

mgr inż. Krzysztof Żak

**Optymalizacja procesu identyfikacji genetycznej NN zwłok
i szczątków ludzkich.**

**Rozprawa na stopień doktora nauk medycznych i nauk o zdrowiu
w dyscyplinie nauki medyczne**

Promotor: dr hab. n. med. Ireneusz Sołtyszewski

Promotor pomocniczy: dr n. med. Maria Szargut

Katedra i Zakład Medycyny Sądowej



Obrona rozprawy doktorskiej przed Radą Dyscypliny Nauk Medycznych
Warszawskiego Uniwersytetu Medycznego

Warszawa 2025 r.

Słowa kluczowe:

Genetyka sądowa, markery identyfikacyjne, identyfikacja genetyczna, symulacje pokrewieństw, analiza biostatystyczna

Keywords:

Forensic genetics, identification markers, genetic identification, kinship simulations, biostatistical analysis

Pragnę złożyć szczególne podziękowania mojemu promotorowi, dr hab. n. med. Ireneuszowi Sołtyśzewskiemu, którego wsparcie i zaangażowanie miały kluczowe znaczenie dla powstania niniejszej pracy. To z Jego inicjatywy podjąłem się jej realizacji, a dzięki konsekwentnej motywacji, merytorycznym wskazówkom oraz nieocenionemu wsparciu możliwe było doprowadzenie pracy do pomyślnego zakończenia.

Serdeczne podziękowania kieruję do dr n. med. Marii Szargut za ogromne wsparcie na każdym etapie realizacji pracy oraz niezmiennie pozytywne nastawienie w trudnych i wymagających momentach. Bez Jej energii, optymizmu i wiary w sens podejmowanych działań praca ta nie przybrałaby obecnego kształtu.

Dziękuję prof. dr hab. n. med. Andrzejowi Ossowskiemu za zaufanie i umożliwienie współpracy przy realizacji projektu, który stał się podstawą niniejszej rozprawy i pozwolił mi na rozwój zarówno naukowy, jak i zawodowy.

Wyrazy wdzięczności składam również dr Wojciechowi Łuczakowi za udostępnienie zestawu KinFinder oraz merytoryczne wsparcie i pomoc w pracy z wykorzystaniem tego narzędzia.

Szczególne podziękowania kieruję do mojej żony Olgi za wyrozumiałość, wiarę w moje możliwości oraz wsparcie w chwilach zwątpienia. Jej motywacja i cierpliwość odegrały istotną rolę w doprowadzeniu pracy do końca.

Spis treści

Spis rycin	7
Spis tabel.....	10
Wykaz stosowanych skrótów.....	11
Streszczenie w języku polskim.....	13
Streszczenie w języku angielskim.....	15
1. Wstęp.....	17
1.1. Kontekst historyczny.....	17
1.1.1. Okres Wielkiego Terroru w Związku Radzieckim	17
1.1.2. Represje na terenie Adżarii - autonomicznym regionie Gruzji	18
1.1.3. Ekshumacje i identyfikacje ofiar Wielkiego Terroru	21
1.2. Metody genetyczne w identyfikacji osobniczej.....	23
1.2.1. Markery STR.....	25
1.2.2. Autosomalne markery STR.....	26
1.2.3. Markery STR chromosomu Y.....	27
1.2.4. Markery STR chromosomu X.....	28
1.2.5. Markery SNP.....	30
1.2.6. Analiza DNA mitochondrialnego.....	30
1.3. Analiza pochodzenia biogeograficznego.....	32
1.4. Proces degradacji DNA	33
1.5. Proces identyfikacji osobniczej NN zwłok i szczątków ludzkich.....	34
1.6. Genetyczne bazy populacyjne	36
2. Założenia i cel pracy	38
3. Materiał i metody	39
3.1. Materiał badawczy	39
3.2. Zastosowane metody badawcze.....	39
3.3. Przygotowanie materiału kostnego do izolacji DNA	41
3.4. Izolacja DNA	42
3.5. Oznaczenie stężenia DNA, stopnia degradacji i obecności inhibitorów reakcji PCR	42
3.6. Amplifikowanie fragmentów STR w systemie GlobalFiler™.....	42
3.7. Amplifikowanie fragmentów STR w systemie KinFinder	43
3.8. Amplifikowanie fragmentów STR w systemie Yfiler™ Plus	43
3.9. Konstruowanie bibliotek SNP w systemie Precision ID Identity	43

3.10.	Analiza profili genetycznych	44
3.11.	Ustalanie haplogrup chromosomu Y	45
3.12.	Wykorzystanie programu Familias	45
3.13.	Wykorzystanie programu FamLink	48
3.14.	Wykorzystanie programu KinBN.....	50
3.15.	Wykorzystanie programu pedbuildr.....	52
4.	Wyniki.....	54
4.1.	Rodzina 1	54
4.2.	Rodzina 2	57
4.3.	Rodzina 3	61
4.4.	Rodzina 4.....	63
4.5.	Rodzina 5	67
4.6.	Rodzina 6.....	69
4.7.	Rodzina 7.....	72
4.8.	Rodzina 8.....	76
4.9.	Rodzina 9.....	78
4.10.	Rodzina 10.....	80
4.11.	Rodzina 11.....	84
4.12.	Rodzina 12.....	86
4.13.	Rodzina 13.....	89
4.14.	Rodzina 14.....	93
4.15.	Rodzina 15.....	96
4.16.	Rodzina 16.....	97
4.17.	Rodzina 17.....	101
4.18.	Rodzina 18.....	104
4.19.	Rodzina 19.....	106
4.20.	Rodzina 20.....	109
4.21.	Dodatkowe obliczenia.....	111
4.21.1.	Rodzina 1 po korekcie	112
4.21.2.	Rodzina 19 po korekcie	115
4.22.	Pochodzenie biogeograficzne	116
4.23.	Opracowanie schematu decyzyjnego doboru paneli markerów genetycznych	118
4.24.	Omówienie wyników badań.....	119

4.25.	Analiza haplotypów chromosomu Y	121
4.26.	Przydatność zestawów markerów	122
4.27.	Przydatność zastosowanych programów	125
4.28.	Ocena jakości DNA i efektywności zestawów	126
5.	Dyskusja	129
5.1.	Ocena stopnia degradacji materiału kostnego i jego wpływ na uzyskane wyniki	129
5.2.	Ocena skuteczności markerów STR w identyfikacji NN zwłok i szczątków ludzkich ...	130
5.3.	Rola markerów Y-STR w analizie linii ojcowskiej i ich przydatność w identyfikacji szczątków ludzkich.....	132
5.4.	Wpływ doboru bazy populacyjnej na analizę statystyczną oraz ocenę wartości dowodowej wyników	134
5.5.	Zastosowanie paneli SNP w identyfikacji NN zwłok i szczątków ludzkich	136
5.6.	Integracja markerów STR, Y-STR i SNP w procesie identyfikacji NN zwłok i szczątków ludzkich z materiału archiwalnego	137
5.7.	Ocena skuteczności zastosowanych narzędzi statystycznych oraz weryfikacja założeń genealogicznych	140
5.8.	Analiza haplotypów Y i pochodzenia biogeograficznego.....	142
5.9.	Znaczenie badań dla praktyki identyfikacyjnej i możliwości wdrożenia modelu	142
5.10.	Socjologiczny i etyczny wymiar badań	144
6.	Wnioski	146
7.	Piśmiennictwo.....	148
8.	Opinia komisji Bioetycznej.....	167

Spis rycin

Rys. 1. Lista egzekucyjna.....	20
Rys. 2. Mapa pogładowa lokalizacji jamy grobowej	23
Rys. 3. Drzewo genealogiczne dla Rodziny 1	55
Rys. 4. Wykres trójkąta IBD dla Rodziny 1 dla zestawu GlobalFiler™ i KinFinder	56
Rys. 5. Symulacje pokrewieństw dla Rodziny 1 dla zestawu GlobalFiler™	56
Rys. 6. Symulacje pokrewieństw dla Rodziny 1 dla zestawu KinFinder	57
Rys. 7. Drzewo genealogiczne dla Rodziny 2	58
Rys. 8. Wykres trójkąta IBD dla Rodziny 2 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity	59
Rys. 9. Symulacje pokrewieństw dla Rodziny 2 dla zestawu GlobalFiler™	59
Rys. 10. Symulacje pokrewieństw dla Rodziny 2 dla zestawu KinFinder	60
Rys. 11. Symulacje pokrewieństw dla Rodziny 2 dla zestawu Precision ID Identity	60
Rys. 12. Drzewo genealogiczne dla Rodziny 3	61
Rys. 13. Wykres trójkąta IBD dla Rodziny 3	62
Rys. 14. Symulacje pokrewieństw dla Rodziny 3	62
Rys. 15. Drzewo genealogiczne dla Rodziny 4	64
Rys. 16. Wykres trójkąta IBD dla Rodziny 4 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity	65
Rys. 17. Symulacje pokrewieństw dla Rodziny 4 dla zestawu GlobalFiler™	65
Rys. 18. Symulacje pokrewieństw dla Rodziny 4 dla zestawu KinFinder	66
Rys. 19. Symulacje pokrewieństw dla Rodziny 4 dla zestawu Precision ID Identity	66
Rys. 20. Drzewo genealogiczne dla Rodziny 5	67
Rys. 21. Wykres trójkąta IBD dla Rodziny 5	68
Rys. 22. Symulacje pokrewieństw dla Rodziny 5	68
Rys. 23. Drzewo genealogiczne dla Rodziny 6	70
Rys. 24. Wykres trójkąta IBD dla Rodziny 6 dla zestawu GlobalFiler™ i KinFinder	71
Rys. 25. Symulacje pokrewieństw dla Rodziny 6 dla zestawu GlobalFiler™	71
Rys. 26. Symulacje pokrewieństw dla Rodziny 6 dla zestawu KinFinder	72
Rys. 27. Drzewo genealogiczne dla Rodziny 7	73
Rys. 28. Wykres trójkąta IBD dla Rodziny 7 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity	74
Rys. 29. Symulacje pokrewieństw dla Rodziny 7 dla zestawu GlobalFiler™	74

Rys. 30. Symulacje pokrewieństw dla Rodziny 7 dla zestawu KinFinder	75
Rys. 31. Symulacje pokrewieństw dla Rodziny 7 dla zestawu Precision ID Identity	75
Rys. 32. Drzewo genealogiczne dla Rodziny 8	77
Rys. 33. Wykres trójkąta IBD dla Rodziny 8 dla zestawu GlobalFiler™ i KinFinder	77
Rys. 34. Symulacje pokrewieństw dla Rodziny 8 dla zestawu GlobalFiler™	78
Rys. 35. Symulacje pokrewieństw dla Rodziny 8 dla zestawu KinFinder	78
Rys. 36. Drzewo genealogiczne dla Rodziny 9	79
Rys. 37. Drzewo genealogiczne dla Rodziny 10	81
Rys. 38. Wykres trójkąta IBD dla Rodziny 10 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity	82
Rys. 39. Symulacje pokrewieństw dla Rodziny 10 dla zestawu GlobalFiler™	82
Rys. 40. Symulacje pokrewieństw dla Rodziny 10 dla zestawu KinFinder	83
Rys. 41. Symulacje pokrewieństw dla Rodziny 10 dla zestawu Precision ID Identity	83
Rys. 42. Drzewo genealogiczne dla Rodziny 11	84
Rys. 43. Wykres trójkąta IBD dla Rodziny 11	85
Rys. 44. Symulacje pokrewieństw dla Rodziny 11	85
Rys. 45. Drzewo genealogiczne dla Rodziny 12	87
Rys. 46. Wykres trójkąta IBD dla Rodziny 12 dla zestawu GlobalFiler™ i KinFinder	88
Rys. 47. Symulacje pokrewieństw dla Rodziny 12 dla zestawu GlobalFiler™	88
Rys. 48. Symulacje pokrewieństw dla Rodziny 12 dla zestawu KinFinder	89
Rys. 49. Drzewo genealogiczne dla Rodziny 13	90
Rys. 50. Wykres trójkąta IBD dla Rodziny 13 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity	91
Rys. 51. Symulacje pokrewieństw dla Rodziny 13 dla zestawu GlobalFiler™	91
Rys. 52. Symulacje pokrewieństw dla Rodziny 13 dla zestawu KinFinder	92
Rys. 53. Symulacje pokrewieństw dla Rodziny 13 dla zestawu Precision ID Identity	92
Rys. 54. Drzewo genealogiczne dla Rodziny 14	94
Rys. 55. Wykres trójkąta IBD dla Rodziny 14 dla zestawu GlobalFiler™ i KinFinder	95
Rys. 56. Symulacje pokrewieństw dla Rodziny 14 dla zestawu GlobalFiler™	95
Rys. 57. Symulacje pokrewieństw dla Rodziny 14 dla zestawu KinFinder	96
Rys. 58. Drzewo genealogiczne dla Rodziny 15	97
Rys. 59. Drzewo genealogiczne dla Rodziny 16	98
Rys. 60. Wykres trójkąta IBD dla Rodziny 16 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity	99

Rys. 61. Symulacje pokrewieństw dla Rodziny 16 dla zestawu GlobalFiler™	100
Rys. 62. Symulacje pokrewieństw dla Rodziny 16 dla zestawu KinFinder	100
Rys. 63. Symulacje pokrewieństw dla Rodziny 16 dla zestawu Precision ID Identity	101
Rys. 64. Drzewo genealogiczne dla Rodziny 17	102
Rys. 65. Wykres trójkąta IBD dla Rodziny 17 dla zestawu GlobalFiler™ i KinFinder	103
Rys. 66. Symulacje pokrewieństw dla Rodziny 17 dla zestawu GlobalFiler™	103
Rys. 67. Symulacje pokrewieństw dla Rodziny 17 dla zestawu KinFinder	104
Rys. 68. Drzewo genealogiczne dla Rodziny 18	105
Rys. 69. Wykres trójkąta IBD dla Rodziny 18	105
Rys. 70. Symulacje pokrewieństw dla Rodziny 18	106
Rys. 71. Drzewo genealogiczne dla Rodziny 19	107
Rys. 72. Wykres trójkąta IBD dla Rodziny 19 dla zestawu GlobalFiler™ i KinFinder	108
Rys. 73. Symulacje pokrewieństw dla Rodziny 19 dla zestawu GlobalFiler™	108
Rys. 74. Symulacje pokrewieństw dla Rodziny 19 dla zestawu KinFinder	109
Rys. 75. Drzewo genealogiczne dla Rodziny 20	110
Rys. 76. Wykres trójkąta IBD dla Rodziny 20	110
Rys. 77. Symulacje pokrewieństw dla Rodziny 20	111
Rys. 78. Drzewo genealogiczne dla Rodziny 1 po korekcie	113
Rys. 79. Wykres trójkąta IBD dla Rodziny 1 dla zestawu GlobalFiler™ i KinFinder po korekcie	114
Rys. 80. Symulacje pokrewieństw dla Rodziny 1 dla zestawu GlobalFiler™ po korekcie ...	114
Rys. 81. Symulacje pokrewieństw dla Rodziny 1 dla zestawu KinFinder po korekcie	115
Rys. 82. Drzewo genealogiczne dla Rodziny 19 po korekcie	116
Rys. 83. Rozkłady haplogrup chromosomu Y oraz oszacowanych haplogrup chromosomu Y szczątków z masowego grobu w Adżarii, Gruzinów, Mingrelian i Svanów [130] [131]	117
Rys. 84. Schemat decyzyjny	118
Rys. 85. Wykres symulacji rozkładu wartości LR i uzyskanych wyników dla zestawu GlobalFiler™	122
Rys. 86. Wykres symulacji rozkładu wartości LR i uzyskanych wyników dla zestawu KinFinder	123
Rys. 87. Wykres symulacji rozkładu wartości LR przy uwzględnieniu markerów oznaczonych dla danych szczątków i uzyskanych wyników dla zestawu KinFinder	124
Rys. 88. Wykres symulacji rozkładu wartości LR i uzyskanych wyników dla zestawu Precision ID Identity	125

Spis tabel

Tab. 1. Zależność szansy odziedziczenia alleli IBD od stopnia pokrewieństwa.....	53
Tab. 2. Symulacja wyników analizy statystycznej dla Rodziny 1	55
Tab. 3. Symulacja wyników analizy statystycznej dla Rodziny 2	58
Tab. 4. Symulacja wyników analizy statystycznej dla Rodziny 3	62
Tab. 5. Symulacja wyników analizy statystycznej dla Rodziny 4	64
Tab. 6. Symulacja wyników analizy statystycznej dla Rodziny 5	68
Tab. 7. Symulacja wyników analizy statystycznej dla Rodziny 6	70
Tab. 8. Symulacja wyników analizy statystycznej dla Rodziny 7	73
Tab. 9. Symulacja wyników analizy statystycznej dla Rodziny 8	77
Tab. 10. Symulacja wyników analizy statystycznej dla Rodziny 9	80
Tab. 11. Symulacja wyników analizy statystycznej dla Rodziny 10	81
Tab. 12. Symulacja wyników analizy statystycznej dla Rodziny 11	85
Tab. 13. Symulacja wyników analizy statystycznej dla Rodziny 12	87
Tab. 14. Symulacja wyników analizy statystycznej dla Rodziny 13	90
Tab. 15. Symulacja wyników analizy statystycznej dla Rodziny 14	94
Tab. 16. Symulacja wyników analizy statystycznej dla Rodziny 15	97
Tab. 17. Symulacja wyników analizy statystycznej dla Rodziny 16	99
Tab. 18. Symulacja wyników analizy statystycznej dla Rodziny 17	102
Tab. 19. Symulacja wyników analizy statystycznej dla Rodziny 18	105
Tab. 20. Symulacja wyników analizy statystycznej dla Rodziny 19	107
Tab. 21. Symulacja wyników analizy statystycznej dla Rodziny 20	110
Tab. 22. Symulacja wyników analizy statystycznej dla Rodziny 1 po korekcie.....	113
Tab. 23. Symulacja wyników analizy statystycznej dla Rodziny 19 po korekcie.....	116
Tab. 24. Zestawienie ilości oznaczonych układów i jakości DNA	128

Wykaz stosowanych skrótów

ASFO - Amerykańskie Stowarzyszenie Odontologii Sądowej (ang. American Society of Forensic Odontology)

CODIS – system baz danych FBI (ang. Combined DNA Index System)

cM – centimorgan

DI – indeks degradacji (ang. Degradation Index)

DVI – identyfikacja ofiar katastrof (ang. disaster victim identification)

ESS – standardowy zestaw loci (ang. European Standard Set)

FTA – Flinders Technology Associates

FIGG – Genetyczna genealogia sądowa (ang. Forensic Investigative Genetic Genealogy)

GAFA - Gruzjińskie Stowarzyszenie Antropologii Sądowej (ang. Georgian Association of Forensic Anthropology)

GRDIP - Gruzjiński Projekt Odzyskiwania, Dokumentacji i Identyfikacji (ang. Georgian Recovery, Documentation, and Identification Project)

HV – regiony hiperzmienne (ang. hypervariable region)

IBD – współdzielenie alleli identycznych z pochodzenia (ang. identity by descent)

IBS – współdzielenie identycznych alleli (ang. identity by state)

ICMP – Międzynarodową Komisją ds. Osób Zaginionych (ang. International Commission on Missing Persons)

ICRC – Międzynarodowym Komitetem Czerwonego Krzyża (International Committee of the Red Cross)

IDFI – Instytut Rozwoju Wolności Informacji (ang. Institute for Development of Freedom of Information)

iiSNP – informatywny identyfikacyjny polimorfizm pojedynczego nukleotydu (ang. identity informative Single Nucleotide Polymorphism)

ISFG – Międzynarodowe Towarzystwo Genetyki Sądowej (ang. International Society for Forensic Genetics)

LR – iloraz wiarygodności (ang. likelihood ratio)

LRc – łączny iloraz wiarygodności (ang. combined likelihood ratio)

MPS – masowe równoległe sekwencjonowanie (ang. Massively Parallel Sequencing)

mtDNA – mitochondrialne DNA (ang. mitochondrial DNA)

NGS – sekwencjonowanie następnej generacji (ang. next generation sequencing)

NRY – region nierekombinujący chromosome Y (ang. Non-recombining Y-chromosome region)

PCR – reakcja łańcuchowa polimerazy (ang. polymerase chain reaction)

PUM - Pomorski Uniwersytet Medyczny w Szczecinie

RM – szybko mutujące markery (ang. rapidly mutating)

RFLP - Polimorfizm długości fragmentów restrykcyjnych (ang. Restriction Fragment Length Polymorphism)

SNP – polimorfizm pojedynczego nukleotydu (ang. single nucleotide polymorphism)

SSR – proste powtórzenia sekwencji (ang. Simple Sequence Repeats)

Stdev – odchylenie standardowe (ang. Standard deviation)

STR – krótkie powtórzenia tandemowe (ang. short tandem repeats)

STRidER – referencyjna baza częstości identyfikacyjnych markerów STR (ang. STRs for Identity ENFSI Reference Database)

SWGAM – Naukowa Grupa Robocza ds. Metod Analizy DNA (ang. Scientific Working Group on DNA Analysis Methods).

X-STR – krótkie powtórzenia tandemowe zlokalizowane na chromosomie X (ang. X-chromosomal Short Tandem Repeats)

YHRD – referencyjna baza haplotypów chromosomu Y (ang. Y Chromosome Haplotype Reference Database)

Y-STR – krótkie powtórzenia tandemowe zlokalizowane na chromosomie Y (ang. Y chromosome short tandem repeats)

Streszczenie w języku polskim

W pracy przedstawiono nowatorskie podejście do identyfikacji NN szczątków ludzkich na przykładzie ofiar masowych egzekucji z okresu Wielkiego Terroru w Adżarii (Gruzja), realizowanego w ramach międzynarodowego projektu Georgian Recovery, Documentation and Identification Project.

Celem badań realizowanych w ramach pracy doktorskiej było opracowanie i weryfikacja schematu analitycznego umożliwiającego skuteczną identyfikację genetyczną NN zwłok i szczątków ludzkich z materiału archiwalnego. W tego typu sprawach problemem jest nie tylko stopień zachowania materiału biologicznego szczątków, ale również ograniczona dostępność próbek referencyjnych od krewnych ofiar, co stanowi wyzwanie dla biegłych wykonujących badania identyfikacyjne.

Materiał badawczy obejmował 28 szkieletów mężczyzn ekshumowanych w 2021 roku na terenie dawnej bazy wojskowej w Adżarii oraz próbki biologiczne pobrane od 58 osób należących do 20 rodzin poszukujących swoich bliskich. Do analizy laboratoryjnej wykorzystano komercyjne zestawy odczynników do amplifikacji i oznaczania markerów STR autosomalnych (GlobalFiler™) oraz markerów zlokalizowanych na chromosomie Y (Yfiler™ Plus). W badaniach wykorzystano również zestaw do amplifikacji markerów iiSNP (Precision ID Identity) i ich sekwencjonowania z zastosowaniem NGS. Producentem tych odczynników jest firma Thermo Fisher Scientific. Ponadto do badań wykorzystano nieskomercjalizowany zestaw odczynników (KinFinder) opracowany przez Laboratorium Diagnostyki Molekularnej GenMed sp. j. Analizy statystyczne zostały wykonane z użyciem oprogramowania Familias, FamLink2, KinBN i pakietu pedbuildr w środowisku R, umożliwiającego symulacje pokrewieństw i ocenę siły dowodu dla złożonych struktur genealogicznych.

Zastosowany schemat badawczy umożliwił uzyskanie jednoznacznych identyfikacji dla wszystkich 17 rodzin poszukujących krewnych znajdujących się na liście egzekucyjnej przypisanej do analizowanej jamy grobowej.

Symulacje IBD oraz analizy pokrewieństw pozwoliły potwierdzić relacje rodzinne w większości przypadków i ujawniły błędne założenia genealogiczne w kilku przypadkach. Przeprowadzona analiza pochodzenia biogeograficznego potwierdziła zgodność genetyczną badanej grupy z populacją gruzińską.

Integracja klasycznych i nowoczesnych metod genetycznych, wsparta poszerzoną analizą biostatystyczną, pozwoliła na uzyskanie wiarygodnych wyników identyfikacyjnych nawet w warunkach ograniczonej ilości i jakości materiału biologicznego. Zastosowanie programów

statystycznych oraz narzędzi symulacyjnych (pedbuildr) umożliwiło dokładniejsze odwzorowanie rzeczywistych relacji rodzinnych, co znacząco zwiększyło moc dowodową analiz. Zastosowane podejście może być z powodzeniem implementowane w sprawach identyfikacyjnych.

Z przeprowadzonych badań wynika, że opracowany zintegrowany model analityczny, łączący analizy STR, Y-STR i SNP z symulacjami pokrewieństw, skutecznie przewyższa ograniczenia tradycyjnych metod identyfikacyjnych genetycznych i pozwala na uzyskanie wiarygodnych wyników nawet przy niskiej zawartości DNA oraz w przypadkach dalszego pokrewieństwa. Symulacje genealogiczne w środowisku R stanowią efektywne narzędzie wspomagające interpretację danych, a zastosowanie kilku niezależnych programów statystycznych potwierdza spójność i powtarzalność uzyskanych wartości LR. Wyniki badań dowodzą, że skuteczna identyfikacja NN zwłok i szczątków ludzkich z materiałów archiwalnych, jest możliwa mimo znacznego upływu czasu od zgonu, pod warunkiem zastosowania kompleksowego, wieloetapowego podejścia analitycznego. Opracowany model może stanowić gotowy schemat działania w przyszłych projektach dotyczących identyfikacji NN szczątków ludzkich.

Streszczenie w języku angielskim

Optimization of Genetic Identification Procedures for Unidentified Human Remains

This study presents a comprehensive approach to the identification of unknown human remains, using the example of victims of mass executions during the Great Terror in Adjara (Georgia), conducted within the framework of the international Georgian Recovery, Documentation and Identification Project. The aim of the study was to develop and verify an analytical scheme enabling effective genetic identification of historical victims. In such cases, the problem is not only the degree of preservation of the biological material of the remains, but also the limited availability of reference samples from the victims' relatives, which poses a challenge for experts conducting identification analyses.

The research material included 28 male skeletons exhumed in 2021 from the site of a former military base in Adjara, as well as biological samples collected from 58 individuals representing 20 families searching for their missing relatives. Laboratory analyses were performed using commercially available reagent kits for the amplification and detection of autosomal STR markers (GlobalFiler™) as well as markers located on the Y chromosome (Yfiler™ Plus). The study also employed a kit for amplifying iiSNP markers and sequencing them using NGS (Precision ID Identity). All commercial reagents were manufactured by Thermo Fisher Scientific. In addition, a non-commercial kit (KinFinder) developed by the Molecular Diagnostics Laboratory GenMed sp. j. was used. Statistical analyses were carried out using Familias, FamLink2, and KinBN, as well as the pedbuildr package in the R environment, which allows for kinship simulations and the assessment of the strength of evidence for complex genealogical structures.

The research scheme used in this study enabled conclusive identification of all 17 families searching for relatives listed on the execution list associated with the analysed burial pit.

IBD simulations and kinship analyses confirmed family relationships in most cases and revealed incorrect genealogical assumptions in several families. In addition, biogeographical ancestry analysis was performed, which confirmed the genetic compatibility of the study group with the Georgian population.

Integration of classical and modern genetic methods, supported by biostatistical analysis, allows for reliable identification results even in conditions of limited quantity and quality of biological material. The use of statistical software that take into account genetic linkage, combined with simulation tools such as pedbuildr, enabled a more accurate representation of actual family

relationships, which significantly increased the evidential value of the analyses. The proposed approach can be successfully implemented in forensic identification cases.

The research shows that the developed integrated analytical model, combining STR, Y-STR and SNP analyses with kinship simulations, effectively overcomes the limitations of traditional identification methods and allows for reliable results even with low DNA quantities and in cases of distant kinship. Genealogical simulations in the R environment proved to be an efficient tool supporting data interpretation. The use of multiple independent statistical programs confirmed the consistency and reproducibility of the obtained LR values. The results of the study prove that effective identification of historical victims is possible despite significant postmortem intervals, provided that a comprehensive, multi-step analytical approach is employed. The developed model can serve as a basis for a standardized workflow for future projects concerning the identification of unknown human remains.

1. Wstęp

Identyfikacja genetyczna nieznanymi zwłok i szczątków ludzkich stanowi jedno z kluczowych wyzwań współczesnej genetyki sądowej. W przypadkach, w których materiał biologiczny ulega znacznej degradacji, a dostępni krewni znajdują się w dalszych liniach pokrewieństwa, skuteczność klasycznych metod analizy genetycznej nie zawsze jest wystarczająca. Dlatego też w pracy skupiono się na optymalizacji procesu identyfikacji, obejmującej dobór metod badawczych, wykorzystanie modeli statystycznych, analizy pochodzenia biogeograficznego oraz integrację danych genealogicznych i historycznych.

1.1. Kontekst historyczny

Dane historyczne stanowią podstawę dla działań ekshumacyjnych i identyfikacyjnych. Informacje zawarte w materiałach archiwalnych, w połączeniu z wynikami badań archeologicznych i antropologicznych, umożliwiają formułowanie hipotez identyfikacyjnych oraz dobór odpowiedniego materiału porównawczego.

Materiał badawczy pochodził ze szczątków 28 osób odkrytych w 2021 roku w masowym grobie na terenie Adżarii (południowo-zachodnia Gruzja). Zarówno analiza kontekstu archeologicznego, jak i dostępne dokumenty archiwalne sugerowały, że mogły to być ofiary represji stalinowskich z lat 1937–1938. Z grobem powiązana została lista egzekucyjna z dnia 27 grudnia 1937 roku, obejmująca 28 nazwisk, odpowiadającą liczbie odnalezionych osób. Materiał porównawczy stanowiły próbki referencyjne pobrane od członków rodzin osób uznanych za zaginione w tamtym okresie.

1.1.1. Okres Wielkiego Terroru w Związku Radzieckim

Związek Radziecki od 1937 do 1938 był miejscem masowych represji, a okres ten nazywany jest okresem Wielkiego Terroru [1]. Brutalne działania państwa przeciwko jego własnym obywatelom miało na celu wzbudzenie powszechnego strachu wśród ludności oraz przeobrażenie wszystkich warstw społeczno-ekonomicznych. Działania aparatu państwa były skierowane zarówno przeciwko rzeczywistym jak i domniemanym wrogom władzy sowieckiej. Represje dotknęły zwykłych obywateli jak również prominentnych działaczy partyjnych. W represjach czynny udział brała każda organizacja i urząd państwowy wraz z przywódcami republik, również takich które posiadały autonomię. W Związku Radzieckim, pomimo iż formalnie był federacją narodowych republik radzieckich, w rzeczywistości władza i gospodarka była bardzo mocno scentralizowana i skupiona w Moskwie. Skala represji była potęgowana także przez podżeganie obywateli do zwracania się przeciwko współobywatelom,

oraz poprzez propagandę w mediach krajowych. Wiedza o wydarzeniach celowo nie była utrzymywana w tajemnicy. Relacje prasowe regularnie informowały o kolejnych aresztowaniach i procesach, przedstawiając je jako pozytywny etap eliminowania wrogów ludu. Na mocy rozkazu nr 00447 z dnia 30 lipca 1937 powołano tzw. „trójki” funkcjonariuszy Ludowego Komisariatu Spraw Wewnętrznych ZSRR (NKWD). Były to specjalnie powołane komisje mające na celu szybkie rozpatrywanie spraw sądowych [2]. Stworzono dwie kategorie oskarżonych: w pierwszej znajdowali się „wrogowie ludu” - osoby które trafiły na tę listę były rozstrzeliwane; w drugiej kategorii znajdowały się osoby wrogie, lecz mniej aktywne niż w pierwszej kategorii - osoby te trafiały do obozów pracy (łagrów), co traktowano jako sposób izolacji „w celu ochrony republiki radzieckiej przed wrogami klasowymi”. Dzięki działaniu „trójek” skazania odbywały się planowo i masowo, często obejmując rozstrzelania i deportacje całych grup społecznych. Niestety ustalenie dokładnej listy ofiar w dalszym ciągu nie jest możliwe.

Aparat bezpieczeństwa kontrolował społeczeństwo, cenzura ograniczała wolność słowa, a wszelkie przejawy sprzeciwu wobec władzy były brutalnie tłumione. Ludzie byli aresztowani bez wyroków, przetrzymywani w więzieniach, torturowani i często mordowani a ich ciała grzebano w bezimiennych dołach. Takie działania miały nie tylko charakter represyjny, lecz także psychologiczny – wprowadzały w społeczeństwie atmosferę strachu i niepewności, utrwalając poczucie bezsilności wobec państwa [3]. Według Wertha w tym okresie skazano na śmierć i rozstrzelano 750 000 osób [4]. Z kolei według Calleja rozstrzelano 1 mln osób a 12 mln zostało uwięzionych w koloniach karnych [5].

1.1.2. Represje na terenie Adżarii - autonomicznym regionie Gruzji

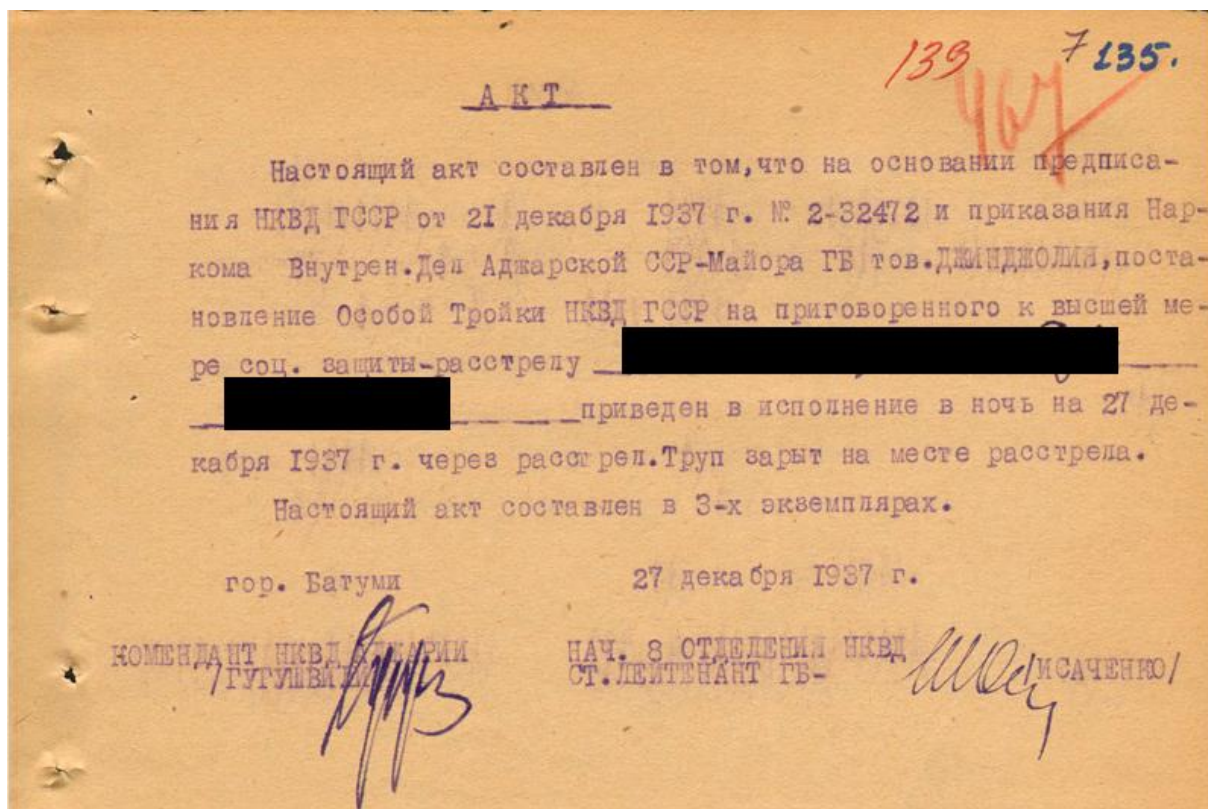
W latach 30. XX wieku Adżaria, wówczas część Gruzjińskiej Socjalistycznej Republiki Radzieckiej, wyróżniała się odmienną strukturą etniczną i religijną niż współczesna Gruzja. Ukształtowanie to było konsekwencją wielowiekowej obecności Imperium Osmańskiego, pod którego panowaniem region pozostawał od XVI aż do końca XIX wieku. W rezultacie znaczna część miejscowej ludności wyznawała islam sunnicki, choć etnicznie mieszkańcy ci stanowili głównie naród gruziński. Lokalna grupa etnograficzna, zwana Adżarami, przez wiele pokoleń zachowała swoją odrębność kulturową, jednocześnie pozostając blisko spokrewnioną i językowo związaną z innymi grupami gruzińskimi.

Po rozpadzie Imperium Rosyjskiego sytuacja polityczna regionu uległa zmianie. Traktat z Karsu z 1921 roku przyznał Adżarii status autonomiczny w ramach tworzącej się Gruzjińskiej Socjalistycznej Republiki Radzieckiej. Autonomia ta została później potwierdzona

w konstytucji z 1922 roku, a następnie utrzymana w konstytucji Związku Socjalistycznych Republik Radzieckich (ZSRR) z 1936 roku, co umocniło jej formalny charakter. Autonomia Adżarii miała charakter terytorialno-polityczny i formalnie gwarantowała regionowi pewien zakres samorządności, obejmujący lokalną administrację, język urzędowy oraz częściową ochronę tożsamości etniczno-religijnej. W praktyce jednak autonomiczność regionu była bardzo ograniczona przez silnie scentralizowaną administrację sowiecką.

Zarządzanie Adżarią odbywało się z wykorzystaniem lokalnych organów partyjnych i państwowych, podlegających bezpośrednio Tbilisi i Moskwie. Najważniejsze decyzje polityczne, społeczne oraz ekonomiczne były podejmowane w oparciu o dyrektywy władz centralnych. Region Adżarii miał znaczenie strategiczne, wynikające głównie z jego geopolitycznego położenia. Bezpośrednie sąsiedztwo z Turcją – krajem należącym do odrębnego, niesowieckiego kręgu polityczno-kulturowego – było powodem szczególnego zainteresowania organów bezpieczeństwa państwowego (NKWD) [6].

W okresie Wielkiego Terroru na terenie Gruzji aresztowano ponad 30 000 osób z czego 9 000 zostało rozstrzelanych. Byli to w głównej mierze członkowie miejskiej inteligencji lub osoby posiadające władzę lokalną [7]. Kary śmierci były wykonywane głównie w nocy na otwartym terenie, w pewnej odległości od ludzkich siedzib. Skazani byli przywożeni z więzień zlokalizowanych w Tbilisi, Batumi, Kutaisi i Sokhumi [8]. Egzekucje przeprowadzano na podstawie list egzekucyjnych (Rys. 1), których treść zawierająca informacje obejmujące m.in. datę, miejsce oraz dane osób skazanych pozwala obecnie na weryfikację danych historycznych i podjęcie prób identyfikacji ofiar.



Rys. 1. lista egzekucyjna, źródło: archiwum GRDIP.

Z uwagi na intensywne kontakty mieszkańców Adżarii z rodzinami mieszkającymi po stronie tureckiej, duża część ludności tego regionu postrzegana była przez NKWD jako potencjalni „szpiegzy” lub „element niepewny”. Fakt ten posłużył jako uzasadnienie do intensyfikacji represji w regionie. Obok dominującej populacji gruzińskiej, w regionie żyły mniejsze grupy etniczne: Ormianie, Rosjanie, Grecy, Turcy meschetyńscy, a także pojedyncze osoby innych narodowości, w tym Ukraińcy, Polacy oraz Niemcy, których status społeczny i prawny w okresie represji był szczególnie zagrożony ze względu na ich obce pochodzenie. Ofiarami stali się nie tylko przedstawiciele inteligencji czy działacze polityczni, ale także dużą część stanowili zwykli mieszkańcy podejrzani o nielojalność wobec władzy radzieckiej.

W związku z powyższym odsetek rozstrzelań i zsyłek w Adżarii był wyższy niż w pozostałych regionach Gruzji. Z zachowanych dokumentów wynika, że w rejonie tym rozstrzelano co najmniej 1050 osób, w większości byli to miejscowi mieszkańcy, jednak na listach egzekucyjnych znaleziono 54 cudzoziemców, w tym m.in. Rosjan, Ukraińców, Polaków, Niemców, Ormian, Greków i Turków.

Okres represji znacząco wpłynął na strukturę demograficzną regionu, powodując drastyczne zmiany. Masowe egzekucje, deportacje oraz emigracje spowodowane strachem przed prześladowaniami doprowadziły do znacznego zmniejszenia populacji oraz osłabienia więzi rodzinnych i społecznych, których efekty są odczuwalne do dziś.

Represje stalinowskie miały długotrwałe, negatywne konsekwencje dla społeczności regionu. Ich pierwszym i bezpośrednim efektem była utrata lokalnych elit społecznych, przedstawicieli inteligencji, duchowieństwa oraz liderów społeczności lokalnych, którzy zostali zabici lub zesłani do obozów pracy. Taki stan rzeczy doprowadził do gwałtownego załamania struktur społecznych i kulturowych, ponieważ wyeliminowano osoby będące nośnikami lokalnej tradycji, historii i pamięci kulturowej. Represje prowadziły także do masowego rozbijania rodzin, co skutkowało znaczącymi problemami społecznymi, w tym utratą poczucia tożsamości oraz głębokimi traumami indywidualnymi i zbiorowymi.

Efekty długofalowe tych wydarzeń są odczuwalne do dziś. Po odzyskaniu niepodległości przez Gruzję, w 1991 roku, mieszkańcy Adżarii zaczęli podejmować próby przywrócenia utraconych elementów kulturowych, odbudowania struktur społecznych oraz odzyskania kontroli nad własną tożsamością poprzez działania obejmujące różnorodne inicjatywy społeczne, edukacyjne oraz naukowe, mające na celu uświadomienie społeczności lokalnej oraz całemu społeczeństwu gruzińskiemu rozmiaru dokonanych represji.

1.1.3. Ekshumacje i identyfikacje ofiar Wielkiego Terroru

Identyfikacja ofiar masowych egzekucji ma bardzo duże znaczenie, zarówno dla zainteresowanych rodzin ofiar jak i z punktu widzenia historycznego. Identyfikacja szczątków ludzkich a co za tym idzie ustalenie losów osób, które padły ofiarą represji, pozwala ich rodzinom na zamknięcie bolesnego rozdziału w ich życiu [9]. Zidentyfikowanie ofiar pozwala również na przeciwdziałanie historycznemu zapomnieniu i negacji faktu i skali Wielkiego Terroru w Gruzji. Ponadto, zgodnie z międzynarodowym prawem humanitarnym, należy zapewnić szacunek i ochronę zwłok ofiar konfliktów zbrojnych i politycznych poprzez zabezpieczenie miejsca pochówku, oraz podjęcie kroków zmierzających do identyfikacji szczątków i przekazania informacji o wynikach żyjącym krewnym ofiar [10].

Pierwsze poszukiwania miejsc pochówków ofiar Wielkiego Terroru rozpoczęły się w latach dziewięćdziesiątych ubiegłego wieku, ale z uwagi na zniszczenie archiwów NKWD nie były one skuteczne. W 2012 r. Instytut Rozwoju Wolności Informacji (IDFI) opublikował listę 3600 osób, które zostały rozstrzelane w latach 1937-1938 na terenie Gruzji [11].

Na terenie Adżarii, jak również w innych regionach Gruzji prowadzone są różne działania na rzecz upamiętnienia ofiar represji stalinowskich oraz szeroko pojętej edukacji historycznej. Przykładem tego typu inicjatywy jest działalność muzealna, m.in. funkcjonujące w Batumi Muzeum Represji Politycznych (Museum of Political Repression), które prowadzi działalność edukacyjną, badawczą oraz upamiętniającą wydarzenia okresu Wielkiego Terroru. Zadaniem

muzeum jest dokumentowanie losów ofiar represji, eksponowanie materiałów archiwalnych oraz organizacja wystaw i wydarzeń publicznych.

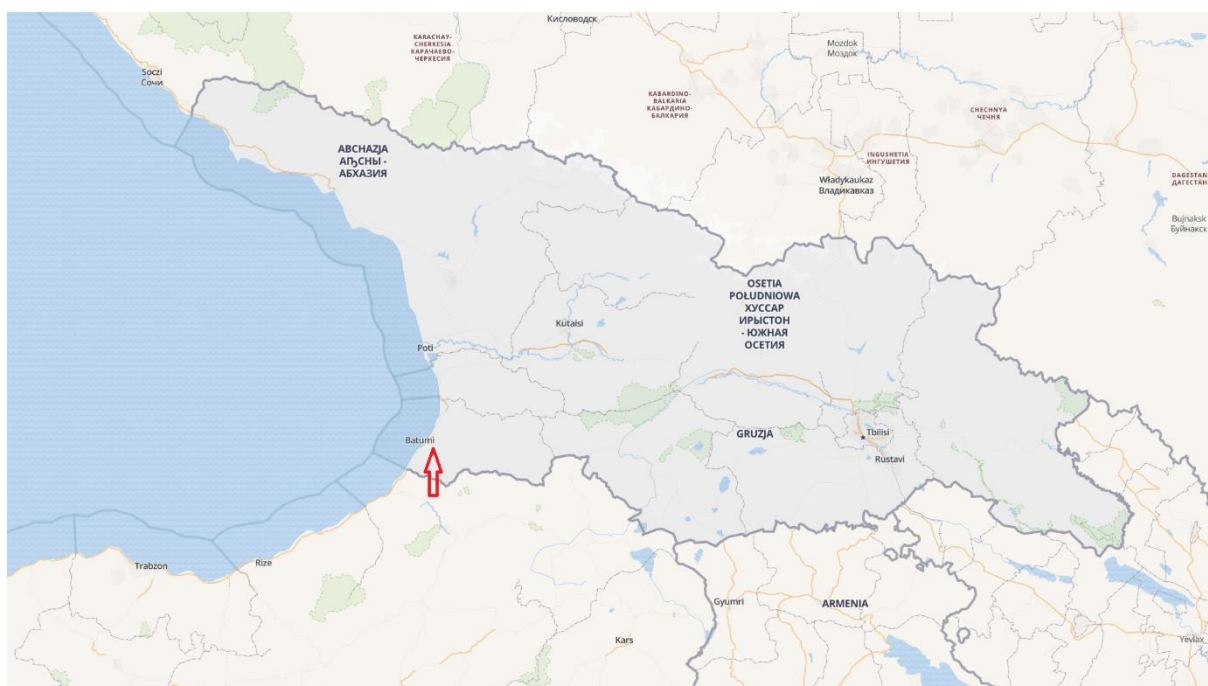
Oprócz działań lokalnych, gruzińskie instytucje prowadzą współpracę z międzynarodowymi organizacjami pozarządowymi i humanitarnymi. W szczególności należy wymienić projekty realizowane we współpracy z Międzynarodową Komisją ds. Osób Zaginionych (International Commission on Missing Persons – ICMP) oraz Międzynarodowym Komitetem Czerwonego Krzyża (International Committee of the Red Cross – ICRC), które pomagają Gruzji w identyfikacji ofiar, odnajdywaniu miejsc masowych pochówków oraz w zapewnieniu wsparcia rodzinom ofiar represji. Dzięki współpracy międzynarodowej możliwe jest wdrażanie nowoczesnych metod antropologii i genetyki sądowej, co znacząco zwiększa skuteczność identyfikacji odnajdywanych szczątków oraz pozwala na godne upamiętnienie ofiar represji.

Georgian Recovery, Documentation, and Identification Project (GRDIP) stanowi inicjatywę kryminalistyczną mającą na celu systematyczne prowadzenie ekshumacji, analiz oraz identyfikacji ofiar represji z okresu sowieckiego w Adżarii, w Gruzji. Projekt zakłada odzyskiwanie i dokumentowanie szczątków ludzkich przy zastosowaniu metod archeologii sądowej, z zapewnieniem właściwej analizy antropologicznej i genetycznej. Poprzez integrację historycznych list egzekucyjnych z danymi przekazanymi przez krewnych ofiar, GRDIP dostarcza kluczowych dowodów dotyczących dawnych naruszeń praw człowieka, jednocześnie dając rodzinom długo oczekiwane odpowiedzi. Ponadto projekt realizuje międzynarodowe standardy ochrony osób zmarłych i umożliwia godny pochówek zarówno ofiar wyznania chrześcijańskiego, jak i muzułmańskiego [12].

Podczas prac ekshumacyjnych prowadzonych przez Gruzkańskie Stowarzyszenie Antropologii Sądowej (GAFA) w sierpniu 2021 roku, na terenie dawnej bazy wojskowej z czasów ZSRR w miejscowości Khelvachauri (41°34'52.5"N 41°38'59.2"E - Rys. 2), przy wsparciu Michigan State University oraz Texas State University, odkryto jamę grobową datowaną na okres Wielkiego Terroru, która zawierała szczątki 28 mężczyzn (1937–1938) [13]. Liczba ta odpowiadała jednej z list egzekucyjnych, co skłoniło międzynarodowy i interdyscyplinarny zespół do poszukiwania rodzin ofiar. Ze względu na rozgłos medialny zgłosiło się 20 rodzin osób które mogły zaginać w tym obszarze i czasie. W trzech przypadkach rodziny poszukiwały osób spoza listy egzekucyjnej (Rodziny 8, 9 i 15) jednakże jako iż nie było pewności co do zgodności listy z jamą grobową pozostawiono te rodziny do dalszych analiz. GAFA zawarło z Pomorskim Uniwersytetem Medycznym w Szczecinie (PUM) porozumienie, na mocy którego materiał kostny pochodzący ze szczątków odkrytych w Adżarii oraz materiał pobrany

od potencjalnych rodzin ofiar został przekazany do laboratorium PUM w celu przeprowadzenia analiz.

Od zbrodni minęło niemal 90 lat, a dostępni krewni byli najczęściej spokrewnieni w drugim i trzecim stopniu (np. wnuki, bratankowie, siostrzenice). Tak odległe pokrewieństwo, w połączeniu z ograniczoną ilością i degradacją DNA jądrowego uzyskanego ze szczątków, stanowiło istotną przeszkodę dla wiarygodnej identyfikacji. Celem niniejszej pracy było zatem opracowanie i przetestowanie strategii umożliwiającej skuteczną identyfikację na dużą skalę w sytuacjach, gdy dostępni są jedynie dalsi krewni.



Rys. 2. Mapa poglądowa lokalizacji jamy grobowej (<https://24timezones.com/Gruzja/mapa>)

1.2. Metody genetyczne w identyfikacji osobniczej

Identyfikacja genetyczna polega na porównaniu profilu DNA osoby badanej z profilem referencyjnym uzyskanym bezpośrednio od niej (np. w ramach postępowania karnego i wprowadzonym do bazy danych) lub z uprawdopodobnionym profilem osoby zaginionej, zabezpieczonym z jej przedmiotów osobistych, takich jak szczoteczka do zębów, maszynka do golenia czy odzież. W przypadkach, gdy materiału takiego nie można uzyskać, porównania dokonuje się z profilami członków najbliższej rodziny, najczęściej rodziców bądź dzieci, rzadziej rodzeństwa [14].

Początki wykorzystania metod genetycznych w identyfikacji osobniczej sięgają roku 1985, kiedy sir Alec Jeffreys opublikował przełomową pracę dotyczącą tzw. „genetycznego odcisku palca” (ang. DNA fingerprint). Pierwsze praktyczne zastosowanie badań DNA w identyfikacji

sprawcy przestępstwa miało miejsce w 1986 roku w Wielkiej Brytanii [15]. Analiza dotyczyła gwałtu i zabójstwa dwóch dziewcząt z miejscowości Narborough (hrabstwo Leicestershire), popełnionych w latach 1983 i 1986. W badaniu wykorzystano polimorfizm długości fragmentów restrykcyjnych (RFLP), technikę opartą na cięciu nici DNA enzymami restrykcyjnymi oraz rozdziale powstałych fragmentów w żelu agarozowym lub poliakrylamidowym metodą elektroforezy [16]. Uzyskany układ prążków stanowił profil DNA danej osoby.

Kolejnym przełomem rozwoju genetyki było opracowanie reakcji łańcuchowej polimerazy (PCR). Metoda ta została po raz pierwszy zaprezentowana w 1984 roku przez Mullisa podczas Zjazdu Naukowców firmy Cetus Corporation (USA), a jej opis opublikowano rok później [17]. Technika PCR, umożliwiającą amplifikację wybranych fragmentów DNA, w krótkim czasie wyparła metodę RFLP z praktyki kryminalistycznej. Współcześnie identyfikacje genetyczne opierają się głównie na analizie sekwencji mikrosatelitarnych krótkich powtórzeń tandemowych (STR) [18]. Loci STR stanowią krótkie, powtarzające się sekwencje składające się zazwyczaj z dwóch do sześciu par zasad, które tworzącą fragmenty o długości od 100 do 450 par zasad. Ze względu na swoje rozmiary mogą być z powodzeniem analizowane nawet w przypadku zdegradowanego DNA.

Dynamiczny rozwój biologii molekularnej oraz technik analitycznych umożliwił znaczące poszerzenie możliwości identyfikacyjnych. W badaniach sądowych jednym z kluczowych elementów pozostaje ocena pokrewieństwa pomiędzy analizowanymi osobami, co ma szczególne znaczenie zwłaszcza w przypadkach identyfikacji szczątków ludzkich [19 - 21]. W takich sytuacjach pozyskanie odpowiedniego materiału porównawczego bywa utrudnione, a szanse powodzenia rosną wraz z bliskością pokrewieństwa. W zależności od rodzaju relacji rodzinnych, dobiera się odpowiednie zestawy markerów STR – autosomalnych lub zlokalizowanych na chromosomach płci (Y, X), a w niektórych przypadkach wykonuje się również analizę mitochondrialnego DNA (mtDNA) [22 - 26].

Niewątpliwą zaletą metod genetycznych jest możliwość ich skutecznego zastosowania nawet w przypadkach, gdy od śmierci osoby do momentu ujawnienia szczątków minęło wiele dziesięcioleci. W takich sytuacjach inne techniki – jak badania daktyloskopijne, wizualne rozpoznanie przez rodzinę czy analiza cech szczególnych – okazują się nieskuteczne z powodu zaawansowanej degradacji tkanek. Jedynie w wyjątkowych przypadkach możliwe jest wykorzystanie elementów trwałych, np. odczytanie numeru seryjnego implantu [27].

Ze względu na wysoki stopień polimorfizmu loci STR, prostotę analizy długości sekwencji metodą PCR oraz relatywnie niski koszt, właśnie ta technika stanowi obecnie podstawową

metodę w badaniach kryminalistycznych i testach pokrewieństwa. Szacuje się, że ponad 99,7% ludzkiego genomu jest identyczne u wszystkich osób, a zróżnicowanie międzyosobnicze wynika z pozostałych 0,3%, obejmujących głównie sekwencje niekodujące DNA [18].

W Polsce pierwsze zastosowanie badań DNA do celów ustalenia ojcostwa miało miejsce w 1990 roku, kiedy Przybylski i Słomski przeprowadzili pionierską analizę genetyczną w tej dziedzinie [28].

1.2.1. Markery STR

Mikrosatelity, określane również mianem prostych powtórzeń sekwencji (SSR) lub krótkich powtórzeń tandemowych, stanowią fragmenty DNA zbudowane z krótkich motywów nukleotydowych powielonych wielokrotnie w układzie tandemowym. Struktury tego typu są powszechnie spotykane w genomach organizmów eukariotycznych [29].

Pojęcie „DNA satelitarnego” zostało po raz pierwszy wprowadzone w 1961 roku w odniesieniu do frakcji DNA wykazującej odmienną gęstość podczas wirowania w gradiencie chlorku cezu, co pozwalało na jej odróżnienie od głównej frakcji genomowej [30]. Termin „mikrosatelita” pojawił się znacznie później – w 1989 roku – gdy Litt i Luty opisali w genomie człowieka tandemowo powtórzoną sekwencję zawierającą motyw TG, charakteryzującą się wysokim stopniem polimorfizmu [31]. Pierwszy opis fragmentu ludzkiego DNA zawierającego powtórzenia tandemowe pochodzi natomiast z roku 1984, kiedy Weller i wsp. zidentyfikowali w genie mioglobiny sekwencję złożoną z powtórzeń GGAT [32].

Ze względu na właściwości biologiczne mikrosatelit – dziedziczenie zgodne z prawami Mendla, wysoki poziom polimorfizmu, częstość mutacji i dużą liczbę kopii w genomie – stanowią one niezwykle użyteczne markery molekularne. Wykorzystuje się je w wielu dziedzinach biologii: w genetyce sądowej do identyfikacji osób i ustalania pokrewieństwa [33, 34], w genetyce populacyjnej do badania zróżnicowania genetycznego i struktury populacji [35], w hodowli roślin i zwierząt do kontroli pochodzenia i selekcji linii hodowlanych [36], a także w celu ochrony bioróżnorodności i monitorowania populacji gatunków zagrożonych [37]. Mikrosatelity są również istotnym narzędziem w biologii ewolucyjnej, gdzie pozwalają śledzić procesy dryfu genetycznego i doboru naturalnego [38], a także w antropologii molekularnej i archeogenetyce, umożliwiając rekonstrukcję pochodzenia i migracji dawnych populacji ludzkich [39]. W zastosowaniach medycznych analiza loci STR wykorzystywana jest w diagnostyce nowotworów [40], ocenie skuteczności terapii, a także w rozpoznawaniu chorób neurodegeneracyjnych, takich jak płasawica Huntingtona czy ataksje rdzeniowo-mózdkowe [41].

Dynamiczny rozwój metod biologii molekularnej, zwłaszcza techniki PCR oraz sekwencjonowania, przyczynił się do szybkiego rozszerzenia wiedzy o rozmieszczeniu mikrosatelit w genomie człowieka. Częstość mutacji w loci mikrosatelitarnych jest od kilku do wielu rzędów wielkości wyższa (10^{-2} – 10^{-6}) niż w przypadku klasycznych mutacji punktowych [42]. W przeciwieństwie do zmian punktowych, które obejmują pojedyncze nukleotydy, mutacje mikrosatelitarne skutkują utratą lub dodaniem całych jednostek powtórzeniowych, niekiedy więcej niż jednej. Wysoka zmienność mikrosatelit koreluje bezpośrednio z ich polimorfizmem populacyjnym, co czyni je szczególnie przydatnymi w analizach identyfikacyjnych i badaniach pokrewieństwa.

W genomie człowieka dominują sekwencje z powtórzeniami dinukleotydowymi, podczas gdy motywy mono- i tetranukleotydowe występują rzadziej, a tri-, penta- i heksanukleotydowe – najrzadziej [43]. Liczba powtórzeń w danym loci stanowi podstawę do określenia allela, czyli wariantu genetycznego opisywanego wartościami liczbowymi (np. 10, 14). Zestaw wszystkich oznaczonych alleli w analizowanych loci STR tworzy profil genetyczny (genotyp) danej osoby [18, str. 43-78].

Zmienność liczby powtórzeń sprawia, że mikrosatelity są niezwykle informatywne i stanowią jeden z najpewniejszych markerów identyfikacyjnych [44]. Dla przykładu, u jednej osoby w określonym locus mogą występować allele zawierające 15 i 10 powtórzeń motywu (np. GATC), podczas gdy inna osoba będzie charakteryzowała się całkowicie odmienną kombinacją tych wartości. Ze względu na ogromną liczbę możliwych wariantów, prawdopodobieństwo wystąpienia identycznych profili STR u dwóch niespokrewnionych osób jest pomijalne [18, str. 87].

1.2.2. Autosomalne markery STR

Genom jądrowy człowieka bez aberracji chromosomowych składa się z 22 par autosomów, czyli chromosomów występujących u obu płci w identycznej liczbie i strukturze, oraz jednej pary allosomów – chromosomów płci. Łącznie w komórce somatycznej znajduje się zatem 46 chromosomów. Taki zestaw cechuje komórki diploidalne, które posiadają podwójny komplet chromosomów homologicznych. Odmienny typ komórek stanowią komórki haploidalne, do których zaliczamy gamety – plemniki i komórki jajowe. Każda z nich zawiera pojedynczy zestaw 23 chromosomów umożliwiając przekazywanie informacji genetycznej w procesie rozmnażania płciowego.

Zgodnie z chromosomową teorią dziedziczności, sformułowaną w 1915 roku przez Morgana, w momencie zapłodnienia w każdej parze chromosomów homologicznych, zarówno

autosomalnych, jak i płciowych, jeden chromosom pochodzi od ojca (z plemnika), a drugi od matki (z komórki jajowej). Skutkiem takiego mechanizmu dziedziczenia jest obecność dwóch alleli dla każdego układu typu STR zlokalizowanego na chromosomach autosomalnych [18, str. 19-42].

Dla przykładu, jeśli matka w obrębie markera TH01 posiada genotyp 6,9, a ojciec 7,9, to ich potomstwo może odziedziczyć jedną z czterech kombinacji alleli: 6,7; 6,9; 7,9 lub 9,9. W sytuacji, gdy badany osobnik otrzyma od każdego z rodziców różny allel, układ taki określa się mianem heterozygotycznego. Z kolei przypadek, gdy oba allele są identyczne, nazywany jest homozygotycznym.

Autosomalne markery STR stanowią podstawę współczesnych analiz identyfikacyjnych oraz badań pokrewieństwa. Wartość dowodowa uzyskana z danego zestawu markerów STR zależy od liczby uwzględnionych loci, ich informatywności oraz jakości zastosowanych baz populacyjnych. W analizach materiału kostnego, gdzie często uzyskuje się jedynie częściowe profile DNA, dla zwiększenia mocy dowodowej, w zależności od analizowanego stopnia pokrewieństwa, łączy się dane pochodzące z autosomalnych STR z markerami zlokalizowanymi na chromosomach Y, X lub z sekwencjami mtDNA.

1.2.3. Markery STR chromosomu Y

W analizach chromosomu Y wykorzystuje się markery obecne wyłącznie w genomie męskim, ponieważ sam chromosom Y występuje jedynie u osobników płci męskiej. Charakterystyczne dla tego chromosomu jest to, że niemal cała jego długość podlega dziedziczeniu w formie niezmienionej, co określa się mianem regionu nierekombinującego (NRY) [45]. Z uwagi na brak homologicznej pary chromosomu Y u mężczyzn markery te określa się jako haploidalne. Otrzymany zestaw alleli, czyli haplotyp chromosomu Y, przekazywany jest z ojca na syna zasadniczo bez udziału rekombinacji, dzięki czemu kolejne pokolenia linii męskiej wykazują niemal identyczne profile. Wyjątek stanowią sporadyczne mutacje, które są źródłem jego naturalnej zmienności [18, str. 363-396].

Z uwagi na taki model dziedziczenia, chromosom Y stanowi charakterystyczny marker genetyczny dla określonej linii rodowej męskiej. Analizy obejmują głównie krótkie powtórzenia tandemowe zlokalizowane na chromosomie Y (Y-STR), analogiczne do autosomalnych STR, lecz zlokalizowane w regionie NRY [46 - 48]. W wyniku analizy uzyskuje się pojedynczy allel dla każdego locus, ponieważ chromosom Y występuje w pojedynczej kopii. Wyjątek stanowią loci wielokrotne, powstałe w wyniku duplikacji

fragmentów chromosomu, jak np. układ DYS385a/b (podwójny) czy DYS464a/b/c/d (poczwórny), w których poszczególne litery oznaczają kolejne kopie powtórzonego regionu. Unikatowość chromosomu Y jako elementu wyłącznie męskiego genomu sprawiła, że markery Y-STR znalazły szerokie zastosowanie w genetyce sądowej.

Główną funkcją markerów Y-STR jest umożliwienie rozróżniania osób wywodzących się z tej samej linii męskiej, co znajduje zastosowanie zarówno w identyfikacji sądowej, jak i w badaniach genealogicznych [49]. Markery haploidalne są szczególnie przydatne w analizach uzupełniających pokrewieństwo, np. w sytuacjach, w których brakuje materiału porównawczego od domniemanych rodziców, a dostępny jest jedynie materiał od krewnych pochodzących z tej samej linii męskiej. Uzyskane haplotypy poddaje się następnie analizie statystycznej, zestawiając ich częstość występowania z danymi referencyjnymi zaczerpniętymi z odpowiednich baz populacyjnych [50].

Choć wartość dowodowa badań Y-STR jest niższa niż analiz autosomalnych (ze względu na możliwość występowania identycznych haplotypów u osób niespokrewnionych), ich znaczenie pozostaje istotne [51]. W praktyce kryminalistycznej markery chromosomu Y stosuje się głównie w przypadkach, gdy w badanej próbce występuje mieszanina DNA pochodzącego od mężczyzny i kobiety [52]. Jak wskazują Yao oraz Qian i wsp. [53, 54], analiza Y-STR powinna być traktowana jako badanie przesiewowe we wszystkich sytuacjach, w których podejrzewa się obecność męskiego DNA, ale nie jest ono możliwe do wykrycia w klasycznej analizie autosomalnych STR. Dzięki dużej czułości tej metody możliwe jest wykrycie nawet niewielkich ilości DNA mężczyzny w próbkach zawierających znaczną przewagę materiału żeńskiego, co ma szczególne znaczenie w sprawach dotyczących przestępstw seksualnych.

Komercyjne zestawy do analizy haplotypów chromosomu Y dostępne na rynku obejmują od kilkunastu do ponad dwudziestu loci [55 - 57]. Zestawy te zawierają zarówno markery o umiarkowanym poziomie mutacyjności jak i szybko mutujące markery, które pozwalają odróżniać od siebie haplotypy mężczyzn blisko spokrewnionych [52, 58, 59]. Zastosowanie tych markerów znacząco zwiększa rozdzielczość analiz w obrębie pojedynczych linii męskich, poprawiając ich wartość dowodową w praktyce sądowej.

1.2.4. Markery STR chromosomu X

Dziedziczenie chromosomu X ma specyficzny charakter, odmienny od mechanizmu obserwowanego dla autosomów. W przypadku dziedziczenia w linii matka–cóрка, proces przebiega analogicznie do dziedziczenia chromosomów autosomalnych – matka przekazuje swojemu potomstwu jeden z dwóch posiadanych chromosomów X. Cóрка dziedziczy natomiast

drugi chromosom X po ojcu. Mężczyzna, posiadający jeden chromosom X i jeden chromosom Y, może przekazać chromosom X wyłącznie swoim córkom, podczas gdy kobieta przekazuje chromosom X wszystkim dzieciom, niezależnie od ich płci [18, str. 389]. W konsekwencji syn nie dziedziczy od ojca chromosomu X, natomiast każda córka otrzymuje od ojca zawsze ten sam chromosom X.

Ustalenie dziedziczego allele po stronie ojca jest więc proste, wiadomo, który allele został przekazany. W przypadku matki, która dysponuje dwoma chromosomami X, nie ma możliwości określenia z góry, z którego z nich pochodzi obserwowany allele. Co więcej, w trakcie mejozy chromosom X może ulegać procesowi crossing-over, podobnie jak chromosomy autosomalne, co dodatkowo komplikuje analizę dziedziczenia. Zjawisko to prowadzi do rekombinacji i zmienności układu alleli na chromosomach X przekazywanych córkom przez matki, co sprawia, że markery chromosomu X nie dziedziczą się w sposób tak prosty, jak markery chromosomu Y, dla którego analizy genetyczno-sądowe obejmują wyłącznie region NRY [60].

Syn dziedziczy swój jedyny chromosom X wyłącznie od matki, dlatego w jego profilu genetycznym można zidentyfikować allele będący bezpośrednim odzwierciedleniem jednego z matczynych chromosomów X [18, str. 87].

Analizy oparte na markerach X-STR (ang. X-chromosomal Short Tandem Repeats) mają szczególną wartość w badaniach genetycznych, gdyż umożliwiają ocenę niezależnego dziedziczenia określonych regionów chromosomu X. Ze względu na bliskie położenie części markerów względem siebie, wyróżnia się tzw. grupy sprzężone (linkage groups), w których loci są dziedziczone wspólnie w formie haplotypów, przekazywanych między pokoleniami [25].

Mimo że markery zlokalizowane na chromosomie X nie są tak powszechnie stosowane w genetyce sądowej jak autosomalne STR czy markery Y-STR, ich przydatność w trudnych przypadkach identyfikacyjnych została potwierdzona [61]. Analiza X-STR ma szczególne znaczenie w sytuacjach, gdy dostępny materiał porównawczy pochodzi od dalszych krewnych lub gdy relacje rodzinne są złożone.

Markery te znajdują zastosowanie m.in. w analizach pokrewieństwa w przypadkach spornych lub nietypowych, takich jak sprawy o kazirodstwo, gdzie domniemanym ojcem dziecka płci żeńskiej jest pozwany mężczyzna bądź jego syn lub w przypadkach spornego ojcostwa, gdy dziecko jest dziewczynką oraz gdy wątpliwości dotyczą ojcostwa wobec jednej z siostr [23].

1.2.5. Markery SNP

Polimorfizm pojedynczego nukleotydu (SNP) stanowi jeden z najpowszechniejszych rodzajów zmienności w genomie człowieka. Polega on na podstawieniach puryn i pirymidyn, a także na insercjach i delecjach w obrębie pojedynczych par zasad. Tego rodzaju zmienność ma szerokie zastosowanie nie tylko w genetyce sądowej, lecz także w badaniach nad uwarunkowaniami genetycznymi wielu chorób [62].

W porównaniu z markerami STR, markery SNP występują w znacznie większej liczbie w genomie ludzkim. Cechują się także niższym współczynnikiem mutacji oraz możliwością analizy bardzo krótkich amplikonów. Te właściwości sprawiają, że markery SNP znalazły zastosowanie w genetyce sądowej, szczególnie w przypadkach dotyczących identyfikacji nieznanymi szczątków ludzkich, gdzie materiał DNA jest silnie zdegradowany [63].

Z uwagi jednak na bialleliczny charakter loci SNP, liczba markerów koniecznych do przebadania w celu uzyskania mocy dyskryminacyjnej porównywalnej z analizą loci STR musi być znacznie większa [64]. Pomimo tego ograniczenia, markery SNP oferują istotne korzyści techniczne, a mianowicie mogą być analizowane przy użyciu różnych metod, w tym technologii mikromacierzy DNA wykorzystujących zjawisko hybrydyzacji. Co istotne, proces analizy SNP może być w dużym stopniu zautomatyzowany, co znacząco zwiększa jego wydajność i powtarzalność [18 str. 343; 65].

1.2.6. Analiza DNA mitochondrialnego

W każdej komórce ludzkiego organizmu występuje od kilkudziesięciu do kilkuset mitochondriów, z których każdy zawiera od czterech do dziesięciu kopii mtDNA [66]. Mitochondrialny DNA ma postać kolistej cząsteczki i charakteryzuje się znacznie większą odpornością na procesy degradacyjne niż DNA jądrowy [67, 68]. Duża liczba kopii tego genomu sprawia, że możliwe jest uzyskanie materiału nadającego się do analizy nawet z bardzo niewielkich ilości zdegradowanego materiału biologicznego. W konsekwencji sekwencjonowanie mtDNA bywa często jedynym sposobem uzyskania profilu genetycznego ze szczątków ludzkich [69].

Mitochondria są półautonomicznymi strukturami komórkowymi, których pochodzenie tłumaczy teoria endosymbiozy – zakładająca, że powstały one z dawnych wolnożyjących bakterii. W ich macierzy znajduje się własny materiał genetyczny o długości 16 569 par zasad, kodujący część białek łańcucha oddechowego oraz cząsteczki RNA: transportujące (tRNA) i rybosomalne (rRNA) [70]. W większości badań sądowych analizie poddaje się jedynie

niewielki, niekodujący fragment genomu mitochondrialnego – tzw. region kontrolny, stanowiący około 7% całkowitej długości mtDNA. W jego obrębie wyróżnia się trzy odcinki o podwyższonym poziomie zmienności, zwane regionami hiperzmiennymi (HV, hypervariable region): HV1 (pozycje nukleotydowe 16024–16365 w odniesieniu do łańcucha L), HV2 (pozycje 73–340) oraz rzadziej analizowany HV3 (438–574) [71, 72].

W przeciwieństwie do genomu jądrowego, DNA mitochondrialny dziedziczony jest wyłącznie w linii matczynej. Oznacza to, że wszystkie osoby spokrewnione poprzez matkę – w tym jej dzieci i rodzeństwo – posiadają identyczną sekwencję mtDNA, o ile nie zaszły w niej mutacje [73]. Ze względu na ograniczoną skuteczność mechanizmów naprawczych mtDNA, jego genom mutuje znacznie szybciej niż DNA jądrowy. Największe międzyosobnicze różnicowanie obserwuje się w regionach kontrolnych HV1 i HV2 [74, 75].

W 2014 roku Komisja DNA Międzynarodowego Towarzystwa Genetyki Sądowej (ISFG) opublikowała zaktualizowane wytyczne dotyczące analizy mtDNA. Dokument zawierał szczegółowe rekomendacje odnoszące się zarówno do dobrej praktyki laboratoryjnej, jak i wyboru analizowanych fragmentów, metod amplifikacji czy sekwencjonowania. Ujęto w nim również zasady stosowania sekwencji referencyjnych, interpretacji heteroplazmii, przypisywania haplogrup oraz prowadzenia baz danych [76, 77]. Zalecenia te są cyklicznie uzupełniane i dostosowywane do pojawiających się metod technologicznych [78].

W rutynowych analizach do sekwencjonowania regionów hiperzmiennych mtDNA wciąż powszechnie stosuje się klasyczną metodę enzymatyczną Sangera [79]. Technika ta umożliwia szybkie odczytanie kolejności nukleotydów w badanym fragmencie i pozwala uzyskać sekwencję o długości od 25 do około 1200 par zasad [80].

W roku 2005 rozpoczęła się nowa era w analizie materiału genetycznego – tzw. sekwencjonowanie następnej generacji (NGS). Metody zwane również masowym sekwencjonowaniem równoległym umożliwiają uzyskanie bardzo licznych odczytów – jednocześnie nawet milionów krótkich fragmentów [81].

Wynik analizy mtDNA przedstawia się w postaci listy pozycji nukleotydowych, w których wykazano różnice względem światowej sekwencji referencyjnej, znanej jako revised Cambridge Reference Sequence (rCRS), określanej również mianem sekwencji Andersona – od nazwiska autora pierwszej publikacji opisującej pełny genom mitochondrialny [18 str. 375-388; 79, 82, 83].

Integralną częścią interpretacji wyniku jest wskazanie przebadanego zakresu sekwencji. Przykładowy zapis mutacji 73G, 315.1C będzie miał różną wartość interpretacyjną

w zależności od tego, czy analiza obejmowała jedynie region HV2, czy też zarówno HV1 jak i HV2 [84].

1.3. Analiza pochodzenia biogeograficznego

Pochodzenie biogeograficzne osobnika można określić na podstawie analizy haplotypów mtDNA lub chromosomu Y. Oba te markery, które są dziedziczone odpowiednio w linii matczynej i ojcowskiej, stanowią istotne narzędzie w badaniach z zakresu genetyki populacyjnej i identyfikacyjnej.

MtDNA, oprócz zastosowania w identyfikacji osób spokrewnionych w linii matczynej, pozwala także na oszacowanie pochodzenia biogeograficznego. Wynika to z dużego zróżnicowania sekwencji mtDNA pomiędzy populacjami, obserwowanego zarówno w skali kontynentów, jak i w obrębie mniejszych regionów [85]. Określenie haplogrupy mtDNA może być dokonane poprzez analizę pełnej sekwencji genomu mitochondrialnego lub identyfikację specyficznych mutacji allelicznych definiujących daną haplogrupę [86]. W praktyce badań identyfikacyjnych najczęściej analizowane są jednak fragmenty HVI i HVII regionu kontrolnego, ze względu na ich wysoki stopień polimorfizmu i bogatą bazę danych porównawczych.

Analogicznie, określenie haplogrupy chromosomu Y umożliwia wyznaczenie pochodzenia biogeograficznego w linii ojcowskiej. Choć najbardziej precyzyjną metodą byłoby pełne sekwencjonowanie chromosomu Y, procedura ta wymaga dużej ilości dobrze zachowanego DNA jądrowego oraz znacznych nakładów finansowych [87]. W przypadkach, gdy nie ma potrzeby ustalenia szczegółowej subhaplogrupy, stosuje się analizę markerów typu Y-STR. Na ich podstawie można oszacować przynależność do określonej haplogrupy przy użyciu ogólnodostępnych narzędzi bioinformatycznych, takich jak NEVGEN Genealogy Tools v1.1. [88].

Haplotypy zarówno mtDNA, jak i chromosomu Y można przyporządkować do konkretnych haplogrup, czyli grup wywodzących się od wspólnego przodka, które występują z różną częstością w poszczególnych populacjach, co umożliwia określenie geograficznego pochodzenia danej osoby w linii matczynej lub ojcowskiej. Przykładowo, rozkład haplogrup mitochondrialnych w populacjach Afryki [89], rdzennych Amerykanów [90] czy w rejonach zachodniej i wschodniej Eurazji [91, 92] różni się znacząco, co pozwala na dość precyzyjne przypisanie próbki do danego obszaru biogeograficznego.

Dodatkowo możliwe jest wyróżnienie bardziej szczegółowych podgrup haplogrup, które pozwalają zawęzić obszar pochodzenia do poziomu regionalnego, np. Europy Środkowej lub

Wschodniej. W interpretacji wyników należy jednak zachować ostrożność w odniesieniu do populacji niejednorodnych genetycznie. W takich przypadkach, jak w populacjach Ameryki Łacińskiej, obserwuje się często zróżnicowanie pochodzenia linii matczynej i ojcowskiej – liniom żeńskim pochodzącym od rdzennych mieszkańców kontynentu towarzyszą linie męskie wywodzące się z Europy [93, 94].

1.4. Proces degradacji DNA

Proces degradacji kwasów nukleinowych rozpoczyna się natychmiast po śmierci organizmu. W środowisku, w którym nie występują warunki sprzyjające konserwacji, takie jak wysuszenie lub zamrożenie, rozkład tkanek postępuje bardzo szybko. Pierwszym etapem jest autoliza, będąca konsekwencją braku tlenu niezbędnego do prowadzenia procesów metabolicznych, co powoduje przejście komórek na oddychanie beztlenowe i w konsekwencji prowadzi do obniżenia pH cytoplazmy. Zakwaszenie wnętrza komórki skutkuje rozerwaniem błon komórkowych i uwolnieniem lizosomów, zawierających enzymy hydrolityczne, które doprowadzają do samorozpadu komórek [95 - 97].

W miarę postępu autolizy dochodzi do uszkodzenia błony jądrowej, a w konsekwencji do uwolnienia DNA, które staje się podatne na działanie nukleaz – enzymów odpowiedzialnych za rozcinanie wiązań fosfodiesterowych w strukturze kwasów nukleinowych. W degradacji tej uczestniczą zarówno endonukleazy, które przecinają łańcuch DNA wewnątrz cząsteczki, jak i egzonukleazy, odcinające nukleotydy od końców nici. Równocześnie aktywność proteaz prowadzi do trawienia histonów, co sprzyja losowej fragmentacji genomu. Procesy te zachodzą najintensywniej w komórkach wrażliwych na niedotlenienie, szczególnie narażone są komórki ośrodkowego układu nerwowego [95, 96].

Wraz z postępującą degradacją tkanek miękkich dochodzi do rozerwania nici DNA, często do długości poniżej 200 par zasad, co znacznie utrudnia analizę genetyczną materiału pochodzącego z wysoko zdegradowanych próbek. Zdecydowanie wolniej rozkładowi ulega DNA obecne w tkankach kostnych, co wynika z ochronnych właściwości struktury kości – niskiej zawartości wody, ograniczonego dostępu enzymów degradacyjnych oraz osłony przed promieniowaniem ultrafioletowym, które może powodować dimeryzację zasad azotowych i uszkodzenia wiązań krzyżowych [98].

Oprócz czynników fizycznych i chemicznych, duży wpływ na rozkład DNA mają mikroorganizmy takie jak bakterie, grzyby, a także drobne bezkręgowce. Produkowane przez nie enzymy przyczyniają się do degradacji kwasów nukleinowych już na wczesnych etapach pośmiertnych [99, 100]. Ponadto, w glebie zachodzą liczne reakcje chemiczne, takie jak

procesy hydrolityczne i oksydacyjne, które, choć postępują stosunkowo powoli, mają charakter nieodwracalny [101, 102]. W odróżnieniu od organizmu żywego, w którym istnieją systemy naprawy DNA, po śmierci uszkodzenia te kumulują się, prowadząc stopniowo do całkowitej utraty informacji o pierwotnej sekwencji nukleotydowej [97].

Zjawiska oksydacyjne mogą również powodować powstawanie zmodyfikowanych zasad azotowych, m.in. 8-hydroksyguaniny, która podczas replikacji jest błędnie rozpoznawana przez polimerazę jako adenina. W efekcie do nowej nici wbudowywana jest tymina, co prowadzi do transwersji typu G→T [103, 104].

Do czynników środowiskowych o największym znaczeniu degradacyjnym zalicza się odczyn i typ gleby, jej wilgotność oraz dostępność tlenu. Parametry te wpływają na przebieg procesów chemicznych modyfikujących strukturę puryn i pirymidyn – podstawowych składników nici DNA [97, 105].

Pomimo znacznego rozkładu materiału biologicznego, technika PCR umożliwia amplifikację zachowanych fragmentów DNA, nawet w przypadku gdy ich długość nie przekracza 100 – 200 par zasad [106]. W praktyce sądowej takie szczątkowe sekwencje stanowią często jedyne źródło materiału genetycznego możliwego do wykorzystania w analizie identyfikacyjnej. Zgodnie z zaleceniami DNA Commission of the ISFG interpretacja wyników uzyskiwanych ze zdegradowanego materiału wymaga szczególnej ostrożności. Powinna zostać przeprowadzona z uwzględnieniem ryzyka wystąpienia zjawisk typu drop-out i drop-in, co pozwala uniknąć błędnej kwalifikacji profili niepełnych i zwiększa wiarygodność wniosków identyfikacyjnych [107, 108].

1.5. Proces identyfikacji osobniczej NN zwłok i szczątków ludzkich

Procedura postępowania w przypadku ujawnienia NN zwłok lub szczątków ludzkich wymaga podjęcia szeregu skoordynowanych działań których nadrzędnym celem jest uzyskanie wystarczającej liczby informacji pozwalających na ustalenie ich tożsamości. W pierwszej kolejności są to: oględziny miejsca ujawnienia zwłok, oględziny sądowo-lekarskie NN zwłok lub szczątków ludzkich, pobranie odcisków linii papilarnych palców i dłoni oraz wykonanie dokumentacji fotograficznej. Kolejnym etapem, po przewiezieniu do prosektorium, jest przeprowadzenie sekcji zwłok w celu ustalenia okoliczności i przyczyny zgonu. W przypadku szczątków ludzkich przeprowadza się oględziny z udziałem lekarza medycyny sądowej oraz w miarę możliwości z udziałem antropologa. Na tym etapie uzasadnione jest sporządzenie pełnej dokumentacji fotograficznej, wykonanie badań obrazowych (RTG lub TK) oraz pobranie

wyselekcjonowanego materiału genetycznego. W niektórych przypadkach, na potrzeby badań identyfikacyjnych, zabezpiecza się i przeprowadza macerację czaszki.

Na potrzeby procesu identyfikacji niezbędne jest również zgromadzenie danych obejmujących wszelkie dostępne informacje mogące przyczynić się do ustalenia tożsamości osoby zmarłej. Zbierane są dane dotyczące relacji rodzinnych, cech szczególnych, dokumentacji stomatologicznej i medycznej, zdjęć osoby zmarłej, przedmiotów osobistych, odcisków palców oraz obrazów radiologicznych. Na tym etapie istotne jest także pozyskanie próbek referencyjnych od krewnych z najbliższej linii pokrewieństwa, ponieważ im dalsze pokrewieństwo, tym mniejsze podobieństwo genetyczne. W wyjątkowych przypadkach w celu uzyskania materiału porównawczego konieczne jest na przykład przeprowadzenie ekshumacji członków rodziny zmarłej osoby.

W zależności od okoliczności znalezienia zwłok stosuje się różne metody identyfikacji. Mają one różny stopień wiarygodności. Identyfikację zwłok należy rozpocząć od metod najbardziej wiarygodnych do tylko sugerujących tożsamość. W pierwszej kolejności jest to identyfikacja daktyloskopijna, odontologiczna i genetyczna. W dalszej kolejności są to dane medyczne, tatuaże, blizny, implanty. Ostatnim etapem jest integracja danych co w założeniu ma umożliwić jednoznaczną identyfikację osoby zmarłej. Ważną rolę w procesie identyfikacji odgrywa również baza danych DNA [109].

W przypadku identyfikacji NN zwłok i szczątków z materiałów archiwalnych, w szczególności pochodzących z masowych egzekucji lub pochówków o charakterze represyjnym, klasyczny schemat identyfikacji osobniczej NN zwłok ulega istotnym modyfikacjom. Proces ten realizowany jest zazwyczaj wiele dekad po śmierci ofiar, w warunkach znacznego ubytku danych *ante-mortem* (łac. przed śmiercią) oraz przy ograniczonej dostępności materiału porównawczego od najbliższych krewnych. W tego typu sprawach źródłem informacji są archiwa, listy represjonowanych, relacje rodzinne oraz dane genealogiczne rekonstruowane na podstawie przekazów ustnych i dokumentów wtórnych – informacje te bardzo często są niepełne, fragmentaryczne. Jak wskazują doświadczenia polskich zespołów prowadzących prace ekshumacyjne i identyfikacyjne ofiar represji totalitarnych, identyfikacja taka ma charakter długofalowy i interdyscyplinarny, a jej skuteczność zależy od ścisłej integracji danych archeologicznych, antropologicznych, historycznych i genetycznych [110].

Należy podkreślić, że w przypadku identyfikacji ofiar katastrof masowych stosuje się wystandardyzowane metody postępowania rekomendowane przez Międzynarodową Organizację Policji (Interpol) dla wszelkiego rodzaju katastrof, niezależnie od ich przyczyny i skali [111, 112].

Wyniki analiz genetycznych są interpretowane z zastosowaniem modeli biostatystycznych, które pozwalają oszacować tzw. siłę dowodu uzyskaną na poparcie hipotezy o tożsamości osoby badanej. Obliczenia te oparte są na szacowanej częstości występowania danego genotypu w populacji oraz na prawdopodobieństwie jego przypadkowego powtórzenia u osoby niespokrewnionej [18, str. 229-258]. Prawdopodobieństwo, że dwie niespokrewnione osoby będą posiadały identyczne profile DNA, jest skrajnie niskie, co stanowi podstawę wysokiej wiarygodności tej metody.

Na podstawie obliczeń biostatystycznych formułowana jest końcowa opinia, której zasadniczym elementem jest jednoznaczne potwierdzenie lub wykluczenie tożsamości osoby badanej, bądź – w przypadkach niejednoznacznych – wskazanie, że dostępne dane nie pozwalają na ostateczne rozstrzygnięcie. Raportowanie wyników analizy jest ujęte w wytycznych Scientific Working Group on DNA Analysis Methods (SWGDM) [113].

1.6. Genetyczne bazy populacyjne

W analizach genetycznych prowadzonych w kontekście sądowym, zasadniczym elementem zapewniającym wiarygodność uzyskiwanych wyników jest zastosowanie właściwych danych referencyjnych, tj. baz populacyjnych, opisujących częstości alleli dla analizowanych markerów genetycznych. Wszystkie obliczenia biostatystyczne wykonywane są na podstawie właśnie tych danych populacyjnych. By uzyskać prawidłowe wyniki należy odnieść się do właściwej populacji referencyjnej, a baza danych musi spełniać rygorystyczne wymagania pod względem reprezentatywności próby oraz spełniać określone kryteria statystyczne [114].

Częstości alleli poszczególnych markerów mogą różnić się pomiędzy populacjami nawet w obrębie tego samego regionu geograficznego, co wynika z czynników historycznych, demograficznych i genetycznych, takich jak dryf genetyczny, efekty założyciela czy izolacja populacyjna [115, 116]. Niewłaściwy dobór bazy referencyjnej może prowadzić do zaniżenia lub zawyżenia wartości wskaźników statystycznych, a w konsekwencji – do błędnej interpretacji siły dowodu.

Zgodnie z zaleceniami DNA Commission of the ISFG populacyjne bazy danych powinny być tworzone w oparciu o reprezentatywne próbki pochodzące od osób niespokrewnionych, zamieszkujących dany obszar geograficzny, przy czym liczebność próby dla każdego markera powinna zapewniać wiarygodne oszacowanie częstości alleli [117]. Dane takie publikowane są w literaturze naukowej i rejestrowane w publicznych bazach, np. referencyjna baza częstości identyfikacyjnych markerów STR (STRidER) [118].

W przypadku analiz pokrewieństwa zaleca się stosowanie bazy populacyjnej możliwie najbardziej zbliżonej etnicznie i geograficznie do osób badanych. W razie braku odpowiedniej bazy można rozważyć użycie danych populacji zbliżonej etnicznie do analizowanej, jednak interpretacja wyników powinna wtedy uwzględniać możliwe ograniczenia wynikające z różnic populacyjnych [115, 119]. Dobór właściwej bazy populacyjnej stanowi zatem integralny element całego procesu analizy genetycznej i bezpośrednio wpływa na rzetelność uzyskanych wniosków w kontekście sądowym i identyfikacyjnym.

Ponadto należy pamiętać, że w przypadku niepełnych profili genetycznych, istnieje ryzyko przypadkowej zgodności profili genetycznych. Im mniej loci oznaczonych w profilu, tym prawdopodobieństwo jest wyższe, co stanowi komplikację przy identyfikacji oraz jest poważnym ograniczeniem badania [120].

2. Założenia i cel pracy

Celem niniejszej pracy była optymalizacja metod identyfikacji genetycznej NN zwłok i szczątków ludzkich, ze szczególnym uwzględnieniem przypadków dalszego pokrewieństwa, charakterystycznych dla badań identyfikacyjnych materiału archiwalnego. Model ten zakładał kompleksowe podejście obejmujące analizę markerów autosomalnych, chromosomu Y oraz markerów SNP, a także wykorzystanie narzędzi biostatystycznych i symulacyjnych do oceny pokrewieństwa.

Badania oparto na rzeczywistym materiale kostnym pochodzącym z grobu masowego z okresu Wielkiego Terroru, odkrytego w 2021 roku na terenie Adżarii w Gruzji, co pozwoliło zweryfikować skuteczność stosowanych metod.

Założono, że zastosowanie kilku niezależnych zestawów markerów (GlobalFiler™, KinFinder, Yfiler™ Plus, Precision ID Identity) oraz wykorzystanie zróżnicowanych algorytmów obliczeniowych (Familias, FamLink, KinBN, pedbuildr) pozwoli na zwiększenie wiarygodności analiz i umożliwi rozstrzygnięcie przypadków, w których klasyczne metody dają niejednoznaczne wyniki.

Przyjęto, że kluczowym elementem procesu identyfikacji będzie porównanie skuteczności różnych konfiguracji markerów oraz ocena ich przydatności w przypadkach gdzie dostępny jest zdegradowany materiał biologiczny i analizowane są dalsze stopnie pokrewieństwa.

W pracy postawiono następujące cele szczegółowe:

1. Opracowanie schematu decyzyjnego pozwalającego na maksymalne wykorzystanie uzyskanego materiału DNA o różnym stopniu degradacji.
2. Ocena skuteczności poszczególnych zestawów markerów w kontekście materiału kostnego.
3. Weryfikacja możliwości zastosowania programu do symulacji pokrewieństw i wspomagania interpretacji wyników w złożonych strukturach genealogicznych.
4. Przeprowadzenie obliczeń przy wykorzystaniu różnych programów statystycznych i ocena ich przydatności.
5. Wypracowanie rekomendacji dotyczących optymalnego zestawu metod i narzędzi dla potrzeb identyfikacji genetycznej NN zwłok i szczątków z materiałów archiwalnych.

3. Materiał i metody

3.1. Materiał badawczy

Materiał kostny pobrany ze szczątków został przetransportowany do laboratorium PUM: 21 fragmentów lewych kości skroniowych i siedem fragmentów prawych kości skroniowych. Dodatkowo do analizy genetycznej przekazano fragment lewej kości piszczelowej i prawej kości ramiennej w celu potwierdzenia powiązań szkieletowych.

Próbki referencyjne od 58 osób należących do 20 rodzin (oznaczonych jako Rodzina 1 – Rodzina 20) również przekazano do PUM.

3.2. Zastosowane metody badawcze

Wszystkie analizy laboratoryjne, z wyjątkiem elektroforezy dla zestawu KinFinder, zostały przeprowadzone w Zakładzie Genomiki i Genetyki Sądowej PUM, działającym zgodnie z wymaganiami normy PN EN ISO/IEC 17025: 2018:02.

Analizę fragmentów za pomocą elektroforezy kapilarnej dla zestawu KinFinder przeprowadzono w laboratorium GenMed działającego zgodnie z wymaganiami normy PN EN ISO/IEC 17025: 2018:02.

Analiza elektroforegramów zestawu KinFinder oraz analizy statystyczne zostały przeprowadzone w Katedrze i Zakładzie Medycyny Sądowej Warszawskiego Uniwersytetu Medycznego.

Prace laboratoryjne zostały przeprowadzone przy zachowaniu rygorystycznych protokołów aseptycznych i antyseptycznych. Aby zminimalizować ryzyko zanieczyszczenia, przez cały czas stosowano środki ochrony osobistej, w tym rękawiczki nitrylowe, zarekawki, fartuchy ochronne, maski jednorazowe i specjalne kombinezony do czyszczenia kości. Ze względu na potencjalnie wysoki stopień degradacji materiału badawczego i niską zawartość DNA, procedury dekontaminacji przeprowadzono zgodnie ze standardami stosowanymi w genetyce sądowej. Na każdym etapie procesu zastosowano kontrole pozytywne (krew suchą zabezpieczoną na kartach Flinders Technology Associates (FTA) oraz kontrolne DNA zawarte w komercyjnych zestawach) i kontrole negatywne (woda dejonizowana klasy molekularnej). Zarówno próbki pobrane z materiału dowodowego, jak i materiały referencyjne były analizowane w dwóch powtórzeniach. Te rygorystyczne środki zapewniły wysokie standardy kontroli zanieczyszczeń, umożliwiając wiarygodną analizę.

O ile nie zaznaczono inaczej, wszystkie etapy laboratoryjne przeprowadzono zgodnie z protokołami odpowiednich producentów.

Obliczenia statystyczne pokrewieństwa zostały wykonane przy użyciu dwóch programów komputerowych – Familias oraz KinBN – które opierają się na odmiennych modelach analizy statystycznej. Zastosowanie dwóch niezależnych narzędzi umożliwiło wzajemną weryfikację uzyskanych rezultatów i zwiększyło wiarygodność wniosków, ograniczając ryzyko błędów wynikających ze specyfiki pojedynczego modelu obliczeniowego. Z uwagi na brak gruzińskich baz populacyjnych do obliczeń statystycznych wykorzystano bazy zgodnie z dostępnością. Dla obliczeń z użyciem GlobalFiler™ wykorzystano bazę Azerbejdżańską [121]. Dla zestawu KinFinder, z uwagi na to, iż niedawno został wprowadzony do produkcji i nie jest jeszcze dostępny komercyjnie, korzystano z bazy populacji Polski (GenMed, dane nieopublikowane) z wykorzystaniem 48 markerów spoza zestawu CODIS-ESS-SE33. Obliczenia oparte o identyfikacyjne markery SNP (iiSNP) zostały wykonane z wykorzystaniem bazy ogólnościowej dostępnej na stronie <https://www.ncbi.nlm.nih.gov/snp/>.

Analizy pokrewieństwa przy wykorzystaniu programu komputerowego Familias [122] wykonano w sposób etapowy. W pierwszej kolejności dokonano porównań wszystkich profili autosomalnych w zakładce programu do identyfikacji ofiar katastrof (DVI), a następnie pogłębiono obliczenia dla najbardziej prawdopodobnych dopasowań rodzin i szczątków. Równolegle w arkuszu Excel zestawiono profile Y-STR, zgodnie z zasadami porównań i raportowania haplotypów SWGDAM [123], co pozwoliło zidentyfikować jeden przypadek zgodności, który nie został wskazany na podstawie markerów autosomalnych. Obliczenia statystyczne haplotypów Y-STR zostały wykonane przy użyciu narzędzia dostępnego w referencyjnej bazie haplotypów chromosomu Y (YHRD) [58], która stanowi międzynarodowy punkt odniesienia i standard w analizach porównawczych profili chromosomu Y [124]. Obliczenia wykonano w oparciu o bazę 90 260 haplotypów oznaczonych w zakresie Y17 z obszaru Euroazjatyckiego. Zakres analizy został wybrany zgodnie z zaleceniami Polskojęzycznej Grupy ISFG [125]. Wyniki obliczeń stanowiły punkt wyjścia do dalszych analiz.

Dane uzyskane z analizy chromosomu Y wykorzystano także do oceny składu haplogrup chromosomów Y osób odnalezionych w grobie zbiorowym oraz porównanie tych wyników z populacjami referencyjnymi z okolicznych regionów. Celem tego etapu była ocena wiarygodności hipotezy historycznej dotyczącej pochodzenia tej grupy. Szacowanie haplogrup chromosomu Y zostało przeprowadzone bez dodatkowych kosztów, ponieważ opierało się wyłącznie na markerach uwzględnionych w zestawie Yfiler™ Plus.

W dalszej kolejności wytypowano rodziny i odpowiadające im szczątki do badań z wykorzystaniem zestawu odczynników KinFinder (GenMed). Kryterium kwalifikacji

stanowiła wartość ilorazu wiarygodności (LR) poniżej 100 000 uzyskana w analizach STR układów z GlobalFiler™. W jednym przypadku (Rodzina 12) z uwagi na niezgodności w układach Y-STR pomiędzy członkami tej rodziny, rozszerzono analizę o zestaw KinFinder pomimo osiągnięcia wartości LR powyżej progu. W przypadkach o najniższych wynikach statystycznych analizę uzupełniono o zestaw Precision ID Identity Panel (Thermo Fisher Scientific) obejmujący 90 autosomalnych markerów SNP. Dla panelu SNP przyjęto kryteria interpretacyjne zgodne z SWGDAM [126].

Wszystkie obliczenia prowadzono zarówno przy założeniu niezależności markerów, jak również z uwzględnieniem sprzężeń genetycznych, co pozwoliło ocenić wpływ tego zjawiska na końcowe wartości LR. Uwzględnianie powiązań pomiędzy markerami jest zagadnieniem istotnym, gdyż loci położone blisko siebie na tym samym chromosomie nie dziedziczą się niezależnie, lecz wykazują tendencję do współdziedziczenia w kolejnych pokoleniach [116, p. 259] Zjawisko to określane jest mianem sprzężenia genetycznego i opisywane jest w jednostkach odległości genetycznej – centymorganach (cM) [127].

Uwzględniono również symulacje alternatywnych struktur genealogicznych w celu weryfikacji uzyskanych dopasowań. Analizy te przeprowadzono z wykorzystaniem pakietu pedbuildr, który umożliwia rekonstrukcję rodowodów oraz modelowanie prawdopodobieństw współdzielenia alleli identycznych z pochodzenia. Równolegle, przy pomocy programu Familias, wykonano symulacje rozkładu wartości LR dla zadanych scenariuszy pokrewieństwa, co pozwoliło na ocenę zgodności uzyskanych wyników z teoretycznie oczekiwanymi przedziałami wartości.

Połączenie obu podejść – rekonstrukcji alternatywnych struktur rodzinnych oraz symulacji wartości LR – umożliwiło nie tylko wskazanie najbardziej prawdopodobnych relacji rodzinnych, ale także ocenę stabilności wyników w świetle potencjalnych wariantów genealogicznych.

3.3. Przygotowanie materiału kostnego do izolacji DNA

Materiał kostny został oczyszczony mechanicznie z wykorzystaniem precyzyjnej miniszlifierki i jednorazowych tarcz szlifierskich. Następnie oczyszczone fragmenty kostne zostały poddane myciu za pomocą 0,5% Tween 20. Tak przygotowany materiał kruszono w ciekłym azocie z użyciem młynka kriogenicznego.

3.4. Izolacja DNA

DNA wyekstrahowano z 25 mg proszku kostnego przy użyciu zestawu Bone DNA Extraction Kit oraz urządzenia Maxwell® FSC Instrument (oba Promega).

DNA z materiału referencyjnego zostało wyekstrahowane przy użyciu zestawu PrepFiler™ (Thermo Fisher Scientific).

3.5. Oznaczenie stężenia DNA, stopnia degradacji i obecności inhibitorów reakcji PCR

Ilość i jakość wyizolowanego DNA oceniano przy użyciu zestawu Quantifiler™ TRIO DNA Quantification Kit (Thermo Fisher Scientific), który umożliwia ilościowe oznaczenie zarówno długich (T. Large, 214 pz), jak i krótkich (T. Small, 80 pz) fragmentów ludzkiego DNA autosomalnego, a także fragmentu specyficznego dla chromosomu Y (75 pz), wykorzystując metodę PCR w czasie rzeczywistym z zastosowaniem sond TaqMan. Marker specyficzny dla płci męskiej został wyłączony z niniejszego badania, ponieważ jego przydatność w analizie materiału zdegradowanego jest istotnie ograniczona. Marker ten został pierwotnie zaprojektowany głównie w celu wykrywania obecności składnika męskiego w próbce, jako element zestawu przeznaczonego do celów sądowych [128]. Wskaźnik degradacji obliczano jako stosunek wartości T. Small do T. Large i wykorzystywano do oceny stopnia degradacji DNA. Wartość wskaźnika degradacji powyżej 10 interpretowano jako wskazujący wysoką degradację w próbce, wartości od 1 do 10 wskazywały degradację od niewielkiej do umiarkowanej. Obecność inhibitorów reakcji PCR w ekstrakcie określano na podstawie wartości Ct wewnętrznej kontroli PCR (IPC), przy czym wartości równe 31 lub wyższe wskazywały na obecność inhibicji, zgodnie z wytycznymi producenta oraz wewnętrzną walidacją. We wszystkich kolejnych etapach, jeśli ekstrakty DNA nie spełniały zalecanych ilości wejściowych określonych w instrukcjach użytkownika, stosowano maksymalną dostępną objętość DNA.

3.6. Amplifikowanie fragmentów STR w systemie GlobalFiler™

Wyizolowany DNA amplifikowano metodą PCR z użyciem zestawu licencjonowanych odczynników GlobalFiler™ (Thermo Fisher Scientific). Zestaw ten umożliwił identyfikację alleli dla 21 autosomalnych markerów STR (D3S1358, D1S1656, D2S441, D10S1248, D13S317, D16S539, D18S51, D2S1338, CSF1PO, TH01, vWA, D21S11, D7S820, D5S818, TPOX, D8S1179, D12S391, D19S433, SE33, D22S1045, FGA), jednego markera Y-STR

(DYS391), markera insercji/delecji chromosomu Y (Y indel) oraz fragmentu genu amelogeniny.

3.7. Amplifikowanie fragmentów STR w systemie KinFinder

System KinFinder (GenMed) umożliwia oznaczenie 50 loci STR (w tym 48 spoza zestawu CODIS-ESS-SE33) zgrupowanych w dwóch multipleksach. Multipleks PCR A umożliwia amplifikację loci: D04L885, D14L953, D06L106, D16L554, D01L267, D03L115, D2N43, D12L794, D02L114, D1S1656, D05L169, D14L699, D3A57, D8S1132, D3N61, D05L140, D13S742, D2L174, D17L432, D1N16, D15L495, D21L291, D14L785, D05L297, D01L569 oraz locus amelogeniny, natomiast multipleks PCR B: D09L159, D02L221, D02L142, D8A26, D05L113, D14L276, D17L255, D20L226, D08L110, D16L732, D01L217, D03L109, D07L144, D7S3048, D12S391, D03L194, D07L101, D10L126, D3N54, D01L215, D12L908, D07L147, D12L630, D10S2325, D07L134. Reakcję obu multipleksów przeprowadzono w objętości 10 µl – na którą składało się 3 µl mieszanki enzymów (HOT FIREPol® MultiPlex Mix, Solis Biodyne), 2 µl odpowiedniego zestawu starterów (0,05 µM) oraz 1 ng matrycowego DNA. Amplifikację przeprowadzono zgodnie z zaleceniami producenta [129].

3.8. Amplifikowanie fragmentów STR w systemie Yfiler™ Plus

Z wykorzystaniem zestawu Yfiler™ Plus (Thermo Fisher Scientific) amplifikowano 27 markerów zlokalizowanych na chromosomie Y (DYS438, DYS627, DYS458, DYS437, DYS391, DYS392, DYS635, DYS19, DYS390, DYS439, DYS456, DYS393, DYS449, DYS387S1 a**b**, DYS576, DYS460, DYS533, DYS389 II, DYS570, DYS385 a**b**, DYS481, YGATA H4, DYS518, DYS448).

3.9. Konstruowanie bibliotek SNP w systemie Precision ID Identity

Precision ID Identity umożliwia oznaczenie 90 SNP autosomalnych (rs2046361, rs10776839, rs354439, rs717302, rs993934, rs901398, rs4364205, rs6955448, rs560681, rs1490413, rs2076848, rs159606, rs10488710, rs9951171, rs729172, rs10773760, rs1335873, rs6444724, rs214955, rs914165, rs2016276, rs740910, rs1736442, rs1357617, rs7520386, rs1523537, rs964681, rs10495407, rs9905977, rs876724, rs2269355, rs2111980, rs10092491, rs321198, rs2342747, rs873196, rs12997453, rs1454361, rs1028528, rs4530059, rs2830795, rs727811, rs1031825, rs1821380, rs1109037, rs1886510, rs7704770, rs2831700, rs1528460, rs740598, rs7041158, rs1355366, rs907100, rs891700, rs917118, rs1463729, rs2292972, rs722098, rs3780962, rs4288409, rs735155, rs445251, rs987640, rs737681, rs2040411, rs338882,

rs430046, rs938283, rs1005533, rs1872575, rs1015250, rs6811238, rs251934, rs576261, rs221956, rs4847034, rs1979255, rs1058083, rs722290, rs733164, rs2056277, rs1493232, rs13218440, rs1360288, rs719366, rs1382387, rs826472, rs1024116, rs1498553, rs1413212) i 34 SNP chromosomu Y (rs20320L298, rs17269816P202, rs2032602, rs17222573, rs16981290, rs2534636, rs3911, rs9341278, rs35284970, rs2032673, rs2032636, rs17250845, rs17842518, rs2032652, rs3848982P256, rs9786139M479, rs2032595, rs2319818, rs16980426, rs3900, rs2032658, rs2033003, rs2032631, rs13447443, rs2032624, rs9786184, rs4141886, rs17306671, rs2032599, rs8179021).

Biblioteki zostały przygotowane przy użyciu Precision ID Identity Panel, Precision ID Library Kit i IonCode™ Barcode Adapters (Thermo Fisher Scientific), z wykorzystaniem AMPure XP Beads (Beckmann Coulter) na etapie oczyszczania. Następnie biblioteki zostały poddane kwantyfikacji przy użyciu zestawu Ion Library TaqMan™ Quantitation Kit, rozcieńczone do 30 pM, poddane dalszej obróbce i załadowane na chipy Ion 530™ przy użyciu zestawu Ion S5™ Precision ID Chef & Sequencing Kit za pomocą urządzenia Ion Chef™ (wszystkie Thermo Fisher Scientific).

3.10. Analiza profili genetycznych

Amplifikaty uzyskane przy wykorzystaniu zestawów GlobalFiler™ oraz Yfiler™ Plus poddano elektroforezie w polimerze POP-4™ w aparacie 3500 Genetic Analyzer (Thermo Fisher Scientific), przy zastosowaniu Data Collection Software v 4.3, a następnie analizowano przy użyciu programu GeneMapper® ID-X Software v1.6 (Thermo Fisher Scientific), z zastosowaniem progu 100 RFU (ang. Relative Fluorescence Unit, względne jednostki fluorescencji) zgodnie ze zwalidowaną metodą.

Produkty amplifikacji zestawu KinFinder zostały poddane elektroforezie kapilarnej w sekwenatorze ABI 3130 XL w polimerze POP-7™ (Thermo Fisher Scientific). Uzyskane elektroforegramy analizowano przy użyciu programu Microsatellite Analysis Software (Thermo Fisher Scientific) dostępnego w chmurze Thermo Fisher Cloud – według ustawień standardowych.

Biblioteki iiSNP utworzone z zastosowaniem Precision ID Identity Panel zostały zsekwencjonowane w systemie Ion GeneStudio™ S5 Prime, a następnie analizowane przy użyciu programu Torrent Suite™ Software v.5.18.1, przy czym analiza wtórna przeprowadzana była z użyciem wtyczki HID Genotyper v.2.2., według ustawień standardowych dla markerów iiSNP.

3.11. Ustalanie haplogrup chromosomu Y

Profile Y-STR zostały przeanalizowane przy użyciu narzędzia statystycznego NEVGEN Genealogy Tools v1.1.. Narzędzie to wykorzystuje bayesowski model częstości alleli, uwzględniając wzajemną zależność wartości markerów Y-STR [88]. Dla celów porównawczych rozkłady haplogrup w populacji gruzińskiej zaczerpnięto z opublikowanych zestawów danych: dane z publikacji Schurra i wsp. [130] zostały użyte bezpośrednio, natomiast zestaw opublikowany przez Nasidze i wsp. [131] został najpierw wykorzystany do predykcji haplogrup przy użyciu NEVGEN, a następnie poddany analizie.

Przy obliczeniach zastosowano ustawienia „Batch calculate”: FTDNA order, wprowadzono dane odnośnie markerów dostępnych w danym ustawieniu: DYS19, DYS385ab, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS437, DYS438, DYS439, DYS448, DYS460, DYS481, DYS533, DYS570, DYS576, DYS635, DYS449, DYS456 i DYS458 oraz GATA H4.

Dla markerów GATA H4 wartości alleli zostały obniżone o 1 (np. jeżeli allel miał wartość 11, wprowadzano 10), z uwagi na niezgodność między sposobem wywoływania alleli w systemach FTDNA i Yfiler™ Plus.

Predykcje uzyskane przy użyciu tego narzędzia dla wymienionych markerów opisywane są zazwyczaj przez dwa parametry:

- Probability (prawdopodobieństwo) – początkowe 100% prawdopodobieństwa przynależności osoby do określonej haplogrupy jest dzielone pomiędzy wszystkie możliwe haplogrupy,
- Fitness score (współczynnik dopasowania) – parametr oceniający poprawność danej predykcji.

NEVGEN sugeruje, że wartość Fitness należy interpretować różnie w zależności od haplogrupy – wartość 25 może być wystarczająca dla „starszych” podgrup, podczas gdy nawet 60 może być niewystarczające dla niektórych „młodszych” podgrup [132].

W niniejszym badaniu do analizy przyjęto jedynie wyniki, dla których Probability przekraczało 50%.

3.12. Wykorzystanie programu Familias

Do obliczeń statystycznych pokrewieństwa pomiędzy analizowanymi osobami wykorzystano program komputerowy Familias version 3.4 (Norwegia). Program, rozwijany przez grupę badawczą z Norwegian University of Life Sciences oraz Norweskiego Instytutu Kryminalistyki, jest szeroko wykorzystywany w genetyce sądowej zarówno w analizach

rutynowych, jak i w złożonych przypadkach pokrewieństwa obejmujących relacje dalszego stopnia [122]. Bardzo istotny jest fakt, że program został zwalidowany co potwierdza prawidłowość uzyskiwanych wyników [133].

Program Familias jest narzędziem wykorzystywanym do szacowania ilorazu wiarygodności oraz wartości prawdopodobieństwa pokrewieństwa. Podczas analizy statystycznej, konfrontowane są ze sobą dwie przeciwstawne hipotezy odnoszące się do możliwości pokrewieństwa pomiędzy analizowanymi osobami. Hipoteza H1, zakładająca określony stopień pokrewieństwa między badanymi osobami, oraz hipoteza H2, zgodnie z którą osoby te nie są spokrewnione. Równanie wartości LR zakłada porównanie prawdopodobieństw ustalenia dowodu z badań genetycznych przy założeniu dwóch alternatywnych hipotez (iloraz dwóch prawdopodobieństw warunkowych).

$$LR = \frac{\text{Pr (dowód|H}_1\text{)}}{\text{Pr (dowód|H}_2\text{)}}$$

gdzie:

LR - likelihood ratio- iloraz wiarygodności

Pr – prawdopodobieństwo

dowód – zestawy alleli (profile DNA analizowanych osób)

H₁: hipoteza 1- określająca szukane pokrewieństwo

H₂: hipoteza 2- określająca brak zakładanego pokrewieństwa

Następnie otrzymaną wartość przekształca się na prawdopodobieństwo pokrewieństwa korzystając ze wzoru:

$$P = \frac{LR}{LR+1}$$

gdzie:

LR - likelihood ratio- iloraz wiarygodności

P – prawdopodobieństwo pokrewieństwa

Funkcje programu:

Jedną z głównych cech programu jest zdolność do wykonywania obliczeń dla dowolnych struktur rodowodowych, obejmujących relacje pierwszego (np. ojciec–dziecko), drugiego (np. pełne rodzeństwo, wuj–bratanek) i dalszego stopnia (np. kuzynostwo). Ponadto oprogramowanie umożliwia wykonywanie obliczeń z uwzględnieniem ewentualnych mutacji, alleli zerowych czy wystąpienia zjawiska drop out/drop in, co jest szczególnie istotne w analizach rodzin wielopokoleniowych, przypadkach katastrof masowych oraz w sytuacjach, w których dochodzi do identyfikacji szczątków ludzkich o zdegradowanym materiale

genetycznym. Uwzględnia także strukturę populacyjną (populational substructure) za pomocą współczynnika korekcyjnego θ [134].

Oprócz wykonania obliczeń statystycznych pokrewieństwa, program umożliwia także wykonanie symulacji przedziałów wartości LR dla zadanych relacji rodzinnych, dzięki czemu można zweryfikować czy wynik uzyskany z obliczeń mieści się w przedziale oczekiwanych rezultatów. Funkcja ta umożliwia również sprawdzenie, bez wykonania badań genetycznych, jak kolejne osoby z drzewa genealogicznego wpłyną na uzyskany wynik analizy pokrewieństwa. Dzięki czemu można ograniczyć badania do osób, które mają realny wpływ na identyfikację.

W programie dostępne są także moduły dedykowane analizom w przypadkach ofiar katastrof masowych umożliwiające dokonanie przeszukań większej liczby ofiar przy wykorzystaniu zadanych relacji pokrewieństwa oraz moduł Blind Search pozwalający na efektywne analizowanie przypadków o nieznanym stopniu pokrewieństwa.

Modele mutacyjne:

W przypadku stwierdzenia niezgodności alleli pomiędzy analizowanymi osobami zakładano możliwość wystąpienia mutacji. Program daje możliwość wybrania spośród pięciu różnych modeli mutacyjnych:

Model 1: Equal probability (Simple): jest najprostszym modelem mutacji dostępnym w programie. Zakłada, że wystąpienie mutacji jest niezależne od częstości danego allela oraz od jego wariantu, np. mutacja wariantu 13 na 14 będzie tak samo prawdopodobna jak mutacja wariantu 13 na 17.

Model 2: Probability proportional to frequency (Stationary): prawdopodobieństwo mutacji z jednego wariantu allela na drugi jest wprost proporcjonalne do jego częstości występowania oraz jest niezależna od jego wariantu, np. częstość wystąpienia w populacji allela 14 wynosi 0,01, a allela 15 wynosi 0,5, dlatego model ten zakłada, iż mutacja z allela 13 na allel 15 jest bardziej prawdopodobna niż z wariantu 13 na 14.

Model 3 i 4: Stepwise (Unstationary) (Stationary): model ten bierze pod uwagę zmianę liczby sekwencji repetytywnej. Mutacja z wariantu 13 na warianty 12 lub 14 jest bardziej prawdopodobna niż mutacja wariantu 13 na warianty 11 lub 15.

Model 5: Extended Stepwise: rozszerzona wersja modelu stepwise, dodatkowo uwzględniająca różnice między mutacjami mikrowariantów.

Modele stacjonarne to takie, w których proces mutacji nie zmienia rozkładu częstości alleli w kolejnych pokoleniach – populacja pozostaje w stanie równowagi względem macierzy mutacji. Zapewniają one stabilność obliczeń i spójność rozkładów alleli w analizach

wielopokoleniowych, podczas gdy modele niestacjonarne mogą prowadzić do niewielkich odchyłen w rozkładzie częstości alleli w kolejnych generacjach [135].

Podczas obliczeń stosowano standardowe wartości parametrów dla alleli zerowych (null=0) z uwagi na ich znikomy wpływ na wyniki [136].

Program korzysta z plików wsadowych w formacie tekstowym rozdzielonym znakami tabulacji, które można wyeksportować bezpośrednio z zastosowanego programu do analizy profili genetycznych. Bazy częstości również są publikowane w formie którą można zaimportować do programu niemalże bez modyfikacji.

Pomimo szerokiej możliwości analitycznych, program Familias posiada również pewne ograniczenia:

- brak wbudowanego modułu do analizy markerów sprzężonych. W przypadku konieczności uwzględnienia sprzężeń genetycznych należy korzystać z zewnętrznego oprogramowania, dedykowany dla Familiasa jest program FamLink, a uzyskane współczynniki należy następnie samodzielnie przemnożyć przez wartości LR wyliczone w programie Familias. Procedura ta, wykonywana ręcznie, zwiększa ryzyko popełnienia błędów obliczeniowych.
- program nie umożliwia bezpośredniego tworzenia graficznych reprezentacji hipotez. Zamiast tego generuje skrypt tekstowy zawierający odpowiednie polecenia, które można zaimportować do środowiska R, gdzie możliwe jest uzyskanie wizualizacji analizowanych relacji rodzinnych.

3.13. Wykorzystanie programu FamLink

FamLink jest narzędziem biostatystycznym wykorzystywanym w analizach pokrewieństwa, umożliwiającym uwzględnienie sprzężenia genetycznego markerów w obliczeniach prawdopodobieństwa relacji rodzinnych. Zastosowanie programu w badaniach identyfikacyjnych pozwala na precyzyjniejsze oszacowanie prawdopodobieństwa pokrewieństwa pomiędzy analizowanymi osobami poprzez uwzględnienie współdziedziczenia alleli w obrębie chromosomów [137]. Oprogramowanie to zostało opracowane w celu rozszerzenia standardowych metod stosowanych w analizie relacji rodzinnych, w których domyślnie zakłada się niezależne dziedziczenie markerów. Uwzględnienie tego efektu pozwala na uniknięcie błędnych oszacowań wartości ilorazu wiarygodności i prawdopodobieństwa pokrewieństwa, które mogą pojawić się, gdy ignorowane jest wspólne dziedziczenie markerów znajdujących się blisko siebie na tym samym chromosomie, co może prowadzić do zaniżenia lub zawyżenia wartości LR oraz błędnej interpretacji siły dowodu [138].

Działanie programu opiera się na algorytmie Lander–Green, powszechnie stosowanym w analizach markerów powiązanych, który umożliwia modelowanie prawdopodobieństwa

rekombinacji pomiędzy loci w obrębie analizowanego rodowodu. Algorytm ten pozwala na precyzyjne uwzględnienie zjawiska rekombinacji i dokładne oszacowanie wartości LR dla różnych scenariuszy pokrewieństwa. W klasycznych analizach genetycznych zakłada się niezależne dziedziczenie loci markerów STR, co uzasadnia stosowanie reguły mnożenia przy obliczaniu łącznego LR. Założenie to może jednak zostać naruszone w przypadku loci zlokalizowanych blisko siebie na tym samym chromosomie, gdzie występuje sprzężenie genetyczne, rozumiane jako ograniczona częstość rekombinacji między markerami [139].

Odległość genetyczna między dwoma loci wyrażana w cM, odpowiada prawdopodobieństwu 1% wystąpienia rekombinacji podczas mejozy. Loci położone blisko siebie mogą wykazywać niską częstość rekombinacji, co skutkuje współdziedziczeniem określonych układów alleli (haplotypów) częściej, niż wynikałoby to z założenia niezależności. Zjawisko to ma istotne znaczenie w analizach pokrewieństwa, szczególnie przy zastosowaniu rozbudowanych zestawów STR, w których obecność sprzężonych loci została udokumentowana (np. D12S391 i vWA na chromosomie 12). Nieuwzględnienie tego efektu może prowadzić do przeszacowania wartości LR i zawyżenia siły dowodu [140].

W niniejszej pracy zastosowano oprogramowanie FamLink2, będące rozwinięciem pierwotnej wersji programu. W odróżnieniu od FamLink, nowa wersja umożliwia jednoczesne uwzględnienie kilku par sprzężonych loci. W FamLink 2 należy wprowadzić pozycje markerów w cM i przypisać je do odpowiednich chromosomów, a program sam wylicza wartość rekombinacji między markerami [141]. W przypadkach, w których odległości w cM nie były dostępne, zastosowano standardowy przelicznik między fizyczną a genetyczną odległością $1 \text{ Mb} \approx 1 \text{ cM}$ [142].

Wyniki obliczeń w FamLink2 prezentowane są w postaci dwóch wartości ilorazu wiarygodności: LR exact (z uwzględnieniem sprzężeń) oraz LR unlinked (dla markerów niezależnych). W celu korekty wyników uzyskanych w programie Familias można wykorzystać tzw. multiplication factor, obliczany jako iloraz LR exact i LR unlinked. Wartość ta odzwierciedla kierunek i skalę wpływu sprzężeń na wynik:

- $\text{MF} > 1$ oznacza, że wynik bez sprzężeń jest niedoszacowany,
- $\text{MF} < 1$ wskazuje na przeszacowanie LR przy założeniu niezależności markerów.

Uzyskany współczynnik można następnie zastosować do korekty wyników z programu Familias poprzez pomnożenie otrzymanej wartości LR przez MF.

Do ograniczeń programu należy zaliczyć stosunkowo niewielki zakres zaimplementowanych modeli pokrewieństwa – w przypadku bardziej złożonych relacji użytkownik musi samodzielnie zdefiniować hipotezy, co nie jest objęte walidacją autorów. W konsekwencji dla

dalszych stopni pokrewieństwa, większej liczby osób lub bardzo rozbudowanych zestawów markerów, program może nie generować wyników.

Wpływ sprzężeń na uzyskiwane wartości LR zależy od kilku czynników, w tym od liczby i rozmieszczenia markerów w multipleksie oraz od analizowanego stopnia pokrewieństwa [139, 143].

Buckleton i Triggs wskazują, że istotny wpływ sprzężeń obserwuje się zwykle przy odległościach poniżej 5 cM, choć w niektórych przypadkach zauważalne efekty mogą występować także przy dystansach do 50 cM [144]. W niniejszej pracy przyjęto wartość 100 cM jako graniczną dla potencjalnego wystąpienia efektów sprzężenia.

Z kolei Egeland i Sheehan zwracają uwagę, że mimo iż sprzężenia mają ograniczone znaczenie w standardowych analizach identyfikacyjnych, mogą być niezbędne przy rozróżnianiu pewnych typów pokrewieństw, np. w przypadkach, gdy dostępne są jedynie profile braci przyrodnych analiza wyłącznie niezależnych markerów nie pozwala odróżnić relacji typu rodzeństwo przyrodnie od relacji typu wuj-siostrzeniec lub dziadek-wnuk [145].

3.14. Wykorzystanie programu KinBN

KinBN jest oprogramowaniem opracowanym przez Morimoto i wsp. [146] w Zakładzie Medycyny Sądowej Uniwersytetu w Kioto, przeznaczonym do analizy pokrewieństwa z uwzględnieniem sprzężeń genetycznych i mutacji w oparciu o sieci bayesowskie. Program posiada graficzny interfejs użytkownika i został napisany w języku R. Umożliwia obliczanie ilorazów wiarygodności dla relacji rodzinnych z jednoczesnym uwzględnieniem sprzężeń pomiędzy loci, mutacji oraz efektu drop-out. W najnowszej wersji wprowadzono dodatkowo możliwość pracy w środowisku R oraz funkcję zapisu i odczytu projektów, co ułatwia prowadzenie analiz w sposób powtarzalny i uporządkowany.

Podstawą działania programu jest model sieci bayesowskiej, w którym allele i genotypy reprezentowane są jako węzły, a zależności między nimi jako krawędzie. Program tworzy probabilistyczny model dziedziczenia alleli pomiędzy osobami w analizowanym rodowodzie, uwzględniając zarówno zmienność mutacyjną, jak i możliwość rekombinacji pomiędzy markerami zlokalizowanymi blisko siebie. Dzięki zastosowaniu uogólnionego modelu opartego na twierdzeniu Bayesa, KinBN umożliwia porównanie dowolnych hipotez pokrewieństwa (nie tylko pokrewieństwo vs. brak pokrewieństwa, ale także np. pełne rodzeństwo vs. rodzeństwo przyrodnie) oraz symulację rozkładów wartości LR dla określonych relacji rodzinnych.

Użytkownik przygotowuje trzy pliki wejściowe w formacie CSV: profile osób analizowanych, liczby obserwowanych alleli w populacji (na podstawie których program estymuje częstości alleli z wykorzystaniem rozkładu Dirichleta), częstości mutacji (zastosowany jest podobny model mutacji do Stepwise z Familias – należy podać prawdopodobieństwo wystąpienia przesunięcia o konkretną liczbę powtórzeń dla kobiet i mężczyzn). Hipotezy H1 i H2 tworzy się poprzez ręczne zdefiniowanie struktury pokrewieństwa w postaci rodowodu. Wyniki obliczeń prezentowane są w formie tabelarycznej oraz graficznej, z możliwością eksportu raportu w formacie CSV. Program może również przeprowadzać symulacje rozkładu LR dla założonych relacji [147].

Wartość recombination rate dla analizowanych zestawów została obliczona na podstawie wzoru przedstawionego w pracy Chatre [148].

Zalety programu:

- intuicyjna obsługa i przejrzysty interfejs graficzny ułatwiający rozpoczęcie pracy,
- możliwość uwzględnienia sprzężeń genetycznych, mutacji oraz allele drop-out,
- obliczenia dla dowolnych hipotez pokrewieństwa,
- możliwość przeprowadzenia symulacji oraz wizualizacji wyników w formie graficznej.

Ograniczenia programu:

- baza częstości alleli musi być przygotowana w formie liczby wystąpień alleli (nie częstości), co wymaga dodatkowego przetworzenia danych, gdyż ogólnodostępne bazy są zazwyczaj w formie częstości wystąpienia,
- brak możliwości importu profilu zawierającego allel nieobecny w bazie — wymaga to ręcznej edycji pliku źródłowego,
- brak podglądu wczytanych profili i bazy alleli w interfejsie programu,
- brak możliwości zmiany nazw osób – program narzuca nazwy dodawanych osób bez profili jako UK, co dla początkujących użytkowników może stanowić utrudnienie,
- konieczność wprowadzenia wartości recombination rate, jeśli nie są dostępne w literaturze należy je policzyć samodzielnie,
- ograniczenie do analizy maksymalnie dwóch hipotez w jednym projekcie,
- W sytuacji gdy wykonujemy analizy dla większej liczby osób program wykonuje obliczenia tylko dla tych markerów dla których posiada wszystkie dane. W sytuacji gdy jedna osoba nie posiada wartości układ w całej hipotezie jest pomijany. Z kolei w programie Familias wykonywane są obliczenia dla tego układu z pominięciem osoby dla której nie mamy pełnych danych,

- brak możliwości dodania kolejnych osób w trakcie analizy – wszystkie profile muszą znajdować się w jednym pliku wejściowym.

KinBN stanowi istotne uzupełnienie dostępnych narzędzi biostatystycznych, umożliwiając równoczesne uwzględnienie mutacji, sprzężeń i efektu drop-out w analizie pokrewieństwa.

3.15. Wykorzystanie programu pedbuildr

Zagadnienie rekonstrukcji drzew genealogicznych odgrywa istotną rolę w analizach kryminalistycznych, zwłaszcza w przypadkach, gdy na podstawie profili genetycznych trzeba ustalić czy osoby były ze sobą spokrewnione. Dzięki wykorzystaniu dedykowanych narzędzi bioinformatycznych możliwe jest odtwarzanie struktur rodzinnych sięgających nawet pięciu pokoleń wstecz [149].

Standardowe analizy pokrewieństwa nie uwzględniają pewnego stopnia spokrewnienia ze sobą członków populacji, również na poziomie jej założycieli (tj. chowu wsobnego), co sprawia, że programy oparte na współdzieleniu alleli identycznych z pochodzenia – w tym pedbuildr – nie odnoszą się do tego zagadnienia [150].

Dopasowania uzyskane w programie Familias zostały poddane weryfikacji za pomocą symulacji alternatywnych struktur pokrewieństwa. Symulacje przeprowadzono z użyciem pakietu pedbuildr – narzędzia w języku R opracowanego przez Vigelanda, służącego do rekonstrukcji struktur rodowodowych na podstawie danych o relacjach pokrewieństwa, profili genetycznych oraz parametrów populacyjnych (<https://github.com/magnusdv/pedbuildr>).

Podczas interpretacji wyników konieczna jest znajomość zasad dziedziczenia markerów genetycznych oraz rozróżnienie alleli identity by state (IBS) i identity by descent (IBD). W przypadku osób niespokrewnionych posiadanie tych samych alleli oznacza IBS, natomiast IBD odnosi się do sytuacji, w której allele u dwóch lub więcej osobników są identyczne, ponieważ zostały odziedziczone po wspólnym przodku bez mutacji w kolejnych pokoleniach. Na podstawie danych wejściowych wyznaczane są prawdopodobieństwa współdzielenia 0, 1 lub 2 alleli przez parę diploidalnych osobników. Prawdopodobieństwa te oznaczają się odpowiednio jako:

- k_0 – brak wspólnych alleli IBD,
- k_1 – jeden allel IBD,
- k_2 – dwa allele IBD

przy czym zachodzi zależność $k_0 + k_1 + k_2 = 1$.

Uzyskane wyniki przedstawiane są graficznie w postaci tzw. trójkąta IBD (IBD triangle) – wykresu, na którym pozycja punktu odpowiada konkretnej konfiguracji IBD dla analizowanej

pary. Na osi x odwzorowana jest wartość k_0 , na osi y – wartość k_2 . Wierzchołki trójkąta reprezentują trzy skrajne typy relacji:

- (0,0): pokrewieństwo rodzic–dziecko ($k_1 = 1$),
- (1,0): brak pokrewieństwa ($k_0 = 1$),
- (0,1): relacja identycznych bliźniąt ($k_2 = 1$).

Położenie punktu w obrębie trójkąta umożliwia wizualną ocenę, które scenariusze pokrewieństwa są najbardziej zgodne z uzyskanymi wartościami IBD. Przy analizie pełnego rodzeństwa istnieje 25% prawdopodobieństwa, że jedno pełne rodzeństwo odziedziczy oba allele drugiego rodzeństwa od wspólnych rodziców, 50% prawdopodobieństwo, że rodzeństwo odziedziczy IBD jeden allel rodzicielski drugiego, oraz 25% prawdopodobieństwo, że rodzeństwo nie odziedziczy żadnych alleli rodzicielskich drugiego rodzeństwa. Przykładowe stopnie pokrewieństwa wraz z odpowiadającymi szansami odziedziczenia alleli IBD przedstawiono w tabeli poniżej (Tab. 1).

Tab. 1. Zależność szansy odziedziczenia alleli IBD od stopnia pokrewieństwa

Relacja	stopień pokrewieństwa	k_0	k_1	k_2
bliźniak jednojajowy	-	0	0	1
rodzic - dziecko	1	0	1	0
rodzeństwo pełne	1	1/4	1/2	1/4
rodzeństwo przyrodnie	2	1/2	1/2	0
wuj - bratanek	2	1/2	1/2	0
dziadek - wnuk	2	1/2	1/2	0
kuzyn	3	3/4	1/4	0
niespokrewniony	-	1	0	0

źródło: [116, str. 380]

Program pedbuildr często wskazuje hipotezy rodzeństwa przyrodniego i dziadek – wnuk jako równie prawdopodobne, co jest zgodne z teoretycznymi wartościami IBD. W celu rozróżnienia krewnych drugiego i dalszego stopnia pokrewieństwa konieczne jest zwiększenie liczby analizowanych loci lub uwzględnienie dodatkowych osób w analizie [151].

Pakiet umożliwia również symulację wszystkich możliwych struktur pokrewieństwa pomiędzy analizowanymi osobami oraz ich ocenę według wartości prawdopodobieństwa dopasowania. Dzięki temu pedbuildr stanowi cenne narzędzie wspomagające ocenę wiarygodności przyjętych hipotez pokrewieństwa i umożliwia ich ewentualną korektę.

Jako ograniczenia programu wskazano:

- analiza możliwa jedynie dla markerów posiadających pełne dane dla wszystkich osób,
- brak graficznego interfejsu użytkownika – obsługa odbywa się poprzez polecenia w konsoli R.

4. Wyniki

4.1. Rodzina 1

W analizie pierwszej rodziny wzięto pod uwagę czterech wnuków zaginionego (osoby 219, 221, 222 i 267) oraz jedną wnuczkę (osoba 220), relacje przedstawiono poniżej (Rys. 3). W przeszukaniu DVI wskazano dopasowanie do szczątków oznaczonych numerem 94. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 18\,729$, która po uwzględnieniu sprzężeń genetycznych obniżyła się do $13\,150$. Równoległa analiza w programie KinBN dała wartości odpowiednio $22\,600$ (bez sprzężeń) i $16\,000$ (ze sprzężeniami).

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestaw KinFinder, a obliczona wartość LR w programie Familias wynosi 10 (bez sprzężeń) oraz 24 dla sprzężonych układów, natomiast w KinBN $LR = 11$ (bez sprzężeń) oraz $LR = 29$ (ze sprzężeniami).

Dodatkowo, analiza haplotypów chromosomu Y wykazała pełną zgodność i wartość $LR = 190\,117$. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 3,56 \cdot 10^{10}$. Symulacje rozkładu LR (Tab. 2) wskazują, że uzyskane wyniki dla zestawu GlobalFiler™ mieszczą się w oczekiwanych przedziałach, choć znajdują się nieco poniżej mediany. Wynik uzyskany dla zestawu KinFinder znajduje się poniżej wskazanych w symulacjach.

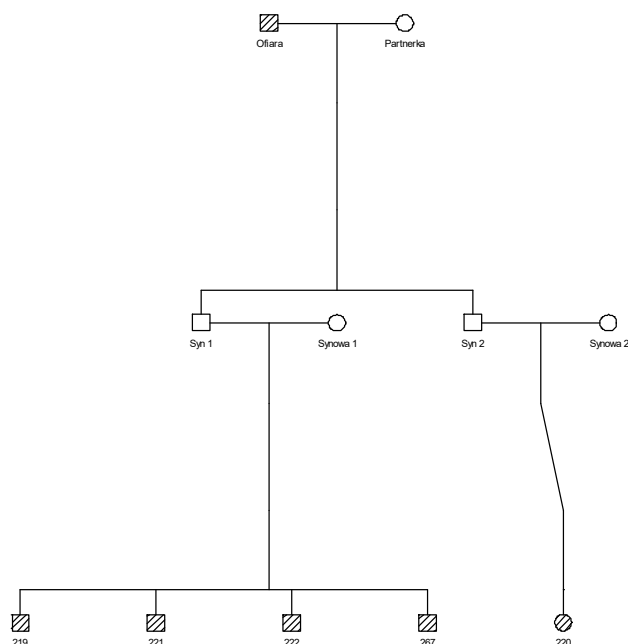
Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (uwzględniając niepełne oznaczenie profilu – 15 układów z 50 uzyskał porównywalną moc z zestawami GlobalFiler™ i PowerPlex® Fusion 6C), a najniższą Precision ID Identity.

Analizy IBD przeprowadzone w pakiecie pedbuildr potwierdziły zgodność hipotezy o relacji wnuk–dziadek pomiędzy osobami 219, 221, 222, 267 a szczątkami 94. W przypadku osoby 220 uzyskano wyniki wskazujące na możliwość alternatywnej relacji – siostry przyrodniej pozostałych wnuków – wspólna matka różni ojcowie (Rys. 4, Rys. 5 i Rys. 6).

Obliczenia statystyczne dla zestawu GlobalFiler™ wskazały, że hipoteza rodzeństwa przyrodniego jest 154 razy bardziej prawdopodobny niż hipoteza że osoba 220 jest kuzynką osób 219, 221, 222, 267, oraz $1,8 \cdot 10^6$ razy bardziej prawdopodobne niż hipoteza, że osoba 220 jest niespokrewniona z 219, 221, 222, 267. Wykonując obliczenia statystyczne dla hipotezy wskazanej w symulacji uzyskano wynik $LR = 1,01 \cdot 10^6$.

Obliczenia statystyczne dla zestawu Kinfinder wskazały, że hipoteza rodzeństwa przyrodniego jest $5,15 \cdot 10^6$ razy bardziej prawdopodobny niż hipoteza że osoba 220 jest kuzynką osób 219, 221, 222, 267, oraz $2,61 \cdot 10^{23}$ razy bardziej prawdopodobne niż hipoteza, że osoba 220 jest

niespokrewniona z 219, 221, 222, 267. Wykonując obliczenia statystyczne dla hipotezy wskazanej w symulacji uzyskano wynik $LR = 1,86 \cdot 10^6$.

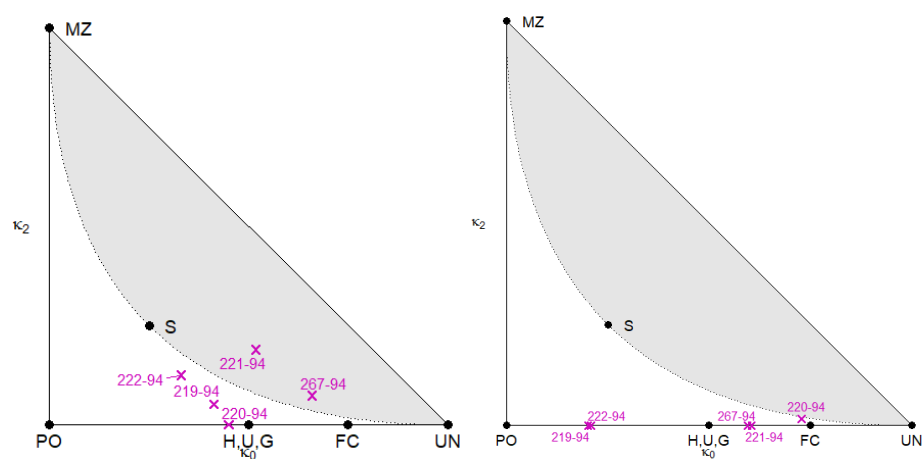


Rys. 3. Drzewo genealogiczne dla Rodziny 1

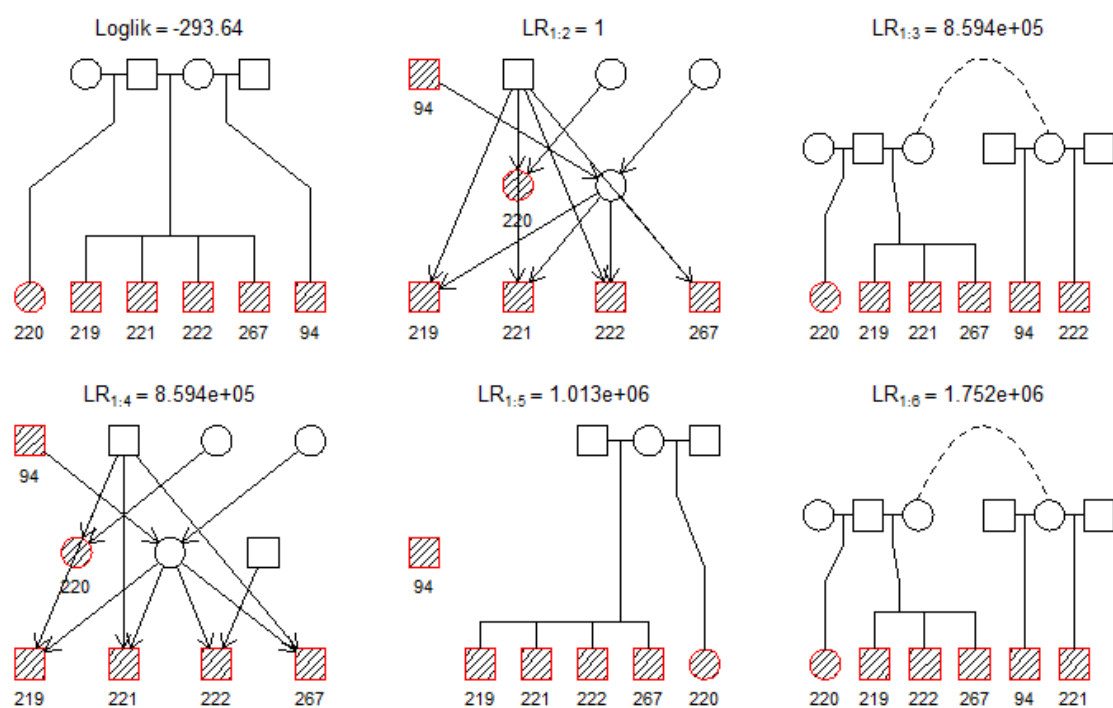
Tab. 2. Symulacja wyników analizy statystycznej dla Rodziny 1

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$3,55 \cdot 10^5$	$2,89 \cdot 10^9$	$5,88 \cdot 10^8$	406,4	$5,59 \cdot 10^{10}$
	H 1	H 2 (TRUE)	0	0,3039	$6,33 \cdot 10^{-8}$	0	8,4
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$1,45 \cdot 10^6$	$7,06 \cdot 10^{11}$	$4,31 \cdot 10^9$	1349	$2,18 \cdot 10^{13}$
	H 1	H 2 (TRUE)	0	0,0109	$1,19 \cdot 10^{-8}$	0	0,1697
KinFinder	H 1 (TRUE)	H 2	$7,97 \cdot 10^{17}$	$1,21 \cdot 10^{27}$	$1,86 \cdot 10^{23}$	$6,71 \cdot 10^{12}$	$3,64 \cdot 10^{28}$
	H 1	H 2 (TRUE)	0	0	0	0	0
KinFinder (układy obecne w profilu szczątków nr 94)	H 1 (TRUE)	H 2	$1,2 \cdot 10^6$	$1,81 \cdot 10^9$	$1,23 \cdot 10^9$	1395	$2,66 \cdot 10^{10}$
	H 1	H 2 (TRUE)	0	0,0077	0	0	0,1592
Precision ID Identity	H 1 (TRUE)	H 2	3001	$3,90 \cdot 10^5$	$6,42 \cdot 10^5$	5,759	$6,37 \cdot 10^6$
	H 1	H 2 (TRUE)	$3,39 \cdot 10^{-5}$	2,44	0,091	$1,76 \cdot 10^{-9}$	70,44

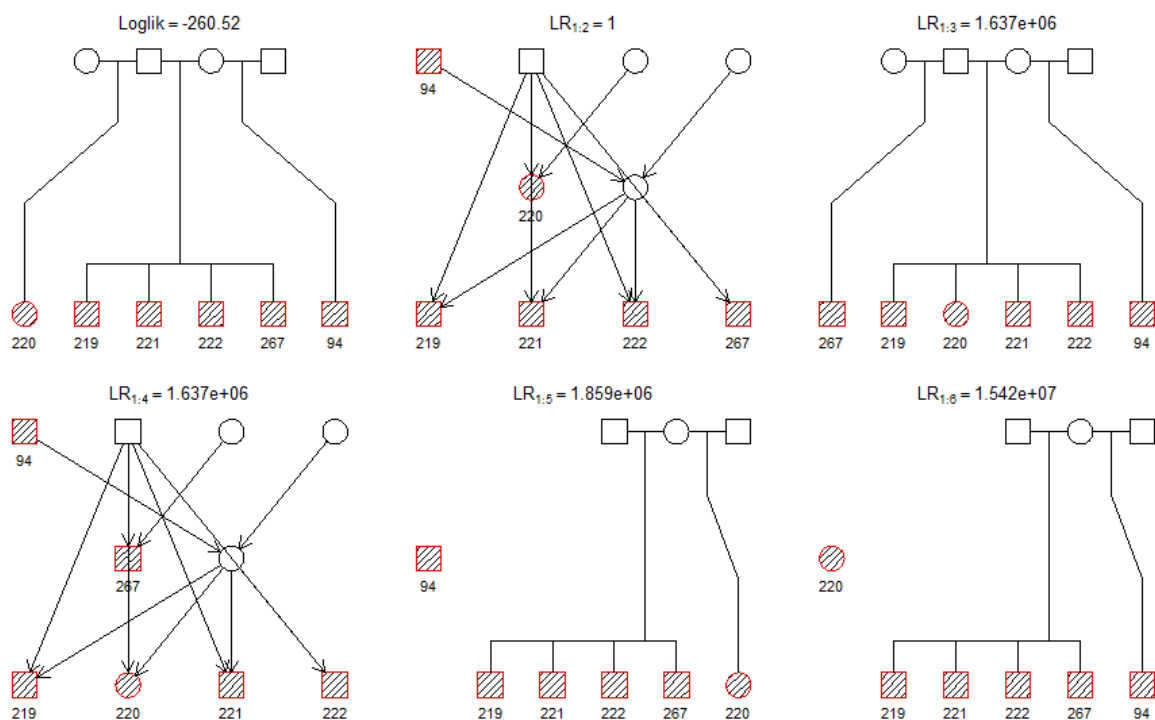
Źródło: opracowanie własne



Rys. 4. Wykres trójkąta IBD dla Rodziny 1 dla zestawu GlobalFiler™ i KinFinder



Rys. 5. Symulacje pokrewieństw dla Rodziny 1 dla zestawu GlobalFiler™



Rys. 6. Symulacje pokrewieństw dla Rodziny 1 dla zestawu KinFinder

4.2. Rodzina 2

Rodzina druga obejmuje wnuka osoby zaginionej (nr 223), relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 7).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerami 83, 84, 93, 97 i 107. Analiza haplotypów chromosomu Y wykazała pełną zgodność między profilami 97 i 223, z wartością $LR = 189\,665$. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 48$, która po uwzględnieniu sprzężeń genetycznych wzrosła do 49. Analiza w programie KinBN dała wartości odpowiednio 54 (bez sprzężeń) i 55 (ze sprzężeniami).

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestawy KinFinder oraz Precision ID Identity.

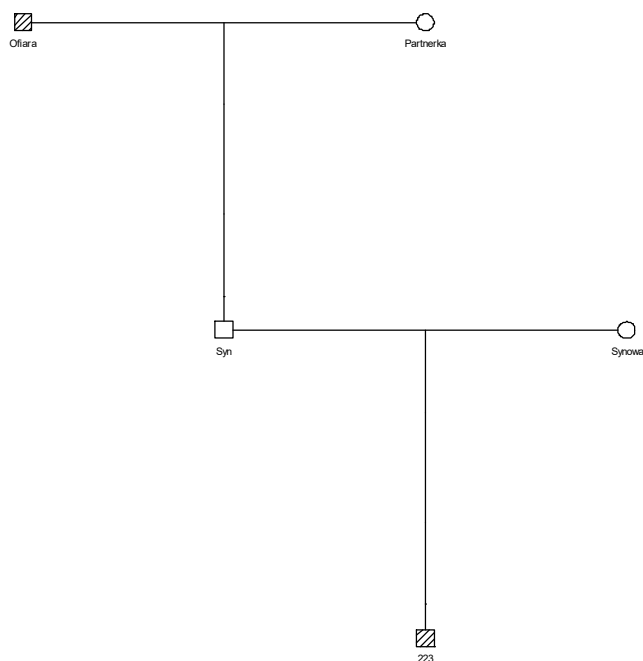
Dla zestawu KinFinder obliczenia w programie Familias $LR = 335\,399$ (bez sprzężeń; dla tego zestawu FamLink nie wykonał obliczeń), natomiast w KinBN $LR = 367\,000$ (bez sprzężeń) oraz $LR = 672\,000$ (ze sprzężeniami).

Dla zestawu Precision ID Identity obliczenia w programie Familias $LR = 32$ (bez sprzężeń) i $LR = 24$ (ze sprzężeniami), a w KinBN odpowiednio $LR = 32$ i $LR = 30$.

Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 9,77 \cdot 10^{13}$.

Symulacje rozkładu LR (Tab. 3) wskazują, że uzyskane wyniki dla wszystkich zestawów mieszczą się w oczekiwanych przedziałach. Dla zestawu GlobalFiler™ wartości znajdowały się poniżej mediany, natomiast dla zestawów KinFinder i Precision ID Identity – powyżej mediany. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (nawet uwzględniając niepełne oznaczenie profilu – 29 układów z 50), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 8) oraz symulacje możliwych pokrewieństw dla wszystkich zestawów (Rys. 9, Rys. 10 i Rys. 11) wykazują zgodność z zakładanym pokrewieństwem.

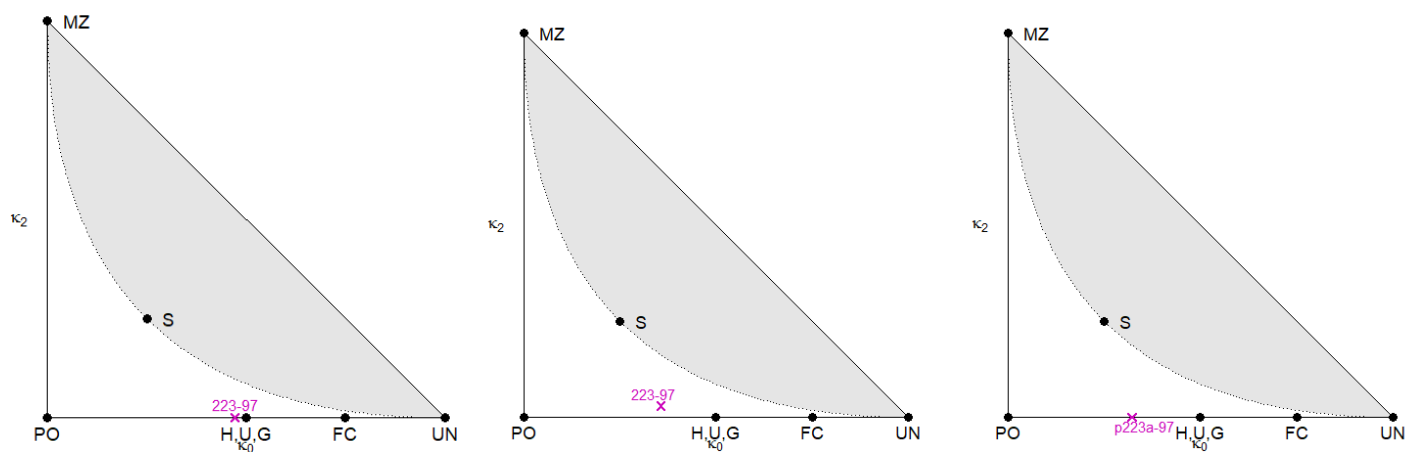


Rys. 7. Drzewo genealogiczne dla Rodziny 2

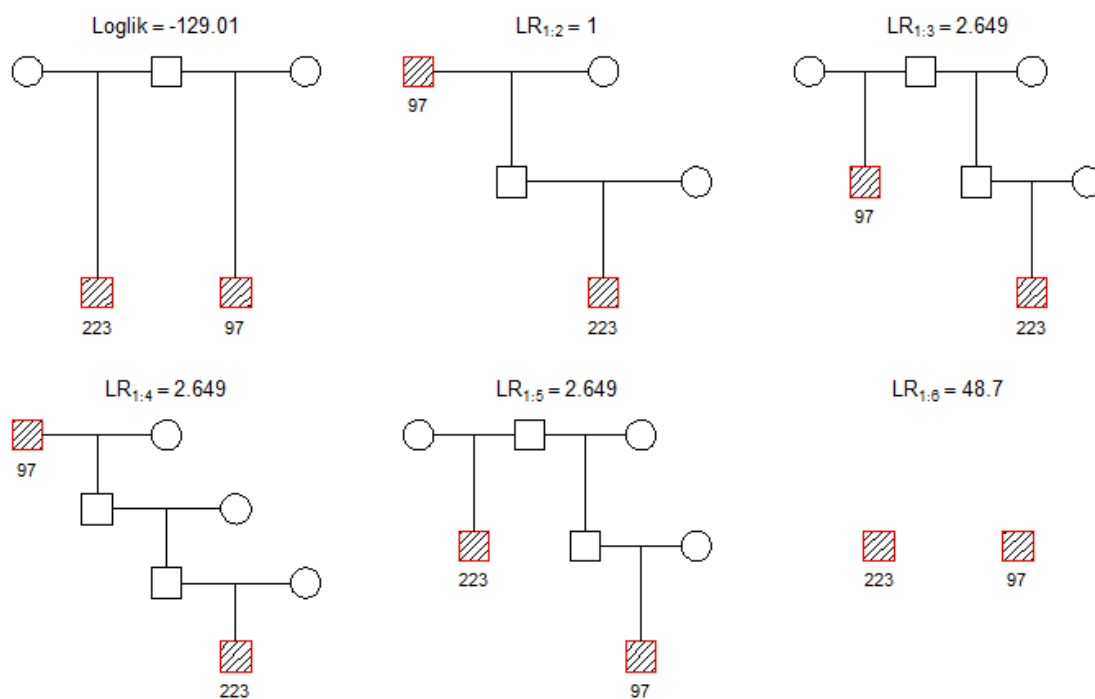
Tab. 3. Symulacja wyników analizy statystycznej dla Rodziny 2

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	60,95	$2,43 \cdot 10^4$	$1,42 \cdot 10^4$	0,3762	$2,81 \cdot 10^5$
	H 1	H 2 (TRUE)	0,0264	1,151	2,478	0,0005	9,058
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	84,17	$6,05 \cdot 10^4$	$1,84 \cdot 10^4$	0,5267	$1,34 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0175	0,8611	1,658	0,0003	10,57
KinFinder	H 1 (TRUE)	H 2	$2,31 \cdot 10^7$	$8,7 \cdot 10^{14}$	$1,01 \cdot 10^{12}$	1313	$2,73 \cdot 10^{16}$
	H 1	H 2 (TRUE)	$4,65 \cdot 10^{-7}$	0,034	0,0016	$2,11 \cdot 10^{-10}$	0,8902
KinFinder (układy obecne w profilu szczątków nr 97)	H 1 (TRUE)	H 2	$1,52 \cdot 10^5$	$7,41 \cdot 10^9$	$1,14 \cdot 10^9$	46,27	$8,54 \cdot 10^{10}$
	H 1	H 2 (TRUE)	$3,61 \cdot 10^{-5}$	0,348	0,0712	$3,9 \cdot 10^{-8}$	6,167
Precision ID Identity	H 1 (TRUE)	H 2	19,55	222,7	997,9	0,3719	801,2
	H 1	H 2 (TRUE)	0,0442	0,8663	3,097	0,0006	5,57

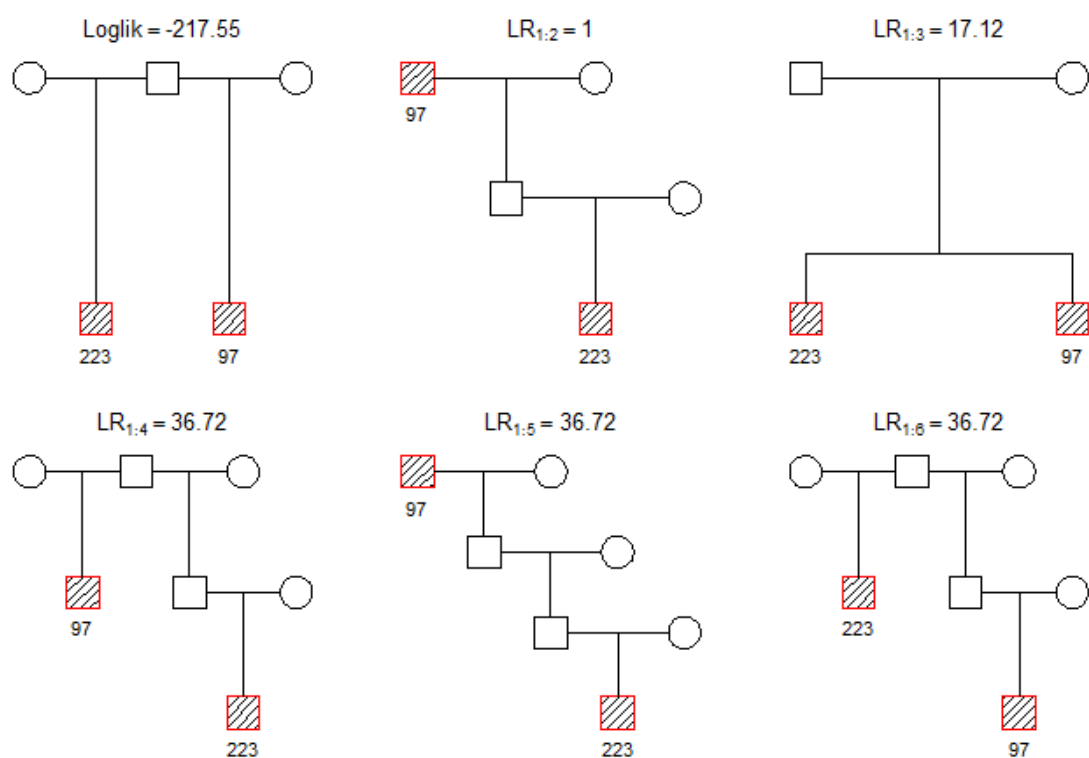
Źródło: opracowanie własne



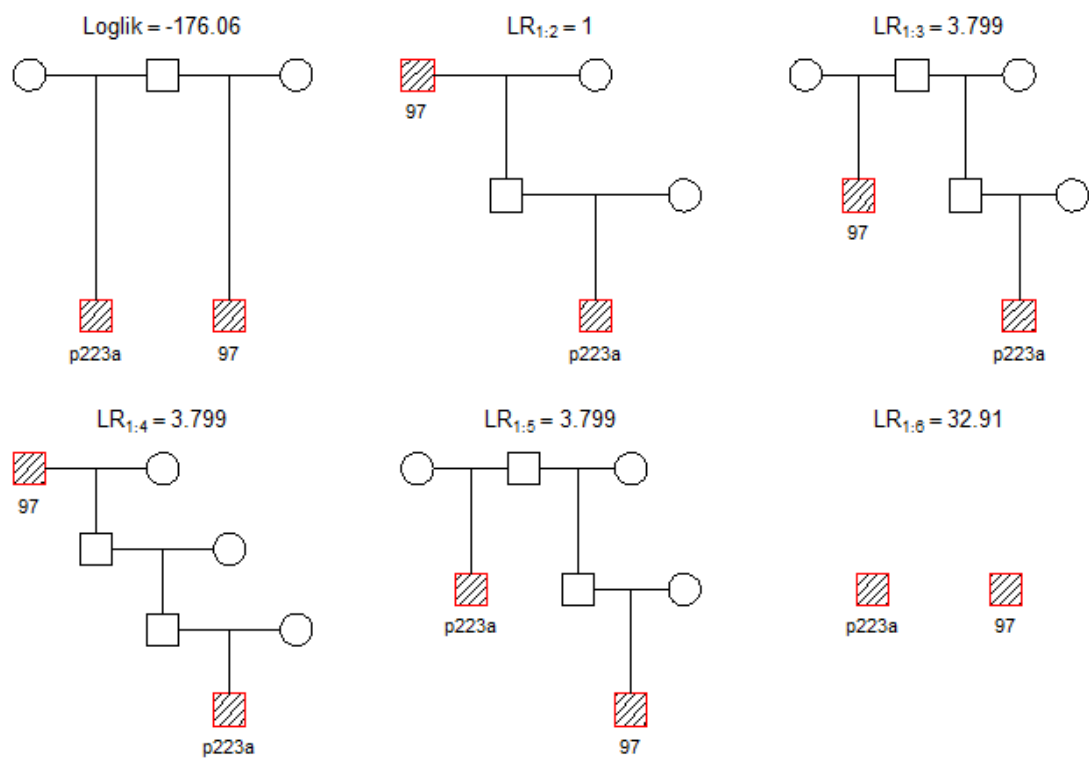
Rys. 8. Wykres trójkąta IBD dla Rodziny 2 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity



Rys. 9. Symulacje pokrewieństw dla Rodziny 2 dla zestawu GlobalFiler™



Rys. 10. Symulacje pokrewieństw dla Rodziny 2 dla zestawu KinFinder



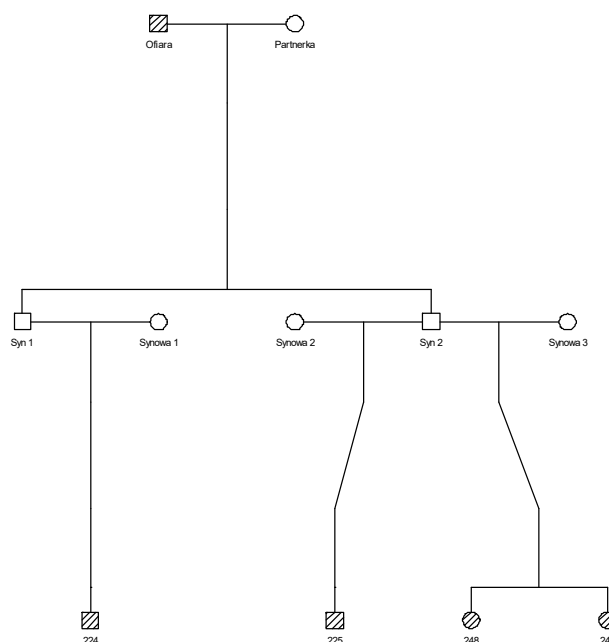
Rys. 11. Symulacje pokrewieństw dla Rodziny 2 dla zestawu Precision ID Identity

4.3. Rodzina 3

Rodzina trzecia obejmuje czworo wnuków osoby zaginionej: osobę 224, będącą synem pierwszego syna osoby zaginionej, oraz osoby 225, 248 i 249, będące dziećmi drugiego syna osoby zaginionej. Osoba 225 pochodzi z jednego związku, natomiast osoby 248 i 249 z innego związku drugiego syna, relacje przedstawiono poniżej (Rys. 12).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 106. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 4,1 \cdot 10^6$, która po uwzględnieniu sprzężeń genetycznych wzrosła do $7,28 \cdot 10^6$. Równoległa analiza w programie KinBN dała wartości odpowiednio $3,78 \cdot 10^6$ (bez sprzężeń) i $6,53 \cdot 10^6$ (ze sprzężeniami). Analiza haplotypów chromosomu Y wykazała pełną zgodność i wartość $LR = 94\ 832$. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 3,89 \cdot 10^{11}$.

Symulacje rozkładu LR (Tab. 4) wskazują, że uzyskane wyniki mieszczą się w oczekiwanych przedziałach i znajdują się nieco powyżej mediany dla zestawu GlobalFiler™. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder, a najniższą Precision ID Identity. Analiza trójkąta IBD (Rys. 13) wykazuje zgodność z zakładanym pokrewieństwem. Symulacje możliwych pokrewieństw (Rys. 14) wykazały drobne różnice jednakże analizując wszystkie możliwości zakładane na wstępie pokrewieństwo zostało poparte.

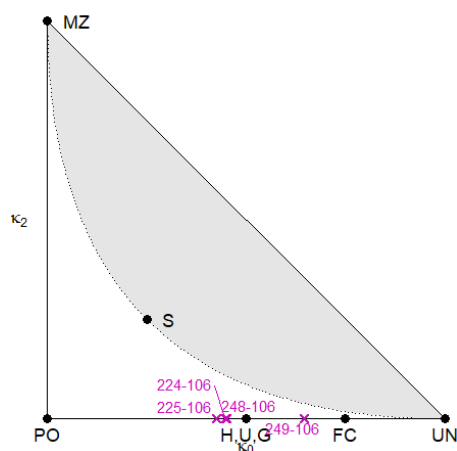


Rys. 12. Drzewo genealogiczne dla Rodziny 3

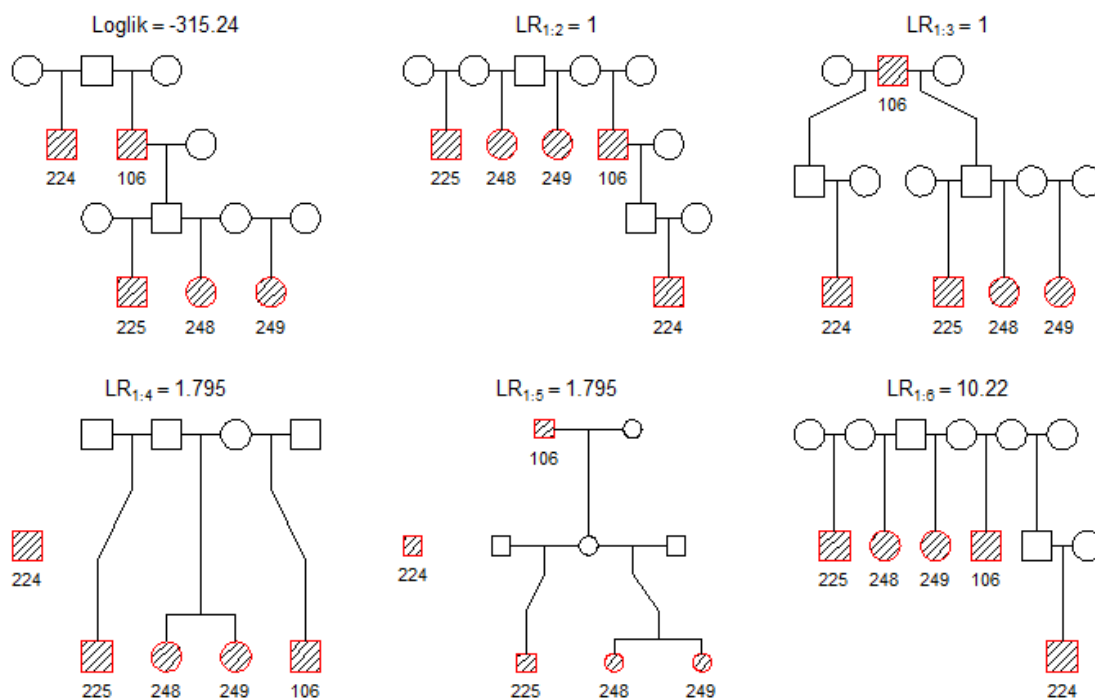
Tab. 4. Symulacja wyników analizy statystycznej dla Rodziny 3

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$5,11 \cdot 10^5$	$2,25 \cdot 10^9$	$1,29 \cdot 10^9$	394,5	$4,65 \cdot 10^{10}$
	H 1	H 2 (TRUE)	0	0,0007	$2,63 \cdot 10^{-6}$	0	0,0146
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$2,22 \cdot 10^6$	$3,12 \cdot 10^{12}$	$4,77 \cdot 10^9$	1201	$9,8 \cdot 10^{13}$
	H 1	H 2 (TRUE)	0	0,0063	0	0	0,1182
KinFinder	H 1 (TRUE)	H 2	$6,03 \cdot 10^{18}$	$4,99 \cdot 10^{25}$	$3,38 \cdot 10^{24}$	$1,96 \cdot 10^{13}$	$6,47 \cdot 10^{26}$
	H 1	H 2 (TRUE)	0	0	0	0	0
Precision ID Identity	H 1 (TRUE)	H 2	2931	$1,36 \cdot 10^6$	$1,01 \cdot 10^6$	4,194	$2,06 \cdot 10^7$
	H 1	H 2 (TRUE)	$6,49 \cdot 10^{-5}$	0,1884	0,0795	$7,1 \cdot 10^{-9}$	2,82

Źródło: opracowanie własne



Rys. 13. Wykres trójkąta IBD dla Rodziny 3



Rys. 14. Symulacje pokrewieństw dla Rodziny 3

4.4. Rodzina 4

Rodzina czwarta obejmuje dwóch wnuków zaginionego (osoby 226 i 227), relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 15).

Po przeszukaniu DVI nie stwierdzono dopasowań. Dodatkowo wykonano przeszukanie profili Y-STR w wyniku którego stwierdzono zgodność pomiędzy profilami 226, 227 i 80 a wyliczona wartość $LR = 189\,665$.

Obliczenia statystyczne wykonane w programie Familias wykazały dla w/w dopasowania wartość $LR = 0,32$ (bez sprzężeń) oraz $LR = 0,19$ (ze sprzężeniami). Analiza w programie KinBN dała wartość $LR = 0,38$ (bez sprzężeń) oraz $LR = 0,23$ (ze sprzężeniami).

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestawy KinFinder oraz Precision ID Identity.

Dla zestawu KinFinder obliczenia w programie Familias $LR = 269$ (bez sprzężeń – niepełny profil DNA, brak oznaczenia układów sprzężonych), natomiast w KinBN $LR = 103$.

Dla zestawu Precision ID Identity obliczenia w programie Familias $LR = 34\,061$ (bez sprzężeń – program FamLink nie wykonał obliczeń dla tej rodziny), a w KinBN odpowiednio $LR = 34\,100$ (bez sprzężeń) i $LR = 28\,300$ (ze sprzężeniami).

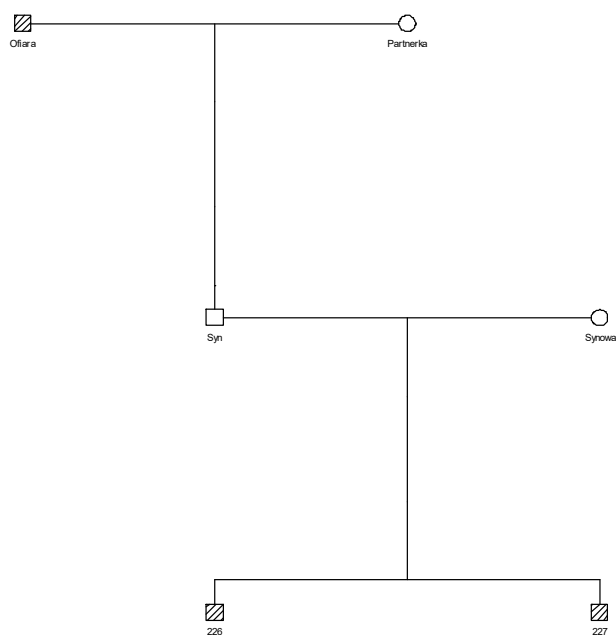
Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 5,56 \cdot 10^{11}$.

Symulacje rozkładu LR (Tab. 5) wskazują, że uzyskane wyniki dla zestawu KinFinder mieszczą się w oczekiwanych przedziałach, znajdują się jednak znacznie poniżej mediany ale powyżej progu 5%. Zestaw GlobalFiler™ bardziej wspiera hipotezę o braku pokrewieństwa i mieści się w przedziale 95% wyników dla braku pokrewieństwa, natomiast wynik uzyskany zestawem Precision ID Identity był znacznie powyżej wskazanych w symulacjach.

Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (uwzględniając niepełne oznaczenie profilu – 16 układów z 50 zestaw ten miał niższe wartości niż PowerPlex® Fusion 6C ale wyższe niż GlobalFiler™), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 16) oraz symulacje możliwych pokrewieństw dla zestawów KinFinder i Precision ID Identity (Rys. 18 i Rys. 19) wykazują zgodność z zakładanym pokrewieństwem.

Trójkąt IBD dla zestawu GlobalFiler™ wykazał brak pokrewieństwa pomiędzy profilem 80 a 226, w symulacjach pokrewieństwa (Rys. 17) jako najbardziej prawdopodobne pokrewieństwo wskazano brak pokrewieństwa, zakładany stopień pokrewieństwa został wskazany jako drugi.

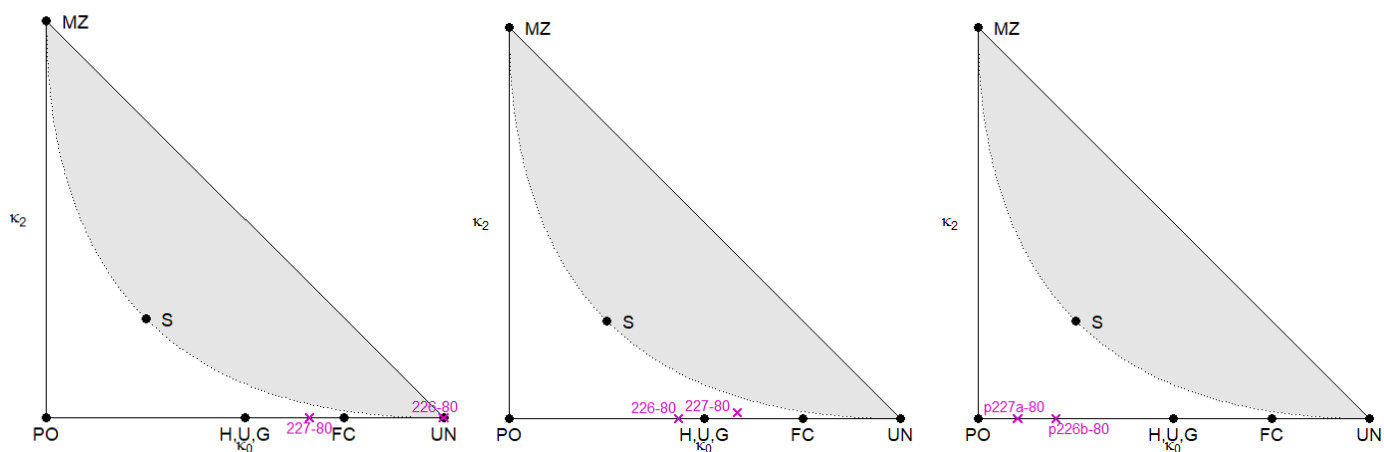


Rys. 15. Drzewo genealogiczne dla Rodziny 4

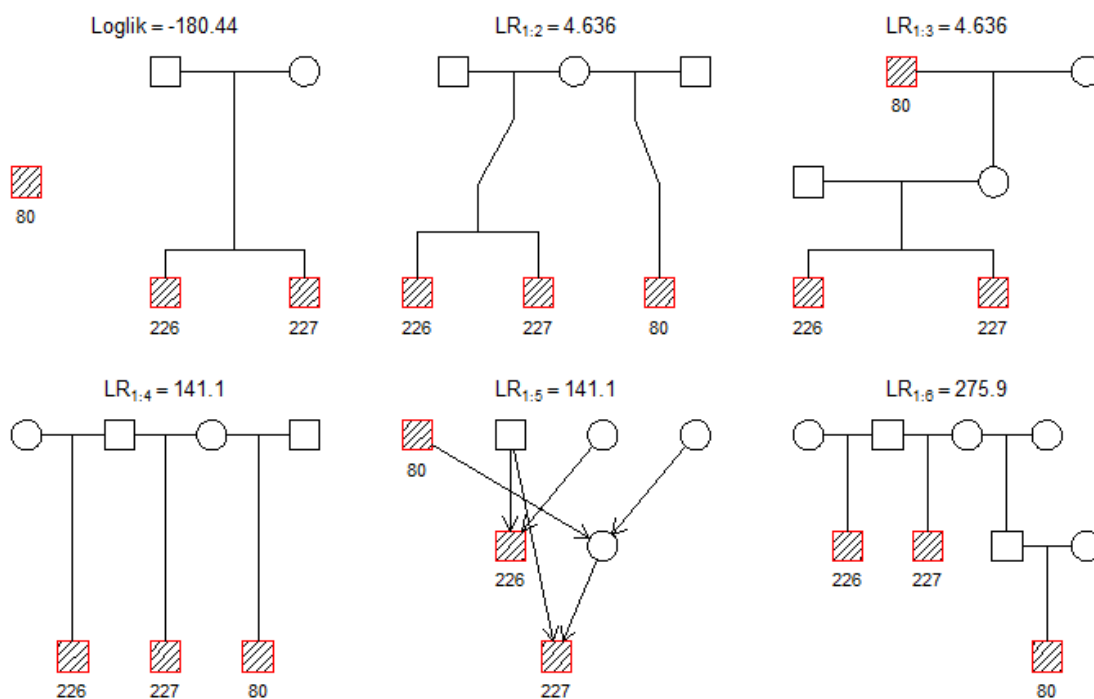
Tab. 5. Symulacja wyników analizy statystycznej dla Rodziny 4

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	848,2	$1,54 \cdot 10^5$	$1,97 \cdot 10^5$	2,866	$1,6 \cdot 10^6$
	H 1	H 2 (TRUE)	$2,64 \cdot 10^{-5}$	1,529	0,4651	0	23,98
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	1573	$9,39 \cdot 10^5$	$7,49 \cdot 10^5$	4,033	$1,36 \cdot 10^7$
	H 1	H 2 (TRUE)	$8,27 \cdot 10^{-5}$	1,663	0,2439	0	43,38
KinFinder	H 1 (TRUE)	H 2	$2,35 \cdot 10^9$	$2,93 \cdot 10^{15}$	$2,98 \cdot 10^{13}$	$1,2 \cdot 10^5$	$7,53 \cdot 10^{16}$
	H 1	H 2 (TRUE)	0	0,0048	$5,03 \cdot 10^{-10}$	0	0,1495
KinFinder (układy obecne w profilu szczątków nr 80)	H 1 (TRUE)	H 2	1561	$2,02 \cdot 10^5$	$2,93 \cdot 10^5$	4,992	$2,56 \cdot 10^6$
	H 1	H 2 (TRUE)	0	0,3063	0,3071	0	3,748
Precision ID Identity	H 1 (TRUE)	H 2	53,28	1787	3301	0,8298	$2,22 \cdot 10^4$
	H 1	H 2 (TRUE)	0,0119	1,828	2,029	$3,31 \cdot 10^{-5}$	31,5

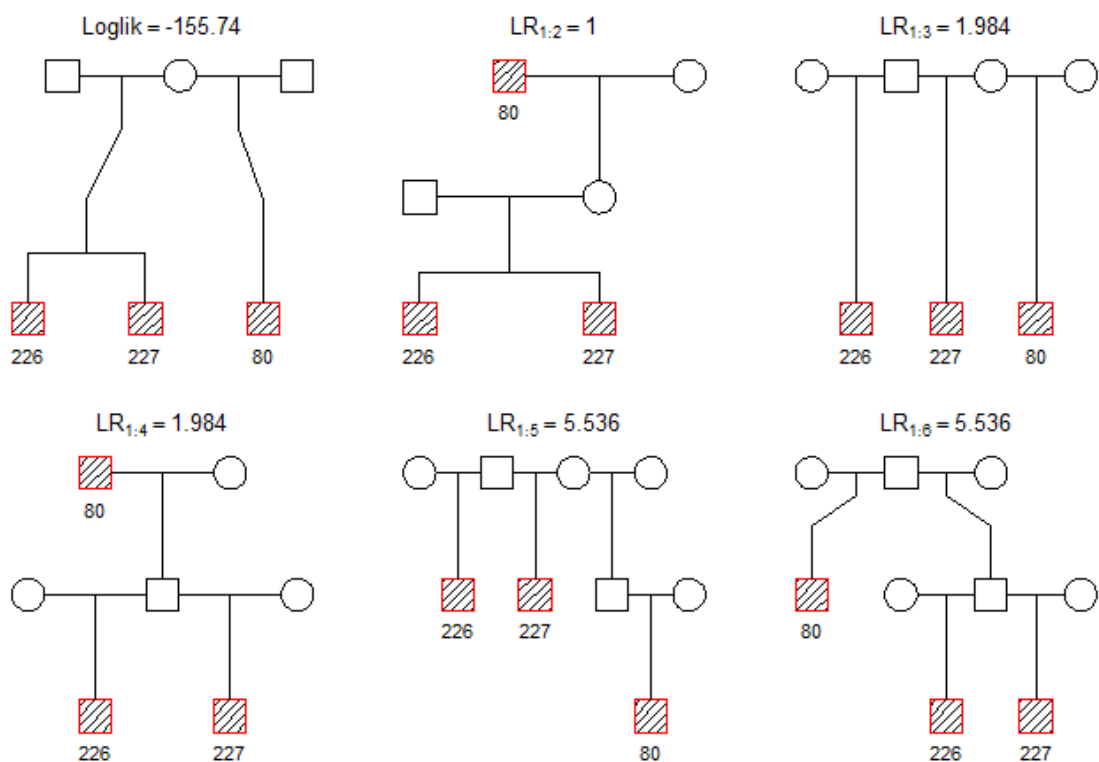
Źródło: opracowanie własne



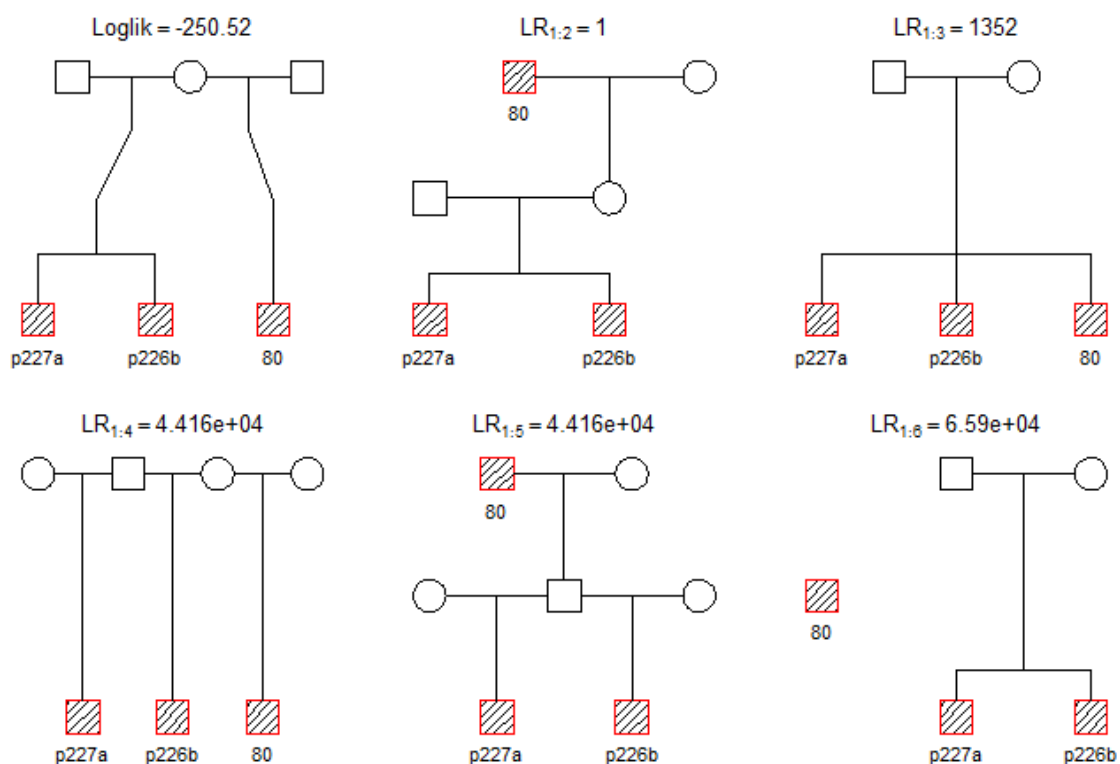
Rys. 16. Wykres trójkąta IBD dla Rodziny 4 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity



Rys. 17. Symulacje pokrewieństw dla Rodziny 4 dla zestawu GlobalFiler™



Rys. 18. Symulacje pokrewieństw dla Rodziny 4 dla zestawu KinFinder



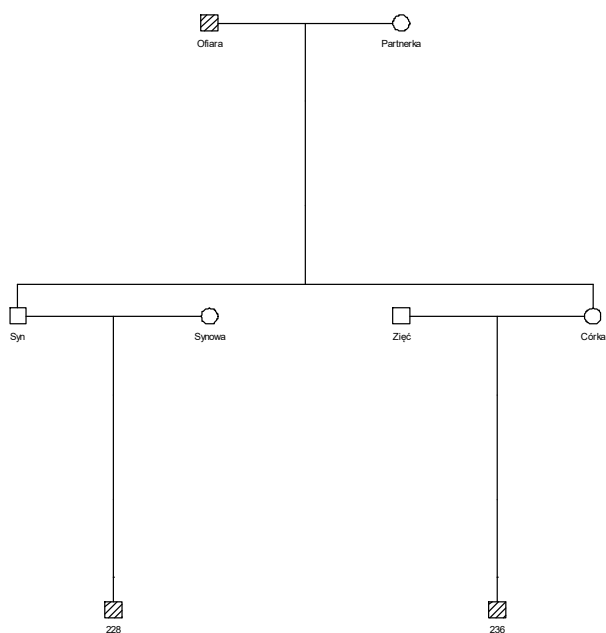
Rys. 19. Symulacje pokrewieństw dla Rodziny 4 dla zestawu Precision ID Identity

4.5. Rodzina 5

W skład piątej rodziny wchodzi dwóch wnuków zaginionego będących kuzynami (osoby 228 i 236), relacje między nimi przedstawiono poniżej (Rys. 20).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 102. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 330\,014$, która po uwzględnieniu sprzężeń genetycznych wzrosła do 438 688. Analiza w programie KinBN dała wartości odpowiednio 338 000 (bez sprzężeń) i 447 000 (ze sprzężeniami). Analiza haplotypów chromosomu Y wykazała pełną zgodność a wartość $LR = 189\,665$. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 6,26 \cdot 10^{10}$.

Symulacje rozkładu LR (Tab. 6) wskazują, że uzyskane wyniki mieszczą się w oczekiwanych przedziałach i znajdują się powyżej mediany dla zestawu GlobalFiler™. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder, a najniższą Precision ID Identity. Analiza trójkąta IBD (Rys. 21) wykazuje zgodność z zakładanym pokrewieństwem. Symulacje możliwych pokrewieństw (Rys. 22) wykazały drobne różnice jednakże analizując ograniczenia programu można uznać wskazania za trafne.

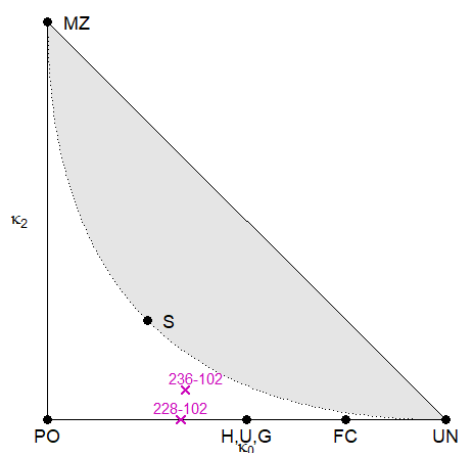


Rys. 20. Drzewo genealogiczne dla Rodziny 5

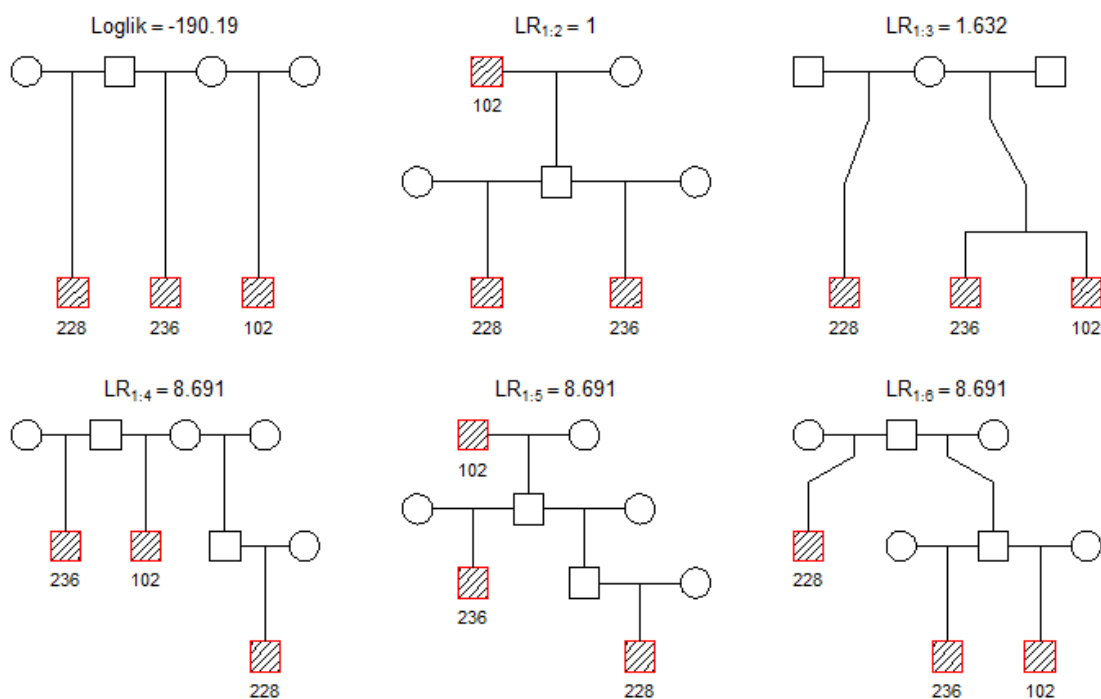
Tab. 6. Symulacja wyników analizy statystycznej dla Rodziny 5

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	2093	$1,58 \cdot 10^7$	$2,38 \cdot 10^6$	2,829	$3,49 \cdot 10^8$
	H 1	H 2 (TRUE)	0,0007	0,7654	0,3949	$2,08 \cdot 10^{-6}$	12,78
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	4826	$4 \cdot 10^7$	$5,04 \cdot 10^6$	4,234	$8,65 \cdot 10^8$
	H 1	H 2 (TRUE)	0,0003	0,6908	0,1833	$6,22 \cdot 10^{-7}$	15,53
KinFinder	H 1 (TRUE)	H 2	$3,19 \cdot 10^{10}$	$1,32 \cdot 10^{17}$	$3,5 \cdot 10^{15}$	$3,52 \cdot 10^5$	$2,85 \cdot 10^{18}$
	H 1	H 2 (TRUE)	$1,9 \cdot 10^{-10}$	0,0002	$3 \cdot 10^{-6}$	$5,1 \cdot 10^{-15}$	0,0059
Precision ID Identity	H 1 (TRUE)	H 2	245,7	$1,28 \cdot 10^4$	$3,55 \cdot 10^4$	1,272	$7,09 \cdot 10^4$
	H 1	H 2 (TRUE)	0,003	2,184	1,021	$7,65 \cdot 10^{-6}$	60,11

Źródło: opracowanie własne



Rys. 21. Wykres trójkąta IBD dla Rodziny 5



Rys. 22. Symulacje pokrewieństw dla Rodziny 5

4.6. Rodzina 6

Rodzina szósta obejmuje troje wnucząt osoby zaginionej: wnuczkę (osoba 229) oraz wnuka (osoba 231) pochodzących z jednego związku, a także wnuka (osoba 230) pochodzącego z innego związku tego samego syna zaginionego, relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 23).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 88. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 8\,639$, która po uwzględnieniu sprzężeń genetycznych wzrosła do $9\,439$. Analiza w programie KinBN dała wartości odpowiednio $9\,710$ (bez sprzężeń) i $11\,500$ (ze sprzężeniami). Analiza haplotypów chromosomu Y wykazała pełną zgodność a wartość $LR = 189\,665$.

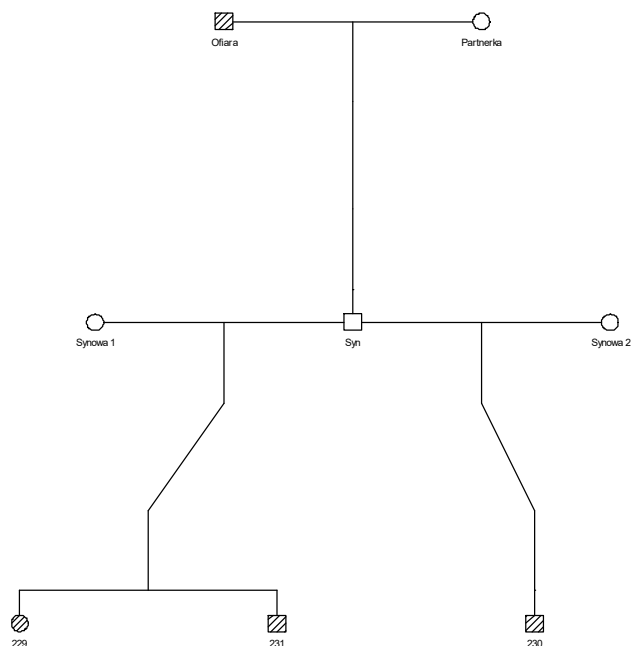
Oznaczenia profili STR autosomalnych zostały rozszerzone o zestaw KinFinder. Obliczenia w programie Familias wykazały znaczny wzrost wartości LR do $237\,619$ (bez sprzężeń) i $998\,000$ (ze sprzężeniami). Analiza w programie KinBN dla zestawu KinFinder dała wartości odpowiednio $1,32 \cdot 10^6$ (bez sprzężeń) i $6,66 \cdot 10^6$ (ze sprzężeniami).

W trakcie szczegółowej analizy układów STR zwrócono uwagę na wystąpienie niezgodności w locus D17L2556. W analizowanym układzie wnuki posiadają wartości: 229 – 20, 24; 230 – 21, 21; 231 – 19, 20, co sugeruje, że ich ojciec powinien mieć genotyp 21, 20, podczas gdy w profilu szczątków stwierdzono wartości 19, 22. Aby zachować możliwość zakładanego pokrewieństwa, założono wystąpienie mutacji w tym locus z prawdopodobieństwem równym 0,002. W programie Familias zastosowano model mutacji Stepwise (Stationary).

Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 3,89 \cdot 10^{14}$.

Symulacje rozkładu LR (Tab. 7) wskazują, że uzyskane wyniki zarówno dla zestawu GlobalFiler™ jak i KinFinder mieszczą się w oczekiwanych przedziałach i znajdują się poniżej mediany. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (nawet uwzględniając niepełne oznaczenie profilu – 15 układów z 50), a najniższą Precision ID Identity.

Analiza trójkąta IBD (Rys. 24) oraz symulacje możliwych pokrewieństw (Rys. 25) dla zestawu GlobalFiler™ wykazują zgodność z zakładanym pokrewieństwem, natomiast trójkąt IBD (Rys. 24) i symulacje pokrewieństw (Rys. 26) dla zestawu KinFinder wykazały drobne różnice.

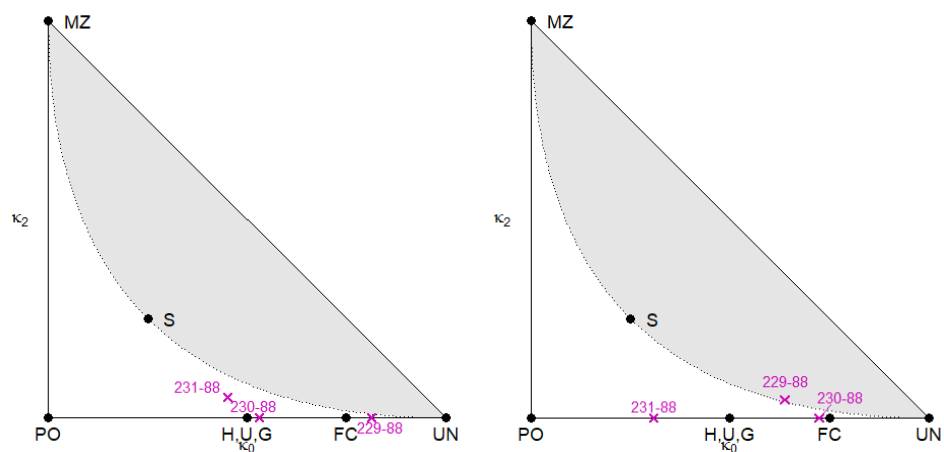


Rys. 23. Drzewo genealogiczne dla Rodziny 6

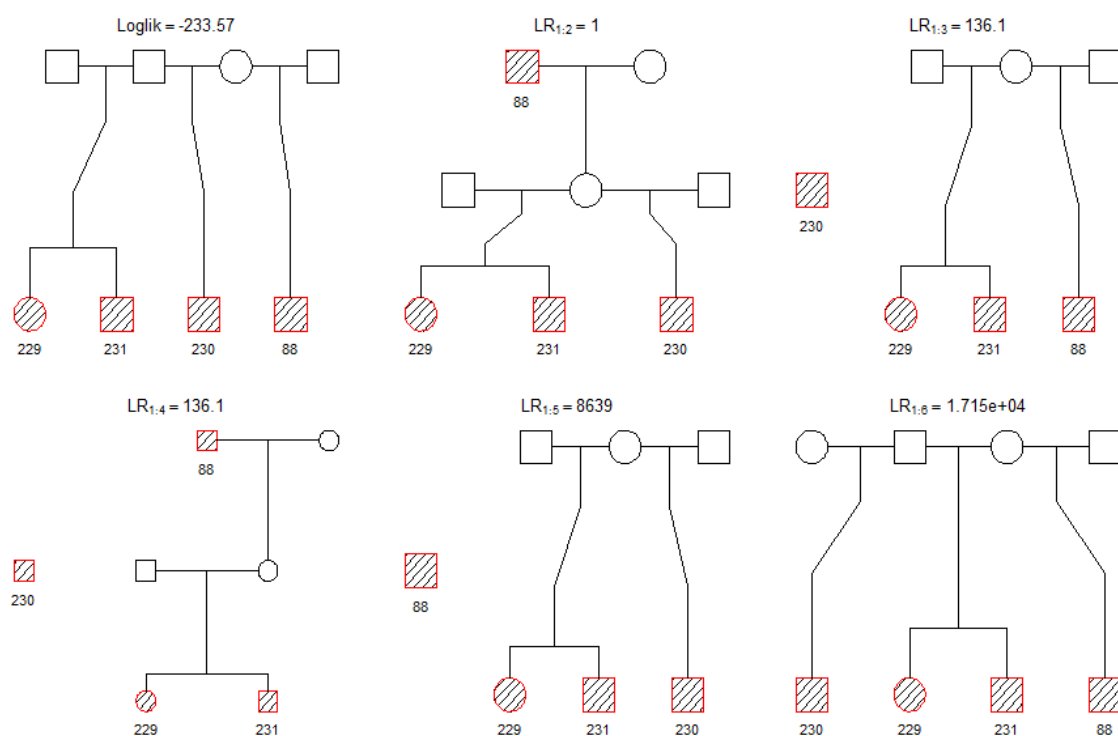
Tab. 7. Symulacja wyników analizy statystycznej dla Rodziny 6

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$2,56 \cdot 10^4$	$1,54 \cdot 10^7$	$1,6 \cdot 10^7$	71,8	$1,25 \cdot 10^8$
	H 1	H 2 (TRUE)	0	0,511	0,0002	0	9,198
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$8,89 \cdot 10^4$	$6,75 \cdot 10^8$	$8,31 \cdot 10^7$	149,3	$1,62 \cdot 10^{10}$
	H 1	H 2 (TRUE)	0	0,0075	$1,525 \cdot 10^{-5}$	0	0,0938
KinFinder	H 1 (TRUE)	H 2	$1,99 \cdot 10^{14}$	$5,73 \cdot 10^{20}$	$5,54 \cdot 10^{18}$	$7,63 \cdot 10^9$	$1,43 \cdot 10^{22}$
	H 1	H 2 (TRUE)	0	$2,56 \cdot 10^{-8}$	0	0	$8,11 \cdot 10^{-7}$
KinFinder (układy obecne w profilu szczątków nr 88)	H 1 (TRUE)	H 2	$1,64 \cdot 10^6$	$9,52 \cdot 10^8$	$1,78 \cdot 10^9$	1342	$1,09 \cdot 10^{10}$
	H 1	H 2 (TRUE)	0	0,0207	0	0	0,5064
Precision ID Identity	H 1 (TRUE)	H 2	390,6	$3,71 \cdot 10^5$	$5,44 \cdot 10^4$	1,735	$1,09 \cdot 10^7$
	H 1	H 2 (TRUE)	0,0009	3,216	0,6129	$4,98 \cdot 10^{-7}$	95,31

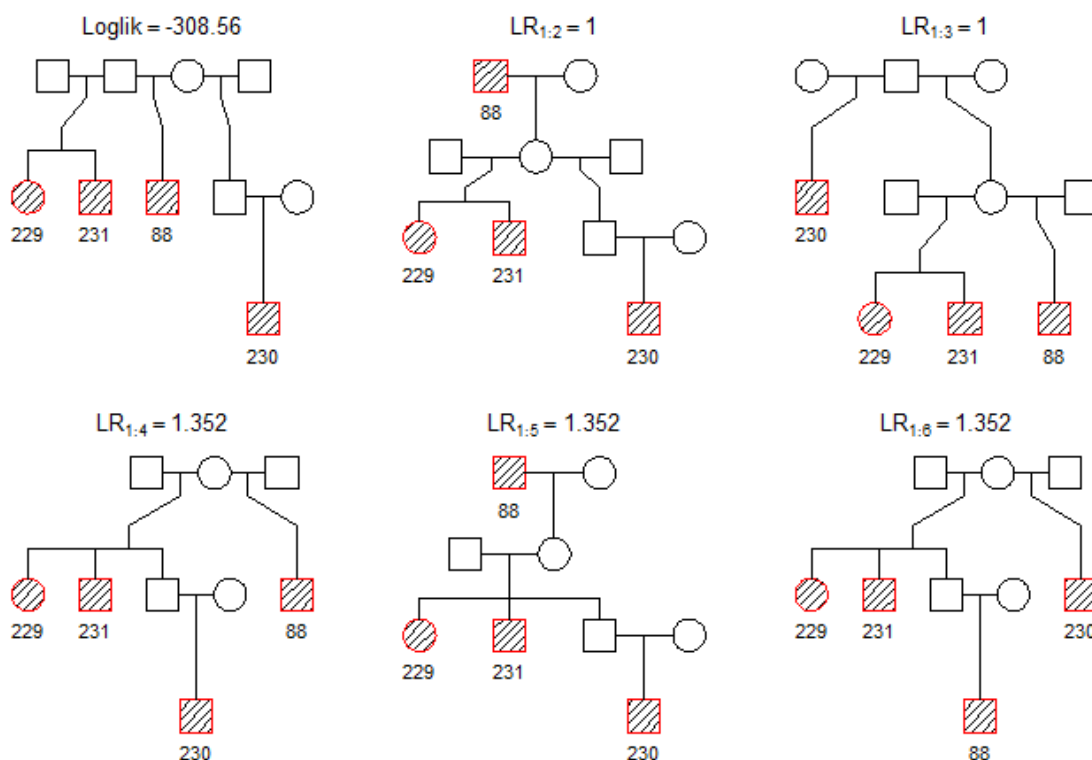
Źródło: opracowanie własne



Rys. 24. Wykres trójkąta IBD dla Rodziny 6 dla zestawu GlobalFiler™ i KinFinder



Rys. 25. Symulacje pokrewieństw dla Rodziny 6 dla zestawu GlobalFiler™



Rys. 26. Symulacje pokrewieństw dla Rodziny 6 dla zestawu KinFinder

4.7. Rodzina 7

Rodzina siódma obejmuje dwóch wnuków zaginionego (osoby 256 i 257), relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 27).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 98. Analiza haplotypów chromosomu Y wykazała pełną zgodność ($LR = 190\,556$). Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 509$, która po uwzględnieniu sprzężeń genetycznych zmalała do 429. Analiza w programie KinBN dała wartości odpowiednio 595 (bez sprzężeń) i 502 (ze sprzężeniami).

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestawy KinFinder oraz Precision ID Identity.

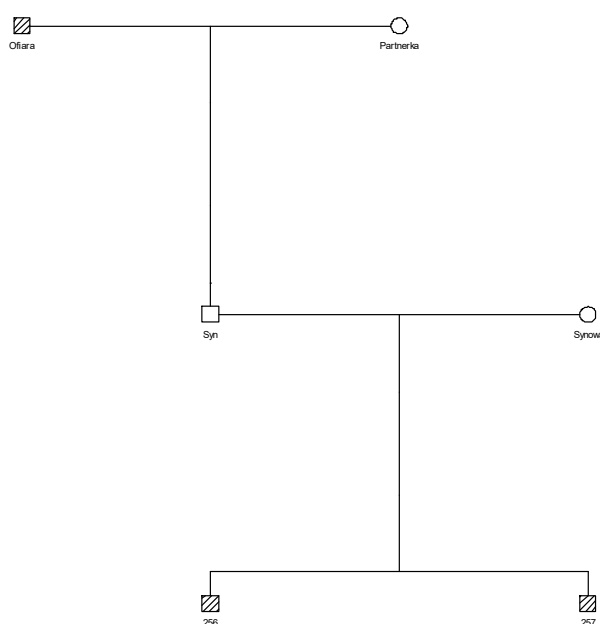
Dla zestawu KinFinder obliczenia w programie Familias $LR = 305\,342$ (bez sprzężeń; w profilu dowodowym brak układów sprzężonych), natomiast w KinBN $LR = 263\,000$ (bez sprzężeń).

Dla zestawu Precision ID Identity obliczenia w programie Familias $LR = 34$ (bez sprzężeń – program FamLink nie wykonał obliczeń sprzężeń dla tej rodziny), a w KinBN odpowiednio $LR = 35$ i $LR = 53$.

Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 1,01 \cdot 10^{15}$.

Symulacje rozkładu LR (Tab. 8) wskazują, że uzyskane wyniki dla wszystkich zestawów mieszczą się w oczekiwanych przedziałach. Dla zestawów GlobalFiler™ i Precision ID Identity wartości znajdowały się poniżej mediany, natomiast dla zestawu KinFinder– powyżej progu 95%. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (uwzględniając niepełne oznaczenie profilu – 15 układów z 50 zestaw ten miał niższe wyniki niż GlobalFiler™), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 28) oraz symulacje możliwych pokrewieństw dla wszystkich zestawów (Rys. 29, Rys. 30, Rys. 31) wykazują zgodność z zakładanym pokrewieństwem.

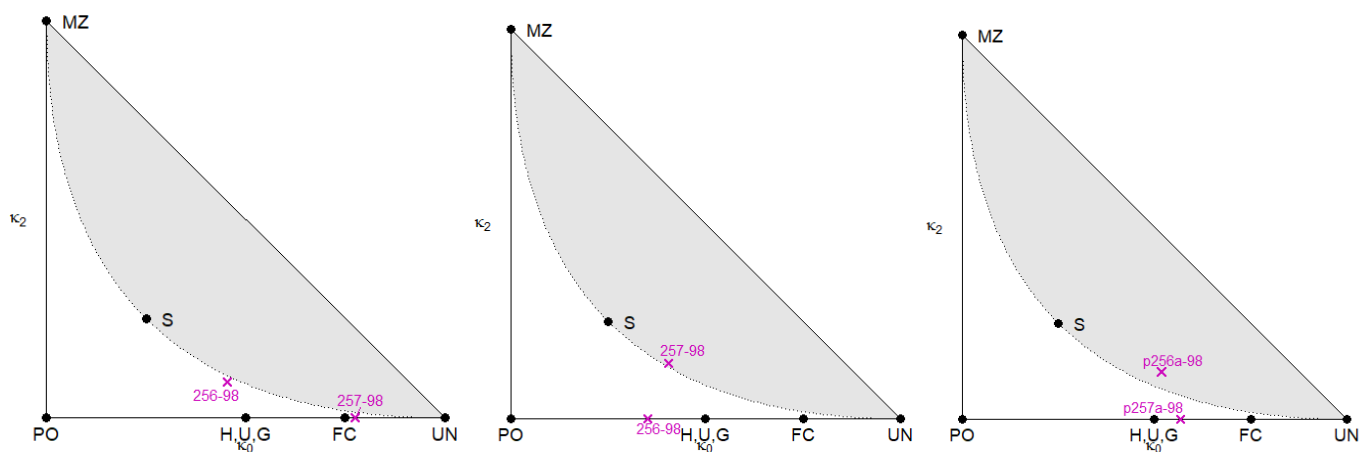


Rys. 27. Drzewo genealogiczne dla Rodziny 7

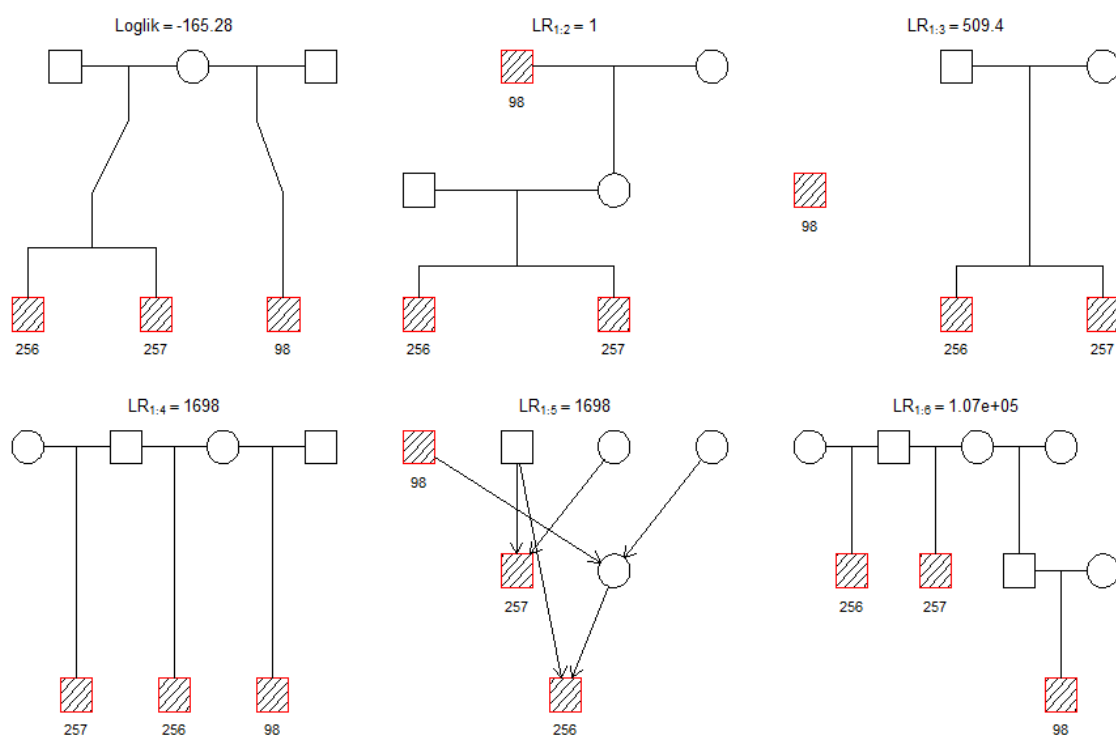
Tab. 8. Symulacja wyników analizy statystycznej dla Rodziny 7

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	785,7	$2,64 \cdot 10^5$	$2,32 \cdot 10^5$	3,731	$2,4 \cdot 10^6$
	H 1	H 2 (TRUE)	$3,54 \cdot 10^{-5}$	0,3305	0,5225	0	3,369
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	1646	$1,49 \cdot 10^6$	$6,79 \cdot 10^5$	4,605	$2,13 \cdot 10^7$
	H 1	H 2 (TRUE)	0	0,74	0,1284	0	15,83
KinFinder	H 1 (TRUE)	H 2	$2,4 \cdot 10^9$	$1,73 \cdot 10^{15}$	$2,95 \cdot 10^{13}$	$6,53 \cdot 10^4$	$3,16 \cdot 10^{16}$
	H 1	H 2 (TRUE)	0	$8,49 \cdot 10^{-5}$	$2,5 \cdot 10^{-12}$	0	0,0023
KinFinder (układy obecne w profilu szczątków nr 98)	H 1 (TRUE)	H 2	778	$5,71 \cdot 10^4$	$1,18 \cdot 10^5$	2,963	$6,98 \cdot 10^5$
	H 1	H 2 (TRUE)	0	0,5704	0,5196	0	7,319
Precision ID Identity	H 1 (TRUE)	H 2	67,9	1419	4176	0,5341	9577
	H 1	H 2 (TRUE)	0,009	0,7288	2,11	$3,52 \cdot 10^{-5}$	5,074

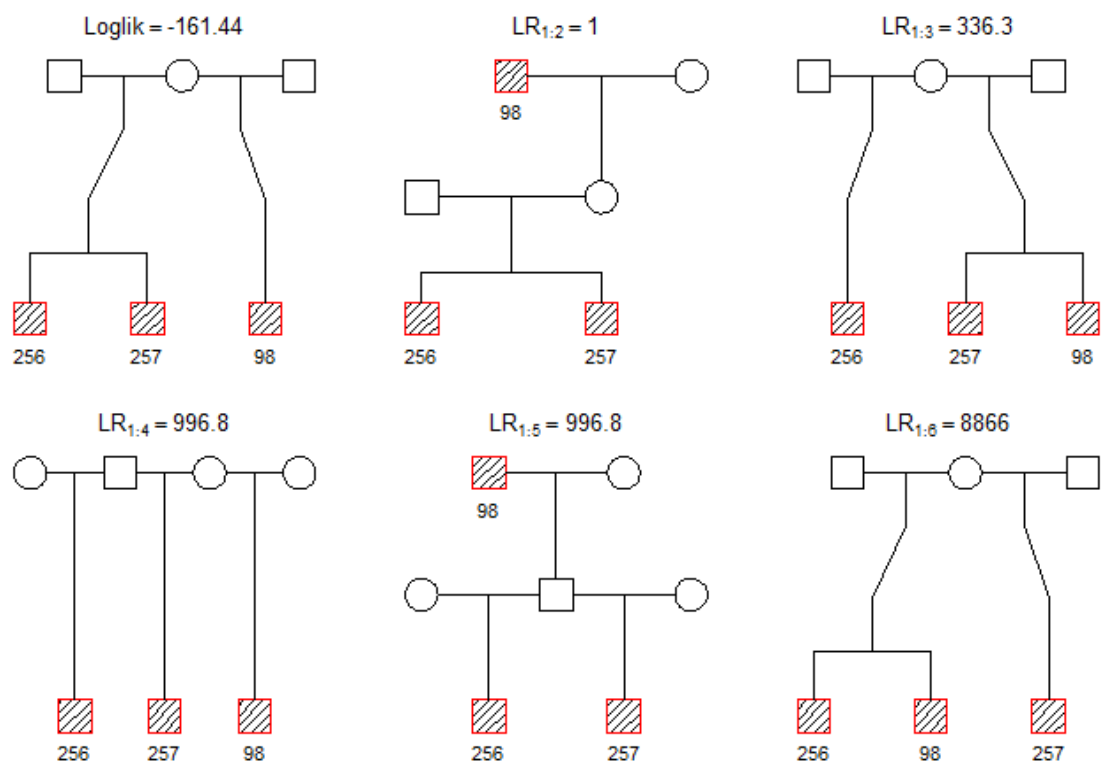
Źródło: opracowanie własne



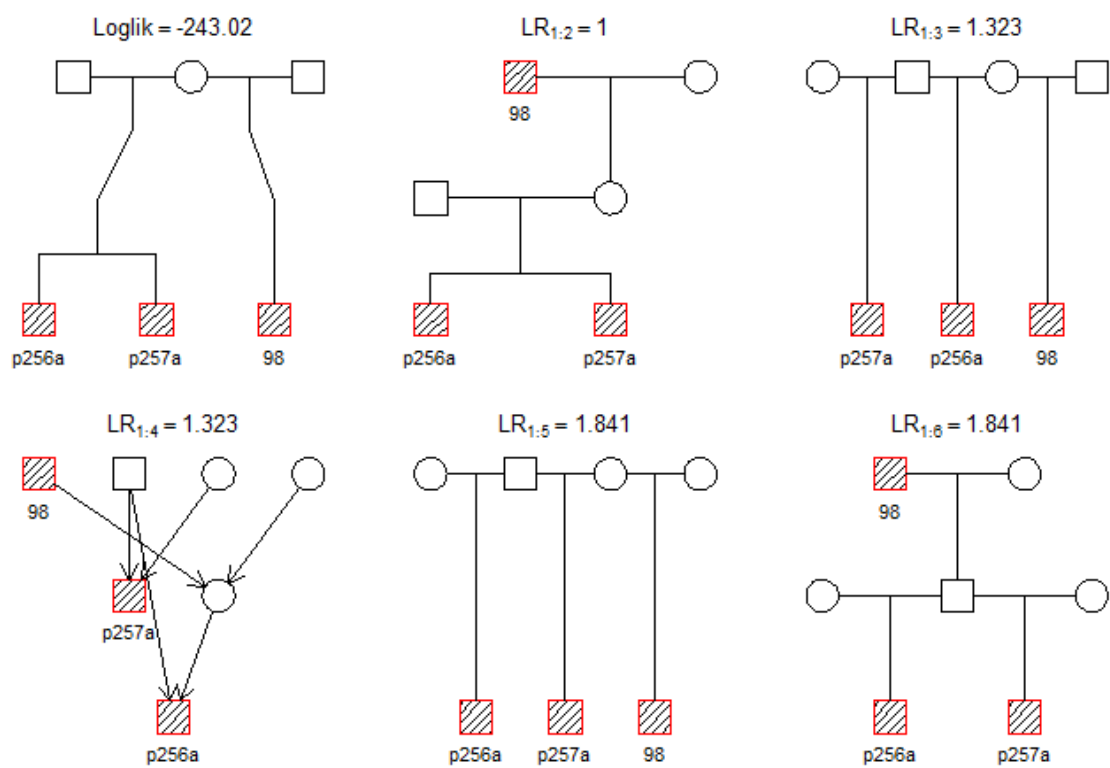
Rys. 28. Wykres trójkąta IBD dla Rodziny 7 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity



Rys. 29. Symulacje pokrewieństw dla Rodziny 7 dla zestawu GlobalFiler™



Rys. 30. Symulacje pokrewieństw dla Rodziny 7 dla zestawu KinFinder



Rys. 31. Symulacje pokrewieństw dla Rodziny 7 dla zestawu Precision ID Identity

4.8. Rodzina 8

Rodzina ósma obejmuje dwie wnuczki zaginionej (osoby 233 i 235), które są siostrami, relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 32).

Po przeszukaniu DVI stwierdzono dopasowanie do szczątków nr 86. Z uwagi na brak krewnych w linii męskiej nie można potwierdzić dopasowania za pomocą markerów chromosomu Y.

Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 47$, która po uwzględnieniu sprzężeń genetycznych wzrosła do 55. Analiza w programie KinBN dała wartości odpowiednio 39 (bez sprzężeń) i 38 (ze sprzężeniami).

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestaw KinFinder.

Dla zestawu KinFinder obliczenia w programie Familias $LR = 0,14$ (bez sprzężeń) i $0,44$ (ze sprzężeniami), natomiast w KinBN $LR = 0,31$ (bez sprzężeń) i $0,98$ (ze sprzężeniami).

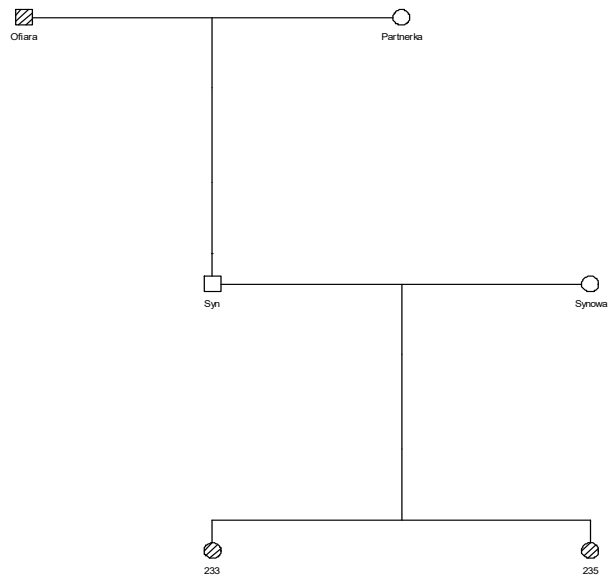
Obliczono łączny iloraz wiarygodności $LR_c = 6,6$.

Symulacje rozkładu LR (Tab. 9) wskazują, że uzyskane wyniki dla zestawu GlobalFiler™ znajdują się poniżej mediany, ale powyżej progu 5%. Wynik uzyskany dla zestawu KinFinder wspiera drugą hipotezę, o braku pokrewieństwa pomiędzy Rodziną a szczątkami, i znajduje się poniżej wartości średniej dla takich wyników.

Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (uwzględniając niepełne oznaczenie profilu – 17 układów z 50 zestaw ten miał porównywalne wyniki z zestawem GlobalFiler™), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 33) oraz symulacje możliwych pokrewieństw dla wszystkich zestawów (Rys. 34, Rys. 35) wykazują niezgodność z zakładanym pokrewieństwem – pokrewieństwo pomiędzy osobami 233 i 235 jest potwierdzone, natomiast program wskazał inne pokrewieństwo pomiędzy Rodziną a szczątkami.

Z uwagi na otrzymane wyniki oraz fakt, iż Rodzina ósma poszukuje osoby spoza listy egzekucyjnej, nie rozszerzano badań o zestaw Precision ID Identity dla tej rodziny, a dopasowanie uznano za negatywne.

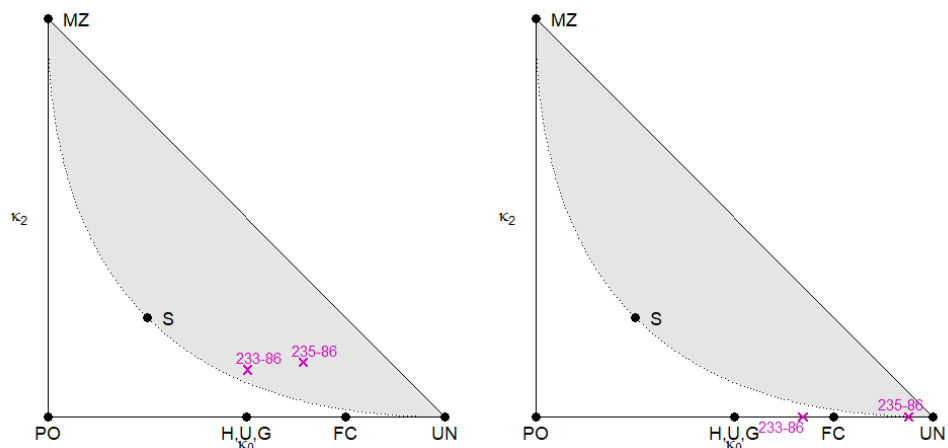


Rys. 32. Drzewo genealogiczne dla Rodziny 8

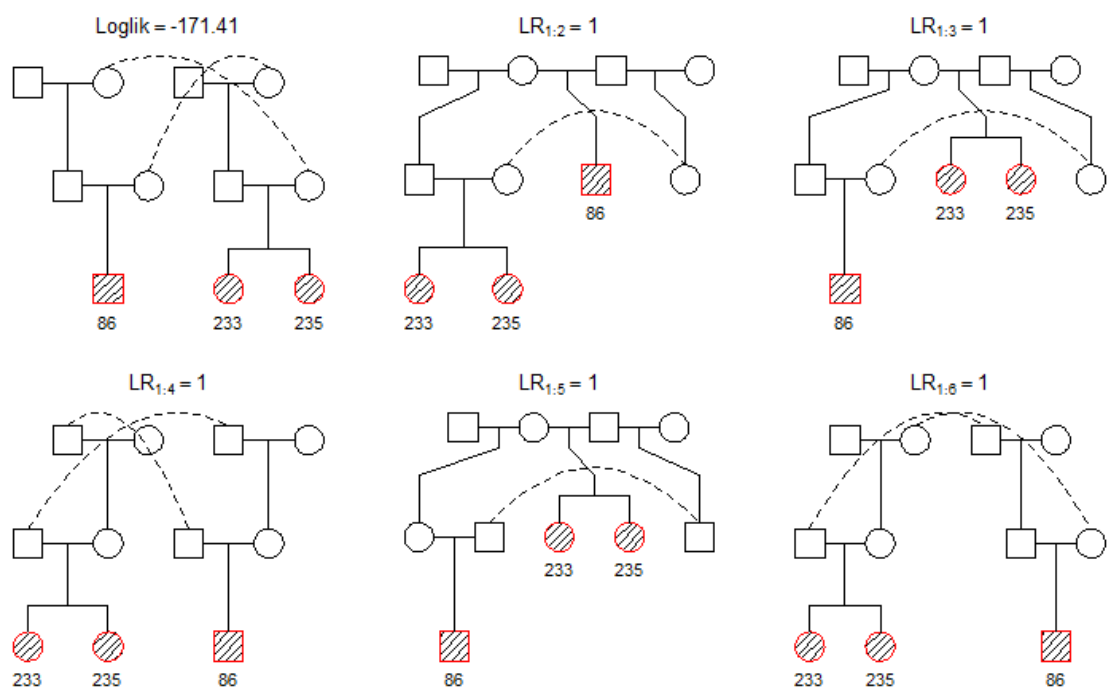
Tab. 9. Symulacja wyników analizy statystycznej dla Rodziny 8

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	813,2	$1,46 \cdot 10^5$	$2,61 \cdot 10^5$	2,814	$1,38 \cdot 10^6$
	H 1	H 2 (TRUE)	$1,66 \cdot 10^{-5}$	0,6671	0,5937	0	13,48
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	1812	$5,86 \cdot 10^5$	$8,29 \cdot 10^5$	4,447	$6,1 \cdot 10^6$
	H 1	H 2 (TRUE)	0	6,488	0,2326	0	201,6
KinFinder	H 1 (TRUE)	H 2	$3,87 \cdot 10^9$	$3,1 \cdot 10^{14}$	$5,04 \cdot 10^{13}$	$2,38 \cdot 10^5$	$7,24 \cdot 10^{15}$
	H 1	H 2 (TRUE)	0	0,0002	$5,67 \cdot 10^{-12}$	0	0,0055
KinFinder (układy obecne w profilu szczątków nr 86)	H 1 (TRUE)	H 2	2365	$1,46 \cdot 10^6$	$4,57 \cdot 10^5$	6,713	$3,42 \cdot 10^7$
	H 1	H 2 (TRUE)	0	0,8801	0,215	0	20,9
Precision ID Identity	H 1 (TRUE)	H 2	47,78	1222	4989	0,6084	5199
	H 1	H 2 (TRUE)	0,0094	0,8958	1,73	$2,28 \cdot 10^{-5}$	6,738

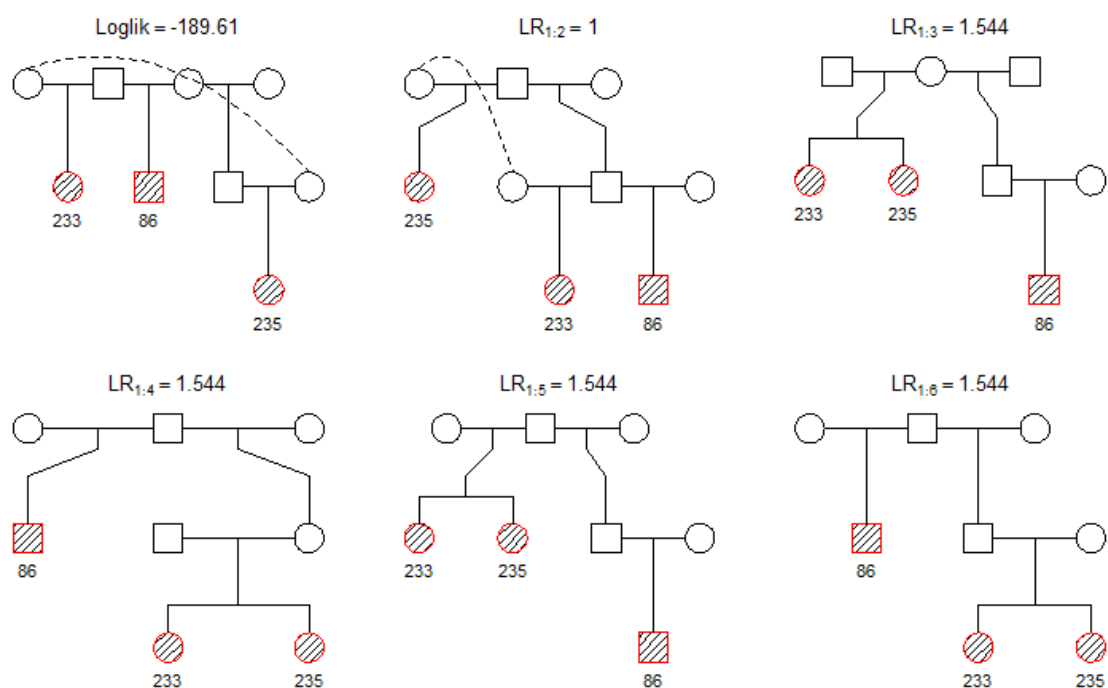
Źródło: opracowanie własne



Rys. 33. Wykres trójkąta IBD dla Rodziny 8 dla zestawu GlobalFiler™ i KinFinder



Rys. 34. Symulacje pokrewieństw dla Rodziny 8 dla zestawu GlobalFiler™



Rys. 35. Symulacje pokrewieństw dla Rodziny 8 dla zestawu KinFinder

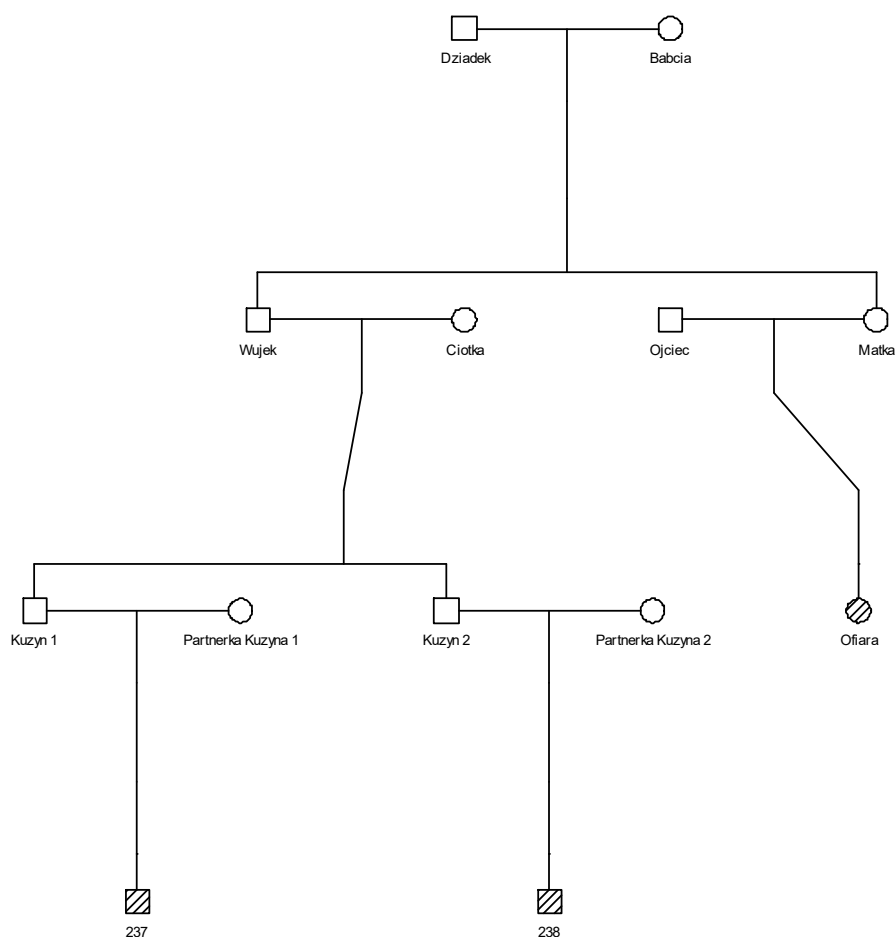
4.9. Rodzina 9

Rodzina dziewiąta jako jedyna poszukiwała osoby spokrewnionej w linii bocznej. Matka osoby zaginionej była siostrą dziadka osób 237 i 238. Osoby 237 i 238 są kuzynami (ich ojcowie są braćmi), a relacje przedstawiono w formie drzewa genealogicznego (Rys. 36).

Ustalono, że rodzina poszukuje ofiary płci żeńskiej. Wobec oznaczenia płci wszystkich analizowanych szczątków jako męskiej, wykluczono możliwość dopasowania genetycznego w tym przypadku.

Jednakże, w celu oceny przydatności zastosowanych zestawów markerów, wykonano symulacje wyników analizy statystycznej możliwych do uzyskania przy analizowanym stopniu pokrewieństwa (Tab. 10).

Wyniki symulacji wskazują, że oczekiwana wartość LR dla tego pokrewieństwa powinna mieścić się w przedziale od około 1 do 17 dla GlobalFiler™, od 1 do 241 dla KinFinder oraz od 1 do 5 dla Precision ID Identity. Przy tak odległym pokrewieństwie wszystkie zestawy wykazują niewielką siłę dyskryminacji.



Rys. 36. Drzewo genealogiczne dla Rodziny 9

Tab. 10. Symulacja wyników analizy statystycznej dla Rodziny 9

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	1,611	5,392	16,54	0,3157	16,61
	H 1	H 2 (TRUE)	0,5774	1,024	2,896	0,1358	2,248
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	1,641	11,52	23,42	0,327	174,9
	H 1	H 2 (TRUE)	0,5396	1,004	2,906	0,1259	2,042
KinFinder	H 1 (TRUE)	H 2	6,381	94,85	241,7	0,2362	873
	H 1	H 2 (TRUE)	0,1993	0,8275	3,227	0,0128	2,246
Precision ID Identity	H 1 (TRUE)	H 2	1,321	1,797	4,648	0,362	1,602
	H 1	H 2 (TRUE)	0,7531	1,003	2,752	0,1895	0,9122

Źródło: opracowanie własne

4.10. Rodzina 10

Rodzina 10 obejmuje dwie wnuczki osoby zaginionej (osoby 239 i 240), relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 37).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerami 93, 94, 95. Dalsze obliczenia wykonano dla najbardziej prawdopodobnego wskazania.

Obliczenia statystyczne wykonane w programie Familias wykazały dla dopasowania z profilem nr 93 wartość $LR = 590$, która po uwzględnieniu sprzężeń genetycznych wzrosła do 612. Analiza w programie KinBN dała wartości odpowiednio 677 (bez sprzężeń) i 701 (ze sprzężeniami).

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestawy KinFinder oraz Precision ID Identity.

Dla zestawu KinFinder obliczenia w programie Familias $LR = 4\,930$ (bez sprzężeń) i 2 958 (ze sprzężeniami), natomiast w KinBN $LR = 6\,050$ (bez sprzężeń) oraz $LR = 4\,170$ (ze sprzężeniami).

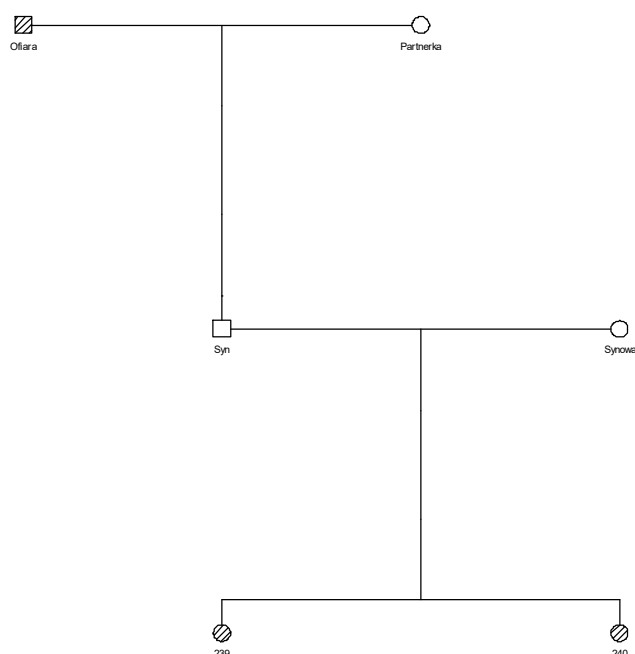
Dla zestawu Precision ID Identity obliczenia w programie Familias $LR = 491$ (bez sprzężeń- program FamLink nie wykonał obliczeń sprzężeń dla tej rodziny), a w KinBN odpowiednio $LR = 491$ i $LR = 545$.

Obliczono łączny iloraz wiarygodności dla STR autosomalnych $LR_c = 1,43 \cdot 10^9$.

Symulacje rozkładu LR (Tab. 11) wskazują, że uzyskane wyniki dla wszystkich zestawów mieszczą się w oczekiwanych przedziałach. Dla zestawów GlobalFiler™ i KinFinder wartości znajdowały się poniżej mediany, natomiast dla zestawu Precision ID Identity – powyżej mediany, ale poniżej średniej. Najwyższą moc dowodową w symulacji wyników wykazał

zestaw KinFinder (nawet uwzględniając niepełne oznaczenie profilu – 21 układów z 50), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 38) oraz symulacje możliwych pokrewieństw dla wszystkich zestawów (Rys. 39, Rys. 40 i Rys. 41) wykazują zgodność z zakładanym pokrewieństwem.



Rys. 37. Drzewo genealogiczne dla Rodziny 10

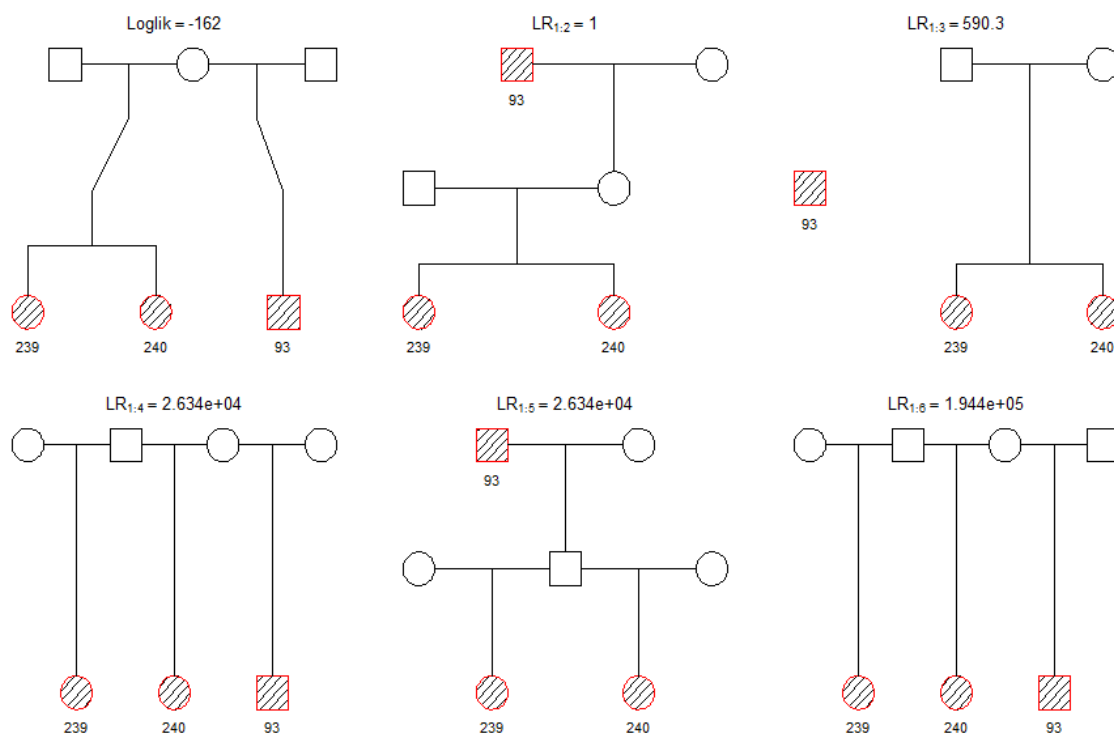
Tab. 11. Symulacja wyników analizy statystycznej dla Rodziny 10

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	762,8	$4,15 \cdot 10^5$	$2,8 \cdot 10^5$	3,563	$4,79 \cdot 10^6$
	H 1	H 2 (TRUE)	$1,1 \cdot 10^{-5}$	0,4963	0,2897	0	8,536
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	1631	$6,03 \cdot 10^5$	$7,59 \cdot 10^5$	3,627	$7,26 \cdot 10^6$
	H 1	H 2 (TRUE)	0	0,7673	0,2191	0	17,25
KinFinder	H 1 (TRUE)	H 2	$2,12 \cdot 10^9$	$2,7 \cdot 10^{14}$	$4,12 \cdot 10^{13}$	$1,67 \cdot 10^5$	$4,6 \cdot 10^{15}$
	H 1	H 2 (TRUE)	0	$3,81 \cdot 10^{-5}$	$2,36 \cdot 10^{-10}$	0	0,001
KinFinder (układy obecne w profilu szczątków nr 93)	H 1 (TRUE)	H 2	$1,05 \cdot 10^4$	$4,64 \cdot 10^6$	$4,2 \cdot 10^6$	13,69	$6,32 \cdot 10^7$
	H 1	H 2 (TRUE)	0	0,0699	0,0658	0	0,7517
Precision ID Identity	H 1 (TRUE)	H 2	58,9	1512	2729	0,4359	$2,64 \cdot 10^4$
	H 1	H 2 (TRUE)	0,0111	0,897	1,343	$3,75 \cdot 10^{-5}$	11,06

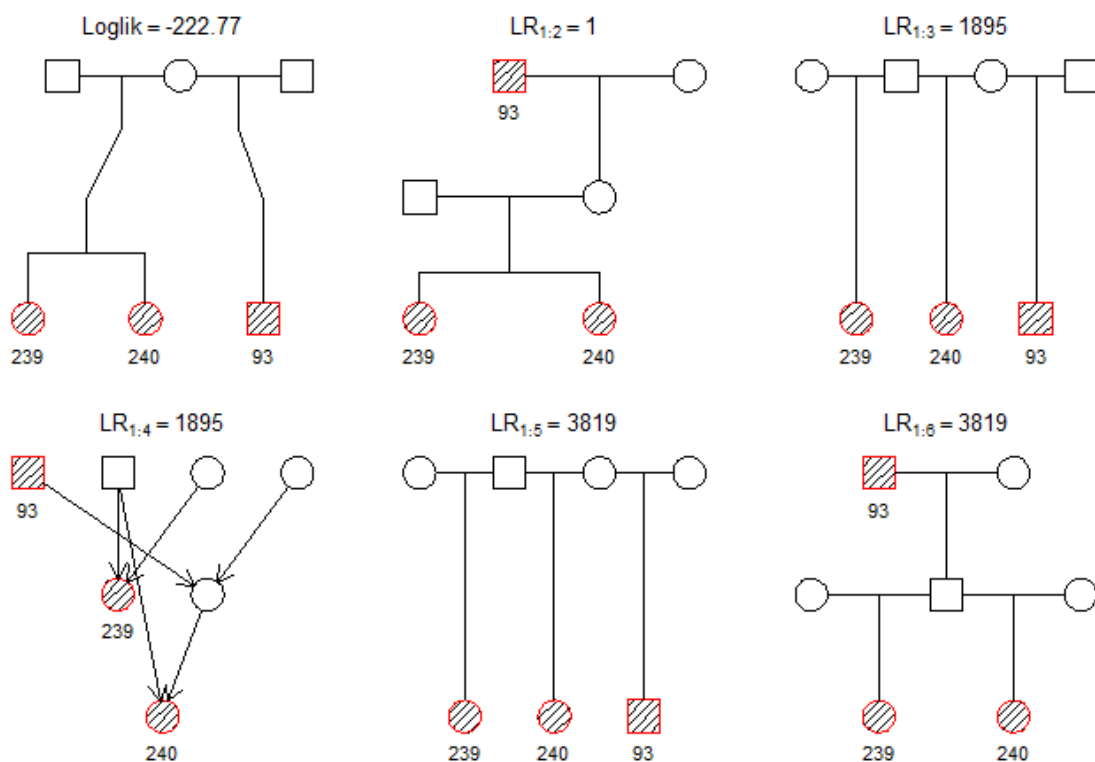
Źródło: opracowanie własne



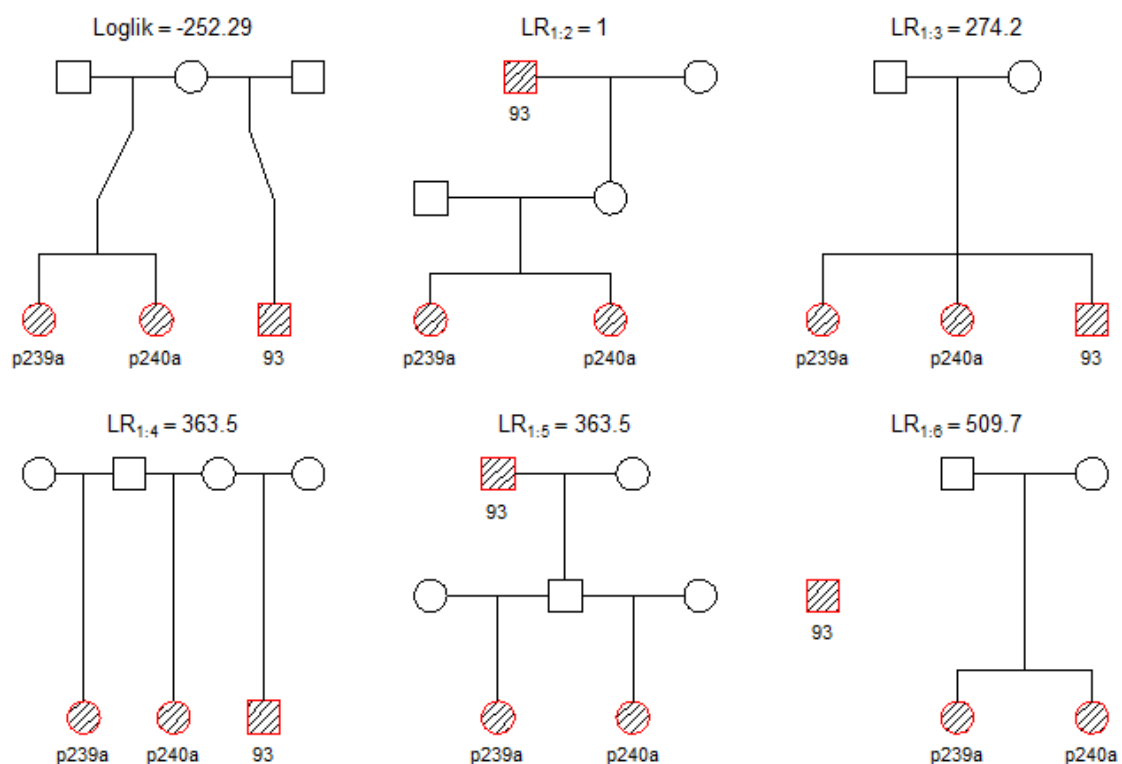
Rys. 38. Wykres trójkąta IBD dla Rodziny 10 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity



Rys. 39. Symulacje pokrewieństw dla Rodziny 10 dla zestawu GlobalFiler™



Rys. 40. Symulacje pokrewieństw dla Rodziny 10 dla zestawu KinFinder



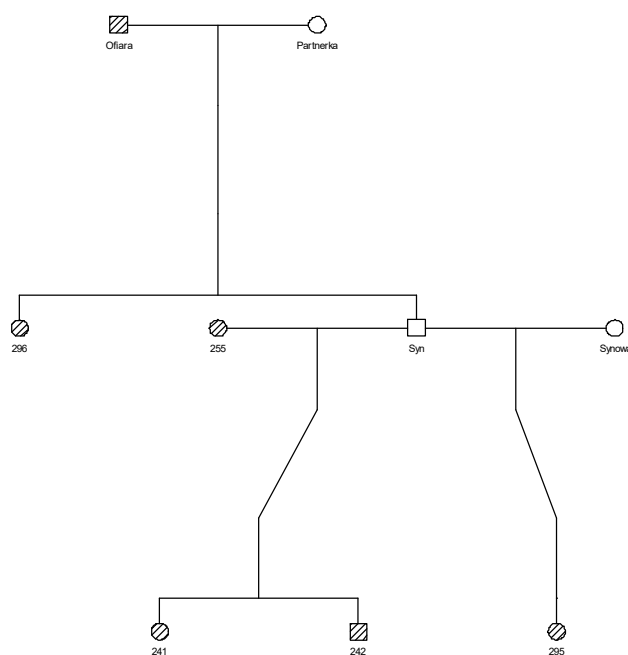
Rys. 41. Symulacje pokrewieństw dla Rodziny 10 dla zestawu Precision ID Identity

4.11. Rodzina 11

Rodzina 11 obejmuje córkę osoby zaginionej (osoba 296) oraz troje wnucząt: dwoje dzieci pochodzących ze związku syna osoby zaginionej z osobą 255 (osoby 241 i 242) oraz jedno dziecko z innego związku syna osoby zaginionej (osoba 295), relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 42).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 99. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 4,62 \cdot 10^{13}$, która po uwzględnieniu sprzężeń genetycznych wzrosła do $8,58 \cdot 10^{13}$. Analiza w programie KinBN dała wartości odpowiednio $4,24 \cdot 10^{13}$ (bez sprzężeń) i $7,94 \cdot 10^{13}$ (ze sprzężeniami). Analiza haplotypów chromosomu Y wykazała pełną zgodność i wartość $LR = 189\,665$. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 8,76 \cdot 10^{18}$.

Symulacje rozkładu LR (Tab. 12) wskazują, że uzyskane wyniki mieszczą się w oczekiwanych przedziałach i znajdują się powyżej mediany dla zestawu GlobalFiler™. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder, a najniższą Precision ID Identity. Analiza trójkąta IBD (Rys. 43) oraz symulacji możliwych pokrewieństw (Rys. 44) wykazały pełną zgodność z zakładanym pokrewieństwem.

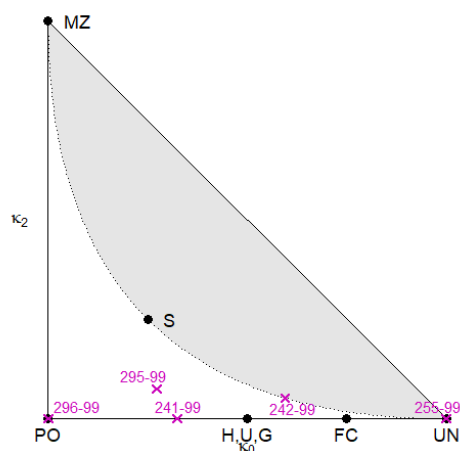


Rys. 42. Drzewo genealogiczne dla Rodziny 11

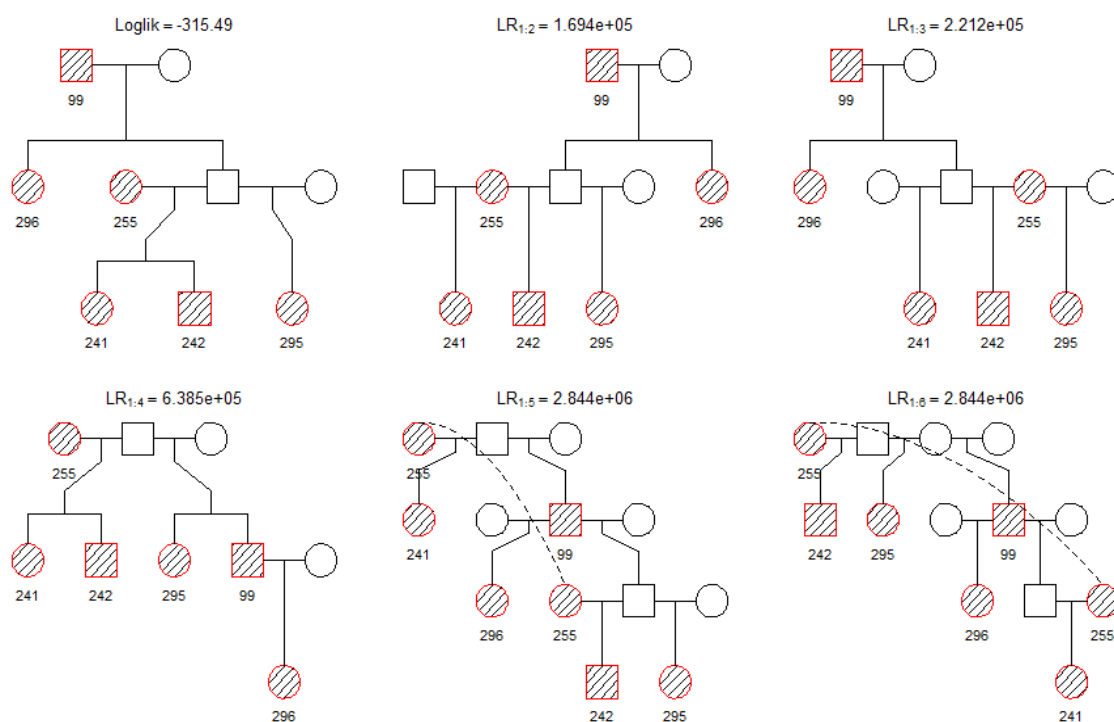
Tab. 12. Symulacja wyników analizy statystycznej dla Rodziny 11

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$3,27 \cdot 10^{11}$	$3,99 \cdot 10^{16}$	$5,43 \cdot 10^{15}$	$8,87 \cdot 10^7$	$8,88 \cdot 10^{17}$
	H 1	H 2 (TRUE)	0	0	0	0	0
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$5,36 \cdot 10^{12}$	$1,49 \cdot 10^{20}$	$1,72 \cdot 10^{17}$	$1,29 \cdot 10^9$	$3,45 \cdot 10^{21}$
	H 1	H 2 (TRUE)	0	0	0	0	0
KinFinder	H 1 (TRUE)	H 2	$1,7 \cdot 10^{36}$	$2,26 \cdot 10^{45}$	$3,85 \cdot 10^{42}$	$1,15 \cdot 10^{30}$	$4,79 \cdot 10^{46}$
	H 1	H 2 (TRUE)	0	0	0	0	0
Precision ID Identity	H 1 (TRUE)	H 2	$2,66 \cdot 10^8$	$4,49 \cdot 10^{11}$	$3,67 \cdot 10^{11}$	$3,44 \cdot 10^5$	$7,22 \cdot 10^{12}$
	H 1	H 2 (TRUE)	0	0	0	0	0

Źródło: opracowanie własne



Rys. 43. Wykres trójkąta IBD dla Rodziny 11



Rys. 44. Symulacje pokrewieństw dla Rodziny 11

4.12. Rodzina 12

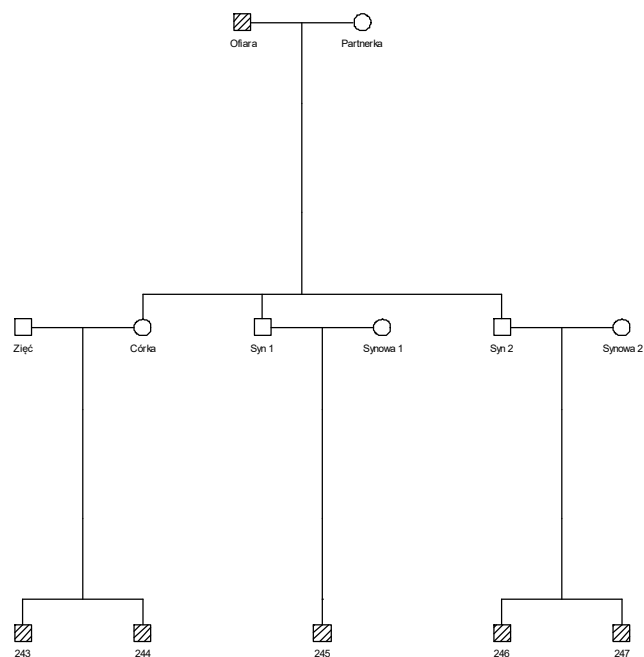
Rodzina 12 obejmuje pięcioro wnuków osoby zaginionej: osoby 243 i 244, będące synami córki osoby zaginionej, osobę 245, będącą synem jednego z synów osoby zaginionej, oraz osoby 246 i 247, będące synami drugiego syna osoby zaginionej, relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 45).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 92. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 2,31 \cdot 10^6$, Analiza w programie KinBN dała $LR = 2,28 \cdot 10^6$ (bez sprzężeń) i $4,5 \cdot 10^6$ (ze sprzężeniami). W programie FamLink nie udało się obliczyć sprzężeń dla tej rodziny. Analiza haplotypów chromosomu Y wykazała pełną zgodność między profilami 92 a 246, wartość $LR = 189\,665$. Pomiedzy profilami 92 a 245 stwierdzono dwie niezgodności w układach DYS437, DYS570, a wyliczona wartość $LR = 225$. Pomiedzy profilami 92 a 247 stwierdzono dwie niezgodności w układach DYS627, DYF387S1, a wyliczona wartość $LR = 15$ (wyliczenia wykonano dla zestawu Y27, gdyż Y17 nie obejmuje niezgodnych układów).

Z uwagi na niezgodności w profilach Y-STR oznaczenie dla tej rodziny zostały rozszerzone o zestaw KinFinder. Obliczenia w programie Familias wykazały znaczny wzrost wartości LR do $2,38 \cdot 10^{10}$. Analiza w programie KinBN dała wartość $1,87 \cdot 10^{10}$. Dla Zestawu KinFinder nie udało się obliczyć sprzężeń w żadnym programie. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR $LR_c = 1,04 \cdot 10^{22}$.

Symulacje rozkładu LR (Tab. 13) wskazują, że uzyskane wyniki zarówno dla zestawu GlobalFiler™ jak i KinFinder mieszczą się w oczekiwanych przedziałach i znajdują się okolicy mediany. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (nawet uwzględniając niepełne oznaczenie profilu – 21 układów z 50), a najniższą Precision ID Identity.

Analiza trójkąta IBD (Rys. 46) oraz symulacji możliwych pokrewieństw (Rys. 47, Rys. 48) zdają się potwierdzać pokrewieństwo w relacji wnuki – dziadek pomiędzy analizowanymi osobami, jednakże można zauważyć różnice pomiędzy pokrewieństwem pomiędzy poszczególnymi wnukami. Po analizie uzyskanych wyników można jednak przypuszczać, że pokrewieństwo wskazane przez wywiad środowiskowy jest prawdopodobne.

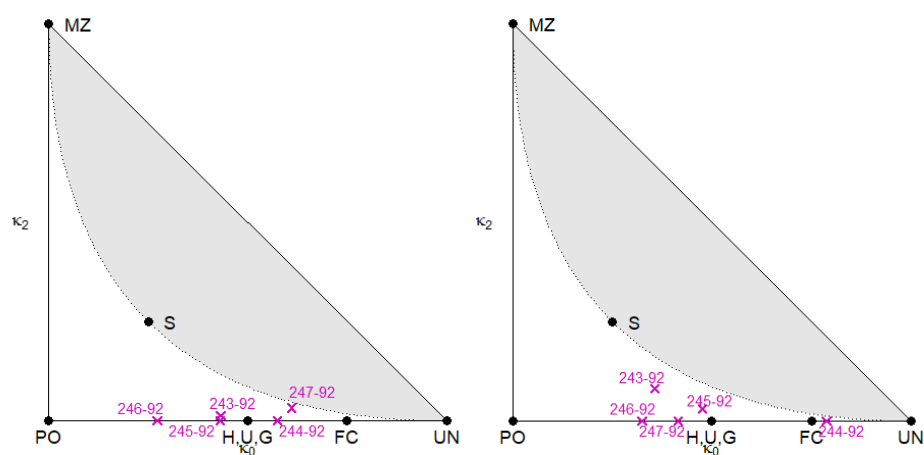


Rys. 45. Drzewo genealogiczne dla Rodziny 12

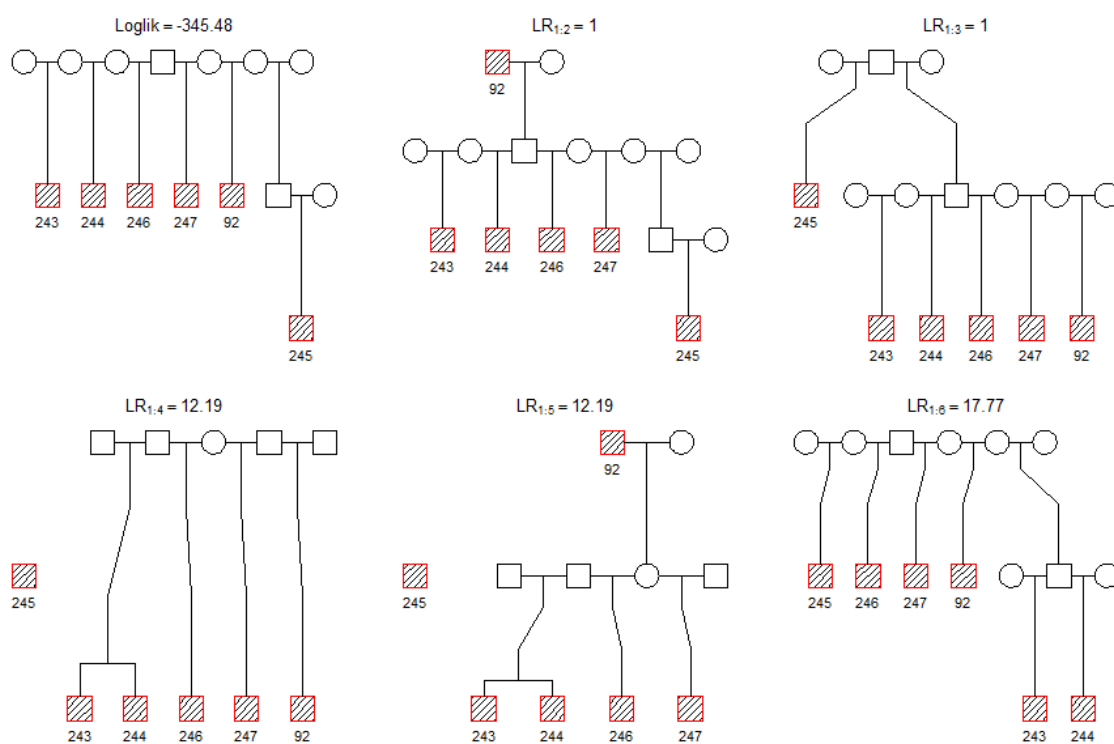
Tab. 13. Symulacja wyników analizy statystycznej dla Rodziny 12

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$7,51 \cdot 10^6$	$1,11 \cdot 10^{11}$	$5,03 \cdot 10^{10}$	1532	$1,73 \cdot 10^{12}$
	H 1	H 2 (TRUE)	0	0,0532	$2,99 \cdot 10^{-6}$	0	1,571
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$3,55 \cdot 10^7$	$8,59 \cdot 10^{12}$	$3,89 \cdot 10^{11}$	$1,2 \cdot 10^4$	$1,83 \cdot 10^{14}$
	H 1	H 2 (TRUE)	0	0,0006	$1,74 \cdot 10^{-7}$	0	0,0088
KinFinder	H 1 (TRUE)	H 2	$1,89 \cdot 10^{22}$	$3,86 \cdot 10^{33}$	$1,54 \cdot 10^{28}$	$1,75 \cdot 10^{16}$	$1,22 \cdot 10^{35}$
	H 1	H 2 (TRUE)	0	$2,12 \cdot 10^{-34}$	0	0	$6,73 \cdot 10^{-33}$
KinFinder (układy obecne w profilu szczątków nr 92)	H 1 (TRUE)	H 2	$6,26 \cdot 10^9$	$7,62 \cdot 10^{15}$	$9,95 \cdot 10^{13}$	$5,69 \cdot 10^5$	$1,98 \cdot 10^{17}$
	H 1	H 2 (TRUE)	0	$1,06 \cdot 10^{-7}$	0	0	$1,91 \cdot 10^{-6}$
Precision ID Identity	H 1 (TRUE)	H 2	$1,37 \cdot 10^4$	$3,33 \cdot 10^6$	$5,71 \cdot 10^6$	20,95	$3,26 \cdot 10^7$
	H 1	H 2 (TRUE)	$1,37 \cdot 10^{-5}$	0,6008	0,034	$3,94 \cdot 10^{-10}$	15,39

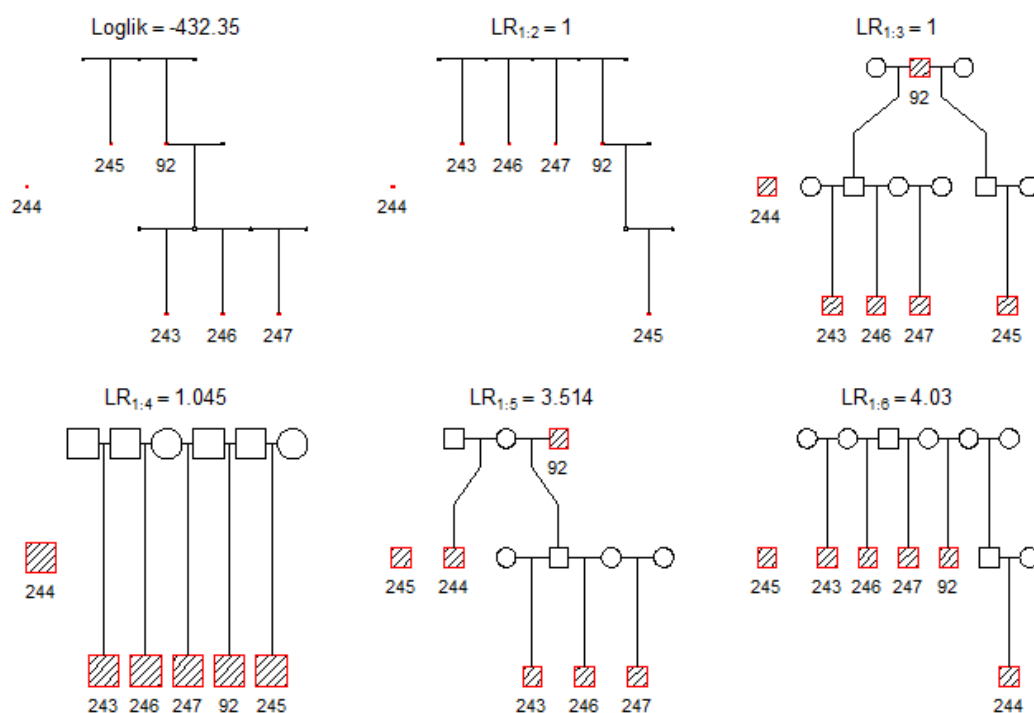
Źródło: opracowanie własne



Rys. 46. Wykres trójkąta IBD dla Rodziny 12 dla zestawu GlobalFiler™ i KinFinder



Rys. 47. Symulacje pokrewieństw dla Rodziny 12 dla zestawu GlobalFiler™



Rys. 48. Symulacje pokrewieństw dla Rodziny 12 dla zestawu KinFinder

4.13. Rodzina 13

Rodzina 13 obejmuje wnuka osoby zaginionej (osoba 250), relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 49).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerami 81, 86, 91, 95, 96 i 105. Analiza haplotypów chromosomu Y wykazała jedną niezgodność w układzie DYS456 profilu szkieletu 96 z profilem osoby 250, a wartością $LR = 788$.

Obliczenia statystyczne wykonane w programie Familias wykazały dla w/w dopasowania wartość $LR = 13$ taką samą zarówno uwzględniając sprzężenia jak i zakładając niezależność układów. Analiza w programie KinBN dała wartości 14 bez sprzężeń i ze sprzężeniami.

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestawy KinFinder oraz Precision ID Identity.

Dla zestawu KinFinder obliczenia w programie Familias $LR = 98$ (bez sprzężeń) i 110 (ze sprzężeniami), natomiast w KinBN $LR = 107$ (bez sprzężeń) oraz $LR = 120$ (ze sprzężeniami).

Dla zestawu Precision ID Identity obliczenia w programie Familias $LR = 3\,299$ (bez sprzężeń) oraz $LR = 3\,355$ (ze sprzężeniami), a w KinBN odpowiednio $LR = 3\,300$ i $LR = 2\,990$.

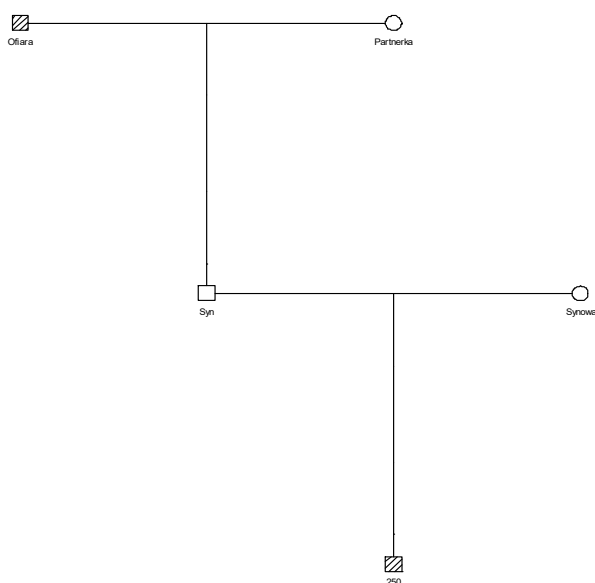
Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 3,31 \cdot 10^9$.

Symulacje rozkładu LR (Tab. 14) wskazują, że uzyskane wyniki dla zestawów GlobalFiler™ i KinFinder mieszczą się w oczekiwanych przedziałach. Zestaw GlobalFiler™ poniżej

mediany, natomiast KinFinder powyżej mediany ale poniżej średniej. Wynik uzyskany zestawem Precision ID Identity był wyższy niż wskazany w symulacjach.

Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (uwzględniając niepełne oznaczenie profilu – 9 układów z 50 zestaw uzyskał niższe wyniki niż GlobalFiler™), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 50) oraz symulacje możliwych pokrewieństw dla zestawów GlobalFiler™ i Precision ID Identity (Rys. 51 i Rys. 53) wykazują zgodność z zakładanym pokrewieństwem. Pedbuildr dla zestawu KinFinder (Rys. 52) wskazał zakładany stopień pokrewieństwa na drugiej pozycji, jako bardziej prawdopodobne uznał pokrewieństwo w relacji rodzeństwo.

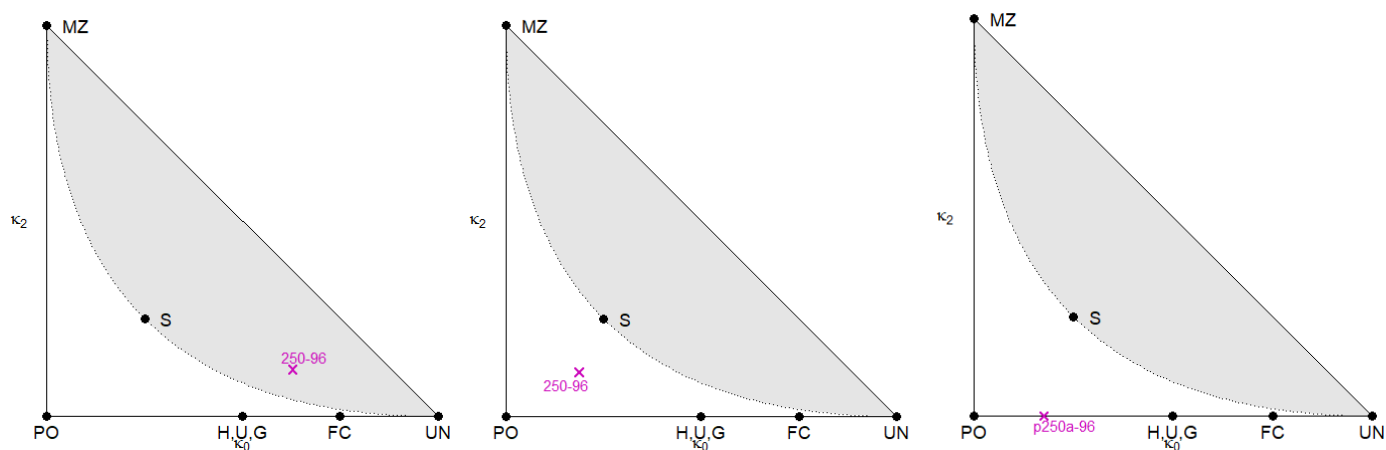


Rys. 49. Drzewo genealogiczne dla Rodziny 13

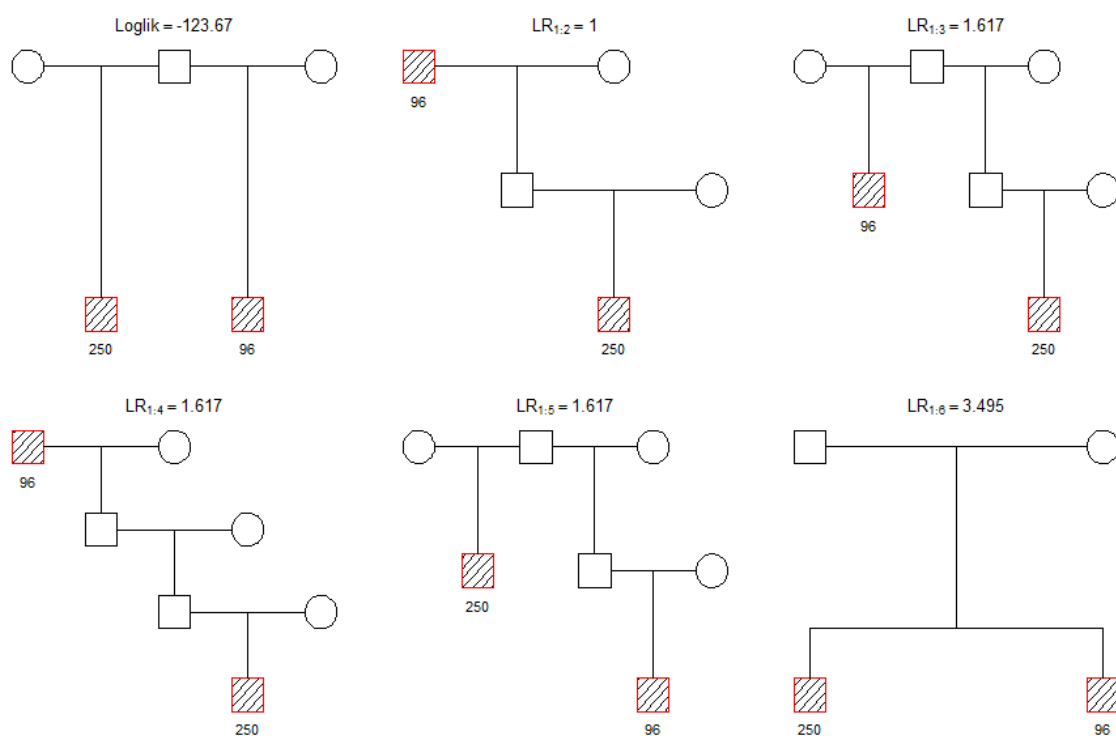
Tab. 14. Symulacja wyników analizy statystycznej dla Rodziny 13

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	72,2	$6,52 \cdot 10^4$	$1,22 \cdot 10^4$	0,4554	$1,57 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0208	0,5022	1,58	0,0005	3,014
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	94,34	$1,69 \cdot 10^4$	$1,86 \cdot 10^4$	0,4951	$1,48 \cdot 10^5$
	H 1	H 2 (TRUE)	0,0133	0,8851	1,303	0,0003	12,49
KinFinder	H 1 (TRUE)	H 2	$4,83 \cdot 10^5$	$1,28 \cdot 10^{11}$	$4,61 \cdot 10^9$	75,06	$2,67 \cdot 10^{12}$
	H 1	H 2 (TRUE)	$1,35 \cdot 10^{-5}$	0,235	0,02217	$1,32 \cdot 10^{-8}$	4,56
KinFinder (układy obecne w profilu szczątków nr 96)	H 1 (TRUE)	H 2	9,756	130,5	414,5	0,2554	758,4
	H 1	H 2 (TRUE)	0,111	1,779	2,972	0,0055	24,48
Precision ID Identity	H 1 (TRUE)	H 2	19,48	203,7	696,6	0,4025	1055
	H 1	H 2 (TRUE)	0,0374	1,116	2,277	0,0004	12,19

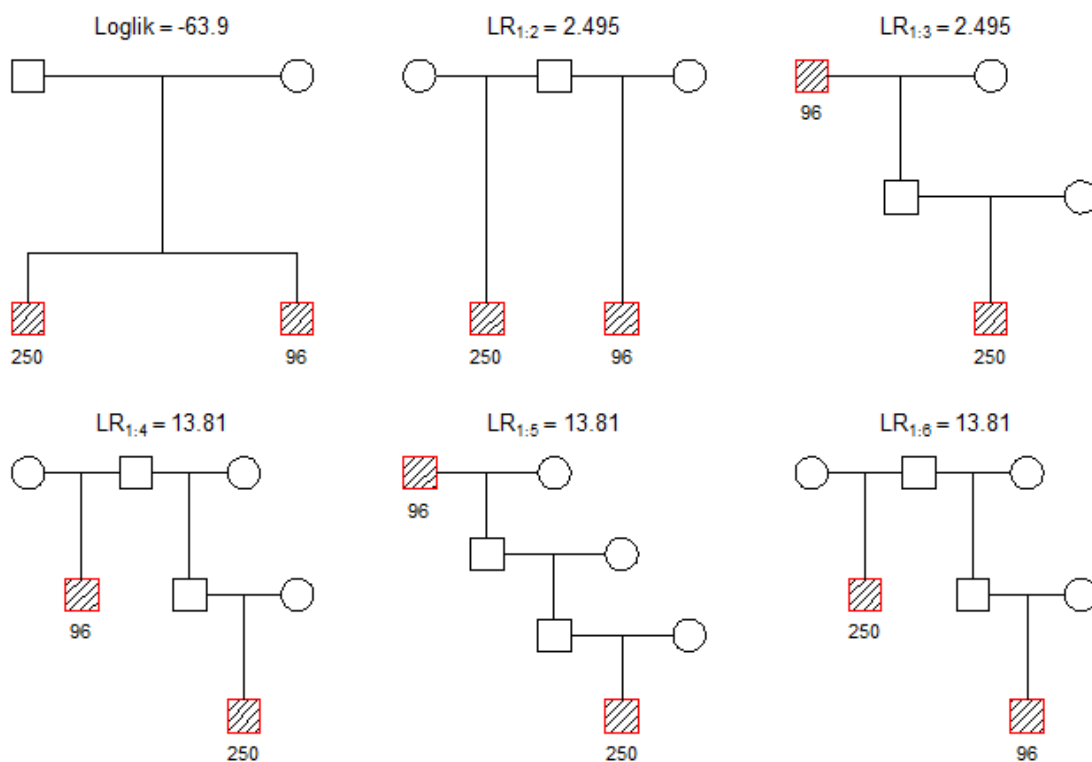
Źródło: opracowanie własne



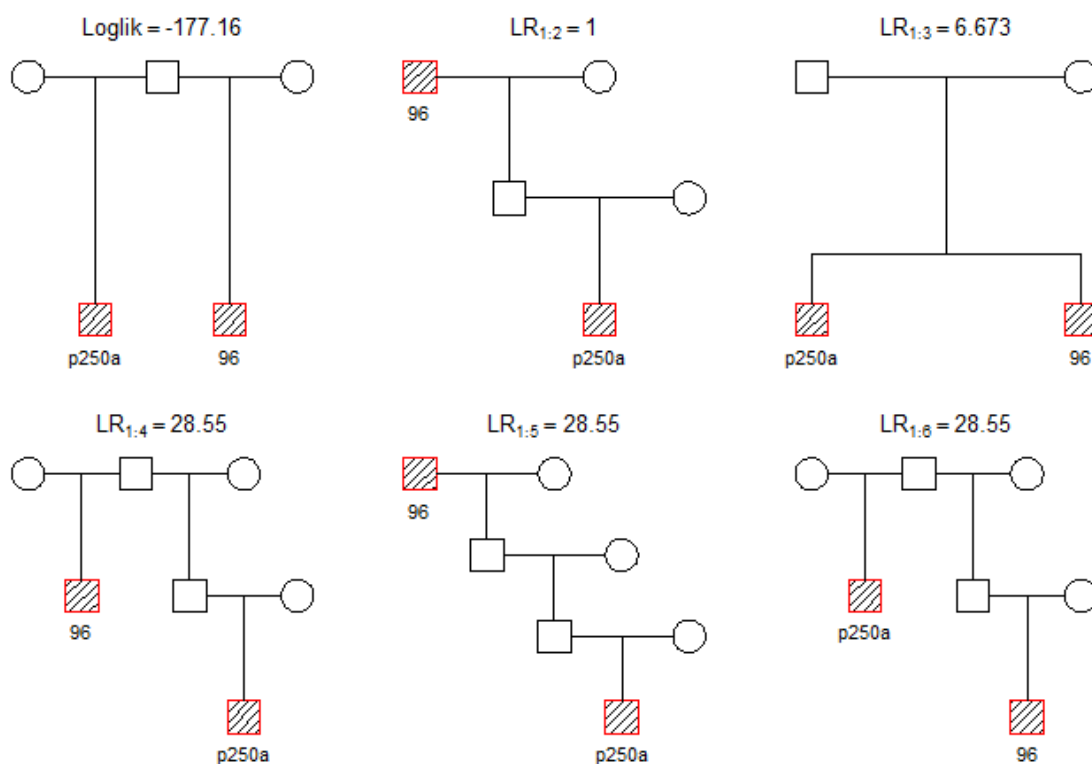
Rys. 50. Wykres trójkąta IBD dla Rodziny 13 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity



Rys. 51. Symulacje pokrewieństw dla Rodziny 13 dla zestawu GlobalFiler™



Rys. 52. Symulacje pokrewieństw dla Rodziny 13 dla zestawu KinFinder



Rys. 53. Symulacje pokrewieństw dla Rodziny 13 dla zestawu Precision ID Identity

4.14. Rodzina 14

Rodzina 14 obejmuje wnuka osoby zaginionej (osoba 801) oraz dwoje prawnuków: osoby 251 i 252, którzy są kuzynami, relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 54).

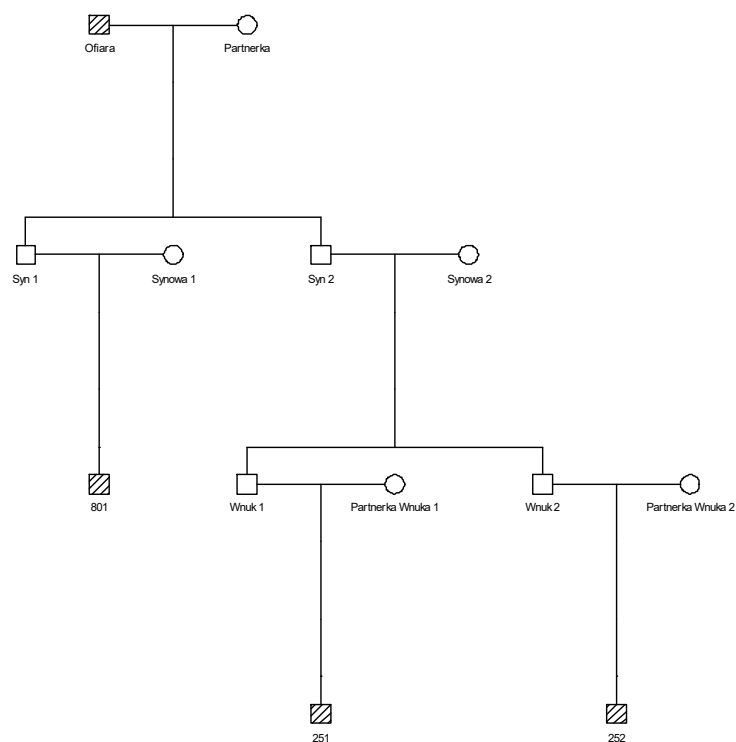
Po przeszukaniu DVI dopasowano szczątki oznaczone numerami 81 i 95. Analiza haplotypów chromosomu Y wykazała pełną zgodność między profilami 95, 801 i 251, wartość $LR = 190\,556$. Pomiedzy profilami 95 a 252 stwierdzono jedną niezgodności w układzie DYS570, a wyliczona wartość $LR = 1\,259$ (wyliczenia wykonano dla zestawu Y27, gdyż Y17 nie obejmuje niezgodnego układu).

Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 132$, która po uwzględnieniu sprzężeń genetycznych wzrosła do 135. Analiza w programie KinBN dała wartości odpowiednio 155 (bez sprzężeń) i 158 (ze sprzężeniami).

Oznaczenia profili STR autosomalnych zostały rozszerzone o zestaw KinFinder. Obliczenia w programie Familias wykazały wartość $LR = 98$, która po uwzględnieniu sprzężeń genetycznych wzrosła do 205. Analiza w programie KinBN dała wartość odpowiednio 107 (bez sprzężeń) i 225 (ze sprzężeniami). Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR $LR_c = 2,47 \cdot 10^9$.

Symulacje rozkładu LR (Tab. 15) wskazują, że uzyskane wyniki zarówno dla zestawu GlobalFiler™ jak i KinFinder mieszczą się w oczekiwanych przedziałach ale znajdują się poniżej mediany. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (uwzględniając niepełne oznaczenie profilu – 16 układów z 50 zestaw ten uzyskał niższe wartości niż GlobalFiler™), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 55) wykazuje na brak pokrewieństwa szkieletu z osobą 251, symulacje możliwych pokrewieństw dla zestawu GlobalFiler™ (Rys. 56) sugerują wykonanie obliczeń tylko z osobą 801, natomiast symulacje pokrewieństw zestawu KinFinder (Rys. 57) sugerują wykonanie obliczeń bez profilu 251. Wykonano obliczenia uwzględniające wskazania pedbuildr i uzyskano następujące wyniki: Zestaw GlobalFiler™ bez profilu 251 $LR = 630$, natomiast uwzględniając tylko profil 801 w obliczeniach uzyskano wartość 445. Dla zestawu KinFinder bez profilu 251 uzyskano $LR = 352$. Program pedbuildr wskazał pokrewieństwo które znacząco zwiększyło wartość LR, jednakże przy obliczeniach końcowych korzystano z pełnych danych środowiskowych odnośnie struktury pokrewieństwa w tej rodzinie.

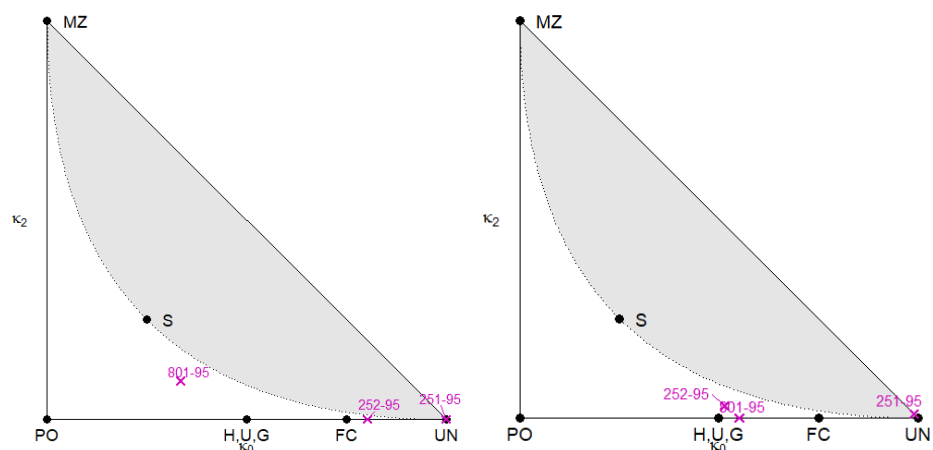


Rys. 54. Drzewo genealogiczne dla Rodziny 14

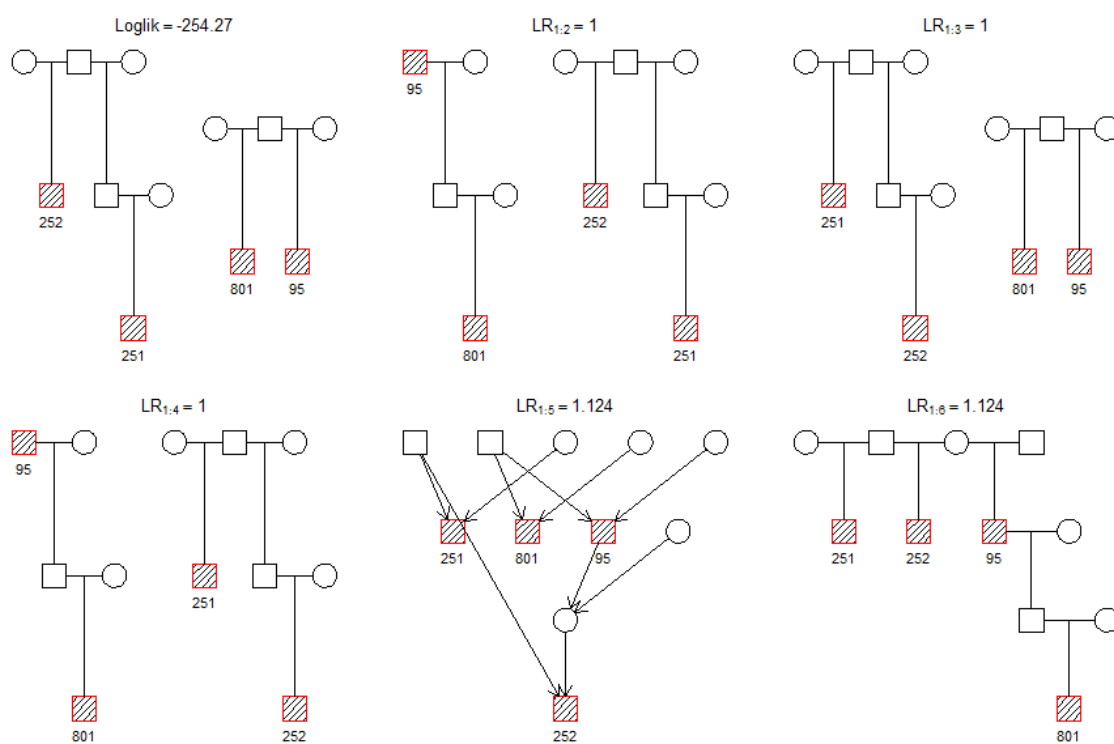
Tab. 15. Symulacja wyników analizy statystycznej dla Rodziny 14

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	406,8	$7,29 \cdot 10^5$	$1,49 \cdot 10^5$	1,337	$1,92 \cdot 10^7$
	H 1	H 2 (TRUE)	0,004	0,6031	0,9281	$3,28 \cdot 10^{-5}$	7,319
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	594,6	$4,62 \cdot 10^6$	$3,19 \cdot 10^5$	1,805	$7,29 \cdot 10^7$
	H 1	H 2 (TRUE)	0,0021	2,386	0,4013	$1,46 \cdot 10^{-5}$	47,96
KinFinder	H 1 (TRUE)	H 2	$1,62 \cdot 10^8$	$4,75 \cdot 10^{13}$	$6,19 \cdot 10^{12}$	5080	$8,9 \cdot 10^{14}$
	H 1	H 2 (TRUE)	$3,69 \cdot 10^{-8}$	0,0053	0,0002	$7,42 \cdot 10^{-12}$	0,1361
KinFinder (układy obecne w profilu szczątków nr 95)	H 1 (TRUE)	H 2	302	$2,07 \cdot 10^5$	$1,93 \cdot 10^5$	1,504	$2,73 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0038	0,6501	0,6605	$2,65 \cdot 10^{-5}$	10,25
Precision ID Identity	H 1 (TRUE)	H 2	54,43	1789	6036	0,5408	$1,16 \cdot 10^4$
	H 1	H 2 (TRUE)	0,0121	0,7705	1,289	$5,83 \cdot 10^{-5}$	8,478

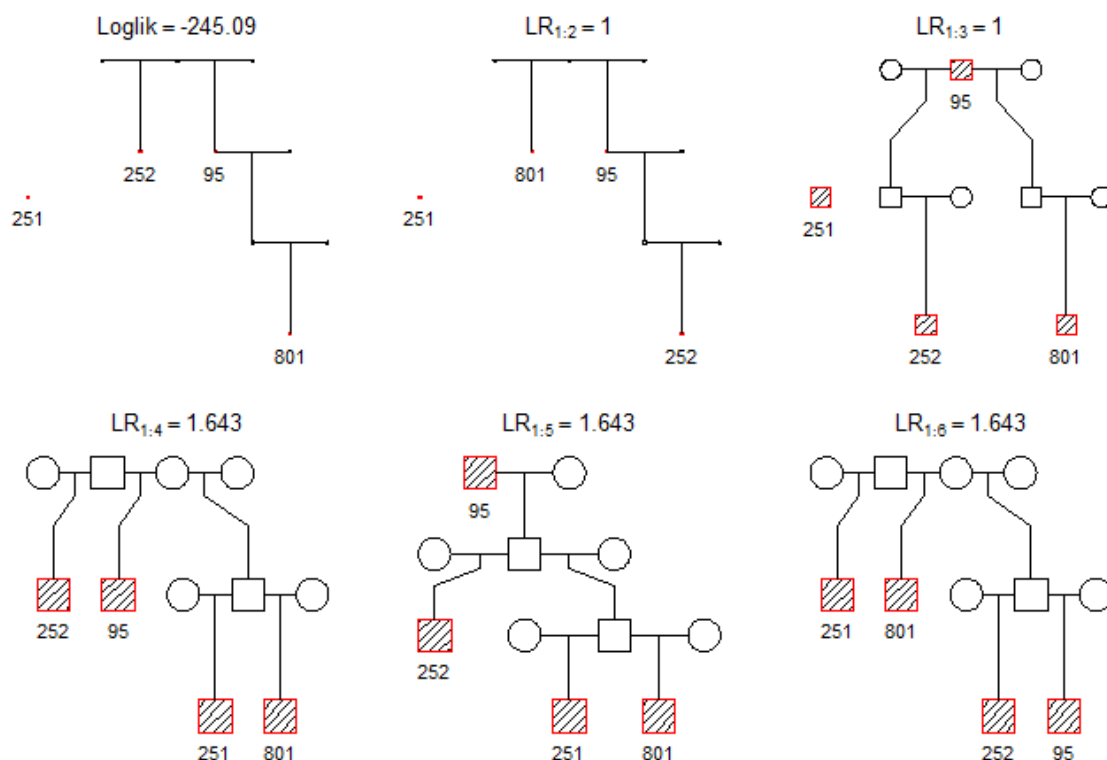
Źródło: opracowanie własne



Rys. 55. Wykres trójkąta IBD dla Rodziny 14 dla zestawu GlobalFiler™ i KinFinder



Rys. 56. Symulacje pokrewieństw dla Rodziny 14 dla zestawu GlobalFiler™



Rys. 57. Symulacje pokrewieństw dla Rodziny 14 dla zestawu KinFinder

4.15. Rodzina 15

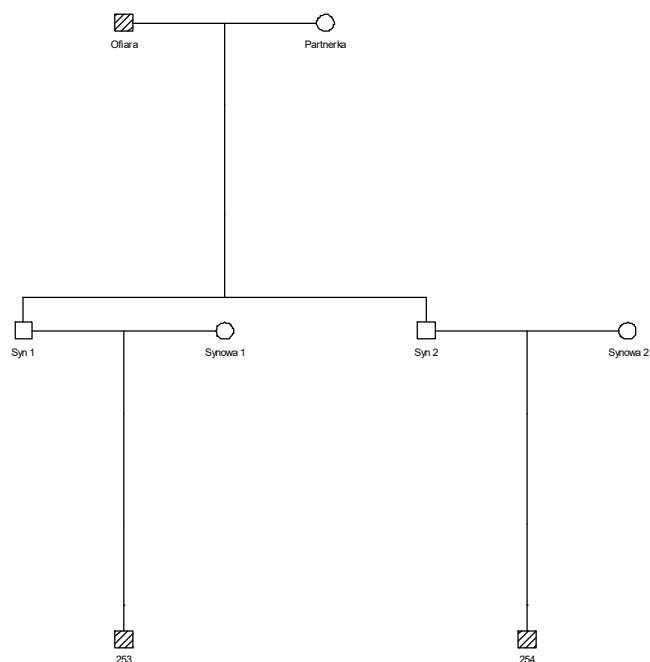
Rodzina 15 obejmuje dwóch wnuków (osoby 253 i 254), którzy są kuzynami, relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 58).

Po przeszukaniu DVI nie stwierdzono żadnego dopasowania. Również przeszukanie profili Y-STR nie przyniosło pozytywnych wyników.

Z uwagi na otrzymane wyniki oraz fakt, iż rodzina 15 poszukuje osoby spoza listy egzekucyjnej, nie rozszerzano badań dla tej rodziny.

W celu oceny przydatności zastosowanych zestawów, wykonano symulacje wyników analizy statystycznej możliwych do uzyskania przy analizowanym stopniu pokrewieństwa (Tab. 16).

Wyniki symulacji wskazują, że oczekiwana wartość LR powinna mieścić się w przedziale od około 2 do $2,45 \cdot 10^6$ dla GlobalFiler™, od $3,68 \cdot 10^5$ do $1,48 \cdot 10^{16}$ dla KinFinder oraz od 1 do $3,21 \cdot 10^4$ dla Precision ID Identity.



Rys. 58. Drzewo genealogiczne dla Rodziny 15

Tab. 16. Symulacja wyników analizy statystycznej dla Rodziny 15

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	2835	$6,42 \cdot 10^7$	$2,45 \cdot 10^6$	2,548	$1,96 \cdot 10^9$
	H 1	H 2 (TRUE)	0,0006	0,5286	0,2936	$2,17 \cdot 10^{-6}$	7,915
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	5070	$1,31 \cdot 10^7$	$1,14 \cdot 10^7$	5,434	$1,41 \cdot 10^8$
	H 1	H 2 (TRUE)	0,0003	0,1566	0,1789	$4,3 \cdot 10^{-7}$	1,724
KinFinder	H 1 (TRUE)	H 2	$5,72 \cdot 10^{10}$	$1,83 \cdot 10^{18}$	$1,48 \cdot 10^{16}$	$3,68 \cdot 10^5$	$5,41 \cdot 10^{19}$
	H 1	H 2 (TRUE)	$1,45 \cdot 10^{-10}$	0,0001	$3,77 \cdot 10^{-6}$	$4,39 \cdot 10^{-15}$	0,0023
Precision ID Identity	H 1 (TRUE)	H 2	190,9	$1,522 \cdot 10^4$	$3,21 \cdot 10^4$	0,7689	$1,21 \cdot 10^5$
	H 1	H 2 (TRUE)	0,0024	0,6973	1,081	$5,39 \cdot 10^{-6}$	7,47

Źródło: opracowanie własne

4.16. Rodzina 16

Rodzina 16 obejmuje wnuka osoby zaginionej (osoba 258), relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 59).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerami 88 i 91. Analiza haplotypów chromosomu Y wykazała pełną zgodność profilu szczątków 91 z profilem osoby 258, a wartością $LR = 189\,665$.

Obliczenia statystyczne wykonane w programie Familias wykazały dla w/w dopasowania wartość $LR = 18$ (bez sprzężeń) oraz $LR = 13$ (ze sprzężeniami). Analiza w programie KinBN dała te same wartości.

W dalszej kolejności oznaczenia autosomalnych profili STR zostały rozszerzone o zestawy KinFinder oraz Precision ID Identity.

Dla zestawu KinFinder obliczenia w programie Familias LR = 318 (bez sprzężeń) i 224 (ze sprzężeniami), natomiast w KinBN LR = 323 (bez sprzężeń) oraz LR = 227 (ze sprzężeniami).

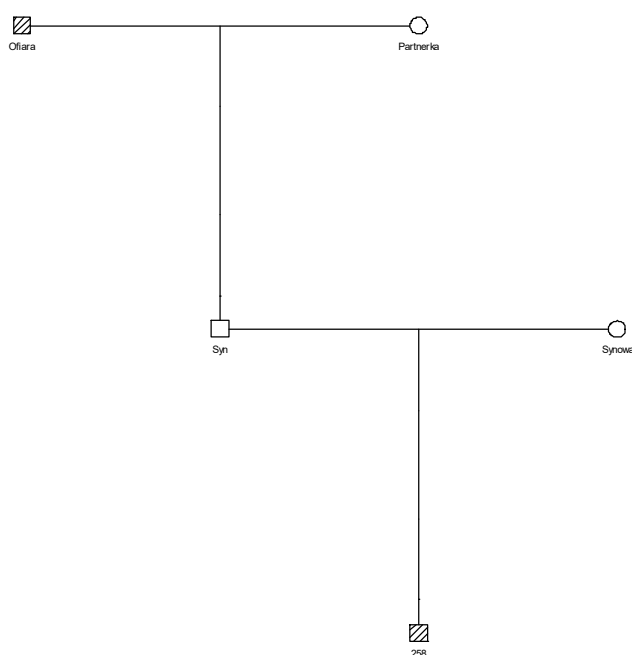
Dla zestawu Precision ID Identity obliczenia w programie Familias LR = 5 094 (bez sprzężeń) oraz LR = 3 652 (ze sprzężeniami), a w KinBN odpowiednio LR = 5 090 i LR = 3 920.

Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 5,53 \cdot 10^{12}$.

Symulacje rozkładu LR (Tab. 17) wskazują, że uzyskane wyniki dla zestawów GlobalFiler™ i KinFinder mieszczą się w oczekiwanych przedziałach. Zestaw GlobalFiler™ poniżej mediany, natomiast KinFinder powyżej mediany ale poniżej średniej. Wynik uzyskany zestawem Precision ID Identity był wyższy niż wskazany w symulacjach.

Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (nawet uwzględniając niepełne oznaczenie profilu – 18 układów z 50), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 60) oraz symulacje możliwych pokrewieństw dla zestawów KinFinder i Precision ID Identity (Rys. 62 i Rys. 63) wykazują zgodność z zakładanym pokrewieństwem. Pedbuildr dla zestawu GlobalFiler™ (Rys. 61) wskazał zakładany stopień pokrewieństwa na drugiej pozycji, jako bardziej prawdopodobne uznał pokrewieństwo w relacji wujek - syn brata przyrodniego lub prawnuk.

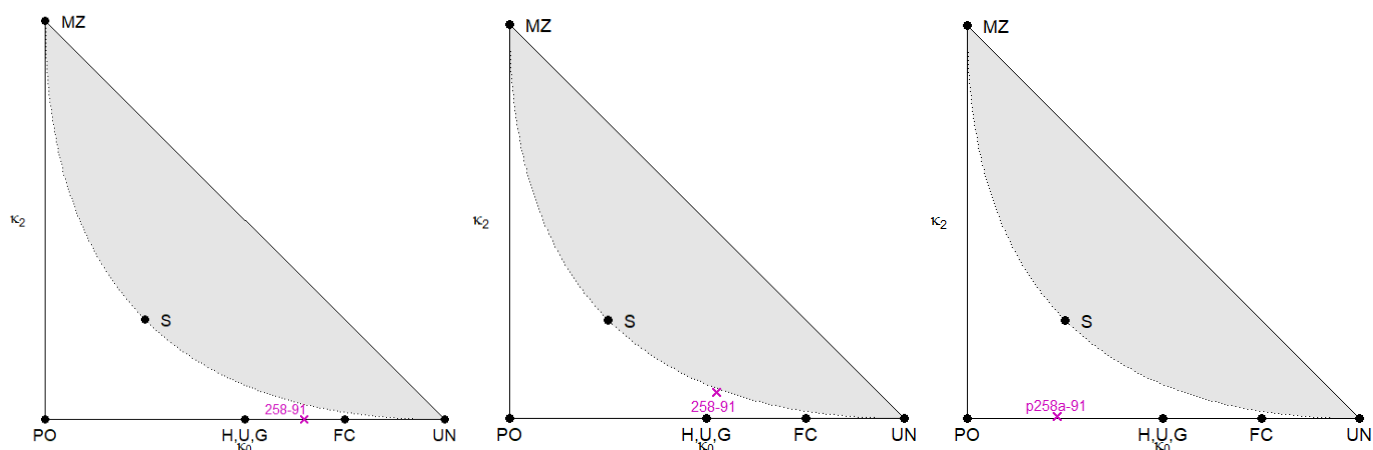


Rys. 59. Drzewo genealogiczne dla Rodziny 16

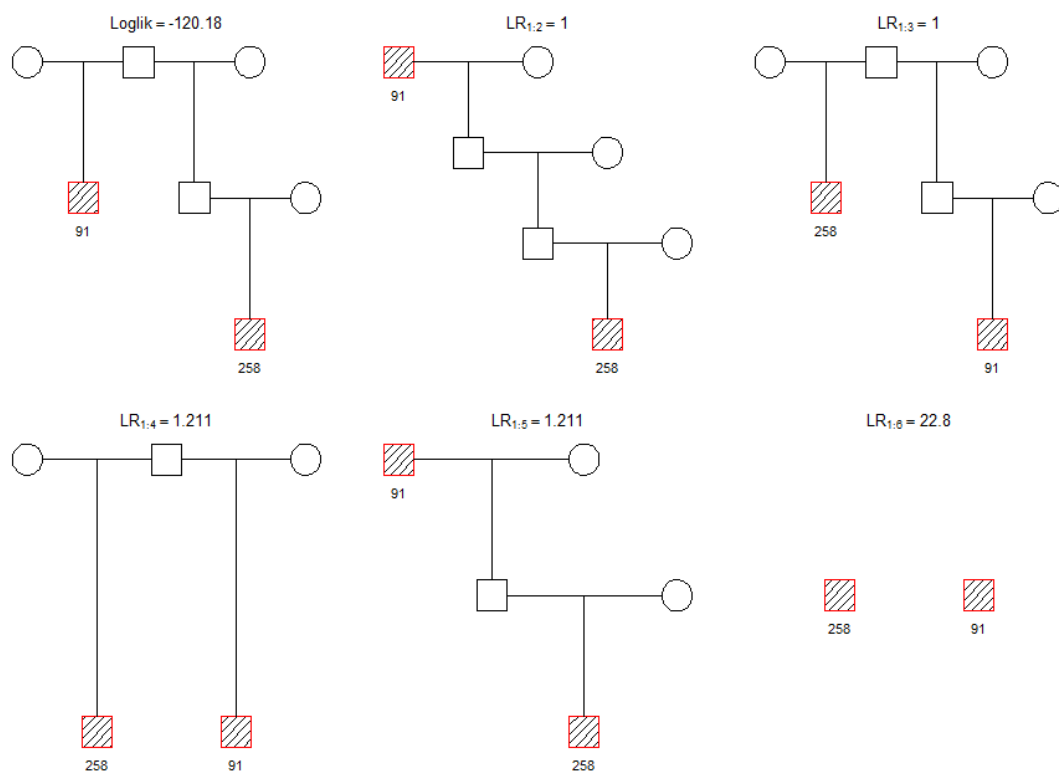
Tab. 17. Symulacja wyników analizy statystycznej dla Rodziny 16

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	49,38	$1,93 \cdot 10^4$	$1,39 \cdot 10^4$	0,5423	$3,29 \cdot 10^5$
	H 1	H 2 (TRUE)	0,0281	2,991	1,667	0,0005	78,52
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	76,47	$7,59 \cdot 10^4$	$2,7 \cdot 10^4$	0,4351	$1,99 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0162	1,424	1,655	0,0002	23,49
KinFinder	H 1 (TRUE)	H 2	$3,39 \cdot 10^5$	$4,11 \cdot 10^9$	$5,23 \cdot 10^9$	58,49	$4,55 \cdot 10^{10}$
	H 1	H 2 (TRUE)	$9,29 \cdot 10^{-6}$	0,03835	0,01757	$1,42 \cdot 10^{-8}$	0,5938
KinFinder (układy obecne w profilu szczątków nr 91)	H 1 (TRUE)	H 2	156,5	$4,79 \cdot 10^4$	$4,69 \cdot 10^4$	0,8415	$1,08 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0133	0,6266	0,9844	0,0002	6,58
Precision ID Identity	H 1 (TRUE)	H 2	21,5	259,6	909	0,3528	1409
	H 1	H 2 (TRUE)	0,0472	0,647	2,538	0,0006	2,705

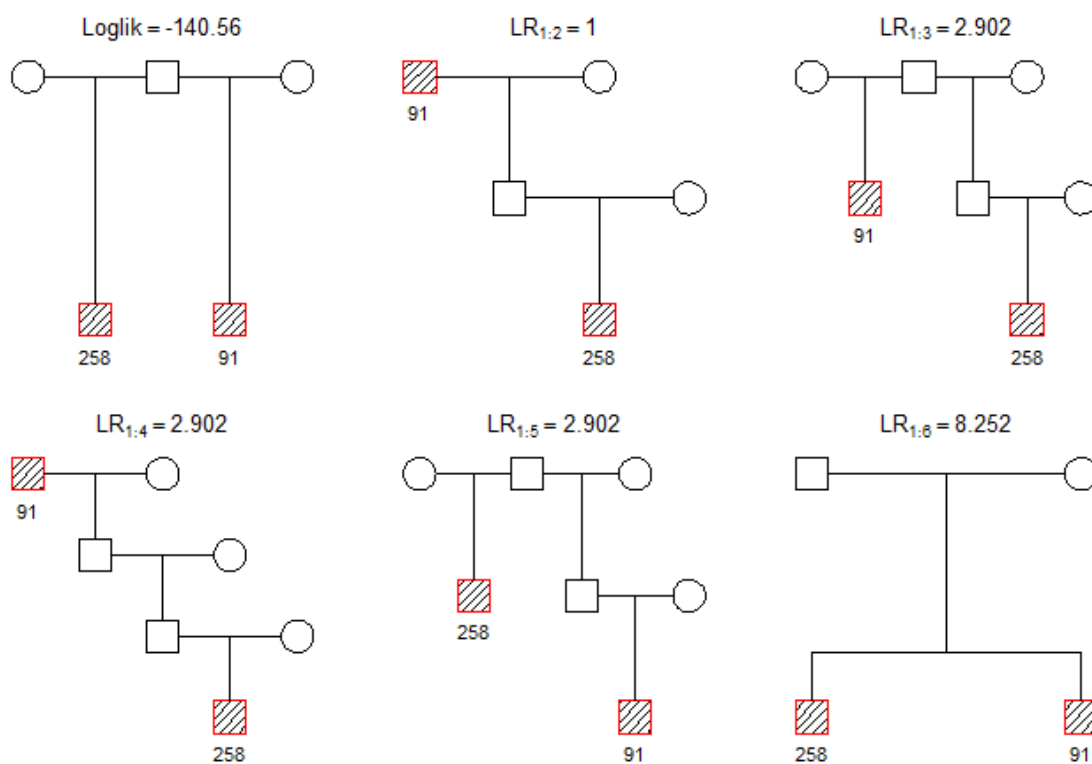
Źródło: opracowanie własne



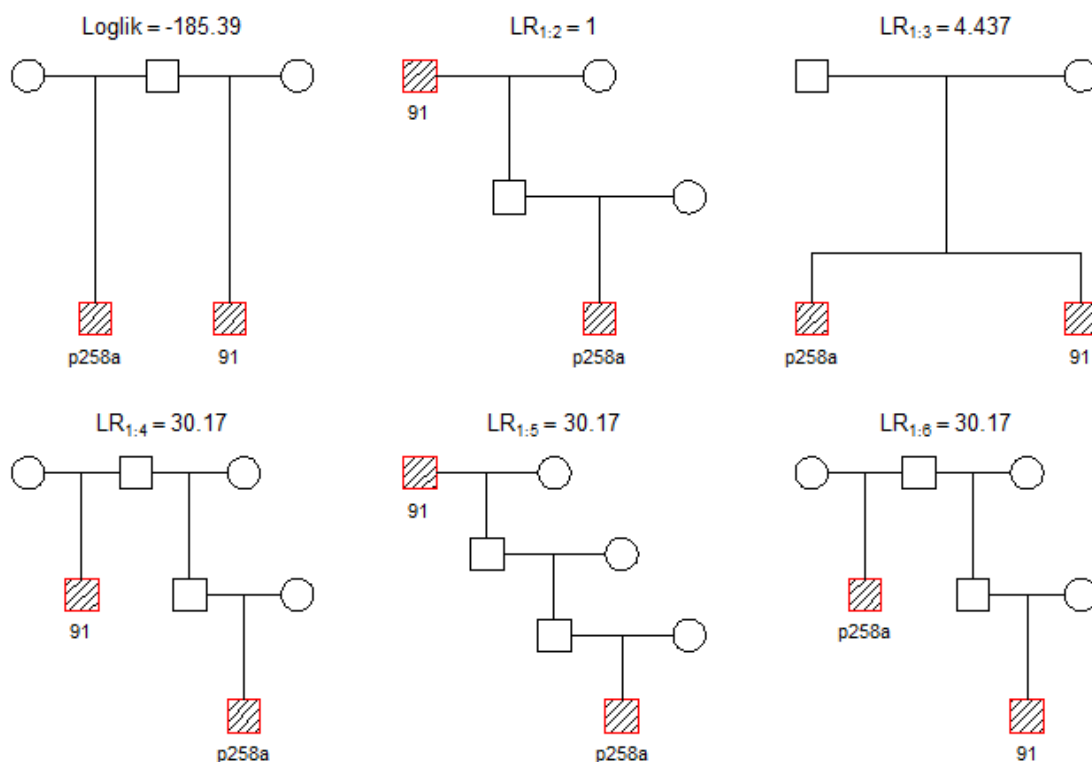
Rys. 60. Wykres trójkąta IBD dla Rodziny 16 dla zestawu GlobalFiler™, KinFinder i Precision ID Identity



Rys. 61. Symulacje pokrewieństw dla Rodziny 16 dla zestawu GlobalFiler™



Rys. 62. Symulacje pokrewieństw dla Rodziny 16 dla zestawu KinFinder



Rys. 63. Symulacje pokrewieństw dla Rodziny 16 dla zestawu Precision ID Identity

4.17. Rodzina 17

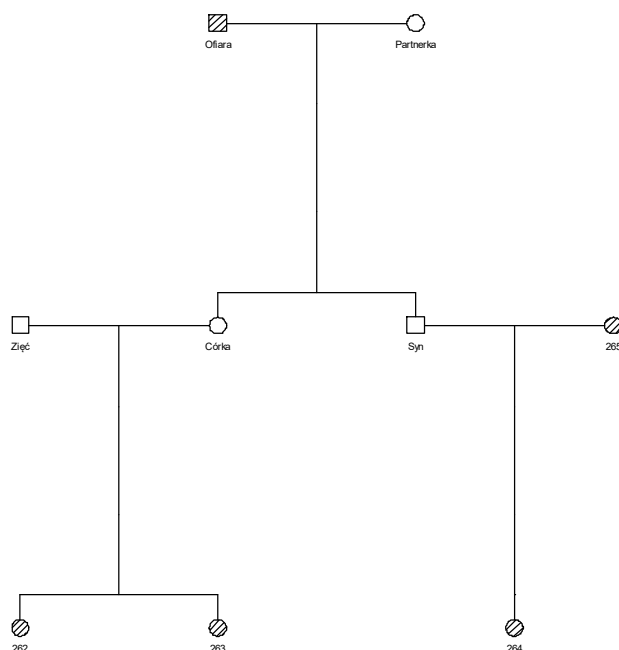
Rodzina 17 obejmuje dwie wnuczki osoby zaginionej (osoby 262 i 263), będące córkami córki osoby zaginionej, wnuka (osoba 264) będącego synem syna osoby zaginionej oraz osobę 265 – matkę osoby 264, relacje przedstawiono w formie drzewa genealogicznego poniżej (Rys. 64). Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 105. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 82\,155$, która po uwzględnieniu sprzężeń genetycznych zmalała do $72\,896$. Analiza w programie KinBN dała wartości odpowiednio $89\,900$ (bez sprzężeń) i $80\,100$ (ze sprzężeniami).

Oznaczenia profili STR autosomalnych zostały rozszerzone o zestaw KinFinder. Obliczenia w programie Familias wykazały wartość $LR = 804\,697$ (program FamLink nie wykonał obliczeń sprzężeń dla tej rodziny). Analiza w programie KinBN dała wartość odpowiednio $1,06 \cdot 10^6$ (bez sprzężeń) i $8,26 \cdot 10^6$ (ze sprzężeniami). Obliczono łączny iloraz wiarygodności dla STR autosomalnych $LR_c = 6,61 \cdot 10^{10}$.

Symulacje rozkładu LR (Tab. 18) wskazują, że uzyskane wyniki dla zestawu GlobalFiler™ mieszczą się w oczekiwanych przedziałach ale znajdują się poniżej mediany, natomiast wyniki uzyskane dla zestawu KinFinder znajdują się poniżej 5%. Najwyższą moc dowodową

w symulacji wyników wykazał zestaw KinFinder (nawet uwzględniając niepełne oznaczenie profilu – 30 układów z 50), a najniższą Precision ID Identity.

Analiza trójkątów IBD (Rys. 65) oraz symulacje możliwych pokrewieństw dla obu zestawów (Rys. 66 i Rys. 67) wykazały zgodność z zakładanym pokrewieństwem (za wyjątkiem pokrewieństwa pomiędzy osobami 262 i 263 dla których wskazano możliwość pokrewieństwa w relacji rodzeństwa przyrodniego).

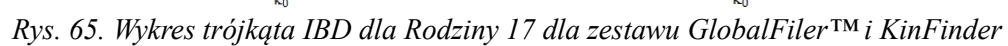


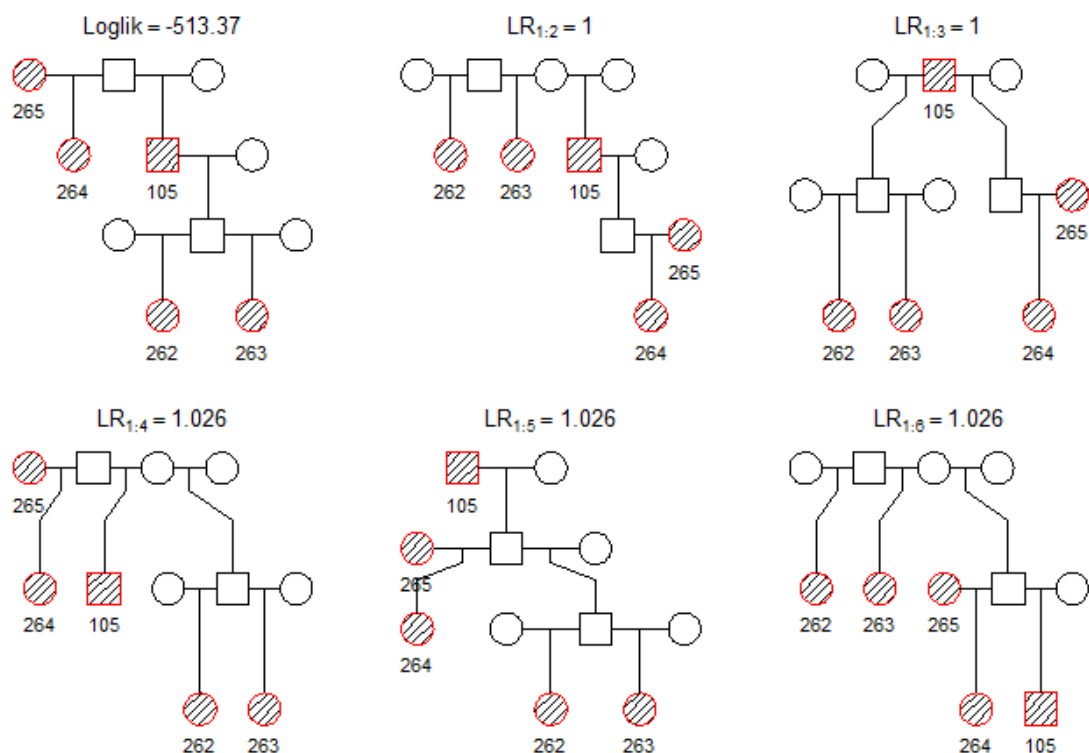
Rys. 64. Drzewo genealogiczne dla Rodziny 17

Tab. 18. Symulacja wyników analizy statystycznej dla Rodziny 17

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$2,46 \cdot 10^5$	$6,02 \cdot 10^9$	$6,45 \cdot 10^8$	72,19	$9,65 \cdot 10^{10}$
	H 1	H 2 (TRUE)	$1,79 \cdot 10^{-10}$	0,2781	0,0098	0	7,743
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$6,21 \cdot 10^5$	$1,94 \cdot 10^{10}$	$3,81 \cdot 10^9$	203,1	$3,1 \cdot 10^{11}$
	H 1	H 2 (TRUE)	0	0,0136	0,0025	0	0,1674
KinFinder	H 1 (TRUE)	H 2	$6,85 \cdot 10^{16}$	$2,48 \cdot 10^{27}$	$4,81 \cdot 10^{22}$	$1,78 \cdot 10^{11}$	$7,68 \cdot 10^{28}$
	H 1	H 2 (TRUE)	0	$4,02 \cdot 10^{-8}$	$4,13 \cdot 10^{-23}$	0	$1,26 \cdot 10^{-6}$
KinFinder (układy obecne w profilu szczątków nr 105)	H 1 (TRUE)	H 2	$6,86 \cdot 10^{10}$	$4,77 \cdot 10^{17}$	$2,9 \cdot 10^{15}$	$1,72 \cdot 10^6$	$1,41 \cdot 10^{19}$
	H 1	H 2 (TRUE)	0	$7,53 \cdot 10^{-5}$	$5,53 \cdot 10^{-9}$	0	0,0017
Precision ID Identity	H 1 (TRUE)	H 2	1612	$4,55 \cdot 10^5$	$8,07 \cdot 10^5$	3,714	$4,53 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0002	0,4265	0,2802	$5,33 \cdot 10^{-8}$	4,537

Źródło : opracowanie własne





Rys. 67. Symulacje pokrewieństw dla Rodziny 17 dla zestawu KinFinder

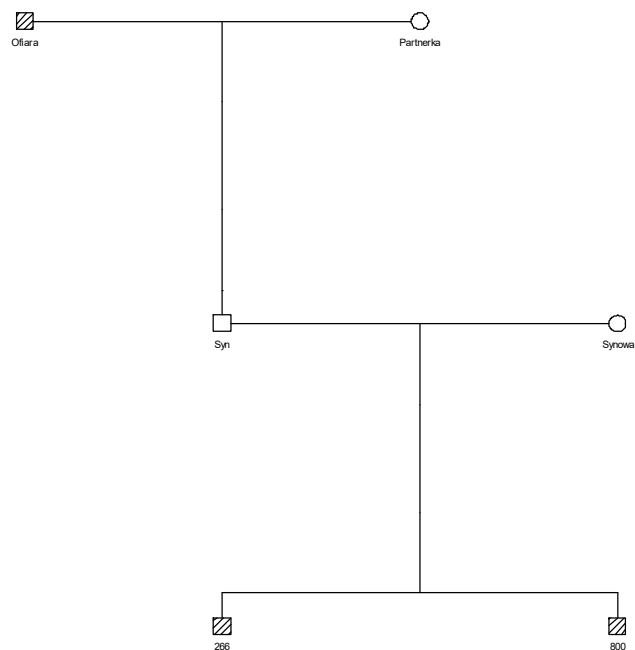
4.18. Rodzina 18

W skład 18 rodziny wchodzi dwóch wnuków osoby zaginionej (osoby 266 i 800), relacje między nimi przedstawiono poniżej (Rys. 68).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 84. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 134\,502$, która po uwzględnieniu sprzężeń genetycznych wzrosła do $161\,779$. Analiza w programie KinBN dała wartości odpowiednio $141\,000$ (bez sprzężeń) i $169\,000$ (ze sprzężeniami). Analiza haplotypów chromosomu Y wykazała pełną zgodność a wyliczona wartość $LR = 189\,665$. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 2,55 \cdot 10^{10}$.

Symulacje rozkładu LR (Tab. 19) wskazują, że uzyskane wyniki mieszczą się w oczekiwanych przedziałach, znajdują się nieco poniżej progu 95% wyników dla zestawu GlobalFiler™. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder, a najniższą Precision ID Identity.

Analiza trójkąta IBD (Rys. 69) oraz symulacji możliwych pokrewieństw (Rys. 70) wykazały zgodność z zakładanym pokrewieństwem.

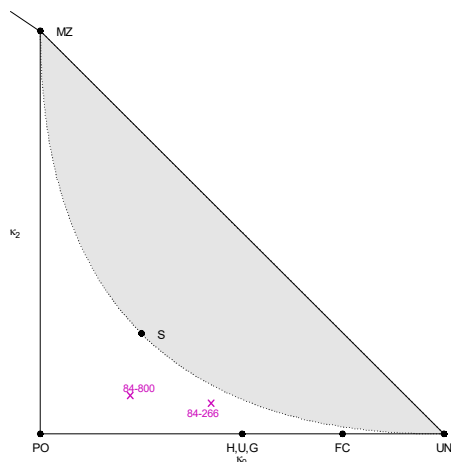


Rys. 68. Drzewo genealogiczne dla Rodziny 18

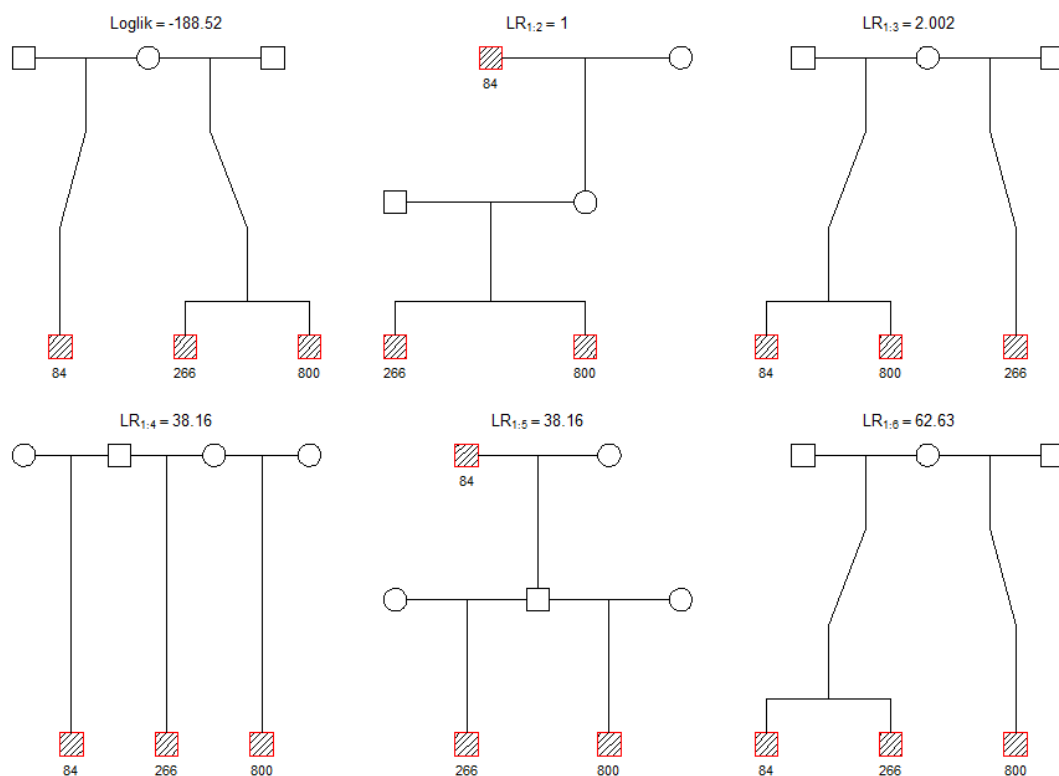
Tab. 19. Symulacja wyników analizy statystycznej dla Rodziny 18

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	768,3	$5,04 \cdot 10^5$	$2,29 \cdot 10^5$	2,297	$8,02 \cdot 10^6$
	H 1	H 2 (TRUE)	$9,67 \cdot 10^{-6}$	0,1833	0,3081	0	1,888
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	1665	$4,26 \cdot 10^5$	$6,23 \cdot 10^5$	5,034	$5,17 \cdot 10^6$
	H 1	H 2 (TRUE)	0	0,379	0,3357	0	6,26
KinFinder	H 1 (TRUE)	H 2	$2,34 \cdot 10^9$	$3,08 \cdot 10^{14}$	$3,34 \cdot 10^{13}$	$6,44 \cdot 10^4$	$5,06 \cdot 10^{15}$
	H 1	H 2 (TRUE)	0	0,0868	$5,02 \cdot 10^{-11}$	0	2,744
Precision ID Identity	H 1 (TRUE)	H 2	60,47	1553	4722	0,6415	$1,59 \cdot 10^4$
	H 1	H 2 (TRUE)	0,0091	1,21	1,942	$3,89 \cdot 10^{-5}$	9,789

Źródło: opracowanie własne



Rys. 69. Wykres trójkąta IBD dla Rodziny 18



Rys. 70. Symulacje pokrewieństw dla Rodziny 18

4.19. Rodzina 19

W skład 19 rodziny wchodzi trzech wnuków (osoby 122, 802 i 803) którzy są kuzynami, relacje między nimi przedstawiono poniżej (Rys. 71).

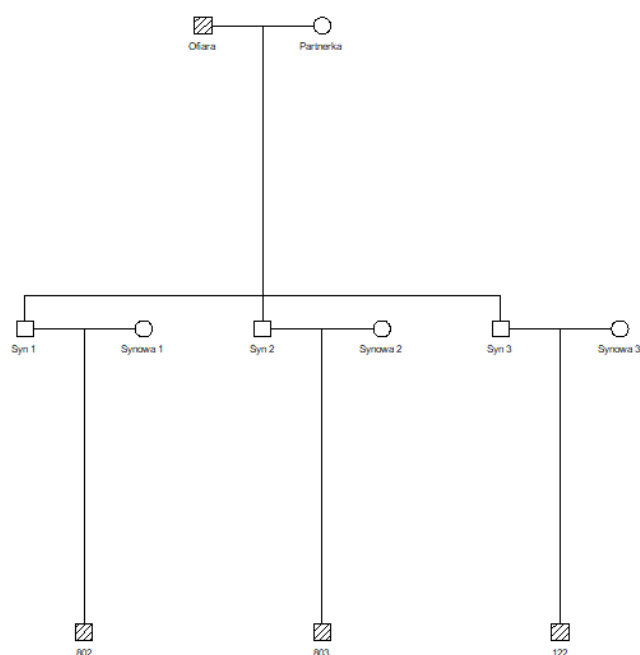
Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 83. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość LR = 56 021, która po uwzględnieniu sprzężeń genetycznych zmalała do 42 301. Równoległa analiza w programie KinBN dała wartości odpowiednio 63 400 (bez sprzężeń) i 47 900 (ze sprzężeniami). Analiza haplotypów chromosomu Y wykazała pełną zgodność i wartość LR = 189 665.

Oznaczenia profili STR autosomalnych zostały rozszerzone o zestaw KinFinder. Obliczenia w programie Familias wykazały wartość LR = 214, która po uwzględnieniu sprzężeń genetycznych spadła do 24. Analiza w programie KinBN dała wartość odpowiednio 1 (bez sprzężeń) i 0,5 (ze sprzężeniami).

Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 2,27 \cdot 10^{12}$.

Symulacje rozkładu LR (Tab. 20) wskazują, że uzyskane wyniki mieszczą się w oczekiwanych przedziałach i znajdują się nieco powyżej mediany dla zestawu GlobalFiler™. Dla zestawu KinFinder uzyskany wynik znajduje się poniżej wartości dla 5% (jednakże powyżej 1%). Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (nawet

uwzględniając niepełne oznaczenie profilu – 21 układów z 50), a najniższą Precision ID Identity. Analiza trójkąta IBD (Rys. 72) oraz symulacje alternatywnych struktur rodowodowych (Rys. 73) w pakiecie *pedbuildr* dla zestawu GlobalFiler™ wykazały zgodność pokrewieństwa pomiędzy profilami 83 i 122, natomiast w przypadku profili 83 i 802 oraz 83 i 803 sugerowany był odmienny stopień relacji (rodzeństwo, prawnukowie ofiary). Natomiast symulacje dla zestawu KinFinder (Rys. 74) wykazały za najbardziej prawdopodobną relację pomiędzy osobami 802 i 803 jako rodzeństwo, wnuki ofiary, natomiast wskazały na brak pokrewieństwa pomiędzy osobą 122 a ofiarą.

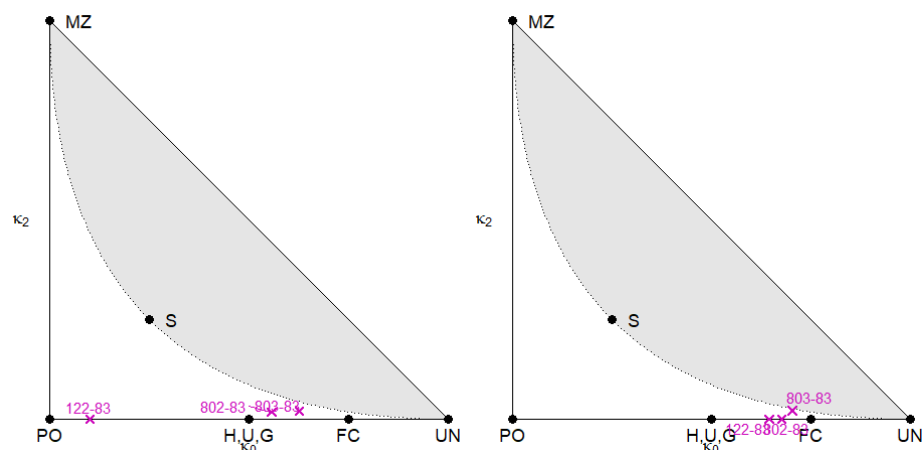


Rys. 71. Drzewo genealogiczne dla Rodziny 19

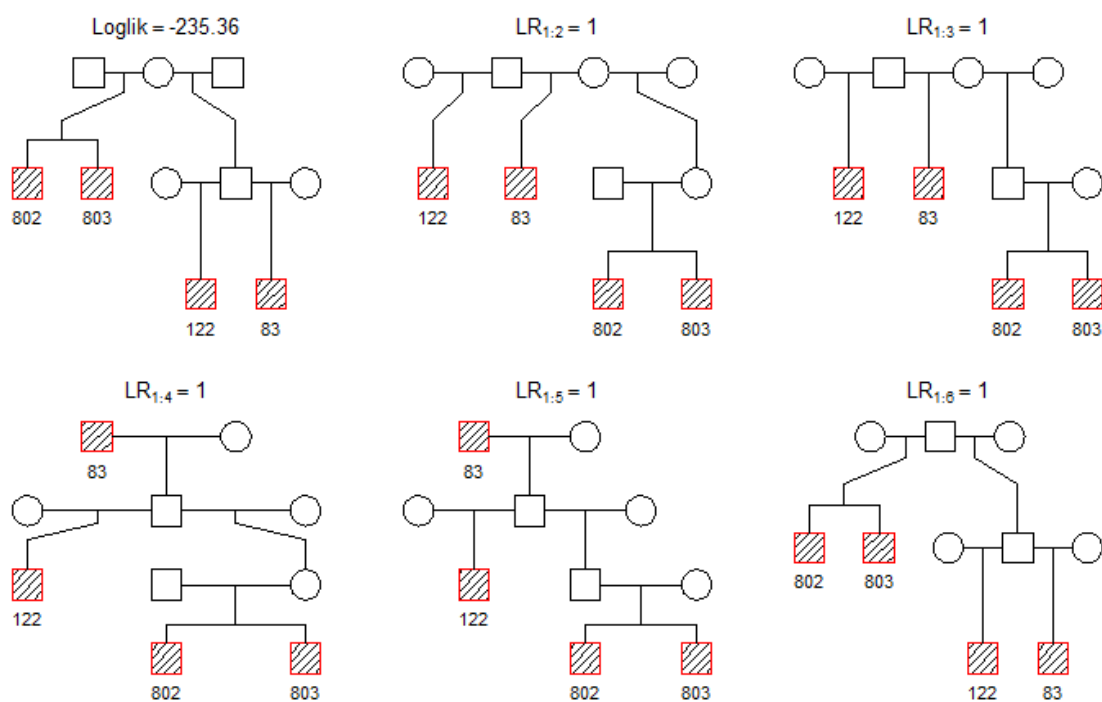
Tab. 20. Symulacja wyników analizy statystycznej dla Rodziny 19

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$5,21 \cdot 10^4$	$2,99 \cdot 10^8$	$1,92 \cdot 10^8$	28,46	$3,92 \cdot 10^9$
	H 1	H 2 (TRUE)	$9,89 \cdot 10^{-6}$	0,1745	0,0536	0	3,285
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$2,09 \cdot 10^5$	$4,91 \cdot 10^9$	$8,63 \cdot 10^8$	89,6	$8,98 \cdot 10^{10}$
	H 1	H 2 (TRUE)	$2,16 \cdot 10^{-6}$	0,017	0,012	0	0,1871
KinFinder	H 1 (TRUE)	H 2	$1,05 \cdot 10^{15}$	$8,65 \cdot 10^{26}$	$1,24 \cdot 10^{21}$	$2,21 \cdot 10^9$	$2,73 \cdot 10^{28}$
	H 1	H 2 (TRUE)	0	$3,07 \cdot 10^{-6}$	$6,87 \cdot 10^{-11}$	0	$9,59 \cdot 10^{-5}$
KinFinder (układy obecne w profilu szczątków nr 83)	H 1 (TRUE)	H 2	$4,07 \cdot 10^6$	$1 \cdot 10^{12}$	$2,3 \cdot 10^{10}$	322	$3,02 \cdot 10^{13}$
	H 1	H 2 (TRUE)	$2,31 \cdot 10^{-8}$	0,0187	0,0016	0	0,3567
Precision ID Identity	H 1 (TRUE)	H 2	1555	$7,87 \cdot 10^5$	$4,7 \cdot 10^5$	2,569	$8,24 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0002	0,581	0,1652	$1,19 \cdot 10^{-7}$	14,09

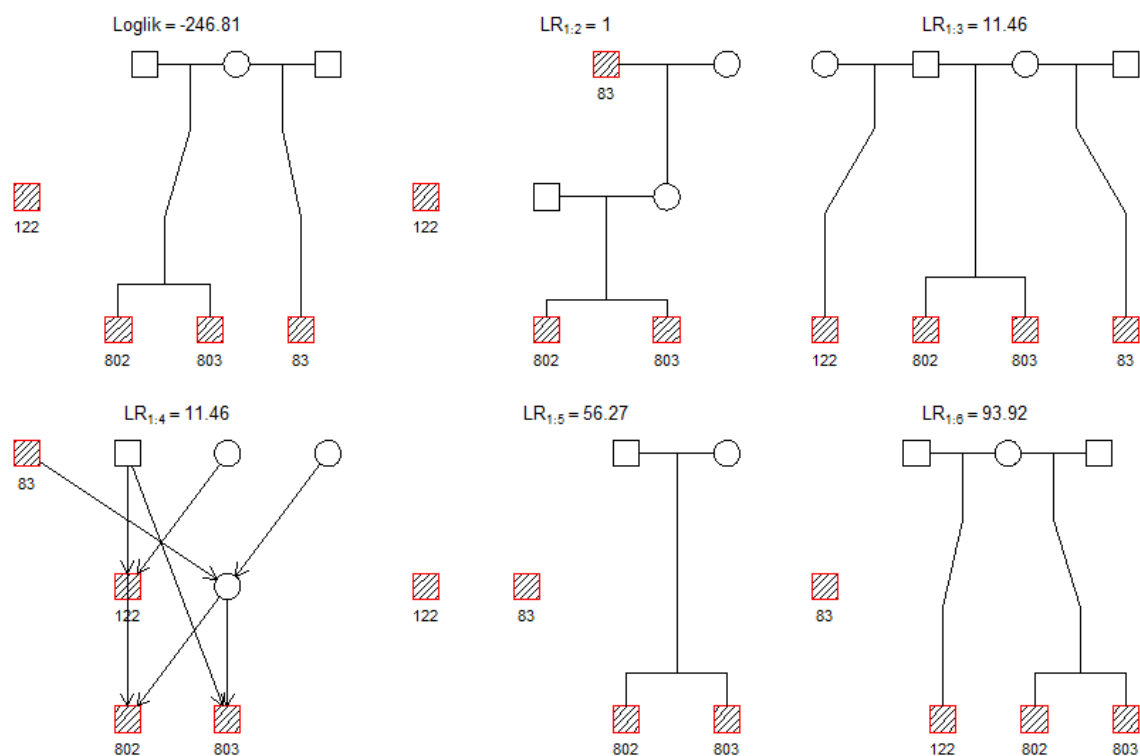
Źródło: opracowanie własne



Rys. 72. Wykres trójkąta IBD dla Rodziny 19 dla zestawu GlobalFiler™ i KinFinder



Rys. 73. Symulacje pokrewieństw dla Rodziny 19 dla zestawu GlobalFiler™



Rys. 74. Symulacje pokrewieństw dla Rodziny 19 dla zestawu KinFinder

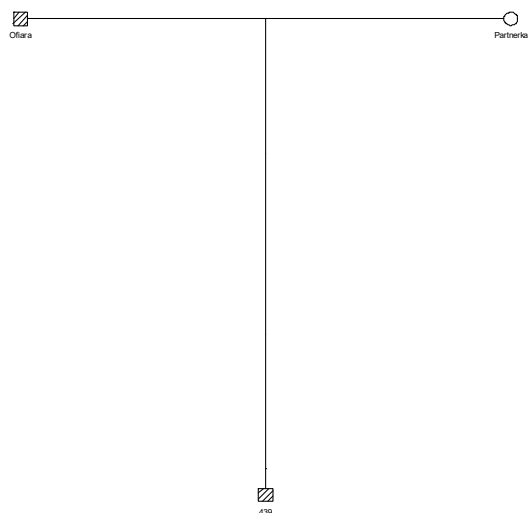
4.20. Rodzina 20

W skład 20 rodziny wchodzi syn osoby zaginionej (osoba 439), relacje między nimi przedstawiono poniżej (Rys. 75).

Po przeszukaniu DVI dopasowano szczątki oznaczone numerem 85. Obliczenia statystyczne wykonane w programie Familias wykazały dla tego dopasowania wartość $LR = 898\,654$, która po uwzględnieniu sprzężeń genetycznych wzrosła do $900\,451$. Równoległa analiza w programie KinBN dała wartość $LR = 1,05 \cdot 10^6$. Program KinBN nie wykazał zmian wartości LR dla obliczeń ze sprzężeniami i bez sprzężeń. Analiza haplotypów chromosomu Y wykazała pełną zgodność i wartość $LR = 198\,098$. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 1,78 \cdot 10^{11}$.

Symulacje rozkładu LR (Tab. 21) wskazują, że uzyskane wyniki mieszczą się w oczekiwanych przedziałach, znajdują się nieco poniżej mediany, nieco powyżej progu 5% dla zestawu GlobalFiler™. Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder, a najniższą Precision ID Identity.

Analiza trójkąta IBD (Rys. 76) oraz symulacji możliwych pokrewieństw (Rys. 77) wykazały zgodność z zakładanym pokrewieństwem.

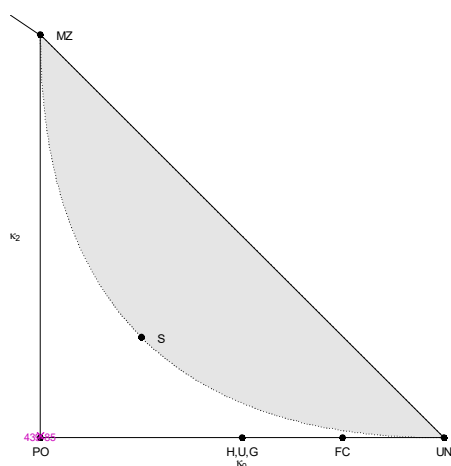


Rys. 75. Drzewo genealogiczne dla Rodziny 20

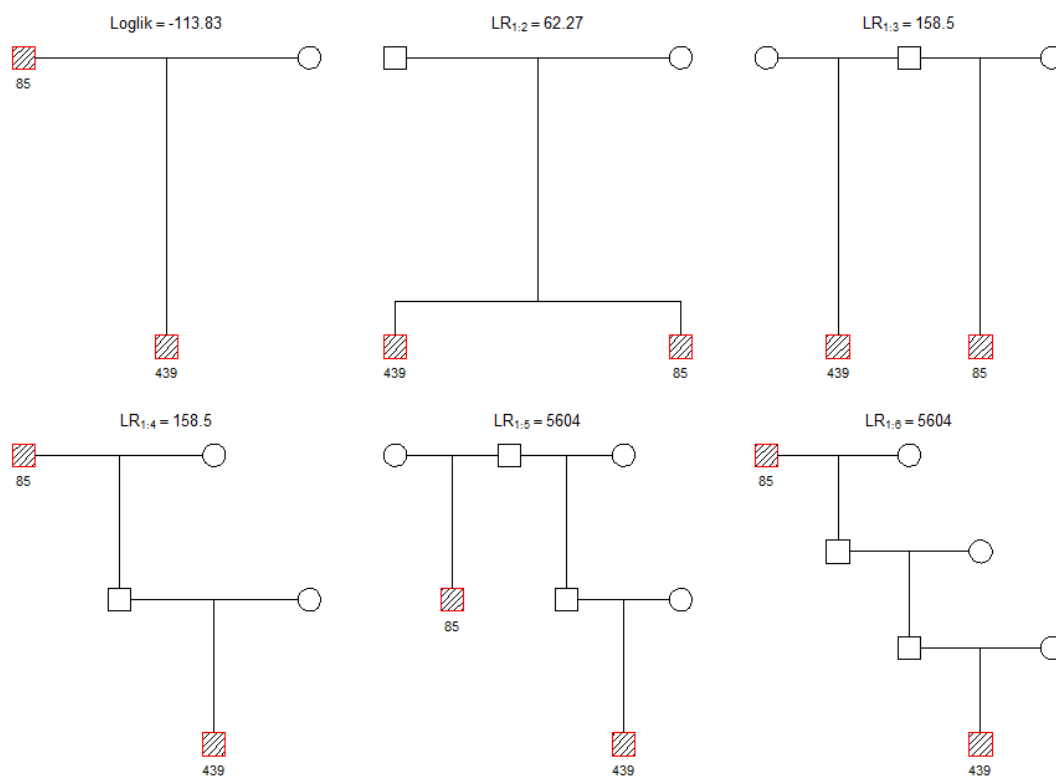
Tab. 21. Symulacja wyników analizy statystycznej dla Rodziny 20

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$5,02 \cdot 10^7$	$6,32 \cdot 10^{11}$	$1,45 \cdot 10^{10}$	$3,01 \cdot 10^5$	$1,39 \cdot 10^{13}$
	H 1	H 2 (TRUE)	0	0	0	0	0
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$4,13 \cdot 10^8$	$3,06 \cdot 10^{12}$	$1,58 \cdot 10^{11}$	$2,81 \cdot 10^6$	$9,27 \cdot 10^{13}$
	H 1	H 2 (TRUE)	0	0	0	0	0
KinFinder	H 1 (TRUE)	H 2	$1,14 \cdot 10^{24}$	$1,26 \cdot 10^{29}$	$4,79 \cdot 10^{27}$	$5,99 \cdot 10^{20}$	$3,13 \cdot 10^{30}$
	H 1	H 2 (TRUE)	0	0	0	0	0
Precision ID Identity	H 1 (TRUE)	H 2	$2,34 \cdot 10^6$	$2,39 \cdot 10^8$	$3,84 \cdot 10^8$	$1,26 \cdot 10^4$	$2,41 \cdot 10^9$
	H 1	H 2 (TRUE)	0	0	0	0	0

Źródło: opracowanie własne



Rys. 76. Wykres trójkąta IBD dla Rodziny 20



Rys. 77. Symulacje pokrewieństw dla Rodziny 20

4.21. Dodatkowe obliczenia

Ze względu na uzyskanie niejednoznacznych wyników analiz pokrewieństwa w odniesieniu do części rodzin, skierowano zapytanie do strony gruzińskiej w celu weryfikacji i ewentualnej korekty założeń genealogicznych. W odpowiedzi kierownik projektu po stronie gruzińskiej przekazała następujące informacje:

- Rodzina 1 – wskazano, że w przypadku osoby nr 220 stopień pokrewieństwa z poszukiwaną ofiarą pozostaje niepewny, a jego jednoznaczne określenie nie jest możliwe. W celu uniknięcia potencjalnych nieporozumień zalecano wyłączenie tej osoby z dalszych analiz statystycznych.
- Rodzina 8 – potwierdzono, iż rodzina ta nie poszukuje ofiary pochodzącej z analizowanej jamy grobowej.
- Rodzina 9 – potwierdzono, że rodzina poszukuje ofiary płci żeńskiej, spoza listy egzekucyjnej
- Rodzina 15 – potwierdzono iż rodzina ta nie poszukuje osoby znajdującej się na liście egzekucyjnej przypisanej do badanego grobu.
- Rodzina 19 – zaktualizowano informacje dotyczące struktury genealogicznej; osoba nr 122 została zidentyfikowana jako syn pierwszego syna ofiary, natomiast osoby nr 802 i 803 jako bracia, synowie drugiego syna ofiary.
- Pokrewieństwa pomiędzy pozostałymi rodzinami zostały potwierdzone.

W związku z powyższymi ustaleniami dokonano korekty przyjętych założeń pokrewieństwa oraz powtórzono obliczenia statystyczne dla rodzin nr 1 i 19. Zaktualizowane wyniki przedstawiono poniżej.

4.21.1. Rodzina 1 po korekcie

Po konsultacjach ze stroną gruzińską w Rodzinie 1 podczas analizy wzięto pod uwagę czterech wnuków zaginionego (osoby 219, 221, 222 i 267) a relacje przedstawiono poniżej (Rys. 78).

Obliczenia statystyczne wykonane w programie Familias dla zestawu GlobalFiler™ wykazały wartość $LR = 202\,032$, która po uwzględnieniu sprzężeń genetycznych wzrosła do $202\,739$. Równoległa analiza w programie KinBN dała wartości odpowiednio $221\,000$ (bez sprzężeń) i $222\,000$ (ze sprzężeniami).

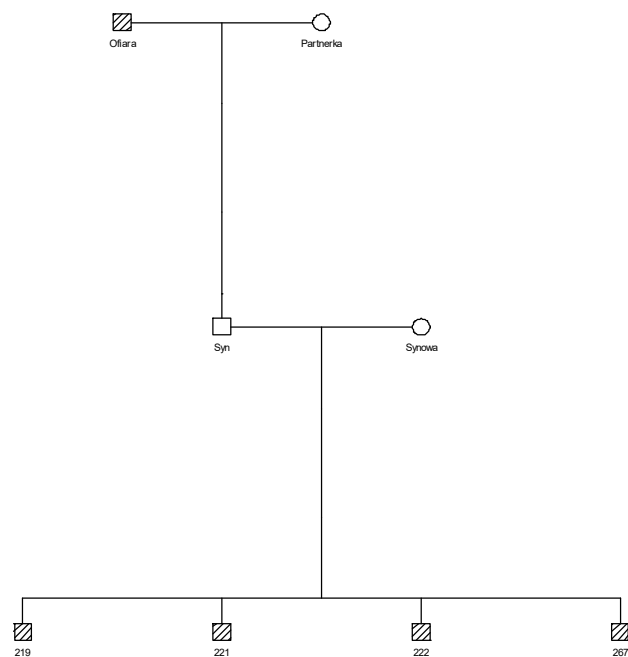
Dla zestawu KinFinder obliczona wartość LR w programie Familias wynosi $9\,209$ (bez sprzężeń) oraz $18\,150$ dla sprzężonych układów, natomiast w KinBN $LR = 8\,870$ (bez sprzężeń) oraz $LR = 17\,800$ (ze sprzężeniami).

LR_{Y-STR} pozostał bez zmian $= 190\,117$. Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LRC = 3,54 \cdot 10^{14}$.

Symulacje rozkładu LR (Tab. 22) wskazują, że uzyskane wyniki dla obu zestawów mieszczą się w oczekiwanych przedziałach, i znajdują się powyżej mediany (KinFinder dla niepełnego profilu).

Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (uwzględniając niepełne oznaczenie profilu – 15 układów z 50 uzyskał porównywalną moc z zestawami GlobalFiler™ i PowerPlex® Fusion 6C), a najniższą Precision ID Identity.

Analizy IBD przeprowadzone w pakiecie pedbuildr potwierdziły zgodność zakładanej hipotezy (Rys. 79, Rys. 80, Rys. 81).

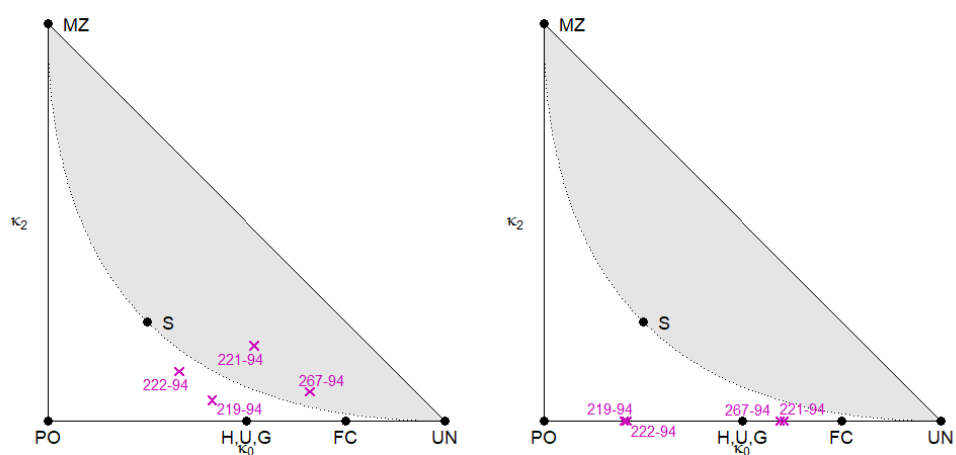


Rys. 78. Drzewo genealogiczne dla Rodziny 1 po korekcie

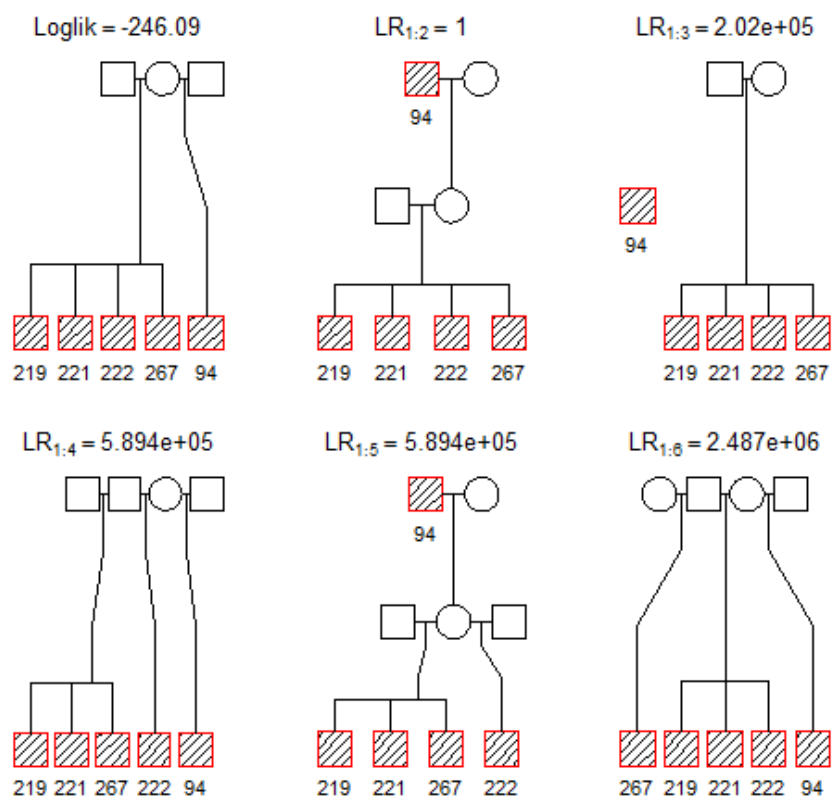
Tab. 22. Symulacja wyników analizy statystycznej dla Rodziny 1 po korekcie

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	8306	$1,23 \cdot 10^6$	$4,3 \cdot 10^6$	42,85	$9,67 \cdot 10^6$
	H 1	H 2 (TRUE)	0	1,575	0,0012	0	33,04
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$2,25 \cdot 10^4$	$3,57 \cdot 10^7$	$1,15 \cdot 10^7$	56,36	$7,96 \cdot 10^8$
	H 1	H 2 (TRUE)	0	0,0142	0	0	0,4024
KinFinder	H 1 (TRUE)	H 2	$1,07 \cdot 10^{13}$	$4,22 \cdot 10^{17}$	$7,59 \cdot 10^{16}$	$9,31 \cdot 10^8$	$9,13 \cdot 10^{18}$
	H 1	H 2 (TRUE)	0	0	0	0	0
KinFinder (układy obecne w profilu szczątków nr 94)	H 1 (TRUE)	H 2	3480	$2,82 \cdot 10^5$	$3,17 \cdot 10^5$	24,46	$5,5 \cdot 10^6$
	H 1	H 2 (TRUE)	0	1,301	0	0	29,88
Precision ID Identity	H 1 (TRUE)	H 2	211,8	$2,22 \cdot 10^4$	$2,69 \cdot 10^4$	1,545	$4,69 \cdot 10^5$
	H 1	H 2 (TRUE)	0,0007	1,406	1,431	$1,71 \cdot 10^{-7}$	15,4

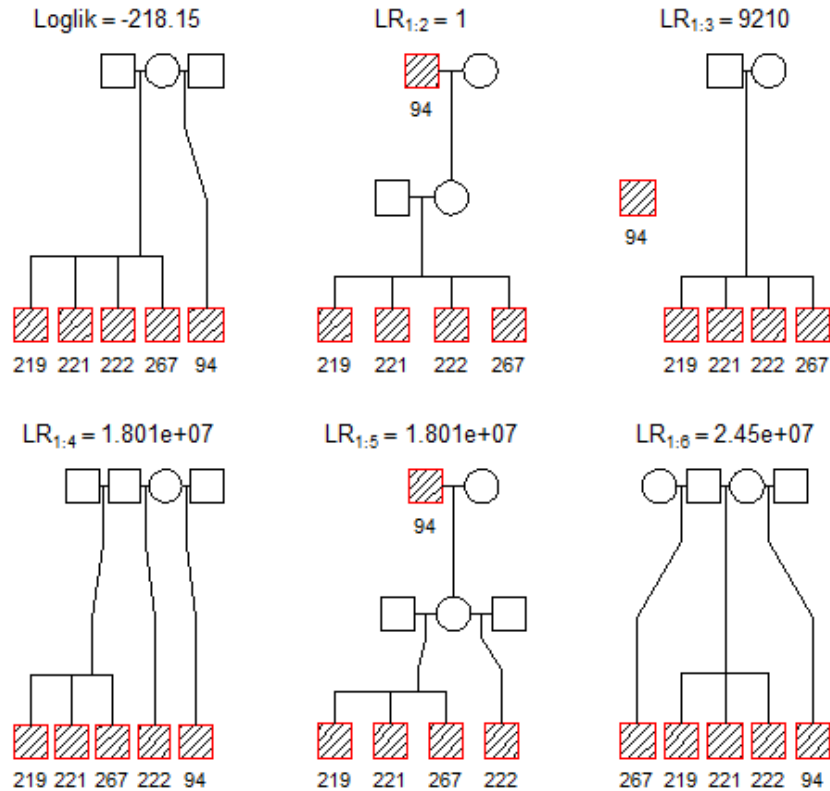
Źródło: opracowanie własne



Rys. 79. Wykres trójkąta IBD dla Rodziny 1 dla zestawu GlobalFiler™ i KinFinder po korekcji



Rys. 80. Symulacje pokrewieństw dla Rodziny 1 dla zestawu GlobalFiler™ po korekcji



Rys. 81. Symulacje pokrewieństw dla Rodziny 1 dla zestawu KinFinder po korekcie

4.21.2. Rodzina 19 po korekcie

W skład 19 rodziny wchodzi trzech wnuków ofiary: osoba 122 będąca synem pierwszego syna ofiary oraz osoby 802 i 803 będące synami drugiego syna ofiary, relacje przedstawiono poniżej (Rys. 82).

Obliczenia statystyczne wykonane w programie Familias wykazały wartość $LR = 62\,988$ która po uwzględnieniu sprzężeń genetycznych wzrosła do $75\,352$. Równoległa analiza w programie KinBN dała wartości odpowiednio $70\,600$ (bez sprzężeń) i $84\,300$ (ze sprzężeniami).

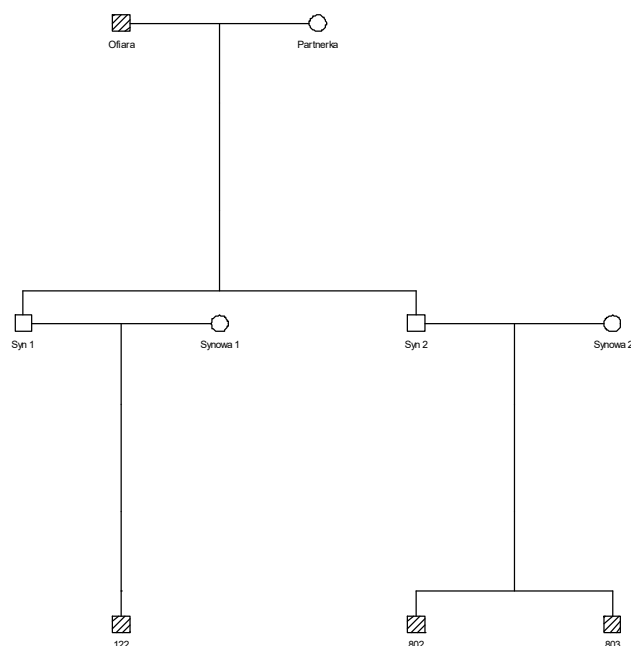
Obliczenia w programie Familias dla zestawu KinFinder wykazały wartość $LR = 21\,220$, która po uwzględnieniu sprzężeń genetycznych spadła do $19\,238$. Analiza w programie KinBN dała wartość odpowiednio 91 (bez sprzężeń) i 86 (ze sprzężeniami).

LR Y-STR nie uległ zmianie i wynosi $189\,665$.

Obliczono łączny iloraz wiarygodności dla STR autosomalnych i Y-STR, $LR_c = 2,54 \cdot 10^{14}$.

Symulacje rozkładu LR (Tab. 23) wskazują, że uzyskany wynik jest nieco wyższy niż mediana dla zestawu GlobalFiler™. Dla zestawu KinFinder uzyskany wynik znajduje się poniżej mediany ale powyżej progu 5% (dla symulacji dla niepełnego profilu). Najwyższą moc dowodową w symulacji wyników wykazał zestaw KinFinder (nawet uwzględniając niepełne oznaczenie profilu – 21 układów z 50), a najniższą Precision ID Identity.

W przypadku tej rodziny trójką IBD i symulacje pokrewieństw nie ulegną zmianie, gdyż nie zmieniła się liczba osób w rodzinie.



Rys. 82. Drzewo genealogiczne dla Rodziny 19 po korekcie

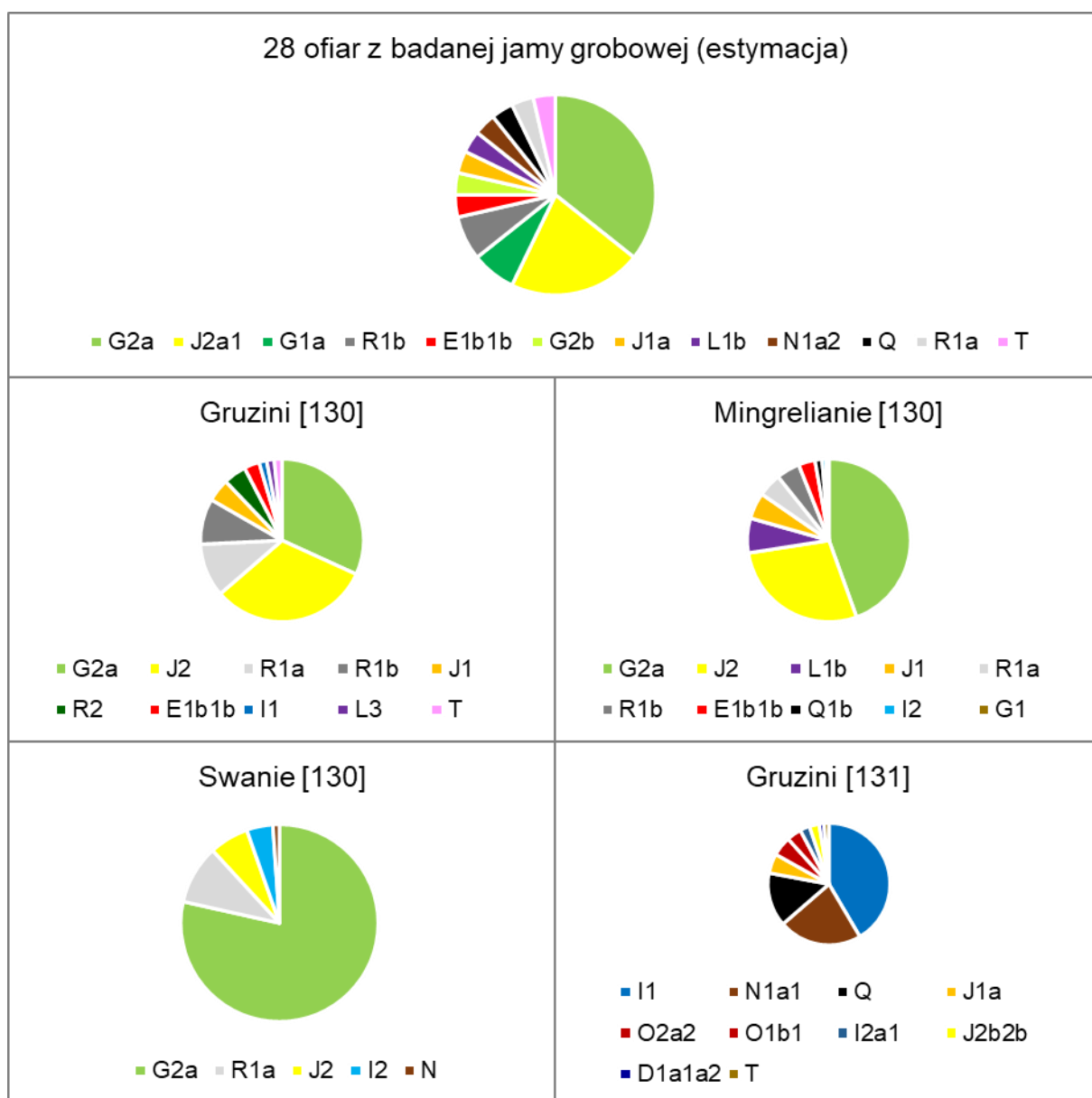
Tab. 23. Symulacja wyników analizy statystycznej dla Rodziny 19 po korekcie

Zestaw	Hypothesis	Versus	Median	Mean	95%	5%	Stdev
GlobalFiler™	H 1 (TRUE)	H 2	$2,36 \cdot 10^4$	$9,28 \cdot 10^7$	$3,51 \cdot 10^7$	16,17	$1,51 \cdot 10^9$
	H 1	H 2 (TRUE)	$1,34 \cdot 10^{-8}$	0,1422	0,0277	0	2,326
PowerPlex® Fusion 6C	H 1 (TRUE)	H 2	$8,95 \cdot 10^4$	$5,77 \cdot 10^8$	$1,5 \cdot 10^8$	58,77	$1,29 \cdot 10^{10}$
	H 1	H 2 (TRUE)	0	0,1041	0,0122	0	1,516
KinFinder	H 1 (TRUE)	H 2	$1,79 \cdot 10^{14}$	$1,43 \cdot 10^{21}$	$2,21 \cdot 10^{19}$	$7,12 \cdot 10^8$	$3,36 \cdot 10^{22}$
	H 1	H 2 (TRUE)	0	$8,91 \cdot 10^{-10}$	$8,86 \cdot 10^{-17}$	0	$2,29 \cdot 10^{-8}$
KinFinder (układy obecne w profilu szczątków nr 83)	H 1 (TRUE)	H 2	$1 \cdot 10^6$	$2,04 \cdot 10^{10}$	$4,26 \cdot 10^9$	411,2	$3,74 \cdot 10^{11}$
	H 1	H 2 (TRUE)	0	0,0162	0,0004	0	0,328
Precision ID Identity	H 1 (TRUE)	H 2	509	$1,81 \cdot 10^5$	$1,46 \cdot 10^5$	2,053	$4,21 \cdot 10^6$
	H 1	H 2 (TRUE)	0,0005	1,57	0,4969	$4,48 \cdot 10^{-7}$	32,67

Źródło: opracowanie własne

4.22. Pochodzenie biogeograficzne

Rozkład haplogrup chromosomu Y zaobserwowany u wszystkich 28 szczątków z masowego grobu w Adżarii silnie wskazywał na ich pochodzenie gruzińskie (Rys. 83).



Rys. 83. Rozkłady haplogrup chromosomu Y oraz oszacowanych haplogrup chromosomu Y szczątków z masowego grobu w Adżarii, Gruzinów, Mingrelian i Svanów [130, 131].

Dominującymi haplogrupami wśród analizowanych szczątków były G2a (35%) i J2a1 (20%), obie szeroko rozpoznawane jako charakterystyczne dla populacji posługujących się językami kartwelskimi i często występujące u współczesnych Gruzinów oraz Mingrelian [130, 152]. Niewielki odsetek stanowiły haplogrupy typowe dla niegruzińskich populacji sąsiednich, takie jak R1a, Q i N1a2 [153]. Rozkład obserwowanych haplogrup okazał się bardzo zbliżony do struktury genetycznej populacji mingrelskiej, w której odsetki G2a i J2a wynoszą odpowiednio około 45% i 25% co jest zgodne z ogólnym zestawem danych dla Gruzinów, w którym dominują linie G2a i J2a [130]. Różnił się on od profilu Svanów, gdzie G2a osiąga około 70%. [130]. Pozorna niezgodność wyników uzyskanych przez Nasidze i wsp. [131] była

najprawdopodobniej spowodowana bardzo ograniczonym zestawem markerów (zaledwie 9 Y-STR) użytych w tamtym badaniu.

Na podstawie tych wyników potwierdzono wstępną hipotezę dotyczącą gruzińskiego pochodzenia osób pochowanych w tej mogile, dostarczając mocnych dowodów genetycznych zgodnych z danymi historycznymi i demograficznymi.

4.23. Opracowanie schematu decyzyjnego doboru paneli markerów genetycznych

Dobór zestawów markerów zastosowanych w analizie został przeprowadzony zgodnie z opracowanym schematem decyzyjnym, przedstawionym na Rys. 84. Schemat ten odzwierciedla sekwencyjne podejście, w którym kolejne systemy były wdrażane w zależności od wyników uzyskanych na wcześniejszych etapach obliczeń statystycznych.



Rys. 84. Schemat decyzyjny

We wszystkich przypadkach analizę rozpoczęto od paneli podstawowych: 21 autosomalnych STR (GlobalFiler™) oraz 27 markerów Y-STR (Yfiler™ Plus). Uzyskanie wartości LR przekraczającej 100 000 oraz jednocześnie potwierdzenie w pełni zgodnego haplotypu Y-STR traktowano jako przesłankę do uznania wyniku za wystarczająco silny. Natomiast przypadki, w których wartość LR nie przekraczała tego progu i/lub nie obserwowano jednoznacznego dopasowania Y-STR, kierowano do rozszerzonej analizy z użyciem 48 dodatkowych markerów panelu KinFinder.

W sytuacjach, w których nawet rozszerzona analiza STR nie była wystarczająca, zastosowano dodatkowy panel obejmujący 90 iiSNP (Precision ID Identity).

Zastosowany schemat umożliwił uzyskanie jednoznacznych dopasowań genetycznych dla wszystkich 17 rodzin poszukujących krewnych znajdujących się na liście egzekucyjnej. Natomiast żadna z rodzin poszukujących osób nieujętych na liście egzekucyjnej nie uzyskała wartości LR wskazujących na potencjalne dopasowanie już na etapie panelu GlobalFiler™. W Rodzinie 9 dodatkowym czynnikiem wykluczającym był fakt poszukiwania kobiety, podczas gdy analiza antropologiczna oraz oznaczenie płci genetycznej potwierdziły, że wszystkie 28 szczątków należały do mężczyzn. Pozostałe dwie rodziny spoza listy (R8 i R15) skierowano do analizy rozszerzonej KinFinder, co również nie dostarczyło dowodów wskazujących na pokrewieństwo – w połączeniu z brakiem zgodności haplotypów Y-STR oraz z danymi historycznymi sugerującymi niskie prawdopodobieństwo pochówku tych osób w badanej jamie doprowadziło do ich ostatecznego wykluczenia z dalszych analiz.

Z uwagi na szeroki zakres zastosowanych paneli wykonano dodatkowo analizę z uwzględnieniem i bez uwzględnienia sprzężeń genetycznych pomiędzy markerami. Włączenie informacji o sprzężeniu nie powodowało zmian wartości LR przekraczających jeden rząd wielkości, a w niektórych przypadkach prowadziło nawet do niewielkiego wzrostu ocenianych wartości. Dla zachowania przejrzystości prezentacji wyników wszystkie wartości LR odnoszą się do modeli zakładających niezależność dziedziczenia markerów.

Przedstawiony schemat decyzyjny pozwolił na prowadzenie analiz zgodnie z zasadą progresywnego zwiększania mocy dowodowej – od paneli podstawowych, przez rozszerzone STR, aż po SNP – co umożliwiło maksymalizację skuteczności identyfikacji przy jednoczesnym zachowaniu efektywności ekonomicznej procesu analitycznego.

4.24. Omówienie wyników badań

W wyniku przeprowadzonych badań zidentyfikowano 17 szczątków ludzkich. Pierwszy etap przyjętego schematu zakończył się identyfikacją siedmiu osób, po drugim kroku liczba ta wzrosła do 14, a dzięki dodaniu kroku trzeciego zakończono na 17 identyfikacjach. Nie stwierdzono dopasowania do rodzin poszukujących osób spoza listy egzekucyjnej przypisanej do analizowanej jamy grobowej.

Identyfikacje uzyskane w pierwszym kroku schematu decyzyjnego:

- Rodzina 3 – szczątki 106,
- Rodzina 11 – szczątki 99,
- Rodzina 12 – szczątki 92.
- Rodzina 5 – szczątki 102,
- Rodzina 18 – szczątki 84,

- Rodzina 20 – szczątki 85.

- Rodzina 1 – szczątki 94 – po konsultacjach ze stroną gruzińską, uzyskano wartość dowodową dla zestawu GlobalFiler™ na poziomie $> 100\,000$ w związku z czym, pomimo wcześniejszego oznaczenia zestawu KinFinder dla tej rodziny można ją przypisać do tej kategorii.

W pierwszych trzech przypadkach do identyfikacji wystarczyłoby oznaczenie oparte wyłącznie na GlobalFiler™. W pozostałych czterech przypadkach identyfikację wsparto na połączonych zestawach GlobalFiler™ i Yfiler™ Plus.

Identyfikacje uzyskane w drugim kroku schematu decyzyjnego:

- Rodzina 2 – szczątki 97,
- Rodzina 6 – szczątki 88,
- Rodzina 7 – szczątki 98,
- Rodzina 10 – szczątki 93,
- Rodzina 14 – szczątki 95,
- Rodzina 17 – szczątki 105,
- Rodzina 19 – szczątki 83

moc dowodowa uzyskana z zestawu GlobalFiler™ była w powyższych przypadkach niewystarczająca. Rozszerzenie badań o panel KinFinder umożliwiło osiągnięcie wartości identyfikacyjnych dopiero po połączeniu wyników z obu zestawów (moc identyfikacyjna samego zestawu KinFinder również była zbyt słaba by stwierdzić identyfikację).

W przypadku Rodziny 14 wyniki uzyskane na podstawie GlobalFiler™ i KinFinder były zbyt niskie, aby przeprowadzić identyfikację. Dopiero uzupełnienie analiz o panel Yfiler™ Plus pozwoliło na uzyskanie wartości LR powyżej progu identyfikacyjnego.

W przypadku rodzin 2, 7 i 10 w celu weryfikacji poprawności opracowanego schematu decyzyjnego wykonano dodatkowo oznaczenia zestawem Precision ID Identity. Połączenie zestawów GlobalFiler™ i Precision ID Identity nie zapewniło wystarczającej mocy dowodowej, co potwierdziło zastosowaną kolejność wykorzystania zestawów odczynników.

Identyfikacje uzyskane po trzecim kroku schematu decyzyjnego:

- Rodzina 4 – szczątki 80,
- Rodzina 13 – szczątki 96,
- Rodzina 16 – szczątki 91

W powyższych przypadkach moc dowodowa uzyskana z połączonych zestawów GlobalFiler™ i KinFinder była zbyt niska. Dopiero po dodaniu trzeciego panelu – Precision ID Identity – uzyskano wartości LR wystarczające do identyfikacji. Analiza połączonych zestawów KinFinder i Precision ID Identity wykazała moc dowodową powyżej progu dla rodzin 4 i 16,

natomiast dla rodziny 13 wartość ta pozostawała poniżej wymaganego poziomu. Należy podkreślić, że w przypadku rodziny 4, opierając się wyłącznie na zestawie GlobalFiler™, nie uzyskano by nawet wskazania możliwego dopasowania.

Symulacja wyników obliczeń statystycznych dla Rodziny 9 wskazała na ograniczoną przydatność analizowanych zestawów w przypadku rozpatrywania pokrewieństwa IV stopnia (dziecko kuzyna analizowanej osoby).

4.25. Analiza haplotypów chromosomu Y

Spośród 17 dopasowań, w 15 przypadkach zasadne było oznaczenie profili Y-STR za pomocą zestawu Yfiler™ Plus.

W 13 rodzinach wartości LR oparte na haplotypach Y-STR przekraczały 100 000, w jednym przypadku wyniosły 94 000, a w jednym przypadku 788 (ze względu na niezgodność w układzie DYS456). Po uwzględnieniu mocy identyfikacyjnej połączonych zestawów GlobalFiler™ i Yfiler™ Plus stwierdzono, że wartości LRc powyżej 1 000 000 uzyskano dla 13 rodzin. W pozostałych dwóch przypadkach dopiero rozszerzenie badań o KinFinder lub Precision ID Identity umożliwiło osiągnięcie progu identyfikacji.

W 11 rodzinach stwierdzono pełną zgodność haplotypów Y-STR.

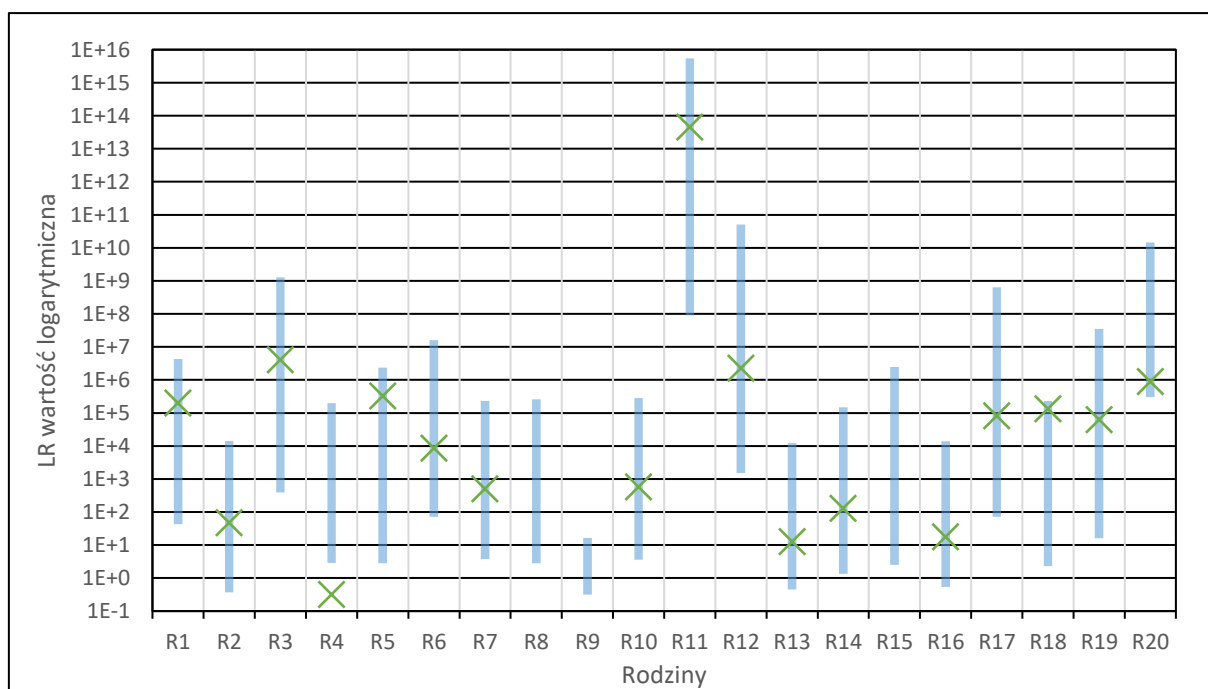
W pozostałych przypadkach zaobserwowano pojedyncze niezgodności:

- Rodzina 6 – zgodność profilu 231 ze szczątkami 88; jedna niezgodność między profilami 230 i 231 w układzie DYS390 (przyrost o jedno powtórzenie);
- Rodzina 12 – zgodność profilu 246 ze szczątkami nr 92;
 - niezgodność między profilami 246 a 245: DYS570 (–1 powtórzenie), DYS437 (+1 powtórzenie),
 - niezgodność między profilami 246 a 247: DYS627 (–1 powtórzenie), DYF387S1 (+1 powtórzenie).
- Rodzina 13 – niezgodność między profilem 250 a szczątkami nr 96 w układzie DYS456 (–1 powtórzenie).
- Rodzina 14 – zgodność profili 801, 251 i szczątków nr 95; niezgodność z profilem 252 w układzie DYS570 (–1 powtórzenie).

Zaistniałe różnice można interpretować jako skutki mutacji mikrosatelitarnych, co zgodnie z danymi literaturowymi jest naturalnym procesem, który mimo, że obniżył uzyskane wartości LR jednakże nie wpłynął na proces identyfikacji [125].

4.26. Przydatność zestawów markerów

W analizach identyfikacyjnych podstawowym punktem odniesienia był zestaw GlobalFiler™, szeroko stosowany w praktyce genetyki sądowej. Uzyskane z jego wykorzystaniem wartości LR charakteryzowały się stabilnością i powtarzalnością, a w większości przypadków mieściły się w oczekiwanych przedziałach symulacyjnych (Rys. 85.). Jedynie w przypadku Rodziny 4 uzyskano wynik poniżej wartości symulowanej. W sytuacjach bardziej złożonych lub przy dalszych relacjach rodzinnych wartości te lokowały się jednak częściej w dolnym zakresie rozkładów, co ograniczało ich moc dowodową.

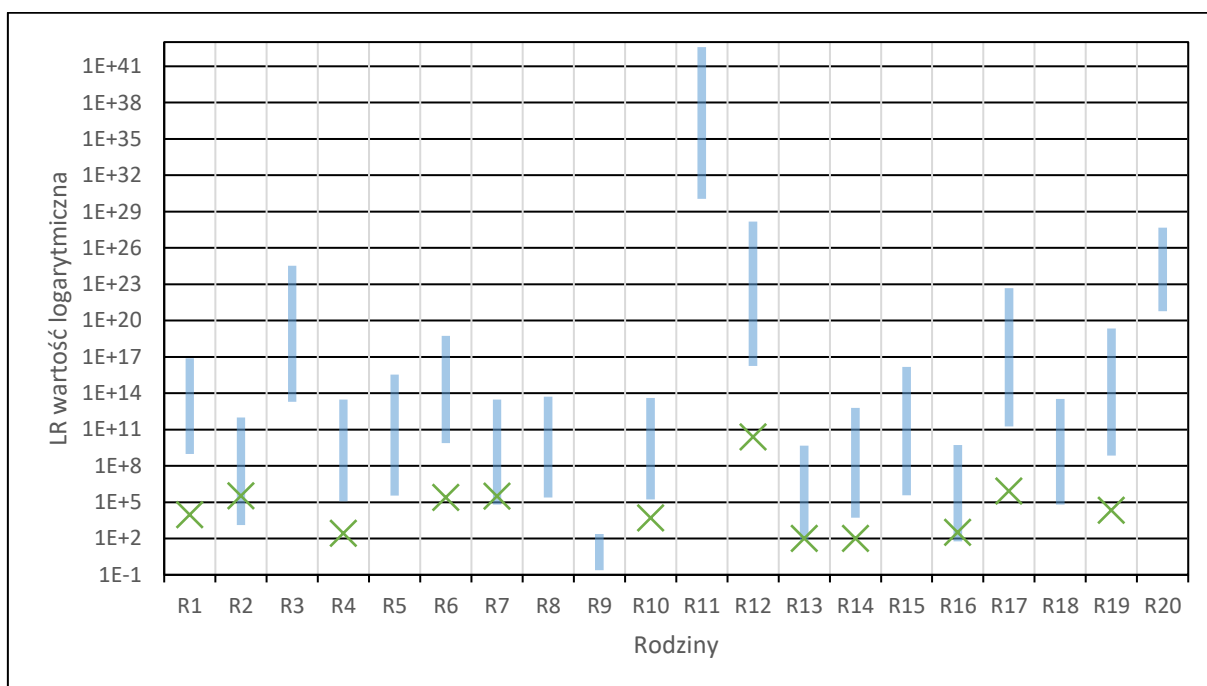


Rys. 85. Wykres symulacji rozkładu wartości LR i uzyskanych wyników dla zestawu GlobalFiler™

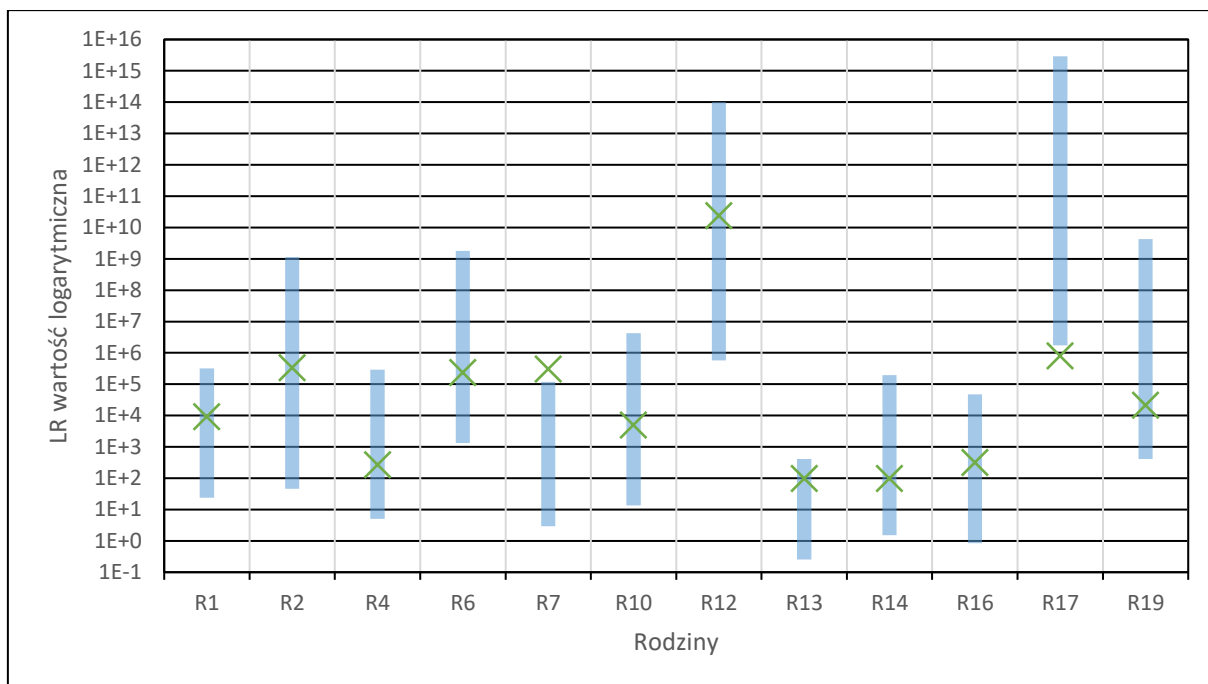
Zestaw PowerPlex® Fusion 6C został włączony do analiz jako przykład innego, powszechnie stosowanego komercyjnie panelu, pozwalającego porównać uzyskane wyniki z odmiennym systemem markerów. Uzyskane wartości LR były zbliżone lub nieco wyższe niż w przypadku GlobalFiler™, co potwierdziło spójność analiz pomiędzy zestawami i podkreśliło przydatność tego panelu jako punktu odniesienia w badaniach identyfikacyjnych.

Najwyższą moc dowodową spośród zastosowanych paneli wykazywał KinFinder, co potwierdziły zarówno wyniki symulacji, jak i obliczenia rzeczywistych wartości LR. KinFinder został zwalidowany w GenMed do rutynowych badań z wykorzystaniem materiału porównawczego dobrej jakości, jego zastosowanie do badań DNA pochodzącego z kości nie zostało do tej pory ocenione, co sprawia, że zastosowanie tego zestawu w niniejszym badaniu ma charakter eksploracyjny. Zestaw ten okazał się szczególnie użyteczny w analizach

złożonych genealogii i dalszych relacji rodzinnych. Należy jednak podkreślić, że zestaw ten był bardziej podatny na występowanie zjawiska drop-out spowodowanego niską jakością materiału genetycznego. W konsekwencji, w przypadku analizy materiału pobranego ze szczątków uzyskiwano niepełne profile genetyczne, przez co w czterech z 12 przypadków uzyskano niższe wartości LR niż wynikające z symulacji (Rys. 86). Dla symulacji przeprowadzonych przy użyciu zestawu markerów faktycznie uzyskanych w profilach analizowanych ofiar uzyskano wyniki mieszczące się w przedziałach (Rys. 87, Rys. 88). Jedynie dla Rodziny 17 uzyskano wynik poniżej progu 5%, należy zauważyć iż wynik ten nadal był bardzo wysoki (LR = 804 697). Mimo tego ograniczenia, nawet częściowe profile dostarczały wartości dowodowych porównywalnych, a nierzadko wyższych niż w przypadku GlobalFiler™, co potwierdza wysoką użyteczność zestawu KinFinder w trudnych przypadkach identyfikacyjnych.

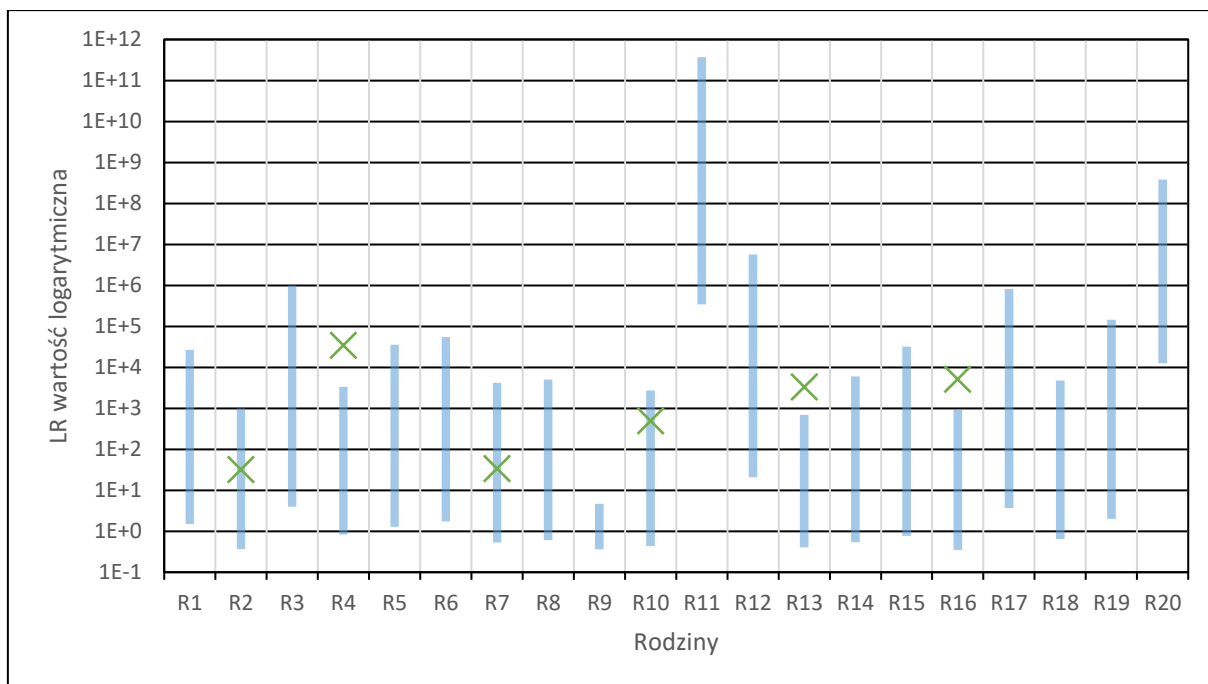


Rys. 86. Wykres symulacji rozkładu wartości LR i uzyskanych wyników dla zestawu KinFinder



Rys. 87. Wykres symulacji rozkładu wartości LR przy uwzględnieniu markerów oznaczonych dla danych szczątków i uzyskanych wyników dla zestawu KinFinder

Zestaw Precision ID Identity, obejmujący panele SNP, charakteryzował się zdecydowanie niższymi wartościami LR niż panele STR. W trzech przypadkach wyniki uzyskane z obliczeń były wyższe niż oczekiwane w wyniku symulacji (Rys. 88), jednak ogólnie uzyskał on niższe wartości w porównaniu z innymi systemami. Jego zastosowanie miało zatem głównie charakter uzupełniający, zwłaszcza w sytuacjach, gdy moc dowodowa markerów STR była ograniczona. Zaletą panelu SNP jest możliwość uzyskania wiarygodnych wyników z próbek o obniżonej jakości DNA. Co istotne wyniki uzyskane z SNP we wszystkich analizowanych rodzinach korespondowały z tymi uzyskanymi z STR autosomalnych potwierdzając zakładane pokrewieństwa.



Rys. 88. Wykres symulacji rozkładu wartości LR i uzyskanych wyników dla zestawu Precision ID Identity

Analizy markerów chromosomu Y z użyciem zestawu Yfiler™ Plus okazały się szczególnie wartościowe przy weryfikacji zgodności w liniach męskich. Z uwagi na mechanizm dziedziczenia haplotypów Y-STR zestaw ten jest mniej informatywny niż zestawy autosomalnych STR i nie pozwala na rozróżnianie stopnia pokrewieństwa. Mimo to w wybranych rodzinach, w tym zwłaszcza w przypadku Rodziny 4, analiza haplotypów Y-STR wносиła informacje, których nie dostarczyły panele autosomalne.

4.27. Przydatność zastosowanych programów

Do analiz identyfikacyjnych wykorzystano kilka narzędzi bioinformatycznych, które wzajemnie się uzupełniały i umożliwiały zarówno prowadzenie obliczeń statystycznych, jak i weryfikację uzyskanych wyników.

Familias stanowił podstawowe narzędzie do obliczeń statystycznych, w tym przeszukań masowych typu DVI, które nie są dostępne we wszystkich programach biostatystycznych. Oprogramowanie pozwalało na szybkie wytypowanie najbardziej prawdopodobnych dopasowań oraz na obliczenia ilorazu wiarygodności przy różnych hipotezach pokrewieństwa. Dodatkowo umożliwiało przeprowadzanie symulacji rozkładów LR dla określonych scenariuszy rodzinnych, co było istotnym punktem odniesienia przy interpretacji uzyskanych wartości. W niniejszym badaniu to właśnie Familias pełnił kluczową rolę w etapie selekcji i weryfikacji wyników.

W analizach uwzględniających sprzężenia genetyczne korzystano z programu FamLink, który umożliwiał uwzględnienie zależności między markerami znajdującymi się blisko siebie na tym samym chromosomie. Dzięki temu uzyskiwane wartości LR mogły być oszacowane z większą precyzją, zwłaszcza w przypadku markerów znajdujących się w niewielkiej odległości fizycznej. W praktyce FamLink był używany w połączeniu z programem Familias, dostarczając współczynników korygujących, które następnie przeliczano na końcowe wartości LR.

KinBN stosowano równolegle do Familiasa w celu niezależnej weryfikacji wyników. Choć program nie oferuje modułu DVI, pozwalał na obliczenia LR zarówno przy założeniu niezależności markerów, jak i przy uwzględnieniu sprzężeń genetycznych. Zastosowanie dwóch niezależnych programów pozwoliło na porównanie i weryfikację uzyskanych wartości LR.

Pedbuildr miał znaczenie przede wszystkim w aspekcie weryfikacyjnym. Dzięki analizom trójkąta IBD oraz symulacjom alternatywnych struktur genealogicznych możliwe było wskazanie najbardziej prawdopodobnych relacji rodzinnych oraz weryfikacja uzyskanych wyników co w niektórych rodzinach umożliwiło wykrycie innych scenariuszy pokrewieństwa niż zakładane. Program ten stanowił cenne uzupełnienie obliczeń statystycznych.

Dzięki takiemu komplementarnemu podejściu uzyskane wyniki były bardziej wiarygodne, a proces identyfikacji oparty na wielu źródłach obliczeniowych pozwolił zminimalizować ryzyko błędnej interpretacji danych.

4.28. Ocena jakości DNA i efektywności zestawów

W celu porównania skuteczności zastosowanych zestawów odczynników porównano ilości wyizolowanego DNA, obliczony indeks degradacji oraz liczbę uzyskanych markerów dla poszczególnych próbek (Tab. 24). Przy porównaniu analizowano tylko układy autosomalne dostępne w zestawach oraz nie uwzględniono amelogeniny.

Dla wszystkich analizowanych próbek nie stwierdzono obecności inhibitorów reakcji (wartości Ct IPC mieściły się w normie).

Analiza wykazała, że dla zestawu GlobalFiler™ w zdecydowanej większości przypadków (20) uzyskano pełne profile genetyczne. W siedmiu przypadkach uzyskano 20 układów z 21. Tylko w jednym przypadku uzyskano jedynie 14 w pełni oznaczonych układów. Była to próbka dla której uzyskano najmniejszą ilość DNA (poniżej 0,02 ng/μl).

Wskazuje to na wysoką czułość zestawu oraz dobrą tolerancję na degradację materiału.

W przypadku zestawu KinFinder liczba oznaczonych układów była bardziej zróżnicowana (od 9 do 34). Ogólnie oznaczono 25 profili z materiału kostnego przy wykorzystaniu tego

zestawu i dla 10 przypadków uzyskano 21 lub więcej w pełni oznaczonych układów. Zaobserwowano, że wyższe stężenia DNA sprzyjały uzyskaniu bardziej kompletnych profili, natomiast w próbkach o niskiej ilości DNA liczba oznaczonych markerów istotnie spadała. Zależność od indeksu degradacji była mniej jednoznaczna – niektóre próbki o wysokim ID dawały lepszy wynik od próbek o umiarkowanej degradacji.

Zestaw Precision ID Identity Panel charakteryzował się najmniejszą wrażliwością na jakość i ilość DNA. Łącznie oznaczono osiem profili dowodowych z czego w czterech przypadkach oznaczono pełne profile, w pozostałych liczba oznaczonych układów wahała się od 87 do 89.

Podsumowując, wyniki wskazują, że:

- ilość DNA pozostaje głównym czynnikiem determinującym liczbę oznaczonych markerów,
- indeks degradacji ma mniejsze znaczenie dla zestawów bazujących na krótkich amplikonach (Precision ID Identity), natomiast w umiarkowanym stopniu wpływa na KinFinder i GlobalFiler™,
- GlobalFiler™ wykazuje bardzo wysoką odporność na degradację i pozwala na uzyskanie dobrych jakościowo profili nawet z próbek o niewielkiej zawartości DNA,
- KinFinder jest bardziej wymagający ilościowo, ale przy dostatecznej ilości DNA dostarcza dużej liczby markerów zwiększających moc dowodową,
- Precision ID Identity okazuje się najbardziej przydatny w próbkach trudnych, zdegradowanych, umożliwiając uzyskanie użytecznych wyników tam, gdzie klasyczne STR dają profile niepełne.

Tab. 24. Zestawienie ilości oznaczonych układów i jakości DNA

nr próbki	Stężenie T. Large [ng/μl]	Stężenie T. Small [ng/μl]	Indeks degradacji	Liczba oznaczonych układów dla zestawu:		
				GlobalFiler™	KinFinder	Precision ID Identity
80	0,04138	0,26404	6,38	21	16	90
81	0,09964	0,27704	2,78	21	24	-
82	0,01511	0,08804	5,83	21	15	89
83	0,04091	0,10469	2,56	21	21	-
84	0,10950	0,58375	5,33	21	-	-
85	0,09536	0,60510	6,35	21	-	-
86	0,03709	0,15419	4,16	21	17	-
87	0,05801	0,48381	8,34	20	17	-
88	0,13614	0,52307	3,84	21	22	-
89	0,03510	0,27526	7,84	20	18	-
90	0,00678	0,01966	2,90	14	12	-
91	0,03484	0,07042	2,02	20	18	90
92	0,05276	0,61014	11,56	20	21	-
93	0,18653	0,78023	4,18	21	21	90
94	0,02496	0,22888	9,17	21	15	-
95	0,05800	0,22372	3,86	21	16	-
96	0,01545	0,03433	2,22	20	9	87
97	0,19707	0,82559	4,19	21	29	90
98	0,02777	0,19009	6,85	20	15	87
99	0,09636	0,30867	3,20	21	17	-
100	0,04178	0,21382	5,12	21	16	-
101	0,30352	1,45759	4,80	21	34	-
102	0,15153	0,58848	3,88	21	29	-
103	0,00919	0,14728	16,03	21	16	-
104	0,23226	0,73694	3,17	21	26	-
105	0,18092	0,63609	3,52	21	30	-
106	0,11862	0,33375	2,81	21	-	-
107	0,02237	0,12370	5,53	20	13	88

Źródło: opracowanie własne

5. Dyskusja

5.1. Ocena stopnia degradacji materiału kostnego i jego wpływ na uzyskane wyniki

Stopień degradacji DNA w materiale kostnym stanowi jeden z kluczowych czynników determinujących skuteczność analiz genetycznych. Proces degradacji DNA w materiale kostnym zależy w dużym stopniu od warunków środowiskowych, w jakich szczątki były deponowane. Wilgotność gleby, dostęp tlenu, aktywność mikroorganizmów oraz zmiany temperatury w istotny sposób wpływają zarówno na tempo rozkładu struktur kostnych i obecnego w nich DNA. Matthiesen i wsp. wykazali, że degradacja kolagenu i komponentu mineralnego przyspiesza w warunkach wysokiej wilgotności [154]. Niekorzystne warunki środowiskowe prowadzą również do destabilizacji puryn i pirymidyn oraz przyspieszają proces deaminacji, będącej jednym z najczęściej obserwowanych typów uszkodzeń w DNA pochodzącym z materiałów archiwalnych [155]. Dodatkowo, kwaśny odczyn gleby oraz intensywne wahania temperatur sprzyjają hydrolizie DNA i rozbijaniu dłuższych fragmentów genomowych, co odzwierciedla zwiększony indeks degradacji obserwowany w materiale kostnym.

Badania porównawcze fragmentów kostnych pochodzących z różnych części szkieletu wykazały, że ich budowa istotnie wpływa na zachowanie DNA: kości o wysokiej gęstości oraz niskiej porowatości takie jak część skalista kości skroniowej znacznie skuteczniej chronią materiał genetyczny niż kości długie czy żuchwa [156 - 158]. W analizach porównawczych amplifikacji próbek archeologicznych wykonanych przez Pilli i wsp. obejmujących części skaliste kości skroniowych, zęby oraz kości udowe, najwyższy stopień amplifikacji oraz najwyższe stężenia DNA uzyskano właśnie z części skalistych kości skroniowych, a w dalszej kolejności z zębów, natomiast najniższe parametry obserwowano w przypadku kości udowych [159]. Wyniki te pozostają w pełnej zgodności z obserwacjami poczynionymi w niniejszym badaniu, w którym próbki pobrane z części skalistej kości skroniowej charakteryzowały się najwyższą jakością profili genetycznych oraz najniższymi wartościami wskaźnika degradacji, podczas gdy z kości długich nie udało się uzyskać produktów amplifikacji.

Pomimo wysokiej wartości części skalistej kości skroniowej, jej wykorzystanie w badaniach genetycznych może napotykać ograniczenia natury etycznej i konserwatorskiej. Sprzeciw zgłaszany przez rodziny ofiar oraz antropologów wynika przede wszystkim z konieczności fragmentacji czaszki, co wiąże się z nieodwracalnym uszkodzeniem szczątków i klasyfikowane

jest jako procedura inwazyjna. W odpowiedzi na te obawy zaproponowano alternatywne podejście polegające na analizie proszku kostnego uzyskanego w wyniku kontrolowanego wiercenia części skalistej, bez konieczności pobrania całego fragmentu kostnego. Badanie opublikowane przez Sirak i wsp. opisało metodę wwiercania się w podstawę czaszki, pozwalającą na minimalizację uszkodzeń kości przy jednoczesnym zachowaniu wysokiej efektywności izolacji DNA. Autorzy porównali tę technikę z klasycznym pobraniem wypreparowanej części skalistej oraz z materiałem pozaczaszkowym na siedmiu szkieletach, wykazując, że najwyższe ilości DNA uzyskiwano z wypreparowanej części skalistej metodą tradycyjną, natomiast najniższą skuteczność charakteryzował materiał pozaczaszkowy. Jednocześnie podkreślono, że w sytuacjach, w których pełne wypreparowanie części skalistej nie jest możliwe, metoda wiercenia podstawy czaszki stanowi istotnie lepszą alternatywę niż pobieranie próbek z innych fragmentów szkieletu [160].

Znaczenie minimalizacji destrukcji szczątków ludzkich zostało również podkreślone w pracy Alpaslan-Roodenberg i wsp., w której zaproponowano globalne wytyczne etyczne dotyczące badań DNA na szczątkach ludzkich. Autorzy, reprezentujący środowiska antropologów, archeologów i genetyków, wskazali konieczność ograniczania ingerencji w materiał kostny do minimum, szczególnie w kontekście badań szczątków o znaczeniu historycznym i kulturowym [161].

Wyniki uzyskane z analizowanych w niniejszej pracy próbek kostnych wykazały szerokie spektrum wartości DI, od próbek umiarkowanie zdegradowanych po znacznie zdegradowane. W próbkach o wysokim DI obserwowano częstsze występowanie zjawisk typu drop-out szczególnie w obrębie markerów STR o dłuższych amplikonach. Zwiększone ryzyko artefaktów PCR oraz utraty alleli jest charakterystyczne dla śladów typu LT-DNA [108].

Zgodnie z zaleceniami DNA Commission of the ISFG interpretacja profili pochodzących z materiału zdegradowanego wymaga szczególnej ostrożności, z uwzględnieniem progów stochastycznych, prawdopodobieństwa wystąpienia drop-out oraz stosowania metod probabilistycznych przy ocenie wiarygodności wyników [107, 108].

5.2. Ocena skuteczności markerów STR w identyfikacji NN zwłok i szczątków ludzkich

Skuteczność autosomalnych markerów STR w analizie materiału zdegradowanego jest warunkowana przede wszystkim długością amplikonów, odpornością zestawu na działanie inhibitorów oraz jakością zastosowanej chemii PCR. W analizowanym badaniu podstawowym panelem był GlobalFiler™, wykorzystujący 21 autosomalnych STR o zróżnicowanych

długościach ampliconów, w tym kilka loci o charakterze miniSTR, co czyni go dobrze przystosowanym do pracy z materiałem archiwalnym.

Degradacja DNA w materiałach archiwalnych w pierwszej kolejności prowadzi do utraty sygnałów w locus o najdłuższych ampliconach, co jest zgodne z wynikami dotychczasowych badań nad LT-DNA, w których preferencyjna amplifikacja krótszych fragmentów stanowi podstawowe ograniczenie w analizach STR [98, 108]. MiniSTR, czyli markery zaprojektowane tak, aby produkt PCR nie przekraczał 150 pz, wykazują wyraźną przewagę w takich warunkach. W literaturze wykazano, że miniSTR umożliwiają amplifikację DNA o bardzo wysokim stopniu fragmentacji, generując profile bardziej kompletne niż konwencjonalne allele STR o długości 200–450 bp, standardowo stosowane w systemach CODIS i ESS [162].

W świetle tych obserwacji różnica w kompletności profili uzyskiwanych przy użyciu GlobalFiler™ i KinFinder pozostaje w pełni zgodna z przewidywaniami teoretycznymi. Panel GlobalFiler™, zawierający liczne loci zoptymalizowane pod kątem krótkich ampliconów, osiągał wysoką kompletność profili nawet w próbkach o podwyższonym DI, natomiast panel KinFinder okazał się bardziej podatny na drop-out.

Inhibitory obecne w materiale kostnym (m.in. kwas humusowy, jony metali, produkty rozpadu tkanek) mogą znacząco utrudniać przebieg reakcji PCR, prowadząc do uzyskania niepełnych profili lub artefaktów amplifikacji. Zestaw GlobalFiler™ wykazuje wysoką odporność na inhibitory, co potwierdzono w badaniach porównawczych, w których pełne profile uzyskano przy stężeniach inhibitorów przewyższających 150 μ M hematyny i 50–70 ng/ μ L kwasu humusowego [163]. W kontekście analiz materiału archiwalnego, charakteryzującego się często współwystępowaniem niskiej ilości DNA i inhibitorów, jest to istotna zaleta.

Panel KinFinder, jako panel autorski, nie posiada danych walidacyjnych dotyczących tolerancji na inhibitory, a uzyskane wyniki wskazują, że w próbkach o wysokim wskaźniku degradacji i obecności inhibicji amplifikacja w panelu KinFinder ulegała zaburzeniom częściej niż w GlobalFiler™. Różnica ta jest zgodna z mechanizmem degradacji DNA, w którym krótsze amplicony wykazują większą odporność na obecność inhibitorów niż amplicony dłuższe [164]. W ostatnich dekadach rozwój genetyki sądowej znacząco rozszerzył możliwości identyfikacji ludzkich szczątków. Aby wyjść poza ograniczenia standardowych zestawów markerów STR, takich jak 20 loci z Combined DNA Index System (CODIS) czy 13 loci z European Standard Set (ESS) [165], opracowano szereg systemów multipleksowych obejmujących markery spoza tych zestawów [166 - 168]. Panel GlobalFiler™ zapewnia profil o wysokiej wartości informacyjnej w analizach pokrewieństwa pierwszego i drugiego stopnia, jednak w bardziej odległych relacjach może okazać się niewystarczający.

Panel KinFinder, dzięki większej liczbie wysoko polimorficznych markerów STR, ma w założeniu zwiększać moc dyskryminacyjną zestawu, zwłaszcza w złożonych relacjach genealogicznych, gdzie standardowe zestawy markerów wykazują ograniczoną przydatność, lecz w przypadku materiału o wysokiej degradacji jego skuteczność okazała się ograniczona przez wysoki odsetek drop-out.

W próbkach o relatywnie lepszej jakości (niższe DI, brak inhibitorów), panel KinFinder dostarczał dużą ilość dodatkowych danych zwiększających moc statystyczną analiz LR. Natomiast w próbkach o gorszej jakości obserwowano niepełne profile, przez co jego moc dyskryminacji spadała, a do identyfikacji konieczne było zastosowanie paneli SNP.

Jak wskazano w podrozdziale 5.1, w przypadku materiału zdegradowanego szczególnie istotne jest uwzględnienie zjawisk typu drop-out, drop-in i zwiększonego udziału artefaktów amplifikacji [107, 108]. Zjawiska te obserwowano również w markerach STR w niniejszym badaniu i były one znacznie częstsze w panelu KinFinder, co bezpośrednio wynikało z większej długości amplikonów. W panelu GlobalFiler™ zjawiska te występowały rzadziej i obejmowały głównie markery o dłuższych amplikonach, co również pozostaje zgodne z danymi literaturowymi.

5.3. Rola markerów Y-STR w analizie linii ojcowskiej i ich przydatność w identyfikacji szczątków ludzkich

Markery Y-STR stanowią kluczowe narzędzie w analizie pokrewieństwa w linii ojcowskiej, szczególnie w przypadkach, w których dostępne próbki referencyjne pochodzą od członków dalszej rodziny lub gdy autosomalne analizy dostarczają wyników niejednoznacznych. Chromosom Y przekazywany jest niemal niezmieniony z ojca na syna, co umożliwia szybkie wykluczenie lub potwierdzenie zgodności haplotypów, nawet przy braku danych autosomalnych o wystarczającej mocy dyskryminacyjnej. W materiale archiwalnym, silnie zdegradowanym i często charakteryzującym się niską ilością odzyskanego DNA jądrowego, analiza Y-STR może być szczególnie przydatna, ze względu na amplifikację krótszych fragmentów.

Wysoki stopień degradacji DNA prowadzi do fragmentacji genomu i preferencyjnej utraty sekwencji o większej długości amplikonu [99, 100]. W kontekście Y-STR oznacza to, że skuteczność amplifikacji zależy przede wszystkim od długości poszczególnych markerów oraz czułości zastosowanego zestawu. Zestaw Yfiler™ Plus, obejmujący 27 loci Y-STR, zaprojektowano tak, aby zmaksymalizować amplifikację krótkich fragmentów DNA i zwiększyć czułość w analizie próbek zdegradowanych. Dane z literatury potwierdzają,

że Yfiler™ Plus charakteryzuje się dobrą skutecznością również w materiale kostnym o wysokim stopniu degradacji co potwierdziły dane uzyskane w niniejszym badaniu [132].

W niniejszym badaniu profile Y-STR zostały wykorzystane jako narzędzie potwierdzające brak zgodności między szczątkami a rodzinami poszukującymi osoby płci męskiej, jak również do jednoznacznego wykluczania relacji w przypadkach, w których autosomalne profile były niepełne. Zbieżność haplotypu Y jest w takiej sytuacji istotną przesłanką do dalszej analizy statystycznej, zaś niezgodność, przy uwzględnieniu możliwości zajścia mutacji, jest traktowana jako dowód braku pokrewieństwa w linii ojcowskiej.

Yfiler™ Plus obejmuje zarówno markery wolnomutujące, mające istotną wartość informacyjną w analizie populacyjnej, jak i loci o wyższym tempie mutacyjności, przydatne w rozróżnianiu blisko spokrewnionych mężczyzn. W badaniach identyfikacyjnych znaczenie tych ostatnich jest jednak ograniczone z uwagi na ryzyko, że mutacje pojawią się po stronie linii referencyjnej, co może prowadzić do niezgodności profili, przez co należy wykazać szczególną ostrożność w interpretacji pojedynczych niezgodności Y-STR [125].

Nowe generacje paneli Y-STR, takie jak RMplex, umożliwiają znacznie skuteczniejsze różnicowanie osób spokrewnionych w linii męskiej. W badaniach na populacji koreańskiej wykazano, że RMplex pozwolił na odróżnienie aż 50,7% par ojciec–syn na podstawie co najmniej jednej mutacji, co stanowi wartość ponad dziesięciokrotnie wyższą niż w przypadku klasycznego zestawu PowerPlex® Y23 oraz około trzykrotnie wyższą od Yfiler™ Plus. Podobne wyniki uzyskano w badaniach europejskich, gdzie RMplex różnicował 42% par, podczas gdy Yfiler™ Plus — zaledwie 14%. [59, 169]. Te wyniki wskazują, że szybko mutujące markery Y-STR są narzędziem szczególnie wartościowym nie tyle do potwierdzania linii ojcowskiej, ile do odróżniania blisko spokrewnionych mężczyzn. Należy jednak podkreślić, że szybko mutujące markery Y-STR, z uwagi na długość amplikonów wymagają stosunkowo dobrej jakości DNA. Wyniki badań walidacyjnych wskazują na zwiększoną podatność tych markerów na drop-out w przypadku znacznej degradacji materiału. W efekcie, w materiale archiwalnym ich zastosowanie może być ograniczone, natomiast w próbkach referencyjnych mogą stanowić cenne uzupełnienie analizy.

Zjawiska charakterystyczne dla LT-DNA, opisane w rozdziale 5.1 występują również w analizie Y-STR. Zgodnie z zaleceniami DNA Commission of the ISFG interpretacja profili Y-STR uzyskanych z materiału zdegradowanego powinna uwzględniać możliwość mutacji międzygeneracyjnych oraz zwiększoną częstość artefaktów w markerach o dłuższych amplikonach [107, 108].

W niniejszym badaniu obserwowano kompletność profilu Y-STR w większości przypadków. Niezgodności alleli interpretowano jako możliwe mutacje w linii ojcowskiej, a wykluczenie opierano na wynikach obliczeń statystycznych.

W oparciu o uzyskane dane można wyróżnić trzy główne funkcje markerów Y-STR w analizie szczątków z materiału archiwalnego:

- potwierdzenie zgodności linii ojcowskiej tam, gdzie autosomalne LR było niewystarczające.
- wykluczenie pokrewieństwa w linii ojcowskiej - uzyskanie niejednoznacznego wyniku z analizy STR autosomalnych często jest weryfikowane przez profil Y-STR.
- potwierdzenie kierunku analiz w przypadkach niejednoznacznych wyników STR autosomalnych, zwłaszcza przy częściowych profilach.

Jak wskazują badania, obecność zarówno markerów wolnomutujących, jak i szybko mutujących zwiększa jednocześnie zdolność różnicowania jak i ryzyko uzyskania pozornych niezgodności. Mutacje jednokrokowe obserwowane w liniach męskich są naturalną konsekwencją dynamiki markerów Y-STR — co potwierdzono w obu pracach dotyczących RMplex, gdzie 97–98% mutacji stanowiły mutacje jednokrokowe. W świetle tych obserwacji uzyskane w badaniu niewielkie niezgodności Y-STR mogą stanowić naturalną część zmienności haplotypów męskich, a nie wynik błędów analitycznych.

5.4. Wpływ doboru bazy populacyjnej na analizę statystyczną oraz ocenę wartości dowodowej wyników

Właściwy dobór bazy populacyjnej stanowi jeden z ważniejszych elementów analizy statystycznej w badaniach pokrewieństwa i identyfikacji szczątków ludzkich. Wartości częstości alleli wykorzystywane w obliczeniach LR są bowiem specyficzne dla danej populacji, a ich niewłaściwy dobór może prowadzić do błędów interpretacyjnych – różnice w częstości alleli mogą wpływać na końcową wartość LR, szczególnie w przypadkach obejmujących relacje dalszego stopnia lub częściowe profile STR [119, 120].

W badaniach porównawczych wykazano, że bazy populacyjne różnią się pod względem proporcji alleli nie tylko między grupami etnicznymi, lecz również wewnątrz jednego kraju lub regionu. Praca Hedella i wsp. wykazała, że dwie szwedzkie bazy referencyjne - pozornie reprezentujące tę samą populację - wykazywały istotne różnice w częstościach alleli na poziomie pojedynczego locus [170]. Różnice te nie przekładały się istotnie na globalne wskaźniki podobieństwa populacji, jednak miały wpływ na wartość LR w pojedynczych przypadkach, zwłaszcza tam, gdzie allele rzadkie były obecne tylko w jednej z baz.

Brak właściwej populacji referencyjnej może szczególnie utrudniać interpretację wyników w przypadkach, w których LR oscyluje w pobliżu progu decyzyjnego [119]. Dotyczy to m.in. rodzin, których relacje są odległe, niepełne lub obarczone wątpliwościami co do poprawności danych.

Z kolei badania Dash i wsp. wykazują, że wielkość bazy populacyjnej ma bezpośredni wpływ na stabilność estymacji częstości alleli i na występowanie rzadkich alleli w populacji. Analiza czterech zbiorów danych (50, 100, 200 i 500 osób) wykazała, że zwiększenie liczby próbek:

- zwiększa liczbę unikalnych alleli,
- ujawnia rzadkie warianty (w populacji Bahrajnu aż 47 nowych alleli pojawiło się dopiero przy $n = 500$),
- prowadzi do zmian w parametrach takich jak siła dyskryminacji (PD), wskaźnik informacji o polimorfizmie (PIC) i prawdopodobieństwo przypadkowej zgodności (MP) [171].

Dla badań identyfikacyjnych ma to istotne znaczenie, ponieważ małe bazy referencyjne mogą przeszacowywać lub niedoszacowywać częstości rzadkich alleli, co bezpośrednio przekłada się na wartość LR.

W niniejszym projekcie zastosowano bazy częstości alleli zgodnie z dostępnością dla poszczególnych zestawów uwzględniając możliwie najbliższe geograficznie populacje referencyjne: populację Azerbejdżanu dla GlobalFiler™, populację euroazjatycką dla Yfiler™ Plus, bazę ogólnoswiatową dla Precision ID Identity oraz populację polską dla panelu KinFinder. Brak dedykowanych danych dla populacji nie wyklucza możliwości przeprowadzenia wiarygodnej oceny statystycznej, jednak literatura wskazuje, że różnice między populacjami mogą w pewnych przypadkach wpływać na wartości LR, zwłaszcza w analizach z udziałem alleli rzadkich [172]. Hedell i wsp. wykazali, że różnice między bazami mogą prowadzić do zmian LR rzędu jednego–dwóch wielkości, co należy mieć na uwadze przy analizach uzyskanych wyników, szczególnie w sytuacjach uzyskania wartości zbliżonej do granicznej dla rozróżnienia kategorii wyników [170]. W niniejszej pracy najniższy uzyskany wynik wynosił $1,43 \cdot 10^9$, dzięki czemu nawet przy uwzględnieniu obniżenia wartości w związku z użyciem niewłaściwej bazy częstości, uzyskano wartość ponad założonym progiem identyfikacji. Można zatem przyjąć, że wskazane identyfikacje, nawet pomimo braku właściwych baz populacyjnych, nie są fałszywie pozytywne.

5.5. Zastosowanie paneli SNP w identyfikacji NN zwłok i szczątków ludzkich

Postęp technologiczny sekwencjonowania następnej generacji pozwolił na włączenie markerów SNP do badań identyfikacyjnych, w tym w szczególności do analiz pokrewieństwa w dalszych liniach genealogicznych oraz do wspomagania identyfikacji szczątków ludzkich w przypadkach, w których standardowa analiza STR nie dostarcza jednoznacznych wyników [164, 173 - 177]. Panele SNP dedykowane badaniom genealogicznym — zarówno oparte na hybrydyzacji (np. FORCE – występujący zarówno w wersji opartej na hybrydyzacji jak i na amplikonach (QIAGEN)), jak i na amplikonach (np. ForenSeq™ Kintelligence (QIAGEN)) - umożliwiają uzyskanie informacji, które tradycyjne zestawy STR często nie są w stanie dostarczyć w przypadku DNA silnie zdegradowanego, poddanego działaniu warunków środowiskowych lub dostępnego jedynie w ograniczonej ilości [178 - 180].

W niniejszym badaniu, istotne było rozważenie potencjalnej roli paneli SNP jako narzędzia uzupełniającego, zwłaszcza w przypadkach, w których rekonstrukcja pokrewieństwa opiera się na analizie dalekich relacji lub gdy dostępny materiał referencyjny jest ograniczony. Wyniki najnowszych badań wskazują, że zastosowanie tego typu paneli może znacząco zwiększyć prawdopodobieństwo wykrycia zgodności rodzinnej.

Wprowadzenie technik określanych jako FIGG (Forensic Investigative Genetic Genealogy) umożliwiło tworzenie drzew genealogicznych na podstawie IBD. Wymaga to jednak dysponowania dużą liczbą markerów SNP, zwykle od kilkudziesięciu do kilkuset tysięcy, co jest typowe dla komercyjnych mikromacierzy, lecz nie zawsze możliwe do uzyskania z materiału silnie zdegradowanego. Rozwiązaniem tego problemu stało się wykorzystanie paneli o ograniczonej liczbie SNP, przystosowanych do pracy z próbkami niskiej jakości. Przykładem takiego rozwiązania jest zestaw Kintelligence, który obejmuje 9867 informatywnych markerów SNP, zaprojektowany tak, aby generować profile nadające się do porównania z bazami genealogicznymi nawet przy niskiej ilości i jakości DNA. Z badań Daniel i wsp. wynika, że panel ten pozwala poprawnie klasyfikować relacje od pierwszego do piątego stopnia, przy czym relacje szóstego stopnia mogą stanowić wyzwanie, co jest zgodne z oczekiwaniami wynikającymi z teorii dziedziczenia [179]. Metoda ta uchodzi również za odporną na obecność inhibicji oraz degradację materiału, co posiada zasadnicze znaczenie w analizie szczątków archiwalnych.

Alternatywą dla paneli opartych na amplifikacji jest podejście wykorzystujące hybrydyzacyjne wyłapywanie fragmentów DNA, co umożliwia analizę nawet wtedy, gdy amplifikacja

w klasycznych zestawach nie jest możliwa. Przykładem takiego rozwiązania jest panel FORCE, obejmujący ponad 5000 markerów SNP, w tym 3931 SNP przeznaczonych do analiz pokrewieństwa w dalszych liniach genealogicznych. Panel ten został przetestowany z użyciem materiału kostnego pochodzącego z ofiar II wojny światowej, a więc materiału o charakterystyce silnie zdegradowanego DNA. Wyniki wykazały możliwość poprawnego określenia relacji od pierwszego do piątego stopnia, z uzyskaniem wysokich wartości LR oraz wysokiej wiarygodności predykcji [178, 180].

W literaturze wskazuje się, że wykorzystanie szerokich paneli SNP umożliwia identyfikację krewnych do czwartego lub piątego stopnia, nawet w przypadku materiału silnie zdegradowanego. Gorden i wsp. wykazali, że panele obejmujące 25–95 tys. markerów SNP pozwalają na wiarygodne wnioskowanie o pokrewieństwie przy średniej długości fragmentów DNA 32–170 pz. [181]. Udoskonalenia technologiczne, takie jak zastosowanie unikalnych identyfikatorów molekularnych (UMI), istotnie poprawiły czułość analiz SNP, umożliwiając pracę z bardzo niskimi ilościami DNA [180].

W niniejszej pracy zastosowano relatywnie niewielki panel SNP Precision ID Identity, umożliwiający oznaczenie 90 autosomalnych SNP oraz 34 SNP chromosomu Y. Dobór tego panelu wynikał z założeń projektowych ukierunkowanych na maksymalizację efektywności ekonomicznej prowadzonych analiz, przy założeniu, że markery SNP będą pełniły rolę uzupełniającą wobec standardowych oznaczeń autosomalnych STR. Uzyskane wyniki pozostają zgodne z danymi literaturowymi — panel SNP wykazywał wysoką skuteczność w analizie materiału archiwalnego, umożliwiając uzyskanie pełnych lub niemal pełnych profili genetycznych nawet z silnie zdegradowanych próbek, jednocześnie, ze względu na ograniczoną liczbę analizowanych układów, uzyskiwane wartości dowodowe pozostawały relatywnie niskie [65].

5.6. Integracja markerów STR, Y-STR i SNP w procesie identyfikacji NN zwłok i szczątków ludzkich z materiału archiwalnego

W niniejszym projekcie wykorzystano różnorodne narzędzia i techniki w celu maksymalizacji liczby możliwych identyfikacji, niezależnie od stopnia pokrewieństwa. Kompleksowe podejście łączy zaawansowane narzędzia obliczeniowe z rozszerzonym zestawem markerów genetycznych, aby przezwyciężyć ograniczenia standardowej analizy STR w badaniach pokrewieństwa. Zastosowano podejście wieloetapowe: wszystkie próbki od ofiar i krewnych zostały początkowo przeanalizowane z wykorzystaniem 21 autosomalnych markerów STR (GlobalFiler™) oraz 27 markerów Y-STR (Yfiler™ Plus). Przypadki, w których nie uzyskano

wartości $LR > 100\,000$ i jednocześnie nie stwierdzono pełnej zgodności profili Y-STR, zostały następnie ponownie przeanalizowane z wykorzystaniem dodatkowych 48 markerów STR (KinFinder). W przypadkach, które budziły wątpliwości interpretacyjne, zastosowano dodatkowy panel 90 SNP. Schemat ten pozwolił na stopniowe zwiększanie mocy dyskryminacji w przypadkach, w których nie uzyskano wystarczających wartości LR.

Zintegrowane podejście umożliwiło uzyskanie wyników o wysokiej wartości dowodowej nawet w przypadkach, w których pojedyncze systemy markerów okazywały się niewystarczające z powodu degradacji DNA, niepełnych profili lub ograniczonej liczby krewnych dostępnych do porównań. Jak wykazano, włączenie do obliczeń zarówno profili autosomalnych, jak i Y-STR oraz, w wybranych przypadkach, SNP, pozwalało na uzyskanie wyższych wartości ilorazu wiarygodności, zwiększając moc dyskryminacji metody.

Do czynników mogących przyczyniać się do uzyskania wyników niejednoznacznych można zaliczyć: mutacje w STR autosomalnych lub Y-STR [182], niskie lub graniczne wartości LR [183], niekompletne profile NN zwłok [184], odległe relacje biologiczne [185], relacje w których pokrewieństwo może nie być jednoznacznie ustalone (np. najwyższy iloraz wiarygodności odpowiada konfiguracji pokrewieństwa innej niż zadeklarowana) [186] – a także inne niepewności związane z konkretnymi przypadkami.

Wyniki potwierdziły, że w sytuacjach dalekiego pokrewieństwa klasyczne zestawy STR nie dostarczają wystarczającej mocy dowodowej. Podobne wnioski formułowali Zvéigorosky i wsp., wskazując, że brak jednoznacznie określonej hipotezy relacji rodzinnych prowadzi do niespójności statystycznych [187]. W niniejszym badaniu zastosowanie rozszerzonych zestawów markerów, w połączeniu z symulacjami pokrewieństw w środowisku R, pozwoliło ograniczyć tę niepewność.

W przypadku wielu rodzin integracja danych STR autosomalnych i Y-STR miała istotny wpływ na wynik końcowy. Przykładowo, w rodzinach, w których LR dla systemów autosomalnych znajdowało się w zakresie wartości umiarkowanych, zgodność haplotypu Y-STR prowadziła do znaczącego zwiększenia łącznej wartości siły dowodu. W praktyce oznaczało to, że nawet jeśli same STR autosomalne nie dostarczały wartości jednoznacznych, włączenie markerów Y-STR pozwalało na podjęcie decyzji identyfikacyjnej zgodnej z kryteriami stosowanymi w genetyce sądowej.

We wszystkich analizowanych rodzinach gdzie możliwe było zastosowanie panelu Y-STR, został on wykorzystany w celu weryfikacji wskazań uzyskanych z przeszukania DVI z wykorzystaniem profili STR autosomalnych. W jednym przypadku dzięki oznaczeniu profilu Y-STR nie doszło do uzyskania wyniku fałszywie negatywnego, który zostałby uzyskany przy

wykorzystaniu jedynie panelu GlobalFiler™. W przypadku Rodziny 9 analiza markerów chromosomu Y oraz markerów płci stanowiła ważny czynnik potwierdzający brak zgodności z zakładaną płcią poszukiwanej osoby.

Zastosowanie panelu SNP miało znaczenie uzupełniające – zgodnie z przedstawionym schematem decyzyjnym, panel ten wykorzystywano w sytuacjach, w których relacje rodzinne były odległe lub autosomalne STR nie zapewniały wystarczającej mocy dowodowej. W takich przypadkach SNP wzmacniały końcową ocenę statystyczną, zwiększając wartość LR lub ograniczając liczbę możliwych scenariuszy pokrewieństwa.

Porównanie uzyskanych wyników wskazało, że konfiguracje obejmujące system Precision ID Identity nie zawsze dawały najwyższe wartości LR w połączeniu z innymi zestawami markerów, a konfiguracje obejmujące GlobalFiler™ i Precision ID Identity nigdy nie generowały najwyższych wartości. Wynikało to z faktu, że Precision ID Identity stosowano w najtrudniejszych przypadkach, wymagających dodatkowej weryfikacji. Pomimo że panel SNP w tym systemie zapewniał pełne profile nawet w silnie zdegradowanym materiale, jego wartość informatywna dla bardziej odległych relacji była ograniczona. W takich sytuacjach szersze panele SNP, obejmujące dziesiątki tysięcy markerów, prawdopodobnie mogą dostarczać bardziej wiarygodnych wyników [173].

Pomimo że literatura sugeruje istotny wpływ sprzężeń genetycznych na analizę statystyczną przy rozszerzonych zestawach markerów wyniki niniejszego badania wskazują, że uwzględnienie sprzężeń prowadziło do niewielkich zmian wartości LR [138]. W większości analiz różnice mieściły się w granicach jednego rzędu wielkości lub były niższe, a w wielu przypadkach skutkowały nawet nieznacznym wzrostem LR (np. wzrost z 330 014 do 438 688 dla Rodziny 5; z 8 639 do 9 439 dla Rodziny 6).

Przeprowadzone analizy pokazały również, że interpretacja wyników w przypadku loci wykazujących sprzężenie wymaga szczególnej ostrożności. Programy wykorzystane w niniejszej pracy, mimo korzystania z odmiennych założeń modelowych, borykają się z podobnymi ograniczeniami dotyczącymi sposobu uwzględniania sprzężeń. W praktyce może dojść do sytuacji, w której trzy loci znajdują się w niewielkiej odległości genetycznej od siebie, przez co powinny być traktowane jako układy powiązane ze sobą [139, 188]. Jednak ze względu na konstrukcję programów możliwe jest jedynie zadeklarowanie sprzężenia pomiędzy dwoma loci, podczas gdy trzeci, mimo iż biologicznie również sprzężony, zostaje pominięty w analizie, w konsekwencji czego w pewnych przypadkach może to prowadzić do częściowej utraty informacji i wpływać na ostateczną wartość LR. Ograniczenia te odzwierciedlają szersze wyzwanie opisywane w literaturze, zgodnie z którym niepełne hipotezy pokrewieństwa

lub uproszczone założenia genetyczne mogą prowadzić do wyników niejednoznacznych lub wprowadzających w błąd [14, 185, 186]. Ponadto program FamLink wykazuje najlepsze wyniki przy wykorzystaniu pokrewieństw zdefiniowanych przez twórców, w konsekwencji czego nie uzyskano wyników sprzężenia alleli dla niektórych bardziej skomplikowanych pokrewieństw.

W praktyce oznacza to, że chociaż wykorzystywanie sprzężeń jest uznawane za „dobrą praktykę” w analizach STR, to jego realny wpływ na wynik końcowy w badaniach szczątków zdegradowanych, charakteryzujących się częściowymi profilami i wysokim poziomem drop-out, jest ograniczony. Może wynikać to z faktu, że sprzężenia mają największe znaczenie w przypadku pełnych profili i dużej liczby wspólnie dziedziczonych markerów, natomiast w materiale kostnym uzyskuje się zwykle nieregularne pokrycie poszczególnych loci. Dlatego w niniejszym badaniu analizy wykonano przy założeniu niezależności dziedziczenia, a wyniki obliczeń uwzględniających sprzężenia zostały wykorzystane tylko jako narzędzie pomocnicze.

5.7. Ocena skuteczności zastosowanych narzędzi statystycznych oraz weryfikacja założeń genealogicznych

W niniejszym badaniu wykorzystano kilka uzupełniających się narzędzi analitycznych do obliczania ilorazu wiarygodności oraz oceny relacji pokrewieństwa: Familias, KinBN, pedbuildr oraz FamLink. Każde z tych narzędzi wykorzystuje odmienny model statystyczny i różny poziom tolerancji dla niekompletności danych, co w przypadku pracy z materiałem silnie zdegradowanym ma zasadniczy wpływ na uzyskane wartości LR [134, 141, 147].

Skuteczność poszczególnych narzędzi nie wynika wyłącznie z algorytmu statystycznego, ale przede wszystkim z tego, jak oprogramowanie radzi sobie z częściowymi profilami, z rozbieżnościami w kompletności profili między osobami oraz z uwzględnianiem mutacji.

Familias dopuszcza brakujące dane u różnych osób oraz stosuje różne modele mutacyjne. Program ten radzi sobie z sytuacjami gdy profile referencyjne różnią się stopniem kompletności lub gdy badany materiał zawierał liczne drop-outy.

KinBN stosuje podejście bayesowskie i wymaga, aby wszystkie osoby miały identyczny zestaw markerów wejściowych. Oznacza to, że program nie wykona analizy gdy chociaż jedna osoba ma niekompletne dane genetyczne – nawet pojedynczy brak allelu wyklucza możliwość uwzględnienia danego markera, a profile muszą zostać ograniczone do wspólnie w pełni oznaczonych układów, co zmniejsza ilość dostępnej informacji.

W wielu rodzinach skutkowało to znacznym obniżeniem wartości LR w stosunku do programu Familias, co wynika nie z gorszego algorytmu, lecz z redukcji liczby markerów użytych w obliczeniach.

Dodatkowo, model KinBN jest bardziej wrażliwy na obecność rzadkich alleli, nie można wprowadzić do obliczeń allele którego nie ma w bazie częstości.

Pedbuildr jako narzędzie z rodziny pedsuite koncentruje się przede wszystkim na rekonstrukcji drzew genealogicznych, symulacjach IBD, a co za tym idzie na ocenie zgodności struktury rodzinnej z danymi obserwowanymi. Program ten nie służy do obliczeń prawdopodobieństwa pokrewieństwa ani ilorazu wiarygodności, jednak metoda ta jest cenna przy rekonstrukcji rodzin o niejasnej strukturze genealogicznej. Do głównych ograniczeń należy zaliczyć to iż podobnie do KinBN wymaga pełnych profili wsadowych. W efekcie pedbuildr dobrze sprawdza się w analizach genealogicznych, lecz nie może zastąpić klasycznego LR w sensie dowodowym.

Wszystkie ograniczenia programów statystycznych pokazują, że w pracy z archiwalnym materiałem wybór narzędzia ma realny wpływ na wynik, a stosowanie jednego oprogramowania może prowadzić do zaniżenia lub przeszacowania siły dowodowej hipotezy. Wykorzystane programy pozwoliły również na zidentyfikowanie błędnych deklaracji pokrewieństwa w dostarczonych przez rodziny informacjach o pokrewieństwie. Jest to szczególnie istotne przy analizowaniu odległego pokrewieństwa w sprawach archiwalnych, gdzie relacje rzeczywiste – z uwagi na upływ czasu i odległe relacje rodzinne mogą różnić się od relacji przedstawianych przez członków rodziny.

Wykorzystanie zaawansowanych metod biostatystycznych i symulacyjnych okazało się kluczowe – symulacje przeprowadzone w pakiecie pedbuildr umożliwiły identyfikację błędnych założeń genealogicznych oraz rozpoznanie relacji, które w klasycznych analizach mogłyby zostać pominięte. Kwestia niespójności z deklarowanym pokrewieństwem jest przedmiotem dyskusji naukowej od wielu lat [189, 190], DNA Commission of the ISFG zaleca, aby takich informacji nie ujawniać [191]. Badania sugerują, że niezgodności między pokrewieństwem biologicznym a deklarowanym mogą występować w niewielkim odsetku rodzin [192], a prawdopodobieństwo ich wykrycia rośnie wraz z liczbą analizowanych generacji i wzrostem rozdzielczości analitycznej.

W sprawach identyfikacyjnych bardzo często zdarza się, że pokrewieństwo pomiędzy NN zwłokami i szczątkami ludzkimi a materiałem porównawczym jest niepewne lub nawet nie jest znane [120]. Zdarzają się również przypadki gdy wykonując badania identyfikacyjne uzyskiwany jest wynik negatywny, a po wykonaniu dokładniejszej analizy okazuje się, że

zakładany stopień pokrewieństwa różnił się od realnie występującego pomiędzy analizowanymi osobami. Różnica ta niekoniecznie oznacza brak pokrewieństwa pomiędzy osobą referencyjną, ofiarą lub pozostałą częścią rodziny, lecz może odzwierciedlać inny stopień pokrewieństwa, który mógł ulec zatarciu na przestrzeni pokoleń. Jest to zjawisko częste w sprawach, w których od zdarzenia upłynęły dekady — w tym przypadku od lat 1937–1938. W sytuacjach masowych represji lub kontekstu politycznego część rodzin nie posiadała wiedzy o losach krewnych, a niektóre informacje nie były przekazywane następnym pokoleniom. W takich okolicznościach sporadyczne różnice pomiędzy relacjami deklarowanymi a biologicznymi są spodziewane. Stosując wskazaną w pracy metodę analizy można ograniczyć ryzyko popełnienia błędu przy skomplikowanych sprawach.

5.8. Analiza haplotypów Y i pochodzenia biogeograficznego

Istotnym elementem pracy była także analiza haplogrup chromosomu Y, która dostarczyła mocnych dowodów na to, że analizowany pochówek reprezentuje grupę pochodzenia gruzińskiego, a nie próbę wieloetniczną z szerszego obszaru Kaukazu, przy czym jako dominujące haplogrupy zidentyfikowano G2a i J2 [130]. Wynik ten jest spójny z historycznymi przekazami dotyczącymi represji w Adżarii, które dotknęły głównie miejscową ludność gruzińską [193].

Choć etap ten nie stanowił krytycznego elementu procesu identyfikacji, dostarczył on istotnych danych kontekstowych potwierdzających zgodność genetyczną badanej grupy z przewidywanym pochodzeniem ofiar, a jego zastosowanie było szczególnie uzasadnione w przypadku, gdy tło historyczne stanowiło główne źródło danych odniesienia. Ponadto analiza haplogrup Y została wykonana bez dodatkowych kosztów, gdyż opierała się wyłącznie na markerach użytych w innym etapie identyfikacji.

5.9. Znaczenie badań dla praktyki identyfikacyjnej i możliwości wdrożenia modelu

Wyniki uzyskane w niniejszym badaniu mają kluczowe znaczenie dla praktyki identyfikacyjnej NN zwłok i szczątków ludzkich z materiału archiwalnego oraz dla projektowania przyszłych procedur analitycznych w przypadkach masowych grobów, w których materiał biologiczny jest w wysokim stopniu zdegradowany, a dostępność próbek referencyjnych ograniczona. Zaprezentowany model – oparty na zintegrowanej analizie markerów STR autosomalnych, Y-STR oraz selektywnie stosowanych paneli SNP – stanowi propozycję możliwego standardu

postępowania w tego typu projektach, łącząc wysoką skuteczność analityczną, rozsądną ekonomiczność oraz odporność na ograniczenia jakościowe materiału kostnego.

Przeprowadzone analizy wskazują, że autosomalne profile STR pozostają bardzo efektywnym i praktycznym narzędziem identyfikacyjnym, szczególnie w kontekście dużych projektów obejmujących materiał o różnym stopniu zachowania. Wynika to z kilku czynników:

- zestawy takie jak GlobalFiler™ są powszechnie stosowane, zwalidowane, kompatybilne z międzynarodowymi bazami populacyjnymi oraz wykazują wysoką tolerancję na degradację.
- niski koszt uzyskania pojedynczego profilu - pełny panel STR pozostaje tańszy niż większość obecnie stosowanych paneli SNP.

Zastosowanie rozszerzonych paneli STR dodatkowo zwiększa liczbę informatywnych markerów, co może być metodą znacznie bardziej opłacalną i skuteczną niż rutynowe sięganie po SNP. W badaniu wykazano, że większa liczba markerów STR zapewnia większą wartość informacyjną niż stosowanie niedużych paneli SNP.

Markery Y-STR okazały się jednym z ważniejszych elementów prezentowanego modelu, umożliwiając:

- szybkie wykluczenie relacji w linii ojcowskiej,
- znaczące podniesienie wartości LRc,
- weryfikację błędnych deklaracji rodzinnych.

W wielu przypadkach dopiero połączenie STR autosomalnych z Y-STR pozwoliło osiągnąć wartości LR spełniające kryteria jednoznacznej identyfikacji.

Oznacza to, że zestawy Y-STR pozostają nieodzownym komponentem praktycznego modelu identyfikacyjnego, szczególnie tam, gdzie liczba krewnych jest ograniczona lub gdy dostępni są wyłącznie mężczyźni potomkowie.

Wyniki analizy SNP potwierdzają ich przydatność w przypadkach trudnych, zwłaszcza w odległych relacjach genealogicznych lub tam, gdzie badanie STR nie dostarczało wystarczającej liczby markerów. Jednak badanie wykazało również, że stosowanie paneli SNP o niewielkiej liczbie markerów nie poprawia znacząco mocy analitycznej w projektach identyfikacyjnych opartych na materiale archiwalnym.

W konsekwencji modelem najbardziej optymalnym – zarówno logistycznie, jak i ekonomicznie – pozostaje podejście, w którym panele STR stanowią fundament identyfikacji, Y-STR pełnią funkcję narzędzia wspomagającego, a SNP stosuje się w przypadkach uzasadnionych, np. przy dalszych relacjach lub niejednoznacznych profilach STR.

Na podstawie uzyskanych wyników możliwe jest sformułowanie praktycznego modelu, który może stać się rekomendowanym standardem dla projektów identyfikacyjnych obejmujących materiał kostny o niskiej jakości.

5.10. Socjologiczny i etyczny wymiar badań

Identyfikacja ofiar masowych represji w Adżarii stanowi proces o głębokim znaczeniu społecznym, etycznym i historycznym. W przypadku zbrodni popełnionych w latach 1937–1938, których skutki odczuwane są w Gruzji do dziś, działania te nie ograniczają się do rekonstrukcji tożsamości biologicznej - obejmują również przywracanie ciągłości rodzinnej, odbudowę pamięci zbiorowej oraz dokumentowanie naruszeń praw człowieka. W społeczeństwie, które przez dziesięciolecia było pozbawione dostępu do informacji o losie ofiar represji, proces ujawniania prawdy ma fundamentalne znaczenie dla kształtowania tożsamości narodowej oraz budowania zaufania do instytucji państwowych i międzynarodowych.

Rekonstrukcja tożsamości ofiar działań totalitarnych jest jednym z fundamentalnych elementów prowadzących do symbolicznej sprawiedliwości i uznania cierpienia przeszłych pokoleń. Dostarcza rodzinom odpowiedzi, na które w wielu przypadkach czekano od blisko 90 lat, a równocześnie przeciwdziała procesowi zapominania, utrwalając wiedzę o zbrodniach państwa.

Istotnym aspektem socjologicznym jest również integracja rodzin i społeczności lokalnych w proces identyfikacji. Zbieranie materiału porównawczego wymagało współpracy z osobami, które przez całe życie pozostawały w niewiedzy o losie swoich krewnych. Aktywne zaangażowanie mieszkańców regionu, wsparcie organizacji społecznych oraz zainteresowanie mediów stworzyły nowe przestrzenie dialogu i społecznego przepracowania traum historycznych. W efekcie projekt identyfikacyjny stał się nie tylko przedsięwzięciem naukowym, lecz także platformą wzmacniającą lokalne poczucie wspólnoty i odpowiedzialności za pamięć o ofiarach [194, str. 324-361, 244-259].

W wymiarze etycznym identyfikacja ofiar masowych zbrodni wiąże się z koniecznością zachowania szczególnej wrażliwości. Dotyczy to zarówno prowadzenia badań genetycznych, jak i publicznej prezentacji wyników. Każdy etap musi uwzględniać godność ofiar, prywatność rodzin oraz odpowiedzialność wynikającą z możliwości ujawnienia nieoczekiwanych informacji o relacjach biologicznych. W sytuacjach, w których analiza ujawnia rozbieżności pomiędzy deklarowaną a biologiczną strukturą pokrewieństwa, priorytetem pozostaje ochrona

rodzin przed potencjalną stygmatyzacją oraz przestrzeganie międzynarodowych standardów poufności [9, 10].

Projekt identyfikacyjny w Adżarii posiada również podstawowy wymiar humanitarny. Zgodnie z praktyką GAFA i GRDIP, odnalezione i zidentyfikowane osoby podlegają godnemu pochówkowi, uwzględniającemu tradycje religijne rodzin. Dla wielu społeczności jest to pierwszy moment od dekad, w którym możliwe jest zakończenie procesu żałoby [13].

Wyniki przedstawionego projektu stanowią jedynie pierwszy etap szerszego programu badawczego obejmującego lokalizację i analizę kolejnych masowych grobów z okresu Wielkiego Terroru. W rejonie Batumi i okolicznych miejscowości nadal istnieje wiele niezbadanych lokalizacji, w których, zgodnie z przekazami historycznymi, mogą znajdować się kolejne ofiary represji. Doświadczenie zdobyte podczas analiz pierwszego grobu stworzy podstawy metodologiczne do prowadzenia identyfikacji na pełną skalę w przyszłych wykopaliskach. Uzyskane wyniki potwierdzają skuteczność zastosowanej strategii genetycznej w bardzo trudnych warunkach oraz podkreślają znaczenie skoordynowanych działań w zakresie pozyskiwania i weryfikacji materiału porównawczego.

Wyniki niniejszej pracy podkreślają, że integracja klasycznych i nowoczesnych technik genetycznych, wsparta analizą biostatystyczną, pozwala uzyskać wiarygodne wyniki identyfikacyjne nawet przy pracy z materiałem zdegradowanym. Co istotne, większość zastosowanych metod może być realizowana przy użyciu aparatury powszechnie dostępnej w laboratoriach genetyki sądowej, co czyni zaproponowane rozwiązanie realnym do wdrożenia w praktyce. Niniejsze badanie dowodzi, że identyfikacje NN zwłok i szczątków ludzkich z materiałów archiwalnych mogą być skutecznie prowadzone w oparciu o zintegrowany model wykorzystujący systemy STR, Y-STR i SNP oraz symulacje pokrewieństw, stanowiąc modelowe rozwiązanie dla innych projektów o charakterze historycznym i humanitarnym. Opracowany model działania może zostać wykorzystany dla kolejnych etapów działań GRDIP.

6. Wnioski

1. Opracowany w niniejszej pracy schemat decyzyjny, umożliwia efektywne wykorzystanie materiału DNA o zróżnicowanym stopniu degradacji, typowy dla NN zwłok i szczątków ludzkich. Zastosowane podejście, obejmujące analizy autosomalnych markerów STR, markerów Y-STR, rozszerzonych zestawów STR oraz markerów SNP, pozwoliło na przezwycięzenie ograniczeń standardowych procedur identyfikacyjnych i zwiększenie liczby przypadków, w których możliwe było uzyskanie rozstrzygających wyników.
2. Przeprowadzona analiza wykazała zróżnicowaną skuteczność poszczególnych zestawów markerów genetycznych w zależności od stopnia degradacji materiału kostnego. Zastosowany standardowy zestaw markerów genetycznych (GlobalFiler™) dostarczał dobrej jakości profile genetyczne, jednakże z uwagi na daleki stopień pokrewieństwa pomiędzy analizowanymi ofiarami a materiałem porównawczym, uzyskiwane wyniki były niewystarczające by uzyskać identyfikację. Zastosowany rozszerzony panel markerów (KinFinder) był zauważalnie bardziej wrażliwy na jakość DNA, jednakże ze względu na wysoką informatywność oznaczanych markerów stanowił istotne uzupełnienie zestawu standardowego. Panel SNP wykazał bardzo dobrą odporność na niską jakość DNA, dostarczając pełne lub prawie pełne profile genetyczne, jednakże z uwagi na ich stosunkowo niewielką ilość (90) i informatywność stanowił on przede wszystkim uzupełnienie zestawów standardowego i rozszerzonego w sytuacjach, gdy oba te zestawy nie pozwalały na uzyskanie wystarczających wyników. Zintegrowane podejście okazało się skuteczne w przypadku próbek o niskiej zawartości DNA, gdzie konwencjonalne metody zazwyczaj nie umożliwiają uzyskania rozstrzygających wyników, szczególnie w przypadkach gdy jako materiał porównawczy dostępni są jedynie krewni dalszego stopnia.
3. Symulacje pokrewieństw wykonane w środowisku R przy użyciu pakietu pedbuildr umożliwiły odwzorowanie struktur pokrewieństwa i ocenę wiarygodności wyników dla różnych scenariuszy genealogicznych, stanowiąc wartościowe narzędzie wspomagające interpretację wyników analiz genetycznych w sprawach identyfikacyjnych. Zastosowanie symulacji umożliwiło ocenę wiarygodności uzyskiwanych wartości statystycznych dla alternatywnych scenariuszy pokrewieństwa. Analiza przypadków rodzinnych ujawniła, że rozbieżności w deklaracjach genealogicznych mogą być istotnym źródłem błędnych interpretacji, dlatego konieczna jest weryfikacja zakładanego stopnia pokrewieństwa pomiędzy analizowanymi osobami.

4. Zastosowanie kilku niezależnych programów statystycznych do obliczeń pokrewieństwa pozwoliło na porównanie uzyskiwanych wyników oraz ocenę ich spójności. Uzyskane wartości LR charakteryzowały się wysoką zgodnością pomiędzy zastosowanymi narzędziami, natomiast obserwowane niewielkie różnice wynikały głównie z odmiennych modeli statystycznych, sposobów implementacji danych populacyjnych oraz przyjętych założeń obliczeniowych. Wyniki te potwierdzają, że zasadne jest stosowanie programów do obliczeń statystycznych, jednakże wybór konkretnego narzędzia statystycznego leży w gestii osoby wykonującej analizy.

5. Na podstawie uzyskanych wyników, w sprawach identyfikacji genetycznej NN zwłok i szczątków ludzkich zarekomendowano schemat postępowania przedstawiony w rozdziale 4.23. W szczególności wskazano zasadność łączenia analiz markerów STR i SNP z zaawansowaną oceną statystyczną. Przedstawiony sposób postępowania został praktycznie zweryfikowany w trakcie badań identyfikacyjnych ofiar Wielkiego Terroru na terenie Gruzji. Uzyskane wyniki potwierdzają możliwość skutecznej identyfikacji genetycznej nawet po upływie kilkudziesięciu lat od zgonu, pod warunkiem zastosowania kompleksowego podejścia łączącego metody STR, analizę biostatystyczną i modelowanie pokrewieństwa. Opracowana konfiguracja metod i narzędzi zapewnia optymalny kompromis pomiędzy kosztem analiz, ilością uzyskiwanej informacji genetycznej oraz precyzją obliczeń pokrewieństwa, umożliwiając skuteczną identyfikację genetyczną ofiar historycznych.

7. Piśmiennictwo

- [1] Khlevnyuk, O. (1995). The Objectives of the Great Terror, 1937–1938. In: Cooper, J., Perrie, M., Rees, E.A. (eds) *Soviet History, 1917–53*. Palgrave Macmillan, London. DOI: 10.1007/978-1-349-23939-9_7.
- [2] Luarsabishvili, V., A brief history of the great terror in Georgia, *Revista de Historia Actual*, tom 13, nr 14-15, str. 175-183, 2017.
- [3] Tasarz, P. (2022). Funkcjonowanie systemów totalitarnych – wybrane aspekty. *Res Politicae*, 14, str. 141-162. DOI: 10.16926/rp.2022.14.08.
- [4] Werth, N. (2013). Przemyśleć na nowo lata Wielkiego Terroru. W T. Kizny, *Wielki Terror 1937–1938* (str. 181). Warszawa.
- [5] González Calleja, E., (2012) *Los Totalitarismos* (str. 74). Sintesis Madryd.
- [6] Avalishvili, L., Melkonian, E., Ismailov, E., Anchabadze, G. (2010). Stalinist Terror in the South Caucasus. *Caucasus Analytical Digest* (22), DOI: 10.3929/ethz-a-006256278.
- [7] Alawerdaszvili, D., Kukhalaszvili, E., Luarsabiszwili, W., Koller, S., Majewski, M. i Tuszuraszvili, O. (2016). *Wielki Terror w Sowieckiej Gruzji 1937-1938 represje wobec Polaków Polacy w Gruzji - Gruzini w Polsce* (Tom 1). Instytut Pamięci Narodowej.
- [8] IDFI Exhibition - Repression of 1937-1938 in Georgia -Stalin (2012) <https://idfi.ge/en/exhibition-repression-of-1937-1938-in-georgia-stalin>, (17.10.2025).
- [9] Al Jazeera Thrown like animals’: Georgians identify victims in Stalin’s mass graves, 2024, <https://www.aljazeera.com/features/2024/1/28/georgias-mass-graves-the-forensic-experts-uncovering-victims-of-stalins-purge>, (25.05.2025).
- [10] ICRC Humanity after Life: Respecting and Protecting the Dead (2019). <https://www.icrc.org/en/document/humanity-after-life-respect-and-protection-dead>, 15.05.2025
- [11] Mchedlishvili, D. A. (2008). Listy stalinowskie z Gruzji. <http://www.nplg.gov.ge/gwdict/index.php?a=index&d=26>. (15.05.2025).
- [12] Georgian Recovery, Documentation, and Identification Project (2025). <https://gafanthro.github.io/GRDIP>, (15.05.2025).
- [13] IDFI Batumi Uncovering the Truth Mass Graves of the Great Terror (1937-1938) in Batumi/Georgia (2024). [https://idfi.ge/en/theses_international_conference_uncovering_the_truth_mass_graves_of_the_great_terror_\(1937_1938\)_in_batumi_georgia](https://idfi.ge/en/theses_international_conference_uncovering_the_truth_mass_graves_of_the_great_terror_(1937_1938)_in_batumi_georgia), (15.05.2025).
- [14] Budowle, B., Bieber, F., Eisenberg, A. (2005). Forensic aspects of mass disasters: Strategic considerations for DNA-based human identification. *Legal Medicine*, 7(4), str. 230-243. DOI: 10.1016/j.legalmed.2005.01.001.

- [15] Butler, J. M. (2005). *Forensic DNA Typing: Biology & Technology Behind STR Markers*. Elsevier Books.
- [16] Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Molecular Breeding*, 2(3), 225-238. DOI: 10.1007/BF00564200.
- [17] Mullis, K. (1990). The Unusual Origin of the Polymerase Chain Reaction. *Scientific American*, 262(4), str. 56-65. DOI: 10.1038/scientificamerican0490-56.
- [18] Butler, J. M. (2010). *Fundamentals of Forensic DNA Typing*. Academic Press. DOI: 10.1016/C2009-0-01945-X.
- [19] Ossowski, A., Diepenbroek, M., Kupiec, T., Bykowska-Witowska, M., Zielińska, G., Dembińska, T., Ciechanowicz, A. (2016). Genetic Identification of Communist Crimes' Victims (1944-1956) Based on the Analysis of One of Many Mass Graves Discovered on the Powazki Military Cemetery in Warsaw, Poland. *Journal of Forensic Sciences*, 61(6), str. 1450-1455. DOI: 10.1111/1556-4029.13205.
- [20] Ossowski, A., Diepenbroek, M., Kupiec, T., Bykowska, M., Zielińska, G., Jasiński, M., March, A. (2016). The Polish Genetic Database of Victims of Totalitarianisms. *Forensic Science International*, 258, str. 41-49. DOI: 10.1016/j.forsciint.2015.10.029.
- [21] Kuś, M., Ossowski, A., Zielińska, G. (2015). Example of genetic identifications of victims of totalitarian regimes in Poland. *Forensic Science International Genetics Supplement Series*, str. e356-e357. DOI:10.1016/j.fsigss.2015.09.141.
- [22] Schoske, R., Vallone, P., Ruitberg, C., Butler, J. (2003). Multiplex PCR design strategy used for the simultaneous amplification of 10 Y chromosome short tandem repeat (STR) loci. *Analytical and Bioanalytical Chemistry*, 375(3), str. 333-343. DOI:10.1007/s00216-002-1683-2.
- [23] Szibor, R. (2007). X-chromosomal markers: past, present and future. *Forensic Science International Genetics*, 1(2), str. 93-99. DOI:10.1016/j.fsigen.2007.03.003.
- [24] Amorim, A., Fernandes, T., Taveira, N. (2019). Mitochondrial DNA in human identification: a review. *Peer J*, str. e7314. DOI:10.7717/peerj.7314.
- [25] Tillmar, A., Kling, D., Butler, J., Parson, W., Prinz, M., Shneider, P., Egeland, T., Gusmao, L. (2017). DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis. *Forensic Science International: Genetics*, 29, str. 269-275. DOI: 10.1016/j.fsigen.2017.05.005.
- [26] Damato, E., Benjeddou, M., Davison, S. (2009). Evaluation of 21 Y-STRs for population and forensic studies. *Forensic Science International: Genetics Supplement Series*, 2(1), str. 446-447. DOI: 10.1016/j.fsigss.2009.08.091.

- [27] INTERPOL Disaster Victim Identification Guide, 2023, https://www.interpol.int/content/download/589/file/18Y1344%20%20E%20DVI_Guide.pdf, (13.11.2025)
- [28] Berent, J. (2003). Biostatystyczna interpretacja badań genetycznych w dochodzeniu ojcostwa dla przypadków bliskiego pokrewieństwa pozwanego mężczyzny i ojca biologicznego. Łódź : Akademia Medyczna.
- [29] Li, Y. C., Korol, A., Fahima, T., Beiles, A., Nevo, E. (2002). Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology*, 11(12), str. 2453-2465. DOI:10.1046/j.1365-294x.2002.01643.x.
- [30] Kit, S. (1961). Equilibrium sedimentation in density gradients of DNA preparations from animal tissues. *Journal of Molecular Biology*, 3(6), str. 711-716. DOI: 10.1016/S0022-2836(61)80075-2.
- [31] Litt, M., Luty, J. A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics*, 44(3), str. 397-401.
- [32] Weller, P., Jeffreys, A. J., Wilson, V., Blanchetot, A. (1984). Organization of the human myoglobin gene. *EMBO Journal*, 3(2), str. 439-446. DOI:10.1002/j.1460-2075.1984.tb01825.x.
- [33] Zhang, Q., Zhou, Z., Wang, L., Quan, C., Liu, Q., Tang, Z., Liu, L., Liu, Y., Wang, S. (2020). Pairwise kinship testing with a combination of STR and SNP loci. *Forensic Science International Genetics*, 46, str. 102265. DOI:10.1016/j.fsigen.2020.102265.
- [34] Wilkenning, S., Chen, B., Hemminki, K., Forsti, A. (2006). STR markers for kinship analysis. *Human Biology*, 78(1), str. 1-8. DOI:10.1353/hub.2006.0030.
- [35] Schlötterer, C. (2000). Evolutionary dynamics of microsatellite DNA. *Chromosoma Focus* (109), str. 365-371. DOI: 10.1007/s004120000089.
- [36] Xu, L., Haas, R., Sun, J., Zhou, Y., Bickhart, D., Li, J., Song, J., Sonstegard, T., Van Tassell, C., Lewin, H., Liu, G. (2017). Systematic profiling of short tandem repeats in the cattle genome. *Genome biology and evolution*, 9(1), str. 20-31. DOI: 10.2527/jam2016-0309.
- [37] Muneer, A. (2014). Application of Microsatellite Markers in Conservation Genetics and Fisheries Management: Recent Advances in Population Structure Analysis and Conservation Strategies. *Genetic Research International*, str. 691759. DOI:10.1155/2014/691759.
- [38] Takezaki, N., Nei, M. (2009). Genomic drift and evolution of microsatellite DNAs in human populations. *Molecular Biology and Evolution*, 26(8), str. 1835-1840. DOI:10.1093/molbev/msp091.

- [39] Slatkin, M., Racimo, F. (2016). Ancient DNA and human history. *Proceedings of the National Academy of Science*, 113(23), str. 6380-6387. DOI: 10.1073/pnas.1524306113.
- [40] Boland, R., Goel, A. (2010). Microsatellite instability in colorectal cancer. *Gastroenterology*, 138(6), str. 2073-2087. DOI:10.1053/j.gastro.2009.12.064.
- [41] Brouwer, J., Willemsen, R., Oostra, B. (2009). Microsatellite repeat instability and neurological disease. *Bioessays*, 31(1), str. 71-83. DOI:10.1002/bies.080122.
- [42] Hodel, R., Segova-Salcedo, C., Landis, J., Crowl, A., Sun, M., Liu, X., Gitzendanner, M., Douglas, N., Germain-Aubrey, C., Chen, S., Soltis, D., Soltis, P. (2016). The report of my death was an exaggeration: A review for researchers using microsatellites in the 21st century. *Applications in Plant Sciences*, 4(6), str. 1600025. DOI:10.3732/apps.1600025.
- [43] Ellegren, H. (2004). Microsatellites: simple sequences with complex evolution. *Nature Reviews Genetics*, 5(6), str. 435-445. DOI:10.1038/nrg1348.
- [44] Evett, I., Pope, S., Puch-Solis, R. (2016). Providing scientific guidance on DNA to the judiciary. *Science & Justice*, 56(4), str. 278-281. DOI: 10.1016/j.scijus.2016.05.001.
- [45] Cliffe, K. M., Day, A. E., Bagga, M., Siggens, K., Quilter, C. R., Lowden, S., Finlayson, H. A., Palgrave, C. J., Li, N., Huang, L., Blott, S. C., Sargent, C. A. (2010). Analysis of the non-recombining Y chromosome defines polymorphisms in domestic pig breeds: ancestral bases identified by comparative sequencing. *Animal Genetics*, 41(6), str. 619-629. DOI: 10.1111/j.1365-2052.2010.02070.x.
- [46] Ginalski, K., Rychlewski, L., Baker, D., Grishin, N. (2004). Protein structure prediction for the male-specific region of the human Y chromosome. *Proceedings of the National Academy of Sciences*, 101(8), str. 2305-2310. DOI:10.1073/pnas.0306306101.
- [47] Jain, M., Olsen, H., Turner, D., Stodart, D., Bulazel, K., Paten, B., Haussler, D., Huntington, W., Akeson, M., Miga, K. (2018). Linear assembly of a human centromere on the Y chromosome. *Nature Biotechnology*, 36(4), str. 321-323. DOI:10.1038/nbt.4109.
- [48] Krausz, C. G., Casamonti, E. (2017). Spermatogenic failure and the Y chromosome. *Human Genetics*, 136(5), str. 637-655. DOI:10.1007/s00439-017-1793-8.
- [49] Calacal, G. C., Delfin, F. C., Tan, M. M., Roewer, L., Magtanong, D. L., Lara, M. C., Fortun, R., Corazon A De Ungria, M. (2005). Identification of exhumed remains of fire tragedy victims using conventional methods and autosomal/Y-chromosomal short tandem repeat DNA profiling. *The American Journal of Forensic Medicine and Pathology*, 26(3), str. 285-291. DOI: 10.1097/01.paf.0000177338.21951.82

- [50] Dettlaf-Kąkol, A., Pawłowski, R. (2002). First Polish DNA "manhunt"-an application of Y-chromosome STRs. *International Journal of Legal Medicine*, 116(5), str. 289-291. DOI:10.1007/s00414-002-0320-0.
- [51] Roewer, L. (2009). Y chromosome STR typing in crime casework. *Forensic Science, Medicine, and Pathology*, 5(2), str. 77-84. DOI:10.1007/s12024-009-9089-5.
- [52] Gusmão, L., Butler, J. M., Carracedo, A. M., Gill, P., Kayser, M., Mayr, W. R., Morling, N., Prinz, M. K., Roewer, L., Tyler-Smith, C., Shneider, P. M. (2006). DNA Commission of the International Society of Forensic Genetics (ISFG): An update of the recommendations on the use of Y-STRs in forensic analysis. *Forensic Science International*, 157(2-3), str. 187-197. DOI: 10.1016/j.forsciint.2005.04.002.
- [53] Yao, H., Wen, S., Tong, X., Zhou, B., Du, P., Shi, M., Jin, L., Li, H. (2016). Y chromosomal clue successfully facilitated the arrest of Baiyin serial killer. *Science Bulletin*, 61, str. 1715-1717. DOI: 10.1007/s11434-016-1183-y.
- [54] Qian, X., Hou, J., Wang, Z., Ye, Y., Lang, M., Gao, T., Liu, J., Hou, Y. (2017). Next Generation Sequencing Plus (NGS+) with Y-chromosomal Markers for Forensic Pedigree Searches. *Scientific Reports*, 7, str. 11324. DOI: 10.1038/s41598-017-11955-x.
- [55] Thermo Fisher Scientific, 2025, <https://documents.thermofisher.com/TFS-Assets/LSG/manuals/MAN0030230-YfilerPlusPCRAmpKit-UG.pdf>, (25.10.2025).
- [56] Promega, 2025, <https://pl.promega.com/resources/protocols/technical-manuals/d0/powerplex-y23-system-for-use-on-the-applied-biosystems-genetic-analyzers-protocol/>, (24.10.2025).
- [57] Qiagen, 2025, <https://www.qiagen.com/us/products/human-id-and-forensics/investigator-solutions/investigator-argus-y-28-qs-kit>, (25.10.2025).
- [58] Willuweit, S., Roewer, L. (2015). The new Y Chromosome Haplotype Reference Database. *Forensic Science International: Genetics*, 15, str. 43-48. DOI:10.1016/j.fsigen.2014.11.024.
- [59] Neuhuber, F., Dunkelmann, B., Griebner, I., Helm, K., Kayser, M., Ralf, A. (2022). Improving the differentiation of closely related males by RMplex analysis of 30 Y-STRs with high mutation rates. *Forensic Science International: Genetics*, 58, str. 102682. DOI: 10.1016/j.fsigen.2022.102682.
- [60] Gomes, I., Pinto, N., Antao-Sousa, S., Gomes, V., Gusmao, L., Amorim, A. (2020). Twenty Years Later: A Comprehensive Review of the X Chromosome Use in Forensic Genetics. *Frontiers in Genetics*, 11, str. 926. DOI:10.3389/fgene.2020.00926.

- [61] Kapińska, E., Wysocka, J., Cybulska, L., Rębała, K., Juchniewicz, P., Szczerkowska, Z. (2012). Examples of application of X chromosomal markers in familial investigations and personal identification. *Archiwum Medycyny Sądowej i Kryminologii*, 62(3), str. 152-159.
- [62] Branicki, W., Kupiec, T., Wolańska-Nowak, P. (2008). *Badania DNA dla celów sądowych*. Kraków: Instytut ekspertyz sądowych.
- [63] Huang, E., Liu, C., Zheng, J., Han, X., Du, W., Huang, Y., Li, C., Wang, X., Tong, D., Ou, X., Sun, H., Zeng, Z., Liu, C. (2018). Genome-wide screen for universal individual identification SNPs based on the HapMap and 1000 genomes databases. *Scientific Reports*, 8, str. 5553. DOI: 10.1038/s41598-018-23888-0.
- [64] Grandell, I., Samara, R., Tillmar, A. O. (2016). A SNP panel for identity and kinship testing using massive parallel sequencing. *International Journal of Legal Medicine*, 130(4), str. 905-914. DOI:10.1007/s00414-016-1341-4.
- [65] Butler Getting, K., Tillmar, A., Sturk-Andreaggi, K., Marshall, C. (2024). Review of SNP assays for disaster victim identification: Cost, time, and performance information for decision-makers. *Journal of Forensic Science*, 69(5), str. 1546-1557. DOI:10.1111/1556-4029.15585.
- [66] Taanman, J. W. (1999). The mitochondrial genome: structure, transcription, translation and replication. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*, 1410(2), str. 103-123. DOI: 10.1016/S0005-2728(98)00161-3.
- [67] Igawa, K., Manabe, Y., Oyamada, J., Kitagawa, Y., Kato, K., Ikematsu, K., Nakasano, I., Matsushita, T., Rokutanda, A. (2009). Mitochondrial DNA analysis of Yayoi period human skeletal remains from the Doigahama site. *journal of human genetics*, 54, str. 581-588. DOI: 10.1038/jhg.2009.81.
- [68] Holland, M. M., Fisher, D. L., Mitchell, L. G., Rodriquez, W. C., Canik, J. J., Merril, C. R., Weedn, V. W. (1993). Mitochondrial DNA sequence analysis of human skeletal remains: identification of remains from the Vietnam War. *Journal of Forensic Sciences*, 38(3), str. 542-553, DOI: 10.1520/JFS13439J
- [69] Goodwin, S., McPherson, J. D., McCombie, R. W. (2016). Coming of age : ten years of next- generation sequencing technologies. *nature reviews genetics*, 17, str. 333-351. DOI: 10.1038/nrg.2016.49.
- [70] St. John, J. C. (2014). Mitochondrial DNA. Its transmission from Gametes and Embryos. Cibelli, W J., Gurdon, J., Wilmut, I., Jaenisch, R., Lanza, R., West, M. D., Campbell, K. H., *Principles of Cloning Second Edition* (str. 429-439). DOI:10.1016/C2010-0-66663-6.

- [71] Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J., Green, R. E. (2010). Computational challenges in the analysis of ancient DNA. *Genome Biology*, 11, str. R47. DOI: 10.1186/gb-2010-11-5-r47.
- [72] Dabney, J., Knapp, M., Glocke, I., Gansuaga, M. T., Weihmann, A., Nickel, B., Valdiosera, C., Garcia, N., Pääbo, S., Arsuaga, J. L., Meyer, M. (2013). Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. *Proceedings of the National Academy of Sciences*, 110(39), str. 15758-15763. DOI:10.1073/pnas.1314445110
- [73] White, D. J., Wolf, J. N., Pierson, M., Gemmell, N. J. (2008). Revealing the hidden complexities of mtDNA inheritance. *Molecular Ecology*, 17(23), str. 4925-4942. DOI: 10.1111/j.1365-294X.2008.03982.x.
- [74] Meyer, S., Weiss, G., von Haeseler, A. (1999). Pattern of Nucleotide Substitution and Rate Heterogeneity in the Hypervariable Regions I and II of Human mtDNA. *Genetics*, 152(3), str. 1103-1110. DOI: 10.1093/genetics/152.3.1103.
- [75] Davis, C., Peters, D., Warhauser, D., King, J., Budowle, B. (2015). Sequencing the hypervariable regions of human mitochondrial DNA using massively parallel sequencing: Enhanced data acquisition for DNA samples encountered in forensic testing. *Legal Medicine*, 17(2), str. 123-127. DOI:10.1016/j.legalmed.2014.10.004
- [76] Gonçalves, F. T., Cardena, M. M., Gonzalez, R. S., Krieger, J. E., Pereira, A. D., Fridman, C. (2011). The discrimination power of the hypervariable regions HV1, HV2 and HV3 of mitochondrial DNA in the Brazilian population. *Forensic Science International: Genetics Supplement Series*, 3(1), str. e311-e312. DOI: 10.1016/j.fsigs.2011.09.018
- [77] Hwa, H. L., Ko, T. M., Chen, Y. C., Lin, C. Y., Huang, Y. H., Tseng, L. H., Su, N. Y., Lee, J. C. (2012). Sequence polymorphisms of mtDNA HV1, HV2 and HV3 regions in eight population groups living in Taiwan. *Australian Journal of Forensic Sciences*, 44, str. 243-252. DOI: 10.1080/00450618.2011.650208.
- [78] Grzybowski, T., Pawłowski, R., Kupiec, T., Branicki, W., Jacewicz, R. (2018). Recommendations of the Polish Speaking Working Group of the International Society for Forensic Genetics for forensic mitochondrial DNA testing. *Archiwum Medycyny Sądowej i Kryminologii*, 68(4), str. 242-258. DOI:10.5114/amsik.2018.84532.
- [79] Lyons, E. A., Scheible, M. K., Sturk-Andreaggi, K., Irwin, J. A., Just, R. S. (2013). A high throughput Sanger strategy for human mitochondrial genome sequencing. *BMC Genomics*, 14, str. 881. DOI: 10.1186/1471-2164-14-88.

- [80] Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, 74(12), str. 5463-5467. DOI: 10.1073/pnas.74.12.5463
- [81] Bruijns, B., Tiggelaar, R., Gardeniers, H. (2018). Massively parallel sequencing techniques for forensics: A review. *Electrophoresis*, 39(21), str. 2642-2654. DOI:10.1002/elps.201800082.
- [82] Kim, N. Y., Lee, H. Y., Park, S. J., Yang, W. I., Shin, K. J. (2013). Modified midi- and mini-multiplex PCR systems for mitochondrial DNA control region sequence analysis in degraded samples. *Journal of Forensic Science*, 58(3), str. 738-743. DOI:10.1111/1556-4029.12062.
- [83] Chong, D. M., Calloway, C. D., Klein, S. B., Orrego, C., Buoncristiani, M. R. (2005). Optimization of a duplex amplification and sequencing strategy for the HVI/HVII regions of human mitochondrial DNA for forensic casework. *Forensic Science International*, 154(2-3), str. 137-148. DOI: 10.1016/j.forsciint.2004.09.128.
- [84] Parson, W., Gusmão, L., Hares, D. R., Irwin, J. A., Mayr, W. R., Morling, N., Pokorak, W. G., Prinz, M. K., Salas, A., Shneider, P. M., Parson, T. J. (2014). DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing. *Forensic Science International: Genetics*, 13, str. 134-142. DOI: 10.1016/j.fsigen.2014.07.010
- [85] Luo, S., Valencja, A. C., Zhang, J., Lee, N. C., Slone, J., Gui, B., Wang, Li, Z., Dell, S., Brown, J., Chen, S. M., Chien, Y. H., Hwu, W. L., Fan, P. C., Wong, L. J., Atwal, P. S., Huang, T. (2018). Biparental Inheritance of Mitochondrial DNA in Humans. *Proceedings of the National Academy of Sciences*, 115(51), str. 13039-13044. DOI: 10.1073/pnas.1810946115.
- [86] Rock, A. W., Dur, A., van Oven, M., Parson, W. (2013). Concept for estimating mitochondrial DNA haplogroups using a maximum likelihood approach (EMMA). *Forensic Science International Genetics*, 7(6), str. 601-609. DOI:10.1016/j.fsigen.2013.07.005.
- [87] Solé-Morata, N., Garcia-Fernandez, C., Urasin, V., Bekada, A., Fadhlouzi-Zid, K., Zalloua, P., Comas, D., Calafell, F. (2017). Whole Y-chromosome sequences reveal an extremely recent origin of the most common North African paternal lineage E-M183 (M81). *Scientific Report*, 7(1), str. 15941. DOI: 10.1038/s41598-017-16271-y.
- [88] NEVGEN Gentula, M. C., Nevski, A. (2025), <https://www.nevgen.org/>, (15.10.2025).
- [89] Ely, B., Wilson, J. L., Jackson, F., Jackson, B. A. (2006). African-American mitochondrial DNAs often match mtDNAs found in multiple African ethnic groups. *BMC Biology*, 4, str. 34. DOI: 10.1186/1741-7007-4-34.

- [90] Fagundes, N. J., Kanitz, R., Bonatto, S. L. (2008). A Reevaluation of the Native American MtDNA Genome Diversity and Its Bearing on the Models of Early Colonization of Beringia. *PLOS One*, 3(9), str. e3157. DOI:10.1371/journal.pone.0003157.
- [91] Cardinali, I., Bodner, M., Capodiferro, M. R., Amory, C., Migliore, N. R., Gomez, E. J., Myagmar, E., Dashzeveg, T., Carano, F., Woodward, S. R., Parson, W., Perego, U. A., Lancioni, H., Achilli, A. (2022). Mitochondrial DNA Footprints from Western Eurasia in Modern Mongolia. *Frontiers Genetics*, 12(8), str. 819337. DOI:10.3389/fgene.2021.819337.
- [92] Derenko, M., Malyarchuk, B., Denisova, G., Perkova, M., Rogalla, U., Grzybowski, T., Khusnutdinova, E., Dambueva, I., Zakharov, I. (2012). Complete Mitochondrial DNA Analysis of Eastern Eurasian Haplogroups Rarely Found in Populations of Northern Asia and Eastern Europe. *PLOS One*, 7(2), str. e32179. DOI: 10.1371/journal.pone.0032179.
- [93] Battaglia, V., Grugni, V., Perego, U. A., Angerhofer, N., Gomez-Palmieri, E. J., Woodward, S. R., Achilli, A., Myres, N., Torroni, A., Semino, O. (2013). The First Peopling of South America: New Evidence from Y-Chromosome Haplogroup Q. *PLOS One*, 8(8), str. e71390. DOI:10.1371/journal.pone.0071390.
- [94] Chacón-Duque, J. C., Adhikari, K., Fuentes-Guajardo, M., Mendoza-Revilla, J., Acuña-Alonzo, V., Barquera, R., Quinto-Sánchez, M., Gómez-Valdés, J., Everardo Martínez, P., Villamil-Ramírez, H., Hünemeier, T., Ramallo, V., de Cerqueira, C. S., Hurtado, M., Villegas, V., Granja, V., Villena, M., Vásquez, R., Llop, E., Sandoval, J., Salazar-Granara, A. (2018). Latin Americans show wide-spread Converso ancestry and imprint of local Native ancestry on physical appearance. *nature communications*, 9, str. 5388. DOI: 10.1038/s41467-018-07748-z.
- [95] Raszeja, S., Nasiłowski, W., Markiewicz, J. (1993). *Medycyna sądowa. Podręcznik dla studentów*. Państwowy Zakład Wydawnictw Lekarskich.
- [96] Wescott, D. J. (2019). Postmortem change in bone biomechanical properties: Loss of plasticity. *Forensic Science International*, 300, str. 164-169. DOI: 10.1016/j.forsciint.2019.04.017.
- [97] Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, 362, str. 709-715. DOI: 10.1038/362709a0.
- [98] Pokines, J., Symes, S. A. (2013). *Manual of Forensic Taphonomy*. CRC Press.
- [99] Hofreiter, M., Jaenicke, V., Serre, D., von Haeseler, A., Pääbo, S. (2001). DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. *Nucleic Acids Research*, 29(23), str. 4793-4799. DOI: 10.1093/nar/29.23.4793.
- [100] Marota, I., Basile, C., Ubaldi, M., Rollo, F. (2002). DNA decay rate in papyri and human remains from Egyptian archaeological sites. *American Journal of Biological Anthropology*, 117(4), str. 310-318. DOI:10.1002/ajpa.10045.

- [101] Zheng, J., Li, X., Shan, D., Zhang, H., Guan, D. (2012). DNA degradation within mouse brain and dental pulp cells 72 hours postmortem. *Neural Regeneration Research*, 7(4), str. 290-294. DOI:10.3969/j.issn.1673-5374.2012.04.009.
- [102] Liu, L., Shu, X., Ren, L., Zhou, H., Li, Y., Liu, W., Zhu, C., Liu, L. (2007). Determination of the early time of death by computerized image analysis of DNA degradation: Which is the best quantitative indicator of DNA degradation? *Journal of Huazhong University of Science and Technology*, 27(4), str. 362-366. DOI:10.1007/s11596-007-0404-7.
- [103] Gilbert, T., Hansen, A. J., Willerslev, E., Rudbeck, L., Barnes, I., Lynnerup, N., Cooper, A. (2003). Characterization of Genetic Miscoding Lesions Caused by Postmortem Damage. *American Journal of Human Genetics*, 72(1), str. 48-61. DOI:10.1086/345379.
- [104] Gilbert, T., Willerslev, E., Hansen, A., Barnes, I., Rudbeck, L., Lynnerup, N., Cooper, A. (2003). Distribution Patterns of Postmortem Damage in Human Mitochondrial DNA. *American Journal of Human Genetics*, 72(1), str. 32-47. DOI:10.1086/345378.
- [105] Hoss, M., Jaruga, P., Zastawny, T. H., Dizdaroğlu, M., Paabo, S. (1996). Damage and DNA Sequence Retrieval from Ancient Tissues. *Nucleic Acids Research*, 24(7), str. 1304-1307. DOI: 10.1093/nar/24.7.1304.
- [106] Graham, E. A. (2007). DNA reviews: Ancient DNA. *Forensic Science, Medicine and Pathology*, 3(3), str. 221-225. DOI:10.1007/s12024-007-9009-5.
- [107] Gill, P., Gusmão, L., Mayr, W. R., Morling, N., Hande, H., Parson, W., Prieto, L., Prinz, M., Schneider, H., Schneider, P., Weir, B. (2012). DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods. *Forensic Science International: Genetics*, 6(6), str. 679-688. DOI: 10.1016/j.fsigen.2012.06.002.
- [108] Parys-Proszek, A., Marcińska, M., Branicki, W., Pawłowski, R., Kupiec, T., Grzybowski, T., Woźniak, M., Spólnicka, M., Jacewicz, R. (2020). Badanie śladów LT-DNA – przegląd literatury i ogólne zalecenia Polskojęzycznej Grupy Roboczej Międzynarodowego Towarzystwa Genetyki Sądowej (ISFG-PL). *archiwum medycyny sądowej i kryminologii*, 70(2-3), str. 103-123. DOI: 10.5114/amsik.2020.104489.
- [109] INTERPOL How We Work, 2025, <https://www.interpol.int/How-we-work/Forensics/Disaster-Victim-Identification-DVI>, (11.11.2025).
- [110] Szwagrzyk, K. (2017). Poszukiwanie ofiar systemu komunistycznego w Polsce. Doświadczenia i perspektywy. *Ochrona Zabytków*, 2, str. 99-108.

- [111] Brough, A. L., Morgan, B., Rutty, G. N. (2015). The basics of disaster victim identification. *Journal of Forensic Radiology and Imaging*, 3(1), str. 29-37. DOI: 10.1016/j.jofri.2015.01.002.
- [112] Lessig, R., Rothschild, M. (2012). International standards in cases of mass disaster victim identification (DVI). *Forensic Science, Medicine, and Pathology*, 8(2), str. 197-199. DOI:10.1007/s12024-011-9272-3.
- [113] SWGDAM guidelines for reporting likelihood ratios, 2025, <https://www.swgdam.org/publications>, (11.11.2025).
- [114] Gill, P., Fereday, L., Morling, N., Shneider, P. M. (2006). The evolution of DNA databases—recommendations for new European STR loci. *Forensic Science International*, 156(2-3), str. 242-244. DOI: 10.1016/j.forsciint.2005.05.036.
- [115] Buckleton, J., Triggs, C. M., Walsh, S. J. (2005). *Forensic DNA Evidence Interpretation*. CRC Press.
- [116] Butler, J. M. (2015). *Advanced Topics in Forensic DNA Typing: Interpretation*. San Diego: Academic Press.
- [117] Gjertson, D. W., Brenner, C. H., Baur, M. P., Carracedo, A., Guidet, F., Luque, J. A., Lessig, R., Mayr, W. R., Pascali, V. L., Prinz, M., Shneider, P. M., Morling, N. (2007). ISFG: Recommendations on biostatistics in paternity testing. *Forensic Science International: Genetics*, 1(3-4), str. 223-231. DOI:10.1016/j.fsigen.2007.06.006.
- [118] Bodner, M., Bastish, I., Butler, J. M., Fimmers, R., Gill, P., Gusmao, L., Morling, N., Phillips, C., Prinz, M., P. Shneider, M., Parson, W. (2016). Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). *Forensic Science International Genetics*, 24, str. 97-102. DOI:10.1016/j.fsigen.2016.06.008.
- [119] Shneider, P. M. (2007). Scientific standards for studies in forensic genetics. *Forensic Science International*, 165(2-3), str. 238-243. DOI:10.1016/j.forsciint.2006.06.067.
- [120] Jabłońska-Milczarek, M., Frankowski, A. (2020). Latest Interpol reports on disaster victim identification and the process of implementing international DVI standards in Poland. *Archiwum Medycyny Sądowej i Kryminologii*, 70(2-3), str. 163-180. DOI:10.5114/amsik.2020.104493.
- [121] Aliyeva, A., Gellatly, V., Abbasov, M., Iyengar, A. (2021). Population genetic data for 21 autosomal STR loci in the Azerbaijani population using the GlobalFiler™ kit. *International Journal of Legal Medicine*, 13S, str. 1789-1791. DOI: 10.1007/s00414-021-02610-2.

- [122] Kling, D., Tillmar, A., Egeland, T. (2014). Familias 3 – Extensions and new functionality. *Forensic Science International: Genetics*, 13, str. 121-127. DOI: 10.1016/j.fsigen.2014.07.004.
- [123] SWGDAM Interpretation Guidelines for Y-Chromosome STR Typing by Forensic DNA Laboratories, 2022, <https://www.swgdam.org/publications>, (11.11.2025).
- [124] Roewer, L., Andersen, M. M., Ballantyne, J., Butler, J., Caliebe, A., Corach, D., D'Amato, M. E., Gusmão, L., Hou, Y., de Knijff, P., Parson, W., Prinz, M., Schneider, P., Taylor, D., Vennemann, M., Willuweit, S. (2020). DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis. *Forensic Science International: Genetics*, 48, str. 102308. DOI: 10.1016/j.fsigen.2020.102308.
- [125] Rębała, K., Branicki, W., Pawłowski, R., Spólnicka, M., Kupiec, T., Parys-Proszek, A., Woźniak, M., Grzybowski, T., Boroń, M., Wróbel, M., Ciesielka, M., Ossowski, A., Jacewicz, R. (2020). Zalecenia Polskojęzycznej Grupy Międzynarodowego Towarzystwa Genetyki Sądowej dotyczące analizy polimorfizmu chromosomu Y dla celów sądowych. *Archiwum Medycyny Sądowej i Kryminologii*, 70(1), str. 1-18. DOI: 10.5114/amsik.2020.97833.
- [126] SWGDAM Interpretation Guidelines for Single Nucleotide Polymorphism (SNP) Analysis by Forensic DNA Testing Laboratories, 2024, <https://www.swgdam.org/publications>, (11.11.2025).
- [127] Michels, V., Thibodeau, S., Thorland, E. (2005). Mendelian and Mitochondrial Inheritance, Gene Identification, and Clinical Testing: Genetic Linkage. W V. Michels, S. Thibodeau i E. Thorland, *Peripheral Neuropathy* (str. 1545-1560), DOI: 10.1016/B978-0-7216-9491-7.X5001-7.
- [128] Holt, A., Wootton, S. C., Mulero, J. J., Brzoska, P. M., Langit, E. i Green, R. (2016). Developmental validation of the Quantifiler® HP and Trio Kits for human DNA quantification in forensic samples. *Forensic Science International: Genetics*, 21, str. 145-157. DOI: 10.1016/j.fsigen.2015.12.007.
- [129] Łuczak, W. (2024). Opracowanie zestawu polimorficznych markerów STR do genotypowania ludzi oraz jego wdrożenie do analiz pokrewieństwa w dalszych relacjach rodzinnych. Rozprawa Doktorska .
- [130] Schurr, T. G., Shengelia, R., Shamooun-Pour, M., Chitanava, D., Laliashvili, S., Laliashvili Irma, Kibret, R., Kume-Kangkolo, Y., Akhvlediani, I., Bitadze, L., Mathieson, L., Yardumian, A. (2023). Genetic Analysis of Mingrelians Reveals Long-Term Continuity of

Populations in Western Georgia (Caucasus). *Genome Biology and Evolution*, 15(11), str. 198. DOI: 10.1093/gbe/evad198.

[131] Nasidze, I., Schädlich, H., Stoneking, M. (2003). Haplotypes from the Caucasus, Turkey and Iran for nine Y-STR loci. *Forensic Science International*, 137(1), str. 85-93. DOI:10.1016/s0379-0738(03)00272-x.

[132] Szargut, M., Diepenbroek, M., Zielińska, G., Cytacka, S., Arciszewska, J., Jałowińska, K., Piątek, J., Ossowski, A. (2019). Is MPS always the answer? Use of two PCR-based methods for Y-chromosomal haplotyping in highly and moderately degraded bone material. *Forensic Science International: Genetics*, 42, str. 181-189. DOI: 10.1016/j.fsigen.2019.07.016.

[133] Drabek, J. (2009). Validation of software for calculating the likelihood ratio for parentage and kinship. *Forensic Science International: Genetics*, 3, str. 112-118. DOI:10.1016/j.fsigen.2008.11.005.

[134] Kling, D., Mostad, P., Egeland, T. (2017). Manual for Familias 3, <https://familias.name/Files/manualFamilias.pdf>, (13.11.2025).

[135] Simonsson, I., Mostad, P. (2016). Stationary mutation models. *Forensic Science International: Genetics*, 23, str. 217-225. DOI: 10.1016/j.fsigen.2016.04.005.

[136] Yang, D., Li, M. L., Wei, A. Q. (2024). Evaluating the effects of silent genes on pairwise kinship testing. *Forensic Science International*, 365, str. 112290. DOI: 10.1016/j.forsciint.2024.112290.

[137] Kling, D., Egeland, T., Tillmar, A. (2012). FamLink – A user friendly software for linkage calculations in family genetics. *Forensic Science International: Genetics*, 6, str. 616-620. DOI:10.1016/j.fsigen.2012.01.012.

[138] Yang, D., Ma, S. X., Zhao, G. L., Gao, A., Xu, Z. K. (2024). Determining the effects of genetic linkage when using a combination of STR and SNP loci for kinship testing. *Legal Medicine*, 69, str. 102441. DOI: 10.1016/j.legalmed.2024.102441.

[139] Tillmar, A., Phillips, C. (2017). Evaluation of the impact of genetic linkage in forensic identity and relationship testing for expanded DNA marker sets. *Forensic Science International: Genetics*, 26, str. 58-65. DOI: 10.1016/j.fsigen.2016.10.007.

[140] O'Connor, K., Tillmar, A. (2012). Effect of linkage between vWA and D12S391 in kinship analysis. *Forensic Science International: Genetics*, 6, str. 840-844. DOI:10.1016/j.fsigen.2012.03.008.

[141] Kling, D., Mostad, P., Tillmar, A. (2025). FamLink2 – A comprehensive tool for likelihood computations in pedigrees analyses involving linked DNA markers accounting for

- genotype uncertainties. *Forensic Science International: Genetics*, 74, str. 103150. DOI: 10.1016/j.fsigen.2024.103150.
- [142] Nothnagel, M., Schmidtke, J., Krawczak, M. (2010). Potentials and limits of pairwise kinship analysis using autosomal short tandem repeat loci. *International Journal of Legal Medicine*, 124, str. 205-215. DOI:10.1007/s00414-009-0413-0.
- [143] Egeland, T., Slooten, K. (2016). The likelihood ratio as a random variable for linked markers in kinship analysis. *International Journal of Legal Medicine*(130), str. 1445-1456. DOI: 10.1007/s00414-016-1416-2.
- [144] Buckleton, J., Triggs, C. (2006). The effect of linkage on the calculation of DNA match probabilities for siblings and half siblings. *Forensic Science International*, 160, str. 193-199. DOI:10.1016/j.forsciint.2005.10.004.
- [145] Egeland, T., Sheehan, N. (2008). On identification problems requiring linked autosomal markers. *Forensic Science International: Genetics*, 2, str. 219-225. DOI:10.1016/j.fsigen.2008.02.006.
- [146] Morimoto, C., Tsuji, H., Manabe, S., Fujimoto, S., Hirai, E., Hamano, Y., Tamaki, K. (2020). Development of a software for kinship analysis considering linkage and mutation based on a Bayesian network. *Forensic Science International: Genetics*(47), str. 102279. DOI: 10.1016/j.fsigen.2020.102279.
- [147] Morimoto, C. (2024). KinBN v2.1.0 user manual, <https://github.com/ChieMorimoto/KinBN/blob/master/inst/manual/KinBN%20v2.1.0%20user%20manual.pdf>, (14.10.2025).
- [148] Chatre, V. (2013). How to calculate recombination fraction from Kosambi distance and vice versa. Southern Institute of Forest Genetics. Pobrano z lokalizacji <https://vc.popgen.org/articles/Kosambi.pdf>.
- [149] Kirkpatrick, B., Li, S. C., Karp, R., Halperin, E. (2011). Pedigree Reconstruction Using Identity by Descent. *Journal of Computational Biology*, 1811, str. 1481-1493. DOI:10.1089/cmb.2011.0156.
- [150] Vigeland, M. D. (2020). Relatedness coefficients in pedigrees with inbred founders. *Journal of Mathematical Biology*(81), str. 185-207. DOI: 10.1007/s00285-020-01505-x.
- [151] Wei, Y., Zhu, Q., Wang, H., Cao, Y., Li, X., Zhang, X., Wang, Y., Zhang, J. (2024). Pairwise kinship inference and pedigree reconstruction using 91 microhaplotypes. *Forensic Science International: Genetics*(72), str. 103090. DOI: 10.1016/j.fsigen.2024.103090.
- [152] Gavashelishvili, A., Yanchukov, A., Tarkhnishvili, D., Murtskhvaladze, M., Akhvlediani, I., Kazanci, C. (2021). Landscape genetics and the genetic legacy of Upper

Paleolithic and Mesolithic hunter-gatherers in the modern Caucasus. *Scientific Reports*, 11, str. 17985. DOI: 10.1038/s41598-021-97519-6.

[153] Gavashelishvili, A., Chukhua, M., Sakhltkhutsishvili, K., Koptekin, D., Somel, M. (2023). The time and place of origin of South Caucasian languages: insights into past human societies, ecosystems and human population genetics. *Scientific Reports*, 13, str. 21133. DOI: 10.1038/s41598-023-45500-w.

[154] Matthiesen, H., Høier Eriksen, A. M., Hollesen, J., Collins, M. (2021). Bone degradation at five Arctic archaeological sites: Quantifying the importance of burial environment and bone characteristics. *Journal of Archaeological Science*, 125, str. 105296. DOI:10.1016/j.jas.2020.105296.

[155] HøierEriksen, A., KjærgaardNielsen, T., Matthiesen, H., Carøe, C., Hestbjerg Hansen, L., Gregory, D. J., Turner-Walker, G., Collins, M. J., Thomas, M., Gilbert, P. (2020). Bonebiodeterioration—The effect of marine andterrestrial depositional environments on early diagenesis and bone bacterial community. *PloSONE*, 15, str. e0240512. DOI:10.1371/journal.pone.0240512.

[156] Ibrahim, J., Brumfeld, V., Addadi, Y., Rubin, S., Weiner, S., Boaretto, E. (2022). Thepetrous bonecontains high concentrations of osteocytes: One possible reason whyancient DNAisbetter preserved in this bone. *PlosONE*, 17(10), str. e0269348. DOI:10.1371/journal.

[157] Kravanja, P., Golob, A., Concato, M., Leskovar, T., Zupanic Pajnic, I. (2025). Effects of different environmental factors on preservation of DNA in petrous bones: A comparative study of two Slovenian archaeological sites. *Forensic Science International*, 371, str. 112495. doi:doi.org/10.1016/j.forsciint.2025.112495.

[158] Haarkötter, C., Vinueza-Espinosa, D., Gálvez, X., Saiz, M., Medina-Lozano, M. I., Lorente, J. A., Alvarez, J. C. (2023). Acomparisonbetweenpetrous bone andtooth, femur and tibia DNA analysis from degraded skeletal remains. *ELECTROPHORESIS*, 44, strong 1559-1568. DOI:10.1002/elps.202300097.

[159] Pilli, E., Vai, S., Caruso, M. G., D'Errico , G., Berti, A., Caramelli, D. (2018). Neither femur nor tooth: Petrous bone for identifying archaeological bone samples via forensic approach. *Forensic Science International*, 283, strong 144-149. DOI:10.1016/j.forsciint.2017.12.023

[160] Sirak, K. A., Fernandes, D. M., Cheronet, O., Novak, M., Gamarra, B., Balassa, T., Bernert, Z., Cseki, A., Dani, J., Gallina, J. Z., Kocsis-Buruzs, G., Kovari, I., Laszlo, O., Pap, I., Patay, R., Petkes, Z., Szenthe, G., Szeniczey, T., Hajdu T., Pinhasi, R. (2017). A minimally-

invasive method for sampling human petrous bones from the cranial base for ancient DNA analysis. *Biotechniques*, 62(6), strony 283-289. DOI:10.2144/000114558

- [161] Alpaslan-Roodenberg, S., Anthony, D., Babiker, H., Bánffy, E., Booth, T., Capone, P., Deshpande-Mukherjee, A., Eisenmann, S., Fehren-Schmitz, L., Frachetti, M., Fujita, R., Frieman, C. J., Fu, Q., Gibbon, V., Haak, W., Hajdinjak, M., Hofmann, K., Holguin, B., Inomata, T., Kanzawa-Kiriyama, H., Keegan, W., Kelso, J., Krause, J., Kumaresan, G., Kusimba, C., Kusimba, S., Lalueza-Fox, C., Llamas, B., MacEachern, S., Mallick, S., Matsumura, H., Morales-Arce, A. Y., Matuzeviciute, G. M., Mushrif-Tripathy, V., Nakatsuka, N., Nores, R., Ogola, C., Okumura, M., Patterson, N., Pinhasi, R., Prasad, S. P. R., Prendergast, M. E., Punzo, J. L., Reich, D., Sawafuji, R., Sawchuk, E., Schiffels, S., Sedig, J., Shnaider, S., Sirak, K., Skoglund, P., Slon, V., Snow, M., Soressi, M., Spriggs, M., Stockhammer, P. W., Szécsényi-Nagy, A., Thangaraj, K., Tiesler, V., Tobler, R., Wang, C.-C., Warinner, C., Yasawardene, S., Zahir, M. (2021). Ethics of DNA research on human remains: five globally applicable guidelines. *Nature*, 599(7883), strony 41-46. DOI:10.1038/s41586-021-04008-x
- [162] Hatzer-Grubwieser, P., Mühlmann, R., Berger, B., Niederstätter, H., Pavlic, M. i Parson, W. (2005). A new “miniSTR-multiplex” displaying reduced amplicon lengths for the analysis of degraded DNA. *International Journal of Legal Medicine*, 120, strony 115-120. DOI:10.1007/s00414-005-0013-6.
- [163] Elwick, K., Mayes, C., Hughes-Stamm, S. (2018). Comparative sensitivity and inhibitor tolerance of GlobalFiler® PCR Amplification and Investigator® 24plex QS kits for challenging samples. *Legal Medicine*, 32, strony 31-36. DOI:10.1016/j.legalmed.2018.01.006.
- [164] Cho, S., Kim, M. Y., Lee, J. H., Lee, H. Y., Lee, S. D. (2020). Large-scale identification of human bone remains via SNP microarray analysis with reference SNP database. *Forensic Science International: Genetics*, 47, str. 102293. DOI: 10.1016/j.fsigen.2020.102293.
- [165] Ludeman, M. J., Zhong, C., Mulero, J. J., Lagace, R. E., Hennessy, L. K., Short, M. L., Wang, D. Y. (2018). Developmental validation of GlobalFiler™ PCR amplification kit: a 6-dye multiplex assay designed for amplification of casework samples. *International Journal of Legal Medicine*, 132(6), str. 1555-1573. DOI:10.1007/s00414-018-1817-5.
- [166] Huang, Y., Wang, M., Liu, C., He, G. (2024). Comprehensive landscape of non-CODIS STRs in global populations provides new insights into challenging DNA profiles. *Forensic Science International: Genetics*, 70, str. 103010. DOI: 10.1016/j.fsigen.2024.103010.
- [167] Liu, Q.-L., Huang, K.-K., Wu, Y.-D., Zhao, H., Li, C.-T., Lu, D.-J. (2014). Genetic polymorphism of 13 non-CODIS STR loci in three national populations from China. *Electrophoresis*, 35(23), str. 3395-3401. DOI:10.1002/elps.201400290.

- [168] Zhang, J., Zhang, J., Tao, R., Jiang, L., Chen, L., Li, X., Li, X., Li, C., Zhang, S. (2020). A newly devised multiplex assay of novel polymorphic non-CODIS STRs as a valuable tool for forensic application. *Forensic Science International: Genetics*, 48, str. 102341. DOI: 10.1016/j.fsigen.2020.10234.
- [169] Jung, J. Y., Kim, E., Song, Y. W., Lee, D. G., Park, M. J., Lee, H. Y., Kayser, M., Ralf A., Kim, E. (2025). Evaluation of RMplex system for differentiating father–son pairs using Y-STRs in a Korean population. *Forensic Science International: Genetics*, 76, str. 103221. DOI:10.1016/j.fsigen.2025.103221.
- [170] Hedell, R., Nordgaard, A., Ansell, R. (2011). Discrepancies. *Forensic Science International: Genetics Supplement Series*, 3, strony e135-e136. DOI:10.1016/j.fsigss.2011.08.068.
- [171] Dash, H. R., Shetkar, R., Al-Snan, N. R. (2025). Impact of population size on population genetic analysis of Short Tandem Repeat (STR) allelic data, forensic and paternity parameters and its effect on forensic DNA analysis. *Forensic Science, Medicine and Pathology*, 21, strony 511-521. DOI:10.1007/s12024-024-00907-3.
- [172] Steele, C., Balding, D. (2014). Choice of population database for forensic DNA profile analysis. *Science and Justice*, 54, strony 487-493. DOI:10.1016/j.scijus.2014.10.004.
- [173] Turchi, C., Previdere, C., Bini, C., Carnevali, E., Grignani, P., Manfredi, A., Melchionda, F., Onofri, V., Pelotti, S., Robino, C., Sorçaburu-Ciglieri, S., Tagliabracci, A., Fattorini, P. (2020). Assessment of the Precision ID Identity Panel kit on challenging forensic samples. *Forensic Science International: Genetics*, 49, str. 102400. DOI: 10.1016/j.fsigen.2020.102400.
- [174] Antunes, J., Walichiewicz, P., Forouzmand, E., Barta, R., Didier, M., Han, Y., Perez, J. C., Snedecor, J., Zlatkov, C., Padmabandu, G., Devesse, L., Radecke, S., Holt, C. L., Kumar, S. A., Budowle, B., Stephens, K. M. (2024). Developmental validation of the ForenSeq® Kintelligence kit, MiSeq FGx® sequencing system and ForenSeq Universal Analysis Software. *Forensic Science International: Genetics*, 71, str. 103055. DOI: 10.1016/j.fsigen.2024.103055.
- [175] Novorski, N. M., Cihlar, J. C. (2022). Evolution of single-nucleotide polymorphism use in forensic genetics. *WIREs Forensic Science*, 4(6). DOI: 10.1002/wfs2.1459.
- [176] King, J. L., Churchill, J. D., Novorski, N. M., Zeng, X., Warshauer, D. H., Seah, L. H. i Budowle, B. (2018). Increasing the discrimination power of ancestry- and identity-informative SNP loci within the ForenSeq™ DNA Signature Prep Kit. *Forensic Science International: Genetics*, 36, str. 60-76. DOI: 10.1016/j.fsigen.2018.06.005.

- [177] Kim, J. J., Han, B. G., Lee, H. I., Yoo, H. W., Lee, J. K. (2010). Development of SNP-based human identification system. *International Journal of Legal Medicine*, 124(2), str. 125-131. DOI:10.1007/s00414-009-0389-9.
- [178] Tillmar, A., Sturk-Andreaggi, K., Daniels-Higginbotham, J., Thomas, J. T., Marshall, C. (2021). The FORCEPanel: An All-in-One SNP Marker Set for Confirming Investigative Genetic Genealogy Leads and for General Forensic Applications. *Genes*, 12, str. 1968. DOI:10.3390/genes12121968.
- [179] Daniel, R., Raymond, J., Sears, A., Stock, A., Scudder, N., Padmabandu, G., Kumar, S. A., Snedecor, J., Antunes J., Hartman, D. (2024). It's all relative: A multi-generational study using ForenSeq™ Kintelligence. *Forensic Science International*, 364, str. 112208. DOI:10.1016/j.forsciint.2024.112208.
- [180] Staadig, A., Hedman, J., Tillmar, A. (2023). Applying Unique Molecular Indices with an Extensive All-in-One Forensic SNP Panel for Improved Genotype Accuracy and Sensitivity. *Genes*, 14(4), str. 818. DOI:10.3390/genes14040818.
- [181] Gorden, E. M., Greytak, E. M., Sturk-Andreaggi, K., Cady, J., McMahon, T. P., Armentrout, S., Marshall, C. (2022). Extended kinship analysis of historical remains using SNP capture. *Forensic Science International: Genetics*, 57, str. 102636. doi:https://doi.org/10.1016/j.fsigen.2021.102636.
- [182] Żak, K., Konarzewska, M. (2023). Analysis of short tandem repeat mutations in paternity cases from Masovian Voivodeship provinces from years 2018-2022 based on materials of the Department of Forensic Medicine, Medical University of Warsaw. *Archiwum Medycyny Sądowej i Kryminologii*, 73(2), str. 139-148. DOI: 10.4467/16891716AMSIK.23.012.18687.
- [183] Pinto, N., Magalhães, M., Conde-Sousa, E., Gomes, C., Pereira, R., Alves, C., Gusmao, L., Amorim, A. (2013). Assessing paternities with inconclusive STR results: The suitability of bi-allelic markers. *Forensic Science International: Genetics*, 7(1), str. 16-21. DOI: 10.1016/j.fsigen.2012.05.002.
- [184] Mandape, S. N., Budowle, B., Mittelman, K., Mittelman, D. (2024). Dense single nucleotide polymorphism testing revolutionizes scope and degree of certainty for source attribution in forensic investigations. *Croatian Medical Journal*, 65(3), str. 249-260. DOI:10.3325/cmj.2024.65.249.
- [185] Ge, J., Budowle, B. (2020). How many familial relationship testing results could be wrong? *PloS Genetics*, 16(8), str. 1008929. DOI:10.1371/journal.pgen.1008929.

- [186] Egeland, T., Vigeland, M. D. (2025). Kinship cases with partially specified hypotheses. *Forensic Science International: Genetics*, 78, str. 103270. DOI: 10.1016/j.fsigen.2025.103270.
- [187] Zvánigorosky, V., Sabbagh, A., Gonzalez, A., Fausser, J.-L., Palstra, F., Romanov, G., Solovyev, A., Barashkov, N., Fedorova, S., Crubézy, É., Ludes, B., Keyser, C. (2020). The limitations of kinship determinations using STR data in ill-defined populations. *International Journal of Legal Medicine*, 134, str. 1981-1990. DOI: 10.1007/s00414-020-02298-w.
- [188] Fondevila, M., Santos, C., Phillips, C., Corrado, A., Butler, J., Lareu, M. V., Vallone, P. (2011). An assessment of linkage between forensic markers: core STRs, mini-STRs and INDELs. 22th International Symposium on Human Identification. Promega.
- [189] Fasani, G., Gino, S., Sguazzi, G. (2023). Incidental Findings in Forensic Investigations: A Narrative Review. *Forensic Science*, 3(2), strony 345-356. DOI:10.3390/forensicsci3020026.
- [190] Parker, L. S., London, A. J., Aronson, J. D. (2013). Incidental findings in the use of DNA to identify human remains: An ethical assessment. *Forensic Science International: Genetics*, 7(2), strony 221-229. DOI:10.1016/j.fsigen.2012.10.002.
- [191] Prinz, M., Carracedo, A., Mayr, W. R., Morling, N., Parson, T. J., Sajantila, A., Scheithauer, R., Schmitter, H., Schneider, P. M. (2007). DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI). *Forensic Science International: Genetics*, 1(1), strony 3-12. DOI:10.1016/j.fsigen.2006.10.003.
- [192] Voracek, M., Haubner, T., Fisher, M. L. (2008). Recent Decline in Nonpaternity Rates: A Cross-Temporal Meta-Analysis. *Psychological Report*, 103(3), strony 799-811. DOI:10.2466/pr0.103.3.799-811.
- [193] Skourtanioti, E., Jia, X., Tavartkiladze, N., Bitadze, L., Shengelia, R., Tushabramishvili, N., Aslanishvili, V., Gasparyan, B., Kandel, A. W., Naumann, D., Neumann, G. U., Bianco, R. A., Mötsch, A., Prüfer, K., Lamnidis, T., Traverso, L., Ghalichi, A., Ellingvåg, S., Stockhammer, P. W., Krause, J., Ringbauer, H. (2025). The genetic history of the Southern Caucasus from the Bronze Age to the Early Middle Ages: 5,000 years of genetic continuity despite high mobility. *Cell*, 188(19), str. 5278-5294. DOI:10.1016/j.cell.2025.07.013.
- [194] Zwolski, M., (2018). Poszukiwanie i identyfikacja ofiar zbrodni systemów totalitarnych (Tom 35). Białystok: Instytut Pamięci Narodowej.

8. Opinia komisji Bioetycznej

Komisja Bioetyczna Pomorskiego Uniwersytetu Medycznego w Szczecinie zatwierdziła, iż przedmiotowe badania nie wymagają opinii Komisji Bioetycznej (KB.006.206.2024, 31.12.2024).

Uzyskano świadomą zgodę wszystkich badanych i/lub ich opiekunów prawnych.