsche Forschungsgemeinschaft (PF 356/1–1)

**References**

Allen M, Engström A-S, Meyers S, Handt O, Saldeen T, von Haeseler A, Pääbo S, Gyllensten U (1998) Mitochondrial DNA sequencing of shed hairs and saliva on robbery caps: sensitivity and matching probabilities. *J Forensic Sci* 43:453–464

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organisation of the human mitochondrial genome. *Nature* 290:457–464

Aquadro CF, Greenberg BD (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. *Genetics* 103:287–312

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, 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

Brown WM, George M, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. *Proc Natl Acad Sci USA* 76:1967–1971

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

Lutz S, Weisser H-J, 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

Lutz S, Weisser H-J, Heizmann J, Pollak S (1999) Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany (Erratum). *Int J Legal Med* 112:145–150

Nei M (1987) *Molecular evolutionary genetics*. Columbia University Press, New York Chichester, West Sussex, pp 177–179

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

- Piercy R, Sullivan KM, Benson N, Gill P (1993) The application of mitochondrial DNA typing to the study of white Caucasian genetic information. *Int. J Legal Med* 106:85–90
- Pfeiffer H, Steighner R, Fisher R, Yoon C-L, Mörnstad H, Holland MM (1998) Mitochondrial DNA extraction and typing from isolated dentin – experimental evaluation in a Korean population. *Int J Legal Med* 111:309–313
- Richards M, Corte-Real H, Forster P, Macaulay V, Wilkinson-Herbots H, Demaine A, Papiha S et al. (1996) Paleolithic and neolithic lineages in the European mitochondrial gene pool. *Am J Hum Genet* 59:185–203
- Robin ED, Wong R (1988) Mitochondrial DNA molecules and number of mitochondria per cell in mammalian cells. *J Cell Physiol* 136:507–513
- 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
- Torroni A, Bandelt H-J, D’Urbano L, Lahermo P, Moral P, Sellitto D, Rengo C, Forster P, Savontaus M-L, Bonn -Tamir B, Scozzari R (1998) MtDNA analysis reveals a major late palaeolithic population expansion from southwestern to northeastern Europe. *Am J Hum Genet* 62:1137–1152
- Vigilant L, Stoneking M, Harpending H, Hawkes K, Wilson AC (1991) African populations and the evolution of human mitochondrial DNA. *Science* 253:1503–1507
- 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