International Congress Series 1239 (2003) 617–620

The Combined DNA Index System

Kevin W.P. Miller*, Barry L. Brown, Bruce Budowle

DNA Analysis Unit II, Federal Bureau of Investigation, 935 Pennsylvania Avenue, N.W., Washington, D.C. 20535 U.S.A.

Abstract

The FBI Laboratory has developed the COmbined DNA Index System (CODIS) to assist in resolving crimes, particularly violent crimes, and prevent further crimes by quickly identifying recidivists. CODIS enables laboratories, both nationally and internationally, to exchange and compare DNA profiles electronically, thereby linking crimes committed by the same individual and/or to convicted offenders whose DNA profiles reside in the databank. In addition to the core short tandem repeat (STR) loci in CODIS, mitochondrial DNA (mtDNA) population data have recently been added. This new component of CODIS, called CODIS^{MT}, will be particularly useful in resolving crimes as well as assisting in the identification of the remains of missing persons.

Keywords: Mitochondrial DNA; Control region; Database; CODIS

1. Introduction

Databanks that contain DNA profiles of convicted felons, missing persons, and/or profiles from evidence from cases where no suspect has been identified, are useful for providing investigative leads for resolving certain crimes. The FBI Laboratory has developed the COmbined DNA Index System (CODIS) to assist in resolving crimes, particularly violent crimes, and prevent further crimes by quickly providing investigative leads on the perpetrator. CODIS enables local, state, federal, and international crime laboratories to exchange and compare DNA profiles electronically, thereby linking crimes committed by the same individual and/or to convicted offenders whose DNA profiles reside in the databank.

* Corresponding author. Tel.: +1-202-324-5065; fax: +1-202-324-5823.
E-mail address: kwm11@cornell.edu (K.W.P. Miller).

0531-5131/03
PII: S0531-5131(02)00502-2

2. Discussion

CODIS is implemented as a distributed database with three hierarchical levels—local, state and federal. All DNA profiles are generated at the local level (LDIS), and then the data are transmitted to the state level (SDIS). SDIS allows laboratories within a state to exchange DNA profile information. Each state and its respective local agencies operate its database according to specific legislative or legal requirements. The National DNA Index System (NDIS) is the highest level in the CODIS hierarchy and enables laboratories participating in CODIS to exchange and compare DNA profiles on a national level (i.e., between and among states).

DNA records from a number of sources can be obtained, stored, and compared in CODIS. The sources are: (1) convicted felons—which are persons convicted of crimes, defined under state statutes and at the national level according to federal legislation; (2) missing persons and their close biological relatives—which are persons reported missing and their biological relatives, such as parents, siblings and children; and (3) forensic evidence—which are samples from crimes that have not been resolved. Lastly, there is a population allele frequency database(s) index for statistical applications.

In order for DNA profile databanks to be compatible at a national level, standardization of the genetic markers used across laboratories is essential. To take full advantage of the power of short tandem repeat (STR) loci typing and to achieve compatibility of DNA profiles, the following 13 STR loci have been selected as core markers for CODIS: CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11. Laboratories in the US are required to type all 13 core STR loci when profiling convicted offenders whose profiles are sent to NDIS, and the typing of the same loci should be attempted when analyzing casework.

Other countries may obtain the CODIS software, and the software will be provided free of charge to appropriate authorized agencies. Currently, CODIS is installed in more than 131 laboratories in the US and 34 laboratories in 12 other countries. NDIS currently contains more than 520,000 profiles from 34 states, the US Army Crime Laboratory, and the FBI.

In addition to STR typing of nuclear DNA, determining the sequence of a portion of the mitochondrial DNA (mtDNA) genome extracted from hair shaft, bones, or teeth can assist in resolving a number of crimes, especially those involving the remains of missing persons. The analysis of the first and second hypervariable segments (HV1 and HV2) of the control region of the human mitochondrial genome has proven to be a useful tool for forensic identification [1–3], medical genetics [4], population and evolutionary studies [5], and anthropological reconstruction [6] due to the high degree of variation between unrelated individuals.

The mtDNA data also can be entered into CODIS. Indices and software have been developed to facilitate searching of an mtDNA nucleotide sequence developed from an evidentiary sample against one or more sequence databases. The databases house mtDNA and/or STR profiles from (1) unidentified human remains and (2) relatives of missing persons (to include personal items when appropriate). Two types of mtDNA sequence searching are enabled with the software: (1) comparison of a single profile against the profiles in an index; and (2) a pairwise search in which every profile is compared to every

other profile in a selected population database(s). The CODIS missing persons' database and mtDNA profile searching software are known as CODIS^{MT}.

Database information is particularly useful for inferring the relative rarity of a mtDNA profile. As the number of individuals typed increases, it is becoming more difficult to access population sequence data quickly and easily. A concordance (i.e., a compilation/compendium) of HV1 and HV2 sequence data was first constructed in 1996 [7] and made available via the internet (http://shelob.bioanth.cam.ac/uk/mtDNA). Since then, additional mtDNA population data have been developed, thus warranting another update.

To support casework interpretation, an updated mtDNA population database is available and will be periodically posted on the Internet (see Forensic Science Communications at http://www.fbi.gov). The current compendium of human mtDNA control region types contains 7528 population-specific mtDNA nucleotide sequences in a standardized format. Particularly, the database includes Scientific Working Group on DNA Analysis Methods (i.e., SWGDAM) mtDNA population data (currently N=4142 samples) that, at a minimum, contain sequence data for both HV1 and HV2 (nucleotide position np16024–16365 and np73–340, respectively) for each individual. Sequence data are described as those np sites that are different from the Cambridge Reference Sequence (CRS) [8,9]. Extant SWGDAM population data are comprised of all major population groups. The compilation of other public (i.e., published) mtDNA population data (N=6700) might be useful in some forensic cases, as well as in other fields of genetic study. The resulting reference list of mtDNA types is in a standardized format arranged by continent, country of origin, and ethnicity and includes all the population-specific mtDNA. The assembled data are referenced to the publication or source in which they originally appeared. A sequence identifier is used to characterize each sample profile source (for more details, see Miller and Budowle [10]).

We invite submissions of new sequence data for inclusion in future editions of the compendium, and welcome comments from users. Updates will be made when sufficient data become available. Finally, it is our intention to archive whole mtDNA genome information when it becomes available.

Acknowledgements

This is publication number 01-25 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

[1] M.M. Holland, D.L. Fisher, L.G. Mitchell, W.C. Rodriquez, J.J. Canik, C.R. Merril, V.W. Weedn, Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam war, J. Forensic Sci. 38 (1993) 542–553.

[2] M.R. Wilson, M. Stoneking, M.M. Holland, J.A. DiZinno, Guidelines for the use of mitochondrial DNA sequencing in forensic science, Crime Lab. Dig. 20 (1993) 68–77.

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

[4] D.C. Wallace, Mitochondrial DNA in aging and disease, Sci. Am., (August 1997) 40–47.
[5] M. Stoneking, Mitochondrial DNA and human evolution, J. Bioenerg. Biomembranes 26 (1994) 251–259.
[6] R. Cann, Human dispersal and diversity, Trends Evol. Ecol. 8 (1993) 27–30.
[7] K.W.P. Miller, J.L. Dawson, H. Hagelberg, A concordance of nucleotide substitutions in the first and second hypervariable segments of the human mtDNA control region, Int. J. Leg. Med. 109 (1996) 107–113.
[8] S. Anderson, A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Roe, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden, I.G. Young, Sequence and organization of the human mitochondrial genome, Nature 290 (1981) 457–465.
[9] 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, Nat. Genet. 23 (1999) 147.
[10] K.W.P. Miller, B. Budowle, A compendium of human mitochondrial DNA control region sequences: development of an international standard forensic database, Croat. Med. J. 42 (2001) 315–327.