

GENETIC VARIABILITY IN A PORTION OF CODING REGION OF MITOCHONDRIAL DNA ENCOMPASSING NUCLEOTIDE POSITIONS 9995–10995

Wojciech BRANICKI¹, Beata ZASADZKA², Tomasz KUPIEC¹

¹ *Institute of Forensic Research, Cracow*

² *Faculty of Biology and Earth Sciences, Jagiellonian University, Cracow*

ABSTRACT: The main limitation of the identification process based on mitochondrial DNA analysis is its relatively low power of discrimination. A reasonable solution to this drawback seems to be analysis of additional information present out of the routinely analysed control region. Due to the presence of medical information in the coding sequence of mitochondrial DNA, the analysis should encompass only particular polymorphic positions – Single Nucleotide Polymorphisms consisting of silent or neutral mutations and an appropriate variants frequency. The objective of this study was to determine the degree of variation characteristic of a fragment of the coding region of mitochondrial DNA and to find Single Nucleotide Polymorphisms useful for forensic identification studies.

KEY WORDS: Forensic science; mtDNA; Coding region; Sequence analysis; Single Nucleotide Polymorphism.

Z Zagadnień Nauk Sądowych, z. LI, 2002, 29–44

Received 28 October 2002; accepted 18 December 2002

INTRODUCTION

Sequencing of highly polymorphic segments of the control region of mitochondrial DNA (mtDNA) is today a routine method of analysis of biological traces which are not suitable for STR analysis due to insufficient concentration of nuclear DNA or heavy degradation processes. The high sensitivity characterising the analysis of mtDNA markers is mainly due to the fact that the mitochondrial genome is represented in hundreds, even thousands of copies in each cell of the organism [7]. In practice, sequence analysis of mtDNA is applied mainly to hair shafts or bones – the latter mainly in identification cases. One of the significant benefits linked with mtDNA analysis is its maternal inheritance. This mechanism of inheritance enables identification of human remains even in a situation where reference material can be obtained only from distant relatives of the person to be identified. Another important feature of the mitochondrial genome is its fast evolutionary rate, which is assessed to be 5 to 10 times higher than in nuclear DNA [3], provid-

ing a high degree of variation, particularly in highly polymorphic regions, termed HV1 and HV2 (hypervariable 1, 2). The mechanism of inheritance of mtDNA and lack of recombination processes associated with this molecule mean that all descendants of one mother have the same type of mtDNA. For this reason some haplotypes of mtDNA can be relatively often observed in a population and a high mutation rate alone is not able to eliminate the significant drawback associated with mtDNA analysis – the low discrimination power of the marker. Population data obtained for different regions of Europe and USA show that the most frequent haplotype of mtDNA in the Caucasian population (263G, 315.1C) is characteristic for about 4% of individuals [4, 15, 17, 19]. The frequency of this haplotype in a population sample from south Poland equals 5.3% (unpublished data from our laboratory). From the point of view of identification studies, the entire mitochondrial genome should be treated as a single genetic marker. Therefore, sequence analysis of additional segments within the mtDNA cannot increase the power of discrimination to an analogous degree to that obtained by raising the number of analysed unlinked STR squads. In spite of this, analysis of additional segments of the mitochondrial genome was proposed in order to limit the low power of discrimination of mtDNA analysis. One of the possible approaches is sequencing of an additional part of the control region, termed HV3 (following on from the remaining routinely analysed highly polymorphic segments of control region, HV1, HV2) [12]. The variation characteristic for this DNA fragment is lower than for HV1 and HV2 segments, but literature data as well as the practice of the Institute of Forensic Research indicate that this DNA fragment can be a valuable source of polymorphism in cases where differentiation between individuals possessing identical HV1 and HV2 haplotypes is necessary. A study performed on a population sample from southern Poland shows that about 30% of samples that are identical in terms of the HV1 and HV2 sequence can be differentiated by additional analysis of the HV3 region (genetic diversity = 0.68 – unpublished data from our laboratory). It was also ascertained that analysis of polymorphism characterising coding regions of mtDNA enables differentiation of individuals having identical HV1 and HV2 haplotypes [16]. It is well known that the coding DNA sequence constitutes the major part (93%) of the mitochondrial genome. The mtDNA molecule contains information concerning 13 peptides (subunits of enzymes involved in the chain of oxidative fosforylation and ATP production), two types of rRNA (12S and 16S rRNA) and 22 types of tRNA involved in the process of translation of proteins encoded by mtDNA (37 genes in total) [2]. It is not expected that the degree of variation characteristic for coding regions will be as high as in highly polymorphic segments of the control region. However, variation potentially included in a DNA fragment of about 15.5 thousand base pairs long could be high and in this context

analysis of the polymorphism occurring in the coding regions of mtDNA seems to be promising. There is abundant literature data describing the association between many genetic diseases and particular point mutations in the mitochondrial genome [23, 24]. This sort of data must not for ethical reasons be obtained in a forensic laboratory. Therefore, mutations associated with diseases with a genetic background must be avoided during identification studies. Unfortunately, when applying DNA sequencing, it is impossible to avoid certain medically significant nucleotide positions that are placed in the neighbourhood of variable positions valuable for identification purposes. There is also an additional significant objection concerning DNA sequencing as a routine method applied for analysis of coding regions of mtDNA: the sequencing technique itself, in spite of significant automation achieved in the last decade, is still very arduous and time-consuming. It is true that an ever increasing number of scientific papers describing application of DNA sequencing of entire mitochondrial genomes are being published – in the context of population studies as well [5, 9, 18] – but it is hard to imagine that this approach could be practical in forensic genetics. A promising approach in this context seems to be analysis of selected single nucleotide polymorphisms (SNPs) that are useful for identification purposes. Currently, there are very many techniques that enable selective analysis of single nucleotide polymorphisms. Deliberation concerning selection of the most suitable technique is beyond the scope of the present work: a detailed discussion of these techniques can be found in a review fully dedicated to this topic [11]. Before particular mtDNA positions can be used in routine analysis, a study enabling recognition of variable positions useful for forensic applications must be performed. For this purpose, the DNA sequencing technique is irreplaceable. The first reports are appearing in the literature describing a search for single nucleotide positions present in the coding region of mtDNA which would be useful for forensic identification studies [22].

The aim of this study was to sequence the portion of the coding region of mtDNA in order to ascertain the degree of variation present in this fragment and to find those particular polymorphic positions that fulfil conditions necessary for their future application in the identification process.

MATERIALS AND METHODS

Specimens

The study was performed on samples of blood or saliva collected from 50 unrelated individuals living in southern Poland. DNA was isolated using the standard organic method followed by double phenol-chloroform-isoamyl

alcohol extraction. The DNA concentration was measured by the fluorimetric method using the PicoGreen dsDNA Quantitation Kit (Molecular Probes) and Fluoroscan Ascent F1 apparatus (Labsystems).

Amplification

We amplified the portion of coding region encompassing positions 9995–10995, in accordance with the Anderson reference sequence [1]. This portion of DNA was amplified in two overlapping fragments using the following PCR primers designed for the purpose of this study:

- fragment I – 469 bp long encompassing positions 9995–10463; primer 1: 9974–9994; primer 2: 10464–10484, in accordance with the Anderson sequence;
- fragment II – 588 bp long encompassing positions 10408–10995; primer 1: 10387–10407; primer 2: 10996–11016, in accordance with the Anderson sequence.

Conditions of PCR reactions were optimised for the purpose of this study. The reaction mixture for both PCR reactions consisted of 1 U Taq DNA polymerase (Fermentas), 1.5 mM MgCl₂, 200 μM dNTP, 0.62 μM in the case of PCR reaction I and 0.31 μM in the case of reaction II of each primer; 0.4 μl BSA, 2.0 μl 10× concentrated PCR buffer and usually 5–10 ng of DNA template in a total reaction volume of 20 μl. Amplification reactions were carried out in the GenAmp 9700 or GenAmp 9600 thermocycler (Perkin Elmer) under the following PCR conditions: 94 °C/1 min; (94 °C/20 s, 58 °C/30 s, 72 °C/40 s) × 32; 72 °C/10 min; 4 °C. Results of amplifications were assessed by electrophoresis of PCR products in 2.6% agarose gels with 8 μl of product input.

Sequencing and statistics

PCR products were purified using the QiaQuick PCR Purification Kit (Qiagen) in accordance with the manufacturer's recommendations. Purified PCR products were subjected to cycle sequencing using the same primers as for amplification and the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). The sequencing mixture consisted of 4 μl of Ready Reaction Premix, 2 μl of 5× Sequencing Buffer, 3.2 pmol of primer and about 30 ng of purified PCR products. Sequencing reactions were carried out in the GenAmp 9700 or GenAmp 9600 thermocycler (Perkin Elmer). The following parameters of the sequencing program were applied: (96 °C/10 s, 50 °C/5 s and 60 °C/2 min) × 25. Products of sequencing reactions were precipitated and suspended in formamide and then subjected to analysis using genetic analyser ABI 3100 with Sequencing Analysis Software 3.0 and Sequence Navigator 2.1 computer programs.

Genetic diversity for the analysed DNA fragment was calculated according to the formula:

$$D = 1 - \sum_i p_i^2 \quad \{1\}$$

where p – frequencies of the observed haplotypes [10].

RESULTS AND DISCUSSION

The basic aim of this work was to assess the degree of variation characterising a selected segment of the coding region of mtDNA. The obtained population data allowed identification of variable nucleotide positions characterised by frequent occurrence of both variants in a population. This feature enables application of such positions as additional markers in the process of human identification. The analysed DNA fragment encompasses nucleotide positions 9995 and 10995 (in accordance with the Anderson reference sequence) and contains genes for two subunits of NADH dehydrogenase (ND3 and ND4L), tRNA for arginine and parts of genes encoding tRNA for glycine and subunit ND4 of NADH dehydrogenase. A population of 50 unrelated individuals from southern Poland was studied. The study enabled identification of 13 different haplotypes and 15 polymorphic nucleotide positions. The most frequent variant (H1) was consistent with the Anderson sequence. In the studied population the frequency of this haplotype equals 56% (Table I).

All substitutions determined during the study are transitions. This fact is consistent with abundant literature data revealing significant domination of transitions over transversions [3, 25]. Six of them are variations type A → G; nine of them are T → C variations. Ten out of 15 identified variations are synonymous substitutions (so-called silent mutations, which do not cause substitutions in the amino acid chain). This result is consistent with the hypothesis of purifying selection, according to which mutations causing amino acid changes are eliminated [14] and is concordant with literature data [8]. All synonymous substitutions occur in the third codon position. Four determined mutations cause changes in the amino acid sequence (Table II), and one variable position is located in the region encoding tRNA for arginine. The study revealed only one heteroplasmic site, present in a single sample in position 10885, which represents mutation T → C.

Significant assistance for the research was provided by Mitomap computer database, which contains information concerning human mtDNA [13]. This database includes data about currently known variable positions, their possible association with genetic diseases, and references to the literature. There is

TABLE I. HAPLOTYPES DETERMINED DURING THE STUDY AND THEIR FREQUENCIES

	10143	10181	10192	10289	10398	10400	10463	10550	10685	10822	10873	10885	10907	10927	10978	Number of individuals
Anderson	G	C	C	A	A	C	T	A	G	C	T	T	T	T	A	
H1	28
H2	G	.	.	G	3
H3	G	.	.	.	A	1
H4	G	6
H5	C	.	2
H6	C	.	.	1
H7	.	T	.	.	G	T	C	C	.	.	.	1
H8	.	.	T	.	G	1
H9	C	.	.	T	1
H10	A	C	2
H11	G	.	.	G	G	1
H12	C	2
H13	.	.	.	G	G	.	.	G	1
Total	50															

also a simple program called Mito Analyzer attached to the database which enables convenient access to information concerning polymorphic positions. It is also an excellent source of information on mitochondrial DNA.

TABLE II. CHARACTERISTIC OF THE VARIABLE POSITIONS IDENTIFIED DURING THE STUDY

Position	Type of mutation	Amino acid change	Frequency	Presence in Mitomap
10143	G A	Gly Ser	4%	Y
10181	C T	Synonymous	2%	Y
10192	C T	Ser Phe	2%	Y
10289	A G	Synonymous	2%	N
10398	A G	Thr Ala	28%	Y
10400	C T	Synonymous	2%	Y
10463	T C	–	10%	Y
10550	A G	Synonymous	10%	Y
10685	G A	Synonymous	2%	Y
10822	C T	Synonymous	2%	N
10873	T C	Synonymous	2%	Y
10885	T C	Synonymous	2%	N
10907	T C	Phe Leu	2%	N
10927	T C	Synonymous	4%	N
10978	A G	Synonymous	2%	N
15 transitions	4 A G 2 G A 5 T C 4 C T	10 synonymous 4 nonsynonymous 1 tRNA	–	9 present 6 new

Detailed analysis showed that six polymorphic sites identified during the present study are positions not observed until now and therefore not mentioned in the Mitomap database (Table II).

The calculated value $D = 0.67 \{1\}$ of the genetic diversity should be understood as high in the context of coding function of the analysed DNA fragment. As was written in the introduction, sequencing of the coding regions of mtDNA is not the optimal approach for identification purposes, because, amongst other reasons, because it could reveal medical information. However, there are no contraindications to analysis of selected polymorphic nucleotide sites that are not involved in genetic diseases. Therefore, of special interest to forensic genetics are synonymous mutations that do not cause changes in the sequence of amino acids (particular mutations in codon position 3 or 1), and neutral mutations. Although the latter cause a change in an encoded polypeptide, none of the protein variants are associated with a disease.

The second important feature of a polymorphic position, determining its usefulness for identification studies, is the frequency of both variants. The optimal level of polymorphism in this context should be characterised by

equal frequency of both variants in a population. Even an approximation of this state, when the rarer variant is characteristic for a significant number of samples, causes such a position to be useful for forensic sciences. Accepting the above criteria, out of 15 identified variable sites, barely 3 are worth studying further, because the others were discovered in single samples. This result confirms data published in the literature, suggesting that most polymorphisms ascertained in coding segments are alterations characteristic for single samples, whereas changes frequently observed in a significant fraction of the population are rare [16, 22].

Polymorphic position 10398 appears to be the most promising for identification studies. This is a substitution, A → G, present in the first codon position causing the change 114 Thr → Ala in the ND3 subunit of NADH dehydrogenase. No association of this mutation with a disease can be found in the literature. The frequent representation of both variants in the studied population (variant A – 72%, variant G – 28%) also suggests that this mutation is neutral (Table II). Additional analysis of 9 samples taken from unrelated African donors showed that variant G is very frequent (it was identified in 5 samples). Position 10463 (substitution T → C) is located in a gene encoding tRNA for arginine. Furthermore, in the case of this mutation, there is no data suggesting association of any of these haplotypes with a disease. Clinical research indicates [20] that the rarer variant C is characteristic for so-called haplogroup T, in accordance with classification proposed by Torroni et al. [21]. It is known that this haplogroup is fairly frequent in some regions of the world and, for instance, can be found in 22% of Swedes and 12% of Germans [6, 21]. The present study indicates, that in the Polish population it is characteristic for 10% of people. Synonymous transition A → G (10550) in the third codon position is located in the ND4L subunit of NADH dehydrogenase. The rarer variant G is characteristic for 10% of the population.

It seems that the three nucleotide positions described above can in future serve as a good source of polymorphism and be used in forensic identification studies.

SUMMARY

Sequence analysis of the 1001 bp fragment of mtDNA located in the coding region conducted on a population of 50 unrelated individuals enabled identification of 13 haplotypes and 15 polymorphic positions, the most frequent variant being concordant with the Anderson sequence. All identified substitutions represent transitions and most of them are synonymous muta-

tions. Three polymorphic sites reveal features which seem to make them suitable for analysis in forensic identification investigations.

References:

1. Anderson S., Bankier A. T., Barrel B. G. [et al.], Sequence and organization of the human mitochondrial genome, *Nature* 1981, vol. 290, pp. 457–465.
2. Boore J. L., Animal mitochondrial genomes, *Nucleic Acids Research* 1999, vol. 27, pp. 1767–1780.
3. Brown W. M., Prager E. M., Wang A. [et al.], Mitochondrial DNA sequences of primates: tempo and mode of evolution, *Journal of Molecular Evolution* 1982, vol. 18, pp. 225–239.
4. Budowle B., Wilson M. R., diZinno J. A. [et al.], Mitochondrial DNA regions HVI and HVII population data, *Forensic Science International* 1999, vol. 103, pp. 23–35.
5. Cooper A., Lalueza-Fox C., Anderson S. [et al.], Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution, *Nature* 2001, vol. 409, pp. 704–707.
6. Hofmann S., Jaksch M., Bezold R. [et al.], Population genetics and disease susceptibility: characterization of central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease, *Human Molecular Genetics* 1997, vol. 6, pp. 1835–1846.
7. Holland M. M., Parsons T. J., Mitochondrial DNA sequence analysis – validation and use for forensic casework, *Forensic Science Reviews* 1999, vol. 11, pp. 21–48.
8. Ingman M., Gyllensten U., Analysis of the complete human mtDNA genome: methodology and inferences for human evolution, *Journal of Heredity* 2001, vol. 92, pp. 454–461.
9. Ingman M., Kaessmann H., Paabo S. [et al.], Mitochondrial genome variation and the origin of modern humans, *Nature* 2000, vol. 408, pp. 708–713.
10. Jones D. A., Blood samples: probability of discrimination, *Journal of Forensic Science Society* 1972, vol. 12, pp. 355–358.
11. Kwok P. Y., Methods for genotyping Single Nucleotide Polymorphisms, *Annual Reviews of Genomics and Human Genetics* 2001, vol. 2, pp. 235–258.
12. Lutz S., Wittig H., Weisser H. J. [et al.], Is it possible to differentiate mtDNA by means of HVIII in samples that cannot be distinguished by sequencing the HVI and HVII regions?, *Forensic Science International* 2000, vol. 113, pp. 97–101.
13. Mitomap: A Human Mitochondrial Genome Database, <http://www.mitomap.org>.
14. Nei M., Kumar S., Molecular evolution and phylogenetics, Oxford University Press, London 2000.
15. Parson W., Parsons T. J., Scheithauer R. [et al.], Population data for 101 Austrian Caucasian mitochondrial DNA d-loop sequences: application of

- mtDNA sequence analysis to a forensic case, *International Journal of Legal Medicine* 1998, vol. 111, pp. 124–132.
16. Parsons T. J., Coble M., Increasing the forensic discrimination of mitochondrial DNA testing through analysis of entire mitochondrial genome, *Croatian Medical Journal* 2001, vol. 42, pp. 304–309.
 17. Piercy R., Sullivan K. M., Benson N. [et al.], The application of mitochondrial DNA typing to the study of white Caucasian genetic identification, *International Journal of Legal Medicine* 1993, vol. 106, pp. 85–90.
 18. Rand D., Mitochondrial genomics flies high, *Trends in Ecology and Evolution* 2001, vol. 16, pp. 2–4.
 19. Rousselet F., Mangin P., Mitochondrial DNA polymorphisms: a study of 50 French Caucasian individuals and application to forensic casework, *International Journal of Legal Medicine* 1998, vol. 111, pp. 292–298.
 20. Ruiz-Pesini E., Lapena A. C., Diez-Sanchez C. [et al.], Human mtDNA haplogroups associated with high or reduced spermatozoa motility, *American Journal of Human Genetics* 2000, vol. 67, pp. 682–696.
 21. Torroni A., Huoponen K., Francalacci P. [et al.], Classification of European mtDNAs from an analysis of three European populations, *Genetics* 1996, vol. 144, pp. 1835–1850.
 22. Tzen C., Wu T., Liu H., Sequence polymorphism in the coding region of mitochondrial genome encompassing position 8389–8865, *Forensic Science International* 2001, vol. 120, pp. 204–209.
 23. Wallace D. C., Mitochondrial diseases in man and mouse, *Science* 1999, vol. 283, pp. 1482–1488.
 24. Wallace D. C., Brown M. D., Lott M. T., Mitochondrial DNA variation in human evolution and disease, *Gene* 1999, vol. 238, pp. 211–230.
 25. Yang Z., Yoder A. D., Estimation of the transition/transversion rate bias and species sampling, *Journal of Molecular Evolution* 1999, vol. 48, pp. 274–283.

ZMIENNOŚĆ GENETYCZNA FRAGMENTU REGIONU KODUJĄCEGO MITOCHONDRIALNEGO DNA ZAWARTEGO MIĘDZY POZYCJAMI NUKLEOTYDOWYMI 9995–10995

Wojciech BRANICKI, Beata ZASADZKA, Tomasz KUPIEC

WSTĘP

Sekwencjonowanie wysoce polimorficznych fragmentów regionu kontrolnego mitochondrialnego DNA (mtDNA) jest dziś rutynową metodą analizy śladów biologicznych, które ze względu na niewystarczające stężenie jądrowego DNA lub zbyt silne procesy degradacji nie nadają się do badań w oparciu o markery typu STR. Fakt wysokiej czułości analizy markerów mtDNA wynika głównie z tego, iż genom mitochondrialny reprezentowany jest w setkach, a nawet tysiącach kopii w każdej z komórek organizmu [7]. W praktyce analizie sekwencji mtDNA poddawane są głównie trzony włosów oraz kości; te ostatnie badane przede wszystkim w sprawach identyfikacyjnych. Do istotnych zalet analizy mtDNA należy fakt jego dziedziczenia w linii matczynej. Taki mechanizm dziedziczenia sprawia, iż identyfikacja szczątków może być przeprowadzona nawet w sytuacji, gdy materiał porównawczy możliwy jest do uzyskania jedynie od dalekich krewnych identyfikowanej osoby. Szybkie tempo ewolucji charakterystyczne dla genomu mitochondrialnego, które – jak się szacuje – jest 5 do 10 razy wyższe niż w DNA jądrowym [3], decyduje o wysokim stopniu polimorfizmu zwłaszcza w regionach wysoce zmiennych, tzw. HV1 i HV2 (ang. hypervariable 1, 2). Mechanizm dziedziczenia mtDNA połączony z brakiem procesów rekombinacyjnych powoduje, iż wszyscy potomkowie jednej matki posiadają jeden i ten sam typ mtDNA. Niektóre z haplotypów mogą być przez to stosunkowo często obserwowane w populacji i samo szybkie tempo mutacji nie jest w stanie wyeliminować istotnej wady towarzyszącej analizie mtDNA polegającej na stosunkowo niskiej sile dyskryminacji układu. Dane populacyjne uzyskane dla próbek z różnych regionów Europy i Stanów Zjednoczonych pokazują, iż najczęstszy haplotyp mtDNA stwierdzony dla rasy białej (263G, 315.1C) występuje z wysoką częstością ok. 4% [4, 15, 17, 19]. Częstość tego haplotypu w populacji Polski południowej ma wartość 5,3% (nieopublikowane badania własne). Cały genom mitochondrialny należy traktować z punktu widzenia badań identyfikacyjnych jako pojedynczy marker genetyczny, a zatem analiza sekwencji kolejnych fragmentów w obrębie mtDNA nie spowoduje wzrostu siły dyskryminacji w stopniu analogicznym jak przy badaniu kolejnych układów typu STR. Mimo to zaproponowano podejście zmierzające do ograniczenia niekorzystnego aspektu analizy mtDNA związanego z niską siłą dyskryminacji poprzez badanie dodatkowych regionów tego genomu. Jedną z możliwości jest analiza kolejnego obszaru w obrębie regionu kontrolnego nazwanego przez analogię do pozostałych analizowanych regionów wysoce polimorficznych – HV3 [12]. Zmienność w tym odcinku mtDNA jest wprawdzie nieco niższa niż w przypadku regionów HV1 i HV2, ale jak wynika z danych cytowanych w literaturze przedmiotu oraz z praktyki Instytutu Ekspertyz Sądowych, może okazać się bardzo

pomocna w przypadku konieczności różnicowania osobników posiadających identyczne haplotypy HV1 i HV2. Badania przeprowadzone na populacji osób zamieszkujących Polskę południową wskazują, iż dzięki zastosowaniu analizy polimorfizmu regionu HV3 można zróżnicować ok. 30% próbek identycznych pod względem sekwencji HV1 i HV2 (zmiennosc genetyczna $D = 0,68$) (nieopublikowane badania własne). Jak stwierdzono, również analiza polimorfizmu cechującego regiony kodujące mtDNA umożliwia zróżnicowanie osobników posiadających identyczne haplotypy HV1 i HV2 [16]. Jak wiadomo, sekwencja kodująca stanowi większą część (93%) genomu mitochondrialnego. W mtDNA jest zapisana informacja dotycząca 13 peptydów (podjednostek enzymów zaangażowanych w łańcuch fosforylacji oksydacyjnej i produkcję ATP), dwóch rodzajów rRNA (12S oraz 16S rRNA) oraz 22 typów tRNA niezbędnych w procesie translacji białek kodowanych przez mtDNA (łącznie 37 genów) [2]. Nie należy oczekiwać, iż stopień polimorfizmu regionów kodujących będzie równie wysoki, jak ma to miejsce w przypadku wysoce polimorficznych segmentów regionu kontrolnego. Jednak liczba danych zawarta w DNA o długości ok. 15,5 tysiąca par zasad sprawia, iż sięgnięcie do polimorfizmów tkwiących w regionach kodujących wydaje się być niezwykle obiecujące. W literaturze przedmiotu odnaleźć można dużo informacji na temat związków wielu chorób uwarunkowanych genetycznie z poszczególnymi mutacjami w genomie mitochondrialnym [23, 24]. Dane o takim charakterze nie mogą być uzyskiwane w laboratorium sądowym. Istotnym problemem jest zatem konieczność wyeliminowania analiz tych pozycji, w których mutacje są odpowiedzialne za determinowanie fenotypów chorobowych. Niestety zastosowanie techniki sekwencjonowania uniemożliwia ominięcie pozycji niosących informację medyczną, a jednocześnie znajdujących się tuż obok polimorfizmów cennych z punktu widzenia identyfikacji osobniczej. Poza tym istnieje drugie istotne zastrzeżenie co do metody sekwencjonowania w kontekście rutynowej analizy regionów kodujących mtDNA. Sama technika sekwencjonowania, mimo znacznego stopnia automatyzacji, jest metodą żmudną i czasochłonną. Wprawdzie coraz częściej pojawiają się publikacje naukowe na temat sekwencjonowania całych genomów mitochondrialnych również w aspekcie badań populacyjnych [5, 9, 18], ale trudno sobie wyobrazić, aby tego typu podejście było praktyczne z punktu widzenia genetyki sądowej. Obiecujące wydaje się zastosowanie analizy wybranych pojedynczych pozycji polimorficznych (ang. SNP) cennych dla celów identyfikacyjnych.

Obecnie istnieje bardzo wiele metod umożliwiających wybiórczą analizę pojedynczych nukleotydów polimorficznych. Rozważania dotyczące wyboru najbardziej odpowiedniej metody wykraczają poza ramy niniejszej pracy, a ich szczegółowe omówienie można odnaleźć w pracy przeglądowej poświęconej w całości temu zagadnieniu [11]. Zanim poszczególne pozycje mtDNA będą mogły być włączone do standardowej analizy, konieczne jest odnalezienie polimorficznych pozycji, które byłyby użyteczne z punktu widzenia genetyki sądowej, a do tego typu badań technika sekwencjonowania pozostaje metodą niezastąpioną. W literaturze przedmiotu odnaleźć można pierwsze doniesienia na temat poszukiwania użytecznych z punktu widzenia identyfikacyjnych badań sądowych pojedynczych pozycji polimorficznych w wybranych regionach kodujących mtDNA [22].

Celem niniejszych badań była analiza sekwencji fragmentu regionu kodującego mtDNA prowadząca do oznaczenia jego stopnia zmienności oraz odnalezienie pozycji

polimorficznych spełniających warunki umożliwiające ich zastosowanie w przyszłości do celów identyfikacyjnych.

MATERIAŁY I METODY

Material do badań

Badaniu poddano 50 próbek krwi lub śliny pochodzących od niespokrewnionych osób zamieszkujących Polskę południową. Izolację DNA przeprowadzono standardową metodą organiczną z dwukrotną ekstrakcją mieszaniną fenol-chloroform-alkohol izoamylowy. Stężenie DNA określano metodą fluorymetryczną z zastosowaniem zestawu PicoGreen dsDNA Quantitation Kit (Molecular Probes) i aparatu Fluoroscanner Ascent FL (Labsystems).

Amplifikacja

Amplifikacji poddano fragment regionu kodującego mitochondrialnego DNA zawarty między 9995 a 10995 parą zasad, zgodnie z sekwencją odniesienia Andersona [1]. Segment o długości 1001 par zasad amplifikowano w dwóch zachodzących na siebie fragmentach z zastosowaniem zaprojektowanych dla potrzeb niniejszych badań starterów reakcji PCR:

- fragment I o długości 469 pz zawarty między pozycjami 9995–10463; starter 1: 9974–9994; starter 2: 10464–10484, zgodnie z sekwencją Andersona;
- fragment II o długości 588 pz zawarty między pozycjami 10408–10995; starter 1: 10387–10407; starter 2: 10996–11016, zgodnie z sekwencją Andersona.

Warunki reakcji PCR zoptymalizowano dla potrzeb niniejszych badań. Mieszanina reakcyjna składała się w przypadku obu prowadzonych reakcji PCR z 1 U Taq DNA polimerazy (Fermentas), 1,5 mM MgCl₂, 200 μM dNTP, 0,62 μM w przypadku reakcji I i 0,31 μM w przypadku reakcji II każdego ze starterów; 0,4 μg BSA, 2,0 μl 10-stężonego buforu do PCR i zazwyczaj 5–10 ng matrycy DNA w całkowitej objętości 20 μl. Amplifikację prowadzono w termocyklerach GenAmp 9700 lub GenAmp 9600 firmy Perkin Elmer stosując następujące warunki: 94°C/1min; (94°C/20 s, 58°C/30 s, 72°C/40 s) ³²; 72°C/10 min; 4°C. Wyniki amplifikacji oceniano poprzez elektroforezę produktów PCR w 2,6 % żelu agarozowym, nanosząc 8 μl produktu.

Sekwencjonowanie i analiza statystyczna

Produkty PCR poddawano oczyszczaniu, używając zestawu QiaQuick PCR Purification Kit (Qiagen) i stosując warunki podane przez jego producenta. Oczyszczone produkty poddawane były cyklicznemu sekwencjonowaniu z zastosowaniem starterów wykorzystywanych do amplifikacji i zestawu BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Mieszanina sekwencyjna składała się z 4 μl Ready Reaction Premix, 2 μl 5' Sequencing Buffer, 3,2 pmol startera reakcji PCR i około 30 ng oczyszczonych produktów PCR. Reakcje sekwencjonowania prowadzono w termocyklerach GenAmp 9700 lub GenAmp 9600 firmy Perkin Elmer. Stosowano następujące parametry programu: (96°C/10 s, 50°C/5 s i 60°C/2 min) ²⁵. Strącone i rozpuszczone w formamidzie produkty reakcji sekwencjonowania analizo-

wano z zastosowaniem automatycznego analizatora genetycznego ABI 3100 wraz z oprogramowaniem Sequencing Analysis Software 3.0 i Sequence Navigator 2.1.

Zmienność genetyczną dla analizowanego fragmentu obliczono, stosując wzór:

$$D = 1 - \sum_i p_i^2 \quad \{1\}$$

gdzie p – frekwencje obserwowanych haplotypów [10].

WYNIKI I DYSKUSJA

Podstawowym celem niniejszej pracy było oszacowanie stopnia zmienności charakteryzującego wybrany fragment regionu kodującego mtDNA. Uzyskane dane populacyjne pozwoliły na zidentyfikowanie zmiennych pozycji nukleotydowych charakteryzujących się częstym występowaniem w populacji obu wariantów, co w efekcie daje możliwość ich zastosowania jako dodatkowych markerów w procesie identyfikacji człowieka. Badany segment jest zawarty między 9995 a 10995 parą zasad (zgodnie z referencyjną sekwencją Andersona) i zawiera geny dla dwóch podjednostek dehydrogenazy NADH (ND3 i ND4L), tRNA argininy oraz fragmenty genów tRNA glicyny i podjednostki ND4 dehydrogenazy NADH. Badaniu poddano próbkę populacyjną 50 niespokrewnionych osób z regionu Polski południowej. Badania pozwoliły zidentyfikować w badanej grupie 13 różnych haplotypów i 15 polimorficznych pozycji nukleotydowych. Najczęściej obserwowany wariant (H1) był zgodny z sekwencją Andersona. W badanej populacji haplotyp ten występował z częstością 56% (tabela I).

Wszystkie stwierdzone podczas badań substytucje należą do zmian typu tranzycji. Jest to zgodne z licznymi danymi zawartymi w literaturze, a wykazującymi znaczącą przewagę tranzycji nad transwersjami [3, 25]. Sześć z nich to zmiany typu A → G, 9 to zmiany typu T → C. Dziesięć z 15 zidentyfikowanych zmian to substytucje synonimiczne (tzw. mutacje ciche, nie powodujące zmian w łańcuchu aminokwasowym). Taki rezultat zgodny jest z hipotezą działania doboru oczyszczającego (ang. purifying selection), który eliminuje mutacje prowadzące do zmian w sekwencji aminokwasów [14] i pozostaje w zgodzie z danymi zawartymi w literaturze przedmiotu [8]. Wszystkie substytucje synonimiczne występują w trzeciej pozycji kodonu. Cztery stwierdzone mutacje powodują zmiany w łańcuchu aminokwasowym (tabela II), a jedna pozycja zmienna znajduje się w regionie kodującym tRNA argininy. Podczas badań ujawniono zaledwie jedną pozycję heteroplazmatyczną obecną w pojedynczej próbce – pozycja 10885, mutacja T → C.

Narzędziem bardzo użytecznym z punktu widzenia prowadzonych analiz okazała się komputerowa baza danych Mitomap gromadząca materiały dotyczące mtDNA człowieka [13]. Baza zawiera informacje o wykrytych dotychczas pozycjach zmiennych, ich ewentualnym związku z chorobami genetycznymi oraz odniesienia do literatury przedmiotu. Udostępnia ona również prosty program o nazwie Mito Analyzer służący do wygodnego uzyskiwania informacji dotyczących pozycji polimorficznych. Baza stanowi również doskonałe źródło wiadomości na temat mitochondrialnego DNA.

Szczegółowa analiza wykazała, iż sześć pozycji zmiennych zidentyfikowanych podczas niniejszych badań to pozycje dotychczas nie obserwowane i nie odnotowane w bazie Mitomap (tabela II).

Uzyskaną wartość $D = 0,67$ zmienności genetycznej należy uznać za wysoką w kontekście kodującej funkcji analizowanego fragmentu. Zgodnie z tym, co napisano we wstępie, sekwencjonowanie regionów kodujących mtDNA nie jest podejściem optymalnym dla celów identyfikacyjnych, między innymi dlatego, iż może ujawniać informacje o charakterze medycznym. Nie ma jednak żadnych przeciwwskazań, ażeby analizie poddawać tylko wybrane polimorficzne pozycje nukleotydowe, które nie niosą takich informacji. Zatem w polu zainteresowania genetyki sądowej pozostają mutacje synonimiczne, które nie powodują zmian w sekwencji aminokwasów (wybrane mutacje w 3 lub 1 pozycji kodonu) oraz mutacje neutralne. Te ostatnie wywołują wprawdzie zmianę w kodowanym polipeptydzie, ale żaden z wariantów białka nie wiąże się ze stanem chorobowym.

Drugą istotną cechą pozycji polimorficznej stanowiącą o jej użyteczności w badaniach identyfikacyjnych jest częstość obu wariantów. Najbardziej optymalny poziom polimorfizmu powinien charakteryzować się równą reprezentacją obu wariantów w populacji. Już zbliżenie się do tego stanu, a więc sytuacja, w której rzadszy wariant jest charakterystyczny dla znaczącego odsetka badanych, sprawia, iż pozycja taka może okazać się przydatna dla nauk sądowych. Przyjmując powyższe kryteria, z 15 zidentyfikowanych pozycji zmiennych zaledwie 3 warte są bliższej analizy, gdyż pozostałe zostały odnalezione w pojedynczych próbkach. Taki rezultat potwierdza dane opublikowane w specjalistycznej literaturze, z których wynika, iż większość polimorfizmów stwierdzonych w regionach kodujących to zmiany charakterystyczne dla pojedynczych próbek, a zmiany reprezentowane przez znaczący odsetek populacji należą do nielicznych [16, 22].

Najbardziej obiecująca z punktu widzenia użyteczności dla badań identyfikacyjnych wydaje się polimorficzna pozycja 10398. Jest to substytucja typu A → G w pierwszej pozycji kodonu powodująca zmianę 114 Thr → Ala w podjednostce ND3 dehydrogenazy NADH. W literaturze przedmiotu nie ma doniesień o związku tej mutacji z jakimkolwiek stanem chorobowym. Za neutralnością mutacji przemawia również fakt częstej reprezentacji obu wariantów w badanej populacji: wariant A – 72%, wariant G – 28% (tabela II). Dodatkowa analiza 9 próbek pobranych od niespokrewnionych dawców rasy afrykańskiej wykazała, iż wariant G jest bardzo częsty występuje u 5 osobników. Pozycja 10463 (substytucja T → C) znajduje się w genie kodującym tRNA argininy. Również w przypadku tej mutacji nie ma danych wskazujących na związek któregoś z haplotypów ze stanem chorobowym. Badania kliniczne wskazują [20], iż rzadszy wariant C jest charakterystyczny dla tzw. haplogrupy T zgodnie z klasyfikacją zaproponowaną przez Torroni i in. [21]. Jak wiadomo, haplogrupa ta należy do dosyć częstych w niektórych regionach świata, na przykład jest reprezentowana przez 22% Szwedów i 12% Niemców [6, 21]. Niniejsze badania wskazują, iż w populacji polskiej jest charakterystyczna dla 10% osobników. Tranzycja synonimiczna typu A → G (10550) w trzeciej pozycji kodonu ma miejsce w podjednostce ND4L dehydrogenazy NADH. Rzadszy wariant G jest charakterystyczny dla 10% populacji.

Wydaje się, iż scharakteryzowane powyżej trzy pozycje nukleotydowe mogą w przyszłości znaleźć zastosowanie jako źródło polimorfizmu i być wykorzystywane w sądowych badaniach identyfikacyjnych.

PODSUMOWANIE

Analiza sekwencji fragmentu mtDNA o długości 1001 pz położonego w obszarze kodującym przeprowadzona na grupie 50 niespokrewnionych osób pozwoliła na oznaczenie 13 haplotypów oraz 15 pozycji polimorficznych, przy czym najczęściej obserwowany haplotyp był zgodny z sekwencją Andersona. Wszystkie zidentyfikowane substytucje należą do tranzycji, a większość z nich to mutacje typu synonimicznego. Trzy pozycje polimorficzne mają cechy, które – jak się wydaje – pozwalają na zastosowanie ich analizy do celów identyfikacyjnych w badaniach sądowych.