

Marcin Woźniak

Genealogia i genetyka – wspólny przedmiot badań, różne perspektywy

Rocznik Lubelskiego Towarzystwa Genealogicznego 3, 203-228

2011

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Marcin Woźniak

(Zakład Genetyki Molekularnej i Sądowej Katedry Medycyny Sądowej
Collegium Medicum Uniwersytetu Mikołaja Kopernika w Bydgoszczy)

Genealogia i genetyka - wspólny przedmiot badań, różne perspektywy

1. Wprowadzenie
2. Informacja genetyczna i jej dziedziczenie
 - 2.1. Dziedziczenie DNA i jego zmienność
 - 2.2. Markery haploidalne
 - 2.2.1. Chromosom Y
 - 2.2.2. DNA mitochondrialny
3. Badania genetyczne dla potrzeb genealogii – przykłady praktyczne
 - 3.1. Klasyczna analiza ojcostwa/macierzyństwa
 - 3.2. Analiza pokrewieństwa w obrębie najbliższej rodziny
 - 3.3. Analiza pokrewieństwa w obrębie dalszej rodziny i badania grupowe
 - 3.4. Ustalanie pochodzenia biogeograficznego
4. Kilka uwag końcowych dla osób rozważających przeprowadzenie badań genetycznych
5. Podsumowanie
6. Słownik podstawowych pojęć
7. Bibliografia

1. Wprowadzenie

Genetyka, podobnie jak genealogia, zajmuje się badaniem pokrewieństwa, choć pokrewieństwo to postrzegane jest w przypadku genetyki bardzo szeroko, jako związki pomiędzy wszystkimi organizmami żywymi zamieszkującymi Ziemię. Współczesna genetyka, opierająca się na teorii ewolucji, wykorzystująca osiągnięcia biologii molekularnej i zaawansowane metody analityczne, pozwala na analizę podobieństw pomiędzy organizmami na poziomie cząsteczek chemicznych tworzących

żywe komórki, takich jak białka i, przede wszystkim, materiał genetyczny. Kwas deoksyrybonukleinowy, tworzący materiał genetyczny człowieka, znany powszechnie jako DNA, jest nośnikiem informacji niezbędnej komórkom do życia i choćby z tego powodu musi być przekazywany z pokolenia na pokolenie. Przekazywana w ten sposób informacja zawarta w DNA może być z kolei źródłem wiedzy o pokrewieństwie między ludźmi i innymi organizmami.

Przedmiotem badań genealogii jest wycinek świata żywego, jakim są populacje ludzkie, a w zasadzie badanie związków pomiędzy poszczególnymi osobnikami oraz związków w obrębie rodzin tworzących te populacje. Klasyczne źródła informacji wykorzystywane w badaniach genealogicznych to przede wszystkim dokumentacja urodzeń i małżeństw (księgi metrykalne), ale także przekazy, takie jak pamiętniki, „legendy rodzinne” i inne formy opisu historii rodzinnej.

Skoro badanie pokrewieństwa stanowi wspólny mianownik poszukiwań zarówno w genealogii, jak i w genetyce, warto zastanowić się, czy metody stosowane w jednej z tych gałęzi wiedzy mogą być przydatne w drugiej. Niniejszy artykuł ma na celu przybliżenie podstawowej wiedzy o dziedziczeniu materiału genetycznego oraz dostępnych możliwościach wykorzystania tej wiedzy dla potrzeb badań pokrewieństwa biologicznego. Chciałbym również przedstawić na kilku przykładach, w jaki sposób można praktycznie wykorzystać metody biologii molekularnej do rozwiązywania konkretnych problemów z zakresu genealogii.

Zdając sobie sprawę z dość hermetycznej natury języka współczesnej biologii molekularnej, starałem się posługiwać w niniejszym opracowaniu możliwie potocznym językiem oraz objaśniać każdy termin pochodzący z dziedziny genetyki. Dla wygody Czytelników na końcu artykułu załączyłem również krótki słownik terminów używanych w tekście, który – mam nadzieję – pomoże w pełnym zrozumieniu treści zawartej w niniejszym opracowaniu.

2. Informacja genetyczna i jej dziedziczenie

Celem niniejszego rozdziału jest zapoznanie Czytelników z podstawowymi informacjami o budowie i dziedziczeniu ludzkiego DNA. Starałem się ograniczyć prezentowane tu informacje do zakresu koniecznego dla zrozumienia praktycznych aspektów wykorzystania analiz DNA dla potrzeb badania pokrewieństwa, stąd Czytelnicy dobrze zorientowani w podstawach genetyki z pewnością dostrzegą pewne uproszczenia i skróty myślowe. W mojej ocenie informacje zawarte w tym rozdziale stanowią niezbędne minimum dla każdego, kto chciałby zrozumieć istotę pokrewieństwa genetycznego i jej powiązanie z różnymi stopniami pokrewieństwa i powinowactwa wykorzystywanymi powszechnie w genealogii.

Informacja genetyczna zawarta w komórkach żywych organizmów dzielona jest zwyczajowo na jednostki zwane genomami. Pojedynczy genom można sobie wyobrazić jako zestaw wszystkich cząsteczek DNA koniecznych do uzyskania pełnego zbioru genów danego organizmu. Genom jest w swej istocie sposobem zapisu informacji niezbędnej do funkcjonowania żywej komórki. Z punktu widzenia genealogii genom można określić jako jednostkę dziedziczenia zawierającą praktycznie wszystkie informacje biologiczne, jakie rodzic przekazuje swojemu dziecku.

Aby zrozumieć istotę dziedziczenia informacji genetycznej i możliwości, jakie otwierają przed genealogią techniki biologii molekularnej stosowane w badaniach pokrewieństwa, konieczne jest poznanie podstaw funkcjonowania i budowy genomu oraz mechanizmów rządzących jego przekazywaniem z pokolenia na pokolenie. Poszczególne poziomy konstrukcji genomu można spróbować opisać uciekając się do analogii, porównując zbiór informacji zwany genomem do innego rodzaju zbioru informacji, biblioteki pełnej książek. Powszechnie wiadomo, że nośnikiem informacji genetycznej jest DNA. Częsteczka ta przyjmuje fizycznie postać nici złożonej z połączonych ze sobą, jedna za drugą, mniejszych cząsteczek, zwanych nukleotydami. W jaki sposób tak zbudowana cząsteczka chemiczna może kodować informację? Schemat opisujący istotę tego zjawiska przedstawiłem na Rycinie 1, a poniżej znajduje się objaśnienie zawartych tam danych.

Wyobraźmy sobie, że informacja o budowie i funkcjonowaniu naszych komórek i całych organizmów zapisana jest w postaci tekstu. Analogia ta jest bliższa rzeczywistości, niż mogłoby się zdawać, gdyż nić DNA to w praktyce nic innego jak ciąg znaków tworzących „słowa” i „zdania” niosące informację. Zapis informacji genetycznej uderzająco przypomina stworzony przez ludzi sposób zapisu z wykorzystaniem alfabetu, z tą jedynie różnicą, że „alfabet” genetyczny zawiera tylko 4 litery, oznaczane umownie A, C, G i T od pierwszych liter nazw związków chemicznych (nukleotydów) tworzących zapis genetyczny. Tak bowiem, jak w tekstach stworzonych przez ludzi informacja jest zapisywana przez postawienie w odpowiedniej kolejności znaków na papierze, tak informacja genetyczna zapisana jest jako ciąg kolejnych cząsteczek chemicznych, których kolejność w nici DNA określa treść zawartej w tej nici informacji. „Litery” genetycznego alfabetu składają się na „słowa”, które również można porównać do słów składanych z liter, przy czym w „języku” komórki słowa te mają zawsze dokładnie 3 litery i nazywane są kodonami. To ograniczenie wynika z mechanizmów molekularnych wykorzystywanych przy odczycie informacji genetycznej, które to mechanizmy potrafią poprawnie odczytywać jedynie słowa trzyliterowe.

„Język” zapisu genetycznego posiada zatem swój alfabet, swoje słowa, a także swoją specyficzną gramatykę, pozwalającą na budowanie „zdań”. W języku informacji genetycznej „zdaniem” możemy nazwać ciąg kodonów rozpoczynający się od określonej sekwencji liter - cząstek chemicznych budujących DNA - i kończący się swego rodzaju znakiem interpunkcyjnym, trójliterową sekwencją określaną jako kodon STOP i pełniącą rolę analogiczną do kropki stawianej na końcu zdania. „Zdanie” zapisane w języku informacji genetycznej to nic innego jak gen, czyli informacja o strukturze jednego z tysięcy białek kodowanych przez genom. Należy bowiem pamiętać, że istotą informacji genetycznej jest przechowywanie w cząsteczce DNA zapisu struktury wszystkich białek produkowanych w komórce i cząstek RNA wspomagających te białka w ich funkcjach.

Wiedząc jak powstaje zapis genu i jaka jest jego funkcja, wyobraźmy sobie teraz tysiące genów w postaci nici DNA połączonych ze sobą, jedna za drugą w taki sposób, że tworzą jedną, bardzo długą nić. Taka pojedyncza nić DNA zawierająca liczne geny nazywana jest chromosomem. Pozostając przy analogii z językiem pisanym, jeśli pojedynczy gen jest „zdaniem”, to pojedynczy chromosom możemy

wyobrazić sobie jako księgę wypełnioną wieloma zdaniem. Kontynuując ten tok rozumowania, zbiór chromosomów („ksiąg”) zawierających łącznie kompletną informację o wszystkich białkach danego organizmu potrzebnych do prawidłowego funkcjonowania jego komórek nazywamy genomem. Zgodnie z prezentowaną analogią, genom możemy nazwać swego rodzaju „biblioteką”. Komplet ksiąg (czyli chromosomów) jest zlokalizowany w jądrze każdej komórki ciała (Rycina 1). W przypadku genomu człowieka na kompletną informację genetyczną składają się nici DNA tworzące 22 chromosomy nie płciowe, dwa chromosomy płciowe oznaczane konwencjonalnie jako X i Y oraz osobliwa cząsteczka DNA występująca poza jądrem komórkowym, nazywana DNA mitochondrialnym (mtDNA). Tej ostatniej cząsteczce, jak również chromosomowi Y, poświęcę jeszcze odrębne akapity niniejszego opracowania, ze względu na ich ogromne znaczenie dla badania pokrewieństwa, w tym również w badaniach z zakresu genealogii.

2.1. Dziedziczenie DNA i jego zmienność

W tym miejscu warto wspomnieć, że chromosomy nie determinujące płci (tzw. autosomy), oznaczane zwyczajowo numerami od 1 do 22, są w genomowej „bibliotece” zdublowane, tzn. każdy z nich występuje w dwóch kopiach, podczas gdy chromosomy determinujące płeć człowieka, oznaczane literami X i Y, mogą występować w jądrze komórkowym w postaci pary siostrzanych chromosomów X (u kobiet) lub pary składającej się z bardzo od siebie odmiennych chromosomów X i Y (u mężczyzn). Przyczyną takiego stanu rzeczy jest sposób dziedziczenia informacji genetycznej, pokazany schematycznie na Rycinie 2. Każdy rodzic posiada w swoich komórkach zdublowane chromosomy każdej pary, czyli dwa genomy, ale przekazuje dziecku po jednym chromosomie z każdej pary chromosomów nie płciowych (22 chromosomy) oraz jeden z chromosomów płciowych. W sumie zatem każdy rodzic przekazuje swojemu dziecku jeden kompletny genom, który w połączeniu z drugim genomem, pochodzącym od drugiego rodzica, tworzy kompletny garnitur chromosomowy komórki.

W przypadku kobiet przekazywanym chromosomem płciowym jest zawsze chromosom X (z pary XX), podczas gdy mężczyzna może przekazać dziecku albo chromosom X albo chromosom Y. Dziecko, które otrzyma od ojca chromosom X posiada w swoich komórkach dwa chromosomy X, jest więc płci żeńskiej, podczas gdy dziecko otrzymując od ojca chromosom Y w dodatku do matczynego X jest płci męskiej. Ponadto matka przekazuje swoim dzieciom swój DNA mitochondrialny (ojciec nie ma w tym żadnego udziału – patrz Rycina 2 i objaśnienia poniżej). Szczególny sposób dziedziczenia chromosomu Y i mtDNA ma istotne znaczenie dla możliwości wykorzystania naszego materiału genetycznego do badań pokrewieństwa.

Informacja genetyczna, pomimo podobieństwa zasad jej zapisu do zasad używania pisma opracowanych przez ludzi, posiada kilka cech, które znacząco odróżniają ją od ludzkich systemów zapisu informacji, a jednocześnie są istotne dla zrozumienia mechanizmów rządzących dziedziczeniem. Przede wszystkim informacja genetyczna w genomach organizmów posiadających jądro komórkowe jest rozproszona, co oznacza, że poszczególne geny („zdania”) nie sąsiadują ze sobą bezpośrednio lecz są rozdzielone fragmentami nici DNA nie niosącymi żadnej informacji, tzw.

sekwencjami niekodującymi. W przypadku genomu człowieka odcinki niekodujące stanowią jego zdecydowaną większość, ocenia się że nawet ok. 98%. Odwołując się do analogii z książkami, można sobie wyobrazić przykładowy chromosom człowieka jako księgę, w której tylko jedna strona na każde 50 jest zadrukowana tekstem posiadającym sens, natomiast pozostałe 49 stron z każdej pięćdziesiątki jest zadrukowanych losowymi ciągami liter, nie tworzącymi spójnego tekstu. Taki sposób zapisu informacji wydaje się być z ludzkiego punktu widzenia bardzo nieefektywnym, jednak mechanizmom komórkowym nie nastęcza większych trudności odnalezienie na tej informacyjnej „pustyni” fragmentów zawierających „oazy” informacji w postaci genów. Jednym z istotnych skutków rozproszenia genów na nici DNA jest znaczna długość ludzkiego DNA. Pełny genom człowieka liczy ok. 3 mld nukleotydów, co oznacza, że w jądrze każdej komórki człowieka obecne są nici DNA o łącznej długości ok. 6 mld nukleotydów. Przekładając te wielkości na powszechnie stosowany system miar można powiedzieć, że każda komórka człowieka zawiera w sumie ok. 2 m DNA, z czego łącznie tylko ok. 4 cm stanowią geny kodujące białka.

Inną ważną cechą sposobu kodowania informacji genetycznej, wynikającą z opisanych wyżej zależności i szczególnie istotną z punktu widzenia badań pokrewieństwa, jest zmienność. DNA, jako związek chemiczny, jest cząsteczką stosunkowo stabilną, zdarza się jednak, że substancje chemiczne i czynniki fizyczne występujące naturalnie w komórce bądź dostające się do jej wnętrza z otoczenia modyfikują strukturę chemiczną nukleotydów budujących materiał genetyczny. Uszkodzenia takie są naprawiane przez odpowiednie białka występujące w komórce, jednak mechanizmy naprawcze nie zawsze są w pełni efektywne, w wyniku czego w naprawianych miejscach nici DNA pojawiają się nukleotydy inne niż te, które znajdowały w tych miejscach pierwotnie. Informacja ulega zatem zmianie – dochodzi do mutacji. Inną przyczyną pojawiania się zmian w sekwencji nukleotydów tworzących DNA może być np. niedokładność mechanizmów kopiujących materiał genetyczny przed podziałem komórki. Odwołując się do analogii z tekstem pisanym, mutacja w DNA odpowiada zatem sytuacji, w której niektóre litery w tekście są losowo zamieniane na inne w wyniku błędów w kopiowaniu lub w naprawianiu uszkodzeń. Opisane mechanizmy powodują, że kolejność nukleotydów, nazywana sekwencją DNA, nie jest identyczna w genomie każdego przedstawiciela naszego gatunku. Zjawisko zmienności sekwencji DNA w obrębie jednego gatunku nazywamy polimorfizmem genetycznym. Odgrywa ono kluczową rolę w badaniach pokrewieństwa między organizmami, gdyż możliwość poznania i wykrywania charakterystycznych mutacji w cząsteczce DNA pozwala na badanie drogi, jaką ta konkretna cząsteczka jest przekazywana z pokolenia na pokolenie w procesie rozmnażania.

Jak wspomniano wyżej, mutacje w DNA zachodzą zasadniczo w sposób losowy, co oznacza, że większość z nich pojawia się w tych regionach genomu, które nie kodują informacji o białkach. Wynika to przede wszystkim z faktu, iż niemal 98% naszego genomu nie koduje białek, a więc losowe mutacje mają jedynie ok. 2% szansy na pojawienie się w obrębie jakiegoś genu. Ponadto mutacje zachodzące w genach są bardzo często szkodliwe dla organizmu, zatem większość z nich jest eliminowana z populacji w procesie selekcji naturalnej. Ponieważ jednak niekodujące fragmen-

ty DNA nie podlegają tak ścisłej selekcji, obecne w nich mutacje są przekazywane z pokolenia na pokolenie, przez co akumulują się w populacji. Ocenia się, że gdyby porównać materiał genetyczny dwóch losowo wybranych, nie spokrewnionych osób, to średnio co ok. 500 nukleotydów wystąpi różnica w sekwencji DNA. Znając miejsca, w których takie różnice występują najczęściej, genetycy potrafią, po przeprowadzeniu odpowiednich badań metodami biologii molekularnej, opisać DNA konkretnego człowieka, podając listę charakterystycznych dla tej konkretnej osoby mutacji. Tworzony jest w ten sposób tzw. indywidualny profil genetyczny, który jest podstawą wszelkich badań pokrewieństwa (patrz Rycina 3 i objaśnienia poniżej).

Kolejną ważną cechą odróżniającą zapis i dziedziczenie ludzkiego DNA od metod zapisu, przechowywania i przekazywania informacji w formie tekstu pisanego jest występowanie zjawiska nazywanego rekombinacją genetyczną (ang. *crossing over*). Jak wiemy, dziecko otrzymuje od każdego z rodziców po jednym chromosomie z każdej pary, jednak chromosomy rodzicielskie, przed przekazaniem dzieciom, przechodzą przez proces, w trakcie którego chromosomy danej pary wymieniają się fragmentami swojego DNA, tworząc nowy chromosom, zawierający mieszankę cech genetycznych (mutacji) charakterystycznych dla chromosomów danej pary obecnych u rodzica. Taki „wymieszany” chromosom trafia do plemnika lub komórki jajowej i jest przekazywany dziecku (Rycina 2). Odwołując się do analogii przyrównującej chromosomy do książek, rekombinację można sobie wyobrazić jako zdarzenie, w wyniku którego z dwóch niemal identycznych książek (para chromosomów), różniących się od siebie jedynie nieznacznie w wyniku losowych zmian wprowadzonych w tekście (mutacje) powstaje nowa książka, zawierająca po 50% tekstu z każdej z książek wyjściowych, wraz z charakterystycznymi dla tych książek zmianami w tekście, obecnymi w poszczególnych fragmentach, z których została złożona „książka potomna”. Ta właśnie „książka” trafia jako dziedzictwo rodzica do nowej „biblioteki” – genomu dziecka (Rycina 2). Jak łatwo sobie wyobrazić, w wyniku procesu rekombinacji dochodzi do zerwania związków pomiędzy mutacjami obecnymi na poszczególnych chromosomach, gdyż po każdej rekombinacji ok. 50% z tych mutacji trafia do nowopowstałego chromosomu, przekazywanego dalej, gdzie łączą się na jednej nici DNA z mutacjami charakterystycznymi dla drugiego chromosomu danej pary.

Ważna dla zrozumienia istoty genetycznych analiz pokrewieństwa jest informacja, że badaniom nie podlega cały DNA danej osoby, a jedynie jego wybrane fragmenty. We współczesnej genetyce najczęściej stosowane w tego typu badaniach są tzw. markery mikrosatelitarne, zwane również mikrosatelitami lub sekwencjami STR. Podstawową zaletą mikrosatelitów dla badań genetycznych jest fakt, iż ulegają one mutacjom częściej niż inne odcinki DNA oraz mogą występować w populacji w kilkunastu, a czasem nawet kilkudziesięciu wariantach. Dzięki tak dużemu zróżnicowaniu omawianych fragmentów DNA możliwe jest stworzenie unikalnego profilu genetycznego każdego człowieka już w oparciu o analizę kilkunastu mikrosatelitów, stanowiących niewielki ułamek całego genomu. Istota takiej analizy sprowadza się do zbadania, jakie warianty mutacji występują w wybranych mikrosatelitach na poszczególnych chromosomach a jej wyniki zapisuje się w postaci tabeli, w której wymienione są badane mikrosatelity i występujące w nich warianty mutacji (Rycina 3 – lewy panel).

2.2. Markery haploidalne

Warto w tym miejscu wspomnieć o dwóch specyficznych cząsteczkach DNA obecnych w naszych komórkach, których sposób dziedziczenia sprawia, że są one wyjątkowo użytecznymi narzędziami w badaniu pokrewieństwa, szczególnie, gdy to pokrewieństwo jest stosunkowo odległe. Cząsteczki te nazywane są markerami haploidalnymi, co oznacza, że występują w komórkach w formie, która uniemożliwia rekombinację. Fakt ten ma istotne znaczenie dla możliwości wykorzystania owych markerów dla potrzeb genealogii.

2.2.1. Chromosom Y

Pierwszą z takich cząsteczek jest chromosom Y. Jest to najmniejszy z ludzkich chromosomów, gdyż jego nić DNA liczy tylko ok. 60 mln. nukleotydów. Chromosom Y posiada właściwości bardzo podobne do pozostałych chromosomów człowieka, z jednym wyjątkiem: praktycznie nie podlega rekombinacji, gdyż tworzący z nim parę drugi chromosom płciowy (X) posiada całkowicie odmienną sekwencję DNA, co nie pozwala mechanizmom kontrolującym rekombinację na „wymieszanie” DNA chromosomów pary XY, podczas gdy para XX w komórkach kobiet podlega rekombinacji tak samo jak wszystkie inne chromosomy. Ze względu na brak rekombinacji w obrębie chromosomu Y nie dochodzi do zerwania ciągłości nici DNA pomiędzy poszczególnymi fragmentami każdego indywidualnego chromosomu, a co za tym idzie, wszystkie mutacje obecne na jego nici DNA są przekazywane z pokolenia na pokolenie łącznie, jako jeden zestaw, który nazywany jest haplotypem. Należy tu podkreślić, że również inne chromosomy, oprócz chromosomu Y, posiadają swoje indywidualne haplotypy, jednak w toku rekombinacji haplotypy te ulegają permutacjom, które w praktyce uniemożliwiają już po upływie kilku pokoleń odtworzenie haplotypów chromosomów wyjściowych (obecnych np. u pradziadków) w oparciu o analizę chromosomów posiadanych przez potomstwo żyjące obecnie. W przeciwieństwie do pozostałych chromosomów, haplotyp chromosomu Y jest w zasadzie niezmienny w danej linii męskiej, za wyjątkiem nabywanych co jakiś czas nowych mutacji (Rycina 4). Dzięki tej właściwości możliwe jest odtworzenie drogi ewolucyjnej każdego chromosomu Y w oparciu o analizę jego profilu genetycznego. Droga ewolucyjna zaś, to nic innego jak opis dziedziczenia i zmian zachodzących w chromosomie z pokolenia na pokolenie. Profil genetyczny chromosomu Y uzyskuje się przez oznaczenie wariantów sekwencji DNA obecnych w wybranych miejscach cząsteczki chromosomu. Miejscami takimi mogą być zarówno mikrosatelity (Rycina 3, panel środkowy), jak i mutacje punktowe, tzw. polimorfizmy SNP.

Analiza chromosomu Y może zatem odbywać się na dwóch poziomach. Analiza mutacji punktowych chromosomu Y (Y-SNP), czyli zmian pojawiających się stosunkowo rzadko, ale bardzo stabilnych ewolucyjnie, pozwala na odtworzenie pokrewieństwa pomiędzy poszczególnymi chromosomami Y w odległej perspektywie czasowej, sięgającej tysięcy lat. Badania Y-SNP pozwalają wydzielić w populacji grupy chromosomów Y niosące określone, mutacje, świadczące o pochodzeniu od wspólnego, męskiego przodka. Grupy takie nazywane są haplogrupami. Przykładowo, w populacji Polski najczęściej występują chromosomy Y należące do trzech

haplogrup: R1b, R1a1 i I. Haplogrupy te można rozdzielić w oparciu o badania dodatkowych mutacji na podgrupy, jednak nawet te podgrupy liczą najczęściej tysiące mężczyzn, których pokrewieństwo w linii ojcowskiej jest bezsporne lecz sięga tysiące lat wstecz. Z tego względu analiza Y-SNP rzadko znajduje zastosowanie w badaniach genealogicznych jako samodzielna technika badawcza, choć może stanowić istotne uzupełnienie innych metod. Podstawową metodą badawczą pozwalającą na badanie pokrewieństwa w linii ojcowskiej jest analiza haplotypu chromosomu Y w oparciu o sekwencje mikrosatelitarne, znane też jako Y-STR. Sekwencje takie charakteryzują się znacznie większą częstością mutacji niż Y-SNP, co powoduje, że zróżnicowanie chromosomów Y w danej linii ojcowskiej następuje znacznie szybciej i pozwala na wyróżnienie linii potomnych i określenie zależności między nimi już w obrębie grup stosunkowo młodych, liczących kilka lub kilkanaście pokoleń.

Aby zrozumieć istotę analizy pokrewieństwa w oparciu o polimorfizm chromosomu Y przyjrzyjmy się Rycinie 4, na której przedstawiono schematycznie proces powstawania zmienności markerów haploidalnych i jej dziedziczenie. Chromosom wyjściowy, założycielski dla danej linii męskiej, charakteryzuje się obecnością dwóch wariantów sekwencji odróżniających go od innych chromosomów Y w populacji (oznaczone A i B na schemacie). Synowie mężczyzny będącego nosicielem tego chromosomu otrzymają go od ojca w formie niezmienionej, jednak w trakcie przekazywania może dojść do pojawienia się w nici DNA nowych mutacji, które definiują nowe linie ojcowskie. Linie te posiadają zatem zarówno mutacje odcjowskie (A i B), jak i własne, „prywatne” mutacje (C lub D). Los nowych linii ojcowskich może być dwojaki: mogą wygasnąć bezpotomnie (co spotkało na schemacie linię z mutacją D) lub przetrwać i stać się częścią puli chromosomów Y obecnych w danej populacji. Omawiany schemat przedstawia losy jednej linii ojcowskiej w ciągu czterech pokoleń. Widzimy, że w ostatnim pokoleniu wszystkie chromosomy nadal niosą charakterystyczne dla chromosomu wyjściowego mutacje A i B, jednak zdecydowana większość potomków mężczyzny, który był założycielem tej linii posiada na swoim chromosomie również mutację C. Stało się tak dlatego, że jeden z potomków założyciela, niosący właśnie tę mutację, osiągnął sukces reprodukcyjny tzn. miał wystarczająco wielu męskich potomków, aby znacząco zwiększyć częstość charakterystycznej dla siebie sekwencji chromosomu Y w populacji. Analizując mutacje występujące w cząsteczce DNA chromosomu Y u osób żyjących współcześnie, możemy zatem – pod warunkiem, że dysponujemy wystarczającą ilością danych – odtwarzać przebieg procesu akumulacji mutacji na tym chromosomie, sprowadzając obserwowaną zmienność do wspólnego mianownika, jakim jest ostatni wspólny przodek danej grupy mężczyzn. Znając tempo zachodzenia mutacji w DNA możemy również w przybliżeniu obliczyć, kiedy żył ów wspólny przodek. Tego typu analizy nie można przeprowadzić dla żadnego innego chromosomu człowieka, gdyż rekombinacja DNA powoduje zerwanie połączenia pomiędzy mutacjami obecnymi na autosomach, co powoduje, że nowe mutacje pojawiają się za każdym razem na tle innych mutacji rodzicielskich.

Ze względu na swoje charakterystyczne cechy chromosom Y służy do analizy pokrewieństwa w linii ojcowskiej, co do pewnego stopnia ogranicza jego zastosowanie w genealogii, lecz jednocześnie pozwala w pewnych sytuacjach na uzyskanie

wiedzy, której nie można ustalić przy pomocy żadnej z innych dostępnych metod analizy genetycznej.

2.2.2. DNA mitochondrialny

Drugą z cząsteczek DNA dziedziczonych bez rekombinacji jest DNA mitochondrialny (mtDNA). Cząsteczka ta posiada kilka cech, które wyróżniają ją spośród pozostałych fragmentów DNA obecnych w naszych komórkach. Przede wszystkim występuje poza jądrem komórkowym, w organellach zwanych mitochondriami. Taka lokalizacja powoduje, że mtDNA traktowany jest formalnie jako odrębny genom – genom mitochondrialny. DNA mitochondrialny nie podlega dziedziczeniu według schematu DNA jądrowego, lecz dziedziczy się przez cytoplazmę komórkową. Ponieważ mitochondria są obecne w cytoplazmie komórki jajowej i przekazywane dziecku, a mitochondria plemników nie biorą udziału w tym procesie, mtDNA człowieka dziedziczy się wyłącznie w linii matczynej. Cząsteczka mtDNA jest stosunkowo mała, liczy zaledwie nieco ponad 16 500 nukleotydów. W obrębie mtDNA występuje 37 genów, które są ciasno „upakowane” na tym krótkim odcinku DNA, co powoduje, że niemal nie występuje w nim rozproszenie informacji genetycznej, charakterystyczne dla chromosomów obecnych w jądrze komórki. W mtDNA nie występują również typowe sekwencje mikrosatelitarne, a jedynym polimorfizmem dostępnym badaniu jest polimorfizm SNP, występujący przede wszystkim na krótkim odcinku, liczącym ok. 1000 nukleotydów, który nie koduje żadnego genu i nazywany jest „pętlą D”. Ze względu na brak możliwości rekombinacji mutacje obecne w cząsteczce mtDNA dziedziczą się w postaci haplotypów, które można oznaczać wykrywając obecność charakterystycznych mutacji w poszczególnych miejscach sekwencji cząsteczki DNA (Rycina 3). Pozycje tych mutacji określa się według standardowej sekwencji wzorcowej mtDNA, zwanej sekwencją rCRS, liczącej 16569 nukleotydów. Według przyjętej konwencji wystąpienie w profilu mtDNA oznaczenia 263G wskazuje, że badana cząsteczka mtDNA różni się od rCRS tym, że posiada w nukleotydzie znajdującym się na 263 pozycji w sekwencji guaninę (G). Brak rekombinacji powoduje również, że zmienność mitochondrialnego DNA kumuluje się w populacji według schematu identycznego jak zmienność chromosomu Y zilustrowana na Rycinie 4.

Ze względu na praktyczny brak sekwencji niekodujących oraz stosunkowo niewielkie rozmiary cząsteczki ludzkiej mtDNA cechuje się znacznie mniejszą zmiennością niż pozostałe cząsteczki ludzkiego DNA. Podstawowym zadaniem analizy mtDNA z punktu widzenia genealogii jest możliwość potwierdzania pokrewieństwa w linii matczynej. Głównym narzędziem pozwalającym na rozróżnienie sekwencji poszczególnych cząsteczek mtDNA jest szczegółowa analiza kolejności nukleotydów w obrębie wspomnianej wcześniej pętli D zwana sekwencjonowaniem. Analiza taka jest bardziej skomplikowana od badań prowadzonych na mikrosatelitach, dlatego jest nieco droższa i oferowana przez stosunkowo nieliczne, wyspecjalizowane laboratoria. W niektórych przypadkach w analizach pokrewieństwa w linii matczynej stosuje się również analizę całej cząsteczki mtDNA, nazywaną sekwencjonowaniem pełnego genomu mtDNA.

3. Badania genetyczne dla potrzeb genealogii – przykłady praktyczne

Dysponując niezbędną wiedzą o mechanizmach rządzących przekazywaniem informacji zakodowanej w DNA z pokolenia na pokolenie przyjrzyjmy się, jakie praktyczne zastosowania może mieć ta wiedza w aspekcie badań genealogicznych.

3.1. Klasyczna analiza ojcostwa/macierzyństwa

Ponieważ w DNA człowieka jeden chromosom z każdej pary pochodzi od matki a drugi od ojca, jeden z wariantów obserwowanych dla danego mikrosatelity w profilu genetycznym dowolnej osoby musi pochodzić od ojca tej osoby a drugi od matki. Korzystając z tej podstawowej wiedzy można zatem ustalać bezpośrednio pokrewieństwo między ludźmi wykonując stosunkowo prostą analizę niewielkiej części genomu w postaci kilkunastu mikrosatelitów. Badania takie polegają na określeniu profilu genetycznego każdej z badanych osób a następnie porównaniu rezultatów i sprawdzeniu, czy w profilach osób przypuszczalnie spokrewnionych występuje w każdym z badanych mikrosatelitów przynajmniej jeden wspólny wariant (Tabela 1). Jeżeli przynajmniej w części ze zbadanych mikrosatelitów nie wystąpi żaden wariant wspólny między profilami dwóch osób, oznacza to, że osoby te nie mogą być bezpośrednio spokrewnione, tzn. nie stanowią pary rodzic – dziecko. Ze względu na stosunkowo wysoką częstość mutacji mikrosatelitów przyjmuje się, że jedna lub dwie rozbieżności między profilami genetycznymi domniemyanych krewnych nie musi oznaczać wykluczenia pokrewieństwa, jednak trzy i więcej rozbieżności uważa się za praktycznie pewne wykluczenie.

Warto zaznaczyć, że potwierdzenie pokrewieństwa między dwoma osobami przy pomocy badań genetycznych musi być dokonane z wykorzystaniem odpowiednich obliczeń statystycznych. Obliczenia te wykonywane są w oparciu o zebrane wcześniej dane o częstości poszczególnych badanych wariantów mikrosatelitów w populacji. Sednem obliczeń jest określenie, jaka jest szansa, że zgodność wariantów sekwencji obecnych w DNA badanych osób jest przypadkowa. Wyniki tego typu obliczeń przedstawiane są zwykle w postaci wartości liczbowej obrazującej stosunek szansy, że badane osoby posiadają wspólne warianty DNA, ponieważ są bezpośrednio spokrewnione do szansy, że osoby te posiadają wspólne warianty DNA w wyniku czystego przypadku wynikającego z tego, że ilość wariantów poszczególnych mikrosatelitów w populacji jest ograniczona. Uznaje się, że wartość tak obliczonego współczynnika pokrewieństwa musi wynosić więcej niż 100 tys., aby można było mówić o praktycznie pewnym jego potwierdzeniu. Wartość ta oznacza, iż hipoteza, że dane dwie osoby są spokrewnione jest 100 tys. razy bardziej prawdopodobna dla uzyskanego wyniku badań, niż hipoteza, że są to nie spokrewnione osoby, losowo wybrane z populacji. Uzyskanie współczynnika pokrewieństwa o wartości przekraczającej 100 tys. jest stosunkowo łatwe w przypadku badań obejmujących dziecko i jego domniemanego rodzica lub rodziców, zwykle wystarczy przebadanie ok. 15 mikrosatelitów. Wartość 100 tys. uzyskaną dla współczynnika pokrewieństwa można przeliczyć na procentową wartość prawdopodobieństwa pokrewieństwa, która wynosi w tym przypadku 99, 999%

Analizy bezpośredniego pokrewieństwa wykorzystywane są przede wszystkim dla potwierdzania ojcostwa lub macierzyństwa, zatem z punktu widzenia typowych badań genealogicznych nie mają większego znaczenia, choć mogą być przydatne w niektórych sprawach. Z technicznego i matematycznego punktu widzenia badania takie są stosunkowo proste, znacznie większym wyzwaniem dla badacza jest prawidłowe przeprowadzenie i interpretacja wyników badań osób dalej spokrewnionych, czemu poświęciłem dalsze podrozdziały.

3.2. Analiza pokrewieństwa w obrębie najbliższej rodziny

Za osoby blisko spokrewnione uważa się rodzeństwo i rodzeństwo przyrodnie oraz dziadków i wnuków. Udowodnienie pokrewieństwa tego typu badaniami genetycznymi jest trudniejsze, niż w przypadku pokrewieństwa bezpośredniego (rodzic – dziecko) a czasami bywa wręcz niemożliwe. Aby zrozumieć, dlaczego tak się dzieje, należy przyrzeć się sposobowi dziedziczenia DNA w obrębie bliskich krewnych (Rycina 5), a w szczególności uwzględnić wpływ, jaki na przekazywanie cech genetycznych ma proces rekombinacji chromosomów i dopływ „świeżej krwi” w postaci materiału genetycznego wnoszonego przez współmałżonków, nie będących z reguły krewnymi biologicznymi.

Biorąc pod uwagę fakt, że zarówno proces rekombinacji chromosomów rodziców jak i sam proces przekazania dziecku jednego z chromosomów każdej pary są w ogromnej mierze losowe, można oszacować, jaka pula DNA w genomach dwóch osób jest wspólna w wyniku posiadania wspólnej historii genetycznej. Takie wspólne odcinki DNA, otrzymane w toku dziedziczenia od przodków, nazywane są odcinkami identycznymi przez pochodzenie (często oznaczane są skrótem IBD od ang. *identical by descent*). Ich obecność możemy wykrywać np. prowadząc analizę mikrosatelitów i stwierdzając obecność jednego z kilkunastu występujących w populacji wariantów zarówno u dziecka jak i u rodzica. Co do osób nie spokrewnionych również może się zdarzyć, że będą one posiadały w niektórych miejscach swojego DNA identyczne warianty mikrosatelitów, jednak to podobieństwo wynika jedynie z tego, że ilość wariantów w populacji jest ograniczona. Takie odcinki DNA, które zawierają identyczne warianty sekwencji, lecz występują u osób nie spokrewnionych, nazywamy identycznymi przez stan (IBS – ang. *identical by state*). Analiza pokrewieństwa polega w swej istocie na określeniu prawdopodobieństwa, że identyczne warianty sekwencji DNA występujące u badanych osób są identyczne właśnie dlatego, że pochodzą od wspólnych przodków (IBD). W przypadku rodzica i dziecka wspólna pula DNA wynosi 50%, gdyż każde dziecko posiada połowę swojego DNA od jednego z rodziców i drugą połowę od drugiego. W każdym badanym mikrosatelicie musimy zatem zaobserwować przynajmniej jeden wspólny dla dziecka i rodzica wariant sekwencji. W przypadku rodzeństwa z jednych rodziców wspólna dla dowolnej dwójki dzieci pula DNA obejmuje jednak o wiele mniej, bo tylko po ok. 25% materiału genetycznego od każdego z rodziców. Aby to zrozumieć przeanalizujmy widoczną na Rycinie 5 rodzinę składającą się z osób oznaczonych jako U (matka), V (ojciec) oraz X i Y (syn i córka). Jak zaznaczono na rycinie, matka przekazała po połowie swojego DNA każdemu z dzieci. Należy jednak pamiętać, że chromosomy trafiające do komórek potomstwa są losową (w zasadzie) mieszaniną fragmentów

danej pary chromosomów u danego rodzica, powstała w wyniku rekombinacji (Rycina 2). Skoro ta mieszanina jest losowa, to oznacza, że każda komórka rozrodcza danego człowieka posiada chromosomy odmienne od wszystkich innych komórek rozrodczych tej osoby a ich jedyną wspólną cechą jest to, że powstały przez rekombinację z tej samej pary chromosomów wyjściowych. Innymi słowy, skoro szansa, że dowolny fragment każdego z chromosomów trafi do genomu dziecka wynosi 0.5 to szansa, że ten sam fragment trafi do genomów dwojga dzieci wynosi $0.5 \times 0.5 = 0.25$ czyli 25%. W sumie zatem, każde z dzieci danej pary rodziców posiada średnio po 50% DNA wspólnego z pozostałymi dziećmi tej pary. Taka sytuacja rodzi problemy interpretacyjne w badaniach pokrewieństwa domniemanego rodzeństwa, jeśli DNA rodziców nie jest dostępny do badań. Przykładowo, jeśli chcemy ustalić, czy osoby X i Y (Rycina 5) są rodzeństwem, możemy określić ich profile genetyczne, podobnie jak w przypadku standardowych badań bezpośredniego pokrewieństwa, i porównać je ze sobą, jednak interpretacja wyników będzie nieco odmienna. Przede wszystkim, brak wspólnych wariantów mikrosatelitów w profilach osób X i Y nie musi świadczyć o braku pokrewieństwa, gdyż, biorąc pod uwagę opisany wyżej sposób dziedziczenia, średnio tylko w co drugim mikrosatelicie możemy spodziewać się występowania wspólnego wariantu, wynikającego z posiadania wspólnych rodziców. Jeśli badane profile genetyczne obejmowały 20 mikrosatelitów, to możemy się spodziewać, że średnio w 10 z nich wystąpią u osób X i Y warianty wspólne, wynikające z posiadania tych samych rodziców. Problemem w takiej analizie jest jednak to, że nie ma żadnej bezpośredniej metody pozwalającej na stwierdzenie, czy identyczność wariantów obecnych w badanych profilach wynika z faktu posiadania przez badane osoby wspólnych przodków (IBD) czy też po prostu jest wynikiem czystego przypadku związanego z ograniczoną ilością wariantów mikrosatelitów obecnych w populacji (IBS). Rozstrzygnięcie tej kwestii jest możliwe na dwa sposoby. Jednym z nich jest podjęcie próby dotarcia do DNA rodziców badanych osób, co bywa możliwe nawet, jeśli owi rodzice nie żyją od wielu lat. Dzięki rozwojowi technik analizy genetycznej źródłem DNA może być bowiem nie tylko materiał pobrany bezpośrednio od żyjącej osoby, ale także np. próbki biologiczne w postaci przechowywanych w szpitalu bloczków parafinowych z fragmentami tkanek, pukiel włosów, szczoteczka do zębów czy maszynka do golenia. W skrajnych przypadkach, jeśli przeprowadzenie badań jest niezbędne z powodów prawnych, można uciec się nawet do ekshumacji. Jeśli materiał genetyczny rodziców uda się pozyskać i uzyskać z niego profil genetyczny przynajmniej jednego rodzica, analiza pokrewieństwa przybiera formę standardowej analizy ojcostwa/macierzyństwa. Takie rozwiązanie daje najwyższą szansę uzyskania wiarygodnego potwierdzenia pokrewieństwa.

Drugim sposobem rozwiązania kwestii ustalenia pokrewieństwa między domniemanym rodzeństwem jest maksymalne rozszerzenie panelu analizowanych mikrosatelitów i wykorzystanie odpowiednich wzorów matematycznych pozwalających na oszacowanie szansy pochodzenia ujawnionych wariantów od wspólnych rodziców. W tego typu obliczeniach istotną rolę odgrywa informacja o częstościach występowania wszystkich wariantów poszczególnych badanych mikrosatelitów w danej populacji. W praktyce badania takie ogranicza się zwykle do ok. 30 mikrosatelitów i ewentualnie innych markerów genetycznych, dla których dane populacyjne

są znane. Zdarza się jednak, że nawet wykorzystanie szerokiego panelu mikrosatelitów nie pozwala na uzyskanie współczynnika pokrewieństwa o wartości przekraczającej 100 tys.

Podobnie, co do zasady, przebiega analiza pokrewieństwa genetycznego między rodzeństwem przyrodnim. Uzyskanie wysokich wartości współczynnika pokrewieństwa jest w tym przypadku jeszcze większym wyzwaniem, gdyż rodzeństwo przyrodnie posiada tylko po 25% DNA pochodzącego od wspólnego rodzica. Również w tym przypadku najlepszym rozwiązaniem jest przeprowadzenie badań dodatkowego materiału genetycznego, pochodzącego od żyjących rodziców. Przykładowo: jeśli konieczne jest wyjaśnienie, czy dwoje dzieci z różnych matek posiada wspólnego ojca, którego DNA nie jest dostępny, wskazane jest poddanie badaniom genetycznym zarówno domniemanego przyrodniego rodzeństwa, jak i obu matek. Dzięki takiej poszerzonej analizie możliwe jest określenie, które warianty mikrosatelitów obecne w profilach domniemanego rodzeństwa pochodzą od matek i wykluczenie ich z dalszej analizy. W ten sposób pula wariantów DNA w genomach domniemanego rodzeństwa mogących pochodzić od ojca zostaje znacząco zawężona, co przekłada się na wyższy wynik obliczeń współczynnika pokrewieństwa.

W badaniach pokrewieństwa w obrębie bliskiej rodziny często pomocą służą markery haploidalne, tj. chromosom Y i mtDNA. Jeśli istnieje podejrzenie, że dwaj mężczyźni są spokrewnieni w linii ojcowskiej, ustalenie profilu ich chromosomów Y może pomóc w rozstrzygnięciu tej kwestii, podobnie jak analiza mtDNA w przypadku domniemania pokrewieństwa w linii matczynej. Oczywistą zaletą takich analiz jest fakt, że zarówno chromosom Y, jak i mtDNA przekazywane są z pokolenia na pokolenie jako kompletne cząsteczki. Analiza markerów haploidalnych ma jednak również swoje wady. Największą z nich jest niedobór danych populacyjnych opisujących częstości poszczególnych wariantów mtDNA i chromosomu Y. Dane populacyjne dla markerów haploidalnych gromadzone są najczęściej w toku badań naukowych prowadzonych na populacjach ludzkich zdefiniowanych zwykle jako grupy zamieszkujące określone regiony geograficzne. Istotą takich badań jest zebranie DNA od reprezentatywnej grupy ludzi wywodzących się z badanej populacji i określenie profili chromosomów Y bądź mtDNA poszczególnych osób. Uzyskane wyniki poddaje się analizie statystycznej w celu ustalenia, jak często poszczególne profile występują w danej populacji. Ponieważ zmienność mtDNA jest stosunkowo mała, zdarza się, że niektóre profile mtDNA obecne są w populacji z częstością nawet kilku procent, co wpływa negatywnie na możliwość jednoznacznego potwierdzenia pokrewieństwa osób, które taki właśnie częsty profil posiadają. W przypadku chromosomu Y zmienność genetyczna jest większa, jednak standardowo uzyskiwane profile Y-STR obejmują najczęściej kilka do kilkunastu mikrosatelitów, przez co profile te również nie zawsze są unikalne w skali populacji. Warto bowiem pamiętać, że ze względu na brak rekombinacji warianty mikrosatelitów obecne na chromosomie Y nie zapewniają tak wysokiego zróżnicowania profili STR chromosomu, jak w przypadku autosomów, zatem kilkanaście układów mikrosatelitarnych, które pozwoliłyby na otrzymanie profilu genetycznego unikalnego w skali całego świata, gdyby były zlokalizowane na autosomach, nie daje takich możliwości, jeśli znajdują się na chromosomie Y.

Informacje o częstości poszczególnych haplotypów mtDNA i Y-STR dostarczane są do publicznie dostępnych baz danych, które można wykorzystywać do oceny częstości profili markerów haploidalnych otrzymywanych w badaniach pokrewieństwa. Rozbudowa wspomnianych baz danych bazuje na aktywności badaczy z laboratoriów rozsianych na całym świecie, którzy zbierają i publikują informacje o zmienności Y-STR i mtDNA w wielu różnych populacjach. W przypadku chromosomu Y najbardziej rozbudowaną i wiarygodną bazą danych częstości jest YHRD (*Y chromosome Haplotype Reference Database* - www.yhrd.org) prowadzona przez grupę badaczy z Uniwersytetu Charite w Berlinie. Baza ta zawiera obecnie dane dla prawie 97 tys. chromosomów Y z całego świata, z czego ponad 40 tys. wywodzi się z populacji Europy. Dla mtDNA najbardziej godną zaufania bazą danych częstości jest prowadzona przez Instytut Medycyny Sądowej w Innsbrucku baza EMPOP (empop.org), licząca obecnie prawie 15 tys. profili. W Internecie dostępne są również serwisy utrzymywane przez firmy prowadzące analizy zmienności Y-STR i mtDNA dla potrzeb osób zainteresowanych badaniami genealogicznymi. W niektórych przypadkach profile Y-STR obecne w takich bazach danych są oznaczone w zakresie znacznie szerszym niż w bazach typowo naukowych, dają zatem teoretycznie większą szansę potwierdzenia pokrewieństwa w razie zgodności profilu, jednak możliwości wykorzystania tych danych dla potrzeb oceny częstości profili w populacji są ograniczone przez specyficzny sposób organizacji tych baz (np. podział profili na grupy w zależności od nazwiska lub regionu pochodzenia), trudności w jasnym zdefiniowaniu populacji, z których pochodzą profile w bazach i restrykcje związane z dostępem do danych np. tylko dla klientów danej firmy.

3.3. Analiza pokrewieństwa w obrębie dalszej rodziny

Genetyczna analiza pokrewieństwa staje się zadaniem tym bardziej skomplikowanym, im dalsze są domniemane więzy biologiczne między badanymi osobami. Spróbujmy spojrzeć na pokrewieństwo genetyczne pomiędzy dalszymi krewnymi z punktu widzenia genealoga i porównajmy jego perspektywę z perspektywą genetyka. Przede wszystkim musimy zauważyć, że klasyczne drzewo genealogiczne wywodzi członków danej rodziny od wybranej pary rodzicielskiej, która to para założycielska jest punktem wyjścia, niejako „pniem”, z którego wyrastają „gałęzie” w postaci dzieci, wnuków itp. Przyglądając się Rycinie 5 można stwierdzić, że klasyczne drzewo genealogiczne jest swego rodzaju odwróceniem drzewa pokrewieństwa genetycznego. Na omawianej rycinie możemy wskazać sześć różnych drzew genealogicznych (zapoczątkowanych przez pary AB, CD, EF, GU, IJ i KL), przy czym każda z osób z ostatniego pokolenia (dzieci) jest spokrewniona genetycznie z przedstawicielami przynajmniej czterech z owych sześciu drzew. Przykładowo, dla rodziny, której „pniem” jest para GH, możemy wytyczyć drzewo, w skład którego wchodzi: para OP, dzieci tej pary czyli T i U (wraz z małżonkami S i V) oraz dzieci tych dzieci, czyli W, X i Y. Analizując przepływ DNA z pary GH do ich prawnuków (W, X i Y) należy zauważyć, kierując się zasadą przekazywania dzieciom przez każdego rodzica 50% swojego DNA, że osoba W posiada tylko 12,5% DNA od pradziadka, założyciela rodu (H) i drugie tyle do prababci będącej jego żoną (G). W sumie zatem prawnuk odziedziczył 25% swojego DNA od pradziadków, zało-

życieli badanego rodu a ponadto, ponieważ jest mężczyzną, posiada chromosom Y odziedziczony od pradziadka (H). Nie posiada jednak mtDNA swej prababci ze strony ojca (G), gdyż jego dziadek ze strony ojca (P) nie mógł przekazać tej cząsteczki DNA swoim dzieciom. Prawnuk omawianej pary założycielskiej rodu widoczny na drugiej gałęzi drzewa (X) również posiada 25% DNA od tej samej pary swoich pradziadków w linii matczynej (GH), ale nie posiada ani chromosomu Y pradziadka G (matka X, oznaczona na diagramie jako U, nie mogła go przekazać) ani mtDNA prababci G (dziadek X, czyli P, nie przekazał go swojej córce, O). Podsumowując, osoby oznaczone jako W i X posiadają wspólnych pradziadków i są kuzynami, a jednak nie posiadają wspólnego chromosomu Y ani mtDNA, a jeśli chodzi o DNA jądrowy pochodzący od „pary założycielskiej”, to jego część wspólna dla obu tych osób wynosi $0,25 \times 0,25$ czyli 0,0625, tzn. tylko nieco ponad 6%!

Ze względu na wspomnianą wielokrotnie rekombinację DNA oraz z uwagi na ciągły napływ do puli genowej rodziny nowych wariantów DNA od osób wchodzących w związki małżeńskie z poszczególnymi przedstawicielami rodziny, którą przedstawia omawiane klasyczne drzewo genealogiczne, pokrewieństwo genetyczne ulega bardzo szybkiemu „rozmyciu” i już po trzech pokoleniach jest niemal niewykrywalne u potomków dowolnej wybranej pary ludzkiej. Warto przy tym pamiętać, że nawet losowo wybrane osoby z dowolnej, dużej populacji (np. polskiej), co do których nie mamy żadnych dowodów na istnienie pokrewieństwa w sensie genealogicznym, posiadają w swoim DNA ok. 1% fragmentów wspólnych (IBD) tylko dlatego, że w gruncie rzeczy wszyscy ludzie są do pewnego stopnia spokrewnieni, gdyż wywodzą się od stosunkowo niewielkiej liczby osób, których DNA było „materiałem wyjściowym” dla założenia danej populacji. Wyjątkami od opisanej reguły są chromosom Y i mtDNA, które nie podlegają rekombinacji a jedynie losowym mutacjom. Kilka pokoleń jest zwykle okresem zbyt krótkim, aby mutacje zataryły podobieństwo profili w obrębie linii matczynej (mtDNA) lub ojcowskiej (Y-STR), dlatego też to właśnie markery haploidalne są najczęściej stosowane w badaniach służących potwierdzeniu pokrewieństwa osób dalej spokrewnionych. Należy jednak zauważyć, że osoby tworzące nieprzerwane linie męskie i żeńskie to w danej grupie rodzinnej stosunkowo wąskie grono. Przykładowo w grupie, której związki rodzinne ukazano na Rycinie 5 można wyróżnić tylko dwie nieprzerwane linie matczyne (pierwsza, złożone z osób oznaczonych A, M, S i W oraz druga, obejmująca osoby oznaczone E, O, T, U, X i Y) oraz dwie nieprzerwane linie ojcowskie (H, P T i W oraz J, Q, V i X). Warto w tym miejscu zwrócić uwagę na specyficzną pozycję, jaką w opisywanym drzewie pokrewieństwa zajmuje osoba oznaczona jako Z. Status tej osoby w genetycznym drzewie pokrewieństwa ściśle zależy od tego, jaki posiada status w drzewie genealogicznym. Jeżeli jest to oficjalnie uznane dziecko ze związku osób oznaczonych H i J to jej pokrewieństwo z pozostałymi członkami rodziny nie budzi wątpliwości, jeśli jednak, jak to się często zdarza, osoba oznaczona jako Z jest uznawana oficjalnie za dziecko pary IJ, to wśród potomstwa tej pary pojawiają się dwie odmienne linie chromosomu Y, które mogą być przekazywane kolejnym pokoleniom. Analogicznie, jeśli osoba Z zostanie uznana oficjalnie za dziecko pary GH, to w rodzinie tej pojawia się nowa linia mtDNA, która jednak nie zostanie przekazana dalej, gdyż Z jest mężczyzną. Sytuacje, w których wśród osób stanowiących gałęzie

drzewa genealogicznego badanej rodziny pojawia się linia chromosomu Y lub mtDNA nie znajdująca odzwierciedlenia w danych genealogicznych zdarzają się w praktyce tym częściej, im więcej pokoleń obejmuje badanie. Z tego względu podejmując badania genetyczne dużych, wielopokoleniowych rodzin, zalecane jest pozyskanie i poddanie analizie materiału genetycznego od przynajmniej kilku przedstawicieli każdej znanej gałęzi danej rodziny.

Specyficznym zastosowaniem współczesnej biologii molekularnej jest poszukiwanie więzów rodzinnych przez osoby posiadające wspólne nazwisko. Ponieważ w naszej kulturze nazwisko dziedziczy się w linii męskiej, podobnie jak chromosom Y, możliwe jest podjęcie próby potwierdzenia wspólnego pochodzenia mężczyzn noszących dane nazwisko przez analizę profili ich chromosomów Y. Wyniki takiej analizy przedstawia się najczęściej jako tabelę z zaznaczonymi, specyficznymi dla poszczególnych mężczyzn wariantami mikrosatelitów. Znacznie czytelniejszym, choć stosunkowo rzadko spotykanym sposobem prezentacji takich danych jest przedstawienie ich w postaci tzw. sieci filogenetycznej (Rycina 6). Sieć taka w sposób graficzny prezentuje zależności między haplotypami poszczególnych badanych osób i pozwala na pierwszy rzut oka ocenić zróżnicowanie chromosomów Y w obrębie badanej grupy. Na Rycinie 6, przedstawiającej schemat sieci wykonanej w oparciu o analizę haplotypów chromosomu Y 27 mężczyzn noszących jedno nazwisko wyraźnie widać, że w badanej grupie występują potomkowie przynajmniej dwóch linii ojcowskich, rozróżnianych na podstawie ilości mutacji oddzielających poszczególne haplotypy. W prezentowanej sieci niejasny jest jedynie status osoby, której haplotyp oznaczono kolorem szarym. Chromosom Y tej osoby różni się od chromosomów jednej z ujawnionych linii w pięciu z piętnastu badanych miejsc, co może oznaczać, że chromosom ten należy do odrębnej linii ojcowskiej. Nie można jednak wykluczyć, że istnieje dalekie pokrewieństwo między chromosomem Y tej osoby a pozostałymi chromosomami grupy. Aby wyjaśnić tego typu wątpliwości przeprowadzić można wspomniane wcześniej badania Y-SNP, dzięki którym możliwe jest opisanie zmienności DNA chromosomu Y w kategoriach stabilnych ewolucyjnie haplogrup. Badanie Y-SNP przeprowadzone dla opisywanego przypadku wykazuje, że chromosom Y, co do którego pokrewieństwa z badaną grupą były wątpliwości rzeczywiście posiada mutacje Y-SNP kwalifikujące go do haplogrupy I, podczas gdy pozostałe z badanych chromosomów należą do podgrup haplogrupy R. Uprawniony jest zatem wniosek, że w badanej grupie mężczyzn, noszących jedno nazwisko, obecne są chromosomy pochodzące z przynajmniej trzech różnych linii ojcowskich. Dodatkową korzyścią wypływającą z analizy podobieństwa haplotypów Y-STR z wykorzystaniem sieci filogenetycznych jest możliwość oszacowania wieku poszczególnych gałęzi sieci, co pozwala na określenie, z pewnym przybliżeniem, kiedy żył ostatni wspólny, męski przodek badanych osób.

3.4. Ustalanie pochodzenia biogeograficznego

Ostatnim z aspektów analizy markerów haploidalnych, o którym warto wspomnieć, jest możliwość wykorzystania otrzymanych wyników do szacowania tzw. pochodzenia biogeograficznego, czyli populacji pierwotnej, z której wywodzi się dana linia ojcowska lub matczyzna. Dla badań tego typu szczególnie przydatne są ha-

plotypy chromosomu Y, ze względu na swoje stosunkowo wysokie zróżnicowanie oraz specyficzną cechę określaną jako patrilokalność, która polega na tym, że w większości kultur mężczyźni mają tendencje do zamieszkiwania rejonów niezbyt odległych od miejsca swego miejsca urodzenia. W wyniku takiego zachowania następuje gromadzenie się w określonych lokalizacjach geograficznych charakterystycznych dla nich haplotypów. Ujawnienie w DNA danej osoby haplotypu chromosomu Y specyficznego dla danego regionu geograficznego może zatem stanowić wskazówkę dotyczącą pochodzenia linii ojcowskiej, do której ta osoba należy.

Ustalanie pochodzenia biogeograficznego jest techniką, która silnie powiązana jest z dostępnością kompleksowych danych o częstościach poszczególnych haplotypów w wielu różnych populacjach. Biorąc pod uwagę dzisiejszy stan baz danych oraz zakres oznaczanych powszechnie haplotypów chromosomu Y można stwierdzić, że badania pochodzenia znajdują się jeszcze w powijkach, choć w specyficznych przypadkach bywają przydatne nawet przy ograniczonym zestawie danych dostępnych dziś.

Podobne badania pochodzenia linii matczynej napotykać na dwie ważne przeszkody: mniejszą zmienność mtDNA oraz większą skłonność kobiet do zmiany miejsca zamieszkania. W efekcie opisanych zjawisk w przypadku mtDNA efekt akumulacji nowych mutacji w konkretnej lokalizacji geograficznej jest znacznie ograniczony. W przypadku mtDNA badania pochodzenia biogeograficznego są użyteczne zwykle tylko w sytuacjach, w których konkretna linia matczyzna nosi mutację charakterystyczną np. dla konkretnej grupy etnicznej.

4. Kilka uwag końcowych dla osób rozważających przeprowadzenie badań genetycznych

Omawiane w niniejszym rozdziale metody ustalania pokrewieństwa są, w mniejszym lub większym zakresie, oferowane przez wiele różnych laboratoriów, w tym przez instytucje prywatne.

Ze względu na fakt, iż analizy tego typu wymagają rzetelnej wiedzy naukowej i dysponowania odpowiednio wyposażoną pracownią genetyki molekularnej, warto przed podjęciem decyzji o przeprowadzeniu badań genetycznych zasięgnąć szczegółowej informacji odnośnie doświadczenia danego laboratorium w tego typu badaniach i jego rzetelności. Zdarzają się niestety sytuacje, w których osoby poszukujące odpowiedzi na skomplikowane pytania związane z analizą bliższego bądź dalszego pokrewieństwa trafiają do instytucji, których personel nie posiada odpowiedniej wiedzy i doświadczenia do prowadzenia takich analiz, a które reklamują się jako jednostki o dużym potencjale naukowym, używając w swoich nazwach takich słów jak „instytut” czy „zakład”, kojarzących się jednoznacznie z jednostkami naukowymi. Prowadzi to czasem do sytuacji, w których osoba zlecająca badania wydaje duże kwoty, nie uzyskując w zamian oczekiwanych informacji, lub, co gorsza, uzyskując informacje niezgodne ze stanem faktycznym. Często okazuje się potem, że możliwe było zaoszczędzenie pieniędzy przez szczegółową i popartą odpowiednią wiedzą naukową analizę danej sprawy, której nierzetelna placówka wykonująca badanie nie przeprowadziła.

Niestety osobie nie posiadającej specjalistycznej wiedzy trudno jest na podstawie jedynie oferty internetowej czy ogłoszenia w gazecie ocenić rzetelność da-

nej instytucji, tym bardziej, że oferty takie bardzo często przedstawiają w sposób przesadzony, lub nawet nieprawdziwy, możliwości danego laboratorium. Dlatego moim zdaniem najbezpieczniejszym sposobem uniknięcia omawianych problemów jest przeprowadzenie badań genetycznych, szczególnie jeśli nie jest to prosta analiza bezpośredniego pokrewieństwa, w placówce akademickiej o uznanej renomie.

5. Podsumowanie

Celem tego artykułu było wprowadzenie Czytelników w świat genetyki molekularnej i jej różnorodnych zastosowań w badaniach bliższego i dalszego pokrewieństwa. Zdaję sobie sprawę, że tematyka ta jest dość trudna i hermetyczna, przez co często możliwości badań genetycznych są przeceniane, bądź niedoceniane przez ich potencjalnych beneficjentów. Staralem się zatem możliwie obiektywnie i przejrzysto przedstawić możliwości, jakie oferuje genealogii nowoczesna genetyka molekularna, ale również wskazać jej ograniczenia. Mam nadzieję, że niniejszy artykuł pozwoli lepiej zrozumieć potencjalne zyski płynące z badań genetycznych i oszacować ich przydatność dla prowadzonych, własnych badań genealogicznych.

6. Słownik podstawowych pojęć

Autosom – chromosom nie warunkujący płci. Człowiek posiada w swoich komórkach 22 pary autosomów, oznaczone numerami od 1 do 22 i dwa chromosomy płciowe (XX lub XY).

Chromosom – pojedyncza, ciągła cząsteczka DNA zawierająca zestaw genów.

Chromosom płciowy – chromosom biorący udział w warunkowaniu płci organizmu. U człowieka chromosomy płciowe oznaczane są symbolami X i Y, przy czym obecność w komórce chromosomu Y warunkuje płć męską.

DNA – kwas deoksyrybonukleinowy, cząsteczka chemiczna, przyjmująca fizycznie postać nierozgałęziającej się nici, przechowująca informacje o budowie wszelkich białek i innych biomolekuł koniecznych do prawidłowego funkcjonowania komórek.

DNA mitochondrialny – (w skrócie: mtDNA) kolista cząsteczka DNA występująca w komórce odrębnie od reszty genomu, wewnątrz mitochondriów i dziedziczona u ssaków wyłącznie w linii matczynej.

Gen – fragment nici DNA zapisujący informację o budowie jednego z białek (czasem też cząsteczki RNA) potrzebnych w komórce.

Genom – zestaw cząsteczek DNA (chromosomów) zawierających kompletną informację genetyczną, charakterystyczną dla komórek danego gatunku i osobnika danej płci. W skład kompletnego genomu wchodzi po jednym chromosomie z każdej pary, zatem w każdej komórce ciała obecne są dwa genomy, a w komórkach rozrodczych jeden genom.

Genotyp – zapis wariantów sekwencji DNA występujących w konkretnym miejscu (*locus*) na obu chromosomach danej pary. Genotyp może być homozygotyczny (jeśli na obu chromosomach występuje taki sam wariant sekwencji) lub heterozygotyczny (jeśli warianty na obu chromosomach są różne).

Haplotyp – zestaw charakterystycznych cech (mutacji) występujących na pojedynczej, ciągłej nici DNA.

IBD – skrót od wyrażenia *identical by descent*, które oznacza wariant polimorfizmu (patrz: polimorfizm) występujący w genomach badanych osób dlatego, że otrzymały go od wspólnego przodka.

IBS - skrót od wyrażenia *identical by state*, które oznacza wariant polimorfizmu DNA (patrz: polimorfizm) identyczny w genomach badanych osób, lecz nie pochodzący od wspólnego przodka.

Kodon – trójliterowa sekwencja (patrz: sekwencja) nukleotydów (patrz: nukleotydy), rozpoznawana przez mechanizmy komórkowe odczytujące informację genetyczną.

Nukleotydy – cząsteczka chemiczna stanowiąca podstawowy moduł budowy DNA. W ludzkim DNA występują 4 nukleotydy, określane konwencjonalnie jako: adenina, cytozyna, guanina i tymina. Bardzo często stosuje się skrócony zapis nazw nukleotydów w postaci liter A, C, G i T.

Marker – fragment DNA występujący w populacji w więcej niż jednym wariantcie sekwencji (patrz: sekwencja), używany w celu uzyskania profilu genetycznego.

Mikrosatelita – (także: sekwencja STR) wysoce zmienny fragment DNA, występujący w populacji w kilku, kilkunastu lub nawet kilkudziesięciu wariantach.

mtDNA – patrz: DNA mitochondrialny.

Mutacja – losowa zmiana sekwencji DNA.

Mutacja punktowa – patrz: SNP.

Polimorfizm – zmienność sekwencji DNA w obrębie danego gatunku lub populacji.

Populacja – grupa osobników danego gatunku wyróżniona na podstawie charakterystycznych cech lub pochodzenia geograficznego. Populacje ludzkie definiowane są zwykle w oparciu o miejsce zamieszkania lub przynależność etniczną danej grupy ludzi.

Profil genetyczny – opis charakterystycznych cech DNA danej osoby wykonany w oparciu o wyniki badań wariantów sekwencji DNA występujących w wybranych miejscach chromosomów lub jednego chromosomu. Profilem może być np. zbiór genotypów (patrz: genotyp) lub haplotyp chromosomu Y czy mtDNA (patrz: haplotyp).

Rekombinacja – proces wymienia się chromosomów danej pary fragmentami nici DNA. Zjawisko to zachodzi podczas produkcji komórek rozrodczych i prowadzi do powstania nowego chromosomu, niosącego w swojej sekwencji mieszaninę cech genetycznych charakterystycznych dla chromosomów wyjściowych.

RNA – kwas rybonukleinowy, substancja chemiczna podobna do DNA, spełniająca w komórkach szereg funkcji pomocniczych, m.in. pośredniczenie w odczycie informacji zawartej w DNA.

Sekwencja – kolejność ułożenia nukleotydów w danej nici DNA.

SNP – (także: mutacja punktowa) zmienny fragment DNA występujący w populacji w niewielkiej liczbie wariantów, najczęściej w dwóch.

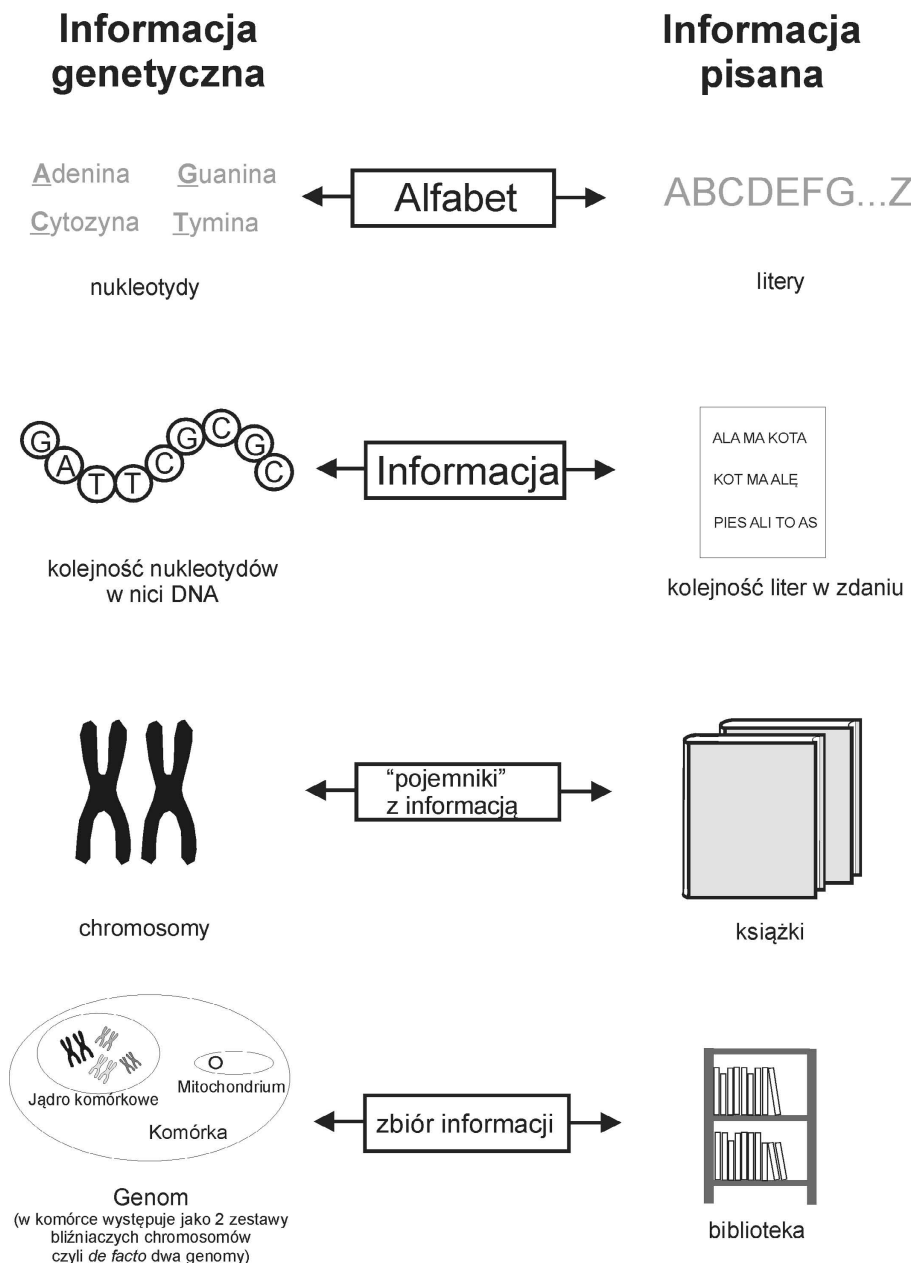
STR – patrz: mikrosatelita.

Y-STR – sekwencja mikrosatelitarna (STR) występująca na chromosomie Y.

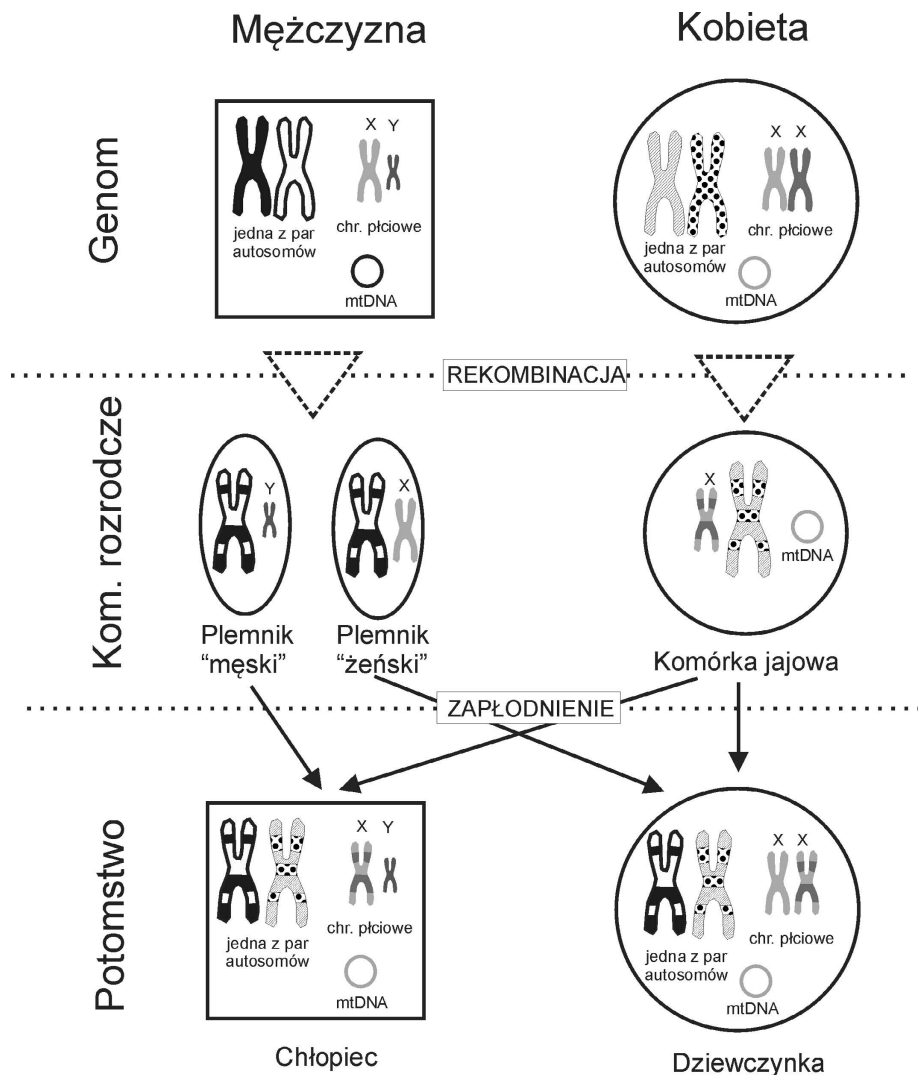
7. Bibliografia

- L. L. Cavalli – Sforza, *Genes, Peoples and Languages*, University of California Press, 2000.
- L. L. Cavalli – Sforza, F. Cavalli – Sforza, *The Great Human Diasporas*, Basic Books, 1995.
- C. Fitzpatrick, A. Yeiser, *DNA & Genealogy*, Rice Books Press, 2005.
- C. Fitzpatrick, *Forensic Genealogy*, Rice Books Press, 2005.
- S. Wells, *Deep Ancestry. Inside the Genographic Project*, National Geographic Society 2006.
- M. Woźniak, T. Grzybowski, *Zastosowanie biologii molekularnej w medycynie sądowej*, [w:] G. Drewa i T. Ferenc (red.), *Genetyka medyczna. Podręcznik dla studentów*. Elsevier Urban & Partner, Wrocław 2011.
- T. Grzybowski, U. Rogalla, *Podstawy genetyki populacyjnej i wybrane zagadnienia z genetyki ewolucyjnej człowieka*, [w:] G. Drewa i T. Ferenc (red.), *Genetyka medyczna. Podręcznik dla studentów*. Elsevier Urban & Partner, Wrocław, 2011.

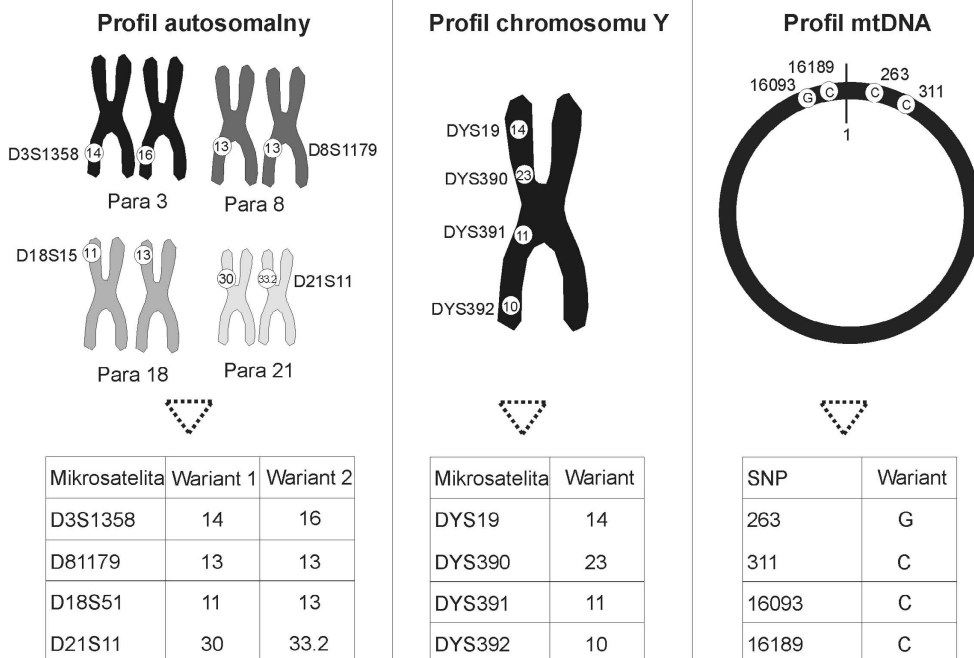
Rycina 1. Schematyczny opis istoty zapisu informacji genetycznej w DNA człowieka.
(Szczegółowy opis w tekście).



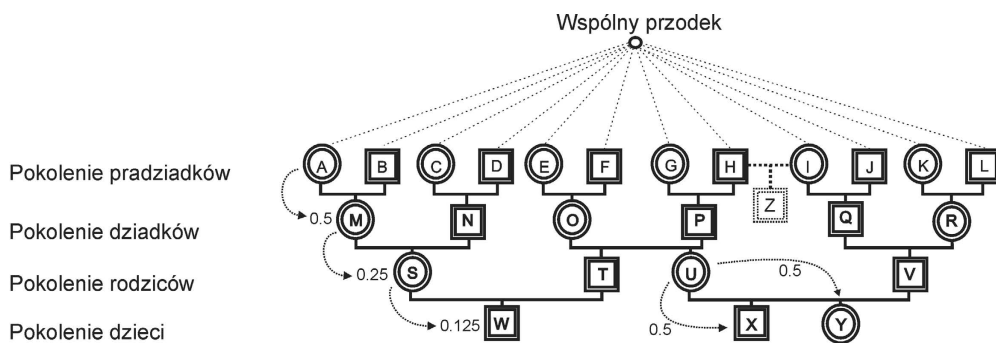
Rycina 2. Schemat przekazywania informacji genetycznej. Aby zilustrować istotę procesu rekombinacji i zachować czytelność, przedstawiono tylko jedną z 22 par ludzkich chromosomów nie płciowych (autosomów) obecnych w komórce. Aby zaznaczyć różnice sekwencji chromosomów, każdy z chromosomów rodzicielskich danej pary wypełniono odmiennym deseniem.



Rycina 3. Ilustracja uzyskiwania profilu genetycznego w oparciu o analizę autosomów, mikrosatelitów chromosomu Y i sekwencji mtDNA. Okręgi symbolizują wybrane miejsca na nici DNA tworzącej dany chromosom, które poddawane są badaniom. Wewnątrz okręgów zaznaczono symbol wariantu sekwencji DNA obecnego w badanym miejscu. W przypadku mikrosatelitów warianty są oznaczane liczbami, w przypadku mtDNA podaje się nukleotydy występujący w badanym miejscu nici.



Rycina 5. Przykładowe drzewo pokrewieństwa obejmujące cztery pokolenia. Strzałki pokazują ilość materiału genetycznego danej osoby przekazywanej kolejnym pokoleniom. Przykładowo, osoba A przekazuje swojej córce 50% swego DNA, ta przekazuje swojemu dziecku 25% DNA utrzymanego od matki, a córka (S) przekazuje swemu dziecku już tylko 12.5% DNA wywodzącego się od prababki w linii matczynej. Dalszy opis i odwołania do tej ilustracji znajdują się w tekście.



Rycina 6. Schemat sieci haplotypów Y-STR obejmującej 27 osób o tym samym nazwisku. Każdy okrąg reprezentuje pewien haplotyp chromosomu Y określony w oparciu o 15 uktądów Y-STR. Średnica każdego okręgu odzwierciedla liczbę osób posiadających dany haplotyp. Najmniejsze okręgi to pojedyncze osoby, w większych okręgach zaznaczono liczbę osób posiadających chromosom Y o danym haplocypie. Linie łączące okręgi reprezentują mutacje, których rezultatem jest zróżnicowanie haplotypów wywodzących się z jednej linii ojcowskiej. Zróżnicowanie to jest proporcjonalne do czasu, jaki upłynął od powstania haplotypu wyjściowego dla danej linii ojcowskiej. Długość linii odzwierciedla liczbę mutacji odróżniających haplotypy (podano również szarą czcionką). Podobne sieci konstruuje się również dla zilustrowania pokrewieństwa pomiędzy haplotypami mtDNA.

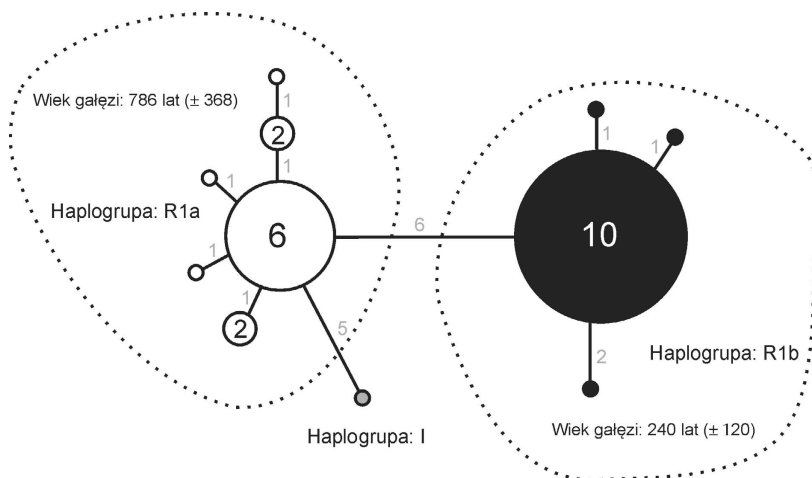


Tabela 1. Analiza bezpośredniego pokrewieństwa na przykładzie badania dziecka i domniemanego ojca. W profilu dziecka pogrubiono te warianty, które mogą pochodzić od domniemanego ojca a w profilu ojca pogrubiono te warianty, które mógł przekazać dziecku. W niektórych mikrosatelitach w profilu dziecka (np. D3S1358) nie można jednoznacznie ustalić, który z wariantów pochodzi od ojca, jednak nie uniemożliwia to wykonania obliczeń.

Mikrosatelita	Dziecko		Domniemany ojciec		Współczynnik pokrewieństwa
	Wariant 1	Wariant 2	Allel 1	Allel 2	
D3S1358	15	15	15	15	4.016064
D19S433	14	14	14	14	2.673797
D2S1338	18	23	18	22	2.450980
D16S539	8	11	8	11	41.185840
D18S51	11	16	14	16	1.295337
D1S1656	17.3	17.3	13	17.3	3.125000
D10S1248	13	16	12	16	1.923077
D2S441	14	14	14	14	3.362475
TH01	7	9.3	9	9.3	0.761035
vWA	17	18	18	18	2.382087
D21S11	31.2	32.2	28	31.2	2.631579
D12S391	17	21	17	19	2.358491
D8S1179	9	15	9	9	33.333333
FGA	22	22	22	24	2.551020
ACTBP2	27.2	27.2	27.2	27.2	13.513514

Współczynnik pokrewieństwa (ojcostwa) $PI = 366\ 841\ 903$
Prawdopodobieństwo pokrewieństwa (ojcostwa) $W = 99.99999727\%$