



archiwum medycyny sądowej i kryminologii

Praca oryginalna
Original paper

Dane populacyjne dla układów: D1S1656, D2S441, D2S1338, D3S1358, D8S1179, D10S1248, D12S391, D16S539, D18S51, D19S433, D21S11, D22S1045, FGA, TH01, vWA, zawartych w systemie NGM na podstawie tysiąca niespokrewnionych osób z regionu łódzkiego Polski Centralnej

Population database on: D1S1656, D2S441, D2S1338, D3S1358, D8S1179, D10S1248, D22S1045, D12S391, D16S539, D18S51, D19S433, D21S11, FGA, TH01, vWA loci included in NGM system based on one thousand unrelated individuals from Lodz region of Central Poland

Renata Jacewicz, Beata Markiewicz, Rafał Wojtkiewicz, Maciej Jędrzejczyk, Jarosław Berent

Pracownia Genetyki Medycznej i Sądowej Zakładu Medycyny Sądowej, Uniwersytet Medyczny w Łodzi, Polska
Medical and Forensic Genetics Laboratory, Department of Forensic Medicine, Medical University of Lodz, Poland

Streszczenie

Przedstawiono dane populacyjne uzyskane na podstawie analizy 1000 niespokrewnionych osób polskiego pochodzenia zamieszkujących region łódzki Polski Centralnej z zastosowaniem fluorescencyjnej reakcji multiplex PCR oraz rozdziału z użyciem elektroforezy kapilarnej. Ocena objęła 15 polimorficznych loci DNA – STR zawartych w zestawie NGM multiplex-PCR, tj. D1S1656, D2S441, D2S1338, D3S1358, D8S1179, D10S1248, D12S391, D16S539, D18S51, D19S433, D21S11, D22S1045, FGA, TH01, vWA. Obliczono rozkład częstości alleli oraz kluczowe parametry statystyczne dla badanych markerów i całego zestawu. Wykazano zgodność badanej populacji z równowagą Hardy’ego i Weinberga, niezależność dziedziczenia oraz wysokie parametry przydatności w aspekcie genetyki sądowej. Porównanie międzypopulacyjne przeprowadzone metodą „najbliższego sąsiada” oraz skalowania wielowymiarowego zobrazowało genetyczne dystanse dzielące badaną polską populację od innych populacji Polski, Europy i świata.

Słowa kluczowe: STR multiplex, NGM system, dane populacyjne, Polska Centralna, porównania międzypopulacyjne, skalowanie wielowymiarowe.

Abstract

A population data obtained on the basis of sample of 1000 unrelated individuals of Polish ancestry living in Lodz region of Central Poland with use of fluorescent multiplex-PCR and capillary electrophoresis were presented. Evaluation included 15 polymorphic loci DNA – STR from NGM multiplex-PCR set, ie. D1S1656, D2S441, D2S1338, D3S1358, D8S1179, D10S1248, D12S391, D16S539, D18S51, D19S433, D21S11, D22S1045, FGA, TH01, vWA. The allele frequency distribution and crucial statistical parameters for the investigated markers and the whole set were calculated.

The compliance of the studied population with Hardy-Weinberg equilibrium, independence of inheritance and high parameters of the usefulness in forensic genetics have been demonstrated. The interpopulation comparison performed by the „neighbor-joining” method as well as multidimensional scaling depicted the genetic distances dividing the examined Polish population from other populations of Poland, Europe and the world.

Key words: STR multiplex, NGM system, population data, Central Poland, interpopulation comparison, multidimensional scaling.

Wprowadzenie

Początki wykorzystania zmiennych regionów tandemowych powtórzeń w genomie człowieka (VNTR) do przeprowadzania identyfikacji osób i analiz pokrewieństwa datuje się na 1985 rok [1]. Wprowadzone w 1991 r. do analizy sekwencje krótkich tandemowych powtórzeń (STR) stanowią nadal podstawowy standard w analizach sądowych [2]. Metodyka badań tych markerów bazuje na fluorescencyjnym znakowaniu oraz jednoczesnej wielolocusowej amplifikacji (*multiplex PCR*). Detekcję produktów PCR we współczesnej genetyce sądowej przeprowadza się, począwszy od 1995 r., przy użyciu techniki elektroforezy kapilarnej (CE), a od 2015 r. również wykorzystując technikę sekwencjonowania wielkoskalowego (MPS) [3]. Uprzywilejowana niezmiennie pomimo upływu czasu pozycja markerów STR wiąże się z ich wykorzystaniem w tworzeniu stale powiększających się baz danych profili DNA, nieocenionego narzędzia w egzekwowaniu prawa [4]. Pierwszą istotną kwestią jest standaryzacja tworzonych baz danych markerów genetycznych, tak aby można było porównywać profile DNA uzyskane ze śladów biologicznych pochodzących z rozdzielonych czasowo i przestrzennie miejsc oraz analizowanych w różnych laboratoriach [5]. Drugą kwestią o dużym znaczeniu jest także dynamiczny rozwój baz tych markerów, biorąc pod uwagę wymogi dotyczące podnoszenia siły dyskryminacji, czułości i poziomu detekcji również w zdegradowanych próbach [6]. W Europie proces ten jest nadzorowany przez istniejącą od 1991 r. w obrębie Międzynarodowego Towarzystwa Genetyki Sądowej (ISFG) Europejską Grupę ds. Profilowania DNA (EDNAP), a także przez istniejącą formalnie od 1995 r. Europejską Sieć Laboratoriów Kryminalistycznych (ENFSI). Określają one standardy i wytyczne dotyczące badań genetyczno-sądowych, zakresy stosowanych markerów i metody analizy biostatystycznej. Stanowią również platformę wymiany doświadczeń w zakresie profilowania DNA w oparciu o międzylaboratoryjne testy jakości oraz międzynarodowe projekty badawcze [7]. W Stanach Zjednoczonych i Kanadzie

Introduction

Variable number tandem repeats (VNTR) in the human genome have been used for personal identification and kinship analysis since 1985 [1]. Short tandem repeat (STR) sequences, which were first used for analysis purposes in 1991, are still considered the basic standard in forensic studies [2]. The methodology of examinations performed for these markers is based on fluorescent labelling and simultaneous multilocus amplification (*multiplex PCR*). The detection of PCR products has been conducted in contemporary forensic genetics using the technique of capillary electrophoresis (CE) since 1995, and with the aid of massive parallel sequencing (MPS) since 2015 [3]. The continued privileged status of STR markers over time is attributable to their application in constantly expanding databases of DNA profiles which are an invaluable tool in law enforcement applications [4]. In this context, the primary aspect is the standardization of compiled databases of genetic markers, so that will be possible to compare DNA profiles obtained from biological traces collected in temporally and spatially separated sites, and analyzed at different laboratories [5]. Another significant factor is the rapid development of databases containing these markers, taking into account requirements related to increasing the power of discrimination, sensitivity and detection level also in degraded samples [6]. In Europe, the process is supervised by the European DNA Profiling (EDNAP) group which has existed as part of the International Society for Forensic Genetics (ISFG) since 1991, and by the European Network of Forensic Science Institutes (ENFSI) which was formally established in 1995. Both bodies define standards and guidelines for forensic genetic studies, ranges of markers used and biostatistical analysis methods. They also represent a platform for the exchange of experiences in DNA profiling based on inter-laboratory quality tests and international research projects [7]. In the United States and Canada, quality standards for forensic DNA studies are laid down by the Scientific Working Group on DNA Anal-

standardy jakości w zakresie sądowych badań DNA wytycza naukowa grupa badawcza SWGDAM (The Scientific Working Group on DNA Analysis Methods) na podstawie rekomendacji dyrektora Federalnego Biura Śledczego (FBI) [8]. Podstawą wszystkich baz danych jest 7 *loci* rekomendowanych przez Interpol, tzw. 7 *loci* ENFSI, tj. D3S1358, D8S1179, D18S51, D21S11, FGA, TH01, vWA [9]. Ten zestaw wzbogacony o kolejne 6 *loci*: D5S818, D7S820, D13S317, D16S539, CSF1PO, TPOX, wyznaczony od 1997 r. w USA 13-locusowy standard FBI, tzw. system CODIS (*Combined DNA Index System*) [10]. Kolejnym krokiem było poszerzenie baz danych o *loci* D2S1338 i D19S433, które razem z systemem CODIS posłużyły do opracowania systemu multipleksowego Identifiler, o powszechnym wykorzystaniu w profilowaniu DNA [11]. Następny krok w standaryzacji markerów STR w Europie stanowiło wprowadzenie zgodnie z zaleceniami ENFSI/EDNAP do bazy danych DNA dodatkowo pięciu markerów o skróconych amplikonach, które dają większą szansę na sukces analizy przy znacznym stopniu degradacji DNA: D1S1656, D2S441, D10S1248, D12S391, D22S1045 [9]. Siedem *loci* Interpolu (ENFSI *loci*) wraz pięcioma nowo wprowadzonymi *loci* ENFSI/EDNAP stały się od 2009 r. nowym rozszerzonym standardem ESS obowiązującym w Europie (*European Standard Set*) [12]. Dwunastolocusowy nowy standard ESS posłużył wraz z trzema *loci*: D2S1338, D16S539 i D19S433, do konstrukcji systemu NGM, a wraz z *locus* SE33 systemu NGM SElect, uznawanego obecnie za jeden z najbardziej dyskryminujących zestawów markerów DNA [13].

Niniejsza praca zawiera wyniki przeprowadzonej na dużą skalę analizy populacyjnej 15 polimorficznych *loci* zawartych w zestawie NGM w populacji Polski Centralnej. W pracy przedstawiono też porównawczą analizę międzypopulacyjną w zakresie badanych markerów w zestawieniu z innymi populacjami Polski, Europy oraz świata.

Material i metody

Próbki DNA pochodziły z populacji 1000 niespokrewnionych osób obojga płci (500 kobiet i 500 mężczyzn) zamieszkujących region łódzki Polski Centralnej, które uczestniczyły w sprawach dochodzenia spornego ojcostwa przeprowadzanych w Pracowni Genetyki Medycznej i Sądowej ZMS UM w Łodzi. Na wykorzystanie prób uzyskano świadomą zgodę badanych oraz zgodę Komisji Bioetyki przy Uniwersytecie Medycznym w Ło-

ysis Methods (SWGDAM) on the basis of recommendations issued by the Director of the Federal Bureau of Investigation (FBI) [8]. The foundation of all databases is constituted by seven *loci* recommended by Interpol: the so-called seven ENFSI-approved *loci*, i.e. D3S1358, D8S1179, D18S51, D21S11, FGA, TH01 and vWA [9]. The set, expanded by another six *loci* (D5S818, D7S820, D13S317, D16S539, CSF1PO and TPOX), was adopted in the USA in 1997 as the 13-locus FBI standard called CODIS (*Combined DNA Index System*) [10]. The next step was to extend the databases by adding two *loci* (D2S1338 and D19S433) which, along with CODIS, were used for developing Identifiler, a multiplex system which is commonly used for DNA profiling purposes [11]. A yet another step forward in the standardization of STR markers in Europe was the addition of five markers with reduced-size amplicons (D1S1656, D2S441, D10S1248, D12S391, D22S1045) to the DNA database in conformity with ENFSI/EDNAP recommendations. The new markers increase the chance for a successful analysis when handling considerably degraded DNA specimens. In 2009, seven Interpol *loci* (ENFSI *loci*) together with five newly introduced ENFSI/EDNAP *loci* were adopted as a new extended ESS standard which is used in Europe (*European Standard Set*) [12]. The new ESS standard, comprising 12 *loci*, was used along with another three *loci* (D2S1338, D16S539 and D19S433) for building the NGM system, and with SE33 *locus* – for compiling NGM SElect which is currently recognized as one of the most discriminating sets of DNA markers [13].

The present study reports results obtained in a large-scale population analysis of 15 polymorphic *loci* contained in the NGM kit. The analysis was performed for the population of Central Poland. Furthermore, the study includes a comparative inter-population analysis for the studied markers in relation to other populations of Poland, Europe and the world.

Material and methods

DNA samples were collected from 1,000 unrelated individuals of both sexes (500 women and 500 men), inhabiting the Lodz region of Central Poland, who have been involved in cases of disputed paternity investigated at the Medical and Forensic Genetics Laboratory, Department of Forensic Medicine, Medical University of Lodz. Written informed consent was obtained from all subjects for the use of the samples. The study was

dzi. Genomowy DNA izolowano z wymazów policzkowych z wykorzystaniem zestawu AX Sherlock zgodnie z protokołem producenta (A&A Biotechnology, Polska) [14]. Analizę jakościową i ilościową wyizolowanego DNA przeprowadzano za pomocą zestawu Power Quant (Promega, Madison, WI, USA) [15] w aparacie Real-Time PCR System 7500 z zastosowaniem oprogramowania HID Real-Time Analysis v.1.1 (Thermo Fisher Scientific, USA). Amplifikację matrycy DNA o stężeniu 0,1–0,3 ng/μl przeprowadzano w objętości 3 μl reakcji *multiplex* PCR przy użyciu zestawu AmpFℓSTR NGM [16]. Produkty reakcji PCR wykrywano za pomocą elektroforezy kapilarnej i detekcji fluorescencyjnej w sekwenatorze 3500 Genetic Analyzer (Thermo Fisher Scientific, USA). Genotypowanie przeprowadzono w obecności kontroli dodatniej oraz kontroli ujemnej z użyciem oprogramowania GeneMapper ID-X v.1.2 w odniesieniu do wzorca wielkości LIZ Size Standard 600 v.2 i drabiny alleli, dostarczonych przez producenta. Analiza statystyczna obejmująca zestawienie częstości alleli, ocenę zgodności badanej populacji z równowagą Hardy’ego i Weinberga (HWE), analizę nierównowagi sprzężeń (LD) w badanych parach *loci*, współczynnik wsobności (*coancestry coefficient* – ϕ), jak również parametry przydatności do badań genetyczno-sądowych, tj. heterozygotyczność (HET), wskaźnik informacji o polimorfizmie (PIC), zróżnicowanie genetyczne (GD), siłę dyskryminacji (PD), prawdopodobieństwo przypadkowej zgodności (MP), siłę wykluczenia (PE) i typowy indeks ojcostwa (TPI) została przeprowadzona przy użyciu arkusza kalkulacyjnego PowerStats v.1.2 [17] oraz oprogramowania GDA v.1.0 [18] i PowerMarker v.3.0 [19]. Ocenę HWE przeprowadzono z zastosowaniem poprawki Bonferroniego przy skorygowanym poziomie istotności ($0,05/15 = 0,0033$) dla wielokrotnego testowania. Dystanse genetyczne pomiędzy populacjami zwiualizowano na podstawie wartości *Fst* obliczonych wg Reynolds i wsp. [20].

Wyniki i dyskusja

Uzyskany rozkład alleli dla badanych układów oraz parametry statystyczne stanowiące ich charakterystykę oraz określające przydatność w analizach sądowych zestawiono w tabeli I. Jak z nich wynika, najwyższe parametry będące wskaźnikami zmienności badanych markerów oraz niskie wartości prawdopodobieństwa przypadkowej zgodności odnotowano dla *loci*: D1S1656, D2S1338, D12S391, D18S51, FGA oraz D21S11. Czyny to te markery najbardziej przydatnymi

approved by the Bioethics Committee at the Medical University of Lodz. Genomic DNA was isolated from buccal swabs with the AX Sherlock kit according to the manufacturer’s protocol (A&A Biotechnology, Poland) [14]. Qualitative and quantitative analyses of isolated DNA were conducted with the Power Quant kit (Promega, Madison, WI, USA) [15] using a Real-Time PCR System 7500, based on HID Real-Time Analysis v.1.1 software (Thermo Fisher Scientific, USA). The amplification of the DNA matrix at the concentration of 0.1–0.3 ng/μl was performed in a 3 μl volume of the multiplex PCR reaction using the AmpFℓSTR NGM kit [16]. Products of the PCR reaction were detected by capillary electrophoresis and fluorescent detection using a 3500 Genetic Analyzer (Thermo Fisher Scientific, USA). Genotyping was conducted in the presence of positive and negative controls with the GeneMapper ID-X v.1.2 software by reference to the LIZ Size Standard 600 v.2 and the allelic ladder supplied by the manufacturer. The statistical analysis, which comprised an estimation of allele frequency, assessment of consistency of the study population with the Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD) in the investigated pairs of loci, inbreeding coefficient – “coancestry coefficient” (ϕ), as well as parameters determining suitability for forensic genetic testing, i.e. heterozygosity (HET), polymorphism information content (PIC), genetic diversity (GD), power of discrimination (PD), random match probability (MP), power of exclusion (PE) and typical paternity index (TPI), was carried out using the PowerStats v.1.2 spreadsheet [17] as well as GDA v.1.0 [18] and PowerMarker v.3.0 [19] software. The HWE assessment was performed applying the Bonferroni correction at a significance level adjusted for multiple testing ($0.05/15 = 0.0033$). Genetic distances between populations were visualized based on *Fst* values calculated according to Reynolds *et al.* [20].

Results and discussion

The distribution of alleles obtained for the systems under study, as well as statistical parameters which characterize them and determine their suitability for forensic analysis, are listed in Table I. What they reveal is that the highest parameters indicating the variability of the studied markers, and the lowest values of random match probability, were noted for the following *loci*: D1S1656, D2S1338, D12S391, D18S51, FGA and D21S11. As a result, the markers should be regarded

Tabela I. Rozkład częstości alleli i parametry statystyczne 15 loci STR systemu NGM dla próby populacyjnej 1000 osób regionu łódzkiego (Polska Centralna)

Table I. The distribution of allele frequencies and statistical parameters of 15 STR loci in a population sample of 1000 individuals of Lodz region (Central Poland)

Allel Allele	D8S1179	D21S11	D18S51	D10S1248	vWA	D16D539	D2S1338	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391
5	0,0005	-	-	-	-	-	-	-	-	0,0015	-	-	-	-	-
6	-	-	-	-	-	-	-	-	-	0,2220	-	-	-	-	-
7	-	-	-	-	-	-	-	-	-	0,1315	-	-	-	-	-
8	0,0095	-	-	-	-	0,0105	-	-	-	0,0935	-	-	-	-	-
9	0,0105	-	-	-	-	0,0885	-	-	-	0,2105	-	-	-	-	-
9.3	-	-	-	-	-	-	-	-	-	0,3350	-	-	-	-	-
10	0,0635	-	0,0065	-	-	0,0505	-	0,0005	-	0,0060	-	0,2085	-	0,0005	-
11	0,0685	-	0,0120	0,0005	-	0,3000	-	0,1460	0,0010	-	-	0,3240	0,0010	0,0965	-
11.3	-	-	-	-	-	-	-	-	-	-	-	0,0470	-	-	-
12	0,1655	-	0,1060	0,0240	-	0,3145	-	0,0220	0,0835	-	-	0,0270	-	0,1405	-
12.2	-	-	-	-	-	-	-	-	0,0015	-	-	-	-	-	-
12.3	-	-	-	-	-	-	-	-	-	-	-	0,0010	-	-	-
13	0,3145	-	0,1120	0,2370	0,0040	0,2020	-	0,0030	0,2205	-	-	0,0195	0,0035	0,0575	-
13.2	-	-	-	-	-	-	-	-	0,0185	-	-	-	-	-	-
14	0,2280	-	0,1470	0,3345	0,1025	0,0325	-	0,0410	0,3550	-	-	0,3305	0,1505	0,0785	-
14.2	-	-	-	-	-	-	-	-	0,0230	-	-	-	-	-	-
14.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0015	-
15	0,1110	-	0,1625	0,2060	0,1060	0,0015	0,0010	0,3540	0,1710	-	-	0,0390	0,2505	0,1170	0,0310
15.2	-	-	-	-	-	-	-	-	0,0415	-	-	-	-	-	-
15.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0520	-
16	0,0255	-	0,1695	0,1550	0,1785	-	0,0565	0,3485	0,0420	-	-	0,0035	0,2225	0,1230	0,0205
16.2	-	-	-	-	-	-	-	-	0,0310	-	-	-	-	-	-
16.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0510	-
17	0,0025	-	0,1240	0,0395	0,2645	-	0,2075	0,0700	0,0025	-	0,0005	-	0,1960	0,0570	0,1060
17.1	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0005	-
17.2	-	-	-	-	-	-	-	-	0,0055	-	-	-	-	-	-
17.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0,1450	0,0155
18	0,0005	-	0,0800	0,0035	0,2420	-	0,0810	0,0100	-	-	0,0130	-	0,1625	0,0035	0,2025

Allel Allele	D8S1179	D21S11	D18S51	D10S1248	vWA	D16D539	D2S1338	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391
18.2	-	-	-	-	-	-	-	-	0,0035	-	-	-	-	-	-
18.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0640	0,0155
19	-	-	0,0400	-	0,0905	-	0,1220	0,0040	-	-	0,0845	-	0,0130	-	0,1115
19.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0105	0,0075
20	-	-	0,0225	-	0,0115	-	0,1350	0,0010	-	-	0,1410	-	0,0005	-	0,1165
20.2	-	-	-	-	-	-	-	-	-	-	0,0010	-	-	-	-
20.3	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0015	0,0010
21	-	-	0,0095	-	-	-	0,0295	-	-	-	0,1970	-	-	-	0,1050
21.2	-	-	-	-	-	-	-	-	-	-	0,0030	-	-	-	-
21.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0010
22	-	-	0,0060	-	0,0005	-	0,0280	-	-	-	0,2125	-	-	-	0,1355
22.2	-	-	-	-	-	-	-	-	-	-	0,0120	-	-	-	-
22.3	-	-	-	-	-	-	-	-	-	-	0,0005	-	-	-	-
23	-	-	0,0015	-	-	-	0,0980	-	-	-	0,1110	-	-	-	0,0795
23.2	-	-	-	-	-	-	-	-	-	-	0,0105	-	-	-	-
24	-	-	0,0010	-	-	-	0,1005	-	-	-	0,1210	-	-	-	0,0285
24.2	-	-	-	-	-	-	-	-	-	-	0,0020	-	-	-	-
24.3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0,0005
25	-	-	-	-	-	-	0,1135	-	-	-	0,0655	-	-	-	0,0165
26	-	0,0035	-	-	-	-	0,0250	-	-	-	0,0210	-	-	-	0,0040
27	-	0,0240	-	-	-	-	0,0025	-	-	-	0,0040	-	-	-	0,0015
28	-	0,1905	-	-	-	-	-	-	-	-	-	-	-	-	0,0005
28.2	-	0,0005	-	-	-	-	-	-	-	-	-	-	-	-	-
29	-	0,1950	-	-	-	-	-	-	-	-	-	-	-	-	-
29.2	-	0,0010	-	-	-	-	-	-	-	-	-	-	-	-	-
30	-	0,2280	-	-	-	-	-	-	-	-	-	-	-	-	-
30.2	-	0,0590	-	-	-	-	-	-	-	-	-	-	-	-	-
31	-	0,0710	-	-	-	-	-	-	-	-	-	-	-	-	-
31.2	-	0,0690	-	-	-	-	-	-	-	-	-	-	-	-	-
32	-	0,0085	-	-	-	-	-	-	-	-	-	-	-	-	-
32.2	-	0,1060	-	-	-	-	-	-	-	-	-	-	-	-	-

Allel Allele	D8S1179	D21S11	D18S51	D10S1248	vWA	D16D539	D2S1338	D22S1045	D19S433	TH01	FGA	D2S441	D3S1358	D1S1656	D12S391
33.2	-	0,0400	-	-	-	-	-	-	-	-	-	-	-	-	-
34	-	0,0005	-	-	-	-	-	-	-	-	-	-	-	-	-
34.2	-	0,0030	-	-	-	-	-	-	-	-	-	-	-	-	-
35	-	0,0005	-	-	-	-	-	-	-	-	-	-	-	-	-
Locus															
Liczba genotypów Number of genotypes	41	59	75	24	34	24	71	34	57	22	73	32	26	85	100
Liczba alleli Number of alleles	12	16	15	8	9	8	13	11	14	7	17	9	9	17	20
MAF	0,3145	0,2280	0,1695	0,3345	0,2645	0,3145	0,2075	0,3540	0,3550	0,3350	0,2125	0,3305	0,2505	0,1450	0,2025
HET	0,7930	0,8460	0,8770	0,7550	0,8080	0,7720	0,8800	0,7440	0,8000	0,7840	0,8410	0,7490	0,8080	0,8880	0,8740
GD	0,7998	0,8469	0,8753	0,7633	0,8095	0,7587	0,8792	0,7247	0,7838	0,7681	0,8569	0,7375	0,8001	0,8989	0,8830
MP	0,0660	0,0410	0,0300	0,0930	0,0630	0,0960	0,0270	0,1270	0,0750	0,0910	0,0370	0,1170	0,0720	0,0190	0,0250
PD	0,9340	0,9590	0,9700	0,9070	0,9370	0,9040	0,9730	0,8730	0,9250	0,9090	0,9630	0,8830	0,9280	0,9810	0,9750
PE	0,5860	0,6870	0,7480	0,5180	0,6140	0,5480	0,7550	0,4990	0,5990	0,5700	0,6770	0,5080	0,6140	0,7710	0,7420
TPI	2,41	3,25	4,06	2,04	2,60	2,19	4,17	1,95	2,50	2,31	3,14	1,99	2,60	4,46	3,95
PIC	0,7732	0,8291	0,8623	0,7254	0,7829	0,7201	0,8674	0,6799	0,7562	0,7317	0,8408	0,6934	0,7692	0,8901	0,8720
ϕ	0,0090	0,0016	-0,0015	0,0114	0,0024	-0,0170	-0,0004	-0,0261	-0,0202	-0,0202	0,0190	-0,0151	-0,0094	0,0126	0,0107
HWE	0,8835	0,2339	0,0883	0,4643	0,0452	0,4939	0,2630	0,5761	0,7030	0,0065*	0,6652	0,8657	0,7500	0,3943	0,6422

MAF – częstość allelu głównego; HET – heterozygotyczność; GD – zróżnicowanie genetyczne; MP – prawdopodobieństwo przypadkowej zgodności; PD – siła dyskryminacji; PE – siła wykluczenia; TPI – typowy indeks ojcostwa; PIC – wskaźnik informacji o polimorfizmie; ϕ – współczynnik wsobności; HWE – rozkład Hardy’ego-Weinberga – wartości prawdopodobieństwa w teście exact na podstawie 2300 permutacji

Łączne wartości dla 15 loci: MP = $1,6 \times 10^{-10}$, PD = 0,99999999999999984, PE = 0,999999789, średnia ϕ = -0,0012

MAF – major allele frequency; HET – heterozygosity; GD – gene diversity; MP – matching probability; PD – power of discrimination; PE – power of exclusion; TPI – typical paternity index; PIC – polymorphism information content; ϕ – coancestry coefficient; HWE – Hardy and Weinberg equilibrium – probability values of exact test based on 2300 shufflings

Combined values of 15 loci: MP: $1,6 \times 10^{-10}$, PD = 0,99999999999999984, PE = 0,999999789, Mean value of ϕ = -0,0012

dowodowo w aspekcie badań genetyczno-sądowych. Z przeprowadzonych obliczeń wynika, że najwyższe wartości parametrów przydatności badanych markerów korelują wyraźnie z najniższymi wartościami częstości populacyjnych najczęstszego allele w obrębie tego markera (MAF). Natomiast w mniejszym stopniu na parametry przydatności markerów ma wpływ liczba genotypów i alleli. Widać to wyraźnie na przykładzie loci D12S391 i D2S1338, które przy znacznych różnicach liczby odnotowanych alleli (20 vs 13) oraz genotypów (100 vs 71), charakteryzują się bardzo zbliżonymi wartościami heterozygotyczności, zmienności genowej, prawdopodobieństwa przypadkowej zgodności, siły dyskryminacji, siły wykluczenia, indeksu ojcostwa oraz wskaźnika polimorfizmu (tabela I). Najmniej polimorficznymi, a tym samym dającymi najmniejszą wartość dowodową okazały się loci: D22S1045, D2S441, D10S1248, D16S539 oraz TH01. Nieuwzględniona w tabeli analiza korelacji przeprowadzona w parach badanych loci nie wykazała braku równowagi sprzężeń zachodzącej dla którychkolwiek markerów badanego multiplexu. Jest to szczególnie istotne dla dwóch par loci syntenicznych, tj. leżących na tych samych chromosomach: VWA i D12S391 oraz D2S1338 i D2S441. Fakt ten uprawnia do mnożenia wartości dowodowych uzyskanych w analizie tych markerów dla oceny całkowitej wartości dowodowej ekspertyzy. Łączna wartość siły wyłączenia dla badanego zestawu loci wynosi 0,9999999, a określony multi-locusowy profil DNA pojawia się w populacji średnio 1 na $6,3 \times 10^{18}$ osób. Średnią wartość współczynnika wsobności ustalono na bardzo niskim poziomie 0,001, co wskazuje na spełnienie w badanej populacji zakładanych warunków niezależnego dziedziczenia, cechujące większość dużych populacji europejskich [21]. Uzyskany rozkład genotypów badanych loci systemu NGM wykazał zgodność z prawem równowagi Hardy'ego i Weinberga. Tylko w układzie TH01 uzyskano wartość prawdopodobieństwa poniżej poziomu istotności, które jednak przy uwzględnieniu poprawki Bonferroniego dla wielokrotnego testowania nie wykazało różnic statystycznie znamiennej w rozkładzie obserwowanym oraz oczekiwanym w tym locus, podobnie do czternastu pozostałych.

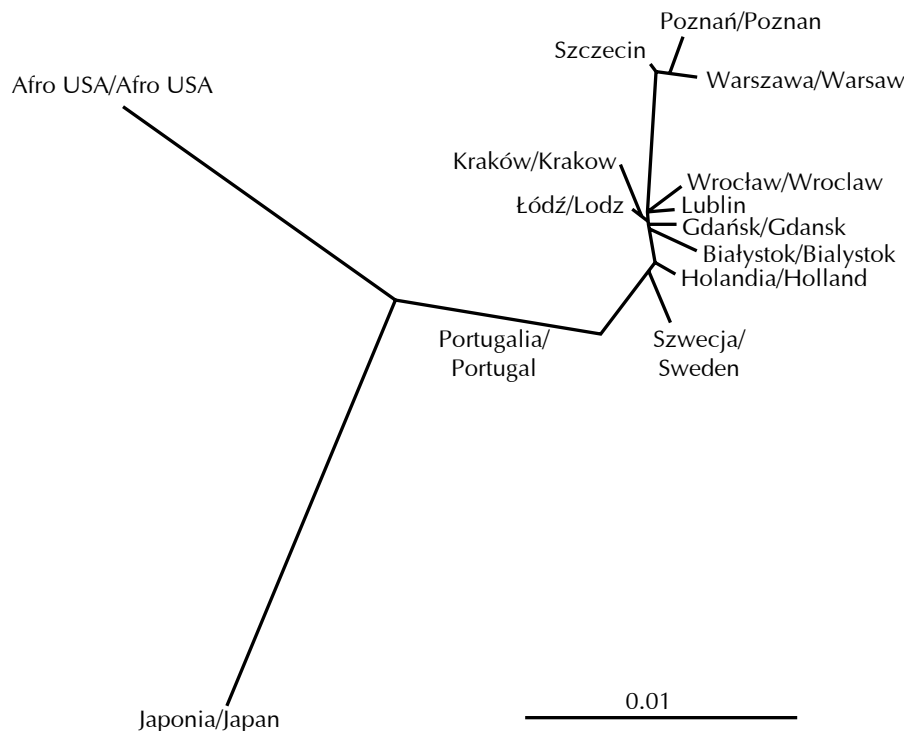
Na podstawie uzyskanych rozkładów cech genetycznych w populacji Polski regionu łódzkiego wyliczono wskaźniki F_{st} wg Reynolds i wsp. [20], a następnie zestawiono je z analogicznymi wskaźnikami uzyskanymi dla innych populacji Polski [24], Europy [23, 25, 26] oraz świata [22, 27]. Wykorzystano do tego

as the most suitable as evidence in forensic genetic investigations. The calculations show a clear correlation between the highest values of suitability parameters of the studied markers and the lowest values of population frequencies noted for the most common allele within the marker (MAF). In contrast, the number of genotypes and alleles has a lesser effect on marker suitability parameters. This can be seen clearly on the example of D12S391 and D2S1338 loci which, despite considerable differences in the number of observed alleles (20 vs. 13) and genotypes (100 vs. 71), share very similar values for heterozygosity, genetic variation, random match probability, power of discrimination, power of exclusion, paternity index and polymorphism index (Table I). The loci that were shown to be the least polymorphic and hence had the lowest value as evidence in forensic investigations included D22S1045, D2S441, D10S1248, D16S539 and TH01. An analysis of correlations conducted in pairs of investigated loci, which is not shown in the table, failed to demonstrate the absence of linkage disequilibrium for any markers of the multiplex under study. The above is pertinent in particular for two pairs of syntenic loci, i.e. lying on the same chromosomes: VWA and D12S391, and D2S1338 and D2S441. This fact justifies the multiplication of evidence values obtained in the analysis of these markers for the assessment of total evidence value of an expert opinion. The total value of the power of exclusion for the investigated set of loci is 0.9999999, which means that a specific multi-locus DNA profile occurs in the population on average in one person per 6.3×10^{18} . The mean value of the inbreeding coefficient was established to be at a very low level (0.001), which shows that the study population meets the conditions of independent inheritance which is found in the majority of large European populations [21]. The genotype distribution obtained for the studied loci included in the NGM kit demonstrated consistency with the Hardy-Weinberg equilibrium law. TH01 was the only system in which the probability value was below the significance level. However, after applying the Bonferroni correction for multiple testing, no statistically significant differences were revealed in the distributions observed and expected in this locus, similarly to the remaining fourteen.

Based on the distributions of genetic traits obtained for the population living in the Lodz region of Poland, F_{st} indices were calculated as proposed by Reynolds *et al.* [20], and compared with equivalent indices determined for other populations of Poland [24], Europe [23, 25, 26] and the world [22, 27]. The comparison

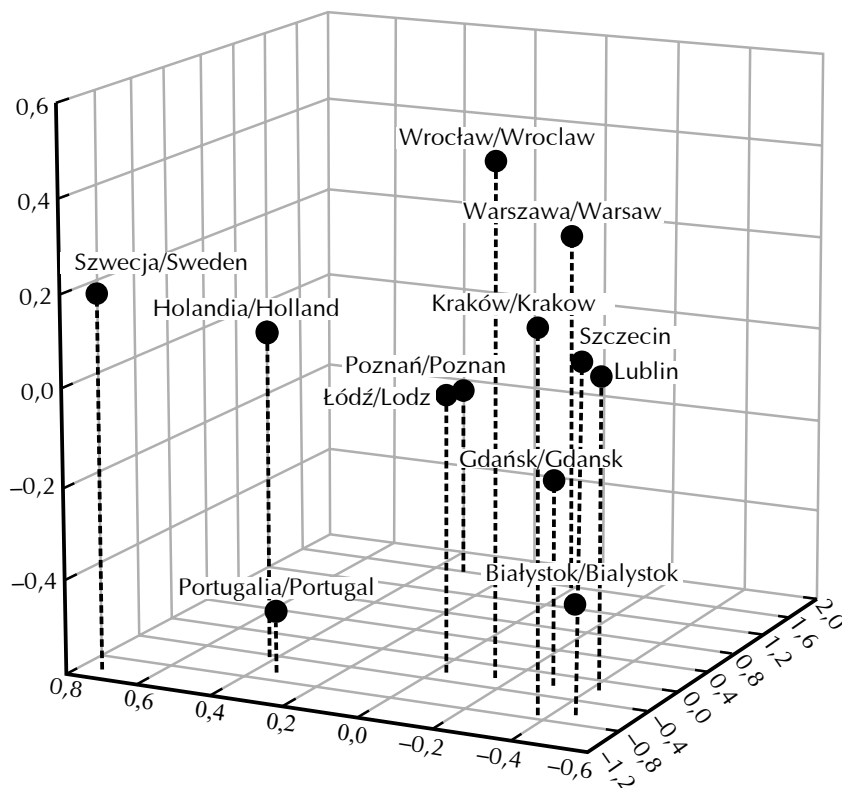
celu wizualizację przy użyciu metody „najbliższego sąsiada” (*neighbour-joining*), na podstawie której skonstruowano drzewo dystansów dla porównywanych populacji świata przedstawione na ryc. 1. Następnie w oparciu o skalowanie wielowymiarowe zobrazowano dystanse genetyczne pomiędzy porównywanymi populacjami Europy (ryc. 2) oraz pomiędzy porównywanymi populacjami Polski (ryc. 3). Z analizy danych wynika, że dystans ujawniony na podstawie matrycy wartości F_{st} w przybliżeniu odzwierciedla dystans terytorialny dzielący badaną populację Polski od innych populacji. Jest to szczególnie widoczne w przypadku najbardziej odległych terytorialnie i genetycznie od naszej populacji afroamerykańskiej i japońskiej. Co ciekawe, różnice statystycznie znamienne ($p < 0,05$) odnotowano nie tylko pomiędzy badaną populacją Łodzi a innymi populacjami świata i Europy, lecz także populacją Warszawy, Szczecina i Poznania. Może to wynikać z faktu, iż analiza częstości w obrębie markerów typu STR nie jest doskonałym narzędziem do analizy biogeograficznego pokrewieństwa. Mogą mieć na to wpływ również stosunkowo niskie liczebności w obrębie porównywanych z łódzką innych

was performed by means of visualization based on the “neighbour-joining” method which was used to create a genetic distance tree for the compared populations of the world, as shown in Fig. 1. Next, based on multidimensional scaling, genetic distances were visualized between the compared populations of Europe in Fig. 2, and between the compared populations of Poland in Fig. 3. The analysis of data shows that the distance revealed on the basis of the F_{st} value matrix approximately reflects the territorial distance between the studied Polish population and other populations. This is evident particularly in the Afro-American and Japanese populations which are separated by the greatest territorial and genetic distance from the Polish population. Interestingly, statistically significant differences ($p < 0.05$) were noted not only between the studied population of Lodz and other populations of Europe and the world, but also the populations of Warsaw, Szczecin and Poznan. The finding may be due to the fact that an analysis of frequencies within STR type markers is not a perfect tool for determining biogeographical ancestry. Another contributing cause may be the fact that other Polish populations which were compared with the Lodz



Ryc. 1. Dystans genetyczny pomiędzy populacją regionu łódzkiego (Centralna Polska) a innymi populacjami świata, zwizualizowany metodą „najbliższego sąsiada”

Fig. 1. Genetic distance between population of Lodz region (Central Poland) and other populations of the world, visualized by the “neighbor-joining” method



Ryc. 2. Skalowanie wielowymiarowe obrazujące dystans genetyczny pomiędzy populacją Łodzi (Polska Centralna) a innymi populacjami Europy na wykresie trójwymiarowym

Fig. 2. Multidimensional scaling revealing genetic distance between population of Lodz region (Central Poland) and other populations of Europe on the three-dimensional graph

populacji Polski skutkujące przeszacowaniem czy też niedoszacowaniem częstości niektórych alleli. Zgodnie z aktualnie obowiązującymi standardami minimalna liczebność, konieczna do opracowania danych populacyjnych z danego regionu, wynosi 500 [28]. Wykorzystanie jak największych baz danych zdecydowanie adekwatniej charakteryzuje poszczególne regiony kraju w porównaniu z danymi zbiorczymi, stanowiącymi sumę danych populacyjnych z tych regionów. Ponadto ma istotne znaczenie w praktyce badawczej z zakresu genetyki sądowej. Chroni przed zaniżaniem częstości alleli, a tym samym zawyżaniem wartości dowodowej ekspertyz.

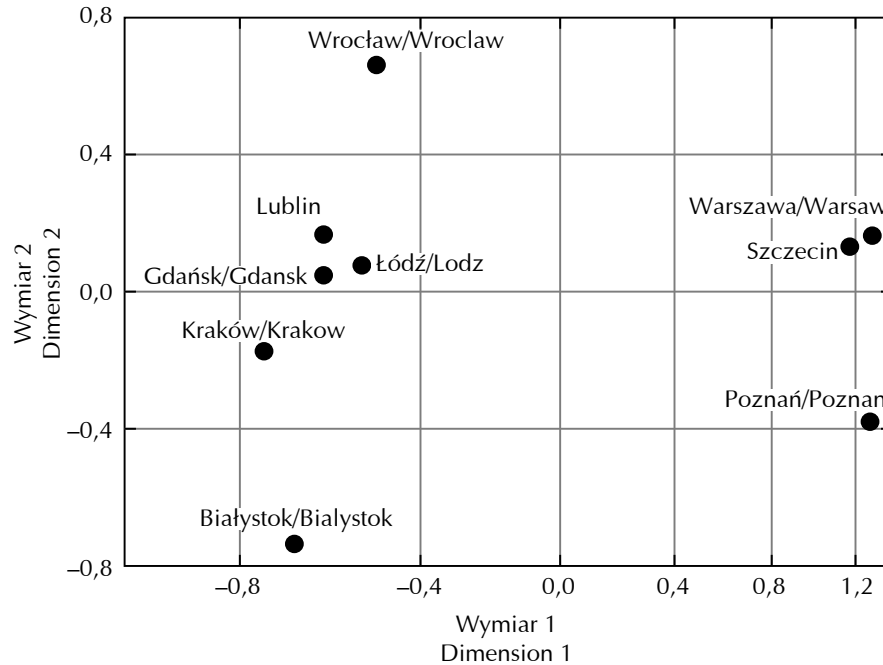
Wnioski

Udokumentowano wysoką przydatność zestawu NGM w badaniach z zakresu genetyki medycznej i sądowej w badanej populacji regionu łódzkiego Polski Centralnej i spełnienie przez nią warunków niezależnego dziedziczenia. Wykazano brak nierównowagi sprzężeń pomiędzy syntenicznymi *loci* zestawu, co

population were relatively small. Consequently, some alleles could have been either over- or underestimated. In accordance with the current standards the minimum number of people within a population that is required for processing population data for a given region is 500 [28]. Databases should be as large as possible to ensure that they characterize individual regions of the country much more adequately than collective data representing a total of population data for respective regions. Furthermore, the use of large databases is very important for forensic genetic research practice. It prevents underestimation of allele frequencies and thus overestimation of the evidence value of expert opinions.

Conclusions

The NGM kit was shown to be highly suitable for medical and forensic genetic studies in the studied population of the Lodz region of Central Poland. The population was found to satisfy the conditions of independent inheritance. Another finding was the absence of linkage disequilibrium between the synthetic *loci*



Ryc. 3. Skalowanie wielowymiarowe obrazujące dystans genetyczny pomiędzy populacją Łodzi (Polska Centralna) a innymi populacjami Polski na dwuwymiarowym wykresie

Fig. 3. Multidimensional scaling revealing genetic distance between population of Lodz region (Central Poland) and other populations of Poland on the two-dimensional graph

daje możliwość ich multipleksowania do oceny wartości dowodowej. Zobrazowano różnice międzypopulacyjne w obrębie badanej populacji z regionu Łodzi i innych regionów Polski. Uzasadniona jest zatem potrzeba konstruowania baz danych o jak największej liczbie, które najadekwatniej charakteryzują populację określonego regionu kraju oraz pozwalają na prawidłowe oszacowanie dowodu z badań DNA w tym regionie.

of the kit, which offers a possibility to multiplex them for the assessment of evidence value. Inter-population differences within the studied population of the Lodz region and other regions of Poland were visualized. Consequently, databases should contain the greatest possible number of records, as large databases provide the most adequate characterization of the population of a particular region of the country, and allow correct assessment of DNA-based evidence in that region.

Autorzy deklarują brak konfliktu interesów.

The authors declare no conflict of interest.

Piśmiennictwo

References

1. Jeffreys AJ, Wilson V, Thein SL. Individual-specific 'fingerprints' of human DNA. *Nature* 1985; 316: 76-79.
2. Edwards A, Civitello A, Hammond HA, Caskey CT. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am J Hum Genet* 1991; 49: 746-756.
3. Parson W, Ballard D, Budowle B, Butler JM, Gettings KB, Gill P, Gusmão L, Hares DR, Irwin JA, King JL, Knijff PD, Morling N, Prinz M, Schneider PM, Neste CV, Willuweit S, Phillips C. Massively parallel sequencing of forensic STRs: Considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements. *Forensic Sci Int Genet* 2016; 22: 54-63.
4. Martin PD, Schmitter H, Schneider PM. A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Sci Int* 2001; 119: 225-231.
5. Schneider PM. Scientific standards for studies in forensic genetics. *Forensic Sci Int* 2007; 165: 238-243.
6. Gill P, Fereday L, Morling N, Schneider PM. The evolution of DNA databases recommendations for new European STR loci. *Forensic Sci Int* 2006; 156: 242-244.

7. Branicki W, Pośpiech E, Kupiec T, Styrna J. Nowy wymiar ekspertyzy DNA – potrzeba szkoleń ekspertów i odbiorców ekspertyz. Arch Med Sąd Kryminol 2014; 64: 175-194.
8. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. Approved at the Scientific Working Group on DNA Analysis Methods meeting, Fredericksburg, Virginia, January 2010, <http://www.swgdam.org/>.
9. Welch LA, Gill P, Phillips C, Ansell R, Morling N, Parson W, Palo JU, Bastisch I. European Network of Forensic Science Institutes (ENFSI): Evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers. Forensic Sci Int Genet 2012; 6: 819-826.
10. Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations J Forensic Sci 2001; 46: 453-489.
11. Jacewicz R, Jędrzejczyk M, Ludwikowska M, Berent J. Population database on 15 autosomal STR loci in 1000 unrelated individuals from the Lodz region of Poland. Forensic Sci Int Genet 2008; 2: e1-e3.
12. Melo F, Amorim A, Alves C. Comparative performance between "next generation" multiplex systems and the new European Standard Set of STR markers in the Portuguese Population. Forensic Sci Int Genet 2014; 8: 137-142.
13. Green RL, Lagacé RE, Oldroyd NJ, Hennessy LK, Mulero JJ. Developmental validation of the AmpFℓSTR® NGM SElect™ PCR Amplification Kit: A next-generation STR multiplex with the SE33 locus. Forensic Sci Int Genet 2013; 7: 41-51.
14. A&A Biotechnology, Sherlock AX. Uniwersalny zestaw do izolacji DNA ze śladów biologicznych oraz trudnych prób. Protokół (wersja 1115). Gdynia, Polska.
15. PowerQuant™ System. Instructions for Use of Products PQ5002 and PQ5008. Technical Manual. 2015 Promega Corporation, Madison, USA.
16. AmpFℓSTR® NGM™ PCR Amplification Kit. User Guide (4425511 Rev. G). Applied Biosystems. 2015 Life Technologies. Carlsbad, USA.
17. Tereba A, Promega Corporation. Tools for analysis of population statistics. Profiles in DNA. Promega 1999; 2: 14-16.
18. Lewis, PO, Zaykin, D. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c) 2001.
19. Liu J, Muse SV. PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 2005; 21: 2128-2129.
20. Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. Genetics 1983; 105: 767-779
21. Weir BS. The second National Research Council report on forensic DNA evidence. Am J Hum Genet 1996; 59: 497-500.
22. Budowle B, Ge J, Charabarty R, Eisenberg AJ, Green R, Mulero J, Lagace R, Hennessy L. Population genetic analyses of the NGM STR loci. Int J Legal Med 2011; 125: 101-109.
23. Ribeiro T, Dario P, Vital N, Sanches S, Espinheira R, Geada H, Costa-Santos J. Population data of the AmpFℓSTR® NGM™ loci in South Portuguese population. Forensic Sci Int Genet 2013; 7: e37-39
24. Soltyszewski I, Pepinski W, Wolanska-Nowak P, Maciejewska A, Paszkowska R, Abreu-Glowacka M, Achrem W, Jonkisz A, Lebioda A, Konarzewska M, Ploski R. Polish population data on 15 autosomal STRs of AmpFℓSTR NGM PCR kit. Forensic Sci Int Genet 2014; 9: 142-149.
25. Gehrig C, Balitzki B, Kratzer A, Cossu C, Malik N, Castella V. Allelic proportions of 16 STR loci-including the new European Standard Set (ESS) loci-in a Swiss population sample. Int J Legal Med 2014; 128: 461-465.
26. Westen AA, Kraaijenbrink T, Robles de Medina EA, Hartevelde J, Willemse P, Zuniga SB, van der Gaag KJ, Weiler NE, Warnaar J, Kayser M, Sijen T, de Knijff P. Comparing six commercial autosomal STR kits in a large Dutch population sample. Forensic Sci Int Genet 2014; 10: 55-63.
27. Fujii K, Watahiki H, Mita Y, Iwashima Y, Kitayama T, Nakahara H, Mizuno N, Sekiguchi K. Allele frequencies for 21 autosomal short tandem repeat loci obtained using GlobalFiler in a sample of 1501 individuals from the Japanese population. Leg Med (Tokyo) 2015; 17: 306-308.
28. Carracedo A, Butler JM, Gusmão L, Linacre A, Parson W, Roewer L, Schneider PM. New guidelines for the publication of genetic population data. Forensic Sci Int Genet 2013; 7: 217-220.

Adres do korespondencji

Renata Jacewicz
Pracownia Genetyki Medycznej i Sądowej
Zakład Medycyny Sądowej
Uniwersytet Medyczny w Łodzi
ul. Sędziowska 18 A
91-304 Łódź, Polska
e-mail: renata.jacewicz@umed.lodz.pl

Address for correspondence

Renata Jacewicz
Medical and Forensic Genetics Laboratory
Department of Forensic Medicine
Medical University of Lodz
Sędziowska 18 A
91-304 Lodz, Poland
e-mail: renata.jacewicz@umed.lodz.pl

