

**Marcin Woźniak, Tomasz Grzybowski, Jarosław Starzyński, Marta Papuga,
Katarzyna Stopińska, Sylwia Łuczak**

Zależności pomiędzy przynależnością haplogrupową a haplotypem Y-STR jako potencjalny element kontroli poprawności analiz w medycynie sądowej

Correlations between haplogroup membership and Y-STR haplotype as a potential measure of quality control in forensic examinations

Z Zakładu Genetyki Molekularnej i Sądowej, Katedry Medycyny Sądowej Collegium Medicum im. Ludwika Rydygiera w Bydgoszczy, Uniwersytet Mikołaja Kopernika w Toruniu
Kierownik: prof. dr hab. n. med. Karol Śliwka

W pracy podjęto próbę określenia związków pomiędzy częstością występowania poszczególnych haplotypów chromosomu Y w zakresie tzw. haplotypu minimalnego a przynależnością danego chromosomu do jednej z trzech haplogrup: R1*, R1a1* i I*. Na podstawie analizy 146 chromosomów obliczono częstości występowania poszczególnych alleli w haplogrupach i stwierdzono, że metodą, która najefektywniej przypisuje chromosom do haplogrupy na podstawie jego haplotypu minimalnego jest metoda oparta na obliczaniu współczynnika podobieństwa haplotypów w oparciu o stosunek częstości allelu występującego w danym haplocie do allelu najczęstszego w danej haplogrupie. Zaproponowano również potencjalne zastosowania tej metody w genetyce sądowej.

A correlation between particular Y-STR alleles from the so-called "minimal haplotype" and haplogroup membership of the Y chromosome was tested. We collected 146 Y chromosomes from haplogroups R1*, R1a1* and I* and estimated the frequency of Y-STR alleles in each haplogroup. We then used different algorithms to assign a haplogroup to a haplotype, and tested their accuracy. Generally, a method based on calculation of haplotype similarity using the highest allele frequencies as modal values and assigning a score to each locus based on a ratio of allele frequencies turned out to give the most precise matches. However, using the same rules for Y chromosomes from other populations did not allow for precise estimation of their Y chromosome haplogroup frequencies. Possible explanations for this failure include

interpopulation differences in haplotypes correlated with particular haplogroups, as well as a relatively small number of chromosomes analyzed. Potential uses for the presented method in forensics were also described.

WSTĘP

W badaniach wykorzystujących techniki identyfikacji osobniczej prawidłowe oznaczenie profilu genetycznego poszczególnych osobników stanowi najistotniejszy element procesu analitycznego, szczególnie w przypadku analiz prowadzonych dla potrzeb medyczno-sądowych. Obserwowana ostatnio w piśmiennictwie krytyka rzetelności wyników umieszczanych w niektórych bazach danych mtDNA wskazuje, iż w wielu wypadkach środki mające zapewnić poprawność generowanych i publikowanych wyników są niewystarczające [1]. Jedną z metod pozwalających na zwiększenie wiarygodności otrzymywanych wyników są analizy filogenetyczne, z powodzeniem stosowane w badaniach mtDNA i wykorzystywane ostatnio szeroko do weryfikacji błędów występujących we wspomnianych wyżej bazach danych [2].

Podstawowym elementem analizy filogenetycznej jest określenie haplotypu danej cząsteczki DNA z wykorzystaniem loci o zróżnicowanej częstości mutacji, np. konserwatywnych mutacji punktowych (SNP) zlokalizowanych w obrębie sekwencji kodu-

jącej genów oraz częściej ulegających mutacjom pozycji występujących w regionach pozagenowych. Nierekombinujące cząsteczki DNA, takie jak mtDNA czy znaczna część chromosomu Y, określana jako region nierekombinujący (NRY) poddają się analizie filogenetycznej, gdyż zachodzące w nich mutacje kumulują się w poszczególnych liniach z pokolenia na pokolenie i pozwalają na odtworzenie kolejności mutacji prowadzących do powstania wybranej linii (matczynej lub ojcowskiej). Wysoce konserwatywne mutacje definiują przynależność danej cząsteczki DNA do klasy podobnych cząsteczek, posiadających identyczne mutacje i określanych jako haplogrupa monofiletyczna. Ze względu na konserwatywny charakter mutacji definiujących poszczególne haplogrupy oraz z powodu oddziaływania dryfu genetycznego, eliminującego z populacji rzadkie warianty sekwencji, uprawnione jest założenie, że każda z obserwowanych obecnie mutacji definiujących haplogrupy powstała i utrzymała się w populacji tylko raz w toku ewolucji. Pamiętając o braku rekombinacji w takich cząsteczkach jak NRY można zatem założyć, że mutacja definiująca daną haplogrupę pojawia się w nich w kontekście konkretnego haplotypu i jako nowy haplotyp jest przekazywana kolejnym pokoleniom. Kolejne mutacje zachodzące w obrębie danego haplotypu, np. w loci STR chromosomu Y, powodują pojawienie się w następujących, po powstaniu nowej haplogrupy, pokoleniach zróżnicowania haplotypów składających się na daną haplogrupę, przy czym zróżnicowanie to będzie tym większe, im starsza i bardziej rozpowszechniona w populacji jest dana haplogrupa [3].

Biorąc pod uwagę przedstawiony wyżej model powstawania zmienności nierekombinujących cząsteczek DNA można zatem założyć, że każda obserwowana haplogrupa powinna posiadać zestaw powiązanych z nią haplotypów, określanych przez wysoko polimorficzne loci znajdujące się na danej cząsteczce DNA. Jeśli powyższe założenia są prawdziwe i haplotypy w obrębie obserwowanej haplogrupy pochodzą od wspólnego przodka to powinno być możliwe, przynajmniej w przypadku stosunkowo młodych haplogrup, statystyczne potwierdzenie związków konkretnych haplotypów lub nawet pojedynczych alleli, z daną haplogrupą. Wykazanie istnienia tego typu związków stanowi podstawę stosowanych coraz szerzej filogenetycznych metod analizy mtDNA, które pozwalają na weryfikację poprawności oznaczenia sekwencji polimorficznych regionów HV1 i HV2 na podstawie obecnych w nich charakterystycznych mutacji powiązanych z przynależnością cząsteczki do określonej haplogrupy [1, 2]. Podobne obserwacje poczyniono w odniesieniu do alleli loci mikrosatelitarnych chromosomu Y (Y-STR) i ich związków z poszczególnymi haplogrupami tego chromosomu [4, 5]. Przydatność

tego typu opracowań dla celów medycyny sądowej nie jest jednak zwykle wysoka, gdyż ich autorzy często posługują się bardzo szerokim lub zbyt wąskim wachlarzem układów Y-STR, znacznie szerszym niż rutynowo stosowany w badaniach śladów biologicznych czy spornego ojcostwa tzw. „haplotyp minimalny”, obejmujący układy: DYS19, DYS385, DYS389I, DYS389II, DYS390, DYS391, DYS392 i DYS393 [6], przez co określenie związków poszczególnych alleli haplotypu minimalnego z konkretnymi haplogrupami bywa kłopotliwe.

Celem niniejszej pracy jest wstępna ocena możliwości zastosowania informacji dotyczącej przynależności haplogrupowej chromosomu Y w genetyce sądowej, w tym przede wszystkim jej ewentualnego wykorzystania do weryfikacji poprawności oznaczenia profilu loci Y-STR w zakresie tzw. haplotypu minimalnego.

MATERIAŁY I METODY

Materiał do badań stanowiło 146 preparatów DNA pochodzących od niespokrewnionych mężczyzn, pozyskanych metodą ekstrakcji organicznej z próbek krwi obwodowej. Stężenie uzyskiwanych preparatów DNA oceniano metodą hybrydyzacyjną z wykorzystaniem zestawu QuantiBlot (Perkin Elmer). Analizę polimorfizmu loci SNP i STR prowadzono zgodnie z opublikowaną wcześniej metodyką [7]. Spośród wybranych do badań próbek, 79 należało do haplogrupy R1a1* (mutacja w markerze M17), 46 należało do haplogrupy I* (mutacja w markerze M170), a 21 należało do haplogrupy R1*(xR1a1* – brak mutacji w markerze M17). Nazewnictwo haplogrup chromosomu Y jest zgodne z nomenklaturą Y Chromosome Consortium (YCC) [8]. Nazewnictwo alleli loci Y-STR zgodne jest z nomenklaturą ISFG [9]. W przypadku locus DYS385 krótszy allel oznaczano jako DYS385a, a dłuższy jako DYS385b; obecność jednego produktu amplifikacji zapisywano jako obecność dwóch alleli o jednakowej długości w DYS385a i DYS385b

Statystyczną istotność różnic częstości alleli i haplotypów obliczano na podstawie dwustronnego testu różnicy pomiędzy wskaźnikami struktury. W obliczeniach wykorzystywano oprogramowanie Statistica v. 7.1, uznając różnice za statystycznie istotne gdy $p < 0.05$.

Parametr określający podobieństwo otrzymanego haplotypu do haplotypu modalnego danej haplogrupy, oznaczany dla potrzeb niniejszej pracy jako pHM^X , gdzie X oznacza symbol haplogrupy, wyliczano korzystając ze wzoru zaproponowanego przez Athey'a [10].

WYNIKI I DYSKUSJA

Przeprowadzono analizę częstości występowania poszczególnych alleli układów Y-STR w obrę-

bie badanych haplogrup. Wyniki obliczeń przedstawiono w tabelach I, II i III.

Allele, których częstości różniły się istotnie pomiędzy haplogrupami zgrupowano w tabeli IV. Jedynie w locus DYS391 nie zaobserwowano żadnych

Tabela I. Częstości alleli w haplogrupie I*. Pierwsza kolumna tabeli: symbole alleli Y-STR. Pierwszy wiersz tabeli: nazwy analizowanych układów (pozbawione przedrostka DYS). Pogrubieniem zaznaczono allel najczęstszy w danym locus w danej haplogrupie.

Table I. Allele frequencies in haplogroup I*. First column: symbols of Y-STR alleles. First row: names of Y-STR loci without the DYS prefix. The most frequent allele in each locus is shown in bold.

	19	389I	389II	390	391	392	393	385a	385b
8	—	—	—	—	—	—	—	—	—
9	—	—	—	—	—	—	—	—	—
10	—	—	—	—	0.543	—	—	—	—
11	—	—	—	—	0.435	0.891	—	—	—
12	—	0.413	—	—	0.022	0.109	0.065	0.022	—
13	0.043	0.587	—	—	—	—	0.913	0.283	0.043
14	0.370	—	—	—	—	—	0.022	0.587	0.326
15	0.174	—	—	—	—	—	—	0.043	0.457
16	0.283	—	—	—	—	—	—	0.065	0.087
17	0.130	—	—	—	—	—	—	—	0.043
18	—	—	—	—	—	—	—	—	0.043
19	—	—	—	—	—	—	—	—	—
20	—	—	—	—	—	—	—	—	—
21	—	—	—	—	—	—	—	—	—
22	—	—	—	0.261	—	—	—	—	—
23	—	—	—	0.152	—	—	—	—	—
24	—	—	—	0.522	—	—	—	—	—
25	—	—	—	0.043	—	—	—	—	—
26	—	—	—	—	—	—	—	—	—
27	—	—	—	0.022	—	—	—	—	—
28	—	—	0.326	—	—	—	—	—	—
29	—	—	0.087	—	—	—	—	—	—
30	—	—	0.217	—	—	—	—	—	—
31	—	—	0.239	—	—	—	—	—	—
32	—	—	0.130	—	—	—	—	—	—

Tabela II. Częstości alleli loci Y-STR w haplogrupie R1a1*. Pierwsza kolumna tabeli: symbole alleli Y-STR. Pierwszy wiersz tabeli: nazwy analizowanych układów (pozbawione przedrostka DYS). Pogrubieniem zaznaczono allel najczęstszy w danym locus w danej haplogrupie.

Table II. Allele frequencies in haplogroup R1a1*. First column: symbols of Y-STR alleles. First row: names of Y-STR loci without the DYS prefix. The most frequent allele in each locus is shown in bold.

	19	389I	389II	390	391	392	393	385a	385b
8	—	—	—	—	—	—	—	0.013	—
9	—	—	—	—	—	—	—	—	—
10	—	—	—	—	0.544	—	—	0.215	—
11	—	—	—	—	0.430	0.949	—	0.709	0.025
12	—	0.025	—	—	0.013	0.051	—	0.063	—
13	—	0.886	—	—	0.013	—	0.949	—	0.051
14	0.038	0.089	—	—	—	—	0.051	—	0.772
15	0.253	—	—	—	—	—	—	—	0.127
16	0.430	—	—	—	—	—	—	—	0.013
17	0.266	—	—	—	—	—	—	—	—
18	—	—	—	—	—	—	—	—	0.013
19	0.013	—	—	—	—	—	—	—	—
20	—	—	—	—	—	—	—	—	—
21	—	—	—	—	—	—	—	—	—
22	—	—	—	—	—	—	—	—	—
23	—	—	—	0.013	—	—	—	—	—
24	—	—	—	0.139	—	—	—	—	—
25	—	—	—	0.759	—	—	—	—	—
26	—	—	—	0.089	—	—	—	—	—
27	—	—	—	—	—	—	—	—	—
28	—	—	—	—	—	—	—	—	—
29	—	—	0.253	—	—	—	—	—	—
30	—	—	0.532	—	—	—	—	—	—
31	—	—	0.190	—	—	—	—	—	—
32	—	—	0.025	—	—	—	—	—	—

Tabela III. Częstości alleli loci Y-STR w haplogrupie R1* (xR1a1*). Pierwsza kolumna tabeli: symbole alleli Y-STR. Pierwszy wiersz tabeli: nazwy analizowanych układów (pozbawione przedrostka DYS). Pogrubieniem zaznaczono allel najczęstszy w danym locus w danej haplogrupie.

Table III. Allele frequencies in haplogroup R1*. First column: symbols of Y-STR alleles. First row: names of Y-STR loci without the DYS prefix. The most frequent allele in each locus is shown in bold.

	19	389I	389II	390	391	392	393	385a	385b
8	—	—	—	—	—	—	—	—	—
9	—	—	—	—	—	—	—	—	—
10	—	—	—	—	0.381	—	—	0.048	—
11	—	—	—	—	0.619	—	0.048	0.762	0.048
12	—	0.048	—	—	—	0.048	0.143	0.190	—
13	—	0.857	—	—	—	0.952	0.810	—	0.048
14	0.857	0.095	—	—	—	—	—	—	0.714
15	0.095	—	—	—	—	—	—	—	0.190
16	—	—	—	—	—	—	—	—	—
17	0.048	—	—	—	—	—	—	—	—
18	—	—	—	—	—	—	—	—	—
19	—	—	—	—	—	—	—	—	—
20	—	—	—	—	—	—	—	—	—
21	—	—	—	—	—	—	—	—	—
22	—	—	—	—	—	—	—	—	—
23	—	—	—	0.286	—	—	—	—	—
24	—	—	—	0.476	—	—	—	—	—
25	—	—	—	0.238	—	—	—	—	—
26	—	—	—	—	—	—	—	—	—
27	—	—	—	—	—	—	—	—	—
28	—	—	0.190	—	—	—	—	—	—
29	—	—	0.524	—	—	—	—	—	—
30	—	—	0.286	—	—	—	—	—	—
31	—	—	—	—	—	—	—	—	—
32	—	—	—	—	—	—	—	—	—

istotnych różnic częstości alleli pomiędzy analizowanymi haplogrupami, natomiast w każdym z pozostałych badanych układów STR ujawniono obecność od 2 do 6 alleli charakteryzujących się statystycznie istotnymi różnicami częstości między co najmniej dwiema haplogrupami.

Wśród obserwowanych alleli, trzy (DYS19-14, DYS389II-29 i DYS390-25) wykazywały statystycznie istotne różnice częstości pomiędzy wszystkimi haplogrupami, co wskazuje na potencjalne istnienie alleli diagnostycznych, będących wskaźnikami przynależności chromosomu do określonej haplogrupy. Allel 14 locus DYS19 występuje np. dwa razy częściej w chromosomach należących do haplogrupy R1* niż w chromosomach haplogrupy I* oraz ponad 20 razy częściej w chromosomach haplogrupy R1* niż R1a1*. Allel ten stanowi zatem, jak się zdaje, dobry wskaźnik pozwalający na dyskryminację pomiędzy haplogrupami R1* i R1a1*. Allel 25 locus DYS390 występuje ok. 3 razy częściej w haplogrupie R1a1* niż w haplogrupie R1*, ponad 5 razy częściej w haplogrupie R1* niż w haplogrupie I* oraz niemal 20 razy częściej w haplogrupie R1a1* niż w haplogrupie I*. Wystąpienie allelu 25 locus DYS390 może zatem stanowić przesłankę za przynależnością danego chromosomu do haplogrupy R1a1* i odrzuceniem hipotezy, że należy on do haplogrupy I*. Allel 29 locus DYS389II obecny jest w ponad połowie chromosomów haplogrupy R1*, w co czwartym chromosomie grupy R1a1*

i tylko w 9 % chromosomów haplogrupy I*. Spośród innych alleli, których częstości różnią się pomiędzy poszczególnymi parami haplogrup, na uwagę zasługują również: allel 16 locus DYS19, którego obecność może stanowić podstawę wykluczenia przynależności do haplogrupy R1*, allel 12 locus DYS389I, wskazujący na przynależność chromosomu do haplogrupy I*, allel 28 locus DYS389II, wykluczający przynależność do haplogrupy R1a1* oraz allel 31 tego locus, wykluczający przynależność do haplogrupy R1*. Zestawienie wszystkich alleli mogących stanowić markery przynależności haplogrupowej przedstawiono w tabeli IV. Należy zauważyć, że obliczona wartość p w przypadku 20 alleli jest niższa niż 0.005, co może świadczyć na korzyść silnych związków tych alleli z haplogrupami, w których występują najczęściej.

Aby zweryfikować możliwość określania przynależności haplogrupowej przy pomocy informacji o częstości występowania poszczególnych alleli w obrębie haplogrup, przygotowano arkusz kalkulacyjny obliczający częstość występowania w badanym haplotypie alleli charakterystycznych dla jednej z badanych haplogrup (I*, R1* i R1a1*) oraz częstość występowania w haplotypie alleli pojawiających się ze zbliżoną częstością w dwóch z trzech badanych haplogrup (stosunek częstości bliski 1 lub częstości niższe niż 5 %) a jednocześnie rzadkich lub nieobecnych w trzeciej z nich (przykładowo, allel 13 locus DYS389I zaliczany był do grupy

Tabela IV. Porównanie częstości alleli Y-STR, których wartości różniły się istotnie pomiędzy poszczególnymi haplogrupami. Istotność statystyczną obserwowanych różnic częstości danego allelu obliczano dla każdej pary haplogrup. Czcionką pogrubioną zaznaczono wartości p niższe od 0.05, czcionką pogrubioną i podkreśleniem oznaczono wartości p niższe od 0.005. W sekcji „przynależność” w kolumnie „+” wskazano haplogrupy, w których dany allel występuje stosunkowo często, w kolumnie „-” wskazano haplogrupy, w których dany allel występuje najrzadziej. Szczegóły w tekście.

Table IV. Comparison of allele frequencies differing significantly between haplogroups. The statistical significance of differences was calculated for each pair of haplogroups. Differences with the p value lower than 0.05 are shown in bold, while those with the p value lower than 0.005 are shown in bold and underlined. The “Hg membership” section shows haplogroups with relatively frequent (“+” column) and relatively rare (“-” column) occurrences of the given allele. Details in the text.

Locus /allele	Częstość allelu w haplogrupie Allele frequency in a haplogroup			Wartość p dla pary haplogrup p value for each haplogroup pair			Przynależność HG membership	
	I*	R1*	R1a1*	I*/R1*	I*/R1a1*	R1*/R1a1*	+	-
DYS19								
14	0.370	0.857	0.038	0.000	0.000	0.000	R1*+I*	R1a1*
16	0.283	0.000	0.430	0.009	0.104	0.004	R1a1**+I*	R1*
17	0.130	0.048	0.266	0.312	0.077	0.035	R1a1**+I*	R1*
DYS389I								
12	0.413	0.048	0.025	0.004	0.002	0.583	I*	R1a1*+R1*
13	0.587	0.857	0.886	0.033	0.000	0.717	R1*+R1a1*	I*
14	0.000	0.095	0.089	0.038	0.039	0.944	R1*+R1a1*	I*
DYS389II								
28	0.326	0.190	0.000	0.256	0.000	0.000	I*+R1*	R1a1*
29	0.087	0.524	0.253	0.002	0.025	0.019	R1*+R1a1*	I*
30	0.217	0.286	0.532	0.541	0.001	0.048	R1a1*	I*+R1*
31	0.239	0.000	0.190	0.017	0.516	0.033	I*+R1a1*	R1*
32	0.130	0.000	0.025	0.088	0.022	0.466	I*	R1a1*
DYS390								
22	0.261	0.000	0.000	0.012	0.000	1.000	I*	R1a1*+R1*
23	0.152	0.286	0.013	0.203	0.003	0.000	R1*+I*	R1a1*
24	0.522	0.476	0.139	0.728	0.000	0.001	R1*+I*	R1a1*
25	0.043	0.238	0.759	0.018	0.000	0.000	R1a1*	I*
26	0.000	0.000	0.089	1.000	0.039	0.159	R1a1*	I*
DYS392								
11	0.891	0.000	0.949	0.000	0.230	0.000	I*+R1a1*	R1*
13	0.000	0.952	0.000	0.000	1.000	0.000	R1*	I*+R1a1*
DYS393								
12	0.065	0.143	0.000	0.303	0.024	0.001	R1*	R1a1*
DYS385a								
10	0.000	0.048	0.215	0.139	0.001	0.080	R1a1*	I*
11	0.000	0.762	0.709	0.000	0.000	0.632	R1*+R1a1*	I*
12	0.022	0.190	0.063	0.018	0.303	0.073	I*+R1*	R1a1*
13	0.283	0.000	0.000	0.009	0.000	1.000	I*	R1a1*+R1*
14	0.587	0.000	0.000	0.000	0.000	1.000	I*	R1a1*+R1*
16	0.065	0.000	0.000	0.236	0.024	1.000	I*	R1a1*+R1*
DYS385b								
14	0.326	0.714	0.772	0.004	0.000	0.582	R1*+R1a1*	I*
15	0.457	0.190	0.127	0.040	0.000	0.462	I*	R1a1*+R1*
16	0.087	0.000	0.013	0.168	0.045	0.601	I*	R1a1*+R1*

„R1* lub R1a1*”). W obliczeniach wykorzystywano jedynie allele, których częstości istotnie różniły się między haplogrupami; w przypadku wystąpienia w haplotypie innego allelu lub występowania allelu obecnego we wszystkich trzech grupach z podobną częstością (np. allel 13 locus DYS393) nie przypisywano go do żadnej haplogrupy. Po określeniu w danym haplotypie liczby alleli charakterystycznych dla poszczególnych haplogrup, haplotyp zaliczany był do tej haplogrupy, dla której zanotowa-

no największą liczbę charakterystycznych alleli. Przykład zastosowania tego algorytmu przedstawia rycina 1. W celu oceny skuteczności przyjętego algorytmu, do przygotowanego arkusza wprowadzono wszystkie 146 haplotypów, na podstawie których obliczano częstości alleli i dokonano przypisania przynależności haplogrupowej na podstawie obecnych w nich alleli, a następnie porównano otrzymane wyniki z rzeczywistą przynależnością haplogrupową chromosomów, z których pochodziły

badane haplotypy. Zastosowano również zmodyfikowaną wersję algorytmu, w którym do obliczeń wykorzystano wszystkie allele. Skuteczność przypisania haplotypów do poszczególnych haplogrup przedstawia tabela V (panele A i B).

Spośród 79 chromosomów haplogrupy R1a1*, 67 zostało przypisanych prawidłowo na podstawie analizy alleli, których częstość różniła się istotnie pomiędzy haplogrupami. W wyniku analizy wszystkich alleli prawidłowo przypisano do haplogrupy 76 haplotypów. Podobnie w przypadku dwóch pozostałych zestawów chromosomów, należących do haplogrup R1* i I* – algorytm oparty na analizie wszystkich alleli pozwolił na prawidłowe przypisanie do haplogrupy większej ilości haplotypów, niż algorytm

oparty na analizie jedynie tych alleli, których częstość różniła się istotnie pomiędzy haplogrupami.

Inna metoda określania przynależności haplogrupowej chromosomu Y na podstawie analizy haplotypów Y-STR, zaproponowana przez Whit Athey'a [10], oparta jest na obliczaniu współczynnika określającego podobieństwo danego haplotypu do haplotypu modalnego dla danej haplogrupy. Parametr ten określono dla potrzeb niniejszej pracy jako pHM^x. Wspomniana metoda opiera się na obliczaniu cząstkowych parametrów podobieństwa dla każdego locus, określonych jako stosunek częstości allelu występującego w danym locus w danym haplocie do częstości najczęstszego allelu w danym locus w danej haplogrupie, a zatem wartości

Tabela V. Skuteczność określenia przynależności haplogrupowej chromosomów Y na podstawie ich haplotypów minimalnych. „n.o.” – brak możliwości przypisania do określonej haplogrupy.

Table V. Efficiency of haplogroup assignment based on minimal haplotype data. Rows represent number of chromosomes assigned to a haplogroup based on Y-STR haplotype. “n.o.” – impossible to assign”.

A	Haplogrupa przypisana na podstawie występowania alleli o częstościach istotnie różnych między haplogrupami Haplogroup assigned based on significantly different Y-STR allele frequencies			
Rzeczywista przynależność haplogrupowa Actual haplogroup	I*	R1*	R1a1*	n.o.
I*	29	3	3	11
R1*	1	15	2	3
R1a1*	1	5	67	6
B	Haplogrupa przypisana na podstawie analizy częstości wszystkich alleli Haplogroup assigned based on allele frequencies of all Y-STR alleles			
Rzeczywista przynależność haplogrupowa Actual haplogroup	I*	R1*	R1a1*	n.o.
I*	34	1	3	8
R1*	—	20	1	—
R1a1*	—	1	76	2
C	Haplogrupa przypisana na podstawie parametru pHM ^x Haplogroup assigned based on pHM ^x value			
Rzeczywista przynależność haplogrupowa Actual haplogroup	I*	R1*	R1a1*	n.o.
I*	46	—	—	—
R1*	—	20	1	—
R1a1*	1	—	78	—

Ryc. 1. Schemat algorytmu obliczeń przynależności haplogrupowej chromosomu Y na podstawie haplotypu Y-STR. Oznaczenia: „n.o.” – allel nie należy do grupy alleli o częstości istotnie odmiennie pomiędzy haplogrupami, „n.b.” locus nie badane ze względu na brak istotnych różnic w częstości alleli.

Fig. 1. Schematic representation of the algorithm assigning haplogroup membership to Y-STR minimal haplotypes. Legend: “n.o.” – allele does not belong to the group of alleles exhibiting a frequency significantly differing between haplogroups, “n.b.” – locus not examined due to no significant differences in allele frequencies between haplogroups.

1. Wprowadzenie haplotypu (haplotype input):

DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b
15	13	30	25	11	11	13	11	14

2. Przypisanie poszczególnych alleli do haplogrup (allele assignment to a haplogroup):

DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385a	DYS385b
n.o.	R1*+R1a1*	R1a1*	R1a1*	n.b.	I*+R1a1*	n.o.	R1*+R1a1*	R1*+R1a1*

3. Zliczenie częstości pojawienia się danej haplogrupy (count of each haplogroup occurrence):

I*	R1*	R1a1*
1	3	6

4. Przypisanie haplogrupy (haplogroup assignment): R1a1*

parametrów cząstkowych wahają się między 0 a 1. Z obliczonych parametrów cząstkowych obliczany jest następnie parametr pHM^X , będący średnią geometryczną parametrów cząstkowych pomnożoną przez 100. Przedstawiony wyżej algorytm wykorzystano do obliczenia wartości parametrów pHM^I ,

pHM^{R1^*} i pHM^{R1a1^*} dla haplotypów minimalnych badanych chromosomów należących do poszczególnych haplogrup. Wyniki przeprowadzonych obliczeń podsumowano w tabeli VI.

Najwyższe średnie wartości parametru pHM^X obliczanego dla częstości alleli poszczególnych

Tabela VI. Wyniki obliczeń parametru pHM^X dla chromosomów z poszczególnych haplogrup. Tabela przedstawia średnią wartość wyniku obliczeń dla chromosomów z poszczególnych haplogrup oraz odchylenie standardowe średniej, medianę, wartości minimalne i maksymalne oraz wartości osiągnięte przez 95 % haplotypów w danej haplogrupie. Najwyższe wartości średniej wyliczone dla częstości alleli w poszczególnych haplogrupach pogrubiono.
Table VI. Results of pHM^X parameter calculation for chromosomes belonging to particular haplogroups. Mean value, standard deviations, median value, maximum and minimum, as well as values reached by 95 % of haplotypes in particular haplogroup are listed. The highest mean value for each haplogroup is shown in bold.

		I*	R1a1*	R1*
		pHM^{R1a1^*}		
średnia	mean	22.41	72.63	13.03
odch. std.	std. dev.	8.64	16.62	8.28
mediana	median	23.73	76.27	11.86
maksimum	max.	43.46	100.00	48.61
minimum	min.	5.02	37.09	1.40
95 centyl	95 percentile	42.71<	>41.99	24.88<
		pHM^{R1^*}		
średnia	mean	10.81	15.67	74.46
odch. std.	std. dev.	4.61	11.98	18.69
mediana	median	9.82	15.71	80.00
maksimum	max.	21.26	59.95	94.75
minimum	min.	3.22	5.03	31.34
95 centyl	95 percentile	21<	25<	>46
		pHM^{I^*}		
średnia	mean	71.52	13.15	4.00
odch. std.	std. dev.	15.28	9.49	2.68
mediana	median	72.57	14.54	3.23
maksimum	max.	93.78	30.13	11.98
minimum	min.	24.31	0.99	0.44
95 centyl	95 percentile	>42.27	28.14<	8.92<

haplogrup otrzymywano zawsze dla chromosomów tej haplogrupy, której częstości alleli Y-STR wykorzystano w obliczeniach. Analiza wykazała również, że w przypadku chromosomów haplogrupy I* 95 % haplotypów osiągało wartość parametru pHM^I wyższą niż 42.27, podczas gdy 95 % haplotypów chromosomów haplogrup R1* i R1a1* osiągało wartości odpowiednio niższe od 8.92 i 28.14. Podobne zależności zaobserwowano w przypadku chromosomów należących do pozostałych haplogrup. Warto zwrócić uwagę, iż w przypadku haplogrupy R1a1* wartość parametru pHM^{R1a1*} jest wyższa dla ok. 2 % chromosomów należących do haplogrupy I* niż dla ok. 2 % chromosomów należących do haplogrupy R1a1*. Przeprowadzono również test poprawności określenia przynależności haplogrupowej chromosomów dokonując ich przypisania do poszczególnych haplogrup na podstawie wartości wyliczonych parametrów pHM^I , pHM^{R1*} i pHM^{R1a1*} , przyjmując że chromosom należy do tej haplogrupy, dla której parametr był najwyższy (tabela V panel C). Metoda oznaczania przynależności haplogrupowej na podstawie obliczenia parametru pHM^X dała w obrębie analizowanych danych najwyższy współczynnik zgodności z rzeczywistą haplogrupą chromosomu Y w porównaniu do pozostałych dwóch algorytmów wykorzystanych w niniejszej pracy. Wszystkie chromosomy haplogrupy I* zostały przypisane do prawidłowej haplogrupy na podstawie obliczeń współczynnika pHM^X , spośród 79 chromosomów haplogrupy R1a1* 78 zostało przypisanych prawidłowo, podobnie spośród 21 chromosomów haplogrupy R1* 20 zostało przypisanych do właściwej haplogrupy.

Rozważając przydatność metody identyfikacji przynależności haplogrupowej opartej na obliczaniu parametru pHM^X należy rozważyć możliwość zastosowania jej do danych innych niż te, na podstawie których wygenerowano tabele częstości alleli w poszczególnych haplogrupach. Podjęto próbę takiej weryfikacji wykorzystując opublikowane haplotypy układów Y-STR: DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 i DYS393 oraz częstości haplogrup I*, R1* i R1a1* dla subpopulacji Polski i Niemiec [7, 11]. Dokonano przypisania haplotypów Y-STR do haplogrup dla danych z populacji Berlina, Freiburga, Gdańska, Greifswaldu, Hamburga, Kolonii, Krakowa, Lipska, Lublina, Magdeburga, Mainz, Muenster, Monachium, Rostocku, Warszawy i Wrocławia. Porównanie rzeczywistych częstości haplogrup I*, R1* i R1a1* w poszczególnych populacjach z częstościami wyliczonymi na podstawie haplotypów w loci DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392 i DYS393 w większości przypadków wykazało różnice istotne

statystycznie. Otrzymany wynik może świadczyć o tym, iż związek poszczególnych haplogrup chromosomu Y z określonymi haplotypami Y-STR może mieć charakter regionalny, kształtowany przez procesy mutacyjne i migracyjne zachodzące na stosunkowo niewielkim obszarze geograficznym. W przypadku haplogrupy R1a1* istnieją np. dane literaturowe wskazujące, że haplotypy Y-STR chromosomów Y należących do tej haplogrupy są odmienne w populacjach Europy wschodniej i zachodniej [12]. Biorąc jednak pod uwagę brak możliwości odtworzenia częstości haplogrup chromosomu Y na podstawie haplotypów Y-STR również w populacjach Polski, o których skądinąd wiadomo, że charakteryzują się wysoką homogennością [7, 13] można wnioskować, że liczba chromosomów analizowanych dla potrzeb niniejszej pracy była zbyt mała, szczególnie w przypadku haplogrupy R1*, aby w pełni oddać złożoność zależności pomiędzy haplotypami Y-STR a ich przynależnością haplogrupową. Istotny dla przedstawionych wyników może być też fakt, iż w badanym zestawie haplotypów populacji Polski i Niemiec nie dysponowano informacją o polimorfizmie locus DYS385, którego allele odgrywają istotną rolę w rozróżnieniu haplogrupy I* od haplogrup R1* i R1a1*. Przedstawione wyżej wyniki należy zatem traktować jako wstępne. Sformułowanie ostatecznych wniosków co do efektywności proponowanej metody kontroli poprawności haplotypów chromosomu Y będzie możliwe po przeanalizowaniu haplotypów minimalnych w znacznie większej liczbie chromosomów o zweryfikowanej przynależności haplogrupowej, w tym również należących do innych, poza badanymi w ramach niniejszej pracy, haplogrup występujących w populacjach Polski oraz w innych światowych populacjach i subpopulacjach.

Podstawowym problemem związanym z wykorzystaniem informacji o przynależności haplogrupowej chromosomu Y może być fakt, iż w większości laboratoriów polimorfizm SNP tego chromosomu oznaczany jest z wykorzystaniem metody RFLP, która ze swej natury wymaga stosunkowo dużych ilości materiału genetycznego, co stanowi poważne utrudnienie w jej rutynowym wykorzystaniu w genetyce sądowej. System weryfikacji poprawności oznaczenia haplotypów chromosomu Y oparty na analizie ich przynależności haplogrupowej wymagałby zatem opracowania czulej, multipleksowej reakcji, pozwalającej na jednoczesne oznaczenie wielu mutacji SNP. Skonstruowanie tego typu reakcji jest, jak się wydaje, możliwe przy wykorzystaniu dostępnych na rynku odczynników [14]. Opracowanie multipleksowego systemu oznaczania mutacji punktowych w loci Y-SNP mogłoby pozwolić na poszerzenie informacji otrzymywanej z próbek analizo-

wanych rutynowo w laboratoriach genetyki sądowej o dane dotyczące przynależności haplogrupowej chromosomu Y. Oznaczanie polimorfizmu Y-SNP oparte jest zwykle na amplifikacji krótkich fragmentów DNA, o długości ok. 100 p.z., zatem system taki mógłby sprawdzić się w analizie materiału zdegradowanego, dającego szczątkowe profile w amplifikacji loci STR. Informacja otrzymana z wykorzystaniem multipleksowego systemu oznaczania haplogrup chromosomu Y mogłaby być wykorzystywana zarówno jako źródło dodatkowych danych przydatnych w dyskryminacji próbek zawierających chromosomy Y pochodzące z różnych linii ojcowskich, jak i w celu weryfikacji otrzymanych profili szczątkowych na podstawie porównania zgodności obserwowanych alleli Y-STR z allelami charakterystycznymi dla haplogrupy, do której należy dany chromosom.

Kolejnym potencjalnym zastosowaniem metod umożliwiających przypisanie haplogrupy do chromosomu na podstawie analizy Y-STR jest weryfikacja danych populacyjnych stanowiących materiał referencyjny. W praktyce genetyki sądowej zdarza się, że konieczne jest określenie częstości występowania danego haplotypu w populacjach innych niż rodzima, niekiedy zamieszkujących odległe obszary geograficzne. Jeśli dla danej populacji dostępne są dane dotyczące częstości poszczególnych haplotypów oraz, w odrębnej publikacji, dane dotyczące częstości poszczególnych haplogrup chromosomu Y, teoretycznie możliwe byłoby zweryfikowanie reprezentatywności próbki populacyjnej wykorzystanej do otrzymania haplotypów Y-STR przez porównanie rzeczywistej częstości haplogrup z obliczoną na podstawie haplotypów Y-STR. Porównanie takie wymagałoby jednak posiadania wiedzy na temat zależności pomiędzy częstościami poszczególnych alleli Y-STR, a przynależnością haplogrupową chromosomów Y w danym regionie geograficznym (np. w populacjach sąsiadujących z badaną) oraz posiadania wiarygodnego źródła danych o częstości poszczególnych haplogrup w danej populacji.

PODSUMOWANIE

Przedstawione w pracy metody przyporządkowywania haplotypów Y-STR do haplogrup wydają się obiecującym narzędziem, które może znaleźć zastosowanie w genetyce sądowej, jednak włączenie tego typu metod do rutynowej praktyki wymaga dobrej znajomości zależności pomiędzy częstością występowania danego allelu a przynależnością grupową chromosomów Y w danym regionie geogra-

ficznym. Koniecznym warunkiem zastosowania tego typu metod w medycynie sądowej jest też opracowanie multipleksowego testu pozwalającego na oznaczanie polimorfizmu SNP, który byłby wystarczająco czuły i wiarygodny, aby znaleźć zastosowanie w badaniach z zakresu genetyki sądowej.

PODZIĘKOWANIA

Autorzy dziękują Paniom: Anecie Jakubowskiej, Ewie Lewandowskiej i Marioli Mrozek za doskonałą asystę techniczną.

Niniejsza praca została sfinansowana z grantu KBN 2PO4C08028.

PIŚMIENNICTWO

1. Bandelt H. J., Lahermo P., Richards M., Macaulay V.: Detecting errors in mtDNA data by phylogenetic analysis, *International Journal of Legal Medicine* 2001, vol. 115, 64-69.
2. Salas A., Bandelt H. J., Macaulay V., Richards M.: Phylogeographic investigations: the role of trees in forensic genetics. *Forensic Sci. Int.* 2006, w druku.
3. Jobling M. A., Hurler M. E., Tyler-Smith C.: *Human Evolutionary Genetics*, 2004, Garland Publishing, New York.
4. Raitio M., Lindroos K., Laukkanen M., Pastinen T., Sistonen P., Sajantila A., Syvänen.: Y-Chromosomal SNPs in Finno-Ugric-Speaking Populations Analyzed by Minisequencing on Microarrays. *Genome Research* 2001, vol. 11, 471-482.
5. Bosch E., Calafell F., Santos F. R. Perez-Lezaun A., Comas D., Benchemsi N., Tyler-Smith C., Bertranpetit J.: Variation in Short Tandem Repeats Is Deeply Structured by Genetic Background on the Human Y Chromosome. *Am. J. Hum. Genet.* 1999, vol. 65, 1623-1638.
6. Roewer L., Krawczak M., Willuweit S., Nagy M., Alves C. i wsp.: Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes. *Forensic Sci Int.* 2001 vol. 118, 106-13.
7. Kayser M., Lao O., Anslinger K.: Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet.* 2005 vol. 117, 428-43.
8. The Y Chromosome Consortium, A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups, *Genome Research* 2002, vol. 12, 339-348.

9. Gill P., Brenner C., Brinkmann B. i wsp.: DNA Commission of the International Society of Forensic Genetics: recommendations on forensic analysis using Y-chromosome STRs, *Forensic Science International* 2001, vol. 124, 5-10.

10. Whit Athey T.: Haplogroup Prediction from Y-STR Values Using an Allele-Frequency Approach. *Journal of Genetic Genealogy* 2005 vol. 1, 1-7.

11. Roewer L., Croucher P. J., Willuweit S., Lu T. T., Kayser M., Lessig R., de Knijff P., Jobling M. A., Tyler-Smith C., Krawczak M.: Signature of recent historical events in the European Y-chromosomal STR haplotype distribution. *Hum Genet.* 2005, vol. 116, 279-91.

12. Passarino G., Cavalleri G. L., Lin A., Cavall-Sforza L. L., Børresen-Dale A., Underhill P. A.: Different genetic components in the Norwegian population revealed by the analysis of mtDNA and Y chromosome polymorphisms. *European Journal of Human Genetics* 2002 vol. 10, 521-529.

13. Płoski R., Woźniak M., Pawłowski R. i wsp.: Homogeneity and distinctiveness of Polish paternal lineages revealed by Y chromosome microsatellite haplotype analysis. *Human Genetics* 2002, vol. 110, 592-600.

14. Grignani P., Peloso G., Achilli A. i wsp.: Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. *Int J Legal Med.* 2006 vol. 120, 151-6.

Adres do korespondencji:

Dr n. med. Marcin Woźniak
Zakład Genetyki Molekularnej i Sądowej
Katedra Medycyny Sądowej CM UMK
ul. Marii Skłodowskiej-Curie 9
85-094 Bydgoszcz
e-mail: marcinw@cm.umk.pl