Problems of Forensic Sciences 2019, vol. 117, 49–61

ISSN 1230-7483

ANALYSIS OF THE FREQUENCY OF OCCURRENCE IN THE POLISH POPULATION OF ALLELES OF 21 GENETIC MARKERS IN THE GLOBALFILER KIT

Maria WRÓBEL, Agnieszka PARYS-PROSZEK, Magdalena MARCIŃSKA, Grażyna BA, Andrzej SEKUŁA, Marek KOWALCZYK, Miłosz JANUŁA, Andrzej DONIEC, Tomasz KUPIEC

Institute of Forensic Research, Kraków, Poland

Abstract

The purpose of this paper was to calculate allele frequencies of markers which are included in the GlobalFiler kit (ThermoFisher) for the Polish population. DNA quantification was carried out followed by PCR amplification and capillary electrophoresis of a set of 164 samples originating from non-related Polish individuals. Statistical analysis was conducted with GenoProof 3 (Qualitype).

The set of samples proved to be a good representation for the Polish population. Statistical analysis showed that parameters are acceptable. Allele frequencies were calculated for all of the autosomal markers. Implications: the database can be used for standard calculations in forensic genetics to estimate the power of evidence.

Keywords

Population study; GlobalFiler; Kinship; Biostatistics.

Received 9 February 2019; accepted 17 December 2019

Introduction

Due to differences in the frequency of occurrence of alleles of STR systems in human populations, the value of statistical factors obtained in genetic analyses of biological traces depends on the reference database used for calculations. The key to the correctness of statistical calculations in this respect is the application of a population database that is consistent with or is as close as possible to the biogeographical origin of the tested sample. The continuous development of methods used in forensic genetics and modifications of marker panels in available kits make it necessary to continuously update the available databases for a given geographical area. Laboratories performing genetic identification tests based on the analysis of STR type markers have at their disposal sets of reagents allowing for the determination of a dozen or so, or even several dozen microsatellite markers (including loci recommended by international organizations) belonging to the ESS (European standard set loci) group containing TH01, vWA, FGA, D21S11, D3S1358, D8S1179, and D18S51, and minisatellite loci, including D10S1248, D1S1656, D12S391, D2S441, and D22S1045 (Parys-Proszek, Kupiec, Wolańska-Nowak and Branicki, 2010; Welch et al., 2012). The new generation of reagents for DNA amplification is characterized by high sensitivity of analysis, significant resistance to inhibition and the ability to analyse even up to 27 STR markers simultaneously. The GlobalFiler kit is amongst the most frequently used commercial kits in forensic laboratories that is available on the market. It enables simultaneous examination of 22 autosomal markers and two markers located on the Y chromo-

some. Statistical analysis of test results is performed by means of computer programmes designed for this purpose, which serve to determine random match probability (RMP) of a given profile in the population or the likelihood ratio (LR). A necessary condition for using a frequency database for a given population is to fulfil the set criteria and to obtain appropriate values of population rates and Hardy-Weinberg equilibrium for particular loci (Butler, 2015; Chakraborty, 1992; Evett, Gill, 1991; Foreman, Evett, 2001; Schneider et al., 2009).

Aim

The aim of the study was to develop a GlobalFiler allele frequency database for the Polish population for individual loci and to determine the key parameters necessary for applying the obtained database in the routine work of experts during the preparation of forensic genetic opinions.

Materials and methods

The material for the study consisted of 164 DNA samples originating from people from the Polish population who were unrelated to each other. 85 samples were collected from men and 79 from women. Samples containing isolated genetic material were subjected to DNA quantification with the use of Quantifiler Human or Quantifiler Trio (by Thermo Fisher) kits and the 7500 Real Time PCR System. Next, DNA amplification was performed using the GlobalFiler kit (by Thermo Fisher), which allows simultaneous amplification of twenty-four STR type markers (including markers that are characteristic for the male sex – DYS391 and Yindel). 1 ng of DNA matrix was used for the reaction. Next, the samples were subjected to capillary electrophoresis using the ABI Prism 3500xL sequencer (from Life Technologies) with a 24-second injection time. The samples were analyzed according to analysis parameters established during validation using GeneMapper® ID-X (version 1.5, by Thermo Fisher) carried out in the Laboratory of Forensic Genetics, Institute of Forensic Research. Statistical analysis was performed using GenoProof 3 (Qualitype) software. The frequency database was created using the *population study* tool. For the purpose of this analysis, statistical parameters such as power of discrimination (PD), level of heterozygosity (HET) and homozygosity (h), compatibility with the Hardy-Weinberg equilibrium, and cumulative random match probability (GenoProof®3 Theory Manual, 2014) were calculated. The frequency value for rare alleles was defined at the level of 0.001, as recommended by the German Stain Commission (Schneider et al., 2009).

Results

The results of calculations of the frequency of occurrence of alleles in the studied Polish population have been presented in Table 1. Statistical parameters for 20 analysed markers are presented in Table 2. 157 samples were used to analyze the frequency of alleles of the SE33 marker, while for the remaining markers all 164 samples were analyzed. The results presented in Table 1 indicate that the examined loci fulfil the Hardy-Weinberg equilibrium criterion.

Discussion

The key indicator of the correctness of the data collected in the analyzed database is compatibility with the Hardy-Weinberg equilibrium criterion, which makes it possible to use the obtained data to calculate the frequency of genotypes in forensic analyses.

As predicted, the obtained results of allele frequency (within shared markers) do not deviate significantly from the parameters obtained for "geographically close" populations (Table 3; Hantschel, Hausmann, Lederer, Martus, Betz, 1999; Mornhinweg, Luckenbach, Ritter, 1998; Reichenpfader, Immel, Klintschar, 2003; Seider, Fimmers, Betz, Lederer, 2010). Statistical analysis using the χ^2 Test of Independence (Table 4) showed a statistically significant correlation between allele frequency in the Polish population and geographically close populations for the markers D21S11 (Czech and Slovak population), D22S1045 (German and Slovak population), D16S539 (German and Slovak population), D18S51 (German population), D19S433 (German and Slovak population) and D2S441 (German population). However, after applying Yates's correction, as suggested by some researchers, the test showed a statistically significant correlation for the marker D16S539. For both approaches, the obtained values support the rejection of the hypothesis about a lack of correlation; they are slightly above the critical value. Therefore, the obtained results do not exclude the use of the frequency of alleles in the Czech, German and Slovakian populations to calculate the value of evidence from genetic analysis, but they indicate even more strongly the advisability of using a dedicated database for the discussed population.

A questionable issue is the application of vWA and D12S391 genetic markers in calculations for genetic identification purposes; these markers are located on chromosome 12 at a distance of 6Mb from each other, which carries with it the risk of lack of inheritance's independence. According to literature reports, it is recommended to perform calculations using both markers with particular caution (O'Connor, Tillmar, 2012) or to exclude one of them from the statistical analysis. The SE33 marker is characterized by a very high degree of heterozygosity, which implies a higher diversity of alleles and by the same token reduces the chance of a random repetition of the genotype. Such a relationship has already been observed for this locus (Borsuk, Gettings, Steffen, Kiesler, Vallone, 2017; Davis et al., 2012) and is its characteristic feature in many populations.

Conclusions

The analyzed population database fulfils criteria for a genetic database of allele frequency in the studied population, which allows it to be used for bio-statistical calculations for forensic genetic purposes. The obtained values of statistical parameters and the number of tested samples are sufficient to relate the obtained results to the entire Polish population.

References

1. Borsuk, L. A., Gettings, K., Steffen, C. R., Kiesler, K. M., Vallone, P. (2017). Sequencing of the highly polymorphic STR locus SE33. *Forensic Science International: Genetics Supplement Series*, *6*, 322–323.
2. Butler, J. M. (2015). *Advanced topics in forensic DNA typing: Interpretation*. San Diego: Elsevier.
3. Chakraborty, R. (1992). Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. *Human Biology*, *6*, 141–159.
4. Davis, C., Ge, J., King, J., Malik, N., Weirich, V., Eisenberg, A. J., Budowle, B. (2012). Variants observed for STR locus SE33: A concordance study. *Forensic Science International: Genetics*, *6*(4), 494–497.
5. Evett, I.W., Gill P. (1991). A discussion of the robustness of methods for assessing the evidential value of DNA single locus profiles in crime investigations. *Electrophoresis*, *12*, 226–230.
6. Foreman, L., Evett, I. (2001). Statistical analyses to support forensic interpretation for a new ten-locus STR profiling system. *International Journal of Legal Medicine*, *114*, 147–155.
7. GenoProof®3 Theory Manual, Qualitype GmbH 3.Edition, Dresden, 2014 Ver.2014-07-24
8. Hantschel, M., Hausmann, R., Lederer, T., Martus, P., Betz, P. (1999). Population genetics of nine short tandem repeat (STR) loci – DNA typing using the AmpFlSTR profiler PCR amplification kit. *International Journal of Legal Medicine*, *112*(6), 393–395.
9. Mornhinweg, E., Luckenbach, C., Ritter, H. (1998). D3S1358 and D8S1179: Analysis and allele frequencies in a South German population. *Progress in Forensic Genetics*, *7*, 315–317.
10. O'Connor, K. L., Tillmar, A. O. (2012). Effect of linkage between vWA and D12S391 in kinship analysis. *Forensic Science International: Genetics*, *6*, 840–844.
11. Parys-Proszek, A., Kupiec, T.,Wolańska-Nowak, P., Branicki, W. (2010). Genetic variation of 15 autosomal STR loci in a population sample from Poland. *Legal Medicine*, *12*, 246–248.
12. Reichenpfader, B., Immel, U., Klintschar, M. (2003). Population data on the AmpFlSTR SGM plus PCR amplification kit in Germans and Austrians. *Forensic Science International*, *132*(1), 84–86.
13. Schneider, P. M., Fimmers, R., Keil, W., Molsberger, G., Patzelt, D., Pflug, W., Rothämel, T., Schmitter, H., Schneider, H., Brinkmann, B. (2009). The German Stain Commission: recommendations for the interpretation of mixed stains. *International Journal of Legal Medicine*, *123*(1), 1–5.
14. Seider, T., Fimmers, R., Betz, P., Lederer, T. (2010). Allele frequencies of the five miniSTR loci D1S1656, D2S441, D10S1248, D12S391 and D22S1045 in a German population sample. *Forensic Science International: Genetics*, *4*(5), 159–160.
15. Welch, L.A., Gill, P., Phillips, C., Ansell, R., Morling, N., Parson, W., Palo, J. U., Bastisch, I. (2012). European Network of Forensic Science Institutes (ENFSI): evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers, *Forensic Science International: Genetics, 6,* 819–826.

Corresponding author
Maria Wróbel
Institute of Forensic Research
ul. Westerplatte 9
PL 31-033 Kraków
e-mail: mwrobel@ies.krakow.pl

Table 1
Frequency of occurrence of alleles of genetic markers included in the Global Filer Kit in the Polish population

	CSF 1PO	D10S 1248	D12 S391	D13 S317	D16 S539	D18 S51	D19 S433	D1S 1656	D21 S11	D22S 1045	D2S 1338	D2S 441	D3S 1358	D5S 818	D7S 820	D8S 1179	FGA	SE33	TH01	TPOX	vWA	
5	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	5
6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.235	–	–	6
7	–	–	–	–	–	–	–	–	–	–	–	–	–	0.003	0.018	–	–	–	0.122	–	–	7
8	0.012	–	–	0.171	0.006	–	–	0.003	–	–	–	–	–	–	0.143	0.018	–	–	0.122	0.588	–	8
8.3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.003	–	–	8.3
9	0.043	–	–	0.067	0.076	–	–	–	–	–	–	0.003	–	0.064	0.143	0.006	–	–	0.183	0.091	–	9
9.3	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	0.332	–	–	9.3
10	0.290	–	–	0.040	0.027	0.009	–	0.003	–	–	–	0.238	–	0.058	0.305	0.079	–	–	0.003	0.030	–	10
11	0.265	–	–	0.363	0.284	–	–	0.101	–	0.146	–	0.314	–	0.348	0.216	0.070	–	–	–	0.253	–	11
11.3	–	–	–	–	–	–	–	–	–	–	–	0.037	–	–	–	–	–	–	–	–	–	11.3
12	0.320	0.021	–	0.232	0.354	0.073	0.122	0.146	–	0.034	–	0.018	–	0.357	0.122	0.168	–	–	–	0.037	–	12
13	0.061	0.235	–	0.076	0.226	0.116	0.226	0.073	–	0.003	–	0.040	0.006	0.165	0.049	0.293	–	0.010	–	–	–	13
13.2	–	–	–	–	–	–	0.012	–	–	–	–	–	–	–	–	–	–	–	–	–	–	13.2
14	0.009	0.308	–	0.049	0.027	0.168	0.299	0.070	–	0.055	–	0.314	0.143	0.006	0.003	0.247	–	0.029	–	–	0.098	14
14.2	–	–	–	–	–	–	0.030	–	–	–	–	–	–	–	–	–	–	–	–	–	–	14.2
15	–	0.247	0.021	0.003	–	0.180	0.210	0.128	–	0.302	0.003	0.030	0.250	–	–	0.098	–	0.032	–	–	0.101	15
15.2	–	–	–	–	–	–	0.049	–	–	–	–	–	–	–	–	–	–	–	–	–	–	15.2
15.3	–	–	–	–	–	–	–	0.034	–	–	–	–	–	–	–	–	–	–	–	–	–	15.3
16	–	0.152	0.024	–	–	0.152	0.034	0.134	–	0.357	0.040	0.006	0.210	–	–	0.021	–	0.025	–	–	0.183	16
16.2	–	–	–	–	–	–	0.015	–	–	–	–	–	–	–	–	–	–	0.003	–	–	–	16.2
16.3	–	–	–	–	–	–	–	0.037	–	–	–	–	–	–	–	–	–	–	–	–	–	16.3
17	–	0.030	0.079	–	–	0.137	–	0.040	–	0.082	0.204	–	0.207	–	–	–	–	0.054	–	–	0.253	17
17.3	–	–	0.027	–	–	–	–	0.159	–	–	–	–	–	–	–	–	–	–	–	–	–	17.3
18	–	0.003	0.171	–	–	0.070	0.003	0.006	–	0.018	0.079	–	0.171	–	–	–	0.009	0.086	–	–	0.241	18
18.3	–	–	0.021	–	–	–	–	0.055	–	–	–	–	–	–	–	–	–	–	–	–	–	18.3
19	–	0.003	0.113	–	–	0.058	–	–	–	0.003	0.128	–	0.009	–	–	–	0.073	0.086	–	–	0.116	19
19.3	–	–	0.012	–	–	–	–	0.012	–	–	–	–	–	–	–	–	–	–	–	–	–	19.3
20	–	–	0.125	–	–	0.018	–	–	–	–	0.125	–	0.003	–	–	–	0.152	0.054	–	–	0.009	20
20.3	–	–	0.003	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	20.3
21	–	–	0.113	–	–	0.015	–	–	–	–	0.027	–	–	–	–	–	0.226	0.010	–	–	–	21

	CSF 1PO	D10S 1248	D12 S391	D13 S317	D16 S539	D18 S51	D19 S433	D1S 1656	D21 S11	D22S 1045	D2S 1338	D2S 441	D3S 1358	D5S 818	D7S 820	D8S 1179	FGA	SE33	TH01	TPOX	vWA	
21.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.016	—	—	—	21.2
22	—	—	0.155	—	—	—	—	—	—	—	0.034	—	—	—	—	—	0.155	0.006	—	—	—	22
22.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.012	0.032	—	—	—	22.2
23	—	—	0.088	—	—	0.003	—	—	—	—	0.104	—	—	—	—	—	0.098	0.003	—	—	—	23
23.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.012	0.025	—	—	—	23.2
24	—	—	0.027	—	—	—	—	—	—	—	0.079	—	—	—	—	—	0.152	—	—	—	—	24
24.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.041	—	—	—	24.2
25	—	—	0.012	—	—	—	—	—	—	—	0.162	—	—	—	—	—	0.067	—	—	—	—	25
25.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.045	—	—	—	25.2
26	—	—	0.003	—	—	—	—	—	—	—	0.015	—	—	—	—	—	0.037	—	—	—	—	26
26.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.061	—	—	—	26.2
27	—	—	—	—	—	—	—	—	0.027	—	—	—	—	—	—	—	0.006	—	—	—	—	27
27.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.118	—	—	—	27.2
28	—	—	—	—	—	—	—	—	0.159	—	—	—	—	—	—	—	—	—	—	—	—	28
28.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.086	—	—	—	28.2
29	—	—	0.003	—	—	—	—	—	0.189	—	—	—	—	—	—	—	—	0.003	—	—	—	29
29.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.073	—	—	—	29.2
30	—	—	—	—	—	—	—	—	0.271	—	—	—	—	—	—	—	—	—	—	—	—	30
30.2	—	—	—	—	—	—	—	—	0.064	—	—	—	—	—	—	—	—	0.048	—	—	—	30.2
31	—	—	—	—	—	—	—	—	0.049	—	—	—	—	—	—	—	—	—	—	—	—	31
31.2	—	—	—	—	—	—	—	—	0.079	—	—	—	—	—	—	—	—	0.019	—	—	—	31.2
32	—	—	—	—	—	—	—	—	0.003	—	—	—	—	—	—	—	—	—	—	—	—	32
32.2	—	—	—	—	—	—	—	—	0.119	—	—	—	—	—	—	—	—	0.025	—	—	—	32.2
33	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.003	—	—	—	33
33.2	—	—	—	—	—	—	—	—	0.040	—	—	—	—	—	—	—	—	—	—	—	—	33.2
36	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.003	—	—	—	36
36.2	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	0.003	—	—	—	36.2
36.3	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	—	36.3
ALLEL	CSF 1PO	D10S 1248	D12 S391	D13 S317	D16 S539	D18 S51	D19 S433	D1S 1656	D21 S11	D22S 1045	D2S 1338	D2S 441	D3S 1358	D5S 818	D7S 820	D8S 1179	FGA	SE33	TH01	TPOX	vWA	

Table 2
Factors of statistical analysis of the Polish population database within studied markers

		CSF1PO	D10S1248	D12S391	D13S317	D16S539	D18S51	D19S433	D1S1656	D21S11	D22S1045	D2S1338	D2S441	D3S1358	D5S818	D7S820	D8S1179	FGA	SE33	TH01	TPOX	vWA
basic analysis parameters	PIC	0.691	0.725	0.878	0.740	0.692	0.856	0.767	0.882	0.817	0.711	0.860	0.698	0.770	0.669	0.774	0.777	0.841	0.934	0.736	0.526	0.785
	homozygosity (h)	0.263	0.236	0.112	0.229	0.264	0.130	0.204	0.109	0.164	0.251	0.127	0.258	0.200	0.283	0.199	0.196	0.143	0.063	0.229	0.421	0.189
	heterozygosity (HET)	0.738	0.764	0.888	0.771	0.736	0.870	0.796	0.891	0.836	0.749	0.873	0.742	0.801	0.717	0.802	0.804	0.857	0.937	0.771	0.579	0.811
	power of exclusion (PE)	0.489	0.535	0.771	0.547	0.487	0.734	0.591	0.778	0.668	0.508	0.741	0.496	0.600	0.456	0.602	0.606	0.709	0.872	0.547	0.267	0.620
	PI	1.905	2.123	4.471	2.186	1.896	3.839	2.450	4.605	3.054	1.992	3.941	1.938	2.506	1.769	2.519	2.547	3.504	7.982	2.186	1.188	2.650
population parameters	combined paternity index (CPI)	825218781.8																				
	combined power of discrimination (CPD)	1																				
	combined power of exclusion (CPE)	0.9999999990																				
PD		0.840	0.873	0.973	0.874	0.836	0.963	0.904	0.975	0.939	0.849	0.965	0.844	0.911	0.809	0.909	0.911	0.955	0.992	0.879	0.522	0.920
MEC		0.498	0.541	0.774	0.570	0.501	0.737	0.604	0.780	0.678	0.532	0.745	0.510	0.600	0.476	0.612	0.619	0.714	0.873	0.556	0.337	0.626
Hardy-Weinberg equilibrium	X^2 (calculated value)	22.46	13.76	141.61	33.17	21.80	85.54	53.71	109.37	48.18	37.56	45.46	42.82	22.31	15.50	39.74	43.88	50.24	414.81	10.71	16.70	12.85
	corresponding with probability	37.3%	98.9%	35.4%	23.0%	41.1%	5.3%	17.5%	36.6%	34.6%	39.7%	97.5%	20.2%	76.7%	79.7%	7.0%	17.2%	92.5%	9.3%	96.8%	8.1%	91.4%
	degrees of freedom	21	28	136	28	21	66	45	105	45	36	66	36	28	21	28	36	66	378	21	10	21
Test of accuracy/ precision (Monte Carlo algorithm)	P value	0.44	0.88	0.43	0.27	0.89	0.03	0.67	0.24	0.35	0.47	0.99	0.55	0.42	0.51	0.34	0.31	0.85	0.55	0.96	0.08	0.94
	deviation	0.0001	0.0009	0.0012	0.0010	0.0008	0.0004	0.0011	0.0012	0.0013	0.0012	0.0002	0.0012	0.0012	0.0013	0.0011	0.0011	0.0009	0.0011	0.0004	0.0006	0.0006

PIC – (polymorphic information content) measure of the distinctiveness of a marker; PI** – paternity index of degree of relationship indicates how much more likely it is that the genotype of the child will support the first hypothesis of the analysis of consanginuity (the alleged parent is the biological parent of the given child) than the alternative hypothesis (another unrelated person is the biological parent of the child); PD – (*power of discrimination*) – the probability that two persons in a given population who are unrelated to each other have different genotypes in a given marker. MEC – (*mean exclusion chance*) – the probability with which population data allow for correct exclusion of an alleged father/mother.

Table 3
Allele frequencies of "geographically close" populations. Data for Czech, German and Slovakian populations originate from: //strider.online/frequencies

VWA				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
11		0.00075529		
13		0.0022659	0.0020243	
14	0.1	0.097432	0.11943	0.097561
15	0.0975	0.10347	0.11943	0.1006098
16	0.175	0.2213	0.19231	0.1829268
17	0.3125	0.25453	0.2753	0.2530488
18	0.2275	0.22054	0.20445	0.2408537
19	0.0725	0.086103	0.076923	0.1158537
20	0.015	0.01284	0.010122	0.0091463
21		0.00075529		

D21S11				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
24.2		0.00075529	0.0020243	
25		0.00075529		
25.2		0.00075529		
26		0.00075529		
27	0.0125	0.03852	0.01417	0.027439
28	0.1525	0.15106	0.13765	0.1585366
28.2		0.00075529		–
29	0.26	0.2077	0.25506	0.1890244
29.2		0.00075529	0.0020243	–
30	0.2	0.22357	0.19231	0.2713415
30.2	0.0275	0.040786	0.07085	0.0640244
31	0.0425	0.077795	0.060729	0.0487805
31.2	0.0875	0.089879	0.078947	0.0792683
32	0.0075	0.017372	0.016194	0.0030488
32.2	0.1375	0.1065	0.11943	0.1189024
33	0.005	0.00075529		–
33.1	0.0025	0.00075529		
33.2	0.06	0.032477	0.032389	0.0396341
34				
34.2	0.005	0.0075529	0.018219	
35.2		0.00075529		

TH01				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
5		0.0015106		
6	0.2375	0.22659	0.23077	0.2347561
7	0.16	0.14804	0.14372	0.1219512
8	0.1	0.13897	0.12348	0.1219512
8.3		0.00075529	0.0020243	0.0030488
9	0.1775	0.17523	0.2004	0.1829268
9.3	0.3225	0.29305	0.29555	0.3323171
10	0.0025	0.015861	0.0040486	0.0030488

FGA				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	199	662	247	164
16			0.0020243	
17		0.0022659	0.0020243	
18	0.01005	0.015861	0.022267	0.0091463
19	0.067839	0.077795	0.093117	0.0731707
20	0.11306	0.13973	0.18219	0.152439
20.2		0.0030211		
21	0.17337	0.18202	0.15992	0.2256098
21.2		0.0022659	0.0020243	
22	0.19347	0.18202	0.16599	0.1554878
22.2	0.015075	0.0067976	0.018219	0.0121951
23	0.15578	0.14577	0.11741	0.097561
23.2	0.020101	0.005287	0.0040486	0.0121951
24	0.14573	0.12915	0.1579	0.152439
24.2		0.00075529		
25	0.082915	0.07855	0.054656	0.0670732
25.2		0.0015106		
26	0.020101	0.021903	0.016194	0.0365854
27		0.0037764	0.0020243	0.0060976
28	0.0025126	0.00075529		
30		0.00075529		

D2S1338				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
1	0.0025	0.00075529		
15	0.0025	0.00075529		0.0030488
16	0.05	0.043807	0.034413	0.0396341
17	0.19	0.22432	0.21862	0.2042683
18	0.0975	0.085347	0.095142	0.0792683
19	0.1075	0.10272	0.09919	0.1280488
20	0.1375	0.14426	0.11336	0.125
21	0.0375	0.035499	0.024292	0.027439
22	0.025	0.024169	0.022267	0.0335366
23	0.115	0.11556	0.12146	0.1036585
24	0.1025	0.092145	0.10729	0.0792683
25	0.11	0.10952	0.13968	0.1615854
26	0.02	0.020393	0.022267	0.0152439
27	0.0025	0.00075529	0.0020243	

D8S1179				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	246	164
8	0.0075	0.01284	0.010163	0.0182927
9	0.005	0.01284	0.010163	0.0060976
10	0.055	0.087613	0.044715	0.0792683
11	0.1	0.077795	0.099594	0.070122
12	0.1525	0.14199	0.14431	0.1676829
13	0.35	0.31269	0.34146	0.2926829
14	0.2125	0.19864	0.25406	0.2469512
15	0.0975	0.11933	0.075203	0.097561
16	0.02	0.030211	0.01626	0.0213415
17		0.0060423	0.004065	

D18S51				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
10	0.0075	0.01284	0.0080972	0.0091463
11	0.0075	0.0090634	0.010122	
12	0.0925	0.14426	0.09919	0.0731707
13	0.1075	0.13293	0.11943	0.1158537
14	0.1575	0.15181	0.16194	0.1676829
15	0.155	0.16465	0.13158	0.179878
16	0.175	0.1216	0.19231	0.152439
16.2			0.0020243	
17	0.1475	0.12009	0.11741	0.1371951
18	0.07	0.071752	0.068826	0.070122
19	0.035	0.035499	0.044534	0.0579268
20	0.0275	0.017372	0.02834	0.0182927
21	0.01	0.010574	0.0080972	0.0152439
22	0.0075	0.0067976	0.0060729	
23		0.00075529		0.0030488
24			0.0020243	

D19S433				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
1		0.00075529		
11	0.0025	0.005287		
12	0.085	0.071752	0.091093	0.1219512
12.2		0.0022659	0.0040486	
13	0.265	0.23338	0.29352	0.2256098
13.2	0.02	0.017372	0.022267	0.0121951
14	0.3125	0.35952	0.32793	0.2987805
14.2	0.0075	0.022659	0.012146	0.0304878
15	0.1825	0.15634	0.15182	0.2103659
15.2	0.055	0.046073	0.020243	0.0487805
16	0.035	0.049094	0.044534	0.0335366
16.2	0.03	0.02568	0.020243	0.0152439
17		0.0030211		
17.2	0.005	0.0037764	0.0080972	
18.2		0.0022659	0.0040486	0.0030488

D22S1045				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
10		0.00075529		
11	0.1325	0.12613	0.20445	0.1463415
12	0.01	0.011329	0.020243	0.0335366
13	0.005	0.0083082	0.0080972	0.0030488
14	0.06	0.050604	0.02834	0.054878
15	0.3525	0.37236	0.37045	0.3018293
16	0.33	0.33761	0.27733	0.3567073
17	0.0825	0.084592	0.076923	0.0823171
18	0.0275	0.0067976	0.010122	0.0182927
19		0.0015106	0.0040486	0.0030488

D1S1656				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
8				0.0030488
10		0.0022659	0.0060729	0.0030488
11	0.0925	0.093656	0.087044	0.1006098
12	0.165	0.12311	0.1498	0.1463415
13	0.0775	0.072508	0.11741	0.0731707
14	0.0675	0.083082	0.064777	0.070122
14.3	0.0025	0.0022659	0.0020243	
15	0.1275	0.14048	0.12146	0.1280488
15.3	0.035	0.057402	0.04251	0.0335366
16	0.0975	0.10498	0.12146	0.1341463
16.3	0.0425	0.060423	0.036437	0.0365854
17	0.06	0.056647	0.060729	0.0396341
17.1		0.0015106		
17.3	0.16	0.13217	0.12348	0.1585366
18	0.005	0.0060423	0.0040486	0.0060976
18.3	0.0575	0.048338	0.060729	0.054878
19		0.00075529		
19.3	0.0075	0.013595	0.0020243	0.0121951
20.3	0.0025	0.00075529		

D3S1358				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
11		0.00075529		
12		0.00075529		
13	0.0025	0.0037764		0.0060976
14	0.105	0.11556	0.089069	0.1432927
15	0.255	0.23112	0.26518	0.25
16	0.2575	0.24622	0.27126	0.2103659
17	0.2225	0.2281	0.19433	0.2073171
18	0.1475	0.15559	0.16802	0.1707317
19	0.0075	0.018127	0.012146	0.0091463
20	0.0025			0.0030488

D12S391				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
15	0.035	0.036254	0.026316	0.0213415
16	0.0125	0.034743	0.016194	0.0243902
16.3		0.00075529		
17	0.105	0.10045	0.11134	0.0792683
17.3	0.015	0.016616	0.022267	0.027439
18	0.1775	0.17296	0.2166	0.1707317
18.3	0.0125	0.017372	0.016194	0.0213415
19	0.1125	0.10347	0.11336	0.1128049
19.3	0.0025	0.0090634	0.0040486	0.0121951
20	0.1175	0.13973	0.12348	0.125
20.2			0.0020243	
20.3	0.0025	0.0030211		0.0030488
21	0.145	0.11027	0.10729	0.1128049
21.3	0.0025		0.0020243	
22	0.1025	0.12009	0.11134	0.1554878
22.1		0.00075529		
23	0.1075	0.080816	0.10526	0.0884146
24	0.0375	0.037764	0.016194	0.027439
25	0.0125	0.014351	0.0040486	0.0121951
26		0.0015106	0.0020243	0.0030488

D10S1248				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
1		0.00075529		
11	0.005	0.0060423		
12	0.0325	0.027946	0.030364	0.0213415
13	0.2175	0.26586	0.21255	0.2347561
14	0.2925	0.30438	0.32186	0.3079268
15	0.215	0.20166	0.24291	0.2469512
16	0.1875	0.15408	0.11943	0.152439
17	0.05	0.035499	0.064777	0.0304878
18		0.0037764	0.0080972	0.0030488
19				0.0030488

D16S539				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
8	0.0075	0.021148	0.0080972	0.0060976
9	0.1	0.11178	0.13765	0.0762195
10	0.07	0.064199	0.072874	0.027439
11	0.2925	0.29079	0.28542	0.2835366
12	0.3225	0.28852	0.28745	0.3536585
13	0.1825	0.19335	0.18016	0.2256098
14	0.0225	0.030211	0.02834	0.027439
15	0.0025			

D2S441				
Allele	CZECH REP.	GERMANY	SLOVAKIA	GF PL
	200	662	247	164
8		0.0015106		
9	0.0025			0.0030488
10	0.15	0.18504	0.18016	0.2378049
10.3			0.0020243	
11	0.3325	0.3497	0.34818	0.3140244
11.3	0.06	0.05287	0.060729	0.0365854
12	0.0275	0.044562	0.032389	0.0182927
12.3		0.0015106	0.0040486	
13	0.0225	0.023414	0.02834	0.0396341
14	0.3475	0.29381	0.30162	0.3140244
15	0.0525	0.046073	0.038462	0.0304878
16	0.005	0.0015106	0.0040486	0.0060976

SE33				
Allele	CZECH REP.	HUNGARY	SLOVAKIA	GF PL
	200	223	245	157
11		0.0022422		
12	0.0025	0.0022422	0.0020408	
13	0.0025	0.0067265	0.0061224	0.0095541
13.2		0.0044843	0.0020408	
14	0.0325	0.049327	0.014286	0.0286624
14.2	0.0025	0.0044843		
15	0.055	0.044843	0.04898	0.0318471
15.2	0.0025	0.0022422		
16	0.045	0.058296	0.042857	0.0254777
16.2			0.0020408	0.0031847
16.3	0.0025	0.0044843		
17	0.0725	0.071749	0.079592	0.0541401
18	0.0775	0.076233	0.069388	0.0859873
18.3		0.0022422		
19	0.0775	0.087444	0.093878	0.0859873
19.2	0.0025	0.0044843	0.0040816	
20	0.0725	0.044843	0.069388	0.0541401
20.2	0.0025	0.0044843	0.0061224	
21	0.02	0.033632	0.022449	0.0095541
21.2	0.01	0.0067265	0.014286	0.0159236
22	0.015	0.0022422	0.0061224	0.0063694
22.2	0.0375	0.020179	0.020408	0.0318471
23			0.0020408	0.0031847
23.2	0.0425	0.035874	0.040816	0.0254777
24			0.0020408	
24.2	0.0425	0.022422	0.042857	0.0414013
25	0.0025			
25.2	0.03	0.044843	0.028571	0.044586
26.2	0.0475	0.051569	0.057143	0.0605096
27.2	0.05	0.08296	0.04898	0.1178344
28			0.0020408	
28.2	0.08	0.060538	0.085714	0.0859873
29			0.0061224	0.0031847
29.2	0.0475	0.06278	0.059184	0.0732484
29.3			0.0020408	
30.2	0.0425	0.051569	0.05102	0.0477707
31.2	0.0425	0.015695	0.032653	0.0191083
32.2	0.02	0.020179	0.010204	0.0254777
33	0.01	0.0067265	0.0081633	0.0031847
33.2	0.005	0.0022422	0.0081633	
34		0.0022422	0.0020408	
34.2	0.005	0.0067265	0.0040816	
35			0.0020408	
36				0.0031847
36.2				0.0031847

Table 4

Chi-Square Test of Independence of two populations (together with Yates's correction). Data for the Polish population are juxtaposed with data for the Czech (Cz), German (Ger) and Slovakian (Sl) populations. Values above tabulated data are shown in red, indicating rejection of the hypothesis about a lack of independence

marker		Chi^2	Chi^2 + Yates	df	tabular value of Chi^2
Vwa	Pl *vs* Cz	6.64	5.50	6	12.6
	Pl *vs* Ger	5.39	4.89	9	16.9
	Pl *vs* Sl	7.04	5.42	7	14.1
TH01	Pl *vs* Cz	3.94	2.54	6	12.6
	Pl *vs* Ger	8.20	5.25	7	14.1
	Pl *vs* Sl	2.02	1.83	6	12.6
D21S11	Pl *vs* Cz	22.54	15.71	12	21.0
	Pl *vs* Ger	19.73	18.74	19	30.1
	Pl *vs* Sl	22.98	17.30	12	21.0
D2S1338	Pl *vs* Cz	9.91	6.49	13	22.4
	Pl *vs* Ger	13.11	10.66	13	22.4
	Pl *vs* Sl	9.08	5.02	12	21.0
D8S1179	Pl *vs* Cz	8.25	6.47	8	15.5
	Pl *vs* Ger	10.28	7.31	9	16.9
	Pl *vs* Sl	12.29	8.64	9	16.9
D22S1045	Pl *vs* Cz	9.00	5.88	8	15.5
	Pl *vs* Ger	18.52	14.94	9	16.9
	Pl *vs* Sl	17.26	14.26	8	15.5
D1S1656	Pl *vs* Cz	8.91	3.52	16	26.3
	Pl *vs* Ger	18.49	12.51	18	28.9
	Pl *vs* Sl	14.70	8.99	15	25.0
D10S1248	Pl *vs* Cz	9.03	4.06	8	15.5
	Pl *vs* Ger	10.48	5.68	9	16.9
	Pl *vs* Sl	9.66	6.25	7	14.1
D16S539	Pl *vs* Cz	10.84	8.50	7	14.1
	Pl *vs* Ger	17.88	15.69	6	12.6
	Pl *vs* Sl	18.93	17.07	6	12.6
FGA	Pl *vs* Cz	17.35	12.35	12	21.0
	Pl *vs* Ger	20.05	14.37	18	28.9
	Pl *vs* Sl	18.40	11.87	14	23.7
D18S51	Pl *vs* Cz	11.74	5.76	13	22.4
	Pl *vs* Ger	24.64	19.04	13	22.4
	Pl *vs* Sl	17.63	10.03	15	25.0
D19S433	Pl *vs* Cz	15.88	9.60	11	19.7
	Pl *vs* Ger	24.30	18.74	14	23.7
	Pl *vs* Sl	24.09	18.00	11	19.7
D3S1358	Pl *vs* Cz	5.33	4.35	7	14.1
	Pl *vs* Ger	10.53	6.28	8	15.5
	Pl *vs* Sl	13.23	8.78	7	14.1
D12S391	Pl *vs* Cz	17.39	11.13	16	26.3
	Pl *vs* Ger	11.14	9.03	17	27.6
	Pl *vs* Sl	16.86	9.92	17	27.6
D2S441	Pl *vs* Cz	14.64	12.61	8	15.5
	Pl *vs* Ger	22.49	14.30	10	18.3
	Pl *vs* Sl	12.51	7.29	10	18.3
SE33	Pl *vs* Cz	45.56	22.81	36	51.0
	Pl *vs* Ger	42.39	26.07	46	~61
	Pl *vs* Sl	46.73	27.07	38	53.4

Problems of Forensic Sciences 2019, vol. 117, 49–61

ANALIZA CZĘSTOŚCI WYSTĘPOWANIA W POPULACJI POLSKIEJ ALLELI 21 MARKERÓW GENETYCZNYCH WCHODZĄCYCH W SKŁAD ZESTAWU GLOBALFILER

Wstęp

Ze względu na różnice częstości występowania alleli układów STR w populacjach ludzkich wartość współczynników statystycznych otrzymanych w badaniach genetycznych śladów biologicznych zależy od zastosowanej do obliczeń bazy referencyjnej. Kluczowe dla poprawności obliczeń statystycznych w tym aspekcie jest zastosowanie bazy populacyjnej zgodnej lub możliwie najbliższej zgodności z pochodzeniem biogeograficznym badanej próbki. Ciągły rozwój metod stosowanych w genetyce sądowej oraz modyfikacie w dostępnych zestawach paneli markerów powodują konieczność aktualizacji dostępnych baz danych dla danego obszaru geograficznego. Laboratoria wykonujące genetyczne badania identyfikacyjne w oparciu o analizę markerów typu STR mają do dyspozycji zestawy odczynników pozwalające na oznaczenie kilkunastu, a nawet kilkudziesięciu markerów mikrosatelitarnych (w tym loci rekomendowane przez organizacje międzynarodowe) należących do grupy ESS (*European standard set loci*) zawierających TH01, vWA, FGA, D21S11, D3S1358, D8S1179, D18S51 oraz loci minisatelitarne, w tym D10S1248, D1S1656, D12S391, D2S441, D22S1045 (Parys-Proszek, Kupiec, Wolańska-Nowak, Branicki, 2010; Welch i in., 2012). Nową generację odczynników do amplifikacji DNA charakteryzuje wysoka czułość analizy, znaczna odporność na inhibicję oraz możliwość analizy nawet do 27 markerów STR jednocześnie. Zestaw GlobalFiler należy do grup najczęściej stosowanych w laboratoriach sądowych komercyjnych produktów dostępnych na rynku. Umożliwia on jednoczesne badanie 22 markerów autosomalnych oraz dwóch markerów zlokalizowanych na chromosomie Y. Analiza statystyczna wyników badań wykonywana jest za pomocą przeznaczonych do niej programów komputerowych, które służą do określenia prawdopodobieństwa powtórzenia się danego zgodnego profilu w populacji PRM (*probability of random match*) lub ilorazu wiarygodności LR (*likelihood ratio*). Warunkiem koniecznym do stosowania bazy częstości dla danej populacji jest spełnienie zadanych kryteriów i uzyskanie odpowiednich wartości współczynników populacyjnych oraz równowagi Hardy'ego-Weinberga dla poszczególnych loci (Butler, 2015; Chakraborty, 1992; Evett, Gill, 1991; Foreman, Evett, 2001; Schneider i in., 2009).

Cel

Celem pracy było opracowanie bazy częstości alleli zestawu GlobalFiler w populacji polskiej dla poszczególnych loci oraz określenie kluczowych parametrów niezbędnych do stosowania uzyskanej bazy w rutynowej pracy biegłego podczas przygotowywania opinii z zakresu genetyki sądowej.

Materiały i metody

Materiał do badań stanowiły 164 próbki DNA pochodzące od niespokrewnionych ze sobą osób z populacji polskiej. 85 próbek pochodziło od mężczyzn, a 79 – od kobiet. Próbki zawierające wyizolowany materiał genetyczny poddano pomiarowi ilości DNA za pomocą zestawów Quantifiler Human lub Quantifiler Trio (firmy Thermo Fisher) i aparatu 7500 Real Time PCR System. Następnie wykonano amplifikację DNA z zastosowaniem zestawu GlobalFiler (firmy Thermo Fisher), który pozwala na jednoczesną amplifikację dwudziestu czterech markerów typu STR (w tym markerów charakterystycznych dla płci męskiej – DYS391 oraz Yindel). Do reakcji użyto 1 ng matrycy DNA. Kolejno próbki poddano elektroforezie kapilarnej za pomocą sekwenatora ABI Prism 3500xL (firmy Life Technologies) przy czasie nastrzyku 24 sekundy. Próbki analizowano według parametrów analizy ustalonych w trakcie przeprowadzonej w Pracowni Genetyki Sądowej IES walidacji z zastosowaniem programu GeneMapper® ID-X (wersja 1.5, firmy Thermo Fisher). Analizę statystyczną przeprowadzono w programie GenoProof 3 (firmy Qualitype). Bazę częstości utworzono za pomocą narzędzia *population study*. Na potrzeby niniejszej analizy obliczono parametry statystyczne, takie jak siła dyskryminacji (PD), poziom heterozygotyczności (HET) oraz homozygotyczności (h), zachowanie równowagi Hardy'ego-Weinberga, łączne prawdopodobieństwo przypadkowej zgodności (GenoProof®3 Theory Manual, 2014). Wartość częstości dla rzadkich alleli określono na poziomie 0,001 rekomendowanym przez German Stain Comission (Schneider i in., 2009).

Wyniki

Wyniki obliczeń częstości występowania alleli w badanej populacji polskiej zamieszczono w tabeli 1. W tabeli 2 zestawiono parametry statystyczne dla 20 analizowanych markerów. Do analizy częstości alleli markera SE33 użyto 157 próbek, natomiast dla pozostałych markerów wszystkie 164 próbki poddano analizie. Wyniki przedstawione w tabeli 1 wskazują, że badane loci spełniają kryterium równowagi Hardy'ego-Weinberga.

Dyskusja

Kluczowym wskaźnikiem poprawności danych zgromadzonych w analizowanej bazie jest spełnianie kryterium równowagi Hardy'ego-Weinberga, co daje możliwość wykorzystania otrzymanych danych do obliczeń częstości genotypów w badaniach sądowych.

Zgodnie z przewidywaniami uzyskane wyniki częstości alleli (w zakresie wspólnych markerów) nie odbiegają znacząco od parametrów uzyskanych dla populacji „bliskich geograficznie" (Tabela 3; Hantschel, Hausmann, Lederer, Martus, Betz, 1999; Mornhinweg, Luckenbach, Ritter, 1998; Reichenpfader, Immel, Klintschar, 2003; Seider, Fimmers, Betz, Lederer, 2010). Przeprowadzona analiza statystyczna z wykorzystaniem testu χ^2 niezależności (Tabela 4) wykazała istotną statystycznie zależność częstości występowania alleli w populacji polskiej i bliskiej geograficznie dla markerów D21S11 (populacja czeska i słowacka), D22S1045 (populacja niemiecka i słowacka), D16S539 (populacja niemiecka i słowacka), D18S51 (populacja niemiecka), D19S433 (populacja niemiecka i słowacka) oraz D2S441 (populacja niemiecka). Natomiast po zastosowaniu poprawki Yatesa sugerowanej przez niektórych badaczy test wykazał istotną statystycznie zależność dla markera D16S539. Dla obu testów otrzymane wartości przemawiają za odrzuceniem hipotezy o braku zależności, są niewiele powyżej wartości tabelarycznej. Dlatego też otrzymane wyniki nie wykluczają stosowania częstości alleli populacji czeskiej, niemieckiej i słowackiej do obliczeń wartości dowodu z analizy genetycznej, ale wskazują tym bardziej na celowość stosowania dedykowanej bazy danych dla omawianej populacji.

Kwestią dyskusyjną jest stosowanie w obliczeniach do celów identyfikacji genetycznej markerów genetycznych vWA oraz D12S391, które położone są na chromosomie 12 w odległości 6Mb od siebie, co niesie ze sobą ryzyko braku niezależności dziedziczenia. Zgodnie z doniesieniami literaturowymi zaleca się, by przeprowadzanie obliczeń z wykorzystaniem obu markerów wykonywać ze szczególną ostrożnością (O'Connor, Tillmar, 2012) lub z wykluczeniem jednego z nich z analizy statystycznej. Marker SE33 charakteryzuje się bardzo wysokim stopniem heterozygotyczności, co implikuje wyższą różnorodność alleli, a tym samym zmniejsza szansę na przypadkowe powtórzenie genotypu. Taka zależność została zaobserwowana już wcześniej dla tego locus (Borsuk, Gettings, Steffen, Kiesler, Vallone, 2017; Davis i in., 2012) i jest jego cechą charakterystyczną w wielu populacjach.

Wnioski

Analizowana baza populacyjna spełnia kryteria genetycznej bazy częstości alleli w badanej populacji, co pozwala na stosowanie jej do kalkulacji biostatystycznych na potrzeby genetyki sądowej. Otrzymane wartości parametrów statystycznych oraz liczba przebadanych próbek jest wystarczająca do odniesienia otrzymanych wyników do całej populacji polskiej.