International Congress Series 1239 (2003) 505–509

D-Loop-BASE is online now Central European database of mitochondrial DNA

Holger Wittig [a,*], Mike Koecke [b], Kai-Uwe Sattler [b], Dieter Krause [a]

[a] *Institute of Forensic Medicine, Otto-von-Guericke-University, Leipziger Strasse 44, D-39120, Magdeburg, Germany*

[b] *Institute of Technical Information Systems, Otto-von-Guericke-University, Magdeburg, Germany*

Abstract

At the 18th Stain workshop held in Magdeburg in 1998, 15 university institutes of forensic medicine from Germany, Austria and Switzerland agreed to establish a Central European database of forensically secured mitochondrial DNA (mtDNA) sequences as a common project at the Magdeburg institute. Since then, more than 1600 sequences have been included into the D-Loop-BASE data stock which is now a profound basis for both frequency inquiries for expert opinions and scientific investigations into population genetic matters. Due to the consent agreement signed by the institutes involved, it was not possible in the beginning to access the data via the Internet, which made it difficult for all parties interested to use this database. Among the number of mtDNA databases in the Internet, there has been only one to date offering access to its data which, however, could be used for forensic purposes only to a limited extent. Together with the Institute of Technical Information Systems at the Magdeburg Otto-von-Guericke-University and the database administration, an Internet interface was developed which is available now under www.d-loop-base.de. Online inquiries for both the frequency of individual sequences and biostatistical parameters are possible now. But individual sequences will not be published. This paper shows the opportunities and limits of data inquiries.

Keywords: mtDNA; Database; Internet; Human identification

1. Introduction

The introduction of PCR technology into DNA analysis has revolutionised criminal investigations of traces over the past few years [1,2]. Further progress could be

* Corresponding author. Tel.: +49-391-6715812; fax: +49-391-67195810.
E-mail address: loop@d-loop-base.de (H. Wittig).

PII: S0531-5131(02)00585-X

achieved by sequencing mitochondrial DNA, which makes it possible now to identify a perpetrator by means of hair fragments without roots, marks of strangulation at the neck of a victim, the ends of a strangulation tool, or the handle of a knife used in a crime.

Nuclear DNA, on which the PCR systems are localised, are contained in two copies in each cell. However, each cell contains up to several 10,000 mitochondria with at least one copy of identical mitochondrial DNA (mtDNA). In the D-loop region, this DNA is polymorphic. Since the D-loop region exhibits almost exclusively point mutations, these can be identified only by sequencing. The disadvantages of this method are time-consuming and costly analyses and evaluation as well as the lack of sufficient population-genetic data.

When a comparison of the trace and of the suspect reveals three or more differences, the person under investigation, with the required degree of certainty, is not deemed to have caused the biological trace [3]. However, to date it has not been possible to reach a sufficiently high degree of certainty when the sequences are identical as the data available on an international scale are not yet sufficient. To date, there has been no publicly available mtDNA database providing an easy and efficient search algorithm.

Purposes of the central database are the collection and management of unencoded mtDNA sequences which have been exclusively analysed by the participating institutes, and are submitted in the form of signed data sheets (no data from the literature), the organisation of blind trials for improving analytical certainty, the opportunity of online frequency investigations to be used for expert opinions on traces by institutes, public prosecutors and courts, investigations for scientific purposes for all partner institutes in addition to the online features and the collection of data of different ethnic groups in Europe and elsewhere.

2. Current status of the database

At present, the D-Loop-BASE database contains more than 1700 individual mtDNA sequences: 1692 sequences HV1, 1654 sequences HV2, 1628 sequences HV1 and HV2 together, 513 sequences HV3, and another 200 sequences are under preparation.

2.1. Population samples

1610 Caucasians, i.e. 1266 from Germany, 195 from Switzerland, 102 from Austria, and 47 from 18 countries all over the world, 101 Asians (Japan) and several Africans.

3. Internet

The D-Loop-BASE database providing data for inquiries about distinct mtDNA frequencies has been online since August 2001. It is available under the URL http://www.d-loop-base.de.

Please also use this address to get into contact with the database administration.

4. Software system and sequence inquiries

The database is hosted on a mySQL server behind the firewall of the Otto-von-Guericke-University Magdeburg. Access is realised with a special Perl script. The original data are stored and supervised in a separate database system with no online access. It is part of our security system.

This on-line offer is of course focussed on frequency inquiries about individual mtDNA sequences through an input mask. At the beginning of inquiries, the data pool to be searched may be limited by entering specific ethnic and geographical data. An earlier publication [4] demonstrated that a general conformity, at least in Central Europe, can be presumed as the individual sub-populations revealed only slight differences.

This database offers the unique opportunity to indicate mtDNA sequences searched for differences from the Cambridge Reference Sequence (CRS) [5,6], simply by entering the base position and the base different from the reference. Entering the limits of the sequenced ranges with the L-strand positions of the first and the last bases, only those sequences from the database are considered for the inquiry which fully cover the indicated

Table 1
Two parts of the result table of an HV1 example search with 200 sequences

Input	
Individual data	
Ethnic group	Caucasians
Geographic origin	no data
Language	no data
Sequence data HV1	
Region analysed	16055–16390
Differences from CRS	16126.0 C
	16294.0 T
	16295.0 T
Sequence data HV2	
Region analysed	no data–no data
Differences from CRS	no data
Result	
Sequences analysed	
No. of sequences	200
0 differences	1
1 differences	13
2 differences	16
3 differences	46
4 differences	42
5 differences	34
6 differences	34
7 differences	11
8 differences	2
9 differences	1

range. The HV1 and HV2 regions can be searched for jointly or separately. The hypervariable region 3 [7], which has been described by only several hundred sequences to date, is integrated in the HV2 region.

The result table (Table 1) summarises all details in the inquiry form to enable input mistakes be identified. Thereafter, the number of the sequences available in the database and the number of sequences considered for the inquiry are displayed. The actual result of inquiries is a frequency distribution of pairwise differences of all mtDNA sequences considered with regard to the sequence searched for, serving also to establish the genetic distance to the overall population.

5. Summary

At the 18th Stain workshop held in Magdeburg in 1998, 15 university institutes of forensic medicine from Germany, Austria and Switzerland agreed to establish a Central European database of forensically secured mtDNA sequences as a common project at the Magdeburg institute.

More than 1600 sequences of the D-loop region of Central European Caucasians have been collected and made available to all interested parties for both expert opinions of traces and scientific purposes. Due to the consent agreement signed by the institutes involved, it was not possible in the beginning to access the data via the Internet, which made it difficult for all parties interested to use this database.

The D-Loop-BASE is now a profound basis for both frequency inquiries for expert opinions and scientific investigations into population genetic matters. Among the number of mtDNA databases in the Internet, there has been only one to date offering access to its data which, however, could be used for forensic purposes only to a limited extent.

Together with the Institute of Technical Information Systems at the Magdeburg Otto-von-Guericke University and the database administration, an Internet interface was developed which is available now under http://www.d-loop-base.de.

Online inquiries for the frequency of individual sequences are now possible. Additional biostatistical parameters are under preparation (e.g. haplotype frequencies), but individual sequences will not be published.

References

[1] K.M. Sullivan, R. Hopgood, B. Lang, P. Gill, Automated amplification and sequencing of human mitochondrial DNA, Electrophoresis 12 (1) (1991) 17–21.

[2] M.R. Wilson, J.A. DiZinno, D. Polanskey, J. Replogle, B. Budowle, Validation of mitochondrial DNA sequencing for forensic casework analysis, Int. J. Legal Med. 108 (2) (1995) 68–74.

[3] R. Piercy, K.M. Sullivan, N. Benson, P. Gill, The application of mitochondrial DNA typing to the study of white Caucasian genetic identification, Int. J. Legal Med. 106 (2) (1993) 85–90.

[4] H. Wittig, C. Augustin, A. Baasner, U. Bulnheim, N. Dimo-Simonin, J. Edelmann, S. Hering, S. Jung, S. Lutz, M. Michael, W. Parson, M. Poetsch, P.M. Schneider, G. Weichhold, D. Krause, 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 (1–3) (2000) 113–118.

[5] S. Anderson, A.T. Bankier, B.G. Barrell, M.H. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, Nature 290 (5806) (1981) 457–465.

[6] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, [letter], Nat. Genet. 23 (2) (1999) 147.

[7] S. Lutz, H.J. Weisser, J. Heizmann, S. Pollak, Location and frequency of polymorphic positions in the mtDNA control region of individuals from Germany, [published errata appear in Int. J. Legal Med, 1998; 111(5):286 and 1999; 112(2):145–150], Int. J. Legal Med. 111 (2) (1998) 67–77.