

Patrycja Daca, Marta Mielnik-Sikorska, Jarosław Bednarek, Tomasz Grzybowski

## Ocena stopnia wysycenia bazy danych mitochondrialnego DNA dla populacji Polski\*

### Saturation of the Polish mitochondrial DNA database

Z Zakładu Genetyki Molekularnej i Sądowej  
Katedry Medycyny Sądowej Uniwersytetu Mikołaja Kopernika w Toruniu,  
Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy  
Kierownik Katedry: dr hab. med. T. Grzybowski, prof. UMK

Celem niniejszej pracy była statystyczna ocena stopnia wysycenia polskiej populacyjnej bazy danych, liczącej 1253 osoby niespokrewnione w linii matczynej, z siedmiu subpopulacji Polski. W ramach oceny przeprowadzono analizę czterech parametrów: zróżnicowania nukleotydowego, haplotypowego oraz liczby miejsc polimorficznych i liczby różnych haplotypów obserwowanych w populacji. Uzyskane wyniki pozwalają stwierdzić, iż polska populacyjna baza danych mtDNA osiągnęła poziom wysycenia w przypadku tylko dwóch z analizowanych parametrów (zróżnicowania haplotypowego oraz nukleotydowego), natomiast poziom wysycenia dla dwóch pozostałych zmiennych – liczby różnych haplotypów oraz miejsc polimorficznych, podobnie jak w przypadku innych baz profili mtDNA w różnych krajach, nie został osiągnięty. Wskazuje to na potrzebę dalszego poszerzania zasobów omawianej bazy.

The main objective of the study was a statistical evaluation of Polish mitochondrial DNA database, consisting of 1253 maternally unrelated individuals from seven different regions of Poland. Four relevant parameters were examined: haplotype diversity, nucleotide diversity, number of polymorphic positions and number of haplotypes, including HVS I and HVS II mtDNA regions. The results show that while haplotype and nucleotide diversities have already reached saturation level, the number of haplotypes and polymorphic positions rises continuously inside the database. These results indicate a need for a substantial increase in the number of haplotypes in Polish mitochondrial DNA database.

#### Słowa kluczowe:

mitochondrialny DNA, filogenetyka, zróżnicowanie haplotypowe i nukleotydowe, liczba miejsc polimorficznych, liczba haplotypów, wysycenie bazy danych

#### Key words:

mitochondrial DNA, phylogenetics, haplotype diversity, nucleotide diversity, number of polymorphic positions, number of haplotypes, database saturation

## WSTĘP

Za sprawą dużej liczby cząsteczek w komórce oraz wysokiej odporności na degradację, mitochondrialny DNA (mtDNA) stał się szczególnie użytecznym markerem stosowanym w badaniach śladów biologicznych zawierających zdegradowany materiał genetyczny, a także w identyfikacji ofiar przestępstw lub katastrof [1, 2]. Z punktu widzenia zastosowań polimorfizmu mtDNA dla potrzeb wymiaru sprawiedliwości i organów ścigania, bardzo ważna jest wiarygodna kwalifikacja statystyczna wyników oparta na wykorzystaniu referencyjnych baz haplotypów mtDNA. Niezwykle istotnym czynnikiem jest tutaj wielkość bazy danych – a więc liczba zgromadzonych w niej próbek, ale także reprezentatywność, będąca pochodną wysycenia bazy haplotypami mtDNA obserwowanymi w populacjach [3, 4, 5]. Obserwowana w cząsteczce mtDNA wysoka zmienność sekwencji, spowodowana przede wszystkim wysokim tempem mutacji regionu kontrolnego, ale także działaniem dryfu genetycznego,

\* Poszerzona wersja referatu, przedstawionego podczas XV Zjazdu Naukowego PTMSiK, Gdańsk 16-18.09.2010.

prowadzi do powstania bardzo dużej ilości kombinacji haplotypów obserwowanych w obrębie danej populacji, która przekłada się na wysoką heterozygotyczność wewnątrz populacji. W przypadku, gdy baza danych nie reprezentuje pełnego zestawu profili pochodzących z różnych obszarów geograficznych, może dojść do mylnej interpretacji statystycznej wyników. Taka sytuacja ma miejsce np. w przypadku podejrzanego, którego haplotyp jest częsty na obszarze jego zamieszkania, a nie ma odpowiedniej reprezentatywności w bazie profili DNA [5, 6, 7]. Przeprowadzone na przestrzeni kilku lat przez różne zespoły analizy parametrów opisujących stopień wysycenia światowych baz danych, takich jak: liczba haplotypów i ich zróżnicowanie, a także ilość pozycji polimorficznych oraz zróżnicowanie nukleotydowe, jednoznacznie wskazały na brak pełnego zakresu haplotypów mtDNA obserwowanych w bazach danych [6, 7, 8, 9]. W celu oszacowania stopnia wysycenia polskiej populacyjnej bazy profili mitochondrialnego DNA, w ramach niniejszego opracowania przeprowadzono statystyczną analizę wyżej wymienionych parametrów.

## MATERIAŁ I METODY

Materiał badawczy stanowiły profile regionu kontrolnego mtDNA pochodzące od 1253 osób niespokrewnionych ze sobą w linii żeńskiej, z siedmiu subpopulacji Polski (Kaszub, Podhala, Pomorza i Kujaw, Pomorza Gdańskiego, Suwalszczyzny, Górnego Śląska oraz Kościerzyny) [10, 11, 12]. Statystyczną ocenę stopnia wysycenia polskiej bazy profili mitochondrialnego DNA przeprowadzono na poziomie zmienności sekwencji regionu kontrolnego (HVS I w zakresie sekwencjonowania 15999-16400 p.z. oraz HVS II 30-407 p.z.) w oparciu o serię losowań ze zwracaniem o nominatach: 200, 400, 600, 800, 1000 oraz 1253 z puli całej populacji, zgodnie z metodyką przedstawioną przez Pereira i wsp. [6]. Zastosowanie niezależnych losowań ze zwracaniem nie zmieniało struktury genetycznej analizowanej bazy danych, a tym samym było gwarancją zachowania w pełni losowego charakteru każdej próbki. W celu ekstrapolowania poziomu wysycenia bazy, każdą grupę danych generowano oddzielnie dla regionu HVS I, HVS II oraz łącznie HVS I i HVS II, a następnie badano pod kątem zróżnicowania szeregu parametrów: liczby haplotypów i zróżnicowania

haplotypowego, ilości pozycji polimorficznych oraz zróżnicowania nukleotydowego. Wartości wymienionych parametrów wyznaczano dla każdej próbki z wykorzystaniem programu *Arlequin v.3.1* [13], a uzyskane dane zobrazowano za pomocą wykresów opracowanych w pakiecie Excel (Microsoft).

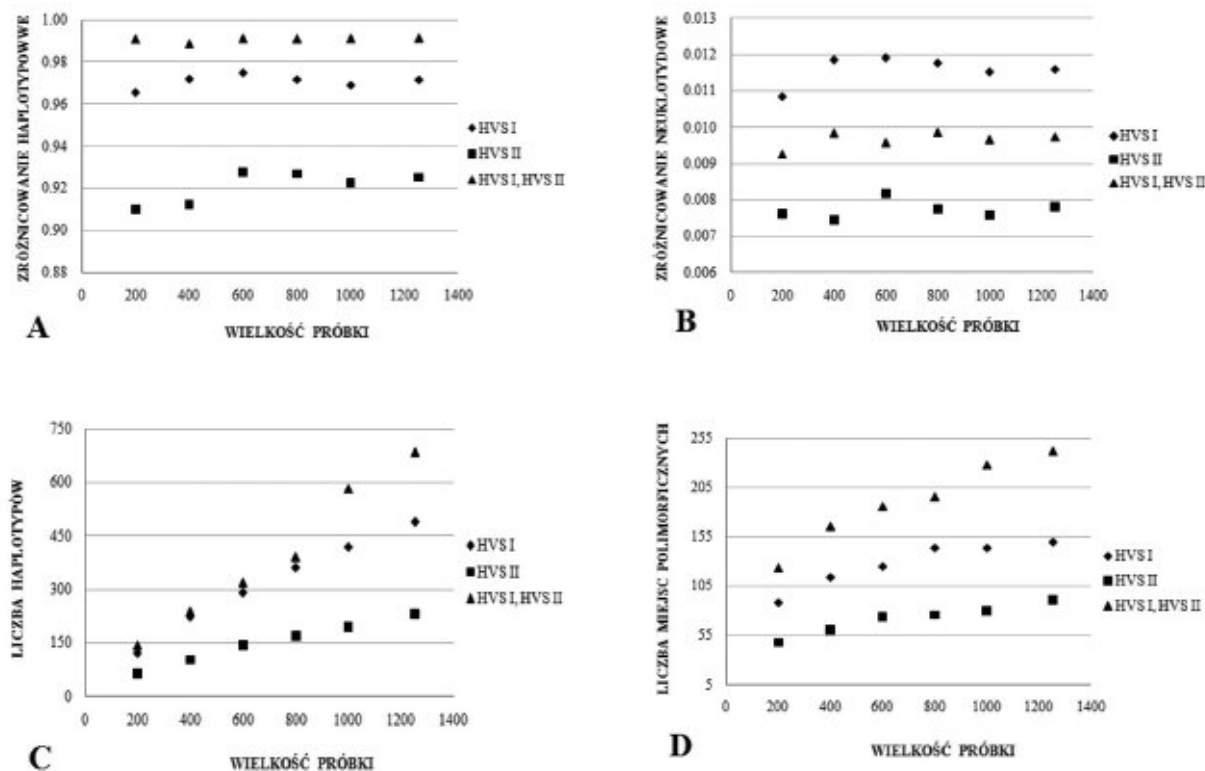
## WYNIKI I DYSKUSJA

Rycina 1 przedstawia wykresy obrazujące zależność pomiędzy wielkością polskiej populacyjnej bazy danych (uzyskaną w wyniku serii losowań ze zwracaniem), a wartościami poszczególnych analizowanych parametrów (liczby haplotypów i zróżnicowania haplotypowego, ilości pozycji polimorficznych oraz zróżnicowania nukleotydowego) wyznaczonych dla regionów HVS I i HVS II oddzielnie oraz łącznie.

W przypadku zróżnicowania haplotypowego (ryc.1A) oraz nukleotydowego (ryc.1B) na poziomie zmienności regionu HVS I oraz HVS II mtDNA rozpatrywanych oddzielnie widoczne jest, że wraz ze wzrostem wielkości populacyjnej bazy danych rośnie również wartość obu parametrów, by przy liczności równej 600 osiągnąć poziom wysycenia (*plateau*).

Z kolei gdy rozważany jest parametr zróżnicowania haplotypowego (ryc. 1A) na poziomie obu fragmentów regionu kontrolnego (HVS I oraz HVS II łącznie), widoczne jest niewielkie wahanie omawianego parametru w granicach 0.989-0.993, bez wyraźnego maksimum. Podobną tendencję zanotowano również w przypadku zróżnicowania nukleotydowego (ryc. 1B), z tą różnicą, że niezależnie od tzw. liczności próbkowania, parametr ten osiąga bardzo niewielkie i zbliżone do siebie wartości z przedziału pomiędzy 0.0093 a 0.0099. Nie przekracza tym samym wartości wyznaczonych dla regionu HVS I, zajmując na wykresie obszar pomiędzy regionem HVS I oraz HVS II.

Wykresy ilustrujące liczbę różnych haplotypów obserwowanych w polskiej populacyjnej bazie danych (ryc. 1C) oraz miejsc polimorficznych (ryc. 1D) odnotowanych w obrębie poszczególnych grup losowych (200, 400, 600, 800, 1000 oraz całej bazy) mają bardzo podobny przebieg i wyraźnie wskazują na jednostajny wzrost omawianych parametrów wraz z licznością bazy. W przypadku obu



Ryc. 1. Wykresy zależności pomiędzy wielkością polskiej populacyjnej bazy danych, a wartościami analizowanych parametrów: A- zróżnicowania haplotypowego, B- zróżnicowania nukleotydowego, C- liczbą haplotypów, D- ilością pozycji polimorficznych wyznaczonych dla regionu HVS I oraz HVS II oddzielnie oraz łącznie.

Fig. 1. Correlations between sample size and A: - haplotype diversity, B- nucleotide diversity, C- number of haplotypes, D- number of polymorphic positions, for mtDNA HVS I and HVS II separately and for both regions (HVS I and HVS II).

parametrów nie obserwuje się fazy *plateau*, a ilość różnych haplotypów oraz obserwowanych w populacji miejsc polimorficznych osiąga najwyższe wartości, gdy brany jest pod uwagę cały region kontrolny mtDNA (HVS I oraz HVS II łącznie).

Z przedstawionych powyżej wykresów wynika, że o ile dwa z analizowanych parametrów, a więc zróżnicowanie haplotypowe oraz nukleotydowe, osiągają poziom wysycenia już przy liczebności próbki równej 600 (wartość ta odnosi się do wszystkich analizowanych fragmentów mitochondrialnego DNA), to w przypadku dwóch pozostałych zmiennych, a więc liczby różnych haplotypów obserwowanych w populacji oraz miejsc polimorficznych, poziom *plateau* nie zostaje osiągnięty w żadnej z grup losowych.

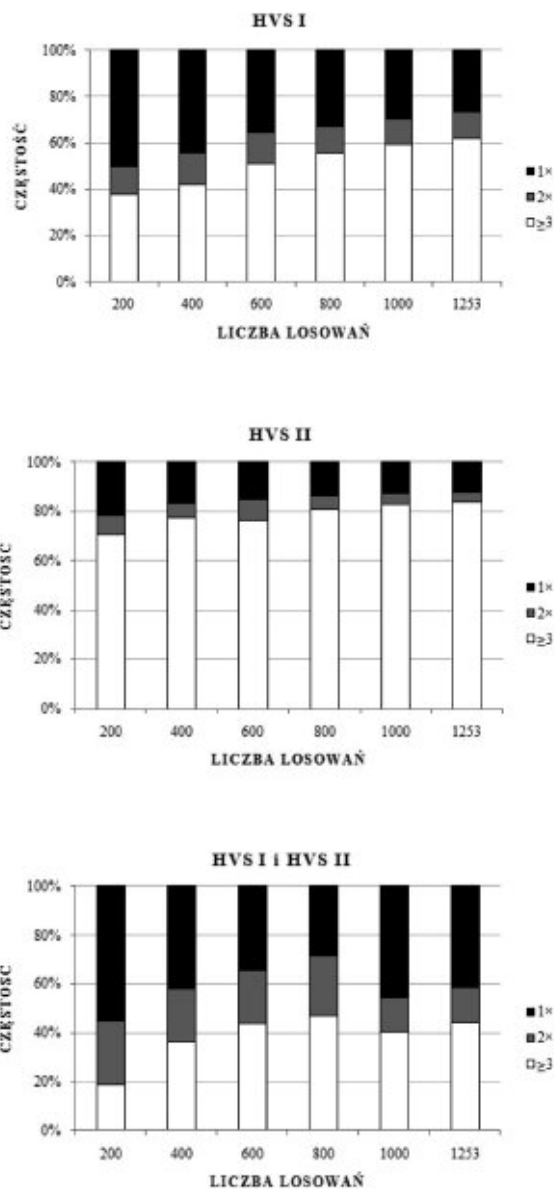
Podobne analizy stopnia wysycenia populacyjnych baz danych przeprowadzone wcześniej dla portugalskiej bazy danych liczącej 549 osób oraz dla społeczności niemieckiej ( $n = 1200$  osób) wskazały na podobną tendencję [6, 9]. Analizując na poziomie HVS I zróżnicowanie nukleotydowe oraz liczbę różnych haplotypów pojawiających się w bazie, Pfeiffer i wsp. (2001) odnotowali, iż podobnie jak to miało miejsce w polskiej bazie danych, liczba haplotypów rosła w sposób liniowy wraz ze wzrostem liczby próbek nie osiągając poziomu *plateau* [9]. Natomiast wysycenie parametru zróżnicowania nukleotydowego (do poziomu 0.981) nastąpiło nieco wcześniej, bo przy liczebności próbki równej 400. Z kolei w portugalskiej

bazie danych poziom wysycenia omawianego parametru został osiągnięty już przy wartości 300 osób [6]. Odnotowane rozbieżności pomiędzy poszczególnymi bazami danych mtDNA wynikają prawdopodobnie z różnego zakresu sekwencjonowania regionu HVS I, który w przypadku portugalskiej oraz niemieckiej bazy danych był identyczny (16024-16365 p.z.), a jednocześnie nieco węższy niż ten, na podstawie którego analizowano polską bazę danych (w polskiej bazie wynosił on 15999-16400 p.z.). Warto zwrócić uwagę, że we fragmencie sekwencji HVS I analizowanym w niniejszej bazie danych, a nie objętym badaniami cytowanych autorów [6, 9] znajdują się pozycje o bardzo szybkim i szybkim tempie ewolucji (16390, 16399, 16368, 16366, 16391). Ponieważ mutacje w tych „gorących miejscach” obserwuje się w haplotypach należących do różnych haplogrup mtDNA [14], ich obecność wpływa na ogólne podwyższenie wartości parametru zróżnicowania nukleotydowego w próbkę polskiej na tle baz danych z populacji niemieckiej i portugalskiej [6, 9]. Szybsze wysycenie portugalskiej bazy danych może być także wynikiem zaobserwowanego przez zespół Pereiry i wsp. [6] niższego poziomu różnorodności mtDNA populacji portugalskiej w stosunku do Europejczyków z centralnej części kontynentu.

Z kolei odnotowane różnice w stopniu wysycenia parametru zróżnicowania haplotypowego oraz brak wysycenia w odniesieniu do liczby różnych haplotypów obserwowanych w populacji, są ściśle związane z reprezentatywnością w bazie danych tzw. haplotypów unikalnych, tj. takich, które pojawiają się w jej zasobach z niską częstością, równą 1 bądź 2.

W celu określenia rozkładu częstości haplotypów obserwowanych w polskiej populacyjnej bazie danych, w tych samych grupach losowych (tj. 200, 400, 600, 800, 1000 oraz 1253), co omówione wcześniej parametry, wyznaczono procentowy udział haplotypów unikalnych oraz takich, które pojawiają się w populacji ze stosunkowo wysoką częstością (ryc.2). Podczas analiz, podobnie jak to miało miejsce wcześniej, rozpatrywano oddzielnie region HVS I, HVS II oraz oba regiony mitochondrialnego DNA łącznie. Uzyskane dane przedstawione zostały na ryc. 2

W omawianej bazie, w regionie HVS I początkowa przewaga rzadkich haplotypów, których częstość sięgała 50% systematycznie spadała, by



Ryc. 2. Procentowy udział haplotypów obserwowanych w polskiej bazie populacyjnej dla regionu HVS I; HVS II oraz HVS I i HVS II: jednokrotnie (1X); dwukrotnie (2X) oraz trzykrotnie i więcej (3).

Fig. 2. Percentage of haplotypes found once (1X), twice (2X) and equal to or more than three times (3) for mtDNA regions HVS I, HVS II and entire control region (HVS I and HVS II) in samples of different sizes.

w całej populacji liczącej 1253 osoby osiągnąć wartość 27%. W regionie HVS II udział haplotypów częstych i unikalnych jest przeciwstawny do tego, jaki odnotowano dla regionu HVS I. Najwyższy udział w polskiej bazie danych osiągają haplotypy notowane z częstością równą lub wyższą niż 3 (71-84%), najniższe zaś sekwencje rzadkie (22-12%). Należy przy tym zauważyć, iż proporcje haplotypów odnotowywanych w populacji z wysoką częstością, jak i tych rzadkich, pozostają praktycznie na stałym poziomie, niezależnie od ilości analizowanych danych. Z kolei biorąc pod uwagę oba fragmenty regionu kontrolnego (HVS I i HVS II), zauważalny był początkowy niewielki spadek ilości rzadkich haplotypów (przy liczności 600 oraz 800), który zapewne przypisać można specyfice losowań, natomiast w kolejnych dwóch grupach losowych podnosił się on do poziomu przeszło 40%, przez co w całej bazie populacyjnej obserwuje się równy udział haplotypów unikalnych i tych notowanych ze stosunkowo wysoką częstością.

Praktycznie identyczny rozkład sekwencji notowanych z wysoką częstością (równą bądź większą niż 3) i unikalnych odnotowany został w portugalskiej bazie danych [6]. Także i w tym przypadku znakomitą większość spośród sekwencji regionu kontrolnego stanowią haplotypy rzadkie, których poziom w przypadku całej bazy liczącej 549 próbek wynosił aż 50%. Utrzymywanie w populacji tak wysokiego poziomu unikalnych haplotypów, i co się z tym wiąże, również dużego zróżnicowania nukleotydowego, zostało zauważone wcześniej, m.in. przez Pfeiffer i wsp. (1999) [15]. Podczas analizy zróżnicowania haplotypowego mtDNA lokalnej społeczności niemieckiej liczba różnych haplotypów, obserwowanych w populacji liczącej 700 osób, utrzymywała się na wysokim poziomie [15]. Taki stan miał miejsce pomimo istnienia wysokiego prawdopodobieństwa natrafienia na spokrewnione ze sobą linie mitochondrialnego DNA. Zjawisko niedoszacowania rzeczywistej różnorodności genetycznej istniejącej na poziomie populacji przypisywane jest niekiedy błędowi tzw. próbkowania (ang. *sampling bias*), a więc zgromadzenia w bazie danych niedostatecznej liczby haplotypów mtDNA lub ich niedostatecznej reprezentatywności, np. na skutek istnienia wyraźnej struktury (rozwarstwienia) w populacji.

Dotychczas przeprowadzone analizy europejskich baz danych, m.in. portugalskiej czy niemiec-

kiej, oparte na równaniu Michaelisa-Mentena, dają przybliżone wyobrażenie co do poziomu wysycenia bazy wszystkimi możliwymi haplotypami mtDNA. Symulacje zespołu Pereiry i wsp. (2004) wskazują, iż poziom wysycenia portugalskiej bazy profili mtDNA dla regionu HVS I z wysokim prawdopodobieństwem zostanie osiągnięty przy wielkości bazy równej 1000 próbek (przy liczbie różnych haplotypów  $n=370$ ) [6]. Dla regionu HVS II zostanie osiągnięty nieco wcześniej, bo przy liczności równej 900 ( $n=169$  haplotypów), a dla całego regionu kontrolnego przy wielkości populacji równej 1300 próbek reprezentowanych przez 638 różnych haplotypów [6]. Z kolei analizy niemieckiej bazy danych, przeprowadzone przez Helgasona i wsp. (2000), wskazują, że poziom wysycenia nastąpi dopiero przy liczebności 1700 osób [8]. Wydaje się, że odnotowane różnice między bazami są pochodną wspomnianego wcześniej nieznacznego rozwarstwienia (struktury) populacji europejskiej, tj. niewielkimi różnicami w poziomie różnorodności mtDNA pomiędzy południową oraz centralną i północną częścią kontynentu, obserwowanymi w niezależnych badaniach sekwencji regionu kontrolnego [11, 16].

Wyniki badań, uzyskane przez zespół naukowców skupionych w Projekcie Genograficznym, wydają się przeczyć przedstawionym powyżej optymistycznym przewidywaniom na temat osiągnięcia poziomu wysycenia baz danych [7]. W ramach wspomnianego projektu przebadano bowiem 78 590 próbek, pochodzących ze światowych populacji zamieszkujących m.in. kontynent euroazjatycki i afrykański, m.in. pod względem zróżnicowania regionu HVS I mtDNA (zakres sekwencjonowania wynosił 16024-16569 p.z.). O ile na poziomie liczby miejsc polimorficznych baza jest bliska poziomowi wysycenia, to liczba haplotypów, na którą bezpośredni wpływ mają m.in. mutacje zachodzące w mitochondrialnym DNA, jak i zjawisko homoplazji, nie osiągnęła fazy *plateau* [7]. Wydaje się jednak, że rozpatrywana całościowo baza mtDNA Projektu Genograficznego w obecnym kształcie nie może być wiarygodnym punktem odniesienia dla analiz wysycenia światowych baz danych. W swojej publicznie dostępnej części baza ta nie zawiera bowiem informacji na temat pochodzenia etnicznego badanych osób. Nie ulega zatem żadnej wątpliwości, że analiza wysycenia

bazy danych Projektu Genograficznego nie będzie uwzględniała rozwarstwienia populacji, istniejącego ponad wszelką wątpliwość zarówno w skali międzykontynentalnej, jak i w różnym stopniu w obrębie danego kontynentu.

## PODSUMOWANIE

Z przedstawionych powyżej analiz wynika, iż polska populacyjna baza profili mitochondrialnego DNA, podobnie jak inne omawiane bazy o zasięgu europejskim lub światowym, nie osiągnęła jeszcze poziomu wysycenia wszystkimi możliwymi haplotypami mtDNA. Baza ta w obecnym kształcie może być wykorzystywana dla celów genetyczno-sądowych, jednakże przy zachowaniu ostrożności w szacowaniu częstości rzadkich i unikalnych haplotypów. Zgodnie z wytycznymi obowiązującymi w międzynarodowym środowisku genetyków sądowych, dla oceny częstości rzadkich haplotypów w populacji należałoby tutaj stosować przedział ufności 95%, z wykorzystaniem logarytmu naturalnego częstości, przybliżenia normalnego do rozkładu dwumianowego oraz antylogarytmu [3, 17]. W ocenie częstości haplotypów można również stosować połączone bazy danych dla różnych grup etnicznych Europy środkowej, wykazano bowiem, że populacja tej części Starego Kontynentu nie nosi widocznych cech rozwarstwienia na poziomie regionu kontrolnego mtDNA [11].

Z rozważań na ten temat wyłania się zatem cel i kierunek działań laboratoriów o profilu genetyczno-sądowym. Poza głównym celem, jakim jest ciągłe wzbogacanie baz danych nowymi zestawami haplotypów pochodzącymi z mało poznanych (lub w ogóle nie poznanych) populacji, istnieje konieczność ustawicznego podnoszenia jakości danych. Jednym z ważnych narzędzi realizacji tego zamierzenia jest weryfikacja wprowadzanych informacji pod względem ich spójności z wiedzą filogenetyczną. Wykorzystanie bazy haplotypów mtDNA

w procedurach sądowych to tylko jeden z obszarów aplikacji. Bazy danych wykorzystywane są również w badaniach populacyjnych, demograficznych, antropologicznych, medycznych itd. Bez względu na charakter badań polimorfizmu mtDNA, właściwa filogenetyczna interpretacja uzyskiwanych wyników w dużej mierze zależy od szeroko pojętej jakości oraz zasobności baz, która stosunkowo szybko rośnie. Niestety nadal zauważalnym problemem pozostaje jakość informacji publikowanych w bazach danych. Ma to związek nie tylko ze stosowanym warsztatem analitycznym, ale także ze stopniem wykorzystania wiedzy filogenetycznej. Nie wszystkie zespoły badawcze dysponują jednakowo efektywnymi technologiami analiz polimorfizmu DNA, a możliwości prowadzenia badań filogenetycznych na najwyższym poziomie rozdzielczości (tj. poprzez sekwencjonowanie pełnych genomów mitochondrialnych) ujawniły się dopiero w ostatnich latach. W środowisku genetyków sądowych zaangażowanych w prowadzenie, rozbudowę oraz weryfikowanie prawidłowości danych kierowanych do bazy mtDNA, podejmowane są wysiłki zmierzające do podniesienia jakości informacji, chociażby poprzez weryfikację ich rzetelności przez zastosowanie odpowiednich narzędzi filogenetycznych [2]. Wydaje się, że najbardziej zaawansowaną pod tym względem pozostaje populacyjna baza danych regionu kontrolnego mtDNA EMPOP [18]. Perspektywnym kierunkiem rozwoju bazy EMPOP byłoby także zrealizowanie zamierzenia włączenia do jej zasobów danych dotyczących polimorfizmów jednonukleotydowych regionu kodującego (SNP). Z przeglądu najnowszego światowego piśmiennictwa wynika, że informacje z tego zakresu mogą znacząco podnieść poziom rozdzielczości filogenezy mitochondrialnego DNA. Ma to duże znaczenie zwłaszcza w przypadku badań zróżnicowania sekwencji mtDNA stosunkowo słabo rozpoznanych populacji zachodniej i wschodniej Eurazji.

## PIŚMIENNICTWO

1. Parson W., Brandstätter A., Alonso A., Brandt N., Brinkmann B., Carracedo A., Corach D., Froment O., Furac I., Grzybowski T., Hedberg K., Keyser-Tracqui C., Kupiec T., Lutz-Bonengel S., Mevag B., Płoski

R., Schmitter H., Schneider P., Syndercombe-Court D., Sorensen E., Thew: The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives. *Forensic Sci Int*, 2004: 139(2-3): 215-226.

2. Parson W. i Bandelt H. J.: Extended guidelines for mtDNA typing of population data in forensic science. *Forensic Sci Int Genet*, 2007: 1(1): 13-19.
3. Holland M. M. i Parsons T. J.: Mitochondrial DNA sequence analysis – validation and use for forensic casework. *Forensic Sci Rev*, 1999: 11,21.
4. Grzybowski T., Malyarchuk B. A., Bednarek J., Woźniak M., Papuga M., Stopińska K., Łuczak S.: Phylogeographic approach in the interpretation of mitochondrial DNA sequencing results in forensics. *Arch Med Sądowej Kryminol*, 2006: 56(3): 191-197.
5. Salas A., Bandelt H. J., Macaulay V., Richards M. B.: Phylogeographic investigations: the role of trees in forensic genetics. *Forensic Sci Int*, 2007: 168(1): 1-13.
6. Pereira L., Cunha C., Amorim A.: Predicting sampling saturation of mtDNA haplotypes: an application to an enlarged Portuguese database. *Int J Legal Med*, 2004: 118(3): 132-136.
7. Behar D. M., Rosset S., Blue-Smith J., Balanovsky O., Tzur S., Comas D., Mitchell R. J., Quintana-Murci L., Tyler-Smith C., Wells R. S.: The Genographic Project public participation mitochondrial DNA database. *PLoS Genet*, 2007: 3(6): e104.
8. Helgason A., Sigurðardóttir S., Gulcher J. R., Stefansson K., Ward R. (2000) Sampling saturation and the European mtDNA pool: implications for detecting genetic relationships among populations. W: Renfrew C., Boyle K. (ed.) *Archeogenetics: DNA and the population prehistory of Europe*. McDonald Institute for Archeological Research, University of Cambridge, Cambridge, 285-294.
9. Pfeiffer H., Forster P., Ortmann C., Brinkmann B.: The results of an mtDNA study of 1200 inhabitants of a German village in comparison to other Caucasian databases and its relevance for forensic casework. *Int J Legal Med*, 2001: 114: 169–172.
10. Malyarchuk B. A., Grzybowski T., Derenko M. V., Czarny J., Woźniak M., Miścicka-Śliwka D.: Mitochondrial DNA variability in Poles and Russians. *Ann Hum Genet*, 2002: 66(Pt 4): 261-283.
11. Grzybowski T., Malyarchuk B. A., Derenko M. V., Perkova M. A., Bednarek J., Woźniak M.: Complex interactions of the Eastern and Western Slavic populations with other European groups as revealed by mitochondrial DNA analysis. *Forensic Sci Int Genet*, 2007: 1(2): 141-147.
12. Dąca P.: Filogeneza mitochondrialnego DNA w subpopulacjach Polski: aspekty etnogenetyczne i identyfikacyjne. Rozprawa doktorska. Bydgoszcz 2010.
13. Excoffier L., Laval G., Schneider S.. Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online*, 2005: 1: 47-50.
14. Soares P., Ermini L., Thomson N., Mormina M., Rito T., Röhl A., Salas A., Oppenheimer S., Macaulay V., Richards M. B.: Correcting for purifying selection: an improved human mitochondrial molecular clock. *Am J Hum Genet* 2009: 84(6): 740-759.
15. Pfeiffer H., Brinkmann B., Hühne J., Rolf B., Morris A. A., Steighner R., Holland M. M., Forster P.: Expanding the forensic German mitochondrial DNA control region database: genetic diversity as a function of sample size and microgeography. *Int J Legal Med*, 1999: 112.
16. McEvoy B., Richards M., Forster P., Bradley D. G.: The Longue Duree of genetic ancestry: multiple genetic marker systems and Celtic origins on the Atlantic facade of Europe. *Am J Hum Genet* 2004: 75(4): 693-702.
17. National Research Council: *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington D.C. 1996.
18. [www.empop.org](http://www.empop.org)

Adres do korespondencji:

Dr hab. Tomasz Grzybowski, prof. UMK

Katedra Medycyny Sądowej, Zakład Genetyki Molekularnej i Sądowej

Collegium Medicum UMK

ul. M. Curie-Skłodowskiej 9

85-094 Bydgoszcz

Tel: +48 52 585 3556

e-mail: [tgrzyb@cm.umk.pl](mailto:tgrzyb@cm.umk.pl)