



Praca oryginalna
Original paper

Rafał Wojtkiewicz, Beata Markiewicz, Maciej Jędrzejczyk, Renata Jacewicz

Polimorfizm 12 STR *loci* zawartych w zestawie Investigator HD-plex w populacji polskiej regionu łódzkiego

Polymorphism of 12 STR *loci* included in the Investigator HD-plex kit in Polish population of Lodz region

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

W pracy przedstawiono dane populacyjne dla Polski i poddano ocenie parametry przydatności 12 autosomalnych *loci* STR z zestawu Investigator HD-plex obejmującego 9 układów niedostępnych w żadnym z innych komercyjnych multipleksów, tj. D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D7S1517, D8S1132, D10S2325 i D21S2055. Do oceny wykorzystano DNA pochodzący od 303 niespokrewnionych osób zamieszkujących region łódzki Polski centralnej. Uzyskany rozkład genotypów jest zgodny z założeniami równowagi Hardy'ego i Weinberga (HWE). Prawdłowo obrazuje strukturę genetyczną badanej populacji w zestawieniu z innymi populacjami Europy i świata. Wskazuje na równowagę sprzężeń w obrębie badanych par *loci* oraz w odniesieniu do innych syntenicznych *loci*. Łączne wartości siły wykluczenia (PE) oraz prawdopodobieństwa przypadkowej zgodności (MP) wyniosły odpowiednio 0,99999988 oraz $5,2 \times 10^{-18}$. Polimorfizm zestawu badanych markerów genetycznych w obrębie multipleksu Investigator HD-plex pozwala zatem na istotne zwiększenie wartości dowodowej. Stanowi tym samym doskonale narzędzie do rozstrzygnięcia trudnych spraw z zakresu genetyki sądowej.

Słowa kluczowe: Investigator HD-plex, multipleks STR, centralna Polska, dane populacyjne, genetyka sądowa, allo-HSCT.

Abstract

In this study Polish population data as well as efficiency parameters of 12 STR *loci* included in the Investigator HDplex set were presented. This set contains 9 systems not available in any other commercial multiplexes, ie.: D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D7S1517, D8S1132, D10S2325 and D21S2055. The evaluation was performed based on DNA samples derived from 303 unrelated individuals living in Lodz region, central part of Poland. The obtained distribution of the genotypes is consistent with the assumptions of the Hardy and Weinberg equilibrium (HWE). It reflects properly genetic structure of the studied population compared with other populations of Europe and the world. It indicates the linkage equilibrium within the pairs of investigated *loci*, as well as with regard to other syntenic *loci*. The total value of the power of exclusion (PE) and the random match probability (MP) were respectively 0.99999988 and 5.2×10^{-18} . Therefore the polymorphism of examined genetic markers within the Investigator HD-plex multiplex allows for a significant increase of the evidence value. Thus it constitutes an excellent tool for resolving difficult cases in the field of forensic genetics.

Key words: Investigator HD-plex, multiplex STR, Central Poland, population data, forensic genetics, allo-HSCT.

Wprowadzenie

Krótkie tandemowe powtórzenia (STR) wykorzystywane są w genetyce sądowej do identyfikacji osobniczej i badania ojcostwa od ponad 20 lat [1]. Uprzywilejowana pomimo upływu czasu pozycja tych markerów wiąże się z ich wieloallelicznością, stosunkowo krótkimi amplikonami, standaryzacją metod badawczych oraz stale powiększającymi się bazami danych [2–4]. Analiza wyników badań profilowania DNA przy ustalaniu źródła pochodzenia śladu w mieszaninach, przy określaniu złożonych relacji rodzinnych oraz przy braku możliwości analizy porównawczej na podstawie materiału pochodzącego od osób blisko spokrewnionych wymaga złożonych obliczeń biostatystycznych. W takich przypadkach konieczne jest zastosowanie jak najszerszego zakresu markerów STR, pozwalających na uzyskanie jak najwyższej siły dyskryminacji [5]. Obecnie dostępnych jest wiele komercyjnych multipleksów, przy czym jednymi z najczęściej stosowanych w praktyce genetyczno-sądowej zestawów są: Identifiler, NGM SElect, GlobalFiler firmy Applied Biosystems (Foster City, CA), PowerPlex 16, PowerPlex 21, Power Plex-Fusion 6C firmy Promega Corporation (Madison, WI) [2, 3]. Swoim zakresem obejmują razem pulę 24 markerów STR autosomalnego DNA. Dodatkowe rozszerzenie zakresu badań o loci niewystępujące w żadnym innym zestawie daje system Investigator HD-plex firmy Qiagen N.V. (Venlo, NL). Multipleks ten, dawniej znany jako Mentype Chimera firmy Biotype Diagnostic GmbH (Dresden, Germany), został udostępniony powszechnie na rynku europejskim w 2010 r. Zestaw zawiera 9 ekskluzywnych markerów: D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D7S1517, D8S1132, D10S2325 i D21S2055, marker płci amelogeninę i 3 markery: D12S391, D18S51 i SE33, zawarte również w innych multipleksach [6]. Jak dotąd nie liczne są dane populacyjne obejmujące w komplecie 12 układów tego zestawu oraz jego łączne wskaźniki przydatności do analiz genetyczno-sądowych. Celem niniejszej pracy było przedstawienie tych danych dla populacji Polski centralnej w zestawieniu z danymi dostępnymi dla innych populacji.

Materiał i metody

Próby DNA pochodziły od 303 niespokrewnionych, zdrowych osób obojga płci zamieszkujących

Introduction

Short tandem repeats (STR) have been used in forensic genetics for personal identification and paternity testing for over 20 years [1]. The privileged status of these markers over time is associated with their multiallelic character, relatively short amplicons, standardization of investigative methods and constantly growing databases [2–4]. An analysis of DNA profiling results when establishing the origin of a trace in mixtures or determining complex family relationships, or in situations with no possibility of performing a comparative analysis based on material derived from closely related individuals, requires complicated biostatistical calculations. In such cases, it is necessary to employ the broadest possible set of STR markers which allow to obtain the highest power of discrimination [5]. At present, there are a lot of commercially available multiplex kits, of which the most commonly used in forensic genetic practice are: Identifiler, NGM SElect, GlobalFiler from Applied Biosystems (Foster City, CA), PowerPlex 16, PowerPlex 21, Power Plex-Fusion 6C from Promega Corporation (Madison, WI) [2, 3]. Overall, their range encompasses a pool of 24 STR markers of autosomal DNA. The scope of investigations can be additionally extended by using Investigator HDplex from Qiagen N.V. (Venlo, NL), a system including loci which are not present in any other kit. This multiplex, formerly known as Mentype Chimera from Biotype Diagnostic GmbH (Dresden, Germany), became widely available on the European market in 2010. The kit contains 9 inclusive markers (D2S1360, D3S1744, D4S2366, D5S2500, D6S474, D7S1517, D8S1132, D10S2325 and D21S2055), sex marker amelogenin and 3 markers (D12S391, D18S51 and SE33) which are also found in other multiplexes [6]. As of now, there are only sparse population data encompassing a complete set of 12 systems available in the kit and its overall determinants of suitability for forensic genetic analyses. The present study was aimed at determining these data for the population of central Poland in relation to data available for other populations.

Material and methods

The DNA samples were derived from 303 unrelated healthy individuals of both sexes, inhabiting the region of central Poland, who have been in-

region centralnej Polski, które uczestniczyły w sprawach dochodzenia spornego ojcostwa przeprowadzanych w pracowni Genetyki Medycznej i Sądowej Zakładu Medycyny Sądowej Uniwersytetu Medycznego w Łodzi. Na wykorzystanie prób uzyskano świadomą zgodę badanych i zgodę Komisji Bioetyki przy Uniwersytecie Medycznym w Łodzi. Genomowy DNA izolowano z wymazów policzkowych z wykorzystaniem zestawu AX Sherlock zgodnie z protokołem producenta (A&A Biotechnology, Polska). Analizę jakościową i ilościową wyizolowanego DNA przeprowadzano za pomocą zestawu Power Quant (Promega, Madison, WI, USA) w aparacie Real-Time PCR System 7500 przy użyciu 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 Investigator HD-plex (Qiagen, Venlo, NL) [6]. 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 ujemnej oraz kontroli dodatniej zgodnie z zaleceniami Komisji DNA ISFG (Międzynarodowego Towarzystwa Genetyki Sądowej) z użyciem oprogramowania GeneMapper ID-X v.1.2 w odniesieniu do wzorca wielkości DNA Size Standard 550 (BTO) dostarczonego przez producenta. Analiza statystyczna obejmująca zestawienie częstości alleli, ocenę zgodności badanej populacji z równowagą Hardy'ego i Weinberga (HWE), analizę korelacji (LD) w badanych parach *loci*, wskaźnik inbredu „coancestry coefficient” (f), jak również parametry przydatności do badań genetyczno-sądowych, tj. heterozygotyczność (Het), wskaźnik informacji o polimorfizmie (PIC), 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 [7] oraz oprogramowania GDA v. 1.0 [8]. Analizę nierównowagi sprzężeń (LD) przeprowadzono również dla syntenicznych *loci* względem *loci* zestawu HD-plex, zawartych w zestawach Identifiler oraz NGM Select, tj. TPOX, D2S1338, D2S441, D3S1358, FGA, D5S816, CSF1PO, D7S820, D8S1179, D10S1248, D21S11, vWA. Ocenę HWE i LD wykonano z zastosowaniem poprawki Bonferroniego przy skorygowanym poziomie istotności dla

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 use of the samples in research. The study was also approved by the Bioethics Committee at the Medical University of Lodz. Genomic DNA was isolated from buccal swabs by means of the AX Sherlock kit, following the manufacturer's protocol (A&A Biotechnology, Poland). Qualitative and quantitative analyses of isolated DNA were conducted with the Power Quant kit (Promega, Madison, WI, USA) 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 a concentration of 0.1–0.3 ng/μl was performed in a 3 μl volume of the multiplex PCR reaction using the Investigator HD-plex kit (Qiagen, Venlo, NL) [6]. Products of the PCR reaction were detected by capillary electrophoresis and fluorescent detection in a 3500 Genetic Analyzer (Thermo Fisher Scientific, USA). Genotyping was performed in the presence of negative and positive controls according to guidelines issued by the DNA Commission of the ISFG (International Society for Forensic Genetics) using GeneMapper ID-X v.1.2 software by reference to the DNA Size Standard 550 (BTO) supplied by the manufacturer. Statistical analysis – comprising an estimation of allele frequency, assessment of compliance of the study population with the Hardy-Weinberg equilibrium (HWE), correlation analysis (LD) in the investigated pairs of loci, inbreeding coefficient – “coancestry coefficient” (f), as well as parameters determining suitability for forensic genetic testing, i.e. heterozygosity (Het), polymorphism information content (PIC), 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 [7] and GDA v. 1.0 software [8]. An analysis of linkage disequilibrium (LD) was also performed for syntenic loci in relation to the loci from the HD-plex kit, contained in the Identifiler and NGM Select kits, i.e. TPOX, D2S1338, D2S441, D3S1358, FGA, D5S816, CSF1PO, D7S820, D8S1179, D10S1248, D21S11, vWA. The HWE and LD assessments were performed applying the Bonferroni correction at a significance level adjusted for multiple testing [9]. Genetic dis-

wielokrotnego testowania [9]. Dystanse genetyczne pomiędzy populacjami zwizualizowano na podstawie wartości F_{st} obliczonych wg Reynoldsa i wsp. [10] z zastosowaniem programu TreeView [11].

Wyniki i dyskusja

Uzyskany rozkład genotypów badanych *loci* systemu Investigator HD-plex wykazał zgodność z prawem równowagi Hardy'ego i Weinberga. Tylko w dwu układach uzyskano wartość prawdopodobieństwa poniżej poziomu istotności 0,05, tj. D6S474 ($p = 0,0174$) i SE33 ($p = 0,0230$). Przy uwzględnieniu poprawki Bonferroni przy skorygowanym poziomie istotności dla wielokrotnego testowania ($p > 0,05/12 = 0,0042$) rozkład obserwowany w tych 2 *loci*, podobnie jak w 10 pozostałych, nie odbiega w sposób statystycznie znamieny od rozkładu oczekiwanego. Uzyskane częstości alleli dla poszczególnych markerów zestawu Investigator HD-plex oraz wyliczone parametry biostatystyczne, określające przydatność w analizach sądowych, zestawiono w tabeli I.

Jak wynika z przeprowadzonych obliczeń, najwyższe wskaźniki zmienności w zakresie heterozygotyczności, siły dyskryminacji, siły wyłączenia, typowego indeksu ojcostwa oraz najniższe wartości przypadkowej zgodności odnotowano dla *loci*: SE33, D21S2055, D18S51, D12S391, D10S1132, D8S1132, D7S1517, co czyni te markery najbardziej przydatnymi dowodowo w aspekcie badań genetyczno-sądowych. Najmniej polimorficznymi, a tym samym dającymi najmniejszą wartość dowodową, okazały się *loci*: D6S474, D4S2366 oraz D2S1360. Nie wykazano sprzężenia pomiędzy żadną z par badanych markerów ($p > 0,05/2 = 0,025$), co jest szczególnie istotne w przypadku D6S474 i SE33 leżących na jednym chromosomie, tzw. *loci* syntenicznych. Całkowita wartość wskaźnika inbredu (f) dla badanej populacji wyniosła 0,003, co jest wartością znacznie niższą od wartości 0,01 rekomendowanej dla populacji spełniającej warunki niezależnego dziedziczenia [12]. Nie istnieją zatem przeszkody w uzyskiwaniu skumulowanej wartości dowodowej zestawu Investigator HD-plex reguły mnożenia wartości dowodowych uzyskanych dla poszczególnych markerów zestawu. Na podstawie dostępnych danych opublikowanych dla tej samej populacji nierównowagę sprzężeń testowano również w parach syntenicznych loci względem *loci*

tances between populations were visualized based on F_{st} values calculated according to Reynolds *et al.* [10], using the TreeView application [11].

Results and discussion

The genotype distribution obtained for the studied loci included in the Investigator HDplex system demonstrated conformity with the Hardy-Weinberg equilibrium law. The probability value was below the 0.05 significance level only in two systems, i.e. D6S474 ($p = 0.0174$) and SE33 ($p = 0.0230$). Applying the Bonferroni correction at a significance level adjusted for multiple testing ($p > 0.05/12 = 0.0042$), the distribution pattern noted in the 2 *loci* – similarly to the other 10 – did not diverge in a statistically significant manner from the expected distribution. The allele frequencies obtained for different markers from the Investigator HDplex kit, as well as calculated biostatistical parameters determining suitability for forensic analysis, are listed in Table 1.

As shown by the calculations, the highest variability indices for heterozygosity, power of discrimination, power of exclusion and typical paternity index, and the lowest values of random match probability were noted for the following *loci*: SE33, D21S2055, D18S51, D12S391, D10S1132, D8S1132 and D7S1517. Consequently, these markers are the most suitable for use as evidence in forensic genetic investigations. The *loci* that were shown to be the least polymorphic and, as a result, the least suitable as evidence in forensic investigations, included D6S474, D4S2366 and D2S1360. There was no linkage between any of the pairs of the investigated markers ($p > 0.05/2 = 0.025$), which is particularly significant for D6S474 and SE33 lying on the same chromosome, the so-called syntenic *loci*. The total value of the inbreeding coefficient (f) for the studied population was 0.003, which is a much lower value than 0.01 recommended for a population meeting the criteria of independent inheritance [12]. Therefore, there are no obstacles to obtaining a cumulative evidence value of the Investigator HDplex kit based on the rule of multiplying evidence values recorded for individual markers in the kit. Based on available data published for the same population, linkage disequilibrium (LD) was also tested in pairs of syntenic loci in relation to HDplex loci contained in the Identifiler and NGM kits, i.e.

Tabela I. Rozkład częstości występowania alleli i parametry biostatystyczne dla 12 *loci* STR w próbie populacyjnej 303 osób z regionu łódzkiego (centralna Polska)Table I. The distribution of allele frequencies and biostatistical parameters for 12 STR *loci* in a population sample of 303 individuals of Lodz region (Central Poland)

Allele Allels	D2S1360	D3S1744	D4S2366	D5S2500	D6S474	D7S1517	D8S1132	D10S2325	D12S391	D18S51	D21S2055	SE33
6								0,0020				
7								0,1450				
8			0,3560	0,0100				0,0530				
9								0,1070	0,0020			
10			0,2190	0,0860				0,1550	0,0070			
11			0,0660	0,3000				0,1490	0,0130			
12			0,1170	0,1780				0,1730	0,0990			0,0030
13		0,0120	0,1260	0,0480	0,2210			0,1290	0,1020			0,0050
13.2												0,0020
13.3			0,0020									
14	0,0940		0,1020	0,0530	0,2360			0,0530		0,1380		0,0330
14.1							0,0020					
14.2												0,0020
15		0,0610	0,0100	0,2180	0,1620			0,0310	0,0310	0,1530		0,0340
15.3												0,0020
16		0,1190	0,0020	0,0890	0,2840	0,0100	0,0030	0,0030	0,0200	0,1660		0,0490
16.1											0,0500	
17		0,3120		0,0150	0,0860	0,0020	0,0810		0,1020	0,1070		0,0660
17.1											0,0280	
17.3									0,0200			
18		0,1680		0,0030	0,0110	0,0360	0,2160		0,1620	0,0960		0,0660
18.1											0,0100	
18.3									0,0120			0,0020
19	0,0080	0,1580				0,1060	0,1350		0,1250	0,0500		0,0510
19.1											0,2310	
19.3									0,0130			
20	0,1270	0,0560				0,1020	0,1490		0,1370	0,0350		0,0510
20.1											0,0400	
20.2												0,0050
20.3									0,0030			0,0020
21	0,0790	0,0200				0,1320	0,1040		0,1100	0,0130	0,0030	0,0200
21.1											0,0030	
21.2												0,0260
22	0,3200					0,1250	0,1420		0,1440	0,0100		0,0120
22.2												0,0460
23	0,1520					0,0610	0,1300		0,0740	0,0050	0,0100	0,0020
23.2												0,0480
24	0,0380					0,1420	0,0330		0,0280	0,0020	0,0030	
24.2												0,0570
24.3									0,0020			

Tabela I. Ciąg dalszy
Table I. Continued

Allele Allels	D2S1360	D3S1744	D4S2366	D5S2500	D6S474	D7S1517	D8S1132	D10S2325	D12S391	D18S51	D21S2055	SE33
25	0,1070					0,1980	0,0050		0,0120		0,1020	
25.2												0,0380
26	0,0710					0,0610			0,0030		0,0940	
26.2												0,0540
27	0,0660					0,0200			0,0020		0,0150	
27.2												0,0560
28	0,0220					0,0050				0,0020	0,0070	
28.2												0,0840
29	0,0030										0,0330	
29.2												0,0530
30	0,0020										0,0450	
30.2												0,0560
31	0,0030										0,0430	
31.2												0,0360
32											0,0360	0,0020
32.2												0,0210
33	0,0020										0,0830	0,0050
33.2												0,0020
34											0,0870	0,0020
34.2												0,0050
35											0,0540	
36											0,0130	0,0020
37											0,0080	
39											0,0020	
	D2S1360	D3S1744	D4S2366	D5S2500	D6S474	D7S1517	D8S1132	D10S2325	D12S391	D18S51	D21S2055	SE33
P	0,1857	0,6113	0,3983	0,5822	0,0174	0,4761	0,4474	0,6557	0,5769	0,3447	0,9969	0,0230
Het	0,7990	0,8320	0,8090	0,8220	0,7820	0,8780	0,8980	0,8380	0,8840	0,8780	0,9240	0,9540
PIC	0,8100	0,8000	0,7500	0,7900	0,7500	0,8600	0,8400	0,8500	0,8800	0,8700	0,8900	0,9500
PD	0,9490	0,9450	0,9210	0,9350	0,9160	0,9700	0,9590	0,9660	0,9750	0,9730	0,9810	0,9920
MP	0,0510	0,0550	0,0790	0,0650	0,0840	0,0300	0,0410	0,0340	0,0250	0,0270	0,0190	0,0080
tj. 1 na	19,5	39,9	24,5	15,5	11,9	33,4	24,5	29,5	39,9	36,4	51,7	119,7
PE	0,5970	0,6590	0,6150	0,6400	0,5660	0,7510	0,7910	0,6720	0,7640	0,7510	0,8450	0,9060
TPI	2,4800	2,9700	2,6100	2,8100	2,3000	4,0900	4,8900	3,0900	4,3300	4,0900	6,5900	10,8200

P – prawdopodobieństwo testu exact (HWE); Het – heterozygotyczność obserwowana; PIC – wskaźnik informacji o polimorfizmie; PD – siła dyskryminacji; MP – prawdopodobieństwo przypadkowej zgodności, tj. 1 na...; PE – siła wykluczenia; TPI – typowy indeks ojcostwa

P – exact test probability (HWE); Het – observed heterozygosity; PIC – polymorphism information content; PD – power of discrimination; MP – matching probability-expressed as 1 in...; PE – power of exclusion; TPI – typical paternity index



HD-plex, zawartych w zestawach Identifiler oraz NGM, tj. TPOX, D2S1338, D2S441, D3S1358, FGA, D5S816, CSF1PO, D7S820, D8S1179, D10S1248, vWA, D21S11 [13, 14]. Najniższe wartości prawdopodobieństwa w teście *exact* odnotowano dla par *loci* leżących na drugim chromosomie, tj. D2S1360-TPOX ($p = 0,046$) oraz D2S1360-D2S441 ($p = 0,029$). Po zastosowaniu korekty Bonferroni nie wykazano jednak braku równowagi sprzężeń pomiędzy żadnym z syntenicznych *loci* zestawów HD-plex, Identifiler czy NGM. Uzyskane wyniki pozostają w zgodzie z wynikami badań Western i wsp. [15]. Wobec powyższego nie istnieją przeszkody w zwiększaniu siły dowodowej ekspertyzy z zastosowaniem reguły mnożenia w oparciu o uzupełniającą analizę markerów STR w systemach Identifiler, NGM oraz HD-plex.

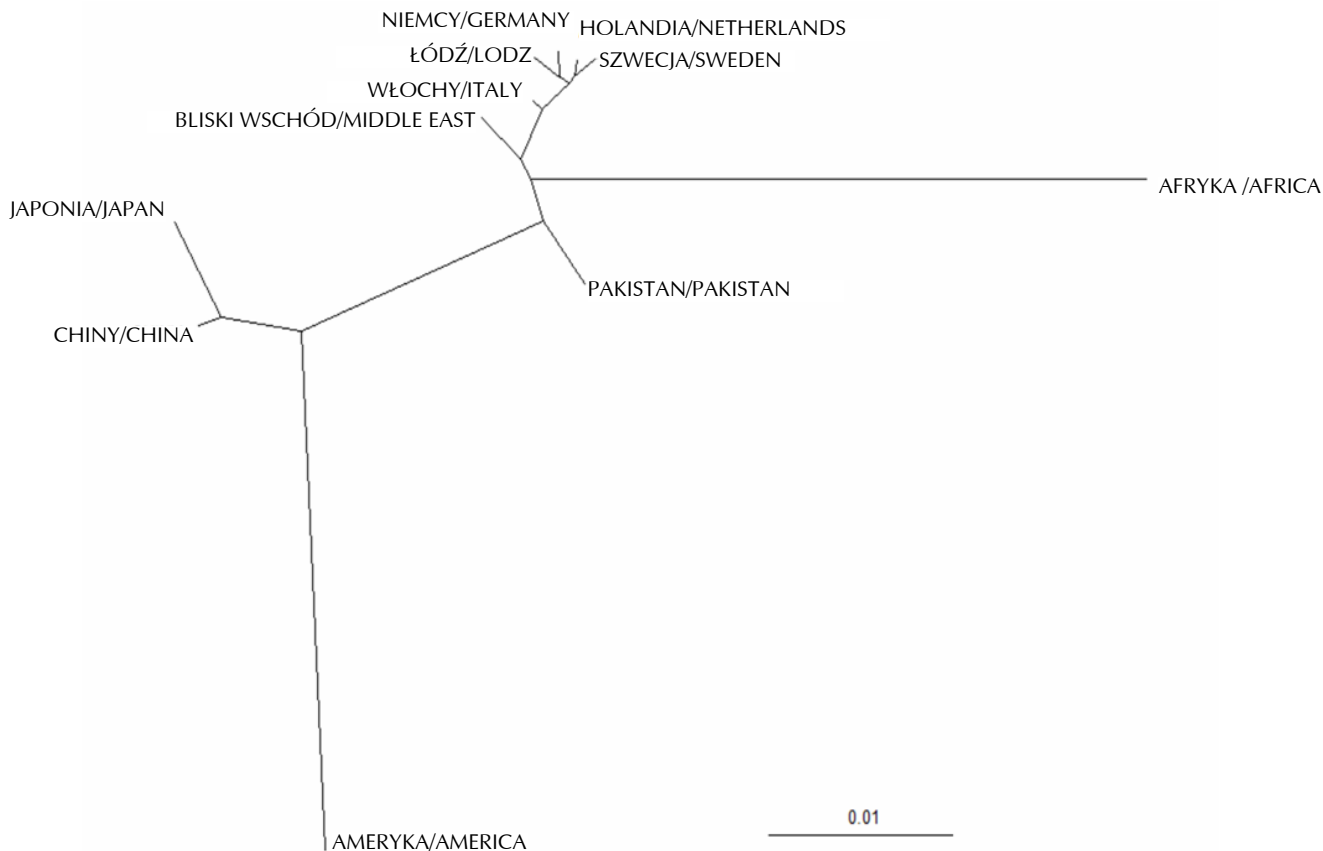
Łączna wartość siły dyskryminacji dla badanego zestawu *loci* wynosi 0,99999999999999995, co oznacza, że określony profil DNA pojawi się w badanej populacji średnio u 1 osoby na $1,93 \times 10^{17}$. Łączna siła wykluczenia dla zestawu badanych markerów (PE) wyniosła 0,999999879, a średni indeks ojcostwa dla pojedynczego markera 4,26. Uzyskane wysokie wartości parametrów biostatystycznych dla populacji Polski centralnej są zbliżone do analogicznych parametrów uzyskanych dla innych populacji [15–20]. Świadczy to o bardzo wysokiej przydatności badanego zestawu Investigator HD-plex do badań zarówno z zakresu identyfikacji osób, jak i analizy ojcostwa. Uzyskane rozkłady alleli w populacji Polski zestawiono z rozkładami alleli uzyskanymi dla innych populacji i z zastosowaniem metody najbliższego sąsiada (*neighbour-joining*) skonstruowano drzewo dystansów genetycznych. Wyznaczone długości gałęzi drzewa łączące badane populacje przedstawiono na rycinie 1. Z jej analizy wynika, że dystans ujawniony na podstawie matrycy wartości *Fst* odzwierciedla adekwatnie dystans terytorialny dzielący badaną populację Polski od innych populacji Europy i świata.

Badania walidacyjne i porównawcze wykazały znaczną odporność na inhibitory oraz właściwą analizę mieszanin z wykorzystaniem zestawu Investigator HD-plex, jak również kompatybilność genotypowania w zestawieniu z innymi multiplexami PCR, tj. Identifiler oraz NGM Select [15]. Co więcej, większość markerów zestawu Investigator

TPOX, D2S1338, D2S441, D3S1358, FGA, D5S816, CSF1PO, D7S820, D8S1179, D10S1248, vWA and D21S11 [13, 14]. The lowest probability values in the exact test were noted for the pairs of loci lying on the second chromosome, i.e. D2S1360-TPOX ($p = 0.046$) and D2S1360-D2S441 ($p = 0.029$). After applying the Bonferroni correction, however, no linkage disequilibrium (LD) was demonstrated between any of the syntenic *loci* included in the HD-plex, Identifiler or NGM kits. The results obtained in the study are thus consistent with the findings of research conducted by Western *et al.* [15]. In view of the above, there are no obstacles to increasing the evidence value of forensic expert opinions by using the multiplication rule based on a complementary analysis of STR markers in the Identifiler, NGM and HD-plex systems.

The total value of the power of discrimination for the investigated set of loci is 0.99999999999999995, which means that a specific DNA profile occurs in the study population on average in one person per 1.93×10^{17} . The total power of exclusion (PE) for the investigated marker kit was 0.999999879, and the mean paternity index for a single marker was 4.26. The high values of biostatistical parameters obtained for the population of central Poland are similar to corresponding parameters determined for other populations [15–20]. The finding shows that the studied kit, Investigator HD-plex, displays a high level of suitability both for personal identification and paternity testing. The distributions of alleles determined for Poland's population were compared with the distributions of alleles found in other populations and, based on the neighbour-joining method, a genetic distance tree was created. The lengths determined for the tree branches connecting the populations under study are shown in Fig. 1. The figure demonstrates that the distance revealed on the basis of the *Fst* value matrix accurately reflects the territorial distance between the studied Polish population and other populations in Europe and worldwide.

Validation and comparative studies demonstrated that the Investigator HD-plex kit is considerably resistant to inhibitors and allows accurate analysis of mixtures and compatibility of genotyping in comparison with other PCR multiplexes, i.e. Identifiler and NGM Select [15]. Furthermore, the majority of markers in the Investigator HDplex kit were validated for the



Rycina 1. Dystans genetyczny, który dzieli populację regionu łódzkiego (centralna Polska) od innych populacji Europy i świata, zwizualizowany metodą „najbliższego-sąsiada” na podstawie wartości F_{st} obliczonych wg Reynoldsa i wsp. [10]

Figure 1. Genetic distance, which divides population of Lodz region (Central Poland) and other populations of Europe and the world, visualized by the neighbour-joining method on the basis of F_{st} values calculated according to Reynolds *et al.* [10]

HD-plex została zwalidowana do monitorowania chimeryzmu po przeszczepieniu komórek macierzystych hematopoezy (allo-HSCT) na podstawie analizy informatywności przeprowadzonej u ponad 200 osób z par dawca-biorca [21]. Monitorowanie stopnia chimeryzmu DNA i dynamiki jego zmian u pacjentów w określonych odstępach czasowych po przeszczepie jest istotnym wskaźnikiem diagnostycznym i wpływa na leczenie. Ocena występowania chimeryzmu po allo-HSCT może też mieć istotne zaostwienie w odniesieniu do analiz z zakresu genetyki sądowej [22]. W przypadku przeszczepów od osób blisko spokrewnionych zdarza się duże podobieństwo w zakresie cech DNA. W tej sytuacji badany zestaw 12 markerów, z których 9 jest unikalnych, może nie tylko być stosowany jako uzupełnia-

monitoring of chimerism following allogeneic haematopoietic stem cell transplantation (allo-HSCT), based on an analysis of informativeness conducted in over 200 donor-recipient pairs [21]. Monitoring of the degree of DNA chimerism and the dynamics of its changes in patients at specified time intervals after transplantation is a significant diagnostic indicator and has an impact on treatment. An assessment of the occurrence of chimerism following allo-HSCT can also have a significant role in forensic genetic analyses [22]. Transplants from closely related individuals occasionally involve a high degree of similarity in DNA features. In such circumstances, the investigated kit of 12 markers, of which nine are unique, can be used not only as a complementary set in calculations of the evidence value in forensic genetic investigations, but

jący do obliczeń wartości dowodowej w badaniach genetyczno-sądowych, lecz także jako dodatkowo dyskryminujący (rozdzielający) w diagnostyce po allo-HSCT.

Wnioski

Wykazano wysoką przydatność zestawu Investigator HD-plex w badaniach z zakresu genetyki medycznej i sądowej. Jest to pierwsza baza opracowana dla kompletnego zestawu 12 markerów tego systemu w populacji polskiej. Uzyskany rozkład genotypów i alleli wskazuje na niezależność dziedziczenia markerów zarówno w obrębie zestawu HD-plex, jak i w odniesieniu do syntenicznych markerów z zestawów Identyfiler i NGM. Badany zakres markerów dobrze obrazuje strukturę genetyczną populacji centralnej Polski w zestawieniu z innymi populacjami Europy i świata. Z uwagi na wysoką siłę dyskryminacji badanych markerów, z których dziewięć nie jest dostępnych w żadnym innym komercyjnym multiplexie, stanowi on doskonałe wsparcie systemów Combined DNA Index System (CODIS) i European Standard Set (ESS) przy rozwiązywaniu najtrudniejszych spraw z zakresu identyfikacji osób i analizy pokrewieństwa, jak również oceny chimerizmu po allo-HSCT.

Autorzy deklarują brak konfliktu interesów.

Piśmiennictwo

References

1. Butler JM, NIST. Fundamentals of forensic DNA typing. Amsterdam: Academic Press, Elsevier, 2010.
2. <http://www.cstl.nist.gov/strbase/> (NIST STRbase). <http://www.cstl.nist.gov/strbase/multiplx.htm>.
3. http://spsmart.cesga.es/popstr.php?dataSet=strs_local (pop.STR). http://spsmart.cesga.es/help.php?dataSet=strs_local.
4. <http://allstr.de/allstr/home.seam?cid=5875416> (ALLSTR database).
5. 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.
6. Investigator HDplex Handbook. QIAGEN 1070461, 11/2012. <https://www.qiagen.com/de/resources/>.
7. Tereba A. Promega Corporation. Tools for analysis of population statistics. Profiles in DNA. Promega 1999; 2: 14-16.
8. Lewis PO, Zaykin D. 2001. Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c).
9. Bland JM, Altman DG. Multiple significance tests: the Bonferroni method. *BMJ* 1995; 310: 170.
10. Reynolds J, Weir BS, Cockerham CC. Estimation of the coancestry coefficient: basis for a short-term genetic distance. *Genetics* 1983; 105: 767-779.
11. Page RD. TreeView: An application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 1996; 12: 357-358.
12. Weir BS. The second National Research Council report on forensic DNA evidence. *Am J Hum Genet* 1996; 59: 497-500.
13. Jacewicz R, Jedrzejczyk 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: e1-3; doi: 10.1016/j.fsigen.2007.08.002.
14. Jedrzejczyk M, Jacewicz R, Szram S, Berent J. Genetic population studies on 15 NGM™ STR loci in central Poland population. *Forensic Sci Int Genet* 2012; 6: e119-20; doi: 10.1016/j.fsigen.2011.10.004.

also as an additional discriminatory tool in post-allo-HSCT diagnostics.

Conclusions

The study has shown the Investigator HD-plex kit to be highly suitable for medical and forensic genetic investigations. It is the first database compiled for the complete kit of 12 markers of the system in the Polish population. The distribution of genotypes and alleles obtained in the study points to an independent pattern of markers inheritance both within the HD-plex kit and in relation to syntenic markers from the Identifiler and NGM kits. The investigated range of markers accurately reflects the genetic structure of the population of central Poland against other populations in Europe and worldwide. Due to the high discrimination power of the investigated markers, nine of which are unavailable in any other commercial multiplexes, the kit represents an excellent support for the Combined DNA Index System (CODIS) and European Standard Set (ESS) in the resolution of the most difficult cases involving personal identification and kinship analysis as well as post-allo-HSCT chimerism assessment.

The authors declare no conflict of interest.

15. Westen AA, Haned H, Grol LJ, Harteveld J, van der Gaag KJ, de Knijff P, Sijen T. Combining results of forensic STR kits: HDplex validation including allelic association and linkage testing with NGM and Identifiler loci. *Int J Legal Med* 2012; 126: 781-789; doi: 10.1007/s00414-012-0724-4.
16. Zhang S, Zhang Z, Zhao S, Zhao Z, Li C. Genetic polymorphisms in 12 autosomal STRs in a Shanghai Han population from China. *Electrophoresis* 2013; 34: 613-617; doi: 10.1002/elps.201200423.
17. Tillmar AO, Nilsson H, Kling D, Montelius K. Analysis of Investigator HDplex markers in Swedish and Somali populations. *Forensic Sci Int Genet* 2013; 7(1): e21-22. doi: 10.1016/j.fsigen.2012.08.006
18. Turrina S, Ferriani M, Caratti S, De Leo D. Investigator HDplex markers: allele frequencies and mutational events in a North Italian population. *Int J Legal Med* 2015; 129: 731-733.
19. Population data for Investigator HDplex. QIAGEN 2010 <https://www.qiagen.com/de/resources/>.
20. Ozeki M, Tamaki K. Allele frequencies of 37 short tandem repeat loci in a Japanese population. *Leg Med (Tokyo)* 2013; 15: 342-346.
21. Thiede C, Bornhäuser M, Ehninger G. Evaluation of STR informativity for chimerism testing-comparative analysis of 27 STR systems in 203 matched related donor recipient pairs. *Leukemia* 2004; 18: 248-254.
22. Jacewicz R, Lewandowski K, Rupa-Matysek J, Jędrzejczyk M, Berent J. Niebezpieczeństwa wynikające z profilowania DNA materiałów biologicznych pochodzących od osób po alogenicznym przeszczepieniu krwiotwórczych komórek macierzystych (allo-HSCT) w odniesieniu do analiz z zakresu genetyki sądowej. *Arch Med Sąd Kryminol* 2015; 65: 225-247.

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

