· Advances in Medical Sciences · Vol. 53(1) · 2008 · pp 64-68 · DOI: 10.2478/v10039-008-0004-1

Database of genetic profiles at the Department of Forensic Medicine, Medical University of Bialystok

Janica J[1], Pepinski W[1*], Niemcunowicz-Janica A[1], Skawronska M[1], Berent J[2], Koc-Zorawska E[1], Soltyszewski I[3]

1 Department of Forensic Medicine, Medical University of Bialystok, Poland
2 Department of Forensic Medicine, Medical University of Lodz, Poland
3 Department of Criminalistics and Forensic Medicine, University of Warmia and Mazury, Olsztyn, Poland

* CORRESPONDING AUTHOR:
Department of Forensic Medicine,
Medical University of Bialystok,
13 Waszyngtona str.,
15-269 Bialystok, Poland,
telephone: +48 85 7485950; fax: +48 85 7485948
e-mail: pepinski@amb.edu.pl (Witold Pepinski)

Received 13.06.2007
Accepted 17.12.2007
Advances in Medical Sciences
Vol. 53(1) · 2008 · pp 64-68
DOI: 10.2478/v10039-008-0004-1

ABSTRACT

Purpose: The objective of this project was to establish an 'in-house' DNA database to store and compare profiles genotyped in the Department of Forensic Medicine, Medical University of Bialystok.
Material and Methods: DNA was extracted using Chelex-100, an organic procedure, or commercial kits. Genetic profiles were obtained using AmpFlSTR SGM Plus, AmpFlSTR Profiler, or AmpFlSTR Identifiler and 310 ABI Prism Genetic Analyzer (Applera). DNAStat v.1.2 software was used to construct the database.
Results: As of the end of 2006 our forensic database stored 1595 profiles genotyped in criminal cases in the years 2000–2006, including 398 non-match samples, 2 non-match fetuses, 5 non-match newborns and 4 non-match corpses.
Conclusion: A DNA database was established that may be used for the purpose of genetic profile comparison in criminal cases.

Key words: forensic science, DNA typing, DNA databases

INTRODUCTION

Deoxyribonucleic acid (DNA) is found in virtually every cell in the body and contains unique genetic information. DNA profiling examines discrete parts of an individual's DNA that vary greatly from one person to another. DNA profiling was first used in criminology in 1986 by Prof. Alec Jeffreys, and quickly became accepted worldwide. Forensic scientists soon wanted to establish a database of DNA profiles, but they needed a more efficient system. The arrival of the polymerase chain reaction (PCR) revolutionised identity testing. In 1988–89 came the first reports of microsatellites, also known as Short Tandem Repeats (STRs). STRs represent highly polymorphic microsatellite markers in the human genome that have tandemly repetitive sequence elements of 2–7 bps in length, located approximately every 10–15 kbs. Multiplex PCR-based STR kits with fluorescence detection technology have been validated to produce rapid and robust amplification of several DNA loci from biological samples, and thus have become one of the most powerful methods for genetic comparison of populations and have provided the most reliable means of personal identification [1,2]. Due to interpopulation variability of particular STR markers' distributions, in order to determine the probability value describing a random genotype occurrence, population data are collected to make a frequency estimation of each observed allele and genotype. Probability estimates are based on known allele frequencies for each STR locus and valid Hardy-Weinberg Equilibrium (HWE) verified by appropriate statistical tests for representative human populations. A DNA database allows analysis and storage of DNA profiles, usually derived from suspects' cheek swabs or blood samples. The resulting profiles can then be checked for matches with casework samples collected at crime scenes now and in the future. The objective of this project was to establish an 'in-house' database to store and compare DNA profiles genotyped in the Department of Forensic Medicine, Medical University of Bialystok.

NATIONAL DNA DATABASES

By a decision of the European Council (97/C 193/02) of June 9, 1997, EU member states were encouraged to establish their own national DNA databases. The United Kingdom's National DNA Database (NDNAD) was established in 1995 using the SGM DNA profiling system (SGM Plus DNA profiling system since 1998) and is currently run by the government-controlled Forensic Science Service (FSS). The data held on the NDNAD is owned by the police authority which submits samples for analysis. The NDNAD is the foremost and largest forensic DNA database of its kind in the world, containing more than 5% of the population, compared with an EU average of 1.13% and 0.5% in the US. A Home Office report has revealed that 5.24% of the UK population now has a DNA profile held on the database. The number of crimes solved through DNA technology has quadrupled over the past five years. There has been a 74% rise in the number of crimes where potential DNA material is collected, and a 75% increase in the number of matches of suspects to crime scenes. Even so, if DNA is found at a crime scene, in about 60% of cases there is no match in the database – the offender is someone without a record. The latest innovative intelligence approach brought forward by the FSS involves the use of familial searching. This is a process that may be carried out in relation to unsolved crime-stains whereby a suspect's DNA profile may not be held on the NDNAD, but that of a close relative is. The inventor of DNA fingerprinting – Alec Jeffreys – believes every citizen's genetic information should be stored on the UK national register. This would solve the problem of some individuals being listed even if they have been cleared of committing a crime. Jeffreys has said that a complete national database should be controlled by an independent body and be limited to storing DNA information that permits identification only – it should not carry DNA data that could be used to infer appearance, or susceptibility to disease.

In 1997 the FBI (Federal Bureau of Investigation) announced the selection of 13 STRs loci to constitute the core of the United States national database – CODIS (The Combined DNA Index System), which has been widely adopted by forensic DNA analysts in the United States [3]. Two commercially available kits, AmpFlSTR Profiler and AmpFlSTR SGM Plus (Applera, USA), were used to amplify the 13 STR loci included in the CODIS STR standardization project: D3S1358, VWA, D8S1179, D21S11, D18S51, TH01, FGA, TPOX, CSF1PO, D5S818, D13S317, D7S820, and D16S539. In its original form, CODIS consisted of two indexes: the Convicted Offender Index and the Forensic Index. The Convicted Offender Index contains profiles of individuals convicted of crimes; state law governs which specific crimes are eligible for CODIS. All 50 states have passed DNA legislation authorizing the collection of DNA profiles from convicted offenders for submission to CODIS. The Forensic Index contains profiles developed from biological material found at crime scenes. As of February 2007 the National DNA Index System (NDIS) contained a total of 4,398,639 profiles, including 4,231,536 convicted offender profiles and 167,103 forensic profiles. The development and expansion of databases that contain DNA profiles at the local, state, and national levels have greatly enhanced law enforcement's ability to solve cases with DNA. CODIS has so far produced over 45,400 hits assisting in more than 46,300 investigations.

Implementation of DNA analysis in German state laboratories started in 1987 when the heads of the Forensic Science Institutes decided to establish a working group to do the evaluation work necessary for introduction of the methodology into forensic casework. At the end of March 2006 the German DNA Database contained 472,000 records, including 18% of forensic evidence profiles. The proportion of matches has reached 26%. At present, DNA databases are fully or partially operational in the Netherlands, Austria, Finland, Norway, Denmark, Switzerland and Sweden.

On December 17, 2004 the Polish Parliament introduced changes to the Police Act of April 6, 1990, obliging the police to establish and administer a DNA database to contain profiles of convicted offenders, forensic evidence and unidentified corpses. Biological specimens used to generate DNA profiles include blood, mouth swabs, hairs, solid tissues, sperm, urine, saliva, etc. According to Article 21C: "Information stored in the DNA database shall be made available free of charge to authorities conducting criminal proceedings and Police authorities engaged in identification activities." A match in the database has no evidential significance, but it carries information about the source of evidence. To acquire legal value the evidence must be submitted to genetic testing and verified against a reference sample collected from the person indicated by the database.

'IN-HOUSE' DNA DATABASE

The application of DNA studies to the administration of justice has led to the need to develop appropriate computer programs. Such programs must address two critical problems, i.e. broadly-defined data processing and archivisation, and biostatistical calculations. The DNAStat v.1.2 software used in our laboratory facilitates the construction of a database to contain the following information: population data of genetic markers used (allele names and frequencies, mutation rates and population numbers), personal and administrative data (genotypes, client names and addresses) [4]. Any data may be modified, added and deleted any time. The database is saved as a single file (*.gdb). The software enables handling of multiple *.gdb files containing different databases. It is possible to switch databases at the software level. Reference or casework genotypes may be keyed in allele by allele or imported as notepad *.txt files or Microsoft Excel *.xls files. The database records are searched according to case number,

Table 1. **Allele frequencies and forensic efficiency parameters of 15 STR loci in a population sample from northeastern Poland, n=968.**

Allele	Loci												Allele	Loci		
	D3S1358	VWA	TH01	D16S539	D18S51	D8S1179	D19S433	TPOX	CSF1PO	D7S820	D13S317	D5S818		FGA	D2S1338	D21S11
5	-	-	0.004	-	-	-	-	-	-	-	-	-	16	0.001	0.048	-
6	-	-	0.248	-	-	-	-	-	-	-	-	-	17	0.004	0.208	-
7	-	-	0.121	-	-	-	-	-	-	0.016	-	-	18	0.016	0.096	-
8	-	-	0.118	0.008	-	0.004	-	0.544	0.008	0.186	0.138	0.004	19	0.096	0.114	-
9	-	-	0.189	0.074	-	0.016	-	0.084	0.024	0.164	0.046	0.064	20	0.148	0.128	-
9.3	-	-	0.312	-	-	-	-	-	-	-	-	-	20.2	0.004	-	-
10	-	-	0.012	0.03	0.024	0.06	-	0.072	0.298	0.268	0.064	0.052	21	0.194	0.048	-
11	-	-	-	0.302	0.012	0.064	0.001	0.268	0.282	0.196	0.368	0.316	22	0.182	0.018	-
12	-	-	-	0.346	0.084	0.186	0.098	0.032	0.318	0.116	0.264	0.384	22.2	0.004	-	-
13	0.001	0.004	-	0.192	0.089	0.324	0.196	-	0.062	0.054	0.084	0.164	23	0.118	0.098	-
13.2	-	-	-	-	-	-	0.032	-	-	-	-	-	23.2	0.004	-	-
14	0.142	0.098	-	0.044	0.162	0.224	0.298	-	0.008	-	0.032	0.016	24	0.136	0.116	-
14.2	-	-	-	-	-	-	0.028	-	-	-	-	-	25	0.046	0.122	-
15	0.269	0.106	-	-	0.174	0.094	0.196	-		-	0.004	-	25.2	0.004	-	-
15.2	-	-	-	-	-	-	0.088	-	-	-	-	-	26	0.042	0.004	
16	0.187	0.182	-	0.004	0.138	0.024	0.044	-	-	-	-	-	27	0.001	-	0.004
16.2	-	-	-	-	-	-	0.008	-	-	-	-	-	28	-	-	0.172
17	0.219	0.284	-	-	0.139	0.004	0.007	-	-	-	-	-	29	-	-	0.164
18	0.156	0.222	-	-	0.088	-	-	-	-	-	-	-	30	-	-	0.23
18.2	-	-	-	-	-	-	0.004	-	-	-	-	-	30.2	-	-	0.068
19	0.026	0.084	-	-	0.044	-	-	-	-	-	-	-	31	-	-	0.072
20	-	0.016	-	-	0.028	-	-	-	-	-	-	-	31.2	-	-	0.112
20.2	-	-	-	-	-	-	-	-	-	-	-	-	32.2	-	-	0.094
21	-	0.004	-	-	0.014	-	-	-	-	-	-	-	33.2	-	-	0.068
22	-	-	-	-	0.004	-	-	-	-	-	-	-	34.2	-	-	0.016
22.2	-	-	-	-	-	-		-	-	-	-	-				
P	0.430	0.667	0.118	0.687	0.816	0.792	0.981	0.536	0.631	0.942	0.557	0.639		0.187	0.092	0.462
PIC	0.75	0.76	0.71	0.68	0.87	0.72	0.72	0.57	0.67	0.77	0.68	0.65		0.84	0.86	0.80
DP	0.902	0.919	0.892	0.869	0.967	0.904	0.898	0.807	0.873	0.932	0.889	0.864		0.958	0.960	0.943
MP	0.098	0.081	0.108	0.131	0.033	0.096	0.102	0.193	0.127	0.068	0.111	0.136		0.042	0.040	0.057
TPI	2.92	2.92	1.52	1.94	3.09	2.19	2.92	1.13	1.50	2.17	1.83	1.27		3.56	4.43	2.65
PE	0.653	0.653	0.385	0.498	0.672	0.547	0.653	0.243	0.379	0.543	0.472	0.298		0.714	0.588	0.621

P: Fisher exact test probability, PIC: polymorphism information content, DP: discrimination power,
MP: matching probability, TPI: typical paternity index, PE: power of exclusion

surname and forename, sample collection date, etc. Searching by genotype is also possible, e.g. entering or importing a genotype of interest causes all matching records, full or partial, to be displayed. When genotypes for a single locus or an allele are compared, the software displays all matching records for the locus or the allele of interest, skipping data for the other loci/alleles. This option is particularly useful in analysis of decomposed or low copy number evidence samples where full profiles are unavailable due to DNA degradation.

DNA templates were extracted using several methods of choice including Chelex-100, organic procedure, or commercial kits, and then quantitated spectrophotometrically. AmpFlSTR SGM Plus, AmpFlSTR Profiler, or AmpFlSTR Identifiler kits (Applera, USA) were used according to the manufacturer's instructions. The kits contain reagents necessary to amplify 15 different STR loci: D3S1358, VWA, D16S539, D2S1338, D8S1179, D21S11, D18S51, D19S433, TH01, FGA, TPOX, CSF1PO, D7S820, D13S317, D5S818 and the gender-specific

Table 2. **Matrix of genetic distances between five populations inhabiting northeastern Poland (pairwise F_{ST} - above diagonal and according to Nei - below diagonal).**

	Poles	Old Believers	Belarussians	Polish Tatars	Lithuanians
Poles	-	0.0116	0.0007	0.0103	0.0017
Old Believers	0.9582	-	0.0115	0.0210	0.0137
Belarussians	0.9975	0.9588	-	0.0093	0.0024
Polish Tatars	0.9623	0.9250	0.9665	-	0.0102
Lithuanians	0.9939	0.9507	0.9914	0.9628	-

marker amelogenin. DNA samples were amplified using GeneAmp PCR System 9700 (Applera, USA). Genotyping was performed in a 310 ABI Genetic Analyzer (Applera, USA) using the GeneScan Analysis v3.7 and Genotyper v3.7 software, in accordance with the DNA recommendations [5]. As of the end of 2006 our forensic database stored 1595 profiles genotyped in criminal cases in the years 2000–2006, including 398 non-match samples, 2 non-match fetuses, 5 non-match newborns and 4 non-match corpses (3 males and 1 female) against which DNA profiles developed from other evidence can be compared. Certain categories of information are collected: 1. DNA identification records of persons related to crimes; 2. Analyses of DNA samples recovered from crime scenes; 3. Analyses of DNA samples recovered from unidentified human corpses or remains; 4. Analyses of DNA samples voluntarily contributed from relatives of missing persons; and 5. Known reference samples from missing persons. The paternity database stored 2160 profiles typed in disputed paternity cases from blood samples or buccal swabs.

Given the recidivistic nature of many crimes a likelihood exists that the individual who committed the crime being investigated has been convicted of a similar crime and already has his/her DNA profile in a searchable DNA database. Moreover, our database permits the cross-comparison of casework profiles. Even if a criminal is not identified through the database, crimes may be linked to each other, thereby aiding an investigation, which may eventually lead to identification of a suspect. The database records have so far made possible the successful identification of 26 male corpses, 2 female corpses and 4 fetuses.

The DNAStat v.1.2 software also provides biostatistical functions for databased genotypes. For the purpose of forensic assessment a profile frequency (f) and probability p(X|X) are calculated with optional coancestry coefficient (F_{ST}) and ceiling principle (CP) for an allelic low frequency limit. Paternity calculations include paternity index (PI) and paternity probability (W) (Ger.: Wahrscheinlichkeit) for full trios and deficiency (motherless) cases. The probability estimates are based on allelic frequencies determined in a population sample of 968 unrelated persons inhabiting northeastern Poland (*Tab. 1*). The genotype frequency distributions showed no deviations from HWE, based on the exact test. Pairwise comparison using the exact test disequilibrium analysis yielded no departures from independence. For calculating the rarity of a DNA profile, the National Research Council (NRC) II Report recommended the use of Wright's F_{ST} statistic [6]. F_{ST} is the correlation between two genes sampled from distinct individuals within a subpopulation or the probability that two alleles are identical by descent (two genes are copies of one of the genes carried by a common ancestor a few generations back). F_{ST} measures the effect of population subdivision, which is the reduction in heterozygosity in a population due to genetic drift. Determination of F_{ST} value in actual human populations is difficult and many laboratories have chosen to work with assigned values, for example 0.01 or 0.03 [7]. The larger values for F_{ST} are consistent with expectations for more isolated groups. These findings are supported by the complexity of the genetic heterogeneity pattern in the ethnic groups of Podlasie [8]. It is estimated that the northeastern corner of Poland is inhabited by 200,000 to 300,000 Belarussians, 20,000 to 30,000 Lithuanians and also 2,500 Polish Tatars and 600 Old Believers. If the allele frequencies for the subgroup are not available, forensic calculations should use the population-structure equations. Otherwise, ignoring F_{ST} would unfairly overstate the strength of the evidence against the defendant. The pairwise population comparisons between autochthonous Poles and the aforementioned minorities revealed statistically significant differences in F_{ST} values and relatively small values of interpopulation variation, which indicated a certain degree of genetic differentiation (*Tab. 2*). We suggest that this variation in distributions of genetic markers in northeastern Polish populations should be considered when evaluating the matching probability of forensic evidence, and consequently a provisional and more conservative F_{ST} value of 0.03 would be appropriate in selected cases.

ACKNOWLEDGEMENTS

The project was supported by the Medical University of Bialystok Grant No. 3-21635.

REFERENCES

1. Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R. Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet. 1994 Jul;55(1):175-89.

2. Ricci U, Sani I, Guarducci S, Biondi C, Pelagatti S, Lazzerini V, Brusaferri A, Lapini M, Andreucci E, Giunti

L, Giovannucci Uzielli ML. Infrared fluorescent automated detection of thirteen short tandem repeat polymorphisms and one gender-determining system of the CODIS core system. Electrophoresis. 2000 Nov;21(17):3564-70.

3. Baechtel FS, Monson KL, Forsen GE, Budowle B, Kearney JJ. Tracking the violent criminal offender through DNA typing profiles--a national database system concept. EXS. 1991;58:356-60.

4. Berent J. DNAStat, version 1.0--a software package for processing a genetic profile database and for biostatistical calculations. Arch Med Sadowej Kryminol. 2006 Jan-Mar;56(1):15-8.

5. Bär W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B. DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. Int J Legal Med. 1997;110(4):175-6.

6. National Research Council Report II. The Evaluation of Forensic DNA Evidence. National Academy Press, Washington, D.C. 1996, pp. 96-97.

7. Curran JM, Buckleton JS, Triggs CM. What is the magnitude of the subpopulation effect? Forensic Sci Int. 2003 Jul 29;135(1):1-8.

8. Pepinski W, Niemcunowicz-Janica A, Skawronska M., Janica JR, Koc-Zorawska E, Janica J, Soltyszewski I. Y-chromosome variation in northeastern Poland. Progress In Forensic Genetics 11, International Congress Series, 2006 Apr;1288:249-251.