



Praca oryginalna
Original paper

Agnieszka Gołaszewska, Małgorzata Skawrońska, Anna Niemcunowicz-Janica, Witold Pepiński

Dane populacyjne dla markerów zestawu Yfiler Plus w populacji Polski północno-wschodniej

Y-STR data of the Yfiler Plus panel in population of north-eastern Poland

Zakład Medycyny Sądowej, Wydział Lekarski, Uniwersytet Medyczny w Białymstoku, Polska
Department of Forensic Medicine, Faculty of Medicine, Medical University of Białystok, Poland

Streszczenie

Wykorzystanie nowych możliwości o wysokiej rozdzielczości i identyfikacji sądowej w zakresie badań populacji, oferowanych przez nowe metody multipleksowe (jak Yfiler Plus), ma kluczowe znaczenie w sprawach genetyczno-sądowych. Opracowanie baz danych z częstością haplotypów jest niezbędne do pełnego wykorzystania nowych możliwości w zakresie oznaczania chromosomu Y. **Cel pracy:** Opracowanie bazy haplotypów zestawu Yfiler Plus dla próby populacyjnej 534 mężczyzn z Polski północno-wschodniej oraz obliczenie parametrów przydatności w badaniach genetyczno-sądowych. **Materiały i metody:** Badanie przeprowadzono na próbie populacyjnej 534 niespokrewnionych mężczyzn z obszaru Polski północno-wschodniej z użyciem panelu Yfiler Plus obejmującego 27 markerów zlokalizowanych na chromosomie Y. **Wyniki:** Cztery haplotypy pojawiły się dwukrotnie. Zdolność dyskryminacyjna (DC) całego zestawu wyniosła 0,9925. Najwyższą wartość zróżnicowania genowego (GD) obliczono dla DYS518 (0,86) należącego do markerów szybko mutujących, zaś najniższą GD dla DYS392 (0,42). **Wnioski:** Wyniki wskazują na konieczność prowadzenia dalszych badań i obserwacji zmian, zarówno w różnych regionach Polski, jak i w całej Europie.

Słowa kluczowe: badania populacyjne, Polska północno-wschodnia, Yfiler Plus

Abstract

The use of new high-resolution and forensic identification capabilities for population studies offered by new multiplex methods (such as Yfiler Plus) is crucial in forensic genetics cases. The development of haplotype frequency databases is essential to take full advantage of the new Y chromosome determination capabilities. **Purpose:** Development of the haplotype database of the Yfiler Plus kit for a population-based sample of 534 males from northeastern Poland and calculation of suitability parameters for forensic genetics studies. **Materials and methods:** The study was conducted on a population sample of 534 unrelated males from the area of northeastern Poland using the Yfiler Plus panel of 27 markers located on the Y chromosome. **Results:** Four haplotypes appeared twice. The Discrimination Capacity (DC) of the entire set was 0.9925. The highest Gene Diversity (GD) value was calculated for DYS518 (0.86) belonging to the fast-mutation markers, while the lowest GD was calculated for DYS392 (0.42). **Conclusions:** The results indicate the need for further research and observation of changes, both in different regions of Poland and across Europe.

Key words: population studies, northeastern Poland, Yfiler Plus

1. Wstęp

Tempo rozwoju genetyki molekularnej w ostatnich 30 latach pozwoliło na poszerzenie badań genetycznych w medycynie sądowej poprzez wykorzystanie markerów STR (krótkich powtórzeń tandemowych) chromosomu Y [1,2]. Nierekombinowany region ludzkiego chromosomu Y – NRY (*Nonrecombinant Region of the human Y chromosome*) jest fragmentem genomu przekazywanym z pokolenia na pokolenie w postaci haplotypów identycznych (z wyjątkiem mutacji) dla wszystkich krewnych w linii męskiej. Z uwagi na to, że NRY zajmuje aż 95% chromosomu Y [3], literatura naukowa opisuje go w sposób uproszczony jako „chromosom Y”, co zastosowano i w naszym opracowaniu. Chromosom Y jest przekazywany w linii męskiej, przez co siła dyskryminacji jego markerów jest ograniczona. Z drugiej strony, krótkie powtórzenia tandemowe na chromosomie Y (Y-STR) w tym regionie zyskują na znaczeniu w ustalaniu pokrewieństwa, identyfikacji osób zaginionych, wnioskowaniu o historii populacji, a także do odkrywania powiązań genealogicznych [4,5]. Dodatkowo, badania Y-STR pomagają w identyfikacji mężczyzn czynnie zaangażowanych w przypadki napaści o charakterze seksualnym. W danym przykładzie zawartością próbek mieszanych jest, obok materiału męskiego, DNA kobiety. Markery Y-STR mogą posłużyć w tym przypadku do identyfikacji męskiego sprawcy lub sprawców [6,7]. Mutacje w obrębie *loci* znajdujących się w chromosomie Y występują typowo z częstością

1. Introduction

With the pace of development of molecular genetics in the last 30 years, forensic genetic studies have been expanded through the use of STR (short tandem repeat) markers of the Y chromosome [1,2]. Nonrecombinant region of the human Y chromosome (NRY) is a fragment of the genome passed down from generation to generation in the form of haplotypes which are identical (except for mutations) for all relatives in the male line. Given that NRY occupies as much as 95% of the Y chromosome [3], the scientific literature describes it in a simplified way as the „Y chromosome,” which has also been applied in the present study. The Y chromosome is passed down in the male line, making the discriminatory power of its markers limited. On the other hand, short tandem repeats on the Y chromosome (Y-STR) in this region are gaining importance in establishing kinship, identifying missing persons, inferring population history, and discovering genealogical connections [4,5]. In addition, Y-STR testing helps identify men actively involved in sexual assault cases. The mixed samples in the given example, in addition to the male material contain the female DNA. Y-STR markers may serve in this case for identification of a male offender or offenders [6,7]. Mutations within *loci* present in the Y chromosome typically occur at an average frequency of 2×10^{-3} per generation [8]. New rapidly mutating (RM) markers have also been identified and are of higher mutation frequency than in the baseline of Y-STR markers.

średnio 2×10^{-3} na pokolenie [8]. Zostały odkryte także nowe markery szybko mutujące RM (*Rapidly Mutating*), których częstość mutacji przewyższa podstawową linię markerów Y-STR. Analiza RM Y-STR wskazuje na zwiększoną siłę dyskryminacyjną występującą nie tylko pomiędzy osobnikami niespokrewnionymi, ale także między mężczyznami z tej samej linii [4,6,9]. Daje to możliwość (w pewnych przypadkach) różnicowania w zakresie pokrewieństwa pierwszego stopnia [6,9,10].

Zastosowanie markerów Y-STR w sprawach śledczych wykazały ich możliwości i pomogły w rozwiązaniu niejednego śledztwa. Spowodowało to powszechne wykorzystywanie Y-STR w laboratoriach genetyki sądowej [11]. Na szczególną uwagę zasługuje obszar północno-wschodniej Polski, ze względu na brak dostępnych danych referencyjnych tej części populacji dla rozszerzonej liczby 27 loci [yhrd.org]. Dotychczasowo prowadzone w Polsce badania z wykorzystaniem markerów Y-STR wykazują ich skuteczność, jednak oczywistym jest potrzeba tworzenia większej bazy danych haplotypów Y-STR, w szczególności dla paneli poszerzonych jak np. Yfiler Plus PCR Amplification Kit (Thermo Fisher Scientific) [11]. Pełna kombinacja 27 markerów Y-STR zawarta w zestawie Yfiler Plus nie została jeszcze wszechstronnie przetestowana na populacji północno-wschodniej Polski ani nie przeprowadzono dotychczas badań z użyciem takiej liczby markerów w pojedynczym zestawie dla tak obszernej grupy badawczej w tej części regionu. Dzięki zwiększonej liczbie uwzględnionych markerów i dodaniu siedmiu loci szybko mutujących w zestawie Yfiler Plus, oczekuje się, że ten multiplex może znacznie zwiększyć zdolność dyskryminacji w populacji północno-wschodniej Polski.

Niewątpliwie istnieje potrzeba powiększenia bazy danych o haplotypy Y-STR, biorąc pod uwagę wzrastające po pandemii COVID-19 statystyki gwałtów w Polsce. Łącznie w roku 2021 w kategorii zgwałcenia zanotowano 1081 przestępstw stwierdzonych (przestępstwo, w którym potwierdzono zaistnienie czynu zabronionego), z czego 955 to przestępstwa wykryte (przestępstwo stwierdzone, w którym ustalono przynajmniej jednego podejrzanego), co daje 88% wykrycia (wskaźnik wykrywalności przestępstw). Dane pochodzą ze statystyk prowadzonych przez policję w Polsce [12].

RM Y-STR analysis indicates increased discriminatory power not only between unrelated individuals, but also between males of the same lineage [4,6,9]. It gives the possibility (in some cases) of differentiation in terms of first-degree kinship [6,9,10].

Application of Y-STR markers in investigation cases proved their capabilities and helped solve numerous cases. This has led to the widespread use of Y-STR in laboratories of forensic genetics [11]. The area of northeastern Poland deserves special attention due to the lack of available reference data for this part of the population for an expanded number of 27 loci [yhrd.org]. To date, studies conducted in Poland with the use of Y-STR markers demonstrate their efficacy, however, it is clear that there is a need to create a larger database of Y-STR haplotypes, especially for expanded panels, e.g. Yfiler Plus PCR Amplification Kit (Thermo Fisher Scientific) [11]. The complete combination of 27 Y-STR markers included in the Yfiler Plus kit has not been up to now comprehensively tested on the population of north-eastern Poland. Owing to the increased number of markers included and the addition of seven rapidly mutating loci in the Yfiler Plus kit it is expected that this multiplex can significantly increase the discriminatory power in the population of northeastern Poland.

Undoubtedly, there is a need to expand the database to include Y-STR haplotypes, given the increasing rape statistics in Poland after the COVID-19 pandemic. A total of 1,081 ascertained crimes (in which a criminal act has been confirmed) were recorded in the category of rape in 2021, of which 955 were detected crimes (an ascertained crime in which at least one suspect is identified), resulting in an 88% detection rate (crime detection indicator). The data is derived from statistics kept by the Police in Poland [12].

2. Cel

Celem badania jest opracowanie populacyjnej bazy danych w zakresie 27 *loci* Y-STR zestawu Yfiler Plus dla próby populacyjnej 534 mężczyzn z Polski północno-wschodniej oraz ocena parametrów przydatności w badaniach genetyczno-sądowych. Uzyskane dane mogą przyczynić się do zbudowania kompleksowej referencyjnej bazy danych haplotypów Y-STR dla populacji Polski oraz posłużą do porównania populacji północno-wschodniej części Polski z innymi populacjami, które zostały wcześniej oznaczone za pomocą tego samego zestawu markerów.

3. Materiały i metody

Yfiler Plus PCR Amplification Kit jest multiplexowym zestawem opartym na sześciokolorowej detekcji, zaprojektowanym w celu optymalnej wydajności, a także by w krótszym czasie uzyskać dokładne wyniki dla wymagających próbek. W jego skład wchodzi 27 markerów Y-STR (17 markerów AmpFISTR Yfiler oraz 10 dodatkowych), w tym 7 markerów szybko mutujących z częstością powyżej 1×10^{-2} . Markery te mają za zadanie pomóc w poprawieniu rozdzielczości linii ojcowskiej, w tym rozróżnieniu blisko spokrewnionych mężczyzn. Zestaw Yfiler Plus jest zatwierdzony przez Krajowy Indeks Danych DNA USA – NDIS (*National DNA Index System*).

Wymazy nabłonka policzków lub próbki krwi zasuszonej na bibule uzyskano anonimowo od 534 niespokrewnionych ochotników zamieszkujących północno-wschodni region Polski, wraz z informacją o miejscu urodzenia i pochodzeniu etnicznym dawcy. Podpisane świadome zgody uzyskano od wszystkich uczestników. Badanie to było zgodne z protokołem zatwierdzonym przez Komisję Bietyczną UMB (Ref: APK.002.40.2021). Ekstrakcję genomowego DNA przeprowadzono przy użyciu GeneMATRIX Bio-Trace DNA Purification Kit (EURx) zgodnie z instrukcją producenta. Ekstrakty DNA oznaczano ilościowo metodą PCR w czasie rzeczywistym przy użyciu zestawu Quantifiler Human DNA Quantification Kit i termocyklera 7500 Real Time PCR System (Applied Biosystems) zgodnie z instrukcją producenta. Zastosowano warunki PCR zgodnie z protokołem zalecanym przez producenta zestawu Yfiler Plus w aparacie GeneAmp

2. Aim

The aim of the present study is to provide with a dataset within the range of 27 Y-STR *loci* of the Yfiler Plus kit for a population sample of 534 males from northeastern Poland and evaluation of the parameters of usefulness in genetic and forensic studies. The data obtained may contribute to the construction of a comprehensive reference database of Y-STR haplotypes for the Polish population, and will be used to compare the northeastern Polish population with other populations that have been previously identified using the same set of markers.

3. Material and methods

Yfiler Plus PCR Amplification Kit is a multiplex set based on six-color detection designed for optimal efficiency and also in order to obtain accurate results for challenging samples in a shorter period of time. It includes 27 Y-STR markers (17 AmpFISTR Yfiler markers and 10 additional ones), of which seven are rapidly mutating markers with frequency above 1×10^{-2} . These markers are designed to help improve paternal line resolution, including differentiation of closely related males. Yfiler Plus kit is approved by the US National DNA Index System (NDIS).

Buccal epithelial swabs or dried blood samples on blotting paper were obtained anonymously from 534 unrelated volunteers residing in the northeastern region of Poland, along with the information on the donor's place of birth and ethnicity. Signed informed consents were obtained from all participants. The present study followed a protocol approved by the Bioethics Committee of the MUB (Ref: APK.002.5.2020). Genomic DNA extraction was performed with the GeneMATRIX Bio-Trace DNA Purification Kit (EURx) according to the manufacturer's guidelines. DNA extracts were determined quantitatively by real-time PCR method using the Quantifiler Human DNA Quantification Kit and the 7500 Real Time PCR System thermocycler (Applied Biosystems) according to the manufacturer's guidelines. PCR conditions were used according to the protocol recommended by the manufacturer of the Yfiler Plus kit in a GeneAmp PCR System 9700 unit (Applied Biosystems). Ampli-

PCR System 9700 (Applied Biosystems). Amplifikację przeprowadzono w 29 cyklach PCR. Wszystkie testy przeprowadzono w obecności kontroli pozytywnej (DNA Control 007) i kontroli negatywnej (test odczynnikowy). Elektroforezę kapilarną i genotypowanie przeprowadzono w sekwenatorze Genetic Analyzer 3500 (Applied Biosystems) przy użyciu oprogramowania GeneMapper® ID-X v.1.5 w oparciu o standard wielkości GS600-LIZ v.2.

Analizę statystyczną, opartą na 534 wygenerowanych profilach, przeprowadzono przy użyciu programu STRAF 2.1.5: *STR Analysis for Forensics* [13]. Częstości alleli DYF387S1, DYS385 i DYS389 obliczono metodą bezpośredniego zliczania. Częstości haplotypów obliczono poprzez bezpośrednie zliczanie. Określono liczbę singletonów, czyli haplotypów zaobserwowanych tylko raz w badanej próbce. Zdolność do dyskryminacji DC (*Discrimination Capacity*) obliczono jako liczbę różnych haplotypów podzieloną przez całkowitą liczbę próbek (pomijając wielokrotność powtarzalnych haplotypów). Prawdopodobieństwo zgodności haplotypu HMP (*Haplotype Match Probability*) wyliczono ze wzoru

$$HMP = \sum p_i^2,$$

gdzie p_i to częstość haplotypu [14]. Zróżnicowanie genowe GD (*Gene Diversity*) w markerach dwuallelicznych DYF387S1, DYS385 i DYS389 zostało wyliczone bezpośrednio, korzystając ze wzoru

$$GD = \left(\frac{N}{N-1} \right) \left(1 - \sum_{i=1}^N (p_i)^2 \right),$$

gdzie N i p_i to odpowiednio liczebność próbki i częstość alleli. Obliczenia zawartości informacji o polimorfizmie PIC (*Polymorphism Information Content*) oparto na wzorze,

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2,$$

gdzie p_i i p_j to częstości alleli. Obliczenia prawdopodobieństwa przypadkowej zgodności MP (*Match Probability*) oparto na wzorze

$$MP = \sum_i (G_i)^2,$$

gdzie G_i jest częstością występowania genotypu i w danym locus w populacji. Siła dyskryminacji

była przeprowadzona w 29 PCR cyklach. Wszystkie testy były przeprowadzane w obecności kontroli pozytywnej (DNA Control 007) i kontroli negatywnej (reagent test). Capillary electrophoresis and genotyping were performed on the Genetic Analyzer 3500 sequencer (Applied Biosystems) using GeneMapper® ID-X v.1.5 software based on the GS600-LIZ v.2 size standard.

Statistical analysis based on 534 generated profiles was conducted with the use of STRAF 2.1.5: *STR Analysis for Forensics* [13]. Haplotype frequencies and DYF387S1, DYS385 and DYS389 allele frequencies were calculated using the direct counting method. The number of singletons, that is, haplotypes observed only once in the sample, was determined. Discrimination capacity (DC) was calculated as the number of different haplotypes divided by the total number of samples (ignoring multiple repeat haplotypes). Haplotype match probability (HMP) was calculated from the formula

$$HMP = \sum p_i^2,$$

where p_i is the frequency of the haplotype [14]. Gene diversity (GD) in diallelic markers DYF387S1, DYS385 and DYS389 was calculated directly using the formula

$$GD = \left(\frac{N}{N-1} \right) \left(1 - \sum_{i=1}^N (p_i)^2 \right),$$

where N and p_i are sample size and allele frequency, respectively. Calculations of polymorphism information content (PIC) were based on the formula

$$PIC = 1 - \sum_{i=1}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2,$$

where p_i and p_j are allele frequencies. Calculations of match probability (MP) were based on the formula,

$$MP = \sum_i (G_i)^2,$$

where G_i is frequency of genotype in a given locus within the population. Power of discrimination (PD) results from the formula $PD = 1 - MP$. Evaluation of genetic distance between populations based on Rst coefficient calculations was performed using analysis of molecular variance (AMOVA). Statistical sig-

PD (*Power Of Discrimination*) wynika ze wzoru $PD = 1 - MP$. Ocena dystansu genetycznego pomiędzy populacjami w oparciu o obliczenia współczynnika R_{st} została wykonana z zastosowaniem analizy wariancji molekularnej (AMOVA). Istotność statystyczną sprawdzono na podstawie 10000 permutacji. Analizą skalowania wielowymiarowego (MDS) posłużono się w celu graficznego przedstawienia dystansów genetycznych między wybranymi populacjami na podstawie zróżnicowania rozkładów haplotypów Y-STR. Na potrzeby obu analiz posługiwano się narzędziami udostępnionymi poprzez stronę internetową bazy YHRD (<https://yhrd.org/>, dostęp 07.03.2023 r.).

4. Wyniki

Częstości alleli markerów Yfiler Plus przedstawiono w Materiale uzupełniającym. Wyniki analiz statystycznych opisanych w tych rozdziale przedstawiono w Tabeli I. W badanej próbie populacyjnej, wśród 534 przeanalizowanych haplotypów ujawniono 526 singletonów (98,5%). Cztery haplotypy pojawiły się dwukrotnie. Ten wysoki odsetek unikalnych haplotypów wpływa korzystnie na wysokość parametrów przydatności w badaniach genetyczno-sądowych w badanej populacji Polski (HMP = 0,002, DC = 0,9925). Największa liczba alleli przypadająca na marker (15) wystąpiła w *locus* DYS481, natomiast najmniejszą (4) w *locus* DYS391, co potwierdza wcześniejsze obserwacje zanotowane w ogólnosiątkowym badaniu różnorodności haplotypów chromosomu Y [15]. Wartości zróżnicowania genowego (GD) przekraczają 0,5 dla większości markerów oprócz DYS392, DYS393, DYS437 i DYS448, gdzie GD zawierało się między 0,420 a 0,481. GD przekracza 0,6 dla 19 markerów (70,1%), a nawet 0,7 dla 13 markerów (48,1%). Warto zauważyć, że w trzech (DYS449, DYS518 i DYS627) spośród siedmiu markerów szybko mutujących, GD przekracza 0,84, zaś w pozostałych 0,76. Najwyższą wartość GD obliczono dla DYS518 (0,86) należącym do grupy RM, zaś najniższą (0,42) dla DYS392 (Tabela I). Należy zaznaczyć, że przy porównywaniu wartości GD z zestawu Yfiler Plus, celowo zostały pominięte trzy markery wielokopijne (DYF387S1a/b, DYS385a/b oraz DYS389 I i II). Wartość PIC zawierała się w zakresie od 0,389 (DYS518, RM) do 0,841 (DYS393).

nificance was checked based on 10000 permutations. Multidimensional scaling analysis (MDS) was used to graphically represent the genetic distances between selected populations based on the variation of Y-STR haplotype distributions. For both analyses, tools provided through the YHRD database website were used (<https://yhrd.org/>, access on 07.03.2023).

4. Results

The allele frequencies of the Yfiler Plus markers are shown in the Supplementary Material. The results of the statistical analyses described in this chapter are shown in Table I. In the studied population sample, among the 534 haplotypes analyzed, 526 singletons (98.5%) were revealed. Four haplotypes appeared twice. This high percentage of unique haplotypes positively influences the amount of parameters of usefulness in genetic and forensic studies in the Polish population studied (HMP = 0.002, DC = 0.9925). The highest number of alleles per marker (15) occurred in the DYS481 *locus*, while the lowest number (4) occurred in the DYS391 *locus*, which confirms previous observations noted in a worldwide survey of Y chromosome haplotype diversity [15]. Gene diversity (GD) values exceeded 0.5 for most markers except DYS392, DYS393, DYS437 and DYS448, where GD ranged between 0.420 and 0.481. GD exceeded 0.6 for 19 markers (70.1%), and even 0.7 for 13 markers (48.1%). It is worth noting that in three (DYS449, DYS518 and DYS627) out of seven rapidly mutating markers, GD exceeded 0.84, while for the others 0.76. The highest GD value was calculated for DYS518 (0.86) belonging to the RM group, whereas the lowest (0.42) for DYS392 (Table I). It should be noted that when comparing GD values from the Yfiler Plus kit, three multi-copy markers (DYF387S1a/b, DYS385a/b, and DYS389 I, II) were deliberately omitted. The value of PIC ranged from 0.389 (DYS518, RM) to 0.841 (DYS393).

Piśmiennictwo z danymi uzyskanymi w poszukiwaniu zmienności populacyjnej regionów Polski nie dostarcza dowodów występowania statystycznie istotnych różnic między nimi. Oznacza to, że obecna populacja Polski jest homogenna w zakresie rozkładów haplotypów Y-STR [16].

The literature containing data obtained in the search for population variation in the regions of Poland fails to provide evidence for statistically significant differences between them. This means that the current Polish population is homogeneous in terms of Y-STR haplotype distributions [16].

Tabela I. Parametry biostatystyczne *loci* zestawu Yfiler Plus dla próby populacyjnej 534 mężczyzn z Polski północno-wschodniej

Table I. Biostatistical parameters of the Yfiler Plus *loci* for a population sample of 534 males from northeastern Poland

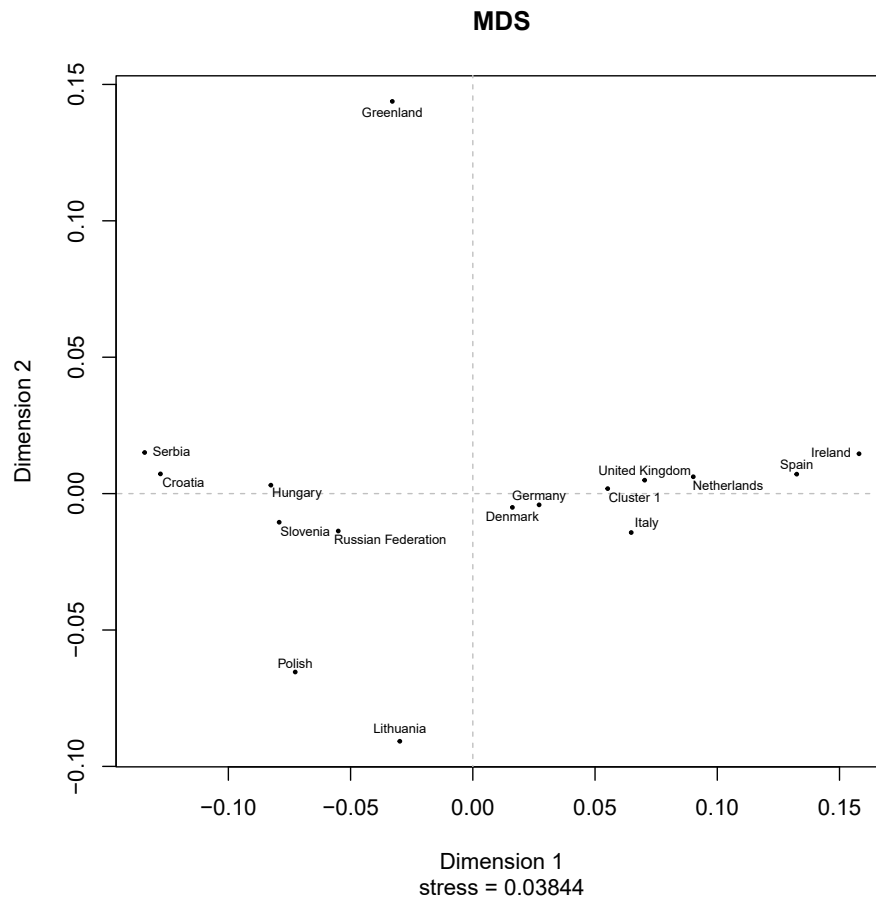
locus	Nall	GD	PIC	MP	PD
DYF387S1*	9	0,781	0,766	0,220	0,780
DYS19	6	0,744	0,699	0,257	0,743
DYS385	13	0,778	0,772	0,223	0,777
DYS389	8	0,633	0,568	0,368	0,632
DYS390	7	0,675	0,623	0,327	0,673
DYS391	4	0,546	0,442	0,456	0,544
DYS392	9	0,420	0,393	0,581	0,419
DYS393	6	0,425	0,389	0,576	0,424
DYS437	5	0,456	0,403	0,545	0,455
DYS438	7	0,611	0,546	0,390	0,610
DYS439	7	0,700	0,646	0,302	0,698
DYS448	8	0,481	0,435	0,520	0,480
DYS449*	12	0,841	0,820	0,161	0,839
DYS456	8	0,747	0,705	0,254	0,746
DYS458	11	0,736	0,687	0,266	0,734
DYS460	6	0,572	0,507	0,429	0,571
DYS481	15	0,843	0,823	0,158	0,842
DYS518*	13	0,859	0,841	0,143	0,857
DYS533	6	0,545	0,494	0,456	0,544
DYS570*	11	0,784	0,749	0,218	0,782
DYS576*	12	0,764	0,729	0,238	0,762
DYS627	13	0,845	0,826	0,157	0,843
DYS635	7	0,599	0,565	0,402	0,598
YGATAH4	5	0,633	0,570	0,369	0,631

Legenda: Nall – liczba obserwowanych alleli w markerze, GD (*Gene Diversity*) – współczynnik zróżnicowania genowego/haplotypowego, PIC (*Polymorphism Information Content*) – wskaźnik informacji o polimorfizmie, MP (*Match Probability*) – prawdopodobieństwo dopasowania, PD (*Power Of Discrimination*) – siła dyskryminacji.

* markery RM (szybkomutujące)

Miarą zróżnicowania między populacjami oraz ich wzajemnych powiązań filogenetycznych jest dystans genetyczny. Wskaźnik R_{ST} jest parametrem dystansów genetycznych między parami populacji, który uwzględnia różnice występujące w wielkości (liczbie powtórzeń) alleli mikrosatelitarnych. W celu zbadania powiązań genetycznych rozkładów haplotypów Y-STR między badaną populacją a wybranymi populacjami europejskimi wykonano analizę skalowania wielowymiarowego na podstawie dystansów genetycznych R_{ST} . Jak przedstawiono na Rycinie 1, mieszkańcy Irlandii [17], Hiszpanii [18], Holandii [19], Wielkiej Brytanii [20], Włoch [9,21], oraz zgrupowani w Klaster 1 mieszkańcy Belgii [15,22], Austrii [23,24] i Szwajcarii [25], znajdują się w znacznej odległości genetycznej od populacji polskiej. W bliższej odległości znajdują się populacje

A measure of the differentiation between populations and their phylogenetic interrelationships is genetic distance. The R_{ST} coefficient is a parameter of genetic distances between pairs of populations that accounts for differences that occur in the size (number of repeats) of microsatellite alleles. In order to examine the genetic associations of Y-STR haplotype distributions between the population under study and selected European populations, a multidimensional scaling analysis was performed based on R_{ST} genetic distances. As shown in Figure 1, residents of Ireland [17], Spain [18], the Netherlands [19], the United Kingdom [20], Italy [9,21], and, grouped into Cluster 1, residents of Belgium [15,22], Austria [23,24] and Switzerland [25], are at a considerable genetic distance from the Polish population. Populations of Serbia [26], Croatia [27], Germany [28],



Rycina 1. Wykres skalowania wielowymiarowego (MDS) populacji polskiej i 17 innych populacji referencyjnych na podstawie odległości R_{ST} .

Figure 1. Multi-dimensional scaling (MDS) plot of the Polish population and 17 other reference populations based on R_{ST} distances

Serbii [26], Chorwacji [27], Niemiec [28], Danii [29], Węgier [30], Słowenii [31] oraz Rosyjskiej Federacji [32,33]. Natomiast Grenlandia [34] jest umieszczona daleko poza bliski obręb reszty przedstawionych na wykresie MDS populacji, wyraźnie wyróżniając się na tle pozostałych grup. Najbliżej populacji polskiej znajduje się populacja litewska [35], co może wynikać ze wspólnej historii państw, które jeszcze w XVIII wieku tworzyły jedno połączone państwo, a część obecnych terenów litewskich przez dłuższy czas należała do Polski [36,37].

5. Omówienie i wnioski

Różne regiony Polski były sukcesywnie analizowane w zakresie zmienności genetycznej chromosomu Y. Oddzielnie przeprowadzono analizę mniejszości narodowościowych przybyłych z sąsiadujących Państw czyli mniejszość litewską, białoruską oraz mniejszość narodową pozbawioną własnego terytorium, jaką jest mniejszość tatarska [38,39,40,41]. Zaprezentowane tu wyniki badań dostarczają informacji o częstości haplotypów zestawu Yfiler Plus, przyczyniając się do budowy kompleksowej referencyjnej bazy danych populacji Polski, która posłuży szacowaniu wartości dowodowej wyników genotypowania Y-STR, jak również będzie stanowić odnośnik w analizach międzypopulacyjnych stanowiąc wartość dla badań z zakresu genetyki sądowej oraz genetyki populacji.

Finansowanie

Badanie było sfinansowane z subwencji dla pracowników oraz doktorantów Uniwersytetu Medycznego w Białymstoku (SUB/1/DN/21/001/1121 UMB).

Materiał uzupełniający w wersji elektronicznej: Częstości alleli markerów Yfiler Plus (Mat. uzup. Tabela I), Haplotypy [534 samples] (Mat. uzup. Tabela II).

Denmark [29], Hungary [30], Slovenia [31], and Russian Federation [32,33] are located in closer distances. In contrast, Greenland [34] is placed far beyond the rest of the populations shown in the MDS chart, clearly standing out from the rest of the groups. The Lithuanian population is most closely related to the Polish population [35], which may be due to the common history of the states that formed a union as recently as in the 18th century, and part of the current Lithuanian territories belonged to Poland for a long time [36,37].

5. Discussion and conclusions

Various regions of Poland have been analyzed consecutively in terms of genetic variation of the Y chromosome. A separate analysis has been conducted on ethnic minorities originating from neighboring countries, e.g. Lithuanian and Belarusian minorities, and the ethnic group devoid of its own territory, which is the Tatar minority [38,39,40,41]. The research results presented here provide information on Yfiler Plus kit haplotype frequencies contributing to the development of complex reference database of the Polish population, which will contribute in estimating evidential value of Y-STR results and provide with a referential dataset for interpopulation analyses.

Funding

The study was supported by grants for employees and doctoral students of the Medical University of Białystok (SUB/1/DN/21/001/1121 UMB).

Supplementary Material in the electronic version: The allele frequencies of the Yfiler Plus markers (Suppl. Table I), The haplotypes [534 samples] (Suppl. Table II).

Piśmiennictwo / References

1. Chakraborty R, Stivers DN, Su B, Zhong Y, Budowle B. The utility of short tandem repeat loci beyond human identification: Implications for development of new DNA typing systems. *Electrophoresis* 1999; 20: 1682-1696.
2. Abreu-Głowacka M, Żaba Cz, Koralewska-Kordel M, Michalak E, Przybylski Z. Polish population data for 17 Y-STRs and 8 Y-SNPs markers. *Arch Med Sąd Krym* 2013; 58: 201-215.
3. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, et al. A physical map of the human Y chromosome. *Nature* 2001; 409: 943-945.
4. Ballantyne KN, Goedbloed M, Fang R, Schaap O, Lao O, Wollstein A, Choi Y, van Duijn K, Vermeulen M, Brauer S, et al. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am J Hum Genet* 2010; 87: 341-53.
5. Mertoglu E, Filoglu G, Zorlu T, Bulbul O. Estimation of the Y-chromosomal short tandem repeat (Y-STR) mutation rates in Turkey. *Turk J Biochem* 2018; 43: 142-150.
6. Ballantyne KN, Keerl V, Wollstein A, Choi Y, Zuniga SB, Ralf A, Vermeulen M, de Knijff P, Kayser M. A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. *Forensic Sci Int Genet* 2012; 6: 208-18.
7. Gopinath S, Zhong C, Nguyen V, Ge J, Lagacé RE, Short ML, Mulero JJ. Developmental validation of the Yfiler(®) Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. *Forensic Sci Int Genet* 2016; 24: 164-175.
8. De Knijff P. Messages through Bottlenecks: On the Combined Use of Slow and Fast Evolving Polymorphic Markers on the Human Y Chromosome. *Am J Hum Genet* 2000, 67: 1055-1061.
9. Rapone C, D'Atanasio E, Agostino A, Mariano M, Papaluca MT, Cruciani F, Berti A. Forensic genetic value of a 27 Y-STR loci multiplex (Yfiler(®) Plus kit) in an Italian population sample. *Forensic Sci Int Genet* 2016; 21: e1-5.
10. Robino C, Ralf A, Pasino S, De Marchi MR, Ballantyne KN, Barbaro A, Bini C, Carnevali E, Casarino L, Di Gaetano C, et al. Development of an Italian RM Y-STR haplotype database: Results of the 2013 GEFI collaborative exercise. *Forensic Sci Int Genet* 2015; 15: 56-63. Erratum in: *Forensic Sci Int Genet* 2018 Feb 7; Erratum in: *Forensic Sci Int Genet* 2018 Apr 4.
11. Rębała K, Branicki W, Pawłowski R, Spólnicka M, Kupiec T, Parys-Proszek A, Woźniak M, Grzybowski T, Boroń M, Wróbel M, et al. Recommendations of the Polish Speaking Working Group of the International Society for Forensic Genetics on forensic Y chromosome typing. *Arch Med Sadowej Kryminol* 2020; 70: 1-18.
12. <https://statystyka.policja.pl/st/przestepstwa-ogolem/przestepstwa-kryminalne/zgwalczenie/122293,Zgwalczenie.html> – Pliki do pobrania (XLSX, DOCX). Data dostępu: 13.12.2022
13. Gouy A, Zieger M. STRAF-A convenient online tool for STR data evaluation in forensic genetics. *Forensic Sci Int Genet* 2017; 30: 148-151.
14. Ashirbekov Y, Sabitov Z, Aidarov B, Abaildayev A, Junissova Z, Cherusheva A, Saidamarova VV, Sharipov K, Ramankulov Y, Zhabagin M. Genetic Polymorphism of 27 Y-STR Loci in the Western Kazakh Tribes from Kazakhstan and Karakalpakstan, Uzbekistan. *Genes* 2022; 13: 1826. <https://doi.org/10.3390/genes13101826>.
15. Purps J, Siegert S, Willuweit S, Nagy M, Alves C, Salazar R, Angustia SM, Santos LH, Anslinger K, Bayer B, et al. A global analysis of Y-chromosomal haplotype diversity for 23 STR loci. *Forensic Sci Int Genet* 2014; 12: 12-23.
16. Abreu-Głowacka M, Pepinski W, Michalak E, Konarzewska M, Zak K, Skawronska M, Niemcunowicz-Janica A, Soltyszewski I, Krajewski P, Zaba C. Population Genetic Data of 30 Insertion-Deletion Markers in the Polish Population. *Genes*. 2022; 13: 1683. <https://doi.org/10.3390/genes13101683>.
17. Aliferi A, Thomson J, McDonald A, Molin V, Ferguson S, Vanhinsbergh D, Syndercombe D, Ballard D. UK and Irish Y-STR population data — A catalogue of variant alleles. *For Sci Int Genetics* 2018; 34: e1-6.
18. García O, Yurrebaso I, Mancisidor ID, López S, Alonso S, Gusmão L. Data for 27 Y-chromosome STR loci in the Basque Country autochthonous population. *Forensic Sci Int Genet* 2016; 20: e10-e12.
19. Westen AA, Kraaijenbrink T, Clarisse L, Grol LJ, Willemse P, Zuniga SB, De R, Schouten R, Der V, Weiler NE et al. Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males. *Forensic Sci Int Genet* 2015; 14: 174-81.
20. Huszar TI, Bodmer WF, Hutnik K, Wetton JH, Jobling MA. Sequencing of autosomal, mitochondrial and Y-chromosomal forensic markers in the People of the British Isles cohort detects population structure dominated by patrilineages. *Forensic Sci Int Genet*. 2022; 59: 102725. doi:10.1016/j.fsigen.2022.102725.
21. Lacerenza D, Aneli S, Di C, Critelli R, Piazza A, Matullo G, Culigioni C, Robledo R, Robino C, Calò C. Investigation of extended Y chromosome STR haplotypes in Sardinia. *Forensic Sci Int Genet* 2017; 27: 172-174.
22. Claerhout S, Roelens J, Der V, Verstraete P, Larmuseau MHD, Decorte R. Ysurnames? The patrilineal Y-chromosome and surname correlation for DNA kinship research. *For Sci Int Genet* 2020; 44: 102204.
23. Pickrahn I, Müller E, Zahrer W, Dunkelmann B, Cemper-Kiesslich J, Kreindl G, Neuhuber F. Yfiler(®) Plus amplification kit validation and calculation of forensic parameters for two Austrian populations. *Forensic Sci Int Genet* 2016; 21: 90-94.
24. Niederstätter H, Berger B, Kayser M, Parson W. Differences in urbanization degree and consequences on the diversity of conventional vs. rapidly mutating Y-STRs in five municipalities from a small region of the Tyrolean Alps in Austria. *Forensic Sci Int Genet* 2016; 24: 180-193.

25. Zieger M, Utz S. The Y-chromosomal haplotype and haplogroup distribution of modern Switzerland still reflects the alpine divide as a geographical barrier for human migration. *For Sci Int Genet* 2020; 48: 102345.
26. Zgonjanin D, Alghafri R, Antov M, Stojiljković G, Petković S, Vuković R, Drašković D. Genetic characterization of 27 Y-STR loci with the Yfiler® Plus kit in the population of Serbia. *Forensic Sci Int Genet* 2017; 31: e48-49.
27. Primorac D, Škaro V, Projić P, Missoni S, Zanki H, Merkaš S, Šarac J, Novokmet N, Ledić A, Makar A et al. Croatian genetic heritage: an updated Y-chromosome story. *Croat Med J* 2022; 63: 273-286.
28. Kayser M, Lao O, Anslinger K, Augustin C, Bargel G, Edelmann J, Elias S, Heinrich M, Henke J, Henke L et al. Significant genetic differentiation between Poland and Germany follows present-day political borders, as revealed by Y-chromosome analysis. *Hum Genet* 2005; 117: 428-43.
29. Olofsson JK, Mogensen HS, Buchard A, Børsting C, Morling N. Forensic and population genetic analyses of Danes, Greenlanders and Somalis typed with the Yfiler® Plus PCR amplification kit. *Forensic Sci Int Genet* 2015; 16: 232-236.
30. Pamjav H, Fóthi Á, Fehér T, Fóthi E. A study of the Bodrogeköz population in north-eastern Hungary by Y chromosomal haplotypes and haplogroups. *Mol Genet Genomics*. 2017; 292: 883-894. doi:10.1007/s00438-017-1319-z.
31. Sterlinko H, Pajnic IZ, Balazic J, Komel R. Human Y-specific STR haplotypes in a Slovenian population sample. *Forensic Sci Int*. 2001; 120: 226-228. doi:10.1016/s0379-0738(01)00390-5.
32. Woźniak M, Derenko M, Malyarchuk B, Dambueva I, Grzybowski T, Miścicka-Sliwka D. Allelic and haplotypic frequencies at 11 Y-STR loci in Buryats from South-East Siberia. *Forensic Sci Int*. 2006; 164: 271-275. doi:10.1016/j.forsciint.2005.11.023.
33. Semikhodskii A, Krassotkin Y, Makarova T, Zavarin V, Ilina V, Sutyagina D. Population genetic data and forensic parameters of the 27 Y-STR panel Yfiler® Plus in Russian population [published correction appears in *Int J Legal Med*. 2021 May 6]. *Int J Legal Med* 2021; 135: 1785-1787. doi:10.1007/s00414-021-02599-8.
34. Hallenberg C, Tomas C, Simonsen B, Morling N. Y-chromosome STR haplotypes in males from Greenland. *Forensic Sci Int Genet* 2009; 3: e145-e146. doi:10.1016/j.fsigen.2008.11.012.
35. Jankauskiene J, Kukiene J, Ivanova V, Aleknaviciute G. Population data and forensic genetic evaluation with the Yfiler™ Plus PCR Amplification kit in the Lithuanian population. *Forensic Sci Int Genet Suppl Ser* 2017; 6: e606-607.
36. Szołtysek M. Rethinking Eastern Europe: Household-formation patterns in the Polish-Lithuanian Commonwealth and European family systems. *Contin Change* 2008; 23: 389-427. doi:10.1017/S0268416008006929.
37. Górska-Szymczak J, Górski G. The Federal Structure of the Polish-Lithuanian Commonwealth in the 16-18 Centuries. Confrontation and Cooperation: 1000 Years of Polish-German-Russian Relations. 2019; 5: 21-28. DOI: 10.2478/conc-2019-0003.
38. Pepinski W, Niemcunowicz-Janica A, Skawronska M, Koc-Zorawska E, Janica J, Soltyszewski I. Allele distribution of 15 STR loci in a population sample of the Lithuanian minority residing in the Northeastern Poland. *Forensic Sci Int* 2004; 144: 65-67. doi:10.1016/j.forsciint.2004.01.023.
39. Pepinski W, Niemcunowicz-Janica A, Skawronska M, Janica J, Koc-Zorawska E, Aleksandrowicz-Bukin M, Soltyszewski I. Genetic data on 15 STR loci in the ethnic group of Polish Tatars residing in the area of Podlasie (Northeastern Poland). *Forensic Sci Int* 2005; 149: 263-265. doi:10.1016/j.forsciint.2004.07.009.
40. Janica J, Pepiński W, Niemcunowicz-Janica A, Skawrońska M, Soltyszewski I, Berent J. Polimorfizm loci Y-STR wśród ludności Polski północno-wschodniej w aspektach zróżnicowania etnicznego i przydatności w badaniach medyczno-sądowych [Ethnic variation and forensic usefulness of Y-STR loci in inhabitants of northeastern Poland]. *Arch Med Sądowej Kryminol* 2008; 58: 17-21.
41. Janica J, Pepinski W, Niemcunowicz-Janica A, Skawronska M, Aleksandrowicz-Bukin M, Ptaszynska-Sarosiek I, Koc-Zorawska E. Y-chromosome STR haplotypes and alleles in the ethnic group of Polish Tatars residing in the Northeastern Poland. *Forensic Sci Int* 2005; 150: 91-95. doi:10.1016/j.forsciint.2004.08.012.

ORCID

Agnieszka Gołaszewska – 0000-0002-1100-9188
Małgorzata Skawrońska – 0000-0002-1237-4240
Anna Niemcunowicz-Janica – 0000-0003-2807-8312
Witold Pepiński – 0000-0003-1244-2247

AUTOR DO KORESPONDENCJI CORRESPONDING AUTHOR

Mgr Agnieszka Gołaszewska
Zakład Medycyny Sądowej, Wydział Lekarski
Uniwersytet Medyczny w Białymstoku
ul. Jerzego Waszyngtona 13
15-269 Białystok, Polska
e-mail: agnieszka11150@wp.pl

