

Ćwiczenie 06

Genetyka populacyjna

Struktura genetyczna populacji. Zróżnicowanie genetyczne. Podobieństwo genetyczne.

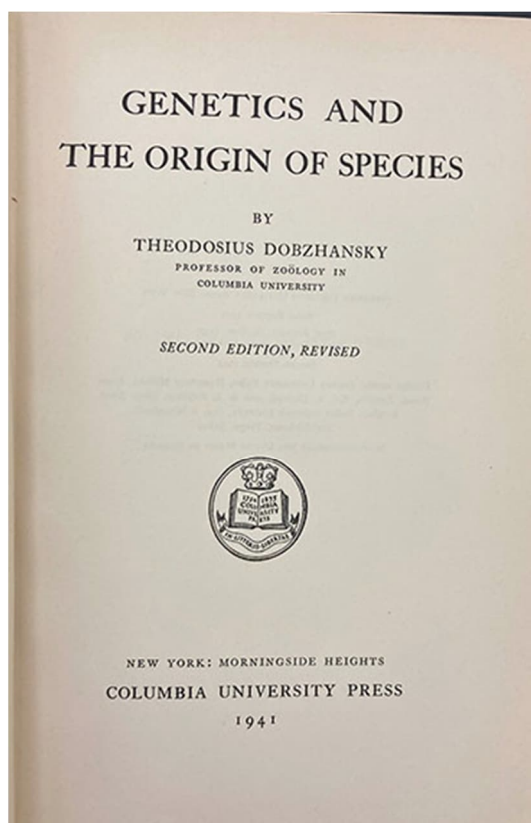
Kornelia Polok

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 procesów 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



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

kreatywności i interdyscyplinarnym umiejętnościom. 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.

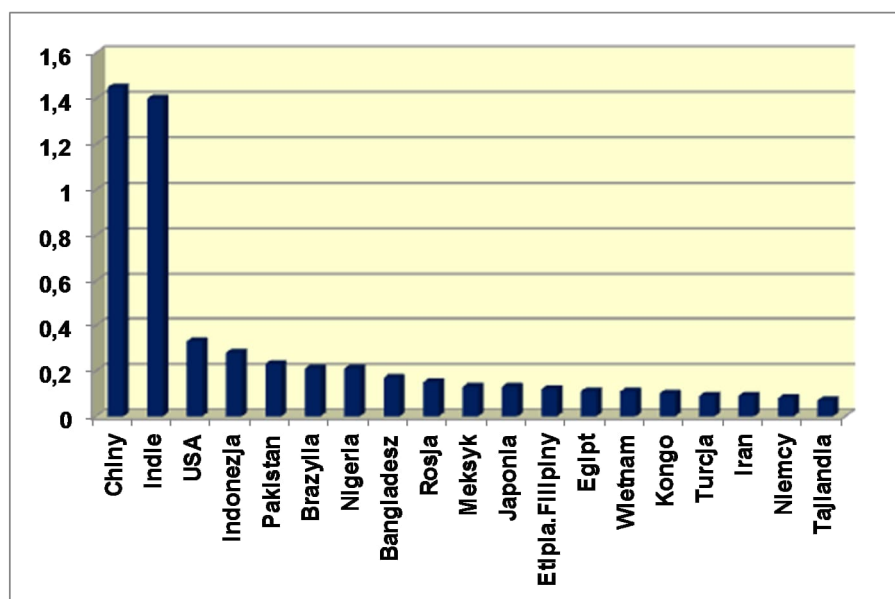
1.2. Populacja biologiczna i jej struktura

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.

1.2.1. Wielkość populacji ludzkiej

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

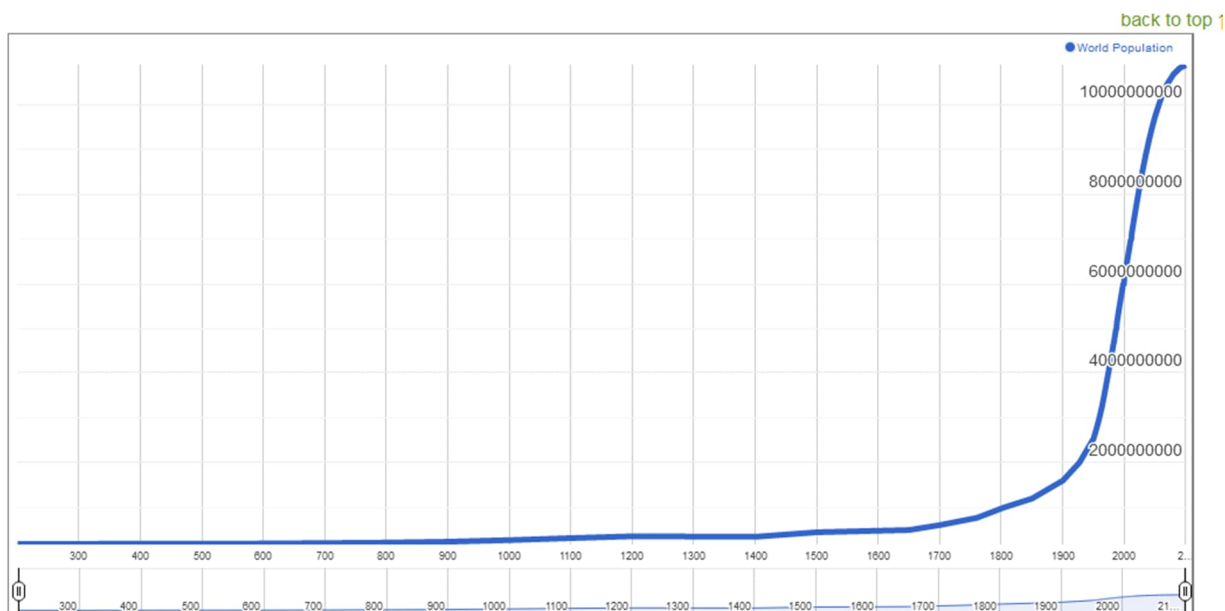
na ogół opisuje tylko populacje ludzkie. Ekologia opisuje populacje za pomocą struktury wiekowej, przestrzennej, tempa wzrostu, zagęszczenia. Natomiast genetyka populacyjna opisuje populację za pomocą parametrów informujących o składzie genetycznym i jego zmianie na przestrzeni czasu.



Rys. 1.2a.. Wielkość populacji najbardziej liczebnych państw.

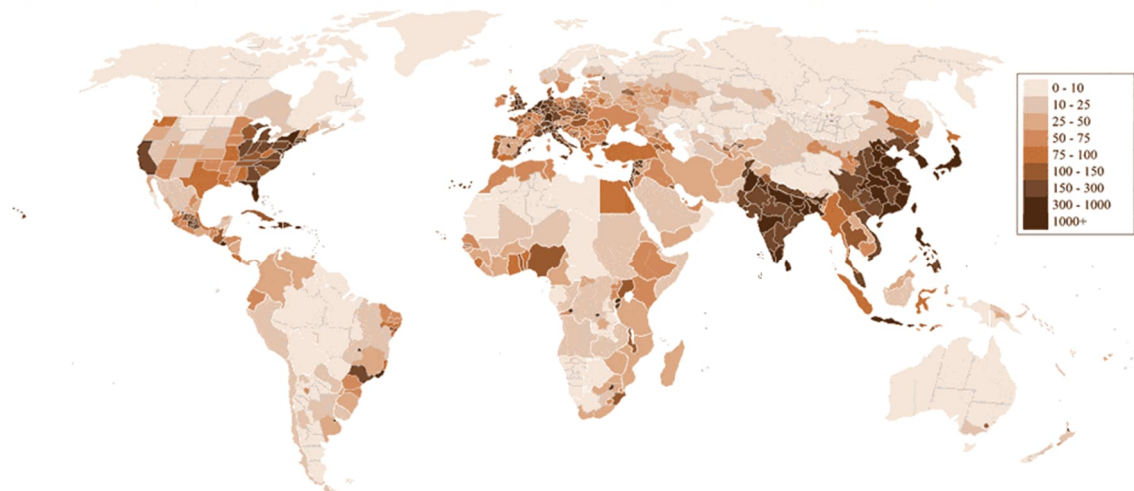
Czynniki takie jak zagęszczenie, struktura wiekowa mają wpływ na parametry genetyczne i w tym zakresie są uwzględniane w genetyce populacyjnej.

Obecnie populacja ludzka na świecie to 7,9 mld (stan na 2021.11.25. Worldometers). Szacuje się, że na Ziemi żyło około 106 mld. ludzi. Największą populację mają Chiny i Indie, po 1,4 mld. Gwałtowny wzrost liczebności populacji ludzkiej obserwuje się od 1900 r. Wartość 1 mld populacja ludzka osiągnęła w 1804 r.



Rys. 1.2b. Tempo wzrostu populacji ludzkiej od średniowiecza.

Population density map of the world showing not only countries but also many subdivisions (regions, states, provinces). See also: [World Map](#)



Rys. 1.2c. Gęstość populacji ludzkiej.

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, do 2 mld w 1930 r. 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. 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.

Na podstawie poniższej tabeli proszę wykreślić wykres struktury wiekowej dla



- A. Ludności Polski w 1980 r. Jaki typ struktury wiekowej wówczas występował w Polsce? (2 punkty)
- B. Ludności Polski w 2019 r. Jaki typ struktury wiekowej występował w 2019 r. w Polsce? (2 punkty)
- C. Proszę porównać strukturę wiekową ludności Polski w 1980 r. oraz w 2019 r. (2 punkty)

1980 w tys.			2019 w tys.		
Klasa	Mężczyźni	Kobiety	Klasa	Mężczyźni	Kobiety
0-4	1664,2	1588,5	0-4	985,5	932,9
5-9	1465,1	1399,4	5-9	1002,5	947,7
10-14	1289,6	1232,2	10-14	1034,8	984,7
15-19	1424,3	1347,9	15-19	923,9	878,0
20-24	1709,3	1630,1	20-24	1036,8	992,7
25-29	1699,1	1650,9	25-29	1278,5	1233,8
30-34	1395,9	1372,8	30-34	1478,5	1436,2
35-39	923,5	931,1	35-39	1633,7	1596,8
40-44	1037,3	1064	40-44	1531,0	1502,2
45-49	1067,4	1104,6	45-49	1304,6	1294,6
50-54	995,4	1108,8	50-54	1125,0	1142,2
55-59	797	986	55-59	1168,0	1234,5
60-64	478,3	617,9	60-64	1295,8	1460,9
65-69	537,8	748,3	65-69	1102,6	1359,8
70-74	440,2	649,4	70-74	741,3	1013,4
>75	410,9	810,8	>75	924,7	1805,1

1.3. Struktura genetyczna populacji



Do parametrów określających strukturę genetyczną należy:

- częstość alleli obliczamy na podstawie częstości obserwowanych

- ▶ dla alleli kodominujących:

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

- ▶ dla alleli dominujących, przy założeniu równowagi Hardy-Weinberga

$$q_i = \sqrt{Q_i}$$

- średnia liczba alleli w locus,

$$A = \frac{\sum A_i}{r}, \text{ gdzie } r \text{ to liczba loci}$$

- efektywna liczba alleli w locus, n_e ,

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

Efektywna liczba alleli uwzględnia częstość z jaką występują dane allele. Przykładowo, jeżeli w locus są dwa allele, każdy z częstością 0,5 to liczba alleli w locus będzie równa liczbie alleli zidentyfikowanych i również wyniesie 2. Jednakże jeżeli jeden z alleli będzie występował częściej niż drugi to efektywna liczba alleli będzie niższa od 2. Dla alleli o częstości 0,9 i 0,1 n_e wyniesie 1,2. Wartość ta informuje, że jeden z alleli występuje stosunkowo rzadko w populacji.

- Częstość genotypów: obliczamy na podstawie częstości obserwowanych. Dla populacji w równowadze Hardy-Weinberga zachodzi częstości genotypów wynoszą:
 - ▶ $P = p^2$
 - ▶ $H = 2pq$
 - ▶ $Q = q^2$

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

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).

1.3.1. W grupie Indian amerykańskich stwierdzono 78 osób z grupą krwi M. 29 osób z grupą MN oraz 4 osoby z grupą N. Proszę podać:



- Częstość genotypów. (1 punkt)
- Częstość alleli M i N (1 punkt)
- Efektywną liczbę alleli w locus (1 punkt)
- Czy obserwowane częstości genotypów są zgodne z rozkładem Hardy-Weinberga? (3 punkty)

1.3.2. 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. (2 punkty)

1.3.3. W locus A występują trzy allele, A1, A2, A3 z częstością odpowiednio 0,6, 0,3, 0,1. Przy założeniu, że krzyżowanie jest losowe proszę podać łączną częstość wszystkich heterozygot w populacji. (1 punkt)

1.3.4. 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? (3 punkty)

2. Zróżnicowanie genetyczne

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 krewniaczych 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. Pomiędzy 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

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

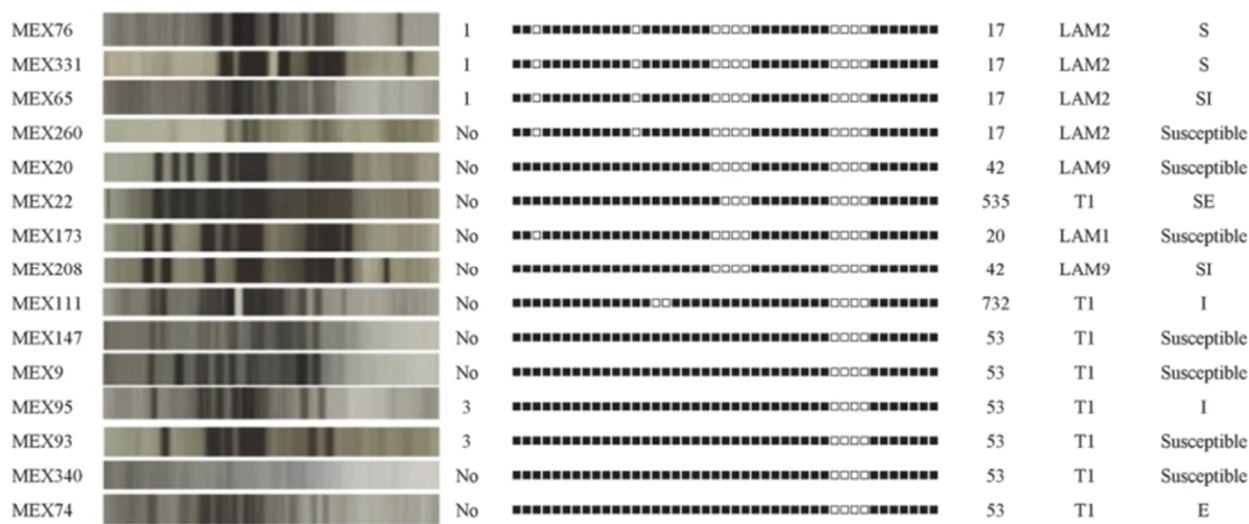
- Średnia heterozygotyczność dla subpopulacji

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

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 wynosząca 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 leki przeciwprątkowe. 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.



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

2.3. Zróżnicowanie genetyczne dla markerów kodominujących

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

- Locus A: A1A1 – 2 osobniki, A1A2 – 4 osobniki, A2A2 – 24 osobniki.
- Locus B: B1B1 – 7 osobników, B1B2 – 3 osobniki, B2B2 – 20 osobników.
- Locus C: C1C1: 10 osobników, C1C2 – 12 osobników, C2C2 – 8 osobników.

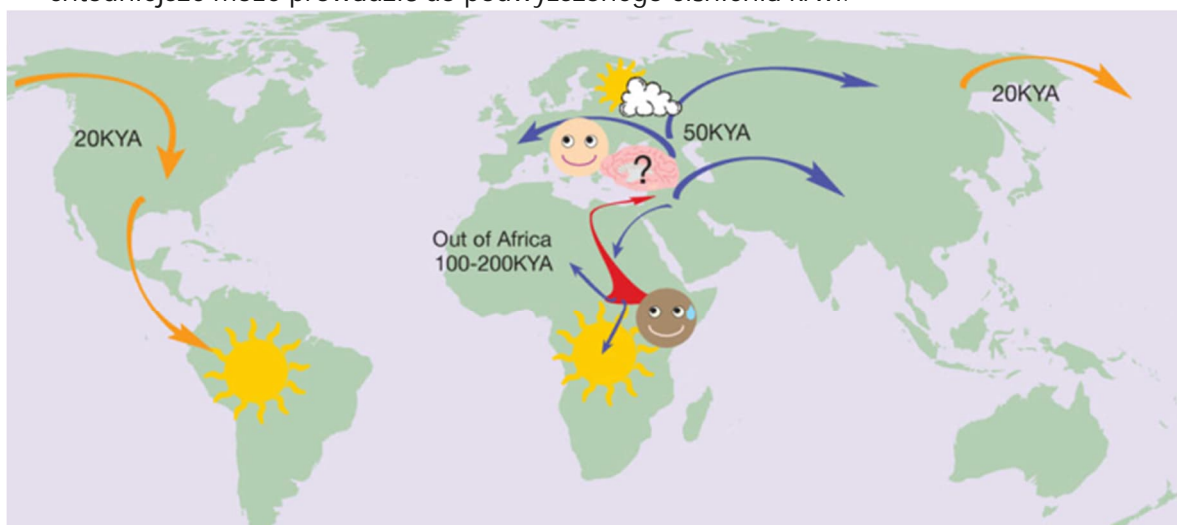
A. Proszę podać heterozygotyczność dla każdego locus według wzoru Nei'a 1973. (2 punkty/locus)

B. Średnią heterozygotyczność w populacji dla wszystkich loci, H . (2 punkty)

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.

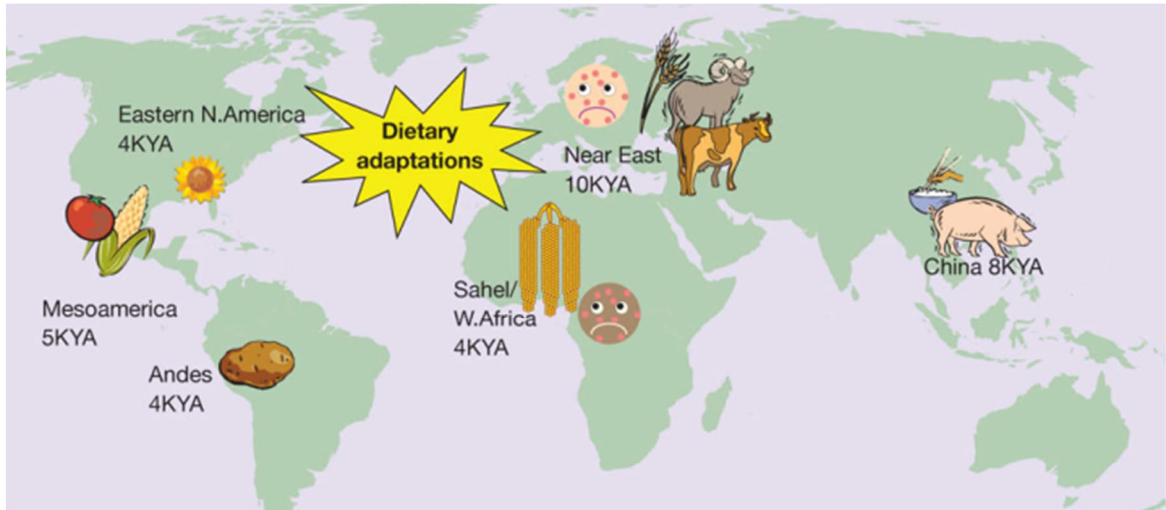
- Wczesna migracja i opanowanie nowych środowisk przez populacje wywodzące się z Afryki. Jednocześnie dochodziło do zastępowania gatunków archaicznych, w większości przypadków bez krzyżowania. W efekcie największe zróżnicowanie genetyczne populacji jest w Afryce, gdzie można zidentyfikować wiele lokalnych 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 pigmentację skóry odpowiada około 20 genów. Ślady selekcji występują u populacji europejskich i wschodnio-azjatyckich. Dotyczą one wielu genów, co potwierdza selekcyjny charakter rozjaśnienia skóry. Jednocześnie, w Afryce tropikalnej 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.



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

- 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. Czynniki

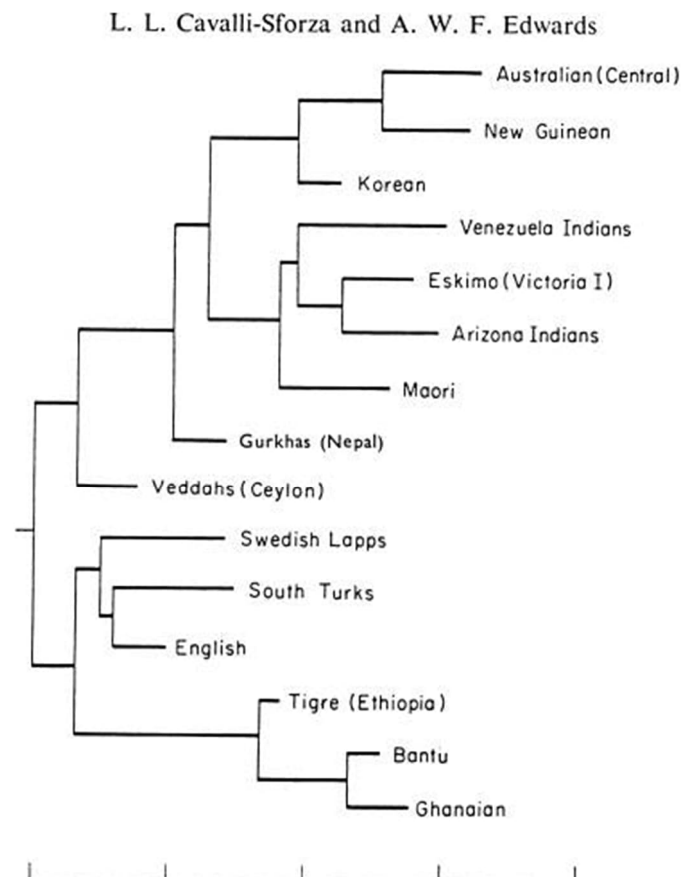
transkrypcyjny *TCF7L2* jest odpowiedzialny za 17-28% przypadków cukrzycy typu 2 w populacjach europejskich, ale jednocześnie odpowiada za obniżenie masy ciała. Pojawienie się wariantu *TCF7L2* jest skorelowane z rozwojem rolnictwa, co wskazuje na znaczenie adaptacyjne, ale istota tej adaptacji jest trudna do określenia. Jedną z najlepiej poznanych adaptacji jest mutacja w czynniku transkrypcyjnym dla genu laktazy umożliwiającą 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

➔ 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.



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

3.1. Podobieństwo i odległość genetyczna dla 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 ona 10^{-7} na locus/na rok. Dla loci mikrosatelitarnych wartość ta wynosi 5×10^{-4} na locus na pokolenie.



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	0,418	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 (2 punkty)

*Uwaga: należy wartości wprowadzić do notatnika, oddzielając klawiszem TAB, należy pominąć przecinek, tzn. wartość 0,0427 należy wprowadzić jako 0427

B. Proszę obliczyć czas dywergencji dla wszystkich wymienionych populacji. (1 punkt/porównanie)