

Ćwiczenie C06 Genetyka populacyjna

Struktura genetyczna populacji Zróżnicowanie genetyczne Podobieństwo genetyczne

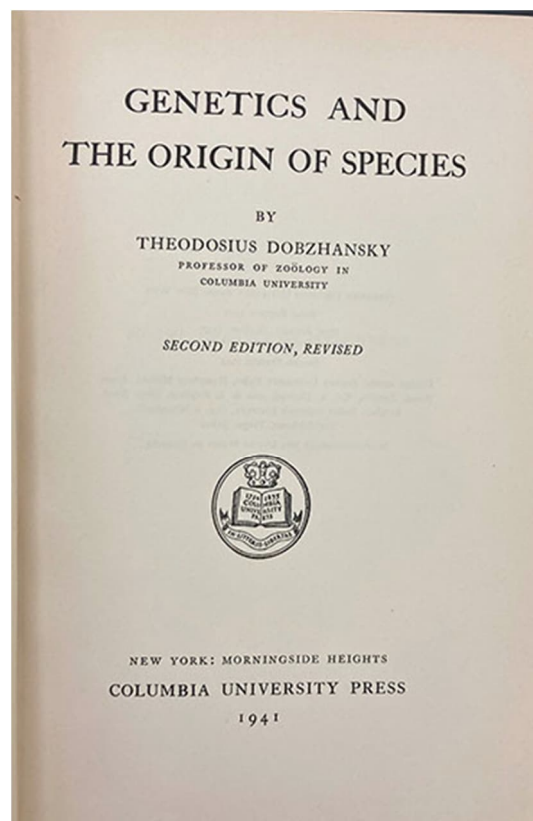
Kornelia Polok

1. Struktura genetyczna populacji

➔ 1.1. Genetyka populacyjna i jej znaczenie

Genetyka populacyjna zajmuje się badaniem składu genetycznego populacji, w tym zmianami częstości alleli, genotypów i fenotypów w odpowiedzi na selekcję, dryf genetyczny, mutacje. Genetyka populacyjna jest bezpośrednio związana z badaniem zjawisk ewolucyjnych, ponieważ selekcja naturalna jest głównym procesem zmieniającym skład populacji.

Obecnie nie ulega wątpliwości, że nie można zrozumieć mechanizmów ewolucji bez zrozumienia zmienności i jej genetycznego podłoża. Jednak nie zawsze było to oczywiste i nadal, dla niektórych nie jest oczywiste. Pierwszym człowiekiem, który powiązał darwinowską teorię powstawania gatunków ze zmiennością genetyczną był Theodosius Dobzhansky. Jako pierwszy zrozumiał on znaczenie zmienności genetycznej w populacjach naturalnych. Wywodząc się jednocześnie ze szkoły Thomasa Morgana (był doktorantem w jego laboratorium), a także posiadając tradycyjną wiedzę i doświadczenie dotyczące historii naturalnej (wykształcił się w Kijowie), Dobzhansky stworzył ramy, które połączyły materialne podstawy ewolucji (geny) opisywane przez genetykę klasyczną ze zróżnicowaniem gatunków obserwowanym przez ewolucjonistów. Było to możliwe dzięki jego kreatywności i interdyscyplinarnym umiejętnościom.



Rys. 1.1. Strona tytułowa II wydania książki T. Dobzhansky'ego.

Był jednym z pierwszych naukowców, który zrozumiał, że do poznania mechanizmów ewolucji niezbędna jest wiedza z wielu obszarów.

W książce „Genetics and the Origin of Species” (1937, 1951; ISBN 0-231-05474-0) T. Dobzhansky pierwszy wykazał związek pomiędzy mikroewolucją (procesy obserwowane na poziomie populacji i gatunku), a makroewolucją (duże zmiany ewolucyjne prowadzące do powstawania wyższych taksonów). Przekonująco zademonstrował on, że badania laboratoryjne materiału genetycznego dostarczają informacji o mechanizmach ewolucji. Książka Dobzhansky’ego nie tylko przyczyniła się do lepszego zrozumienia mechanizmów ewolucji, ale także stworzyła ramy dla współpracy pomiędzy różnymi specjalistami. Dlatego ma ona nie tylko znaczenie naukowe, ale należy ją uznać za podstawę współczesnego rozumienia badań – podejścia interdyscyplinarnego, w którym istotne jest połączenie wysiłków różnych specjalistów. Obserwowane cechy należy widzieć w szerszym kontekście, nawet wówczas, gdy analiza dotyczy takich zagadnień, wydawałoby się odległych od rozważań ewolucyjnych, jak rozprzestrzenianie się chorób w populacjach ludzkich. Nie bez przyczyny, książka Dobzhansky’ego została uznana za jedną z najważniejszych pozycji będących podstawą współczesnej syntetycznej teorii ewolucji i w 1990 roku została zaliczona do najważniejszych książek świata zachodniego (Great books of the Western World) razem z pracami Karola Darwina.

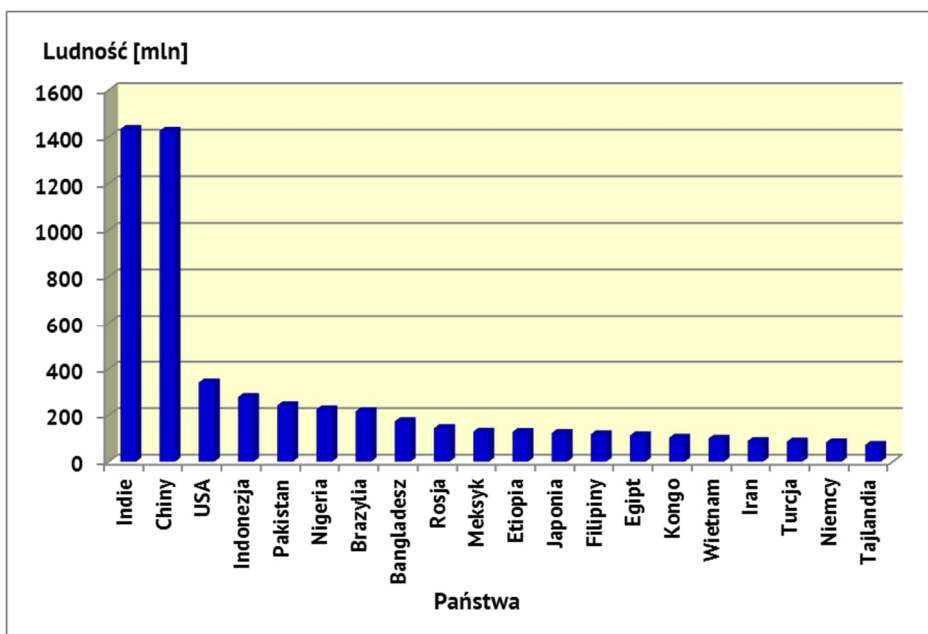
➔ 1.2. Populacja biologiczna i jej struktura

Populacja biologiczna to grupa osobników danego gatunku występująca na danym obszarze i w danym czasie, zdolna do wzajemnego krzyżowania się. Jest to jednostka ewolucyjna podlegająca selekcji naturalnej.

Populacja jest opisywana zestawem parametrów, przy czym różnią się one w socjologii, ekologii i genetyce populacyjnej. W socjologii populację opisuje się za pomocą rasy, narodowości religii i demografii.

Socjologia na ogół opisuje tylko populacje ludzkie. Ekologia opisuje populacje za pomocą struktury wiekowej, przestrzennej, tempa wzrostu, zagęszczenia.

Natomiast genetyka populacyjna opisuje populację za pomocą parametrów informujących o składzie genetycznym i jego zmianie na

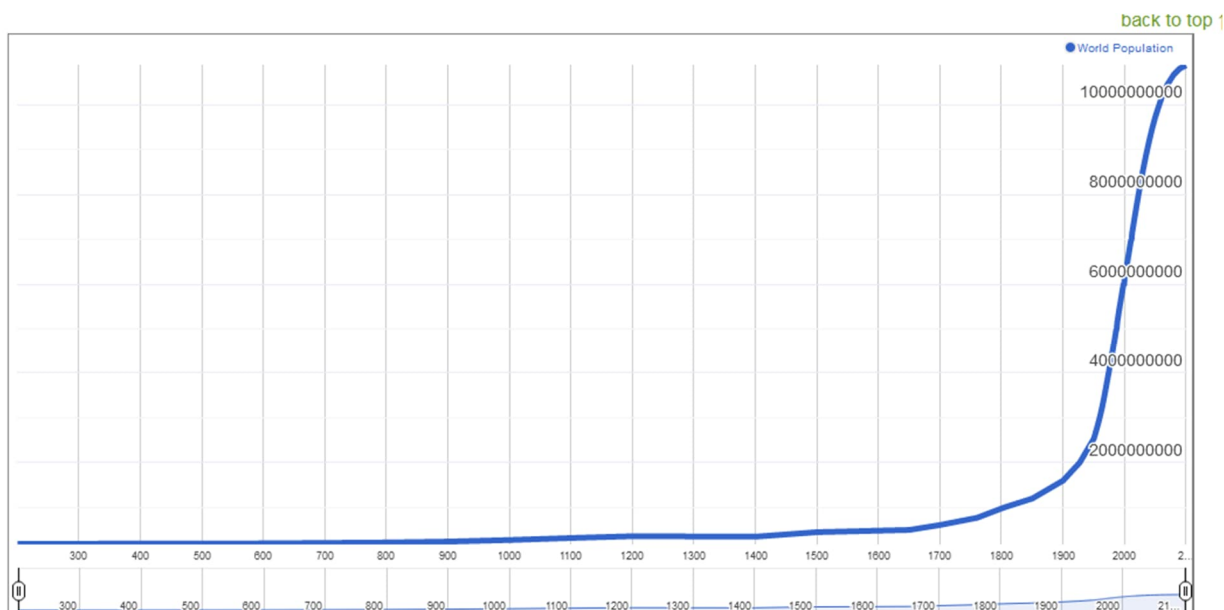


Rys. 1.2.1a. Wielkość populacji najbardziej liczebnych państw.

przestrzeni czasu. Czynniki takie jak zagęszczenie, struktura wiekowa mają wpływ na parametry genetyczne i w tym zakresie są uwzględniane w genetyce populacyjnej.

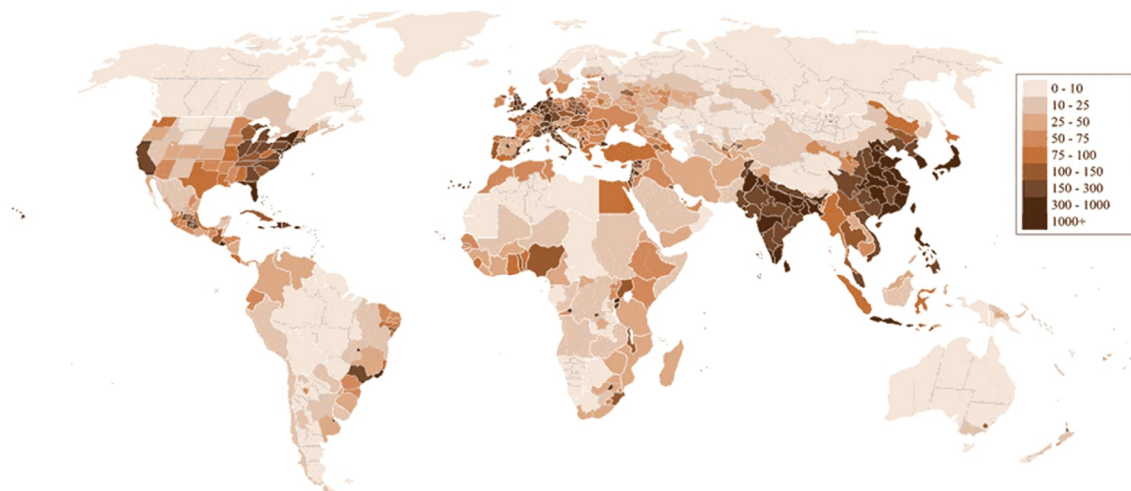
1.2.1. Wielkość populacji ludzkiej

W 2023 r. populacja ludzka na świecie to 8,07 mld (stan na 2023.11.20. Worldometers). W stosunku do roku 2022 r. populacja ludzka zwiększyła się o 0,88% (70 mln). Szacuje się, że na Ziemi żyło około 106 mld ludzi. Największą populację mają Indie i Chiny, po 1,4 mld. Gwałtowny wzrost liczebności populacji ludzkiej obserwuje się od 1900 r. Liczebność 1 mld populacja ludzka osiągnęła w 1804 r.



Rys. 1.2.1.b. Tempo wzrostu populacji ludzkiej od średniowiecza.

Szacuje się, że 8 tys. lat p.n.e. populacja ludzka liczyła około 5 mln. Od tego czasu wzrastała w tempie 0,05% rocznie, do około 200 mln w okresie średniowiecza. Gwałtowny wzrost populacji nastąpił wraz z rewolucją przemysłową. Liczebność populacji ludzkiej podwoiła się w ciągu 130 lat (od 1804 r. do 1930 r.) do 2 mld. Osiągnięcie 4 mld zajęło już tylko 44 lata (1974 r.). W podobnym tempie (57 lat) populacja ludzka podwoiła się do blisko 8 mld w 2021 r. Obecnie



Rys. 1.2.1.c. Gęstość populacji ludzkiej.

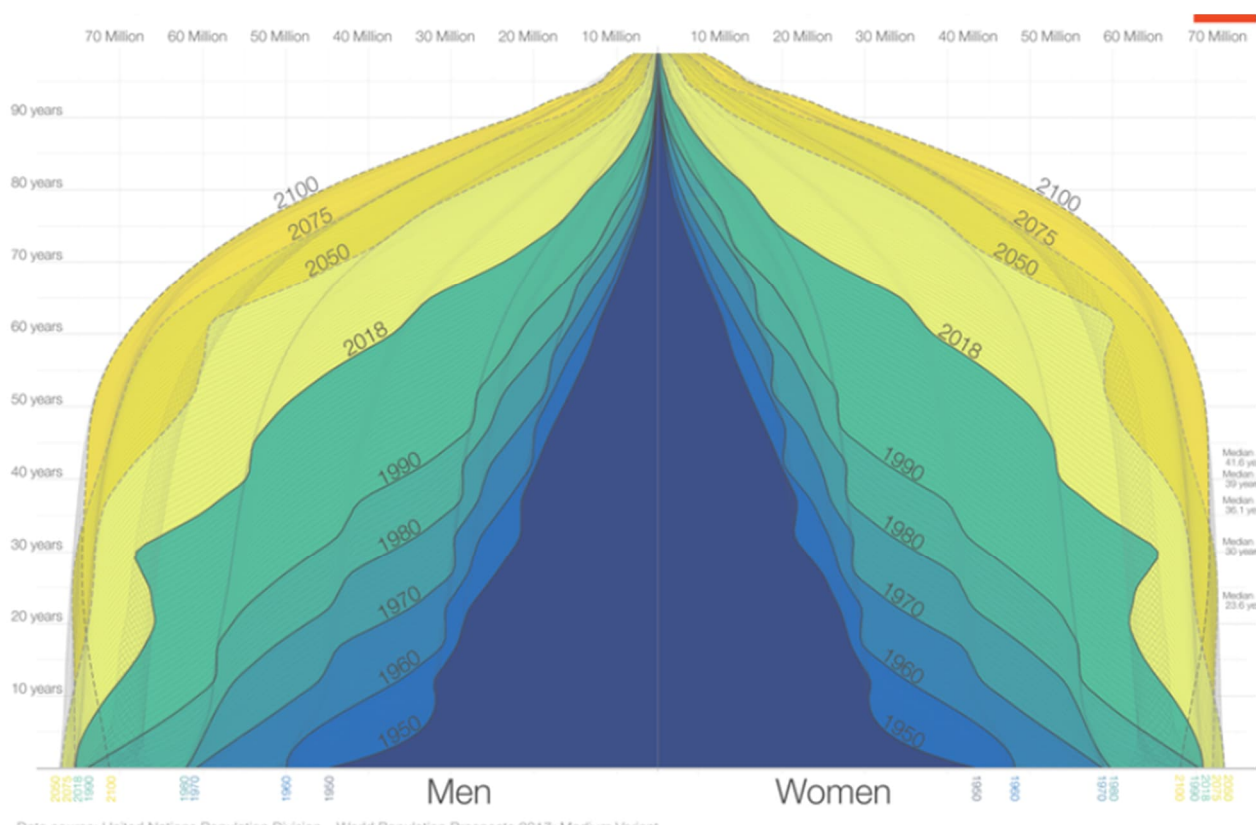
tempo wzrostu spadło. Szacuje się, że kolejne podwojenie będzie miało miejsce za 200 lat. Przewiduje się, że w 2050 r. populacja ludzka będzie liczyła 9,7 mld. Największe zagęszczenie populacji ludzkiej jest w Europie Zachodniej, USA, Chinach i Indiach.

1.2.2. Struktura wiekowa populacji ludzkiej

Struktura wiekowa przedstawia skład populacji z podziałem na grupy wiekowe. W przypadku populacji ludzkiej najczęściej wyróżnia się grupy co 5 lat. Strukturę wiekową przedstawia się na wykresie w postaci poziomych słupków, które tworzą piramidę. Kobiety na ogół przedstawia się po prawej stronie, mężczyzn po lewej stronie.

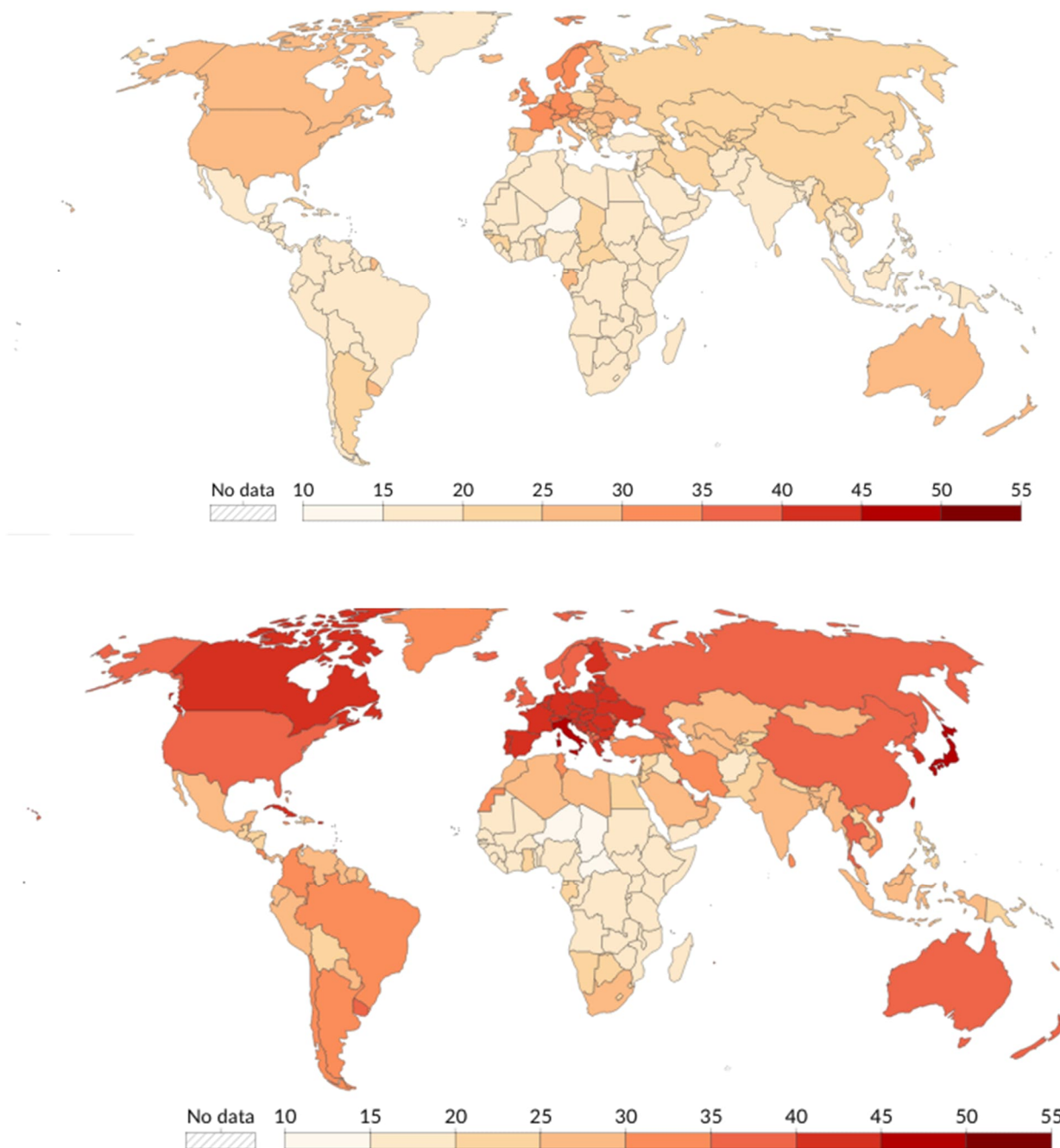
W roku 1950 struktura wiekowa populacji ludzkiej stanowiła typową piramidę (Rys. 1.2.2a, granatowy trójkąt), w której najwięcej było osób młodych. W kolejnych latach struktura wiekowa się zmieniła, wzrastał udział osób w wieku produkcyjnym i poprodukcyjnym. Szacuje się, że w roku 2100 liczba osób w wieku 50 lat będzie prawie równa liczbie nowych urodzeń. Wpływ na taki kształt ma zarówno liczba urodzeń jak i mniejsza śmiertelność.

W roku 1950 urodziło się 97 mln dzieci, w roku 2018 – 143 mln. W tym samym czasie spadła śmiertelność dzieci, co sprawia, że struktura wiekowa nadal przypominała piramidę. Jednakże przewidywany wzrost liczby urodzeń między 2018 r. a 2100 r. jest nieznaczny. W tym samym czasie wzrasta długość życia.



Rys. 1.2.2a. Piramida wiekowa populacji światowej w latach 1950-2100.

Mediana jest kolejnym wskaźnikiem opisującym rozkład wieku w populacji. Mediana wieku dla populacji światowej wynosi 30 lat. Najwyższą medianę wieku ma Japonia, 48,4 lata, najniższą Niger, 14,5 roku. Generalnie, mediana wieku jest wyższa w bogatych, rozwiniętych państwach Ameryki Północnej, Europy i Wschodniej Azji.



Rys. 1.2.2b Porównanie mediany wieku w poszczególnych państwach w roku 1950 (górze) oraz w roku 2021 (dół)



1.2.3. Struktura wiekowa populacji polskiej

Na podstawie poniższej tabeli proszę wykreślić wykres struktury wiekowej dla:

- Ludności Polski w 1980 r. Jaki typ struktury wiekowej wówczas występował w Polsce?
- Ludności Polski w 2023 r. Jaki typ struktury wiekowej występował w 2023 r. w Polsce?
- Proszę porównać strukturę wiekową ludności Polski w 1980 r. oraz w 2023 r.

1980			2023		
Klasa	Kobiety	Mężczyźni	Klasa	Kobiety	Mężczyźni
0-4	1588,5	1664,2	0-4	826,9	870,4
5-9	1399,4	1465,1	5-9	962	1015,5
10-14	1232,2	1289,6	10-14	1007,9	1063,8
15-19	1347,9	1424,3	15-19	910,9	957,2
20-24	1630,1	1709,3	20-24	885,4	922,1
25-29	1650,9	1699,1	25-29	1047,5	1082,2
30-34	1372,8	1395,9	30-34	1286,9	1328,4
35-39	931,1	923,5	35-39	1504,4	1542,7
40-44	1064	1037,3	40-44	1560,3	1589,3
45-49	1104,6	1067,4	45-49	1443,4	1453,5
50-54	1108,8	995,4	50-54	1209,8	1193,9
55-59	986	797	55-59	1123,3	1071,3
60-64	617,9	478,3	60-64	1264,2	1128,8
65-69	748,3	537,8	65-69	1383,9	1131,8
70-74	649,4	440,2	70-74	1201,8	876,7
>75	810,8	410,9	75-79	763,1	490,6



1.3. Struktura genetyczna populacji

Struktura genetyczna populacji jest wynikiem zmian w częstościach alleli między subpopulacjami. W populacji panmiktycznej, w której osobniki swobodnie się krzyżują częstości alleli nie różnią się między grupami, a także między pokoleniami. Ponieważ krzyżowanie na ogół nie jest całkowicie swobodne (np. występowanie góry, rzeki, duża odległość), dochodzi do zmiany częstości alleli między subpopulacjami i tym samym utworzenia struktury genetycznej populacji.

Powstałe różnice w częstościach alleli nie muszą prowadzić do obserwowalnych zmian fenotypowych. Jednakże często różnice w częstościach alleli mogą być przypadkowo korelowane z obserwowanymi zmianami, np. występowaniem choroby. Dlatego struktura genetyczna jest częstą **zmienną zakłócającą (zewnętrzny czynnik wpleciony w badany związek przyczynowo-skutkowy)**.

Struktura populacji jest zjawiskiem złożonym, dlatego jej analiza wymaga kombinacji metod badawczych i modeli statystycznych. Najczęściej strukturę populacji opisuje się za pomocą częstości alleli, liczby alleli efektywnych i częstości genotypów. Ze struktury populacji wynika zróżnicowanie genetyczne, które opisuje zmienność genetyczną wewnątrz populacji oraz między populacjami.

1.3.1. Parametry opisujące strukturę genetyczną populacji

Populacja na ogół jest zbyt duża, aby badać wszystkie osobniki do niej należące, dlatego wybiera się grupę reprezentatywną (próbę). Wybór osobników musi być losowy, a grupa dostatecznie duża. W genetyce populacyjnej przyjmuje się minimalną liczebność próby wynoszącą 20-30 osobników w przypadku analiz morfologicznych i enzymatycznych. Jednakże liczba ta może być mniejsza (np. 10), jeżeli wykorzystujemy dużą liczbę markerów DNA (np. 100).

- Częstość genotypów: obliczamy na podstawie częstości obserwowanych. Dla populacji w równowadze Hardy-Weinberga zachodzi zależność między częstościami genotypów a częstościami alleli:

- ▶ $P = p^2$
- ▶ $H = 2pg$
- ▶ $Q = q^2$;

gdzie P i Q to częstości homozygot, H to częstość heterozygot, p i q to częstości alleli.

Zgodność częstości genotypów z rozkładem Hardy-Weinberga sprawdzamy za pomocą testu χ^2 .

- Częstość alleli

- ▶ Dla alleli kodominujących, częstość alleli można obliczyć bezpośrednio na podstawie liczby alleli danego typu w stosunku do wszystkich alleli występujących w populacji, przy czym należy pamiętać, że homozygoty mają dwa allele danego typu. Częstość alleli kodominujących możemy także obliczyć na podstawie częstości obserwowanych genotypów.

$$p_i = P_{ii} + \frac{1}{2} \sum_{j \neq i} H_{ji}$$

gdzie p_i to częstość i -tego allele w locus i , P_{ii} to częstość homozygot ii , H_{ji} to częstość heterozygot zawierających allel i

- ▶ W przypadku dominacji obliczenie częstości alleli nie jest możliwe na podstawie obserwowanej liczby alleli danego typu nie jest możliwe, gdyż nie można odróżnić homozygot dominujących od heterozygot. Dlatego częstość alleli oblicza się korzystając z prawa Hardy-Weinberga, przy założeniu, że populacja występuje w stanie równowagi.

$$q_i = \sqrt{Q_i}$$

oraz

$$p_i + q_i = 1, \text{ stąd } p = 1 - q;$$

gdzie Q_i to częstość homozygot recesywnych, q_i to częstość allele recesywnego w locus i , p_i to częstość allele dominującego w locus i .

- Średnia liczba alleli w locus, A

$$A = \frac{\sum A_i}{r}$$

gdzie A_i to liczba alleli w i -tym locus, r – liczba loci.

Średnia liczba alleli informuje o różnorodności alleli w populacji, ale nie rozróżnia pomiędzy allelami często występującymi i rzadko występującymi. Przykładowo, dla dwóch loci, w których występują dwa allele, A wyniesie 2. Jeżeli w jednym z loci częstości alleli wyniosą 0,9 i 0,1, a w drugim 0,5 i 0,5, to średnia liczba alleli nie rozróżni tych sytuacji. Tymczasem mimo dwóch alleli w obu loci, zróżnicowanie populacji w locus z częstościami 0,1 i 0,9 będzie znacznie niższe niż dla locus z częstościami 0,5 i 0,5.

- Efektywna liczba alleli w locus, N_e

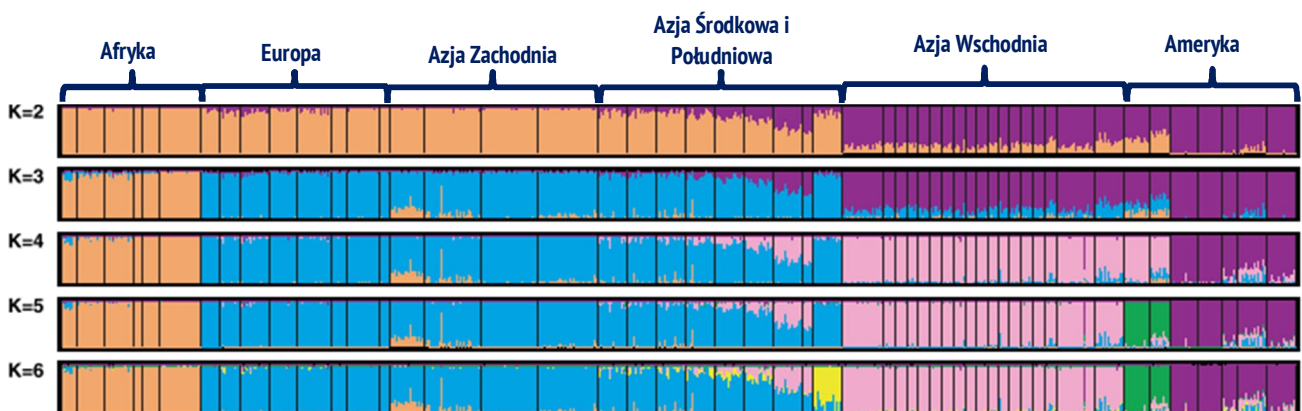
$$N_e = \frac{1}{\sum p_i^2}$$

gdzie p_i to częstość i -tego allele w locus i .

Efektywna liczba alleli uwzględnia częstość, z którą występują dane allele. W przykładzie z punktu powyżej, efektywna liczba alleli dla locus z częstościami 0,9 i 0,1 wyniesie 1,2 podczas gdy dla locus z częstościami 0,5 i 0,5 wyniesie 2. Wartość 1,2 informuje, że jeden z alleli występuje stosunkowo rzadko w populacji.

1.3.2. Struktura genetyczna populacji ludzkich na podstawie loci mikrosatelitarnych

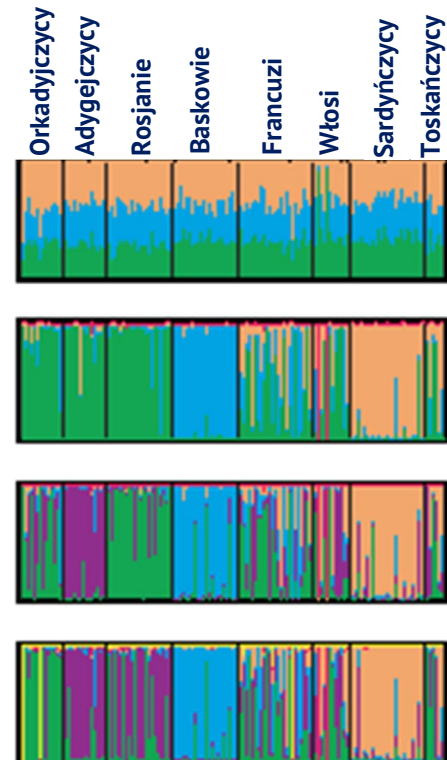
Znajomość struktury genetycznej współczesnych populacji ludzkich dostarcza informacji o procesach ewolucyjnych, ale także wiedza ta jest istotna w rozważaniach epidemiologicznych. Populacja ludzka charakteryzuje się dużym zróżnicowaniem wewnątrzpopulacyjnym przy stosunkowo małych różnicach między populacjami. Analiza 4 199 alleli w loci mikrosatelitarnych wykazała, że 46,7% z nich występuje z dużą częstością we wszystkich badanych regionach Afryki, Europy, Środkowego Wschodu, Południowej Azji, Wschodniej Azji, Oceanii i Ameryki (Rosenberg i inni 2002). Jedynie 7,4% alleli występowało tylko w jednym regionie. Były to w większości allele rzadkie, występujące z częstością jedynie nieznacznie wyższą od 1%.



Rys. 1.3.2a. Struktura genetyczna populacji ludzkiej. Czarne linie oddzielają poszczególne populacje. Allele pokazane są różnymi kolorami. Populacje były grupowane metodą K-clustering. Średnie podobieństwo genetyczne N_{ei} między populacjami wyniosło 0,97 (Rosenberg i inni 2002).

Wykorzystanie alleli rzadkich wraz z różnicami w częstości alleli powszechnych umożliwiło grupowanie populacji światowej na podstawie struktury. Zgodnie z przewidywaniami, populacje sąsiadujące miały podobną strukturę. Obserwowano także gradację częstości alleli wraz z odległością między regionami (Rys. 1.3.2a).

Europa charakteryzuje się najniższą zmiennością pomiędzy poszczególnymi grupami. Grupowa metoda K-clustering w zależności zastosowanych parametrów wykazało b struktury genetycznej różnicującej poszczególne populacje ($K = 3$) lub inną strukturę dla Basków, Sardyńczyków ($K = 4$, $K = 5$, $K = 6$). Grupowanie Rosjan zależało od zastosowanych parametrów. Rosjanie grupowali się z Adygejczykami (Republika Adygei w Rosji) lub Orkadyjczykami (Orkady – archipelag w pobliżu Szkocji, ludność używa języka orkadyjskiego, będącego odmianą szkockiego). Grupowanie Rosjan z Orkadyjczykami może być związane z wspólnym pochodzeniem od Wikingów. Należy jednak podkreślić, że pochodzenie od Wikingów również może dotyczyć Francuzów, Włochów i Toskańczyków. Ponadto wiele grup etnicznych nie było reprezentowane w badaniach (Słowianie). Wreszcie analizowane grupy Europy były małe, co sprawia, że na częstości alleli mogły mieć wpływ procesy losowe, stąd trudności w ustaleniu struktury populacji (Rys. 1.3.2b).



Rys. 1.3.2b. Struktura genetyczna populacji europejskiej. Czarne linie oddzielają poszczególne populacje. Allele pokazane są różnymi kolorami. Populacje były grupowane (metodą K-clustering. Średnie podobieństwo genetyczne N_e między populacjami wyniosło 0,97 dla $K = 3$, i $K = 5$; 0,93 dla $K = 4$, i 0,86 dla $K = 6$ (Rosenberg i inni 2002).

1.3.3. Grupy krwi MN u Indian amerykańskich



W grupie Indian amerykańskich stwierdzono 10 osób z grupą krwi M, 53 osoby z grupą MN oraz 147 osób z grupą N.

- Proszę podać częstość genotypów w locus *MN* w grupie Indian amerykańskich.
- Proszę podać częstość alleli *M* i *N*.
- Proszę podać efektywną liczbę alleli w locus.
- Czy obserwowane częstości genotypów są zgodne z rozkładem Hardy-Weinberga?



1.3.4. Albinizm u Indian Hoopi

W grupie Indian amerykańskich Hoopi zidentyfikowano 26 osób albinotycznych. Cała populacja Hoopi liczy 6000 osób. Zakładając, że populacja jest w równowadze Hardy-Weinberga proszę podać ile nosicieli recesywnego allele warunkującego albinizm występuje w populacji Hoopi. W locus A występują trzy allele, A_1 , A_2 , A_3 z częstością odpowiednio 0,6, 0,3 0,1. Przy założeniu, że krzyżowanie jest losowe proszę podać łączną częstość wszystkich heterozygot w populacji.

1.3.5. Obliczenie częstości heterozygot

W locus A występują trzy allele, A_1 , A_2 , A_3 z częstością odpowiednio 0,6, 0,3 0,1. Przy założeniu, że krzyżowanie jest losowe proszę podać łączną częstość wszystkich heterozygot w populacji.

1.3.6. Ocena ryzyka genetycznego mukowiscydozy

W rodzinie, gdzie zarówno matka jak i ojciec byli zdrowi urodziło się dwoje chłopców, jeden zdrowy, a drugi z mukowiscydozą. Jakie jest ryzyko genetyczne wystąpienia mukowiscydozy w potomstwie chłopca zdrowego ze zdrową kobietą bez historii choroby w rodzinie, jeżeli częstość mukowiscydozy w populacji wynosi 1/500?

2. Zróżnicowanie genetyczne populacji



2.1. Miary zróżnicowania genetycznego

Zróżnicowanie genetyczne określa zmienność w obrębie populacji. Jest to podstawowy parametr w genetyce populacyjnej. Zróżnicowanie występuje na różnym poziomie organizacji, począwszy od grup krewniczych przez populacje do zróżnicowania pomiędzy gatunkami. Poziom zróżnicowania zależy od cyklu życiowego danego organizmu i sposobu rozmnażania. Zróżnicowanie fenotypowe nie zawsze odpowiada zróżnicowaniu genetycznemu. Przykładowo, gatunki bliźniacze mają podobne fenotypy, a różnią się genetycznie. Pomiedzy nimi występuje bariera reprodukcyjna.

Najprostszą miarą zróżnicowania genetycznego jest polimorfizm. Polimorfizm oblicza się jako procent loci, w których występują różne allele w populacji. Przykładowo, jeżeli w pięciu analizowanych genach, dla trzech stwierdzamy wystąpienie co najmniej dwóch alleli, to poziom polimorfizmu wyniesie $3/5$ (60%). Polimorfizm informuje tylko o wystąpieniu różnych alleli, ale nie mówi nic o ich częstości. Tymczasem na zróżnicowanie genetyczne osobników w populacji wpływa zarówno obecność różnych alleli w locus jak i ich częstość. Allel rzadki o częstości 0,06 tylko nieznacznie wpłynie na częstość heterozygot. Większość osobników w populacji będzie homozygotami względem częstszego allele.

Heterozygotyczność (gene diversity) ocenia nie tylko występowanie różnych alleli w locus, ale także ich częstość. Dzięki temu stanowi bardziej miarodajną ocenę zróżnicowania genetycznego populacji niż polimorfizm. Heterozygotyczność obliczamy dla danego locus oraz jako średnią dla wszystkich loci w populacji. Heterozygotyczność możemy podawać dla populacji, gatunku, populacji podzielonych na subpopulacje. Jeżeli populacja (gatunek) jest podzielona na subpopulacje to można ocenić poziom zmienności wewnątrzpopulacyjnej i międzypopulacyjnej.

- Heterozygotyczność w locus jest najczęściej rozumiana jako prawdopodobieństwo, że losowo wybrana para alleli jest różna. Wówczas oblicza się ją jako:

$$h = 1 - \sum p_i^2 \text{ (Nei, 1973)}$$

- W przypadku, gdy osobniki są powiązane (np. rozmnażanie klonalne) można zastosować nieobciążony wskaźnik Nei'a i Roychoudhry, 1974

$$h = \frac{2n(1 - \sum_{i=1}^m p_i^2)}{2n - 1}, \text{ gdzie } n \text{ to liczba osobników}$$

- Średnia heterozygotyczność w populacji dla wszystkich loci

$$H = \frac{\sum_{i=1}^r h_i}{r}, \text{ gdzie } r \text{ to liczba loci}$$

- Średnia heterozygotyczność dla subpopulacji

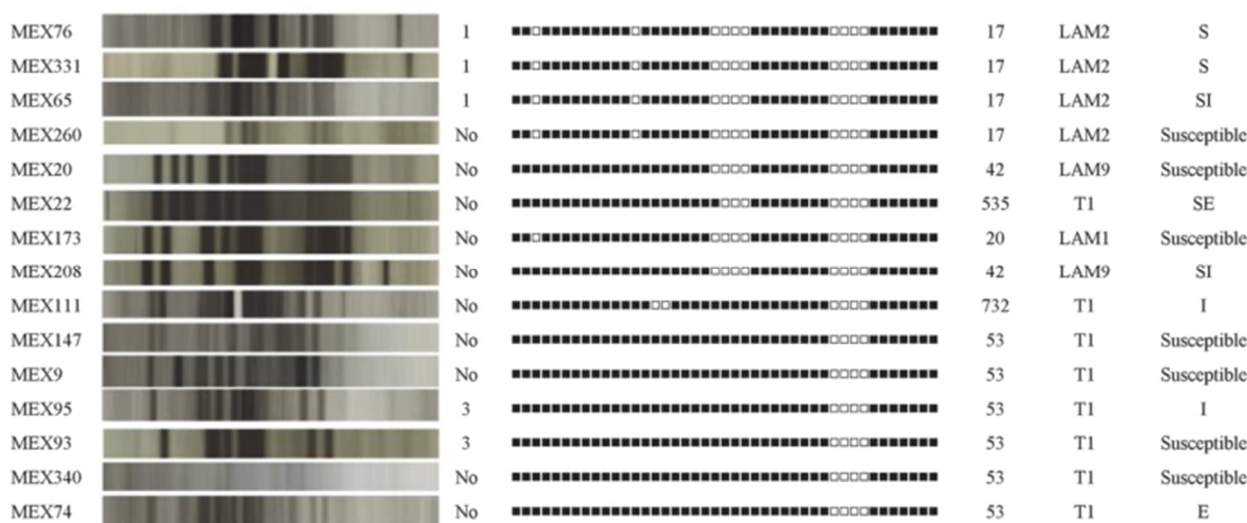
$$H_s = \frac{\sum H_k}{s}, \text{ gdzie } s \text{ to liczba subpopulacji}$$



2.2. Zróżnicowanie genetyczne *Mycobacterium tuberculosis*

Mycobacterium tuberculosis jest wolno rosnącą bakterią rozmnażającą się klonalnie, stąd uważano, że charakteryzuje się ona niskim poziomem zmienności genetycznej. Jednakże analiza genomu prątka gruźlicy przy pomocy markerów DNA wykazała polimorfizm we wszystkich analizowanych loci oraz heterozygotyczność wynoszącą 0,32 dla gatunku. Podobnie wysokie wartości obserwuje się u organizmów diploidalnych. Z drugiej strony, wartość H_T dla *M. tuberculosis* jest blisko trzykrotnie niższa niż dla *E. coli* (0,85).

Ocena zróżnicowania genetycznego *M. tuberculosis* pozwala na identyfikację szczepów dominujących w różnych regionach, a także jest pomocna w rozróżnianiu szczepów opornych na



Rys. 2.2. Zróżnicowanie *M. tuberculosis* na podstawie miejsc restrykcyjnych w sekwencji *IS6110*. I: oporność na izoniazyd, S, oporność na streptomycynę, E: oporność na etambutol. Każdy szczep ma unikalny wzór miejsc restrykcyjnych.

leki przeciwpłatkowe. Do oceny zróżnicowania genetycznego najczęściej wykorzystuje się sekwencję insercyjną *IS4110* oraz geny *KatG* i *rpoB* warunkujące oporność na izoniazyd i ryfampicynę. W przypadku sekwencji *IS6110* analizuje się miejsca restrykcyjne oraz liczbę kopii tej sekwencji. Obecność identycznych genotypów we wszystkich analizowanych fragmentach u różnych pacjentów wskazuje na jedno źródło pochodzenia infekcji.

2.3. Obliczanie zróżnicowania genetycznego dla markerów kodominujących



Analiza trzech loci kodominujących u 30 osobników wykazała występowanie następujących genotypów:

- Locus A: A1A1 – 4 osobniki, A1A2 – 6 osobników, A2A2 – 30 osobników.
- Locus B: B1B1 – 10 osobników, B1B2 – 5 osobników, B2B2 – 25 osobników.
- Locus C: C1C1: 10 osobników, C1C2 – 18 osobników, C2C2 – 12 osobników.

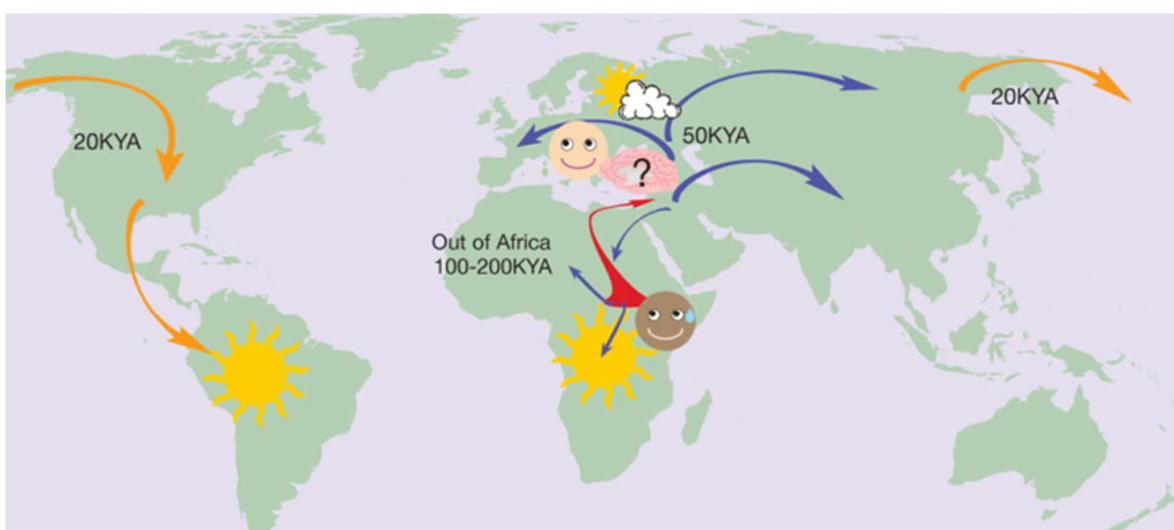
- A. Proszę podać heterozygotyczność dla każdego locus według wzoru Nei'a 1973.
B. Proszę podać średnią heterozygotyczność w populacji dla wszystkich loci, H.



2.4. Zróżnicowanie genetyczne populacji ludzkich

Złożony wzór zróżnicowania genetycznego populacji ludzkiej jest efektem wielu wydarzeń demograficznych i ewolucyjnych. Wczesne zdarzenia, gdy populacja ludzka była mała z pewnością miały zdecydowanie większy wpływ na zróżnicowanie genetyczne niż wówczas, gdy populacja osiągnęła większe rozmiary. Większość danych genetycznych umożliwia śledzenie losów populacji ludzkiej przez ostatnie 200 tys. lat.

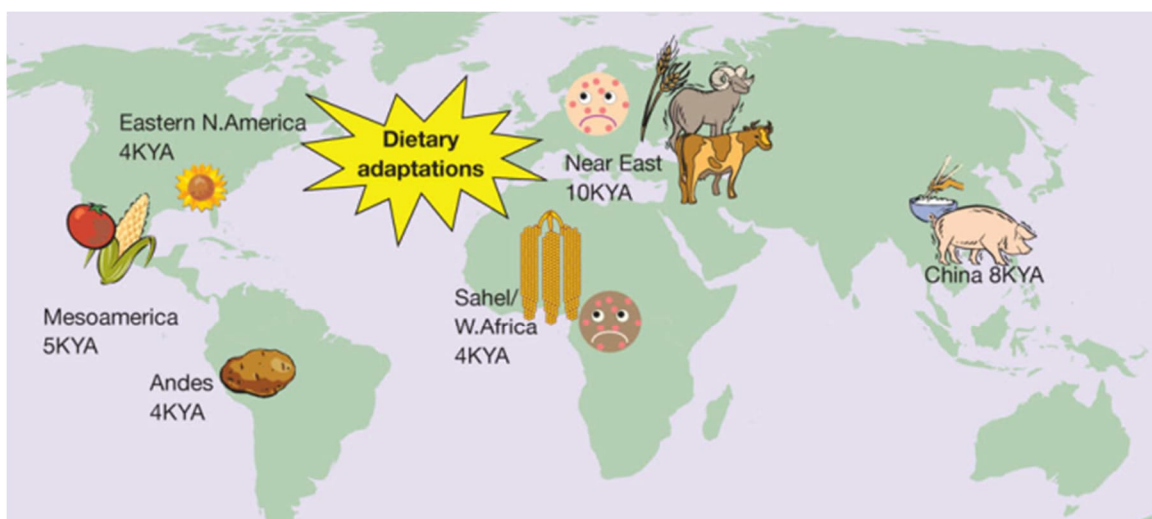
- Wczesna migracja z Afryki była związana z opanowaniem nowych środowisk. Jednocześnie dochodziło do zastępowania gatunków archaicznych. W efekcie największe zróżnicowanie genetyczne populacji jest w Afryce, gdzie można zidentyfikować wiele lokalnych linii. Odległość geograficzna od obszarów Afryki Wschodniej koreluje ze spadkiem zróżnicowania geograficznego populacji.
- Migracja z Afryki wystawiła populacje ludzkie na działanie chłodnego klimatu z mniejszą ilością promieniowania słonecznego. W takich warunkach jest promowana depigmentacja, gdyż ułatwia ona penetrację promieniowania UV niezbędnego do produkcji witaminy D. Za



Rys. 2.4a. Opanowanie nowych środowisk miało przede wszystkim konsekwencje w adaptacjach morfologicznych jak np. pigmentacja skóry.

pigmentację skóry odpowiada około 20 genów. Ślady selekcji występują u populacjach europejskich i wschodnio-azjatyckich. Dotyczą one wielu genów, co potwierdza selekcyjny charakter rozjaśnienia skóry. Jednocześnie, w Afryce tropikalnej doszło do selekcji genotypów ułatwiających zatrzymanie soli. Wynikało to z dużych strat soli przy jednoczesnej małej ilości soli w pożywieniu. Migracja tych genotypów na tereny chłodniejsze może prowadzić do podwyższonego ciśnienia krwi.

- Rewolucja neolityczna, adaptacja do różnej diety. Zróżnicowanie środowisk opanowanych przez populacje zbieracko-łowieckie odzwierciedlają zróżnicowane diety. Najistotniejszym elementem było wprowadzenie rolnictwa około 12-10 tys. lat temu na Bliskim Wschodzie. Wzrost zagęszczenia populacji korelował ze spadkiem zróżnicowania dietetycznego, osiadłym stylem życia i zwiększoną ekspozycją na patogeny. Powszechnie uważa się, że rewolucja neolityczna doprowadziła do selekcji „genotypów oszczędnych” umożliwiających przeżycie w okresie głodu. Naturalną konsekwencją takiej selekcji jest występowanie obecnie cukrzycy typu 2 jako efektu ciągłego dostępu do wysokoenergetycznych pokarmów. Jednakże zależność ta nie jest prosta. Czynnikiem transkrypcyjnym *TCF7L2* jest odpowiedzialny za 17-28% przypadków cukrzycy typu 2 w populacjach europejskich, ale jednocześnie odpowiada za obniżenie masy ciała. Pojawienie się wariantu *TCF7L2* jest skorelowane z rozwojem rolnictwa, co wskazuje na znaczenie adaptacyjne, ale istota tej adaptacji jest trudna do określenia. Jedną z najlepiej poznanych adaptacji jest mutacja w czynniku transkrypcyjnym dla genu laktazy umożliwiająca trawienie mleka krowiego dorosłym osobnikom.



Rys. 2.4.b. Adaptacja do różnorodnej diety w okresie rewolucji neolitycznej. Jest to także okres zmniejszającej się różnorodności gatunkowej diety oraz zwiększonego narażenia na patogeny ze względu na wzrost zagęszczenia populacji ludzkich.

3. Podobieństwo i odległość genetyczna w populacjach ludzkich

Podobieństwo i odległość genetyczna przedstawiają zależności między populacjami na podstawie podobieństw/różnic między genomami. Odległość genetyczną na ogół oblicza się jako ujemny logarytm naturalny z podobieństwa genetycznego. Wartość podobieństwa najczęściej określa się na podstawie częstości alleli. Oczywiście allele różnią się nukleotydami, a więc różnice w częstości alleli to także różnice w sekwencjach nukleotydowych.



3.1. Podobieństwo i odległość genetyczna na podstawie częstości alleli

Podobieństwo genetyczne oznacza w jakim stopniu genomy są podobne. Wartość podobieństwa (odległości) może być wykorzystana do oceny czasu dywergencji między populacjami. Kilka metod jest wykorzystywanych do oceny podobieństwa na podstawie częstości alleli, jednakże najpowszechniej wykorzystuje się odległość genetyczną Nei'a, która jest proporcjonalna do czasu dywergencji.

- Odległość genetyczna Nei'a

$$D = -\ln I, \text{ gdzie } I \text{ to podobieństwo genetyczne}$$

$$I = \frac{\sum p_x p_y}{\sqrt{\sum p_x^2 p_y^2}}$$

- Czas dywergencji

$$D = 2 \alpha t, \text{ gdzie}$$

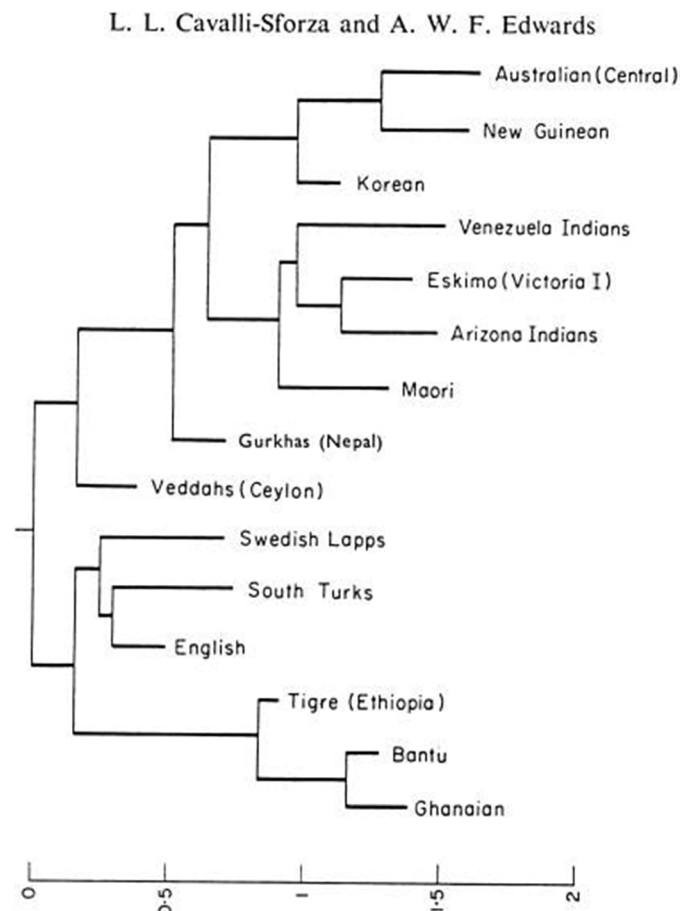
α - tempo substytucji

t - czas

- Jeżeli znamy tempo substytucji oraz odległość genetyczną to możemy obliczyć czas dywergencji:

$$t = \frac{D}{2 \alpha}$$

- ▶ Wartość α jest różna dla poszczególnych loci. Dla loci enzymatycznych u człowieka wartość ta wynosi 10^{-7} na locus/na rok. Dla loci mikrosatelitarnych wartość ta wynosi 5×10^{-4} na locus na pokolenie.



Rys. 3. Pierwsze drzewo filogenetyczne człowieka oparte o częstości alleli.



3.1.1. Poniżej przedstawiono odległość genetyczną między populacjami ludzkimi na podstawie loci enzymatycznych.

	Włochy	Japonia	Chiny	Ghana	Bantu
Włochy	0	0,0319	0,0427	0,0418	0,0393
Japonia		0	0,0047	0,1023	0,0992
Chiny			0	0,1251	0,1228
Ghana				0	0,0048
Bantu					0

A. Korzystając ze strony <http://genomes.urv.cat/UPGMA/> proszę narysować dendrogram grupujący te populacje na podstawie odległości genetycznej.

*Uwaga: należy wartości wprowadzić do notatnika, oddzielając klawiszem TAB, w wartościach zamiast przecinka należy użyć kropki.

B. Proszę obliczyć czas dywergencji dla wszystkich wymienionych populacji.

➔ 4. Analiza zmienności genetycznej w populacji ludzkiej

Retrotranspozony stanowią wygodny marker do analiz populacyjnych ze względu na ich wysoki polimorfizm i równomierne rozmieszczenie w genomie. Remobilizacja elementów ruchomych jest często odpowiedzią na warunki stresowe. Analiza loci, w których doszło do insercji może dostarczyć informacji o ewolucji populacji, a także jej różnicowanie się w czasie teraźniejszym.

U człowieka obecnie obserwuje się niewielką aktywność transpozonów *LINE-1* oraz *Alu*. Oba transpozony były także aktywne w trakcie ewolucji człowieka i mogą być wygodnym narzędziem do analizy struktury i zróżnicowania genetycznego populacji ludzkich.



Tabela poniżej przedstawia dane dotyczące analizy insercji *LINE* w locus *L1* oraz insercji *Alu* w locus *Alu1b* w czterech populacjach.

Populacja	Liczebność	Liczba osób z insercją LINE w locus L1	Liczba osób z insercją Alu w locus Alu1b
Afryka	130	54	60
Azja Wschodnia	59	28	30
Europa	107	68	73
Indie	149	70	82

A. Na podstawie tabeli 4 proszę obliczyć częstości alleli w obu loci (*L1* i *Alu1b*) w badanych populacjach, efektywną liczbę alleli w locus (N_e), wiedząc, że obecność insercji jest dominująca w stosunku do jej braku oraz populacje znajdują się w stanie równowagi Hardy-Weinberga.

- B. Dla badanych populacji proszę podać heterozygotyczność w locus dla analizowanych loci (h), średnią heterozygotyczność w populacji dla wszystkich loci (H) oraz wartość H_s dla wszystkich populacji analizowanych łącznie.
- C. Wyniki dla efektywnej liczby alleli w locus (N_e), heterozygotyczności w locus (h) oraz średniej heterozygotyczności w populacji (H) proszę przedstawić na wykresach.
- D. Proszę porównać obliczone parametry struktury genetycznej i zróżnicowania genetycznego dla analizowanych loci. Proszę wyjaśnić różnice wykorzystując wiedzę dotyczącą genomu człowieka.
- E. Na podstawie wyników z punktu C proszę porównać strukturę i zróżnicowanie genetyczne badanych populacji. Proszę wyjaśnić różnice między populacjami korzystając z informacji o ewolucji człowieka.

Samodzielne wykonanie zadania 4: 5 punktów.

Termin: 15.12.2023., 23:59

**Uwaga; do obliczeń najlepiej wykorzystać arkusz Excel, a wyniki przedstawić w załączonej tabeli. Wartości w tabeli proszę zaokrąglić do wartości tysięcznych. Ponadto zakładamy, że obserwowane różnice są istotne statystycznie.*

Parametry struktury genetycznej i zróżnicowania genetycznego populacji ludzkich na podstawie insercji LINE i Alu.						
Populacja	Locus	Allel	Częstość alleli	N_e	h	H
Afryka	L1	p				
		q				
	Alu1b	p				
		q				
Azja Wschodnia	L1	p				
		q				
	Alu1b	p				
		q				
Europa	L1	p				
		q				
	Alu1b	p				
		q				
Indie	L1	p				
		q				
	Alu1b	p				
		q				
					H_s	