

Katarzyna Bąbol-Pokora, Adam Prośniak, Renata Jacewicz, Jarosław Berent

## Baza 500 alleli SNP w populacji centralnej Polski<sup>1</sup>

### The central Poland population database of 500 SNP alleles

Z Katedry i Zakładu Medycyny Sądowej Uniwersytetu Medycznego w Łodzi  
Kierownik: prof. nadzw. dr hab. n. med. J. Berent

Analiza SNP jest jedną z najnowszych metod stosowanych do identyfikacji osobniczej w genetyce sądowej. Jest coraz częściej stosowana w praktyce, zwłaszcza w analizach silnie zdegradowanego materiału dowodowego, dlatego też wymaga odpowiednich danych populacyjnych. W pracy przedstawiono wyniki badań populacyjnych, przeprowadzonych metodą minisekwencjonowania, w postaci bazy populacyjnej regionu Polski centralnej, obejmującej 500 alleli w zakresie 5 loci SNP (rs2294067, rs2282160, rs2070764, rs2277216, rs1063739).

SNP analysis is one of the most contemporary methods for personal identification in forensic genetics. It is increasingly more frequently used in forensic practice, especially for analyses of highly degraded DNA samples from crime scenes and thus it requires suitable population data. The aim of this study was to develop a central Poland population database consisting of 500 alleles in a range of 5 SNP biallelic loci (rs2294067, rs2282160, rs2070764, rs2277216, rs1063739). DNA fragments were amplified in one multiplex PCR reaction and SNPs were identified in a minisequencing reaction. The combined PD of the pentaplex was 0.9907147657. This makes the pentaplex a good screening method for forensic applications.

Słowa kluczowe: SNP, baza populacyjna, Polska centralna

Key words: SNP, population database, central Poland

#### WSTĘP

Markery SNP (Single Nucleotide Polymorphism) to utrwalone mutacje punktowe DNA, powstające w wyniku tranzycji lub transwersji. Występują w całym genomie ludzkim, w DNA jądrowym i mitochondrialnym, zarówno w regionach kodujących, jak i niekodujących. Są doskonałym narzędziem do identyfikacji osobniczej oraz badań pokrewieństwa, dzięki czemu są wykorzystywane w genetyce sądowej. Przede wszystkim mają znacznie niższy wskaźnik mutacyjności niż powszechnie stosowane w genetyce sądowej markery STR i VNTR, co ma istotne znaczenie w analizach spornego ojcostwa [1, 2]. Ponadto charakteryzują się polimorfizmem binarnym, dzięki czemu są znacznie łatwiejsze do walidacji niż markery STR i VNTR, gdyż dokładne określenie częstości alleli w przypadku markerów SNP nie wymaga zgenotypowania tak dużej grupy osób, jak to ma miejsce w przypadku STR i VNTR [3]. SNP mogą być genotypowane przy użyciu nowoczesnych technologii i bardzo wydajnych metod, których jest coraz więcej, a ich automatyzacja ma zastosowanie w tworzeniu dużych baz profili genetycznych przestępców [4]. Największą zaletą tych markerów jest jednak to, że podczas ich analizy amplifikacji poddawane są bardzo krótkie fragmenty, najczęściej poniżej 100 pz, co sprawia, że są one idealnym narzędziem do badań antropologicznych i śladów biologicznych z miejsc przestępstw, umożliwiają bowiem analizę bardzo małych ilości DNA, który dodatkowo może być silnie zdegradowany [5]. Metoda z użyciem SNP może zatem zminimalizować problemy związane z identyfikacją zdegradowanego materiału biologicznego, gdzie zastosowanie metod

<sup>1</sup> Temat opracowany w ramach prac własnych Uniwersytetu Medycznego w Łodzi nr 502-11-373

opartych na analizie STR nie daje zazwyczaj zadowalających wyników.

Celem prezentowanej pracy było opracowanie bazy populacyjnej centralnej Polski, składającej się z 500 alleli, w zakresie 5 loci SNP (rs2294067, rs2282160, rs2070764, rs2277216, rs1063739) [6].

## MATERIAŁ I METODY

Materiał do utworzenia bazy populacyjnej stanowiły próbki krwi pobrane od 250 niespokrewnionych osób biorących udział w standardowych badaniach identyfikacyjnych w Katedrze i Zakładzie Medycyny Sądowej Uniwersytetu Medycznego w Łodzi. DNA wyizolowano metodą kolumnkową z zastosowaniem zestawu Sherlock AX (A&A Biotechnology). Kompleksowa reakcja amplifikacji została przeprowadzona z użyciem pięciu par starterów, obejmujących następujące loci SNP: rs2294067, rs2282160, rs2070764, rs2277216, rs1063739 [6]. Produkty PCR zostały oczyszczone za pomocą zestawu kolumnek ultrafiltracyjnych MiniElute® (Qiagen). Minisekwencjonowanie przeprowadzono z użyciem zestawu ABI Prism® SnaPshot™ (Applied Biosystems). Produkty minisekwencjonowania, po oczyszczeniu zestawem DyeEx® (Qiagen), zostały rozdzielone w poliakrylamidowym żelu denaturującym w sekwenatorze ABI Prism 377 z użyciem LIZ™ 120

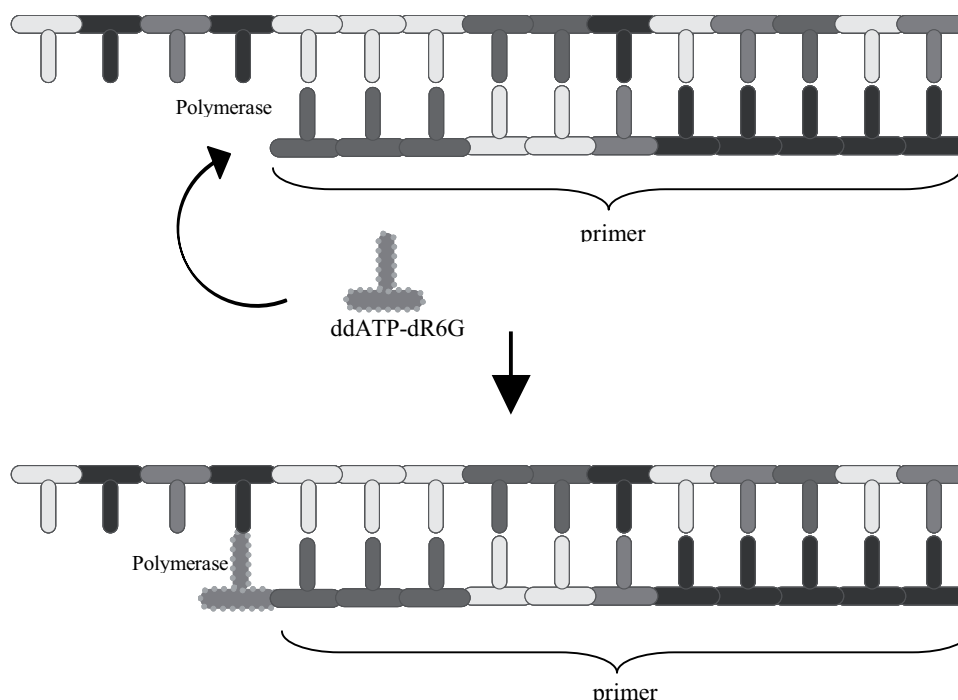
jako wewnętrznego standardu wielkości. Genotypowanie przeprowadzono za pomocą programu Gene Scan™. Analiza statystyczna, którą przeprowadzono przy użyciu programów komputerowych GDA [7] i DLP [8], objęła zgodność z prawem Hardy'ego-Weinberga (exact p) oraz parametry przydatności pentapleksu SNP w standardowych analizach genetyki sądowej: tj. heterozygotyczność obserwowaną i oczekiwaną (Het obs.; Het exp) oraz jej błąd [9], współczynnik informacji o polimorfizmie (PIC) [10], minimalną i maksymalną [11] oraz średnią szansę ojcostwa (PI) [12], siłę dyskryminacji (PD) [13] oraz siłę wykluczenia dla pełnych ( $PE_{trio}$ ) [14] i niepełnych trójek ( $PE_{duo}$ ) [15].

## WYNIKI I DYSKUSJA

Opracowano bazę, składającą się z 500 alleli. Tabela I przedstawia uzyskany rozkład częstości występowania poszczególnych wariantów allelicznych badanych SNP w populacji regionu łódzkiego. Ocena testem exact wykazała zgodność z prawem Hardy'ego-Weinberga we wszystkich loci. Łączna siła dyskryminacji pentapleksu SNP wyniosła:  $PD_{total} = 0,9907147657$ . Łączna siła wykluczenia dla trójek osiągnęła wartość:  $PE_{trio} = 0,628877578$ , natomiast w sprawach bez matki wyniosła:  $PE_{duo} = 0,4495078358$ .

Ryc. 1. Schemat reakcji minisekwencjonowania, czyli wydłużania startera o jeden komplementarny znakowany dideoksyrybonukleotyd.

Fig. 1. Minisequencing process, i.e. Single Base Extension Reaction scheme.



Parametry statystyczne, obliczone dla każdego SNP, zostały przedstawione w tabeli I. Najczęstszy profil, jaki otrzymano w zakresie 5 badanych SNP, miał częstość:  $f_{\text{total}} = 0,033$  czyli genotyp taki występował u jednej na 30 osób, natomiast profil najrzadszy miał

częstość:  $f_{\text{total}} = 0,00012$ . Oznacza to, że występował on u jednej na 8297 osób. Średnia częstość profilu wyniosła 0,0089, czyli profil ten oznaczono u 1 na 283 osoby

Tabela I. Parametry statystyczne bazy populacyjnej 500 alleli dla poszczególnych markerów SNP, obliczone za pomocą programów GDA i DLP.

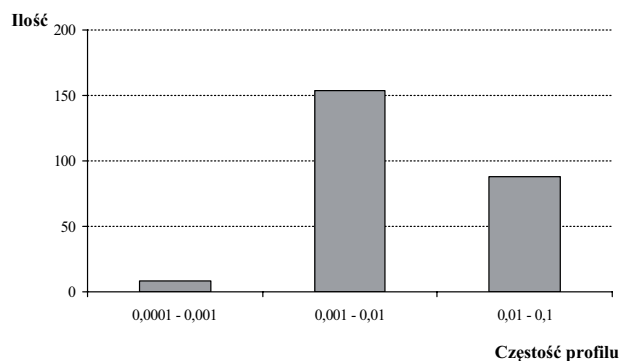
Table I. Statistical parameters of every SNP counted for population database of 500 alleles using GDA and DLP programs.

Locus	Allel Allele	p exact test	PD	Het			PIC	PE		PI		
				exp	obs	błąd		trio	duo	min	średnia	max
rs2294067	C	0,803	0,621	0,493	0,484	0,02	0,371	0,186	0,121	0,890	0,987	2,283
	G											
rs2070764	T	0,900	0,608	0,488	0,488	0,02	0,369	0,184	0,119	0,862	0,977	2,281
	A											
rs1063739	A	0,602	0,625	0,501	0,520	0,02	0,375	0,187	0,125	0,992	1,002	2,016
	C											
rs2282160	G	0,381	0,625	0,499	0,468	0,02	0,374	0,187	0,124	0,943	0,998	2,128
	A											
rs2277216	C	0,404	0,494	0,382	0,360	0,02	0,308	0,154	0,073	0,672	0,810	3,906
	T											

Wykres na ryc. 2 pokazuje rozkład częstości profili w zakresie 5 SNP w populacji regionu łódzkiego. Najwięcej profili, występujących w bazie 250 osób, mieści się w zakresie częstości 0,001–0,01.

Ryc. 2. Rozkład częstości profili w zakresie 5 SNP w populacji 250 mieszkańców centralnej Polski.

Fig. 2. A distribution of genetic profiles frequencies in the range of 5 SNP loci among 250 subjects from the central Poland population.



Badane w pracy markery SNP różniły się częstościami alleli występującymi w populacji Polski centralnej w porównaniu z innymi populacjami, np. badanymi w ramach projektu HapMap Project, tj.: 60 Europejczyków (HapMap-CEU) [16], 45 Chińczyków z Pekinu (HapMap-HCB) [17] i 60 Nigeryjczyków (HapMap-YRI) [18]; populacją japońską (JBIC) [19], czy grupą CEPH (Centre d'Etude du Polymorphisme Humain) [20]. Różnice te przedstawiono w tabeli II.

Powyższe dane świadczą o przydatności markerów SNP do badań populacyjnych. Bialleliczny charakter układów SNP sprawia, że baza składająca się z 500 alleli jest wystarczającą bazą populacyjną, która z powodzeniem może zostać wykorzystana do obliczeń statystycznych zarówno w badaniach dowodów rzeczowych, jak i ojcostwa.

Tabela II. Porównanie częstości alleli pięciu loci SNP w populacji centralnej Polski oraz 6 innych populacjach.  
Table II. A comparison of allele frequency of five SNP loci between the central Poland population and six other populations.

		Częstości wg baz populacyjnych Frequencies according to population databases						
Bazy populacyjne Population databases		1	2[3]	3 [24]	4 [19]	5[20]	6[22]	7[23]
Liczba alleli w bazie Number of alleles		500	204	184	120	90	120	1276-1496
rs2294067	G	0,562	0,637	-	0,412	0,614	0,170	0,556
	C	0,438	0,363	-	0,588	0,386	0,830	0,444
rs2070764	A	0,420	0,446	-	0,308	0,422	0,525	0,453
	T	0,580	0,554	-	0,692	0,578	0,475	0,547
rs1063739	A	0,496	0,392	0,660	-	-	-	-
	C	0,504	0,608	0,340	-	-	-	-
rs2282160	A	0,530	0,544	-	0,558	0,511	0,746	0,493
	G	0,470	0,456	-	0,442	0,489	0,254	0,507
rs2277216	T	0,256	0,520	-	0,208	0,467	0,500	0,483
	C	0,744	0,480	-	0,792	0,533	0,500	0,517

1 – 250 niespokrewnionych mieszkańców Polski centralnej – populacja badana

2 – 102 niespokrewnionych Japończyków

3 – CEPH – grupa 92 niespokrewnionych mieszkańców stanu Utah

4 – HapMap-CEU – grupa 60 Europejczyków pochodzących z północnej Europy

5 – HapMap-HCB – grupa 45 niespokrewnionych Chińczyków z Pekinu

6 – HapMap-YRI – grupa 60 Nigeryjczyków

7 – JBIC – grupa niespokrewnionych Japończyków, zmienna w zależności od SNP

tj.: 748 dla rs2294067, 653 dla rs2070764, 744 dla rs2282160 i 638 dla rs2277216

1 – 250 unrelated subjects from central Poland

2 – 102 unrelated Japanese

3 – CEPH – group of 92 unrelated subjects from Utah

4 – HapMap-CEU – group of 60 North Europeans

5 – HapMap-HCB – group of 45 unrelated Chinese from Beijing

6 – HapMap-YRI – group of 60 Nigerians

7 – JBIC – unrelated Japanese, different for every SNP – 748 for rs2294067, 653 for rs2070764, 744 for rs2282160 and 638 for rs2277216

## PIŚMIENNICTWO

1. Sobrino B., Brion M., Carracedo A.: SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci. Int.*, 2005, 154, 181-194.

2. Inagaki S., Yamamoto Y., Doi Y.: et al. A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Sci. Int.*, 2004, 144, 45–57.

3. Jacewicz R., Berent J., Bąbol K., Szram S.: Rozkład częstości alleli w 10 loci STR w regionie

centralnej Polski. *Arch. Med. Sąd. Krym.*, 2004, 54(1), 25-28.

4. Martin P.D., Schmitter H., Schneider P.M.: A brief history of the formation of DNA databases in forensic science within Europe. *Forensic Sci. Int.*, 2001, 119, 225-231.

5. Alonso A., Martin P., Albarran C.: et al. Specific quantification of human genomes from low copy number DNA samples in forensic and ancient DNA studies. *Croat. Med. J.*, 2003, 44, 273-280.

6. Bąbol-Pokora K., Prośniak A., Jacewicz R., Berent J.: Pentapleks SNP – rozkład częstości alleli w populacji centralnej Polski. Arch. Med. Sąd. Krym., 2006, 56 (4), 228-231.
7. Lewis P. O., Zaykin D.: Genetic Data Analysis: Computer program for the analysis of allelic data. Version 1.0 (d16c), 2001, <http://lewis.eeb.uconn.edu/lewishome>
8. Berent J., Szram S.: DLP – a computer program for calculation of discrete locus parameters. Forensic Sci., 2003, 1, 46-47.
9. Nei M., Roychoudhury A. K.: Sampling variants of heterozygosity and genetic distance. Genetics, 1974, 76, 379-390.
10. Botstein D., White R. L., Skolnick M., Davis R. W.: Construction of a genetic linkage map in man using restriction fragment length polymorphism. Am. J. Hum. Genet., 1980, 32, 314-331.
11. Berent J. A., Myścicka-Śliwka D., Czarny J.: Średnie wartości szansy ojcostwa – obliczenia dla populacji polskiej. Arch. Med. Sąd. i Kryminol., 1999, 49, 11-15.
12. Brenner C., Morris J. W.: Paternity index calculations in single locus hypervariable DNA probes: validation and other studies. Proceedings from the International Symposium on Human Identification 1989, Promega, Madison, 1990, pp. 21-53.
13. National Research Council Report II. The Evaluation of Forensic DNA Evidence. National Academy Press, Washington, D.C., 1996, pp. 96-97.
14. Weir B. S.: Genetic Data Analysis II. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, 1996, pp. 209-211.
15. Fung W. K., Chung Y. K., Wong D. M.: Power of exclusion revisited: probability of excluding relatives of the true father from paternity. Int. J. Legal Med., 2002, 116, 64-67.
16. [http://www.ncbi.nlm.nih.gov/SNP/snp\\_viewTable.cgi?pop=1409](http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?pop=1409)
17. [http://www.ncbi.nlm.nih.gov/SNP/snp\\_viewTable.cgi?pop=1410](http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?pop=1410)
18. [http://www.ncbi.nlm.nih.gov/SNP/snp\\_viewTable.cgi?pop=1412](http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?pop=1412)
19. [http://www.ncbi.nlm.nih.gov/SNP/snp\\_viewTable.cgi?pop=738](http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?pop=738)
20. [http://www.ncbi.nlm.nih.gov/SNP/snp\\_viewTable.cgi?pop=1303](http://www.ncbi.nlm.nih.gov/SNP/snp_viewTable.cgi?pop=1303)

Adres pierwszego autora:  
Katedra i Zakład Medycyny Sądowej  
Uniwersytetu Medycznego w Łodzi  
ul. Sędziowska 18 a  
91-304 Łódź  
[katarzynababol@pro.wp.pl](mailto:katarzynababol@pro.wp.pl)