

**Renata Jacewicz¹, Paweł Krajewski², Danuta Ulewicz³, Jarosław Piątek⁴,
Maciej Jędrzejczyk⁵, Katarzyna Bąbol-Pokora¹, Adam Prośniak¹, Magdalena Konarzewska²,
Andrzej Ossowski⁴, Mirosław Parafiniuk⁴, Jarosław Berent¹**

Y-STR Polska – baza danych do oceny wartości dowodowej w genetyce sądowej

Y-STR Poland – a database for evaluation of evidence value in forensic genetics

- ¹ Z Katedry i Zakładu Medycyny Sądowej Uniwersytetu Medycznego w Łodzi
Kierownik: prof. dr hab. n. med. J. Berent
- ² Z Katedry i Zakładu Medycyny Sądowej Warszawskiego Uniwersytetu Medycznego
Kierownik: dr hab. n. med. P. Krajewski
- ³ Z Laboratorium Kryminalistycznego Komendy Wojewódzkiej Policji w Łodzi
Naczelnik: mł. insp. R. Ziemecki
- ⁴ Z Zakładu Medycyny Sądowej PUM w Szczecinie
Kierownik: prof. dr hab. n. med. M. Parafiniuk
- ⁵ Z Zakładu Orzecznictwa Sądowo-Lekarskiego i Ubezpieczeniowego UM w Łodzi
p.o. Kierownika: prof. dr hab. n. med. S. Szram

Baza danych „Y-STR Polska” to projekt wielo-
ośrodkowy, którego celem jest stworzenie ogólno-
dostępnej bazy haplotypów chromosomu Y ozna-
czonych w populacji polskiej w zakresie szesnastu
loci systemu AmpFISTR®Y-filer. Baza ta będzie re-
gularnie aktualizowana i ma służyć ocenie wartości
dowodowej w genetyce sądowej. Wyjściowa baza
„Y-STR Polska” zawiera 1600 haplotypów Y-STR
i obejmuje dane zebrane w regionie łódzkim (2 nie-
zależne ośrodki), warszawskim oraz szczecińskim.
Praca zawiera jako załącznik plik z danymi typu
excel służący do obliczeń częstości występowania
w populacji polskiej określonego haplotypu Y. Plik
będzie regularnie aktualizowany wraz z poszerza-
niem bazy danych i ogólnie dostępny na stronie in-
ternetowej www.genetyka-sadowa.pl.

The „Y-STR Poland” is a multicenter project, the
aim of which is the construction of a widely available
database of Y chromosome haplotypes determined
in the Polish population in a range of sixteen loci in
AmpFISTR®Y-filer system. The database will be
regularly updated and it will be used in assessment of
evidence value in forensic genetics. The starting base
„Y-STR Poland” contains 1600 Y-STR haplotypes

and encompasses data collected in Lodz (two
independent centers), Warsaw and Szczecin regions.
The present report contains as an attachment the
data in an Excel-type file, which serves as a tool in
frequency determination of a given Y haplotype in
the Polish population. The file will be updated on
a regular basis along with updating the database,
and will be freely available from www.genetyka-sadowa.pl.

Słowa kluczowe:

Y-STR Polska, baza haplotypów,
genetyka populacyjna, obliczenia częstości

Key words:

Y-STR Poland, haplotype database,
population genetics, frequency estimation

WSTĘP

Nierekombinacyjne markery typu STR zlokalizo-
wane na chromosomie Y i badane w oparciu o sys-
tem multipleksowej detekcji są niezwykle przydatne
w genetyce sądowej [1]. Selektywna amplifikacja
tych markerów pozwala na identyfikację śladów bio-

logicznych pochodzących z przestępstw na tle seksualnym oraz umożliwia wyodrębnienie śladowych ilości DNA mężczyzny w materiale pochodzącym od kobiety [2]. Analiza polimorfizmu Y-STR ujawnia przynależność do określonej męskiej linii rodowej występującej w populacji z daną częstością [3]. Ocena wartości dowodowej analizy genetyczno-sądowej w zakresie układów Y-STR odbywa się z wykorzystaniem ogólnodostępnych baz danych, z których największą jest baza YHRD [4]. Baza ta zawiera aktualnie dane pochodzące z 393 różnych populacji świata (91 601 haplotypów), w tym 21 populacji Polski (4730 haplotypów). Dane te dotyczą jednak wyłącznie 7 loci STR-Y. Natomiast rutynowo stosowany zakres badawczy obejmujący 16 loci Y-STR w bazie YHRD obejmuje zaledwie 3 polskie populacje i zawiera 1266 haplotypów.

CEL PRACY

„Y-STR Polska” to projekt wielośrodkowy, którego celem jest stworzenie, a następnie regularne poszerzanie bazy haplotypów w zakresie 16 loci Y-STR w oparciu o akces kolejnych ośrodków badawczych. Wyjściowa baza „Y-STR Polska” zawiera 1600 haplotypów Y-STR i obejmuje dane zebrane w regionie łódzkim (2 niezależne ośrodki), warszawskim oraz szczecińskim, które nie zostały wcześniej umieszczone w bazie YHRD. W oparciu o wyjściową bazę przeprowadzono analizę homogenności w obrębie badanych populacji Polski, ocenę parametrów przydatności do badań oraz obliczenia częstości alleli, które posłużą do oceny wartości dowodowej ekspertyz w zakresie genetyki sądowej. Łatwe przeszukiwanie zgromadzonej bazy ma umożliwić specjalnie skonstruowany do tego celu plik z danymi typu excel stanowiący załącznik do niniejszej pracy, ogólnie dostępny na stronie internetowej www.genetyka-sadowa.pl.

MATERIAŁ I METODY

Materiał do badań stanowiła krew obwodowa bądź wymazy nabłonka pobrane ogółem od 1600 niespokrewnionych mężczyzn pochodzących z trzech województw: łódzkiego, mazowieckiego oraz zachodnio-pomorskiego. Wykorzystano dane populacyjne zebrane w czterech niezależnych ośrodkach badawczych: Pracowni Genetyki Sądowej Katedry

i Zakładu Medycyny Sądowej Uniwersytetu Medycznego w Łodzi (Łódź, n=400), Laboratorium Kryminalistycznym Komendy Wojewódzkiej Policji w Łodzi (Łódź-2, n=200), Katedrze i Zakładzie Medycyny Sądowej Warszawskiego Uniwersytetu Medycznego (Warszawa, n=760), oraz Pracowni Hemogenetyki Sądowej Zakładu Medycyny Sądowej Pomorskiego Uniwersytetu Medycznego w Szczecinie (Szczecin, n=240). DNA izolowano metodą kolumnkową bądź organiczną. Matrycowe DNA amplifikowano z wykorzystaniem zestawu PCR AmpFISTR®Y-filer zgodnie z zaleceniami producenta (Applied Biosystems).

Detekcję produktów amplifikacji przeprowadzano w genetycznych analizatorach wobec standardów wielkości GeneScan™ LIZ (Applied Biosystems). Statystyczna analiza wyników obejmująca obliczenia częstości występowania poszczególnych alleli w populacji, analizę zmienności molekularnej (AMOVA), ocenę dystansu genetycznego pomiędzy populacjami w oparciu o obliczenia współczynnika R_{st} została wykonana z zastosowaniem oprogramowania Arlequin ver 3.01 [5]. Zróżnicowanie w obrębie układów (GD) i haplotypów (HD) obliczono z zastosowaniem formuły Nei [6].

WYNIKI I DISKUSJA

Mapę Polski uwzględniającą regiony, które wzięły dotąd udział w projekcie „Y-STR Polska” obejmującym łączną liczbę 1600 haplotypów w zakresie 16 loci STR chromosomu Y przedstawiono na rycinie 1. Obserwowany w populacji Polski zakres alleli w obrębie 16 badanych układów systemu PCR AmpFISTR®Y-filer wraz z odpowiednim oznaczeniem GenBank [7], strukturą motywu powtórzeniowego [8] oraz nomenklaturą rekomendowaną przez Komisję DNA Międzynarodowego Towarzystwa Genetyki Sądowej [9, 10] zawarto w tabeli I.

Wysoka zmienność haplotypów wewnątrz czterech polskich populacji (99,89%) w porównaniu z niską zmiennością pomiędzy nimi (0,11%) świadczy o homogenności tych populacji w zakresie loci systemu AmpFISTR®Y-filer (tabela II). Porównanie międzypopulacyjne wykazało na bardzo zbliżony rozkład zmienności genowej w czterech analizowanych populacjach w obrębie poszczególnych układów Y-STR, co zostało zobrazowane graficznie na rycinie 2. Średnia wartość GD mieści się w prze-

dziale od 32,13 do 83,85%. Najwyższą zmiennością w badanych populacjach Polski, podobnie jak w innych populacjach świata, charakteryzuje się bialleliczny locus DYS385 [11, 12, 13]. Następnymi w kolejności po DYS385 pod względem siły dyskryminacji są układy DYS19 oraz DYS458. Natomiast najstabilniej dyskryminującymi w Polsce układami Y-STR są kolejno: DY393, DY392 oraz DY389I. Ujawniona zmienność haplotypowa jest zbliżona w obrębie 4 badanych populacji i waha się w przedziale 99,94-99,97%.



Ryc. 1. Mapa Polski uwzględniająca regiony, które wzięły dotąd udział w projekcie „Y-STR Polska” zawierającym łączną liczbę 1600 haplotypów w zakresie 16 loci systemu AmpFISTR®Y-filer.

Fig. 1. A map of Poland showing the regions which have so far participated in the „Y-STR Poland” project containing the total number of 1600 haplotypes in a range of 16 loci of AmpFISTR®Y-filer system.

Wartości R_{st} zawarte w tabeli III potwierdzają brak statystycznie istotnych różnic ($P > 0.05$) między populacjami z poszczególnych regionów Polski. Dlatego też zasadnym jest połączenie baz danych haplotypów zebranych w obrębie analizowanych populacji z województw: łódzkiego (2 niezależne ośrodki), mazowieckiego oraz zachodnio-pomorskiego (rycina 1) w zbiorczą bazę danych „Y-STR Polska”. Baza ta będzie służyć do szacowania częstości występowania określonego haplotypu

Tabela I. Zakres wielkości alleli w populacji Polski oraz charakterystyka markerów systemu AmpFISTR®Y-filer zgodnie z wytycznymi Komisji DNA Międzynarodowego Towarzystwa Genetyki Sądowej [9, 10].

Table I. The allele range in Polish population and the characteristics of AmpFISTR® Y-filer system markers according to the recommendations of the DNA Commission of the International Society of Forensic Genetics [9, 10].

Locus	Oznaczenie GenBank Accession	Motyw powtórzeniowy Repeat motif	Zakres alleli Allele range
DYS19	AC017019	TAGA	13-19
DYS389I	AC004617	(TCTG)(TCTA)(TCTG)(TCTA)	10-15
DYS389II	AC004617	TCTA	26-33
DYS390	AC011289	(TCTA)(TCTG)	21-27
DYS391	AC011302	TCTA	9-13
DYS392	AC011745	TAT	7-15
DYS393	AC006152	AGAT	10-16
DYS437	AC002992	TCTA	10-17
DYS438	AC002531	TTTTTC	7-14
DYS439	AC002992	AGAT	8-14
DYS448	AC025227	AGAGAT	17-23
DYS456	AC010106	AGAT	11-19
DYS458	AC010902	GAAA	12-20.2
DYS635	AC004772	TSTA compound	19-27
GATAH4.1	AC011751	TAGA	17-23
DYS385a/b	AC022486	GAAA	8-20

Tabela II. Rezultaty analizy molekularnej wariacji (AMOVA).

Table II. The results of the analysis of the molecular variance (AMOVA).

Porównywane populacje Compared populations	Źródło wariacji Variation source	Wariancja Variation
Łódź vs. Łódź - 2 vs. Warszawa vs. Szczecin	Pomiędzy populacjami Among populations	0,11%
	W obrębie populacji Within population	99,89%

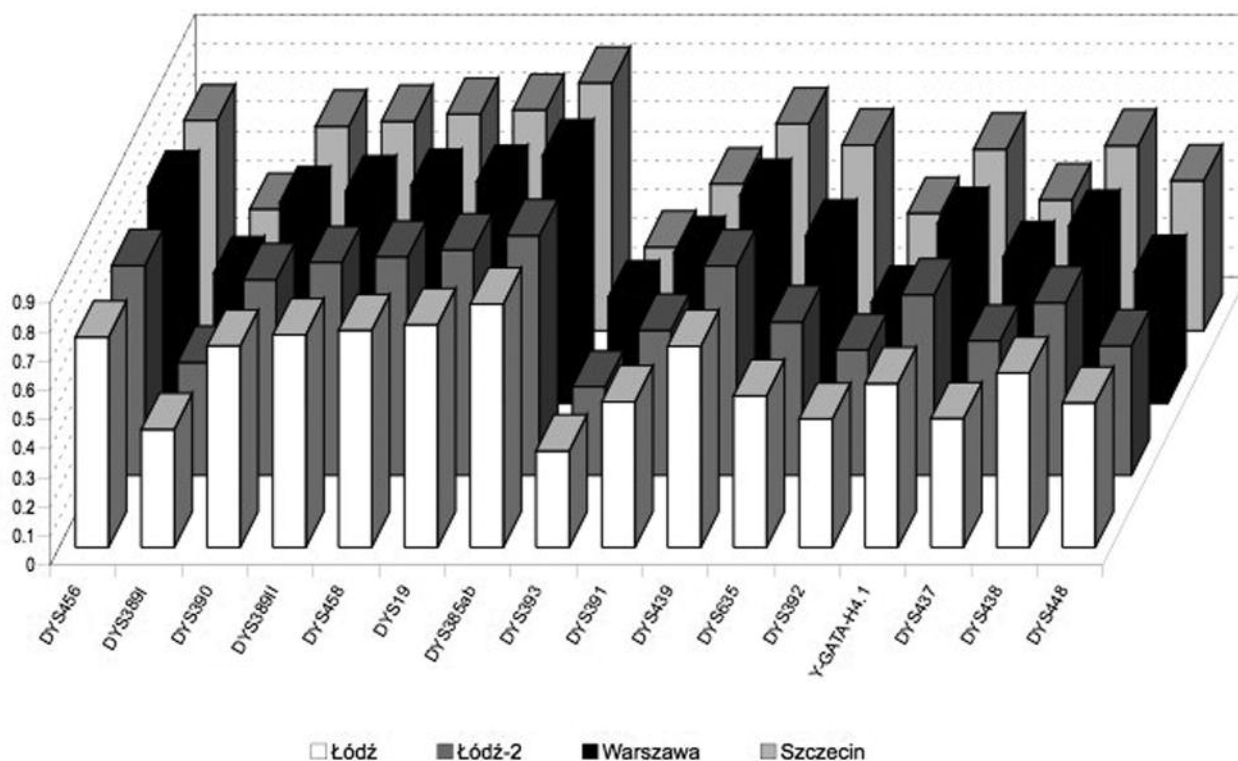
Y-STR, a tym samym oceny wartości dowodowej analiz z zakresu genetyki sądowej.

Wśród analizowanych 1600 haplotypów bazy „Y-STR Polska” zaobserwowano 1355 różnych haplotypów, co dało siłę dyskryminacji 84,69%. Średnia zmienność badanych markerów wynosi 60,01%. Zmienność haplotypowa w analizowanym zakresie 16 układów Y-STR ma wartość 99,95%, co znaczy, iż prawdopodobieństwo, że dwóch niespokrewnionych mężczyzn będzie miało ten sam 16 *locusowy* haplotyp wynosi 0,05%. Uzyskane wartości wskaźników dyskryminacji są zbliżone do analogicznych wartości w innych populacjach Europy i świata [14, 15, 16]. Trzy najczęstsze haplotypy pojawiły się w bazie „Y-STR Polska” 19, 11 oraz 10 razy z częstością odpowiednio 1,19%, 0,69% oraz 0,62%. Tabela IV przedstawia charakterystykę uzyskanych rozkładów alleli i zestawienie ich częstości oraz wartości zmienności w obrębie poszczególnych układów.

Tabela III. Dystans genetyczny oparty na wartościach R_{st} dla każdej pary w obrębie 4 porównywanych populacji. Nie obserwowano różnic statystycznie znamiennej ($P > 0.05$).

Table III. Genetic distance based upon pairwise R_{st} values for each pair in the 4 compared populations. Statistically significant differences ($P > 0.05$) were not observed.

	Łódź	Łódź-2	Warszawa	Szczecin
Łódź	0.000			
Łódź-2	0.003	0.000		
Warszawa	0.003	-0.002	0.000	
Szczecin	0.003	0.001	0.001	0.000



Ryc. 2. Porównanie wartości zmienności genowej (GD) dla poszczególnych markerów systemu PCR AmpFISTR®Y-filer w regionach, które dotąd wzięły udział w projekcie „Y-STR Polska”.

Fig. 2. The gene diversity (GD) values comparison within the markers of PCR AmpFISTR®Y-filer system in the regions which have so far participated in the „Y-STR Poland” project.

Tabela IV. Częstości alleli w zakresie 16 loci w bazie „Y-STR Polska” (n=1600).

Table IV. The allele frequencies of the 16 loci in the “Y-STR Poland” database (n=1600).

Allel/DYS	19	389I	389II	390	391	392	393	437	438	439	448	456	458	635	Y-GATA-H4.1	Allelic class 385a/b			
1																8,14	0.001	14,14	0.046
2																9,13	0.001	14,15	0.069
3																9,14	0.003	14,16	0.015
4																9,15	0.001	14,17	0.002
5																10,10	0.004	14,18	0.001
6																10,11	0.001	14,19	0.001
7						0.001			0.001	0.001						10,13	0.007	15,15	0.011
8						0.001			0.031	0.004						10,14	0.154	15,16	0.009
9					0.007	0.001			0.276	0.364						10,15	0.012	15,17	0.003
10		0.001			0.582	0.006	0.001	0.001	0.544	0.340						10,16	0.001	15,18	0.002
11		0.001			0.391	0.768	0.002	0.001	0.544	0.340		0.002				11,11	0.012	15,19	0.002
12		0.119			0.019	0.041	0.070	0.001	0.134	0.189		0.003	0.001			11,12	0.001	16,13	0.001
13	0.040	0.736			0.001	0.122	0.803	0.002	0.013	0.094		0.026	0.005			11,13	0.054	16,17	0.006
14	0.196	0.141				0.053	0.113	0.687	0.001	0.009		0.114	0.044			11,13.2	0.001	16,18	0.012
15	0.242	0.004				0.007	0.009	0.212				0.299	0.265			11,14	0.339	16,19	0.004
16	0.303						0.001	0.089				0.378	0.326			11,15	0.066	16,21	0.001
16.2													0.001			11,16	0.013	16,22	0.001
17	0.214							0.006		0.001	0.001	0.148	0.262	0.001		11,17	0.001	17,17	0.002
17.2													0.002			11,19	0.001	17,18	0.004
18	0.004									0.011	0.011	0.027	0.074	0.001		12,12	0.002	17,19	0.002
19	0.001									0.227	0.004	0.004	0.018	0.002		12,13	0.004	18,18	0.002
19.2										0.001						12,14	0.036	18,19	0.001
20										0.686			0.001	0.018		12,15	0.018	18,21	0.001
20.2													0.001			12,16	0.003	19,20	0.001
21										0.004	0.068			0.105	0.547	12,17	0.001		
22										0.076	0.004			0.102	0.154	12,18	0.001		
23										0.137	0.001			0.631	0.018	13,13	0.006		
24										0.267				0.116		13,14	0.026		
25										0.461				0.021		13,15	0.015		
26										0.001	0.051			0.004		13,16	0.006		
27										0.002	0.004			0.001		13,17	0.006		
28										0.092						13,17.2	0.001		
29										0.278						13,18	0.002		
30										0.391									
31										0.181									
32										0.046									
33										0.009									
Allel/DYS	19	389I	389II	390	391	392	393	437	438	439	448	456	458	635	Y-GATA-H4.1	Allelic class 385a/b			
GD	0.757	0.417	0.718	0.701	0.504	0.403	0.287	0.449	0.635	0.710	0.515	0.720	0.742	0.636	0.623	0.849			

PODSUMOWANIE

Porównanie międzypopulacyjne przeprowadzone w oparciu o analizę molekularnego dystansu (AMOVA) oraz obliczenia współczynnika R_{st} wykazały homogenność w obrębie badanych polskich populacji pochodzących z województw: łódzkiego, mazowieckiego oraz zachodnio-pomorskiego. Pozwoliło to na łączne wykorzystanie baz danych zgromadzonych haplotypów w zakresie 16 loci systemu AmpFISTR®Y-filer jako reprezentatywnych dla populacji Polski. Analiza polimorfizmu w grupie 1600 niespokrewnionych mężczyzn wykazała średnią wartość zmienności genowej $GD=60,01\%$

oraz łączną wartość zmienności haplotypowej $HD=99,95\%$, co dało niskie prawdopodobieństwo przypadkowego powtórzenia się haplotypu i wysoką siłę dyskryminacji badanego zestawu markerów. Uzyskane dane stanowiły punkt wyjścia do opracowania bazy danych „Y-STR Polska” służącej do oceny wartości dowodowej ekspertyz z zakresu genetyki sądowej. Jako narzędzie do przeszukiwania bazy zgromadzonych haplotypów Y-STR skonstruowano plik z danymi typu excel, który stanowi załącznik do niniejszej pracy i jest ogólnie dostępny na stronie internetowej www.genetyka-sadowa.pl.

Praca sfinansowana z grantu Uniwersytetu Medycznego w Łodzi Nr 502-11-736.

PIŚMIENNICTWO

1. Butler J. M., Schoske R., Vallone P. M., Kline M. C., Redd A. J., Hammer M. F.: A novel multiplex for simultaneous amplification of 20 Y-chromosome STR markers. *Forensic Sci. Int.* 2002, 129: 10-24.

2. Jacewicz R.: Badanie polimorfizmu DNA u osób po allogenicznym przeszczepie komórek macierzystych hematopoezy w aspekcie jego konsekwencji dla analiz z zakresu genetyki sądowej. Rozprawa habilitacyjna. Pracownia Genetyki Sądowej Katedry i Zakładu Medycyny Sądowej Uniwersytetu Medycznego w Łodzi. Biuro Promocji Wydawnictw Uniwersytetu Medycznego w Łodzi. Łódź 2009.

3. Roewer L., Kayser M., de Knijff P., Anslinger K., Betz A., Caglia A., Corach D., Fűredi S., Henke L., Hidding M., Kűrgel H. J., Lessig R., Nagy M., Pascali V. L., Parson W., Rolf B., Schmitt C., Szibor R., Teifel-Greding J., Krawczak M.: A new method for evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. *Forensic Sci. Int.* 2000, 114: 31-43.

4. <http://www.yhrd.org/>. YHRD.Org 3.0, Y-STR Haplotype Reference Database, release 35.

5. Excoffier L., Laval G., Schneider S.: Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* 1. 2005: 47-50.

6. Nei M.: Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*. 1978, 89: 583-590.

7. <http://www.ncbi.nlm.nih.gov/Genbank/>

8. <http://www.cstl.nist.gov/strbase/>

9. Gusmão L., Butler J. M., Carracedo A., Gill P., Kayser M., Mayr W. R., Morling N., Prinz M., Roewer L., Tyler-Smith C., Schneider P. M.: DNA Commission of International Society of Forensic Genetics (ISFG): an update of recommendations on the use of Y-STRs in forensic analysis. *Forensic Sci. Int.* 2006, 157: 187-197.

10. Mulero J. J., Budowle B., Butler J. M., Gusmão L.: Nomenclature and allele repeat structure update for the Y-STR locus GATA H4. *J. Forensic Sci.* 2006, 51: 694.

11. Veselinovic I. S., Zgonjanin D. M., Maletin M. P., Stojkovic O., Djurendic-Brenesel M., Vukovic R. M., Tasic M. M.: Allele frequencies and population data for 17 Y-chromosome STR loci in a Serbian population sample from Vojvodina province. *Forensic Sci. Int.* 2008, 176: 23-28.

12. Sanchez C., Barrot C., Xifro A., Ortega M., Gomez de Aranda I., Huguet E., Corbella J., Gené M.: Haplotype frequencies of 16 Y-chromosome STR loci in the Barcelona metropolitan area population using Y-Filer™ kit, *Forensic Sci. Int.* 2007, 172: 211-217.

13. Alves C., Gomes V., Joao Prata M., Amorim A., Gusmao L.: Population data for Y-chromosome

haplotypes defined by 17 STRs (AmpFISTR YFiler) in Portugal. *Forensic Sci. Int.* 2007, 171: 250-255.

14. Pereira R. W., Monteiro E. H. G., Hirschfeld G. C. R., Wang A. Y., Grattapaglia D.: Haplotype diversity of 17 Y-chromosome STRs in Brazilians. *Forensic Sci. Int.* 2007, 171: 226-236.

15. Hara M., Kido A., Takada A., Adachi N., Saito K.: Genetic data for 16 Y-chromosomal STR loci in Japanese. *Legal Med.* 9. 2007: 161-170.

16. Long K. Li., Kai L., Yuming M.: Y chromosome STR haplotypes of Tibetan Living Tibet Lassa. *Forensic Sci. Int.* 2007, 172: 79-83.

Adres do korespondencji:

Katedra i Zakład Medycyny Sądowej

Uniwersytetu Medycznego w Łodzi

ul. Sędziowska 18 a

91-304 Łódź

e-mail: renata.jacewicz@umed.lodz.pl