



HEPTAPLEX SNP – AN ALLELE FREQUENCIES DATABASE OF THE SOUTH-WEST POLAND POPULATION

Magdalena SEKOWSKA, Dorota MUSIAŁ, Arleta LEBIODA, Iwona CHROMIK, Tadeusz DOBOSZ

Department of Forensic Medicine, Wrocław Medical University, Wrocław, Poland

Abstract

159 samples from the South-West Poland population were typed for seven bi-allelic, autosomal SNP (single nucleotide polymorphism) markers. All forensic parameters, allele frequencies as well as *HWE* were calculated for each SNP. The heptaplex-SNP panel includes the following loci: rs2013526, rs730249, rs1063739, rs727206, rs643788, rs999842 and rs877228. The created SNP multiplex might be considered as a useful alternative method for forensic identification purposes, particularly for challenging samples such as degraded or low copy number DNA present in forensic evidence and archival tissues. The 7-locus SNP multiplex can also act as a supplementary test to the currently available paternity testing techniques.

Key words

Single nucleotide polymorphism (SNP); Population database; Minisequencing.

Received 23 March 2010; accepted 11 May 2010

1. Population samples

To determine the allele frequencies of 7 SNPs, 159 unrelated DNA samples were collected from adult persons living in South-Western Poland (DNA Bank, Molecular Technique Unit of Medical University of Wrocław).

2. Selection of SNP loci

Five SNP loci were selected according to dbSNP (www.ncbi.nlm.nih.gov/SNP) and data published by Vallone et al. [12]. Two SNP loci were selected according to dbSNP (rs1063739: chromosome 8, reference allele T; rs643788: chromosome 11, reference allele T). The SNP selected were at least 50 megabases apart.

3. DNA extraction

Blood samples were extracted using the QIAamp DNA Mini Kit (Qiagen), according to the manufacturer's instructions (QIAamp DNA Mini Kit and QIAamp DNA Blood Mini Kit Handbook 02/2003).

4. Multiplex amplification of DNA fragments

The sets of primers used in PCR and SNaPshot reactions are shown in Table I. The exact chromosomal locations were ascertained using BLAST (<http://genome.ucsc.edu/cgi-bin/hgBlat>) and dbSNP (www.ncbi.nlm.nih.gov/SNP). The heptaplex PCR reaction was performed in a volume of 10 μ l, containing 1.0 μ l template DNA, 5 μ l Multiplex PCR Master Mix (Qiagen), 1.0 μ l mixture of 14 primers with final concentration 0.3 μ M of each and 3 μ l H₂O. The thermal cycling program was carried out on a GeneAmp 9700

TABLE I. SETS OF PRIMERS USED IN PCR AND SNaPshot REACTION

SNP	Sequence of primers	PCR product size [bp]	SNaPshot product size [nt]	Final concentration in SNaPshot reaction [nmol]	Melting temperature
rs1063739F	5' GGAGCCAGGCGGTGAGG 3'	37	18	50.0	60
rs1063739R	5' ACCTTCGCCCTCAGTCCAT 3'		20	60.0	60
rs643788F	5' GATATCGAATGGAGAAGGTGA 3'	40	22	50.0	60
rs643788R	5' CAGATCCTGGGCAGTGCC 3'		19	70.0	60
rs2013526F	5' GAGACTACAAGTTCAAATATACTAAA CTATTC 3'	54	33	65.0	64
rs2013526R	5' GCCAGATTCTACCAGGATCAG 3'		22	70.0	64
rs730249F	5' GACTGACTGACTGAACTTCCTGATAT GGGAAATCAAA 3'	71	37	40.0	66
rs730249R	5' GAGACTGACTGTATATAACCAGAAGG AACACTCA 3'		35	50.0	66
rs877228F	5' GAGACTGACTGACTGACTGACTACAC AGTTTGGAGTGACCCAA 3'	82	44	15.0	62
rs877228R	5' GACTGACTGACTGACTGATGGAGATT CTGGACATTCCA 3'		39	30.0	64
rs727206F	5' GACTGACTGACTGACTGACTGACTGA CTGTCACCCTCAGAGCCCAGAA 3'	91	49	20.0	64
rs727206R	5' GAGACTGACTGACTGACTTCCTGTTTT CTGTTATATTCTGCT 3'		43	30.0	64
rs999842F	5' GAGACTGACTGACTGACTGACTGACTG ACTGACTGACTAGTATGCTTTTGCAAAA GGTCCA 3'	114	62	40.0	64
rs999842R	5' GAGACTGACTGACTGACTGACTGACT GACTGGATGAAAATCTCGCACCATCT 3'		53	50.0	64

(Applied Biosystems) using the following conditions: one cycle at 95°C for 15 min, followed by 30 cycles at 94°C for 30 s, 57°C for 90 s, 72°C for 90 s, and final extension at 72°C for 10 min. After multiplex amplification, unincorporated primers and dNTPs were removed by enzymatic treatment. 1.4 μ l of Exo-SAP enzyme cocktail, consisting of Exonuclease I (Exo I, 0.4 U/ μ l, Fermentas) and Shrimp Alkaline Phosphatase (SAP, 0.2 U/ μ l, Fermentas) in the ratio 1:10 was used to purify each 4 μ l of PCR template. Samples were mixed briefly and incubated at 37°C for 30 min and then at 80°C for 15 min to inactivate the enzymes.

5. Minisequencing

A multiplex primer extension reaction was carried out in a total volume of 5 μ l using 1.5 μ l of ABI Prism[®] SNaPshot[™] Multiplex Ready Reaction Mix (Applied Biosystems), 2.5 μ l water, 0.5 μ l of PCR template, and 0.5 μ l stock solution of extension primers, which contained empirically balanced primers (Table I). Extension reactions were incubated as follows: 25 cycles at 96°C for 10 s, 57°C for 5 s, and 60°C for 30 s. Excess fluorescently-labelled ddNTPs were inactivated by addition of 0.4 μ l SAP (1 U/ μ l). Samples were mixed briefly and incubated at 37°C for 60 min, then at 80°C for 15 min.

6. Electrophoresis

A 1 µl aliquot of each SAP-treated primer extension product was diluted in 10 µl Hi-Di™ formamide and 0.4 µl GS 120 LIZ internal size standard (Applied Biosystems) and after heat denaturation, analyzed on the 3130 Genetic Analyzer (Applied Biosystems) using filter set E5. Samples were injected electrokinetically for 4 s at 1 kV. Separations were performed in approximately 20 min on a 36 cm capillary array using polymer POP-4™ (Applied Biosystems). After electrophoresis, the data were analyzed by GeneMapper ID v. 3.2 software (Applied Biosystems).

7. Analysis of data

The data were analyzed with Power Stats v.12 software [6]. The power of discrimination (PD) was calculated according to Jones [7] and National Research Council Report II (1996). The power of exclusion (PE) was calculated according to Weir [13] and Fung et al. [3]. The polymorphism information content (PIC) was calculated according to Botstein et al. [1]. Divergence from Hardy-Weinberg expectations (HWE) was determined by the χ^2 test. Allele frequencies, MP – match-

ing probability, Ho – observed heterozygosity, He – expected heterozygosity, and p -value were included.

8. Results and discussion

All forensic parameters, allele and genotype frequencies, as well as HWE are provided in Table II. No deviations from Hardy-Weinberg equilibrium were observed in any locus ($P > 0.05$). The following heterozygosities were observed: rs2013526 (60.4%), rs730249 (55.3%), rs1063739 (49.1%), rs727206 (48.4%), rs643788 (46.5%), rs999842 (46.5%) and rs877228 (43.4%). The power of discrimination of the 7-locus SNP multiplex system is $PD_{total} = 0.998$. The cumulative power of exclusion for this test system is $PE_{total} = 0.778$. All calculated forensic parameters (heterozygosity, PD , PIC , PE) confirmed the informative role of the investigated markers. Thus they can be successfully used in DNA typing, paternity testing or population studies.

Currently, STRs (short tandem repeats) markers are widely used in forensic medicine. The advantages of SNP typing in forensic genetics are also well known and include a wider choice of high-throughput platforms, lower mutation rates, lower amplicon size and

TABLE II. ALLELE AND GENOTYPE FREQUENCIES AND OTHER FORENSIC PARAMETERS FOR EACH SNP IN THE SOUTH-WEST POLAND POPULATION

Parameter	rs1063739	rs643788	rs2013526	rs730249	rs877228	rs727206	rs999842
Allele C [%]	50.3	34.6	42.1	49.1	58.2	63.2	59.1
Allele T [%]	49.7	65.4	57.9	50.9	41.8	36.8	40.9
Genotype CC	0.258	0.1132	0.119	0.214	0.365	0.390	0.359
Genotype CT	0.491	0.4654	0.604	0.553	0.434	0.48	0.465
Genotype TT	0.251	0.4214	0.277	0.233	0.201	0.126	0.176
MP	0.370	0.407	0.455	0.406	0.362	0.402	0.376
PE	0.179	0.159	0.295	0.239	0.136	0.174	0.159
PD	0.630	0.593	0.545	0.594	0.638	0.598	0.624
PIC	0.37	0.35	0.37	0.37	0.37	0.36	0.37
Ho	49.1	46.5	60.4	55.3	43.4	48.4	46.5
He	50.0	45.2	48.8	50.0	48.7	46.5	48.3
HWE (p -value)	$p > 0.99$	$p > 0,975$	$p > 0,975$	$p > 0,975$	$p > 0,90$	$p > 0,975$	$p > 0,975$
PD_{total}				0.998			
PE_{total}				0.778			

MP – matching probability, PE – power of exclusion, PD – power of discrimination, PIC – polymorphic information content, Ho – observed heterozygosity, He – expected heterozygosity, PD_{total} – power of discrimination of 7 SNP loci, PE_{total} – power of exclusion of 7 SNP loci.

improved analysis of degraded samples. Therefore SNPs are being considered for a potentially useful role in forensic human identification [2, 8, 9, 10, 11, 12]. The mentioned papers are only a small part of the total publications concerning the role of SNP in forensic genetics. All of them support utilization of SNPs for forensic identification purposes.

The novel Heptaplex-SNP analysis system might be a good alternative method when STRs fail to give results, especially for challenging samples such as bones, teeth and highly degraded tissue commonly encountered in mass-disasters.

References

1. Botstein D., White R., Skolnicki M. [et al.], Construction of a genetic linkage map in man using restriction fragment length polymorphisms, *American Journal of Human Genetics* 1980, 32, 314–331.
2. Dixon L., Murray C., Archer E. [et al.], Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes, *Forensic Science International* 2005, 154, 62–77.
3. Fung W., Chung Y., Wong D., Power of exclusion revisited: probability of excluding relatives of the true father from paternity, *International Journal of Legal Medicine* 2002, 116, 64–67.
4. <http://genome.ucsc.edu/cgi-bin/hgBlat>
5. <http://www.ncbi.nlm.nih.gov/SNP>
6. <http://www.promega.com/geneticidtools/default.htm>
7. Jones D., Blood samples: Probability of discrimination, *Journal of the Forensic Science Society* 1972, 12, 355–359.
8. Kidd K., Pakstis A., Speed W. [et al.], Developing a SNP panel for forensic identification of individuals, *Forensic Science International* 2005, 164, 20–32.
9. Phillips C., Fang R., Ballard D. [et al.], Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel, *Forensic Science International: Genetics* 2007, 1, 180–185.
10. Sanchez J., Phillips C., Børsting C. [et al.], A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis* 2006, 9, 1713–1724.
11. Vallone P. M., Decker A. E., Butler J. M., Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples, *Forensic Science International* 2005, 149, 279–286.
12. Vallone P. M., Decker A. E., Coble M. D. [et al.], The evaluation of an autosomal SNP 12-plex assay, *International Congress Series* 2006, 1288, 6–63.
13. Weir B., Genetic Data Analysis II, Sinauer Associates, Inc. Publishers, Sunderland 1996.

Corresponding author

dr Dorota Musiał
Zakład Medycyny Sądowej
Akademia Medyczna we Wrocławiu
ul. M. Skłodowskiej-Curie 52
PL 50-369 Wrocław
e-mail: dorotac@forensic.am.wroc.pl

HEPTAPLEKS SNP – ROZKŁAD CZĘSTOŚCI ALLELI WYBRANYCH POZYCJI TYPU SNP W POPULACJI DOLNOŚLĄSKIEJ

1. Próbkki populacyjne

W celu określenia częstości występowania alleli 7 pozycji typu SNP zebrano próbki DNA od 159 niespokrewnionych dorosłych osób zamieszkujących teren południowo-zachodniej Polski (Bank DNA, Zakład Techniki Molekularnych Medycznego Uniwersytetu we Wrocławiu).

2. Wybór loci SNP

Pięć loci SNP wybrano na podstawie analizy danych zawartych w bazie dbSNP (www.ncbi.nlm.nih.gov/SNP) oraz danych opublikowanych przez Vallone i in. [12]. Dwa loci SNP wybrano na podstawie analizy danych zawartych w bazie dbSNP (rs1063739: chromosom 8, allel odniesienia – T; rs643788: chromosom 11, allel odniesienia – T). Wybrane loci SNP oddalone są od siebie co najmniej o 50 milionów par zasad.

3. Ekstrakcja DNA

Próbki krwi izolowano z zastosowaniem zestawu QIAamp DNA Mini Kit (Qiagen), stosując się do zaleceń producenta (QIAamp DNA Blood Mini Kit Handbook 02/2003).

4. Amplifikacja DNA metodą multipleks PCR

Zestawy starterów zastosowanych do reakcji PCR oraz SNaPshot przedstawiono w tabeli I. Szczegóły dotyczące ich położenia na chromosomach ustalono przy pomocy programu BLAST (<http://genome.ucsc.edu/cgi-bin/hgBlat>) oraz informacji zawartych w bazie dbSNP (www.ncbi.nlm.nih.gov/SNP). Reakcję PCR umożliwiającą jednoczesną amplifikację 7 loci SNP (heptapleks) prowadzono w całkowitej objętości roztworu 10 μ l i składał się on z 1 μ l matrycowego DNA, 5 μ l mieszaniny reakcyjnej Multiplex PCR Master Mix (Qiagen), 1 μ l mieszaniny 14 starterów o ostatecznym stężeniu każdego równym 0,3 M oraz 3 μ l H₂O. Reakcje amplifikacji prowadzono przy użyciu aparatu GeneAmp 9700 (Applied Biosystems), stosując następujący profil temperaturowy: 95°C przez 15 min (1 cykl), a następnie 30 cykli: 94°C przez 30 s, 57°C przez 90 s, 72°C przez 90 s i wydłużanie końcowe – 72°C przez 10 min. Po reakcji amplifikacji metodą multipleks PCR niewykorzystane startery oraz

nukleotydy usuwano metodą enzymatyczną. W tym celu 1,4 μ l mieszaniny enzymów Exo-SAP złożonej z egzonukleazy I (Exo I, 0,4 U/ μ l, Fermentas) oraz alkalicznej fosfatazy (SAP, 0,2 U/ μ l, Fermentas) połączonych w stosunku 1:10 dodano do 4 μ l produktu PCR. Próbkki krótko mieszano i inkubowano w temperaturze 37°C przez 30 min, a następnie w temperaturze 80°C przez 15 min w celu inaktywacji enzymów.

5. Minisekwencjonowanie

Multipleksową reakcję wydłużania startera prowadzono w całkowitej objętości mieszaniny reakcyjnej 5 μ l składającej się z 1,5 μ l odczynnika ABI Prism[®] SNaPshot[™] Multiplex Ready Reaction Mix (Applied Biosystems), 2,5 μ l wody, 0,5 μ l produktu PCR i 0,5 μ l wyjściowej mieszaniny starterów wydłużających złożonej zawierającej startery o doświadczalnie dobranych stężeniach (tabela I). Reakcje wydłużania startera prowadzono w następujących warunkach temperaturowych: 25 cykli – 96°C przez 10 s, 57°C przez 5 s i 60°C przez 30 s. Nadmiar znakowanych fluorescencyjnie dideoksynukleotydów usuwano, dodając 0,4 μ l SAP (1 U/ μ l). Próbkki po krótkim wirowaniu inkubowano w temperaturze 37°C przez 60 min, a następnie w temperaturze 80°C przez 15 min.

6. Elektroforeza

1 μ l produktu wydłużania startera oczyszczonego za pomocą enzymu SAP rozpuszczano w 10 μ l formamidu (Hi-Di[™] formamide) oraz 0,4 μ l wewnętrznego standardu długości GS 120 LIZ (Applied Biosystems) i po denaturacji cieplnej poddawano rozdzielowi elektroforetycznemu z zastosowaniem analizatora genetycznego 3130 (Applied Biosystems) z filtrem E5. Próbkki podlegały elektrokinetycznej iniekcji przez 4 s przy napięciu 1 kV. Rozdział prowadzono przez 20 min przy użyciu zestawu kapilar o długości 36 cm i żelu POP-4[™] (Applied Biosystems). Po elektroforezie analizowano dane, używając programu GeneMapper ID v. 3.2 software (Applied Biosystems).

7. Analiza danych

Uzyskane dane analizowano z zastosowaniem programu PowerStats v 1.2 [6]. Siłą różnicującą (*PD*) obliczano według Jonesa [7] i wytycznych zawartych w Na-

tional Research Council Report II (1996). Siłę wykluczającą (PE) obliczano według Weira [13], Funga i in. [3]. Współczynnik zawartości polimorfizmu (PIC) obliczano według Botsteina i in. [1]. Odstępstwo od stanu równowagi Hardy’ego-Weinberga (HWE) badano za pomocą testu ². W obliczeniach uwzględniono również częstości alleli, prawdopodobieństwo zgodności (MP), obserwowaną heterozygotyczność (Ho), oczekiwaną heterozygotyczność (He) oraz wartości p .

8. Wyniki i dyskusja

Wszystkie współczynniki istotne dla celów sądowych, częstości alleli i genotypów oraz dane na temat heterozygotyczności obserwowanej i oczekiwanej (HWE) przedstawiono w tabeli II. W żadnym z badanych *loci* nie stwierdzono odstępstwa od stanu równowagi Hardy’ego-Weinberga ($P > 0.05$). Uzyskano następujące wyniki heterozygotyczności analizowanych markerów: rs2013526 (60,4%), rs730249 (55,3%), rs1063739 (49,1%), rs727206 (48,4%), rs643788 (46,5%), rs999842 (46,5%), rs877228 (43,4%). Siła różnicująca panelu multipleksowego złożonego z 7 *loci* typu SNP $PD_{total} = 0.998$.

Łączna siła wykluczenia dla opracowanego testu $PE_{total} = 0.778$. Wyniki uzyskane dla wszystkich obliczonych współczynników sądowych (heterozygotyczność, PD , PIC , PE) potwierdziły informatywność badanych markerów genetycznych, wskazując na ich użyteczność w profilowaniu DNA w kryminalistyce, analizie spornego ojcostwa oraz badaniach populacyjnych. Obecnie w badaniach medyczo sądowych powszechnie stosowane są markery typu STR (ang. short tandem repeats). Zalety badania markerów typu SNP w genetyce sądowej są również znane i należy do nich możliwość większego wyboru platform analitycznych o dużej przepustowości, niższe tempo mutacji, mniejsze rozmiary amplifikowanych fragmentów DNA i większe możliwości analizy próbek narażonych na dużą degradację. Markery typu SNP uznaje się więc jako potencjalnie użyteczne w badaniach identyfikacyjnych człowieka prowadzonych dla celów sądowych [2, 8, 9, 10, 11, 12]. Należy zauważyć, że zacytowane prace stanowią zaledwie niewielki ułamek publikacji poświęconych znaczeniu polimorfizmu SNP w genetyce sądowej. Wszystkie zwracają uwagę na użyteczność analizy polimorfizmu SNP do sądowych badań identyfikacyjnych. Przedstawiony nowy test do jednoczesnej analizy siedmiu pozycji typu SNP może stanowić pomocną alternatywę w sytuacji, gdy badania markerów typu STR kończą się wynikiem negatywnym, zwłaszcza w przypadku próbek szczególnie wymagających, jak kości, zęby lub silnie zdegradowane tkanki, często będące przedmiotem badań w sprawach katastrof z dużą liczbą ofiar.