

GENETIC DATA ON NINE COMPLEX STR SYSTEMS IN THE POPULATION OF SOUTH-EASTERN POLAND AND THEIR USEFULNESS IN PATERNITY TESTING

Piotr KOZIOŁ, Marzanna CIESIELKA, Roman MADRO

Chair and Department of Forensic Medicine, Medical Academy, Lublin

ABSTRACT: Population studies on 9 short tandem repeat (STR) systems (FGA, TPOX, D8S1179, VWA, Penta E, D18S51, D21S11, TH01, D3S1358) were conducted on a sample group of 317 unrelated residents of south-eastern Poland. Genotyping was based on a multiplex PCR reaction using the Promega GenePrint PowerPlex 2.1 system. Electrophoretic separation of the products was done in polyacrylamide denaturing gels, and an FMBIO II was used for detection of fluorochrome-marked fragments.

The distribution of phenotypes and frequency of alleles are presented in the paper. These were compared to those of populations in other regions of Poland. A lack of significant deviations from the Hardy-Weinberg equilibrium (*HWE*) was observed, and all of the PowerPlex 2.1 markers proved highly informative, though they display largely varied polymorphism information content (*PIC*) values. Mean exclusion chance (*MEC*), mean paternity exclusion probability (*MEP*), and mean paternity index (*MPI*) values were considerable, and these studies can be expected to prove decisive in paternity cases. Results from the examination of 9 STR systems were compared with those obtained analysing 4 VNTR loci (D4S139, D5S110, D7S21, D12S11) in 63 paternity cases. Paternity was confirmed in 50 cases; a paternity index value (*PI*) of more than 10 000 was found in 37 of these, using the STR system, and in 42 cases using the VNTR system. In 48 cases, the combined PowerPlex 2.1 and 4 VNTR investigations provided a probability of paternity exceeding 99.99%, which is equivalent to certain paternity.

KEY WORDS: Nine STRs; Population Database; Linkage Equilibrium; Forensic genetics; Paternity testing.

Z Zagadnień Nauk Sądowych, z. XLV, 2001, 66–80

Received 4 December 2000; accepted 28 March 2001

INTRODUCTION

The PowerPlex 2.1 system for multiplex STR amplification (Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, VWA) and fluorescent product detection is of great usefulness both in routine forensic paternity testing, and as a base of PCR genetic markers (with the exception of Penta E) for the European Network of Forensic Science Institutes (ENFSI),

as well as the Combined DNA Index System (CODIS) employed in the United States [13, 15]. These markers are characterized by a high degree of heterozygosity and low frequency of mutation; the data from nine loci can be obtained in a single reaction, and the alleles are easily identified by their number of repeats [13].

Using the multiplex PCR reaction with fluorescent product detection facilitates rapid generation of population-based data for several STR systems simultaneously. However, in order to introduce new markers, population studies should be performed and reliable allele frequencies obtained.

This paper investigates the south-east population data for 9 STR loci. The allele frequencies were compared with the frequencies determined in other regions of Poland. Moreover, the results of 9 STR systems studied in 63 paternity cases were evaluated and compared with those obtained in 4 VNTR systems.

MATERIALS AND METHODS

Whole blood samples were collected from 317 unrelated residents of south-eastern Poland and placed in EDTA tubes. DNA was extracted using the phenol-chloroform method, and quantified using the slot-blot ACES 2.0 Human DNA Quantification System (Life Technologies) [3]. Amplification of STR loci *via* PCR was achieved using the Promega PowerPlex 2.1 system kit as recommended [13]. The total volume of the PCR reaction, in which 1 ng of template DNA was used in a Biometra TRIO-Thermoblock apparatus with the cycle parameters of the PCR Perkin-Elmer 9600 block, equalled 20 μ l. After amplification, a mixture was prepared consisting of 2.5 μ l PCR products, 2.5 μ l loading dye, and 1 μ l Promega CXR 600 internal lane standard. The samples were denatured at 95°C for 2 min and then loaded onto the cathode end of the gel.

Electrophoresis was carried out in a 4.5% polyacrylamide sequencing gel (4.5% T, 6% C, 8 M urea), on an SA 43 gel electrophoresis apparatus (Life Technologies Inc., Gaithersburg, Md, USA) using 1 \times TBE buffer. Separation was initially performed at 80 W for approximately 5 min, and then continued at 60 W. The resulting fluorescent-labelled products were detected using the Hitachi FMBIO II FluoroImager.

Allele designations were determined by comparing the sample fragments with those of the allelic ladder (placed at every fourth lane), using FMBIO II Image Analysis 8.0 and StaRcall 3.0 software.

The nine STR systems were studied using the PowerPlex 2.1 in 63 cases of paternity, in which 4 VNTR loci were also tested (routine determinations in such cases), using the hybridisation method and chemiluminescence de-

tection. Previously documented conditions [8] were applied for the treatment of DNA samples with Hinf I, agar gel electrophoresis, and hybridisation using alkaline phosphatase-labelled probes (PH 30 and LH 1 from Life Technologies, MS 31 and MS 43A from Cellmark Diagnostics). The allele sizes were calculated based on the distribution of a molecular weight marker (MW 100) from Gibco-BRL, using the computerized Biotech Fischer GmbH videoscanner. The fragment size data were sorted out using the float bins method, and a window range of 4.8% bp ($\pm 2.4\%$ bp) was determined experimentally.

The statistical analysis of the results was performed using HWE evaluation methods, i.e. the χ^2 number of observed and expected homozygotes test [11], the exact test [5], and the likelihood ratio test [14]. Moreover, several parameters determining the usefulness of these systems in forensic testing were calculated: power of discrimination (*PD*) [7], polymorphism information content (*PIC*) [5], discriminating index (*DI*) [17], mean paternity exclusion probability (*MEP*), mean paternity index (*MPI*) [2], and power of exclusion (*PE*) [16].

The homogeneity of the examined group with 3 other Polish populations was checked using the 2-way $R \times C$ contingency table according to G. Carmody (Carlton University, Ottawa, Canada).

RESULTS AND DISCUSSION

The distribution of genotype and allele frequencies for 9 STR loci found in south-eastern Poland are presented in Table 1. The table also shows the observed and expected homozygosis and HWE. The genotype frequency distributions for 9 loci show no deviations from HWE in the χ^2 -test (Ho-T), logarithmic likelihood ratio test (LR-T), and exact test (Ex-T).

It is noteworthy that the probability values in the homozygous test are relatively low ($0.2 < p < 0.3$) for the FGA, D21S11, and TH01 loci, while those of the G and exact tests are high, especially considering that the number of homozygotes was lower than expected in every case.

Table 2 presents a comparison of the allele frequencies of 8 STR systems (Polish population statistics for the Penta E system were unavailable) in the population of south-eastern Poland (Lublin) with those of 3 other regions: Pomerania-Kujawy (Bydgoszcz) [9], south-western (Wrocław) [4], and northern Poland (Gdansk) [12]. No significant differences were found in any of the systems studied when comparing our region with northern Poland. Comparison of our population with that of the south-western region, however, showed homogeneity only in the D21S11, D8S1179, D3S1358, VWA, TH01, TPOX systems; with the Pomerania-Kujawy region, only in the

D3S1358, VWA, TH01 and TPOX systems. Statistically significant differences were observed in the FGA and D18S51 systems when comparing the allele frequency distributions of the studied region (Lublin) with those of south-western Poland, and in the FGA system when compared with those of the Pomerania-Kujawy region. Heterogeneity of the populations of Lublin and south-western Poland (Wrocław) in the D18S51 system may stem from the relatively small portion of the population examined in Wrocław (250 alleles, 125 individuals), especially since the significant differences concern only 2 alleles, 14 and 16, which have a 3% higher level of frequency in our region. In the FGA system, only the Gdańsk and Lublin populations were found to be homogenous. Comparison of the remaining populations showed them to be heterozygous, with the number of individuals examined in each population exceeding 200.

Additionally, the results obtained concerning the FGA system show varying levels of heterozygosity (Bydgoszcz – 0.90; Wrocław – 0.87; Gdańsk – 0.85, and Lublin – 0.87). The greatest differences (8–9%) were observed in the frequencies of allele 25 (Bydgoszcz – 0.16; Wrocław – 0.08; Gdańsk – 0.07; Lublin – 0.07), while the smallest ones (but still above 4%) concerned the alleles 19, 20, 21, 22, 24.

TABLE I. DISTRIBUTION OF PHENOTYPES AND ALLELES OF NINE STR LOCI IN THE POPULATION FROM SOUTH-EAST POLAND

PENTA E $n = 317$

Allele	5	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	23	Frequency
5	4																	0.0836
7	6	4																0.1293
8	1	0	1															0.0237
9	2	1	0	0														0.0126
10	8	11	1	0	3													0.1088
11	2	11	1	1	8	4												0.0899
12	7	10	3	1	9	11	4											0.1278
13	6	8	2	1	7	3	10	5										0.1167
14	1	8	1	1	4	6	5	9	3									0.0852
15	5	6	0	0	5	2	1	3	1	0								0.0505
16	2	2	1	0	6	1	7	9	4	2	1							0.0615
17	3	5	1	0	2	1	3	2	1	3	2	1						0.0426
18	2	2	2	1	0	2	4	1	3	1	0	1	0					0.0331
19	0	4	0	0	0	0	2	2	1	1	1	0	0	0				0.0174
20	0	0	0	0	1	0	0	1	2	2	0	1	0	0	0			0.0110
21	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	0		0.0047
23	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0.0016

Obs. Ho. = 0.094; Exp. Ho. = 0.091; HWE; Ho-T ($p=0.839$), LR-T ($p=0.831$), Ex-T ($p=0.392$).

FGA $n = 317$

Allele	16	18	19	20	20.2	21	21.2	22	22.2	23	23.2	24	24.2	25	25.2	26	27	Frequency
16	0																	0.0016
18	0	0																0.0031
19	0	0	1															0.0836
20	0	0	10	4														0.1593
20.2	0	0	0	0	0													0.0016
21	1	0	9	21	0	12												0.1830
21.2	0	0	1	0	0	0	0											0.0016
22	0	1	10	25	0	21	0	14										0.2129
22.2	0	0	0	1	0	1	0	0	0									0.0079
23	0	0	7	12	0	16	0	15	1	2								0.1120
23.2	0	0	0	2	0	0	0	3	0	2	0							0.0142
24	0	1	7	13	1	13	0	15	1	7	2	5						0.1183
24.2	0	0	0	1	0	0	0	1	0	0	0	0	0					0.0047
25	0	0	6	6	0	7	0	11	1	5	0	4	1	1				0.0694
25.2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0			0.0016
26	0	0	1	2	0	1	0	4	0	2	0	1	0	1	0	0		0.0205
27	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0.0047

Obs. Ho. = 0.123; Exp. Ho. = 0.143; HWE; Ho-T ($p = 0.292$), LR-T ($p = 0.99$), Ex-T ($p = 0.99$).D21S11 $n = 317$

Allele	26	27	28	28.2	29	29.2	30	30.2	31	31.2	32	32.2	33	33.2	34.2	Frequency
26	0															0.0016
27	0	0														0.0284
28	0	6	13													0.2114
28.2	0	0	0	0												0.0016
29	0	2	26	0	12											0.1956
29.2	0	0	1	0	0	0										0.0047
30	1	4	25	1	34	1	17									0.2366
30.2	0	0	5	0	5	0	6	0								0.0331
31	0	1	9	0	8	0	4	0	0							0.0567
31.2	0	0	17	0	8	0	17	2	6	2						0.0899
32	0	0	2	0	4	0	4	0	1	1	0					0.0205
32.2	0	3	9	0	5	1	15	2	2	1	1	1				0.0694
33	0	0	1	0	0	0	1	0	0	0	0	0	0			0.0032
33.2	0	2	7	0	8	0	3	1	4	1	0	3	0	0		0.0457
34.2	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0.0016

Obs. Ho. = 0.142; Exp. Ho. = 0.159; HWE; Ho-T ($p = 0.375$); LR-T ($p = 0.997$), Ex-T ($p = 0.993$).

D18S51 $n = 317$

Allele	10	11	12	13	14	15	16	17	18	19	20	21	22	Frequency
10	0													0.0047
11	0	0												0.0189
12	1	2	4											0.0962
13	0	0	8	4										0.0899
14	1	2	6	7	6									0.1562
15	0	4	7	10	17	6								0.1483
16	0	2	13	7	20	14	13							0.1798
17	1	2	7	11	10	14	13	3						0.1278
18	0	0	4	3	9	6	13	7	4					0.0899
19	0	0	3	2	9	3	3	4	3	0				0.0426
20	0	0	0	1	3	3	1	4	4	0	0			0.0252
21	0	0	2	0	2	3	2	0	0	0	0	0		0.0142
22	0	0	0	0	1	1	0	2	0	0	0	0	0	0.0063

Obs. Ho. = 0.126; Exp. Ho. = 0.124; HWE; Ho-T ($p = 0.887$), LR-T ($p = 0.796$), Ex-T ($p = 0.916$).

D8S11 $n = 317$

Allele	8	9	10	11	12	13	14	15	16	Frequency
8	0									0.0142
9	0	0								0.0173
10	1	0	0							0.0678
11	0	1	6	1						0.0694
12	5	3	6	6	10					0.1767
13	3	3	13	10	30	28				0.3044
14	0	2	11	10	26	55	16			0.2366
15	0	1	6	6	15	16	13	1		0.0931
16	0	1	0	3	1	7	1	0	0	0.0205

Obs. Ho. = 0.176; Exp. Ho. = 0.199; HWE; Ho-T ($p = 0.294$), LR-T ($p = 0.149$), Ex-T ($p = 0.200$).

D3S1358 $n = 317$

Allele	13	14	15	16	17	18	19	20	Frequency
13	0								0.0031
14	1	6							0.1467
15	1	17	17						0.2303
16	0	28	40	20					0.2524
17	0	21	31	29	18				0.2161
18	0	13	21	19	15	7			0.1309
19	0	0	2	3	5	0	0		0.0158
20	0	1	0	1	0	1	0	0	0.0047

Obs. Ho. = 0.214; Exp. Ho. = 0.202;
HWE; Ho-T ($p = 0.585$), LR-T ($p = 0.748$),
Ex-T ($p = 0.882$).

VWA $n = 317$

Allele	11	13	14	15	16	17	18	19	20	Frequency
11	0									0.0016
13	0	0								0.0079
14	0	2	1							0.0836
15	0	0	4	2						0.0978
16	0	1	18	13	7					0.2082
17	1	1	11	15	34	22				0.2555
18	0	1	11	21	35	38	18			0.2413
19	0	0	5	4	14	17	9	4		0.0915
20	0	0	0	1	3	1	2	1	0	0.0126

Obs. Ho. = 0.170; Exp. Ho. = 0.192;
HWE; Ho-T ($p = 0.312$), LR-T ($p = 0.809$),
Ex-T ($p = 0.761$).

TH01 $n = 317$

Allele	6	7	8	9	9.3	10	Frequency
6	24						0.2382
7	25	4					0.1451
8	12	9	3				0.0946
9	27	21	10	24			0.2319
9.3	37	29	23	40	24		0.2823
10	2	0	0	1	2	0	0.0079

Obs. Ho. = 0.249; Exp. Ho. = 0.220;
 HWE; Ho-T ($p = 0.204$), LR-T ($p = 0.373$),
 Ex-T ($p = 0.317$).

TPOX $n = 317$

Allele	6	8	9	10	11	12	Frequency
6	0						0.0016
8	1	110					0.5789
9	0	29	2				0.0962
10	0	25	6	3			0.0678
11	0	81	19	5	16		0.2240
12	0	11	3	1	5	0	0.0315

Obs. Ho. = 0.4132; Exp. Ho. = 0.4001;
 HWE; Ho-T ($p = 0.546$), LR-T ($p = 0.747$),
 Ex-T ($p = 0.712$).

Obs. Ho. – observed homozygosity; Exp. Ho. – expected homozygosity, Ho-T – homozygosity test,
 LR-T – likelihood ratio test, Ex-T – exact test, n – number of individuals.

TABLE II. HOMOGENEITY TESTING FOR THE DEGREE OF RELATIONSHIP BETWEEN SOUTH-EAST POLAND AND 3 OTHER POLISH REGIONS, USING THE TEST ACCORDING TO CARMODY

Systems South-east ($n = 634$)	Pomerania-Kujawy 410 < (n) < 492			South-west 250 < (n) < 438			North ($n = 1412$)	
	(n)	P (χ^2)	P (G-test)	(n)	P (χ^2)	P (G-test)	P (χ^2)	P (G-test)
FGA	(420)	$>10^{-3}$	$>10^{-3}$	(442)	0.014	0.004	0.078	0.079
D21S11				(258)	0.036	0.081	0.080	0.109
D18S51				(250)	$>10^{-3}$	$>10^{-3}$	0.475	0.461
D8S1179				(256)	0.310	0.332	0.884	0.838
D3S1358	(420)	0.181	0.173	(438)	0.244	0.218	0.386	0.438
VWA	(432)	0.230	0.241	(438)	0.287	0.189	0.792	0.801
TH01	(410)	0.559	0.564	(362)	0.260	0.269	0.502	0.530
TPOX	(410)	0.315	0.319	(382)	0.031	0.026		

(n) – number of alleles.

At this stage of studies it is difficult to say whether this heterozygosity is related to the presence of sub-populations, or if it stems from other factors. It seems unwarranted at this point to regard this heterogeneity a result of comparing overly small populations. It should be noted that the alleles of the system in question not only differ in the number of recurrences of the 2 and 4 nucleotide motif, but also exhibit variation in their sequences of repetition. Moreover, the population studies were carried out using various starters (the sequences of which are unknown in the case of the Promega and Perkin-Elmer multiplex systems), as well as other separation and amplification product analysis techniques (capillary or plate electrophoresis using the Perkin-Elmer 377 or Hitachi FMBIO II gel sequencer). Thus, inter-laboratory comparative studies should be performed to find out whether the sub-populations indeed exist.

TABLE III. STATISTICAL PARAMETERS OF FORENSIC IMPORTANCE FOR NINE STR LOCI IN SOUTH-EAST POLAND

Locus	$H_{obs(exp)}$		PIC	PD	DI	PE	MPI
PentaE	0.905	(0.910)	0.901	0.982	0.017	0.815	5.165
FGA	0.877	(0.858)	0.840	0.962	0.028	0.712	3.979
D18S51	0.874	(0.878)	0.901	0.971	0.029	0.750	3.881
D21S11	0.858	(0.842)	0.822	0.954	0.037	0.686	3.458
VWA	0.830	(0.809)	0.781	0.933	0.053	0.621	2.900
D8S1179	0.823	(0.802)	0.774	0.928	0.056	0.616	2.801
D3S1358	0.758	(0.799)	0.767	0.929	0.082	0.597	2.339
TH01	0.751	(0.781)	0.744	0.918	0.108	0.565	2.045
TPOX	0.587	(0.601)	0.555	0.791	0.271	0.367	1.380
Total 9 systems						0.99995	14783

$H_{obs(exp)}$ – observed (expected) heterozygosity; PIC – polymorphism information content; PD – power of discrimination; DI – discrimination index; PE – power of exclusion; MPI – mean paternity index.

The statistical parameters defining the usefulness of each of the systems in forensic testing are shown in Table 3. They reveal that the most informative systems were: Penta E, FGA, D21S1 and D18S5, while the least informative one was TPOX. The combined determinations of 9 systems of the multiplex PowerPlex 2.1 make it possible to obtain a power of exclusion = 0.99995, and a paternity index = 14 783.

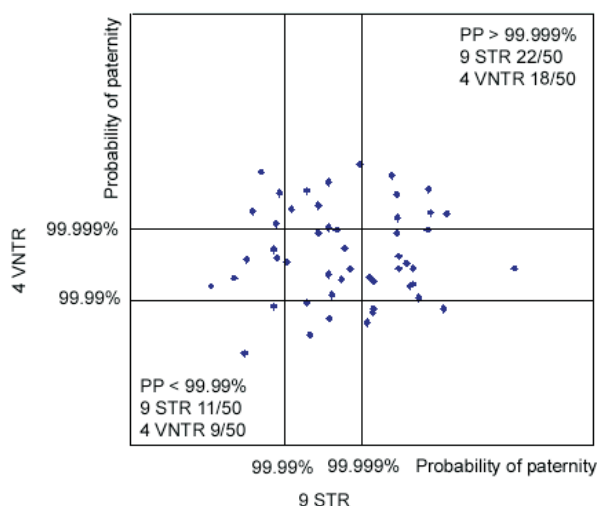


Fig. 1. Comparison of the probabilities of paternity calculated with four VNTR loci and with nine STR loci for the 50 cases.

63 parentage cases were examined using VNTR and STR analysis. In 13 cases, paternity was excluded in at least 3 out of 4 VNTR systems, and 4 out of 9 analysed STR loci. For the remaining 50 cases in which paternity was not excluded, the values of paternity likelihood, calculated on the basis of the VNTR and STR data were compared (Figure 1), since in Poland paternity is confirmed if paternity likelihood exceeds 99.99% (PI above 10 000).

The results of this analysis (Figure 1) show that a paternity likelihood of over 99.99% was obtained in 39 cases in the STR studies, and in 41 cases where VNTR systems were analysed. The high values of PI in the 4 VNTR systems result, among other things, from the use of the float bins method, which allows the determination of allelic fragment frequencies at every 4.8% bp. However, the results should be repeatable, and measurement error should not exceed $\pm 2.4\%$ bp. To achieve this, the electrophoretic separation of the father's, mother's and child's fragments has to be performed in one, 35 cm long agar gel using MW 100 size markers with constant parameters. Only such a method of VNTR phenotyping is fully justifiable and recommended by other authors [6, 10].

Figure 1 shows that in the VNTR investigations, the values of likelihood were lower than 99.99% in 9 cases, while in the STR studies such values were obtained in 11 cases. High values of paternity likelihood, exceeding 99.999%, were found in 18 cases with the VNTR studies, and 22 cases using the STR tests. In the combined 9 STR and 4 VNTR studies, a paternity likelihood of more than 99.99% was found in all cases, which made it possible to render opinions confirming paternity.

The replacement of 4 VNTR systems with multiple sets of STR loci shows several advantages: shorter analysis time, lower cost, and a variety of samples (not only blood). However, our studies show that the value of the 4 VNTR determinations may be compensated for if about 9 STR systems, similar to the ones studied in the multiplex PowerPlex 2.1, are included in the analysis. Nevertheless, the use of 2 different techniques, i.e. amplification in the case of the STR system, and hybridisation in the case of the VNTR, guarantees safe opinions. This is partly due to the fact that VNTR studies use the genomic DNA, which eliminates any errors related to the amplification reaction.

This study was supported by the Medical University of Lublin, Grant DS/380.

References:

1. Botstein D., White R. L., Skolnick M. [et al.], Construction of a genetic linkage map in man using restriction fragment length polymorphism, *American Journal of Human Genetics* 1980, vol. 32, pp. 314–331.
2. Brenner C., Morris J. W., Paternity index calculations in single locus hypervariable DNA probes: validation and other studies, Proceedings from the International Symposium on Human Identification, Promega Corporation 1989, pp. 21–53.
3. Budowle B., Baechtel F. S., Comey C. T. [et al.], Simple protocols for typing forensic biological evidence: chemiluminescent detection for human DNA quantification and RFLP analyses and manual typing of PCR amplified polymorphisms, *Electrophoresis* 1995, vol. 16, pp. 1555–1567.
4. Dobosz T., Częstości wybranych genów warunkujących polimorfizm genetyczny w obrębie populacji dolnośląskiej, Akademia Medyczna we Wrocławiu, Wrocław 2000.
5. Guo S. W., Thompson E. A., Performing the exact test of Hardy-Weinberg proportion for multiple alleles, *Biometrics* 1992, vol. 48, pp. 361–372.
6. Hamilton J. F., Starling L., Cordiner S. J. [et al.], New Zealand population data at five VNTR loci: validation as databases for forensic identity testing, *Science & Justice* 1996, vol. 36, pp. 109–117.
7. Kloosterman A. D., Daselaar P., Budowle B. [et al.], Population genetic study on the HLA-DQ α and the D1S80 locus in Dutch Caucasians, Proceedings from the Third International Symposium on Human Identification, Promega Corporation 1992, pp. 329–344.
8. Kozioł P., Tokarzewska J., Krajka A. [et al.], Polimorfizm lokus D5S110 w populacji Polski południowo-wschodniej i ocena przydatności w ekspertyzie genetycznej ojcostwa, *Postępy Medycyny Sądowej i Kryminologii* 1997, t. 3, s. 279–287.
9. Miścicka-Śliwka D., Czarny J., Grzybowski T. [et al.], Population genetics of 14 STRs: vWA, TH01, TPOX, CSF1PO, D5S818, D13S317, D7S820, D16S539, F13A01, FESFPS, F13B, LPL, D3S1358 and FGA in the Pomerania-Kujawy region of Poland, *Progress in Forensic Genetics* 1998, vol. 7, pp. 261–263.
10. Morling N., Hansen H. E., Paternity testing with VNTR DNA systems, *International Journal of Legal Medicine* 1993, vol. 105, pp. 189–196.
11. Nei M., Sampling variance of heterozygosity and genetic distance, *Genetics* 1974, vol. 76, pp. 379–390.
12. Pawłowski R., Dettlaff-Kąkol A., Jezierski G. [et al.], Genetyka populacyjna dziesięciu loci typu STR z zestawu ProfilerPlus w próbie populacyjnej z obszaru Polski, *Archiwum Medycyny Sądowej i Kryminologii* 2000, t. L, s. 207–213.
13. Promega Corporation. GenePrint™ PowerPlax™ 2.1 System Technical Manual, Part #TMD011(revised 5/99), Madison 1999.
14. Weir B. S., Independence of VNTR alleles defined as fixed bins, *Genetics* 1992, vol. 130, pp. 873–887.

15. Werrett D. J., The national DNA database, *Forensic Science International* 1997, vol. 88, pp. 33–42.
16. Wiegand P., Budowle B., Rand S. [et al.], Forensic validation of the STR systems SE33 and TC11, *International Journal of Legal Medicine* 1993, vol. 105, pp. 315–320.
17. Wong Z., Wilson V., Patel I. [et al.], Characterization of a panel of highly variable minisatellites cloned from human DNA, *Annals of Human Genetics* 1987, vol. 51, pp. 269–288.

BADANIA 9 UKŁADÓW STR W POPULACJI POLSKI POŁUDNIOWO-WSCHODNIEJ ORAZ OCENA ICH PRZYDATNOŚCI W EKSPERTYZIE DOTYCZĄCEJ USTALANIA OJCOSTWA

Piotr KOZIOŁ, Marzanna CIESIELKA, Roman MADRO

WSTĘP

Zestaw Power Plex 2.1 do multipleksowej amplifikacji STR (Penta E, D18S51, D21S11, TH01, D3S1358, FGA, TPOX, D8S1179, VWA) i fluorescencyjnej metody detekcji produktów jest bardzo użyteczny w badaniach genetycznych spornego ojcostwa, jak również jako baza markerów PCR (z wyjątkiem Penta E) europejskiego systemu ENFSI (European Network of Forensic Science Institutes, DNA Working Group) i systemu CODIS (Combined DNA Index System), który obowiązuje w Stanach Zjednoczonych [13, 15]. Układy te charakteryzują się wysokim stopniem heterozygotyczności i niską częstością mutacji oraz tym, że mogą być amplifikowane w jednej reakcji PCR, a ich allele są identyfikowane w oparciu o ilość powtórzeń sekwencji [13].

Zastosowanie multipleksowej reakcji PCR z fluorescencyjną detekcją produktów pozwala na oznaczanie jednocześnie w bardzo krótkim czasie dużej bazy populacyjnej dla kilku układów STR. Wprowadzenie do badań sądowych nowego systemu markerów wymaga jednak wykonania badań populacyjnych i uzyskania wiarygodnych częstości alleli.

W niniejszej pracy przedstawiono wyniki badań 9 układów STR uzyskane dla populacji południowo-wschodniej Polski.

Otrzymane dla poszczególnych układów częstości alleli porównano z wynikami z trzech innych regionów Polski. Przeprowadzono również ocenę wartości badań 9 układów STR w 63 przypadkach dotyczących ustalania ojcostwa oraz porównano je z wynikami badań 4 układów VNTR.

MATERIAŁ I METODY

Materiał do badań stanowiły próbki krwi z dodatkiem EDTA pobrane od 317 nie-spokrwinnionych mieszkańców południowo-wschodniej Polski.

DNA izolowano metodą fenolową, natomiast jego stężenie oznaczano przy użyciu zestawów ACES 2.0 Human DNA Quantitation (Life Technologies) techniką *dot-blot* [3]. Do amplifikacji 9 *loci* STR techniką STR zastosowano zestaw PowerPlex 2.1 (Promega) zgodnie z instrukcją producenta [13]. Całkowita objętość reakcji PCR, do której użyto 1 ng DNA matrycy w aparacie TRIO-Thermoblock (Biometra) wg parametrów cykli przewidzianych dla bloku PCR Perkin-Elmer 9600, wynosiła 20 μ l.

Po amplifikacji przygotowywano mieszaninę złożoną z 2,5 μ l produktu PCR, 2,5 μ l buforu obciążającego oraz 1 μ l standardu wielkości CXR 600 (Promega), którą

(po denaturacji przez 2 minuty w temperaturze 95°C) наносzono na żel w jego górnej, katodowej części.

Elektroforezę wykonywano w 4,5% poliakryloamidowych żelach sekwencyjnych (4,5 % T, 6% C, 8 M mocznik) z użyciem buforu $1 \times$ TBE oraz aparatu SA 43 (Life Technologies Inc., Gaithersburg, Md, Stany Zjednoczone). Wstępny rozdział prowadzono przez 5 min przy mocy prądu 80 W, a następnie zmniejszono moc do 60 W. Detekcję rozdzielonych produktów PCR znakowanych fluorochromami wykonywano skanerem laserowym FMBIO II firmy Hitachi.

Allele były rozpoznawane w wyniku porównania fragmentów próbki z wzorcową drabiną alleli (umieszczaną w każdym żelu co czwartą linię rozdziału) przy pomocy programów do analizy obrazu FMBIO II Image Analysis 8.0 i StaRcall 3.0.

Wykonano również oznaczenia 9 układów STR z użyciem zestawu PowerPlex 2.1 w 63 przypadkach sądowego ustalania ojcostwa, w których rutynowo oznaczono także 4 *loci* VNTR (D4S139, D5S110, D7S21, D12S11) przy użyciu techniki hybrydyzacyjnej oraz systemu detekcji chemiluminescencji. Zastosowano przy tym wcześniej opisane warunki trawienia próbek DNA enzymem Hinf I, rozdziału elektroforetycznego w żelach agarozowych oraz hybrydyzacji z sondami znakowanymi alkaliczną fosfatazą (PH 30 i LH 1, Life-Technologies; MS 31 i MS 43A, Cellmark Diagnostics) [8]. Wielkość allelicznych fragmentów restrykcyjnych była obliczana przy użyciu systemu komputerowej analizy obrazu z wideokamery Biotech Fischer GmbH w oparciu o rozdział molekularnego markera wielkości MW 100 firmy Gibco-BRL. Częstość fragmentów w przedziałach wielkości ustalano metodą *float bins*, a wielkość przedziału tzw. *window* – 4.8% pz ($\pm 2.4\%$ pz) – wyznaczono doświadczalnie.

Analizę statystyczną otrzymanych wyników badań populacyjnych wykonano przy użyciu metod sprawdzających równowagę wg reguły Hardyego-Weinberga, tj. testem χ^2 ilości homozygot obserwowanych i oczekiwanych [11], testem „exact” [5] oraz testem „likelihood ratio” [14]. Obliczono również szereg parametrów, które określają przydatność tych układów do badań sądowych, tj. współczynnik dyskryminacji (*PD*) [7], współczynnik informacji o polimorfizmie (*PIC*) [1], indeks dyskryminacji (*DI*) [17] oraz średni współczynnik ojcostwa (*MPI*) [2] i szansę wykluczenia (*PE*) [16]. Homogenność badanej populacji z trzema innymi populacjami Polski sprawdzono przy użyciu programu 2-way $R \times C$ wg G. Carmody’ego (Uniwersytet Carleton, Ottawa, Kanada).

WYNIKI I DYSKUSJA

Rezultaty badań 9 układów STR w populacji południowo-wschodniej Polski zestawiono w tabeli 1, z której wynika, że rozkład częstości genotypów dla wszystkich *loci* wykazuje zgodność z regułą Hardyego-Weinberga zarówno w teście homozygotycznym χ^2 (Ho-T), jak i największej wiarygodności (LR-T) oraz „exact” (Ex-T). Na uwagę zasługuje przy tym dość niska wartość prawdopodobieństwa w teście homozygotyczności ($0,2 < p < 0,3$) dla lokus FGA, D21S11 i TH01 przy jednocześnie wysokich wartościach prawdopodobieństw uzyskanych w testach LR i Ex-T, tym bardziej, że w każdym przypadku ilość homozygot oczekiwanych jest mniejsza od ilości homozygot obserwowanych.

W tabeli 2 przedstawiono wyniki badań porównawczych częstości alleli 8 układów STR (autorzy nie znaleźli bowiem wyników badań populacji Polskiej w zakresie układu Penta E) populacji Polski południowo-wschodniej (Lublin) z wynikami badań trzech innych populacji polskich: regionu pomorsko-kujawskiego (Bydgoszcz) [9], południowo-zachodniego (Wrocław) [4] i północnego (Gdańsk) [12]. Wykazano w ten sposób homogenność przebadanej populacji z populacją regionu północnego we wszystkich porównywanych układach.

Natomiast porównanie badanej populacji z populacją regionu południowo-zachodniego wykazało homogenność tylko w zakresie układów D21S11, D8S1179, D3S1358, VWA, TH01, TPOX a z regionem pomorsko-kujawskim w zakresie układów D3S1358, VWA, TH01, TPOX.

Istotnie statystycznie różnice uwidoczniły się w układach FGA i D18S51 przy porównaniu częstości alleli badanego regionu z regionem południowo-zachodniej Polski oraz w układzie FGA przy porównaniu z regionem pomorsko-kujawskim.

Heterogenność populacji badanego regionu (Lublin) i południowo-zachodniej Polski (Wrocław) w układzie D18S51 może wynikać z małej liczby przebadanych osób w populacji wrocławskiej (250 alleli, 125 osób), tym bardziej, że istotne różnice dotyczą jedynie dwóch alleli: 14 i 16, których częstość jest o 3% wyższa w Polsce południowo-wschodniej. Jednocześnie stwierdzono, że w przypadku układu FGA jedynie populacje Gdańska i Lublina są homogenne, zaś porównywanie pozostałych populacji wykazało, że są one heterogenne, przy czym w każdej populacji przebadano więcej niż 200 osób. Informacje na temat układu FGA świadczą przy tym o różnicach w jego heterozygotyzmie (0,90 Bydgoszcz, 0,87 Wrocław, 0,85 Gdańsk i 0,87 Lublin) oraz że największe różnice (8–9%) dotyczą częstości występowania allela 25 (0,16 Bydgoszcz, 0,08 Wrocław, 0,07 Gdańsk i 0,07 Lublin), natomiast mniejsze różnicowanie (ale powyżej 4%) dotyczy alleli: 19, 20, 21, 22 i 24.

Na obecnym etapie badań nie można jednoznacznie rozstrzygnąć, czy heterozygotyczność ta dowodzi istnienia subpopulacji, czy też wynika z innych przyczyn. W chwili obecnej nieuzasadnione wydaje się jedynie traktowanie tej heterogenności jako efektu porównywania zbyt małych populacji. Należy więc zwrócić uwagę na fakt, że allele wspomnianego układu różnią się nie tylko ilością powtórzeń 2 i 4 nukleotydowego motywu, ale wykazują również zmienność w sekwencjach powtórzeń. Ponadto badania populacyjne zostały wykonane przy użyciu różnych starterów (których sekwencje nie są znane w przypadku zestawów multipleksowych firmy Promega i Perkin-Elmer), jak również innych technik rozdziału i analizy produktów amplifikacji (elektroforeza kapilarna lub elektroforeza płytowa z użyciem sekwenatora 377 Perkin Elmer lub FMBIO II Hitachi). Konieczne wydaje się zatem wykonanie międzylaboratoryjnych badań porównawczych, które mogą wyjaśnić, czy istotnie mamy do czynienia z subpopulacjami.

Tabela 3 przedstawia te statystyczne parametry, które określają przydatność poszczególnych układów do badań sądowych. Wynika z nich, że najbardziej informacyjne są układy: Penta E, FGA, D21S11 i D18S51, a najmniej układ TPOX. Natomiast łączne oznaczenie 9 układów multipleksu PowerPlex 2.1 pozwala na uzyskanie szansy wykluczenia = 0,99995 oraz indeksu ojcostwa = 14 783.

Badania 4 układów VNTR i 9 STR w 63 sprawach alimentacyjnych pozwoliło na wykluczenie ojcostwa w 13 przypadkach w co najmniej 3 układach VNTR i 4 układach STR. W pozostałych 50 sprawach, w których ojcostwa nie wykluczono, porówna-

no wartości prawdopodobieństwa ojcostwa obliczone w oparciu o badania VNTR i STR (rycina 1), co było uzasadnione ze względu na to, że w Polsce opinia potwierdzająca ojcostwo pozwanego wymaga uzyskania prawdopodobieństwa ojcostwa wynoszącego powyżej 99,99 % (*PI* powyżej 10 000).

Przedstawione graficznie rezultaty tej analizy (rycina 1) dowodzą, że prawdopodobieństwo ojcostwa wynoszące powyżej 99,99% uzyskano w 39 przypadkach wyłącznie na podstawie wyników badań 9 układów STR, a w 41 przypadkach wyłącznie na podstawie 4 układów VNTR. Bardzo wysokie wartości prawdopodobieństwa ojcostwa uzyskane dla 4 VNTR wytłumaczyć można między innymi wynikiem zastosowania metody *float bin*, która pozwala na ustalanie częstości fragmentów allelicznych w przedziałach wielkości co 4,8% pz. Podstawą do użycia takiej metody jest jednak uzyskiwanie powtarzalnych wyników, których błąd pomiaru wielkości fragmentów restrykcyjnych nie przekracza $\pm 2,4\%$ pz. W tym celu rozdziały elektroforetyczne fragmentów restrykcyjnych pozwanego, dziecka i matki należy prowadzić na tym samym żelu agarozowym o długości 35 cm z użyciem markera wielkości MW 100 przy stałych parametrach. Taki sposób fenotypowania VNTR jest bowiem w pełni uzasadniony i polecany także przez innych autorów [6, 10].

Z ryciny 1 wynika również, że w oparciu o badania 4 *loci* VNTR w 9 przypadkach na 50 uzyskano prawdopodobieństwo mniejsze niż 99,99%, podczas gdy badania 9 układów STR dały takie prawdopodobieństwa w 11 przypadkach. Wysokie wartości prawdopodobieństwa ojcostwa (powyżej 99,999%) uzyskano przy tym w 18 sprawach, oceniając jedynie wyniki 4 VNTR, a w 22 przypadkach jego podstawą były badania 9 STR. Natomiast łączne użycie 9 układów STR i 4 VNTR dało prawdopodobieństwo ojcostwa wynoszące powyżej 99,99% we wszystkich przypadkach, co pozwalało na wydanie opinii potwierdzającej ojcostwo pozwanego.

Zastąpienie 4 układów VNTR poszerzonym badaniem układów STR ma szereg uzasadnień, do których można zaliczyć: krótki czas analizy, jej niższy koszt, możliwość pobierania do analizy innego materiału niż krew oraz ujednoczenie nomenklatury genotypów. Z przeprowadzonych przez autorów prac badawczych wynika jednak, że wartość oznaczeń 4 VNTR w ekspertyzie dotyczącej ustalenia ojcostwa można zrównoważyć, włączając do niej około 9 układów STR o charakterystyce podobnej do badanych w multipleksie PowerPlex 2.1. Zastosowanie dwóch różnych technik, tj. amplifikacji w przypadku STR i hybrydyzacji w przypadku VNTR, zapewnia jednak duże „bezpieczeństwo” opiniodawcze między innymi dlatego, że badając układy STR amplifikujemy DNA, z czym wiąże się możliwość powstania błędu, natomiast badanie VNTR nie zawiera tego ryzyka.

Pracę wykonano w ramach projektu badawczego DS/380 finansowanego przez Akademię Medyczną w Lublinie.