

Techniki biochemiczne i molekularne w ekologii



Część slajdów jest autorstwa
prof. Mirosława Ratkiewicza
z Uniwersytetu w Białymstoku
któremu składam
SERDECZNE PODZIĘKOWANIA
za pozwolenie na wykorzystanie jego slajdów

Wykład 15 godzin:

5 x 3 godziny lekcyjne

zaliczenie: test pisemny (wyboru) 60% końcowej oceny

Ćwiczenia 15 godzin, dwie grupy:

5 x 3 godziny lekcyjne

zaliczenie: kolokwium pisemne 40% końcowej oceny

3 x laboratorium molekularne

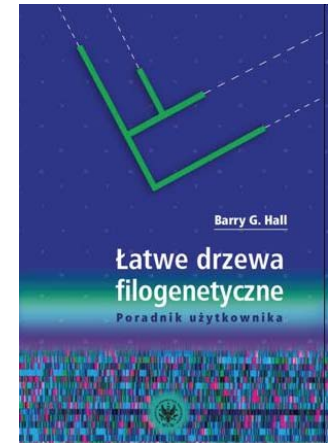
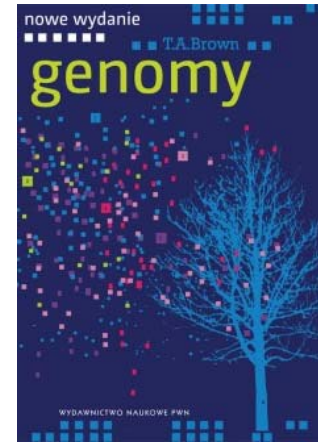
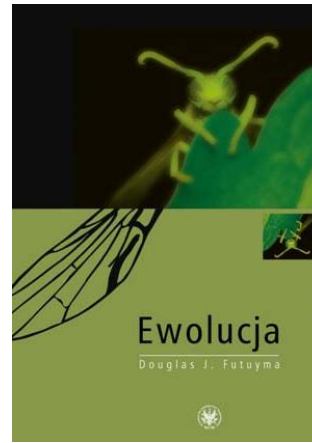
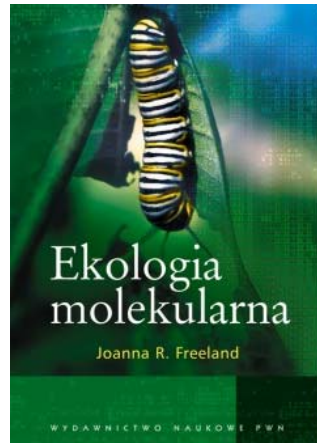
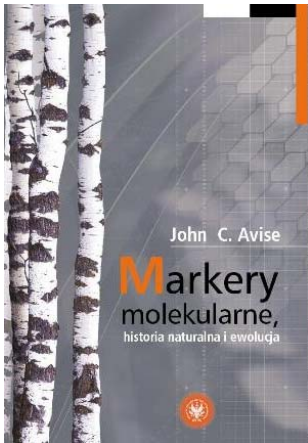
2 x laboratorium komputerowe

www: <http://www.eko.uj.edu.pl/molecol/>

-> Dla studentów -> Biochemiczne i molekularne metody...

Podręczniki

- Avise: Markery molekularne, historia naturalna i ewolucja. WUW 2008
- Freeland: Ekologia molekularna. PWN 2008
- Futuyma: Ewolucja. WUW 2008
- Brown: Genomy. PWN 2009
- Hall: Łatwe drzewa filogenetyczne. WUW 2008



Wykłady

- Ekologia molekularna: integracja ekologii, biologii ewolucyjnej i biologii molekularnej
- Czym są markery molekularne?
- Dlaczego markery molekularne są ważne w badaniach ekologicznych i ewolucyjnych?
- Najważniejsze techniki molekularne
- Zastosowanie technik molekularnych do rozwiązywania problemów badawczych/**uzyskiwania informacji niedostępnych w inny sposób**

Ekologia molekularna

- zastosowanie metod **genetyki molekularnej** do rozwiązywania problemów badawczych z zakresu ekologii, biologii ewolucyjnej, konserwatorskiej i ekologii behawioralnej
- termin pojawia się na początku lat 90tych XX w.
- niektóre techniki stosowano od połowy lat 60tych
- jej szybki rozwój dzięki opracowaniu różnorodnych **markerów molekularnych opartych na DNA**
- nie jest prostym zastosowaniem technik molekularnych do „starych” problemów ekologicznych

Ekologia molekularna

Zagadnienia możliwe do badania wyłącznie lub przede wszystkim technikami molekularnymi:

- analizy rodzicielstwa i pokrewieństwa
- identyfikacja osobników
- metody nieinwazyjne i ślady biologiczne
- śledzenie historii populacji i gatunków nieodróżnionych morfologicznie
- badania genetyczne wymarłych gatunków
- przybliżone lecz bardzo szybkie i uniwersalne metody oceny bioróżnorodności we wszystkich skalach

Ekologia molekularna **jest ściśle powiązana z teorią ewolucji**

Dane molekularne pozwoliły na testowanie kluczowych aspektów Syntetycznej Teorii Ewolucji (Nowoczesnej Syntezy)

Przewidywania wyprowadzone z syntetycznej teorii ewolucji stanowią podstawę wnioskowania o procesach ekologicznych z danych molekularnych

- teoria mutacji neutralnych
- zegary molekularne
- genetyka populacji
- koalescencja
- genomika

Dlaczego techniki biologii molekularnej stosuje się w ekologii?

- analizuje się **cechy dziedziczne**, unika się badania cech które wykazują znaczną plastyczność fenotypową
- są **uniwersalne** – od wirusów do człowieka
- dostarczają zazwyczaj **dokładniejszych** wyników niż tradycyjne metody i trwają znacznie **krócej**
- możliwa jest identyfikacja **genów** odpowiedzialnych za obserwowane cechy i zjawiska – **podstawy genetyczne procesów ekologicznych i ewolucyjnych**
- techniki analizy DNA wymagają **śladowych ilości materiału biologicznego** – mogą być nieinwazyjne i niedestrukcyjne

Czym są i do czego służą markery molekularne?

Wykazujące zmienność, dziedziczne cechy DNA, RNA lub białek użyteczne w identyfikacji:

- pozycji w obrębie genomu (mapy genetyczne)
- komórek i tkanek (np. nowotworowych)
- osobników (analizy rodzicielstwa, medycyna sądowa)
- populacji
- gatunków
- jednostek ewolucyjnych wyższego rzędu

A przez to pozwalające wnioskować o przebiegu procesów ekologicznych i ewolucyjnych

Czym są i do czego służą markery molekularne?

Markery molekularne pozwalają na uzyskanie **różnorodnych informacji biologicznych** na temat:

- systemów kojarzeń organizmów
- kierunków i wzorców migracji
- centrów różnorodności genetycznej gatunków
- bioróżnorodności (mikro)organizmów
- adaptacji na poziomie molekularnym

Wiele typów markerów molekularnych

Wybór markerów molekularnych

Sposób dziedziczenia

Po obu rodzicach:

- zlokalizowane na autosomach

Po jednym z rodziców:

- na **chromosomach płci**

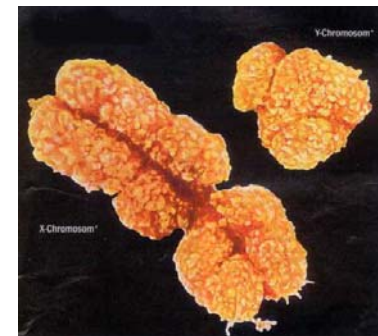
chromosom W u ptaków tylko po samicach

chromosom Y u ssaków tylko po samcach

- w **organellach**

mitochondria u zwierząt dziedziczą się po matce,
niektórych roślin po ojcu

chloroplasty zazwyczaj dziedziczą się po matce



Wybór markerów molekularnych

Tempo ewolucji

W zależności od badanego problemu przydatne będą markery ewoluujące **szybko**:

- identyfikacja osobników
- analiza pokrewieństwa w grupach rodzinnych i populacjach
- migracje (przepływ genów) między populacjami

W **średnim** tempie:

- historia określonego gatunku
- identyfikacja jednostek zasługujących na ochronę
- szybka ocena bioróżnorodności (kody kreskowe DNA)

Wolno:

- pokrewieństwa ewolucyjne większych grup
- ewolucja genomów i cech

Wybór markerów molekularnych

Kodominacja i dominacja

Marker **kodominujący** to taki, dla którego można określić pełny genotyp osobnika – np. rozróżnić heterozygotę od homozygoty

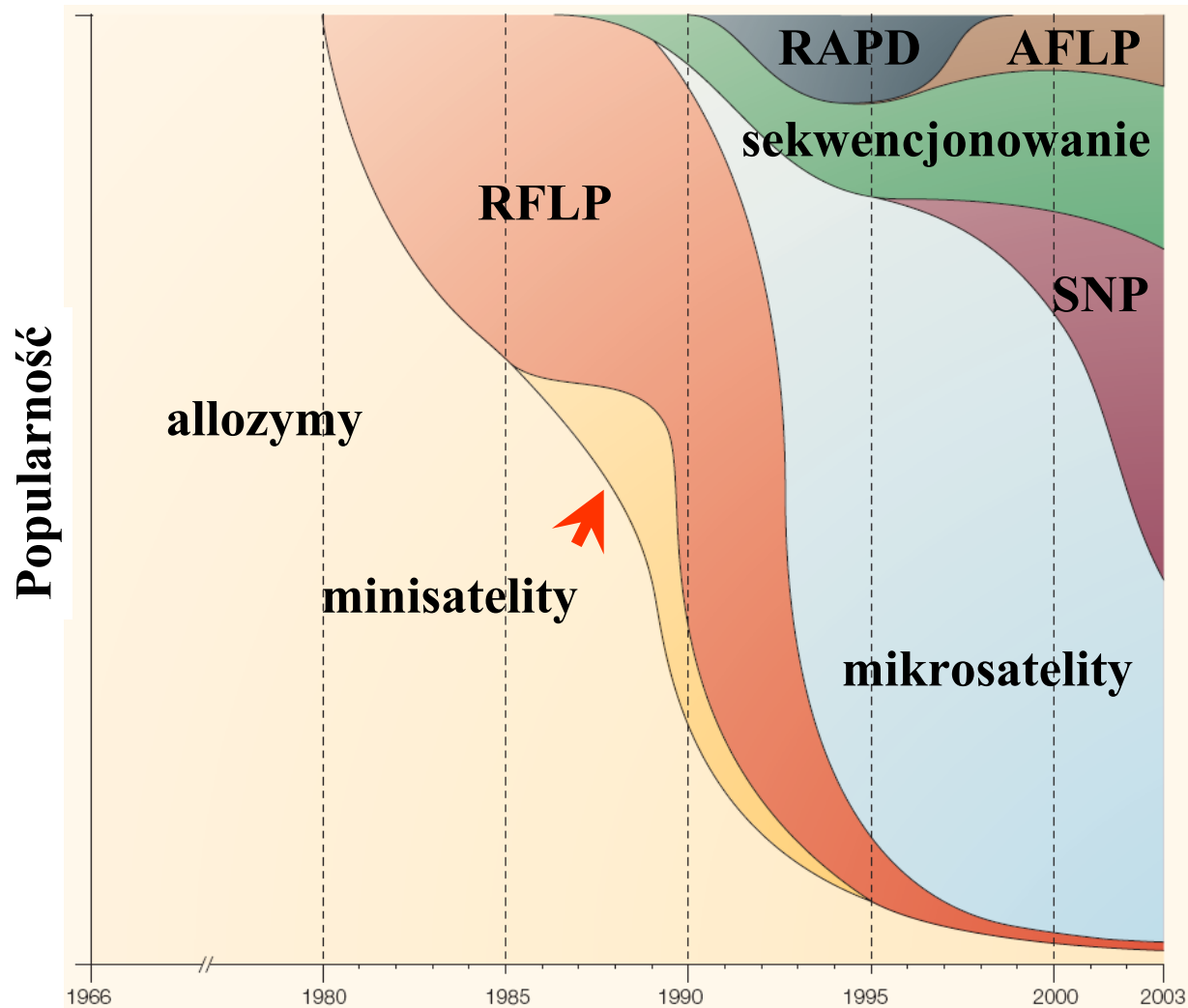
Bardzo użyteczne lecz trudniejsze do uzyskania niż dominujące

Marker **dominujący** to taki, w którym występowanie jednego allelu maskuje inne

Łatwo dostępne, można badając jednocześnie bardzo wiele markerów lecz nie dostarczają tak szczegółowych informacji. Słabo rozwinięte metody analizy

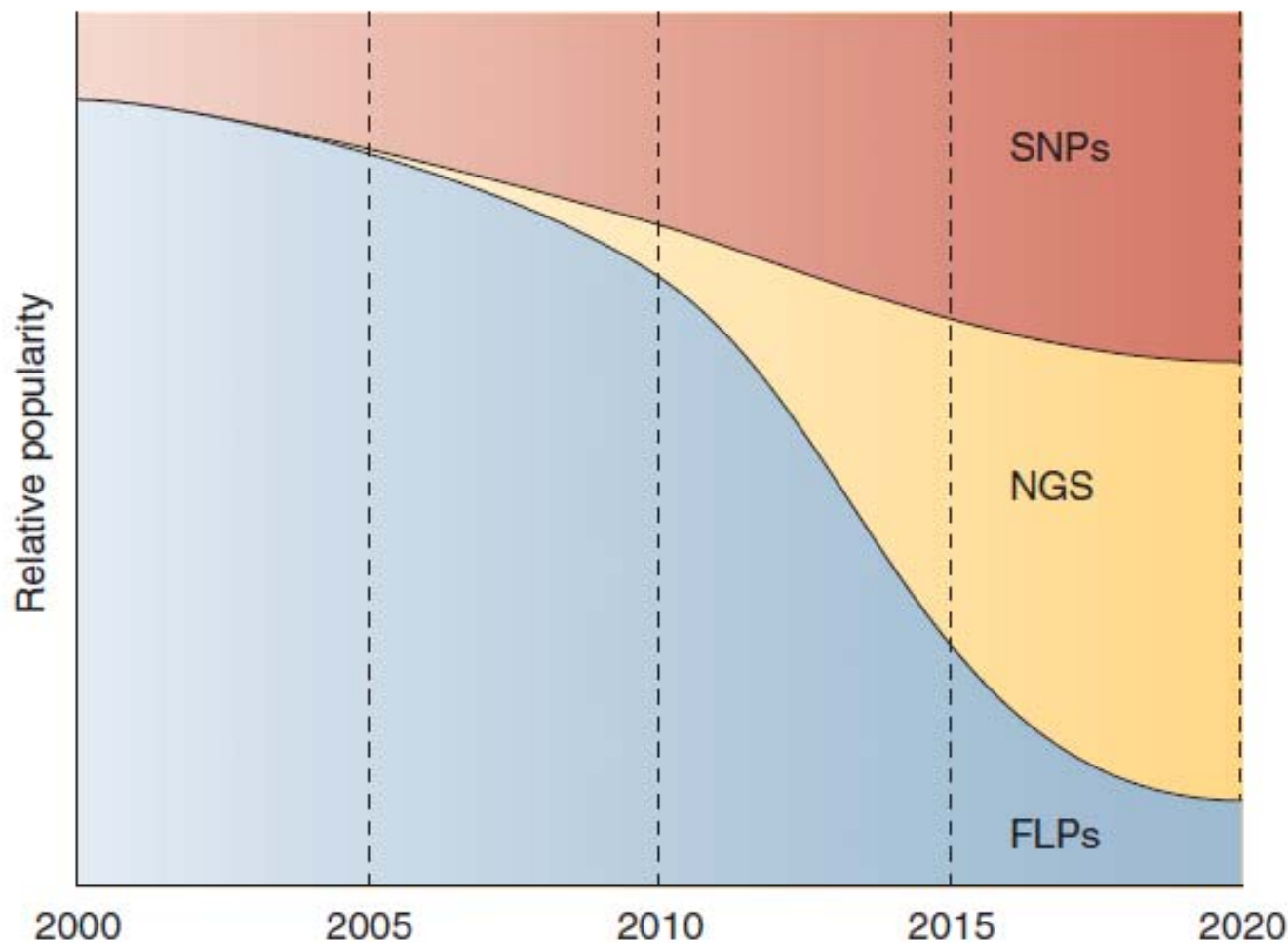
Wybór markerów molekularnych

Możliwości techniczne, koszt, moda



Wybór markerów molekularnych

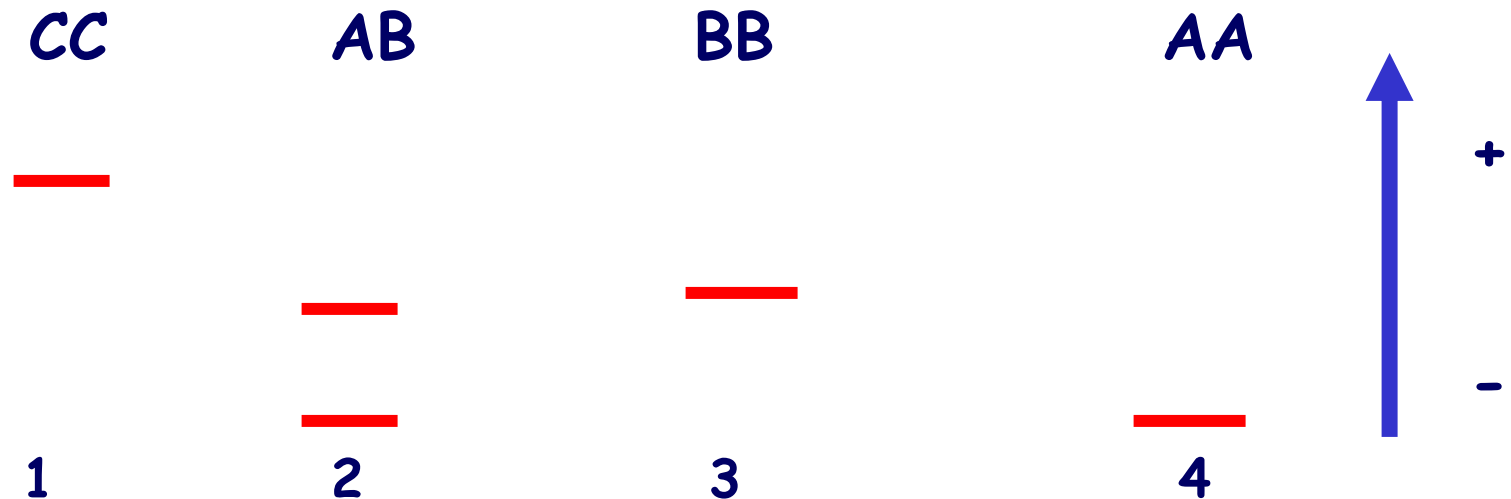
Możliwości techniczne, koszt, moda



Elektroforeza białek – Allozymy, Izozymy

Białka o różnym ładunku będą się przemieszczać w żelu w polu elektrycznym z różną prędkością.

Pierwsza technika biochemiczna szeroko stosowana w genetyce populacji: *Drosophila pseudoobscura* (Lewontin i Hubby 1966); człowiek (Harris 1966)



Elektroforeza białek – Allozymy, Izozymy

- Nawet pojedyncze zmiany aminokwasów mogą spowodować zmianę ładunku białka i wpłynąć na jego ruchliwość elektroforetyczną. Ok. **30% podstawień** aminokwasowych powoduje wykrywalne zmiany.
- Takie rozróżnialne elektroforetycznie formy danego białka to **allele**
- Nie wszystkie podstawienia nukleotydów w kodującej części genów powodują zmiany aminokwasów w białku (bo kod genetyczny jest zdegenerowany) – podstawienia **synonimowe** oraz **niesynonimowe**
- **Elektroforeza białek wykrywa jedynie część zmienności genetycznej**

Elektroforeza białek – Allozymy, Izozymy

Homogenat tkanki lub całego organizmu

Elektroforeza

skrobina

octan celulozy

akrylamid

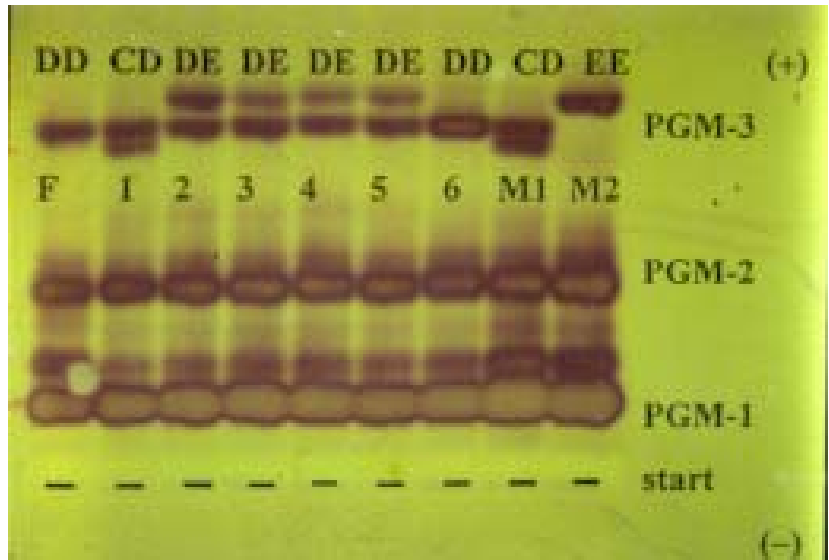


Specyficzne barwienie

enzymy (kilkadziesiąt)

hemoglobina

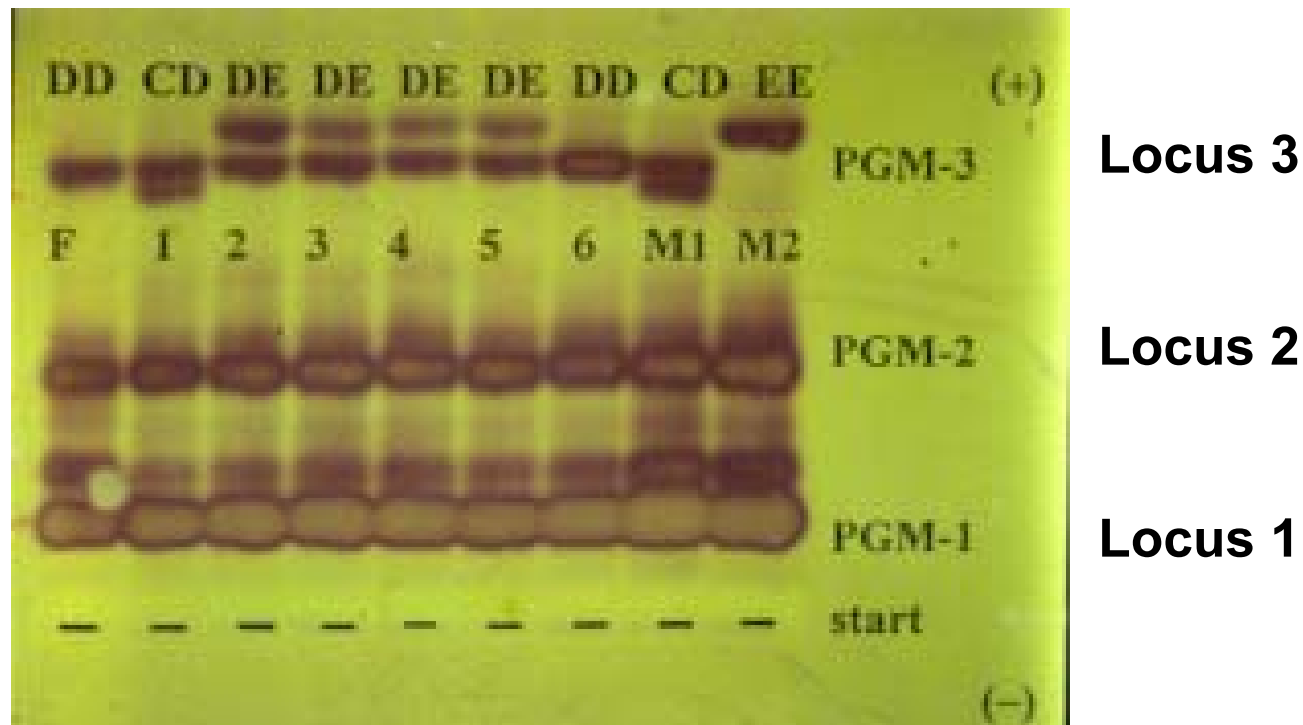
transferyny



Elektroforeza białek – Allozymy, Izozymy

Niektóre enzymy kodowane są przez kilka genów (loci), ich produkty białkowe można rozróżnić na żelu

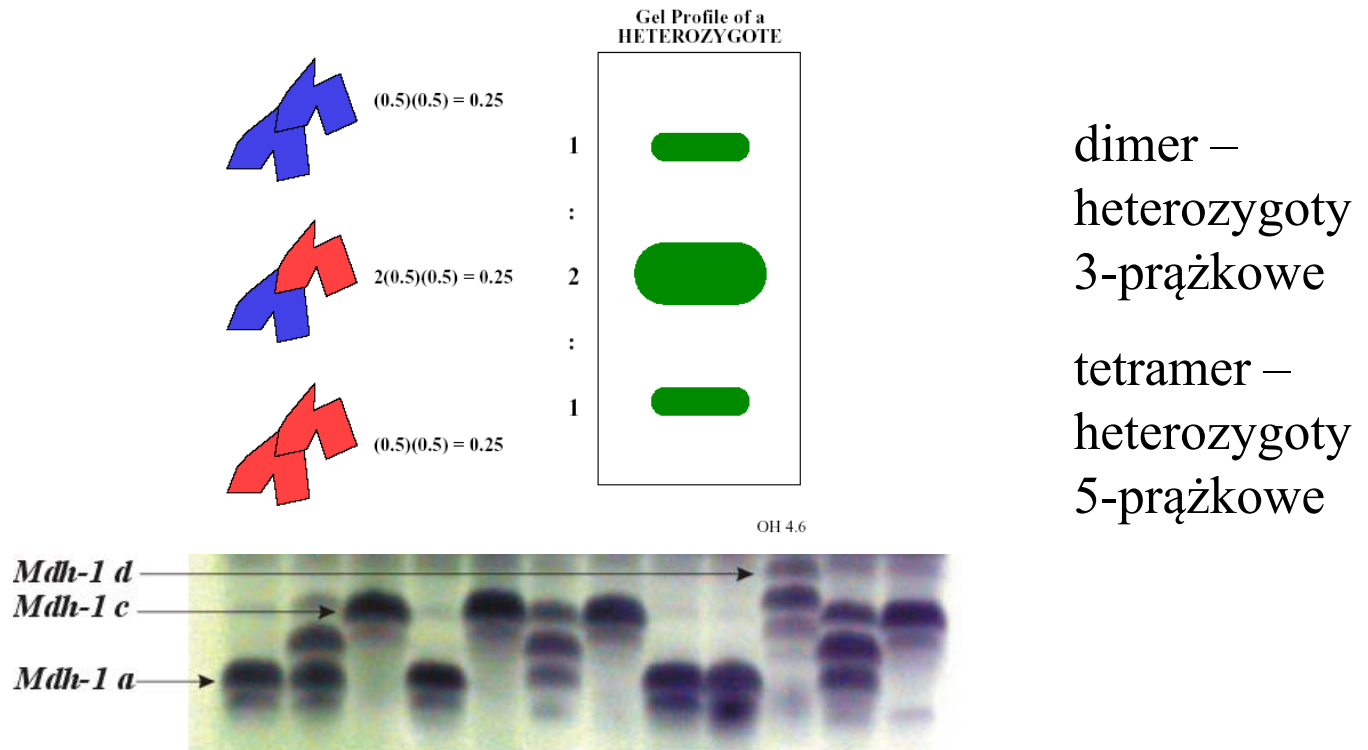
Układ prążków – białko monomerowe



Elektroforeza białek – Allozymy, Izozymy

Wiele enzymów jest dimerami lub tetramerami –
podjednostki kodowane przez jeden gen

Układ prążków w heterozygocie – białko dimerowe



Elektroforeza białek – Allozymy, Izozymy

Zalety:

- Można zbadać stosunkowo dużo loci (30-50)
- Metoda tania
- Prosty sposób dziedziczenia
- Opracowane narzędzia analizy danych
- Można badać niemal wszystkie taksony

Zastosowanie:

- Do oceny zmienności populacji
- Określanie różnic między populacjami
- Badanie efektów doboru naturalnego

Elektroforeza białek – Allozymy, Izozymy

Wady:

- umiarkowany poziom zmienności
- mogą być pod wpływem doboru naturalnego (a wiele zastosowań ekologicznych wymaga markerów neutralnych)
- trudności porównywania wyników różnych prac
- technika nie nadająca się do automatyzacji
- wymaganie świeżej lub zamrożonej tkanki
- technika prawie zawsze destrukcyjna, dlatego trudną ją stosować u gatunków rzadkich i chronionych

Elektroforeza białek – Allozymy, Izozymy

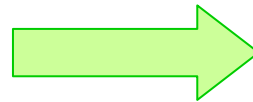
- Pierwsza technika która umożliwiła uzyskanie wiarygodnych informacji na temat **poziomu zmienności genetycznej** w naturze
- Tysiące prac na wszystkich grupach taksonomicznych
- Trudno przecenić jej znaczenie **teoretyczne** i **praktyczne** (teoria mutacji neutralnych, zegary molekularne, badania adaptacji, rodzicielstwa, biologia konserwatorska, kontrola handlu gatunkami zagrożonymi itd.)
- Ciągłe stosowane, choć w bardzo dużym stopniu zastąpione przez markery oparte na DNA

Techniki DNA

- Dają dostęp do zmienności na najbardziej podstawowym poziomie
- Praktycznie nieograniczona liczba markerów
- Stosowane dla wszystkich organizmów
- Wiele rodzajów, w zależności od potrzeb
- Większość opiera się o reakcję PCR
- Zazwyczaj niedestrukcyjne
- Mogą być też nieinwazyjne
- Łatwe do automatyzacji i miniaturyzacji

Typowe etapy analizy DNA

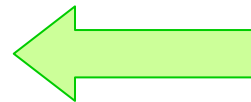
pobranie prób



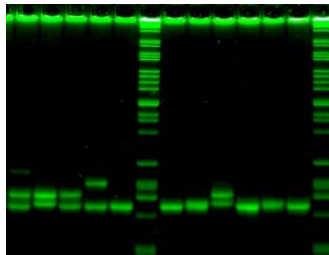
izolacja DNA



namnożenie
(PCR)
określonych
markerów



genotypowanie
markerów



Pobieranie prób do analiz genetycznych

Metody inwazyjne – Dużo DNA, dobrej jakości

- destrukcyjne (zabicie organizmu !)
- niedestrukcyjne (pobranie krwi, fragmentu tkanki, włosów, piór, liści)

Metody nieinwazyjne – Mało DNA, słabej jakości

pobranie tzw. śladów biologicznych pozostawionych przez zwierzę w terenie (pióra, włosy, odchody, wyplówki sów)

Odpowiednie utrwalenie:
wysuszenie, zamrożenie,
alkohol, bufor, itd.

Izolacja DNA

- rozdzielenie komórek, rozbicie/strawienie ścian komórkowych
 - homogenizacja mechaniczna
 - homogenizacja w ciekłym azocie
 - trawienie chitynazą
 - trawienie proteinazą K
- liza komórek
 - detergent w buforze zasadowym
- oczyszczenie DNA – usunięcie białek, lipidów, cukrów
 - ekstrakcja organiczna (fenol/chloroform) lub wysalanie białek, następnie strącanie DNA alkoholem
 - wiązanie DNA na kolumnach wypełnionych krzemionką
- przechowywanie DNA
 - DNA zawiesza się w lekko zasadowym buforze z dodatkiem EDTA (wychwytuje jony Mg^{2+}) i przechowuje w +4 lub -20°C

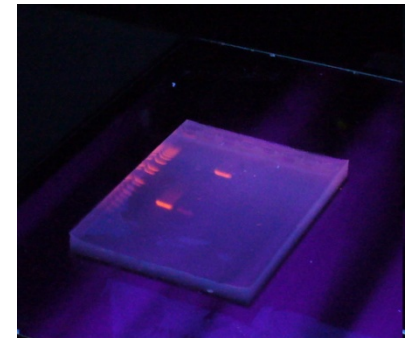
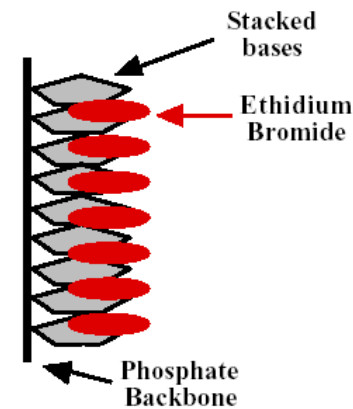
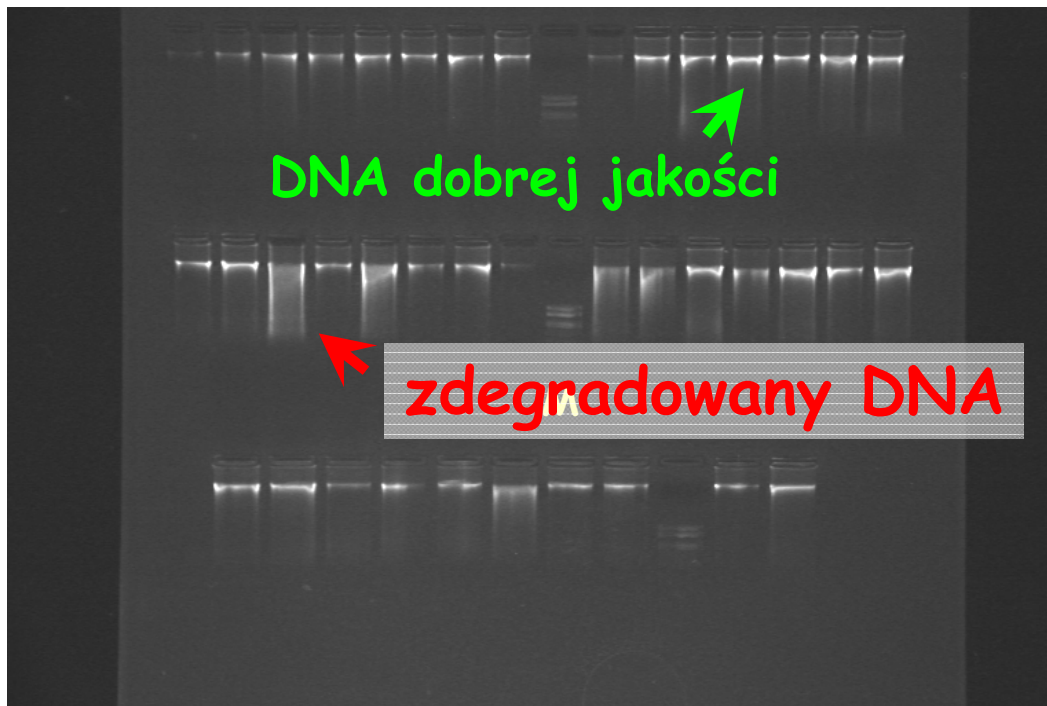
Izolacja DNA

- istnieje mnóstwo **dostępnych komercyjnie zestawów**
- wybór metody izolacji zależy od:
 - organizmu
 - tkanki
 - jakości materiału
 - sposobu utrwalenia
 - żądanej **ilości** DNA
 - żądanej **jakości** DNA
- izolacja DNA możliwa jest również z **materiału kopalnego**
 - ok. 500 tys. lat z rdzeni lodowych, kilkadziesiąt tys. lat ze skamieniałości
 - jedynie **krótkie fragmenty** – kilkadziesiąt pz
 - bardzo trudna, wymaga wyspecjalizowanych laboratoriów
 - konieczność wielostopniowych kontroli –
kontaminacja współczesnym DNA

Ocena ilości i jakości wyizolowanego DNA

Elektroforeza w żelu agarozowym o niskim stężeniu (0.5 – 1 %)

DNA barwi się bromkiem etydyny lub innymi, mniej toksycznymi barwnikami fluorescencyjnymi SYBR Gold, SYBR Safe, Gel Red itd., które świecą w świetle UV



Ocena ilości i jakości wyizolowanego DNA

Pomiar spektrofotometryczny,
absorbancja przy długości fali 260 nm

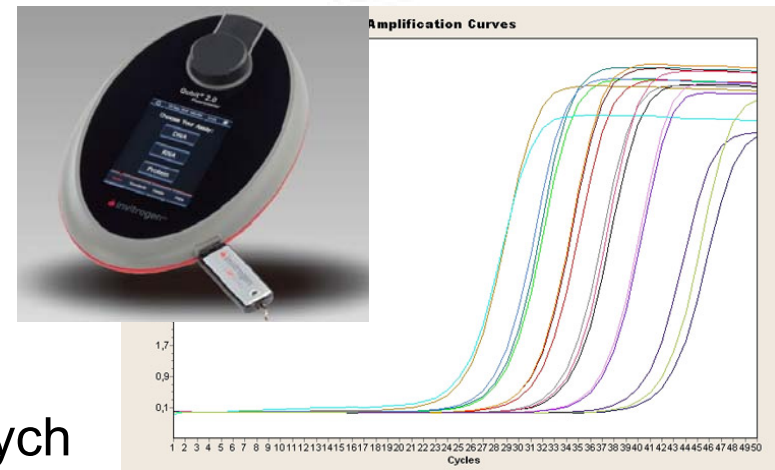
Czystość:

Dla czystego DNA proporcja A_{260}/A_{280}
wynosi 1.8 – 2.0 a $A_{260}/A_{230} > 2.0$

Pomiar można dokonać już z 1 ul roztworu
w zakresie stężeń 3 – 3000 ng/ul
mierzy wszystkie kwasy nukleinowe

Małe stężenia DNA można dokładnie
oznaczyć ilościowym PCR
w czasie rzeczywistym (real-time PCR)
lub używając barwników fluorescencyjnych
we fluorymetrach

Pomiar może być specyficzny dla dwuniciowego DNA

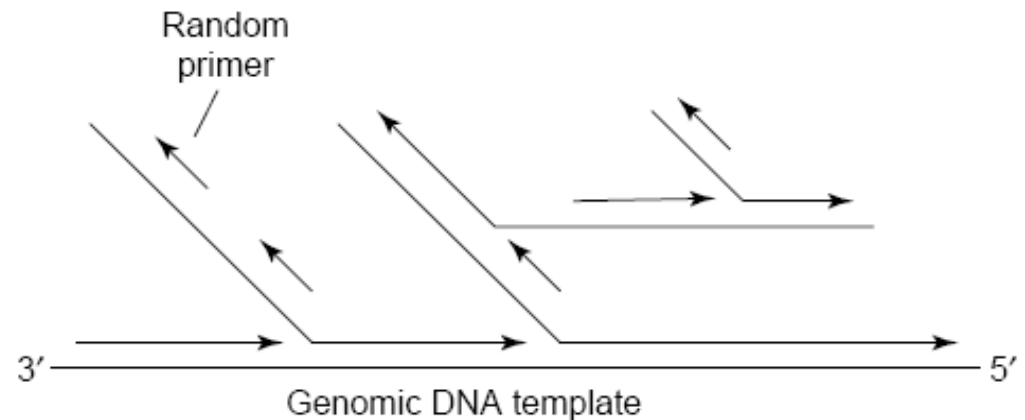


Ile i jakiej jakości DNA potrzeba do analiz?

- liczy się **liczba kopii genomu** w izolacie
- bakteria $\sim 10^6 - 10^7$ bp = 1 – 10 fg ($1 - 10 \times 10^{-15}$ g) DNA
- komórka ludzka – ok. 6×10^9 bp = 6 pg (6×10^{-12} g) DNA
- typowa izolacja „mikro” z ssaczjej tkanki (200 ul krwi, 10^7 komórek, 10-25 mg tkanki) – 4 – 30 ug ($4 - 30 \times 10^{-6}$ g) DNA
- do 1 reakcji PCR potrzeba 1-100 ng ssaczego DNA , więc **izolacja mikro** wystarcza na **> 1000 reakcji** PCR
- długość fragmentów DNA uzyskiwanych standardowymi metodami z dobrej jakości materiału to **$\sim 10 - 20$ kb** – wystarczająca do większości zastosowań
- nawet DNA mocno zdegradowany (fragmenty kilkadziesiąt – kilkaset bp) może być w niektórych sytuacjach wykorzystywany do badań techniką PCR

Amplifikacja całych genomów

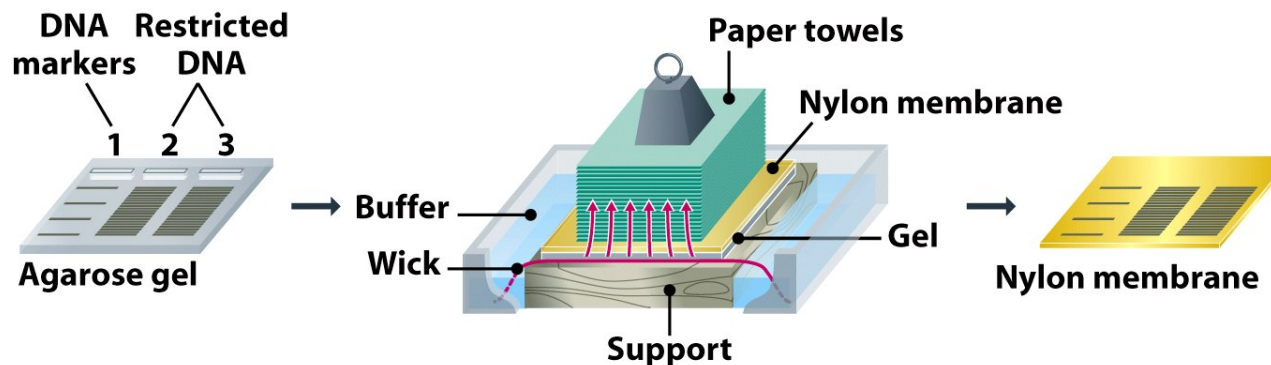
- Jak uzyskać dużą ilość DNA przy zachowaniu „składu” oryginalnego izolatu?
- Namnożony DNA będzie jedynie tak „dobry” jak oryginał
- **Izotermalna amplifikacja** z wykorzystaniem polimerazy bakteriofaga $\phi 29$ i wymiany nici (*strand displacement*)
 - losowe 6-mery jako startery
 - **długie fragmenty** do 100 kb
 - można uzyskać **dowolną ilość** DNA
 - możliwa amplifikacja **z 1 komórki**
 - mało błędów (10^{-6})



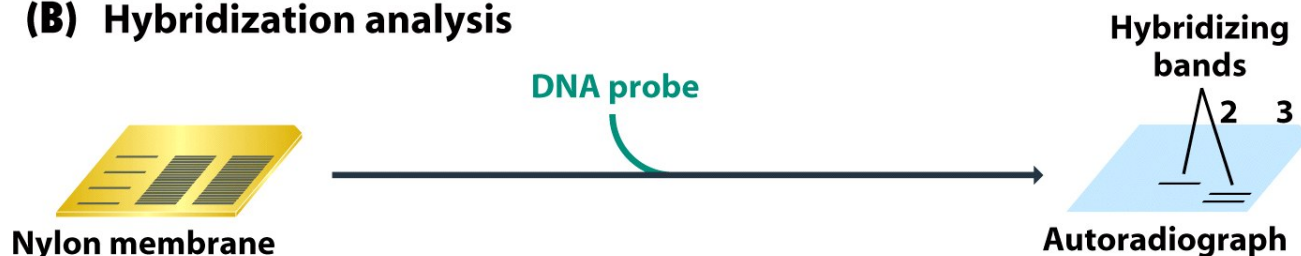
Profilowanie DNA – genetyczne odciski palców

- Całkowity DNA tnę się enzymem restrykcyjnym
- Fragmenty rozdziela się w żelu agarozowym – polimorfizm długości fragmentów restrykcyjnych (RFLP)
- DNA przenosi się na membranę i wiąże do niej
- Hybrydyzacja ze znakowaną sondą

(A) Transfer of DNA from gel to membrane

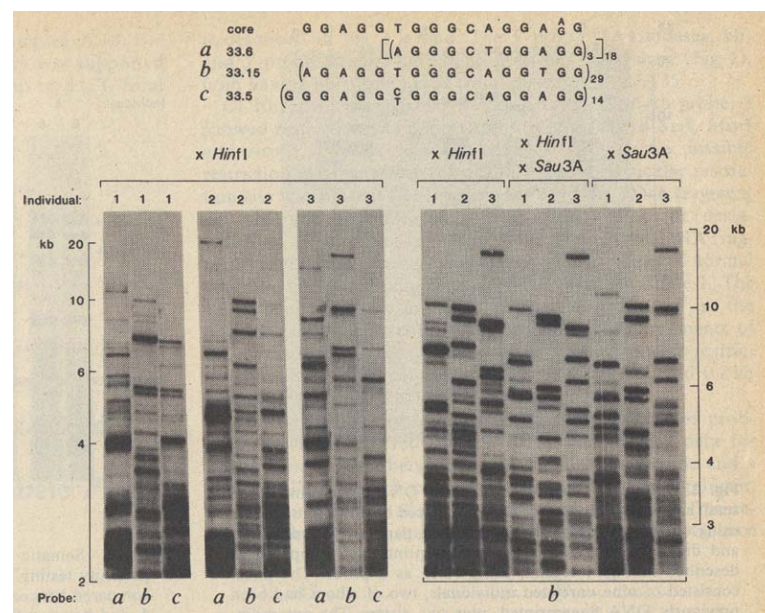


(B) Hybridization analysis



Profilowanie DNA – genetyczne odciski palców

- Gdy jako sondy użyjemy sekwencji minisatelitarnej, tworzącej w genomie wiele powtórzeń, każdy osobnik uzyska indywidualny wzór prążków
- Technika idealna dla identyfikacji osobników (unikatowe profile jak odciski palców)
- Trudna technicznie
- Wymaga dużo, dobrej jakości DNA
- Ma znaczenie historyczne – pierwsza technika stosowana do identyfikacji osobników w erze przed-PCR (od 1985 r.)
- Reprezentuje szerszą rodzinę technik RFLP opartych na hybrydyzacji Southerna

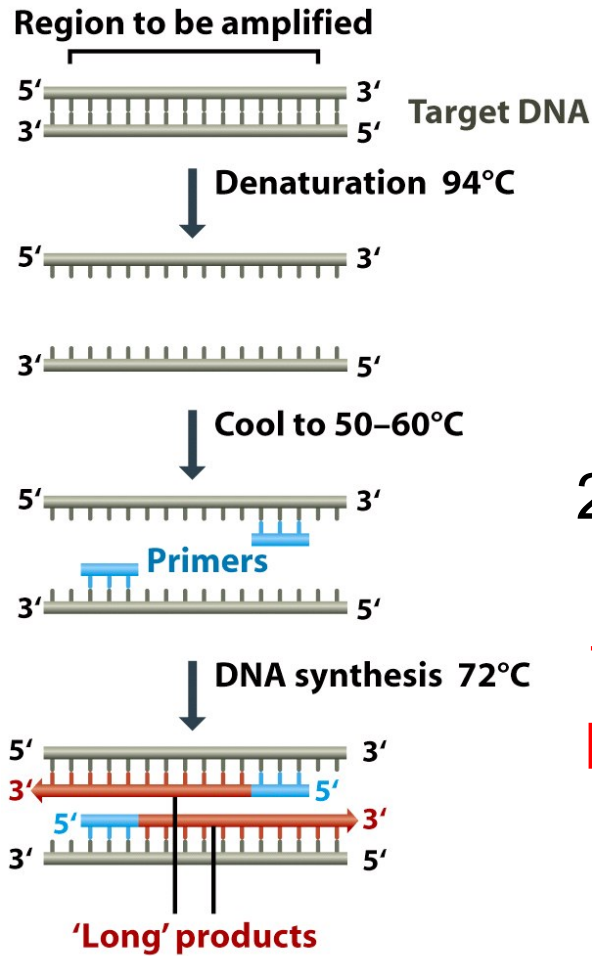


Łańcuchowa reakcja polimerazy - PCR

Najważniejsza technika ekologii molekularnej

- Pozwala uzyskać ze złożonej mieszaniny fragmentów DNA, np. całego genomu człowieka, miliardy kopii ściśle określonego fragmentu DNA
- O tym, który fragment będzie namnożony decydują startery, krótkie (15-30 bp) jednoniciowe fragmenty DNA o znanej sekwencji, które przyczepiają się do matrycy DNA definiując fragment ulegający amplifikacji; startery syntetyzuje się chemicznie
- Amplifikacja może odbywać się z minimalnej ilości wyjściowego DNA, nawet z 1 cząsteczki

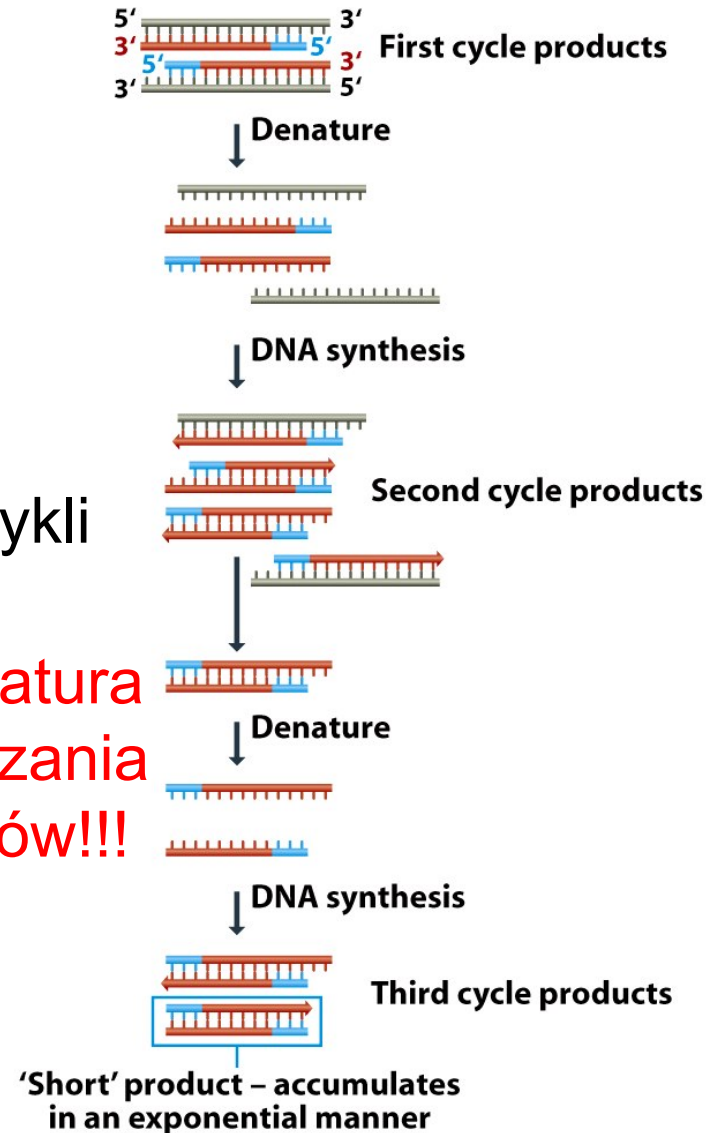
PCR



20-40 cykli

temperatura
przyłączania
starterów!!!

animacja

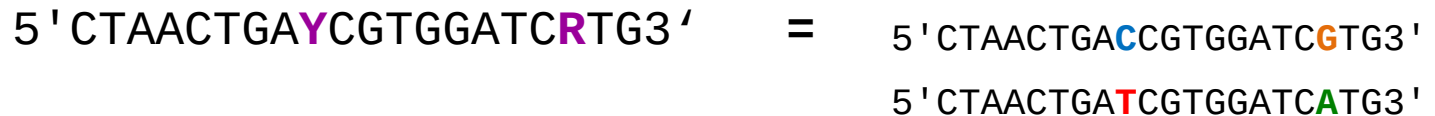


PCR

- **Termostabilna polimeraza *Taq*** lub podobne – bo denaturacja dwuniciowego DNA w temp. $>90^{\circ}\text{C}$
- Amplifikacja fragmentów 100 – 5000 bp
- Zastosowanie specjalnych polimeraz, amplifikacja fragmentów do 20-50 kb
- Odpowiedni dobór starterów – muszą wydajnie amplifikować DNA, wzajemna komplementarność, tworzenie struktur drugorzędowych

- Startery **konserwatywne** i **specyficzne**

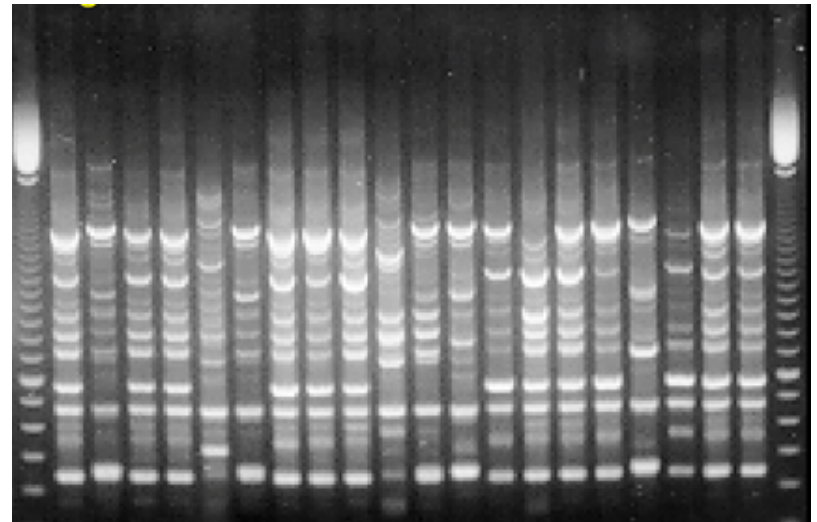
- **Startery zdegenerowane**



- Stosowanie wielu różnych starterów do amplifikacji wielu różnych fragmentów DNA w jednej reakcji – **multipleks PCR**

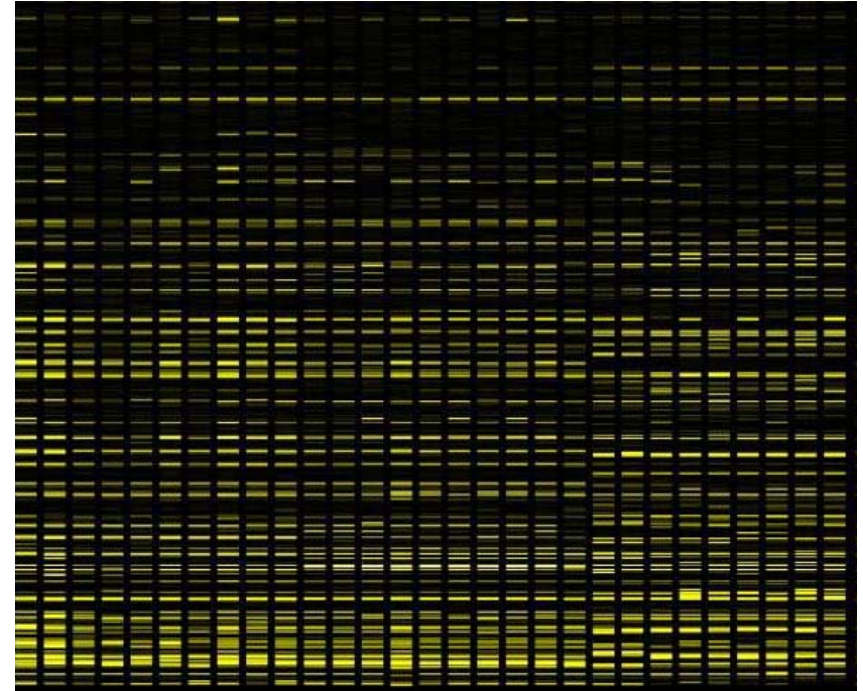
Analiza polimorfizmu losowo namnożonych fragmentów DNA – RAPD

- w technice używa się arbitralnie wybranych starterów PCR o dł. ok. 10 nukleotydów, niska temperatura przyłączania starterów, $\sim 40^{\circ}\text{C}$, namnaża się wiele fragmentów genomu - liczne prążki
- metoda tania i prosta
- sprawdza się obecność lub brak prążka danej długości
- markery **dominujące** trudna analiza
- znaczna zmienność
- **problemy z powtarzalnością** obecnie rzadko stosowana



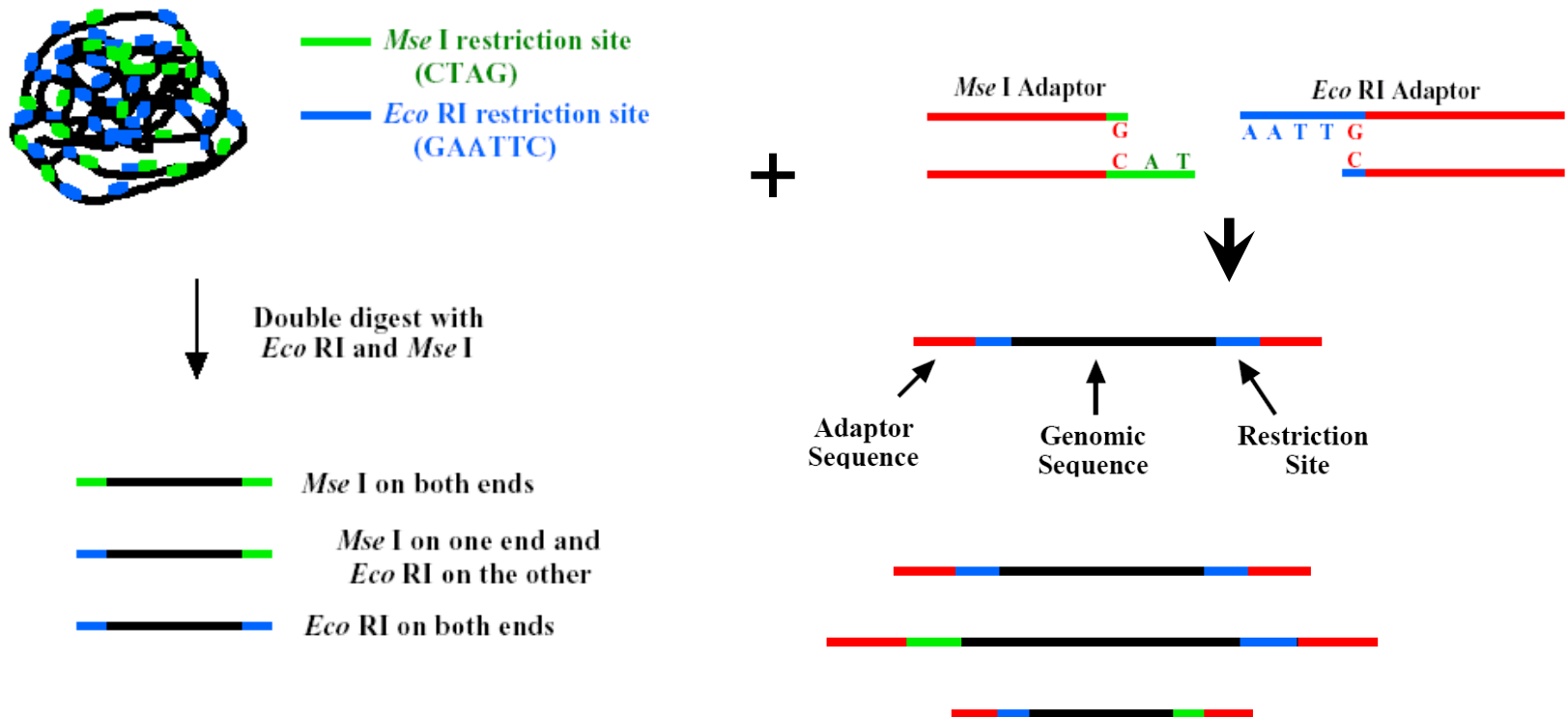
Analiza polimorfizmu długości namnożonych fragmentów DNA - AFLP

- Idea metody podobna do RFLP i RAPD, lecz stosuje się dodatkowo „adaptery” oraz selektywną amplifikację
- wysoka powtarzalność
- uzyskany obraz to setki prążków,
- zakres analizy 50 – 500 bp
- rejestrowanie obecności lub braku prążka o danej wielkości (system 1, 0; dwa allele)
- markery dominujące – nie potrafimy rozróżnić heterozygot od homozygot



AFLP

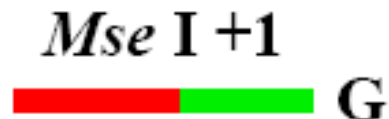
1. Cięcie enzymami restrykcyjnymi i ligacja adapterów



setki tysięcy fragmentów

AFLP

2. PCR I - preselektywna amplifikacja, nieznakowane startery komplementarne do sekwencji adaptera wydłużone na końcu 3' o 1 nukleotyd, 16 x redukcja liczby fragmentów lecz nadal zbyt dużo



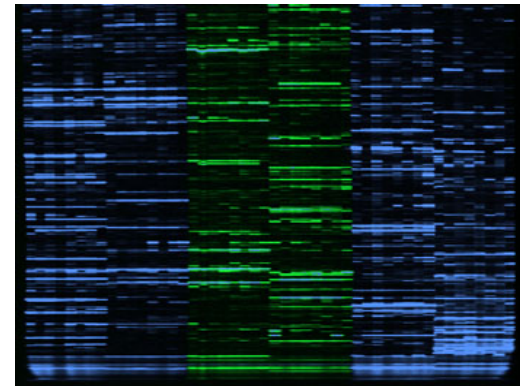
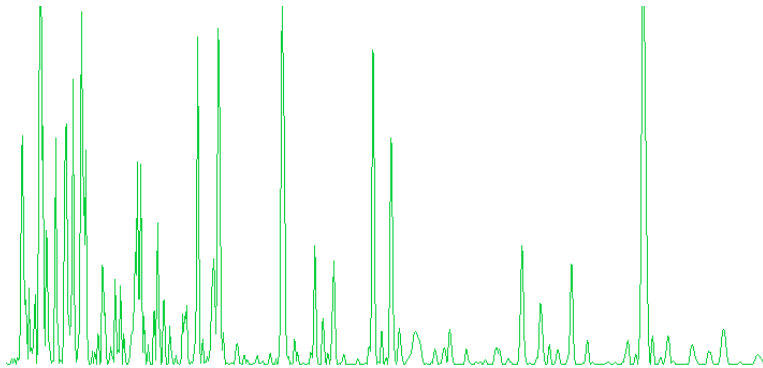
3. PCR II - selektywna amplifikacja, produkt PCR I namnaża się starterami z dodatkowymi dwoma nukleotydami – dalsza redukcja złożoności 256 x



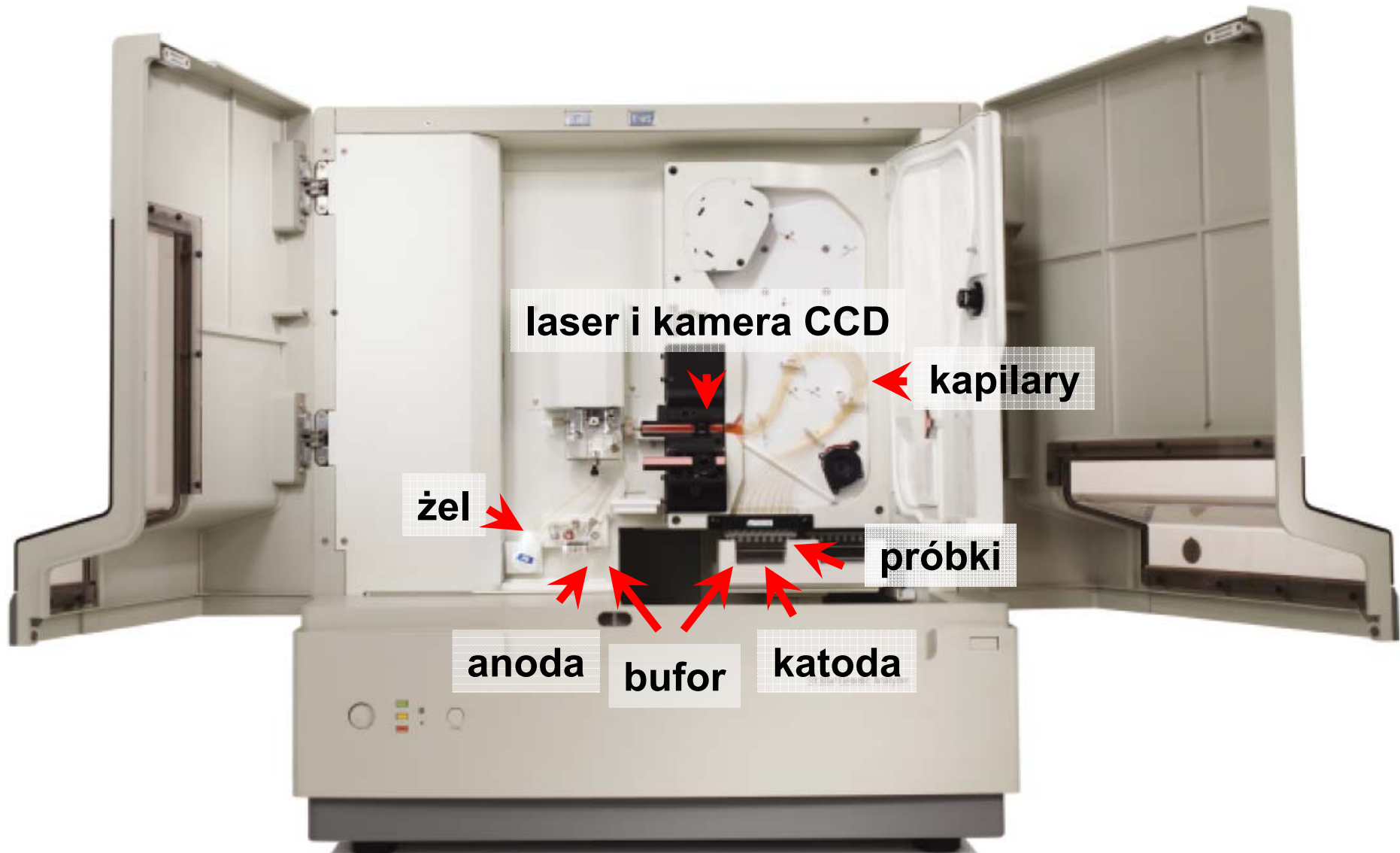
znakowanie tylko jednego startera powoduje że tylko fragmenty zawierające miejsce cięcia *Eco*RI będą widoczne

AFLP

- Dawniej rozdzielano produkty PCR II w żelach akrylamidowych i uwidaczniano przez autoradiografię lub barwienie np. srebrem
- Produkty PCR II rozdziela się w automatycznym analizatorze (sekwenatorze) DNA, uzyskuje się kilkadziesiąt prążków
- Można wykorzystać 64 pary starterów selektywnych – tysiące markerów AFLP dla dowolnego gatunku



Automatyczny analizator (sekwenator) DNA



AFLP

Rodzaje sekwencji

- sekwencje obecne w genomie w jednej, kilku lub wielu kopiach
- kodujące i niekodujące
- losowo rozmieszczone w genomie
- umiarkowanie lub wysoce zmienne

Zastosowanie

- identyfikacja osobników
- poziom zmienności w populacjach
- różnicowanie międzypopulacyjne
- filogeneza blisko spokrewnionych gatunków
- wykrywanie regionów genomu warunkujących adaptacje

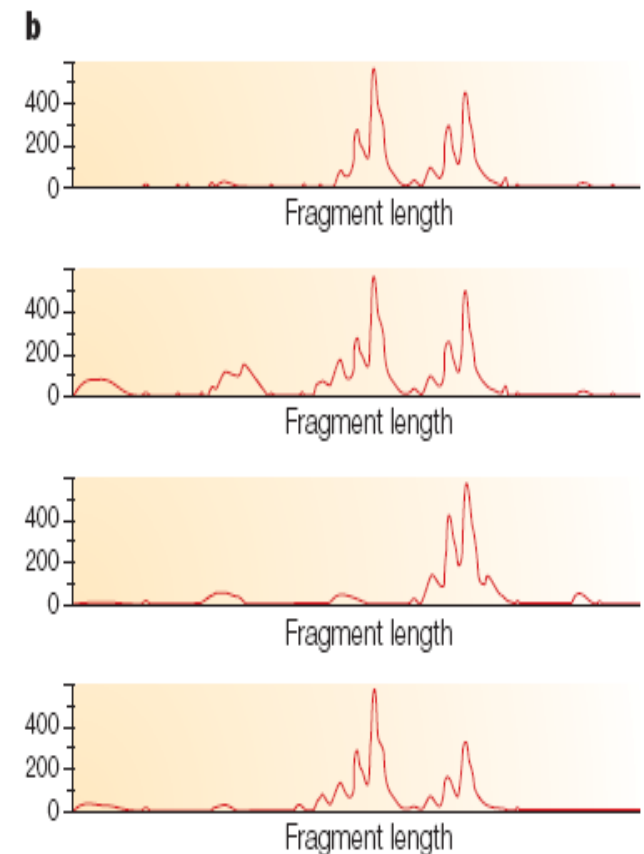
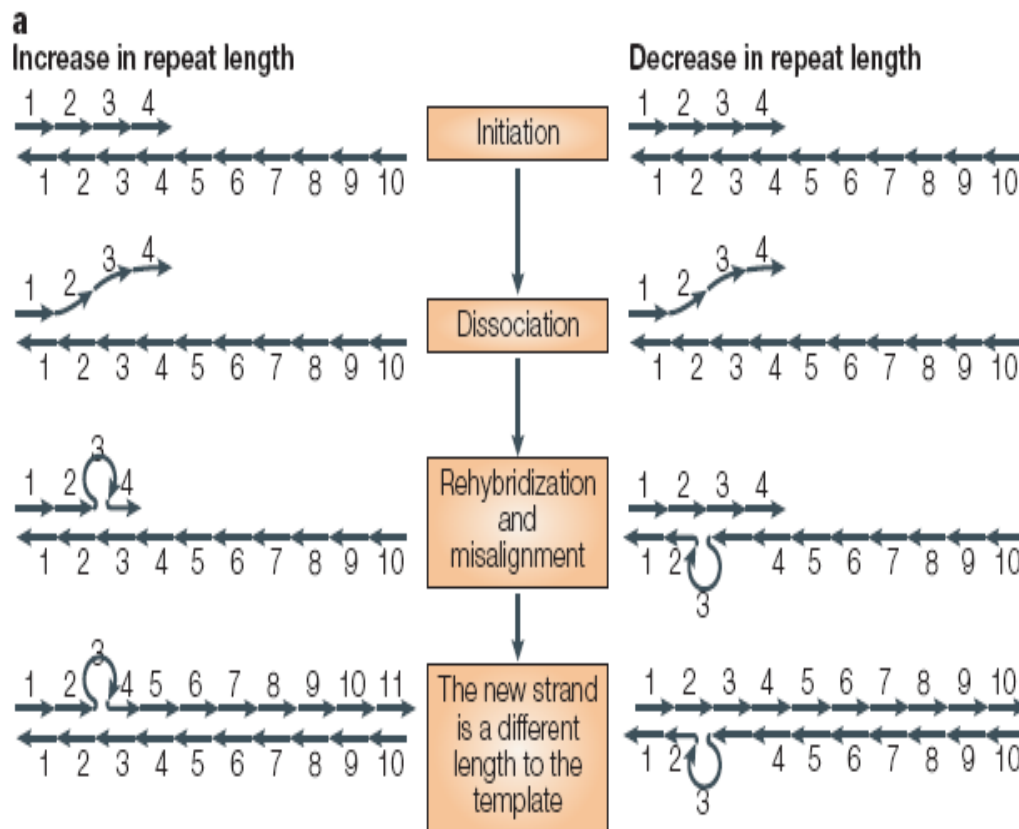
Mikrosatelity

- Wielokrotne powtórzenia 2-5 nukleotydów
- Prawdopodobnie losowo rozmieszczone w genomie
- Doskonałe markery w badaniach biologicznych gdyż wykazują zazwyczaj ogromną zmienność – wiele alleli w populacji
- Bardzo wysokie tempo mutacji (tysiące razy wyższe niż mutacji punktowych)
- Sposób mutacji złożony lecz prawdopodobnie dominuje poślizg polimerazy DNA



Mikrosatelity

Mutacje często w wyniku „poślizgu” polimerazy

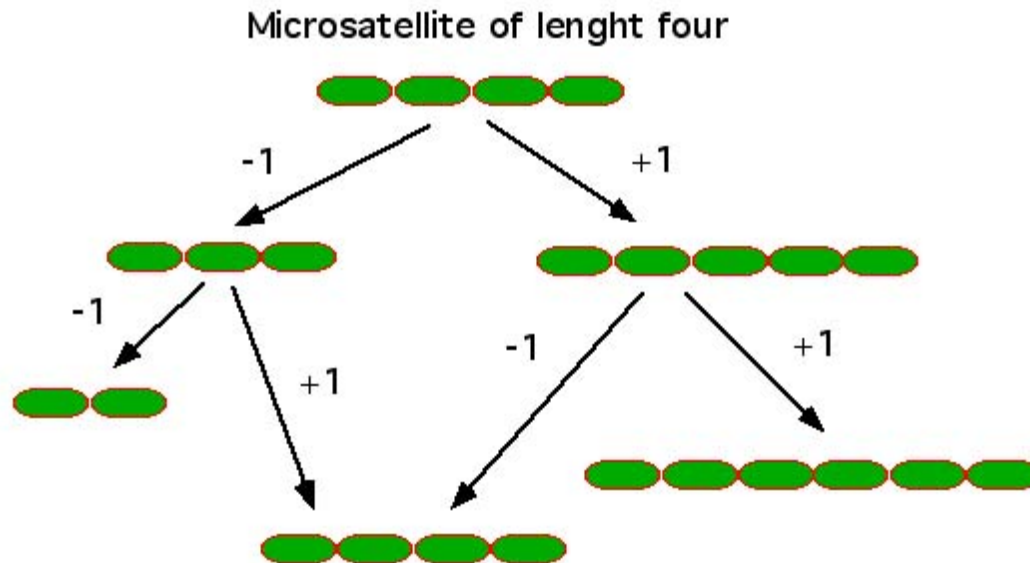


Mikrosatelity

- **kodominujące**, prosty sposób dziedziczenia
- **wysoka zmienność** – nawet kilkaset alleli na locus w populacji, zazwyczaj kilkanaście, kilkadziesiąt, na podstawie kilku-kilkunastu loci możliwa identyfikacja osobników z praktycznie 100% pewnością
- łatwa automatyzacja
- można namnażać kilka-kilkanaście loci w reakcji multiplex
- łatwy sposób wyznaczania pokrewieństwa między osobnikami
- dużo narzędzi do opracowywania i analizy statystycznej

Mikrosatelity

Krokowy model mutacji (SMM) dla mikrosatelitów



Wielkość nowego allelu (liczba par zasad) zależy od wielkości istniejących alleli

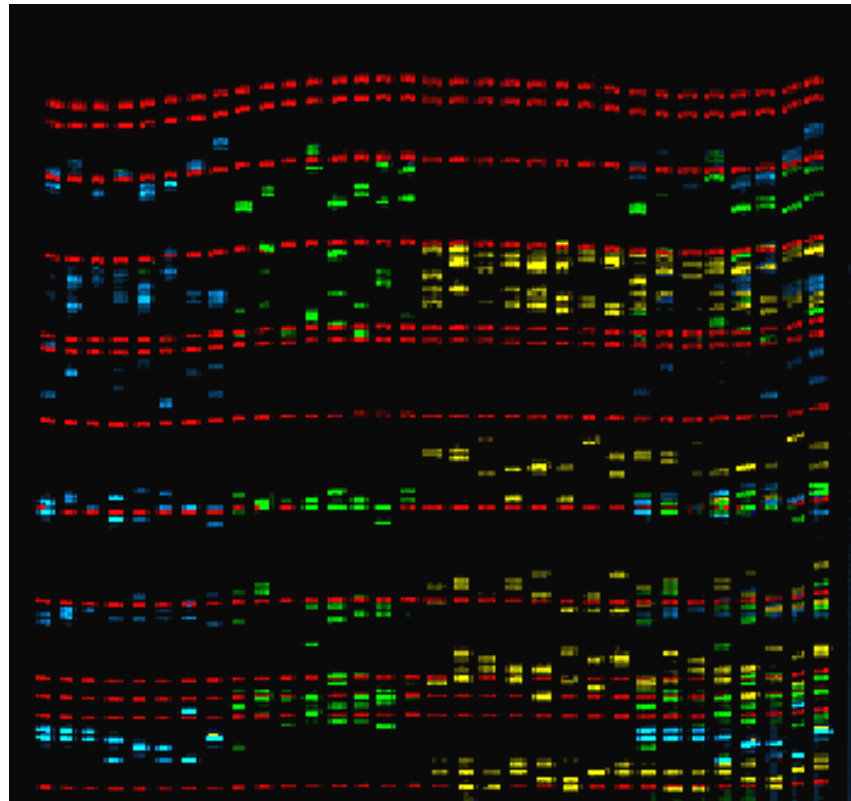
Model mutacji większości mikrosatelitów jest bardziej złożony, „mieszany”

Mikrosatelity

Genotypowanie zazwyczaj w automatycznym analizatorze DNA
PCR ze znakowanymi fluorescencyjnie starterami – 3-4 kolory

żel

długość fragmentu

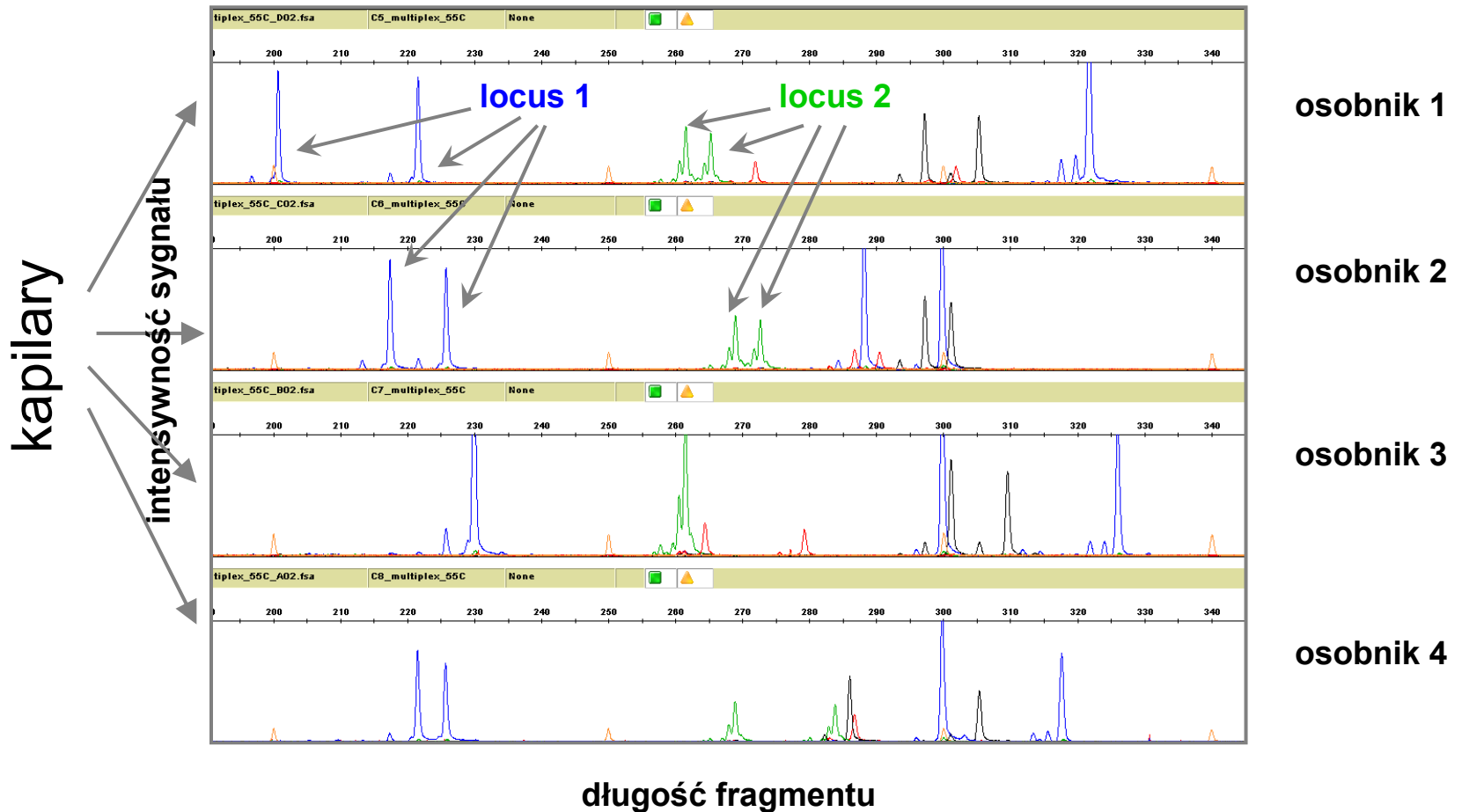


standard wielkości

32 osobniki

Mikrosatelite

Genotypowanie zazwyczaj w automatycznym analizatorze DNA PCR ze znakowanymi fluorescencyjnie starterami – 3-4 kolory



Mikrosatelity - trudności

- Startery są specyficzne dla gatunku lub blisko spokrewnionych gatunków – trzeba je opracowywać od nowa, a to jest dość trudne i kosztowne
- Złożony, mieszany model mutacji
- Występowanie homoplazji (dwa allele identycznej wielkości, np. 120 par zasad powstałe mutacji z 2 różnych alleli 118 pz przez dodanie jednego powtórzenia, oraz ze 122 przez usunięcie jednego powtórzenia)
- Allele zerowe – czyli takie których nasze startery nie namnażają, efektem będzie pozorny nadmiar homozygot w stosunku do częstości oczekiwanych z prawa Hardy’ego-Weinberga

Allele zerowe w mikrosatelitach

allele → ACTGTGCACCTGATCTG(AT)₁₀GTCTGTACTGATCCTA ✓
starter → TGACACGTGGACTAGAC CAGACATGACTAGGAT

ACTGTGCACCTGATCTG(AT)₁₇GTCTGTACTGATCCTA ✓
TGACACGTGGACTAGAC CAGACATGACTAGGAT

ACTGTGCACCTGATCTG(AT)₁₂GTCTGTACTGATCCTA ✓
TGACACGTGGACTAGAC CAGACATGACTAGGAT

ACTGTGCACCTGATCT**C**(AT)₁₂GTCTGTACTGATCCTA
! CAGACATGACTAGGAT zerowy - brak amplifikacji
TGACACGTGGACTAGAC

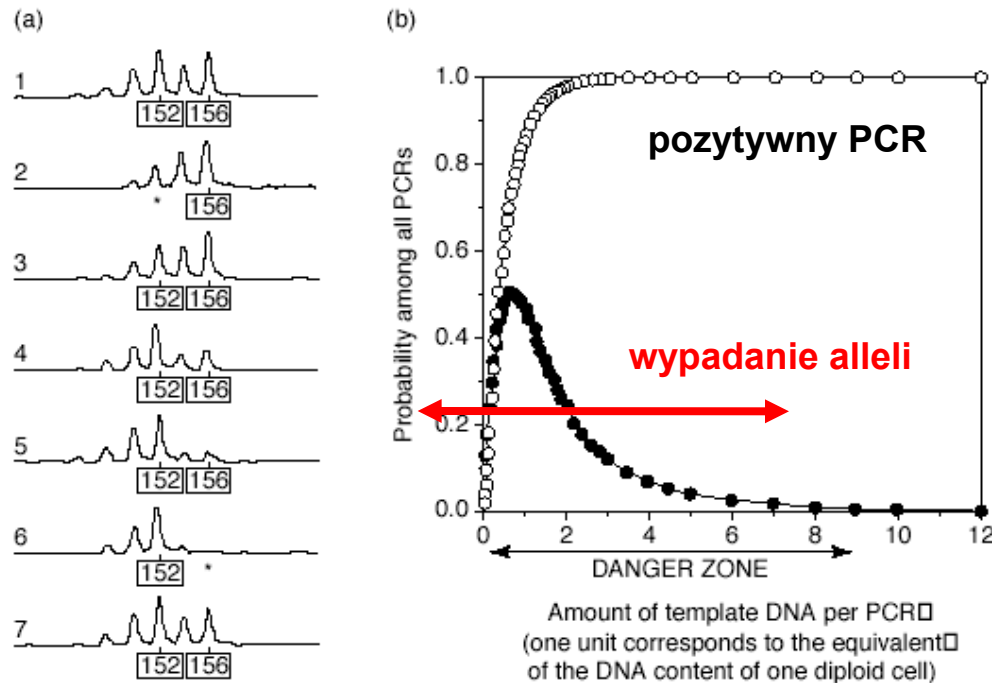
ACTGTGCACCTGATCTG(AT)₁₅GTCTGTACTGATCCTA ✓
TGACACGTGGACTAGAC CAGACATGACTAGGAT

ACTGTGCACCTGATCTG(AT)₁₀GTCTGTACTGATCCTA ✓
TGACACGTGGACTAGAC CAGACATGACTAGGAT

Mikrosatelite

Genotypowanie z nieinwazyjnie pobranych prób

- złej jakości DNA może spowodować brak produktu w reakcji PCR lub wywołać utratę jednego z alleli wypadanie alleli (*allelic dropout*)
- konieczność powtórzeń

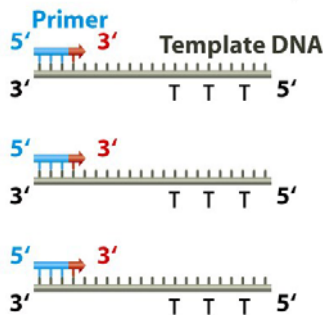


Taberlet i in (1999)

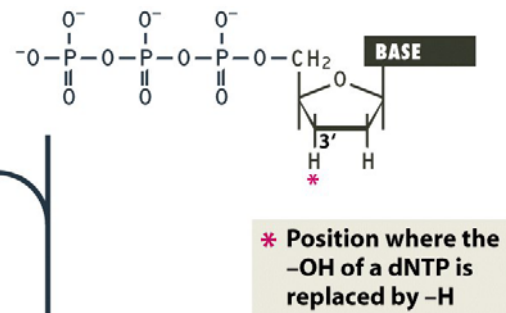
Sekwencjonowanie DNA metodą Sanger

- sekwencjonowanie przez syntezę i terminację łańcucha dideoksynukleotydami

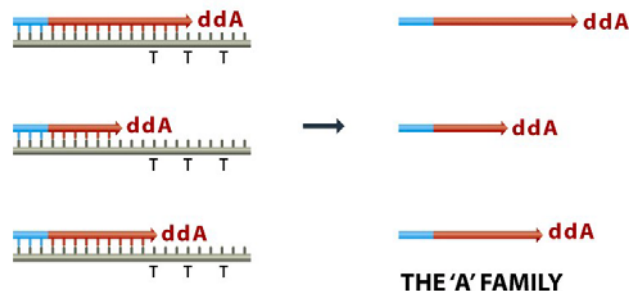
(A) Initiation of strand synthesis



(B) A dideoxynucleotide



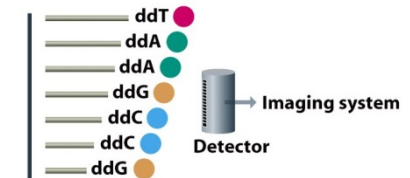
(C) Strand synthesis terminates when a ddNTP is added



(A)

ddA ● ddC ● ddT ● ddG ●
ddNTPs – each with a different fluorescent label

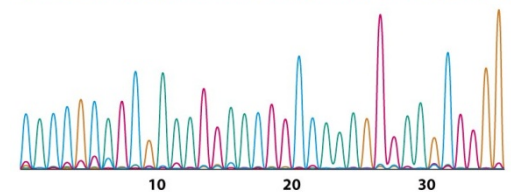
↓ Sequencing reactions, fractionation of products



Fluorescent bands move past the detector

(B)

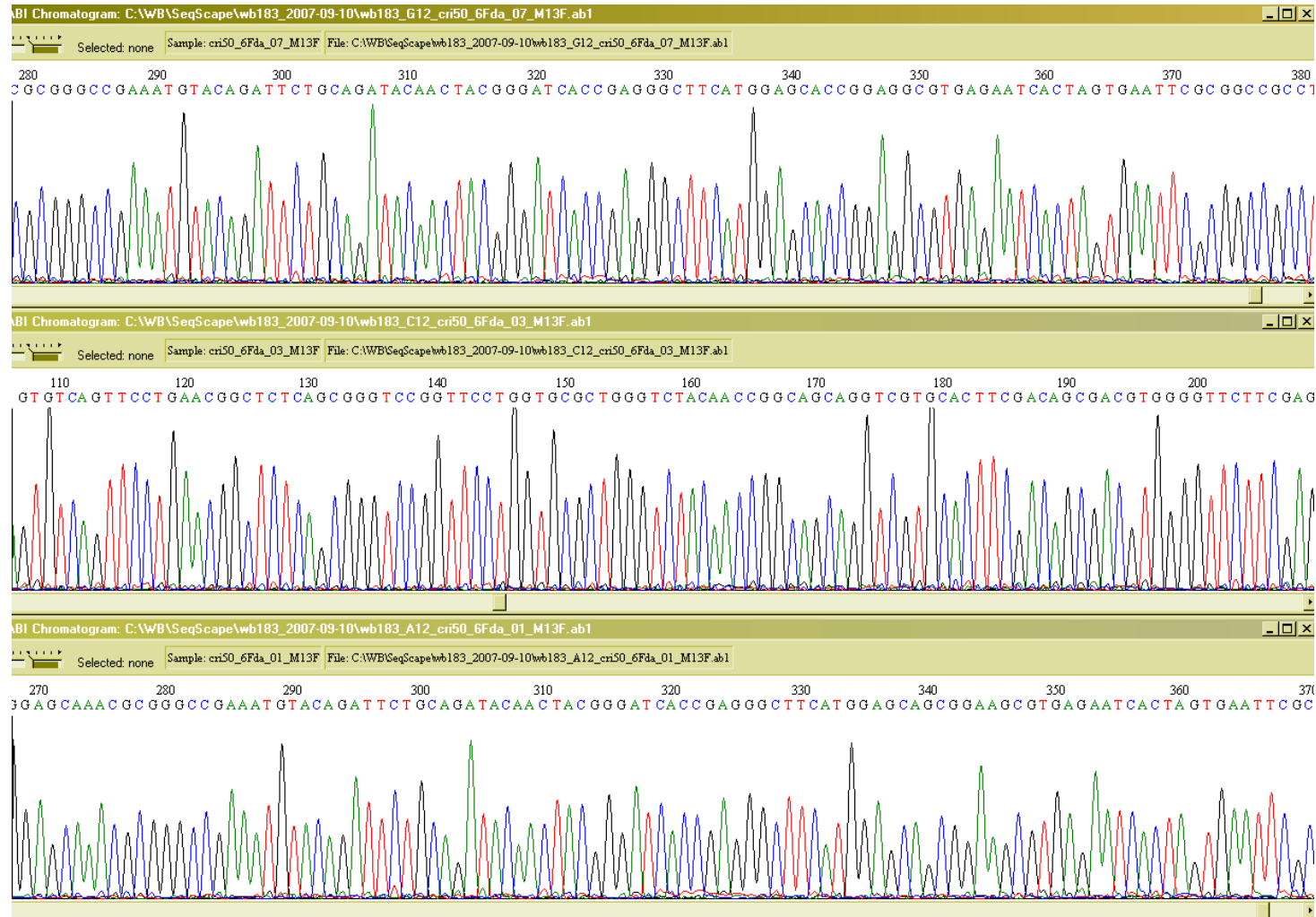
C A C C G C A T C G A A A T T A A C T T C C A A A G T T A A G C T T G G



animacja

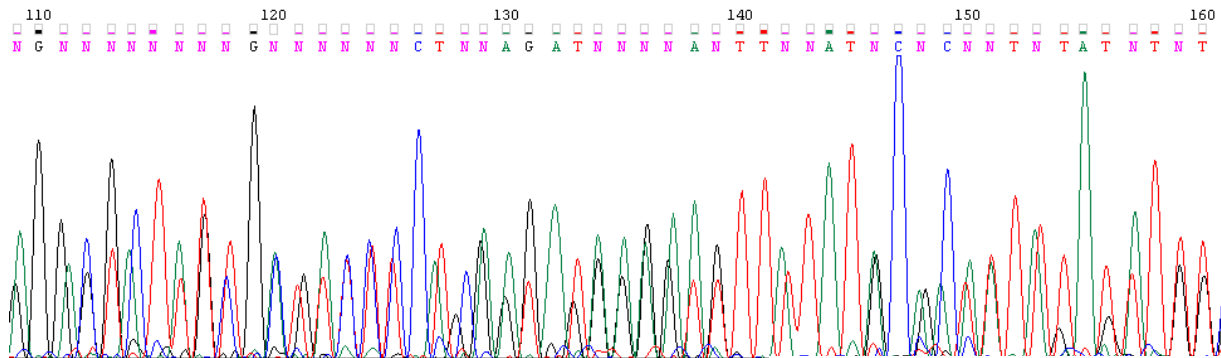
Sekwencjonowanie DNA metodą Sanger

Obraz sekwencji DNA z automatycznego sekwenatora



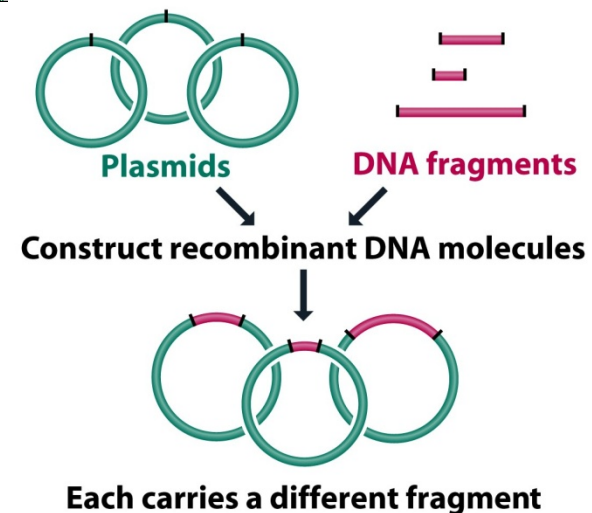
Sekwencjonowanie DNA metodą Sanger

Wymaga oczyszczonej, jednorodnej matrycy, inaczej sekwencja będzie nieodczytywalna



Jeżeli produkt PCR zawiera kilka różnych sekwencji trzeba go klonować w plazmidzie i sekwencjonować wiele klonów

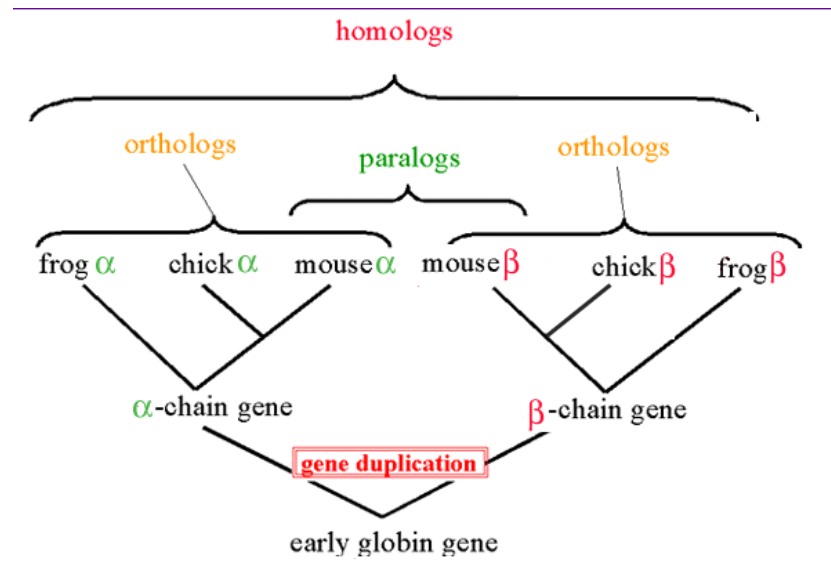
Pracochłonne i kosztowne



Sekwencjonowanie DNA metodą Sanger

- W ogromnej większości zastosowań ekologiczno-molekularnych porównuje się ze sobą jedynie **sekwencje homologiczne** - odpowiadające temu samemu odcinkowi danego genu – idealną techniką do przygotowania takich fragmentów jest PCR

Czasami tylko, do specjalnych problemów wykorzystuje się geny **paralogiczne**, powstałe przez duplikację i dywergencję



Zmienność sekwencji DNA

- sekwencjonowanie daje dostęp do dowolnej części genomu
- analizując zmienność DNA w populacji porównujemy między osobnikami **te same** (homologiczne) **pozycje nukleotydowe** – wyrównanie (alignment) sekwencji

dobre wyrównanie
sekw. homologiczne

```
GGTCGTGCGCTTCGACAGTGATGTGGGCGAGTATCACGCGGTGACCGAGCTGGGTTCGGAGTGATGCTGAGGTC
...A...C...
...T...C...
```

pozycja
nukleotydowa

złe wyrównanie
sekw. homologiczne

```
GGTCGTGCGCTTCGACAGTGATGTGGGCGAGTATCACGCGGTGACCGAGCTGGGTTCGGAGTGATGCTGAGGTC
A.GTCATGCGC.TCGACAGTGATGT..GC..AGTATCACGC.GTGA.CGAGCT..GTC.GAGTGATGCCGA.GT
A.GTCGTGCGC.TCGACAGTGATGT..GCCAGTATCACGC.GTGA.CGAGCT..CTC.GAGTGATGCTGA.GT
...T...A...C...
```

sekwencje
niehomologiczne

```
AGCTCGTTCGCTTCGACAGTGATGTGGGCGAGTATCACGCGGTGACCGAGCTGGGTTCGGAGTGATGCTGAGGT
G.TCGTTCGCT.TCAGCTGATGTG..CAAGTATCACGCG.TGAC.GAGCTG..TCG.AGTGATGC.GAG.TC
T.TGTT.T.CT..T.TATCC.CGTCTCTT.TCCCA.TA.TCC.G..CCT.ATCC..CACCCACC..GCCC.C
CCTA.AGCGCGG.G.GTG.A.CA.GA..G.CCCGAGTACT..CAT.A.GAGAC.CAGA.GACT.A.GAC.CCG
```

Sekwencjonowanie DNA

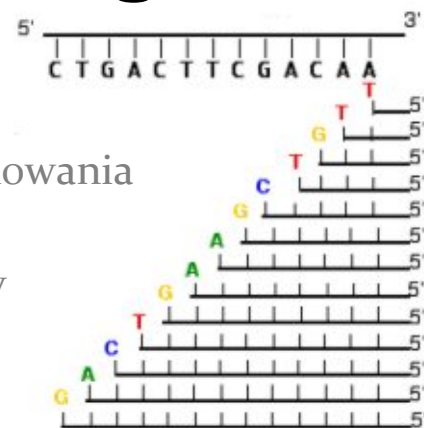
- rekonstrukcja filogenezy
- genomika
- filogeografia
- ocena bioróżnorodności
- ewolucja eksperymentalna
- medycyna sądowa
- medycyna

Znaczenie sekwencjonowania ciągle rośnie a nowe metody pozwalają na skokowe zwiększenie wydajności i zmniejszenie kosztów – genomika personalna, ale też ogromne perspektywy dla badania gatunków niemodelowych

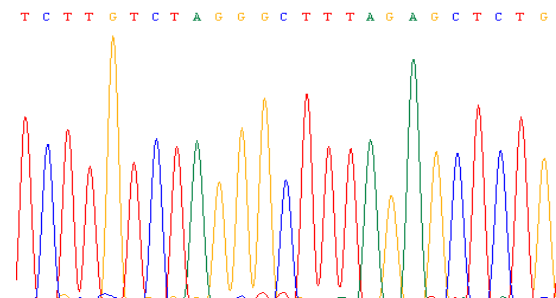
Sekwencjonowanie Sangera

- używane od ponad 30 lat
- uniwersalne
- zautomatyzowane
- wysoka jakość – gold standard
- długie sekwencje (>500 pz)
- trudne do miniaturyzacji i paralelizacji – ograniczona wydajność
- kosztowne

reakcja
sekwencjonowania
znakowane
terminatory



elektroforeza
kapilarna
detekcja
fluorescencyjna

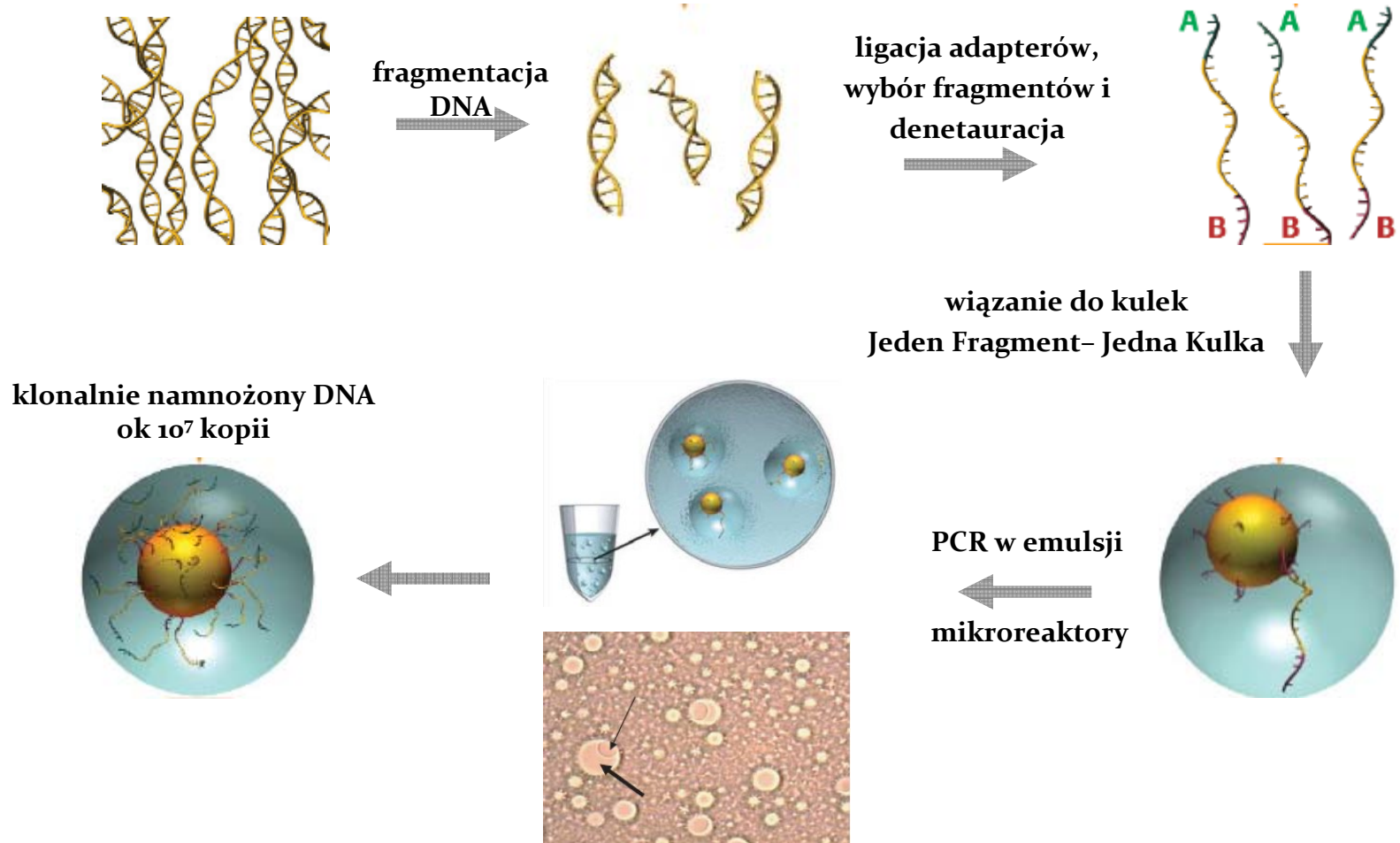


Sekwencjonowanie nowej generacji

- Różnorodne technologie umożliwiające sekwencjonowanie setek milionów lub miliardów par zasad za jednym zamachem – np. całe genomy
- Można sekwencjonować złożoną mieszaninę DNA
- Nie wymaga indywidualnego przygotowania matrycy
- Zazwyczaj stosunkowo krótkie sekwencje
- Obróbka i składanie sekwencji wymagają dedykowanych programów i dużej mocy obliczeniowej
- W praktyce od 2005 roku, cały czas testowane nowe rozwiązania

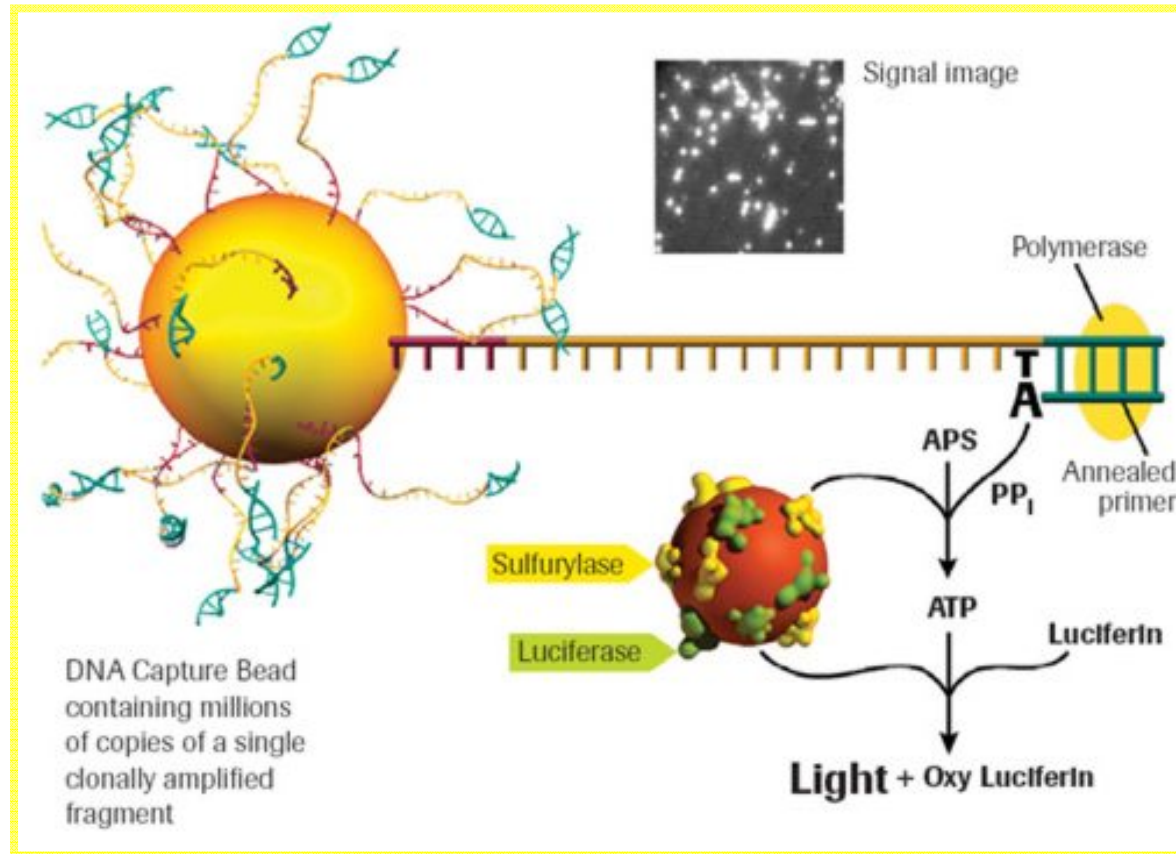
Etapy procedury

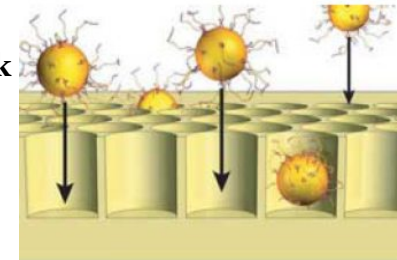
- Przygotowanie matrycy
 - fragmentacja i selekcja fragmentów
 - przyłączenie adapterów
 - selekcja jednoniciowych fragmentów
- (Amplifikacja klonalna)
 - potrzebna dla wzmocnienia sygnału
 - niepotrzebna w technikach sekwencjonowania pojedynczych cząsteczek
- Sekwencjonowanie
 - synteza nici komplementarnej przez polimerazę DNA
 - ligacja oligonukleotydów
 - różne sposoby detekcji sygnału



454

pirosekwencjonowanie





szkiełko z włókna optycznego 3.2 miliona studzienek

dodanie enzymów

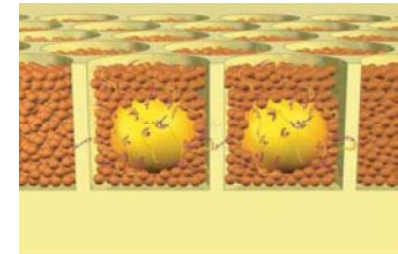
[illegible]

TCAGGTGGCATTGCCGCCTGGCCTTGTGCGA

odczyty 600 - 700 bp, ok. 1 mln sekwencji, 6-7 x 10⁶ bp

odczynniki
jeden dNTP na cykl

>300 cykli



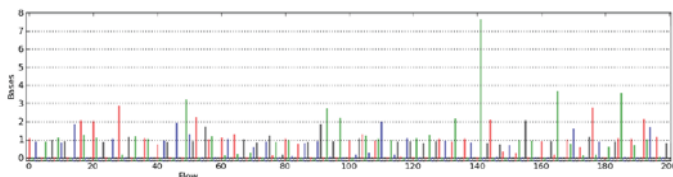
Margulies i in. 2005
Metzker 2010

- Najstarsza technika
- Długie odczyty
- Stosunkowo szybka (kilka – kilkanaście h)
- Wysoki koszt na zasadę
- Problem z regionami homopolimerowymi



Ion Torrent/Ion Proton

- Zasada podobna do 454, lecz zamiast światła, **mierzona** jest **elektronicznie zmiana pH** spowodowana przyłączeniem nukleotydu
- Zamiast płytki z włókna optycznego, **półprzewodnikowy układ scalony**
- Stosunkowo tani i skalowalny:
 - bez modyfikowanych nukleotydów
 - elektroniczna detekcja sygnału
 - nie ma potrzeby obrazowania
- Szybki postęp, 4 generacja, odczyty 200 bp, 1-10 Gb/analizę

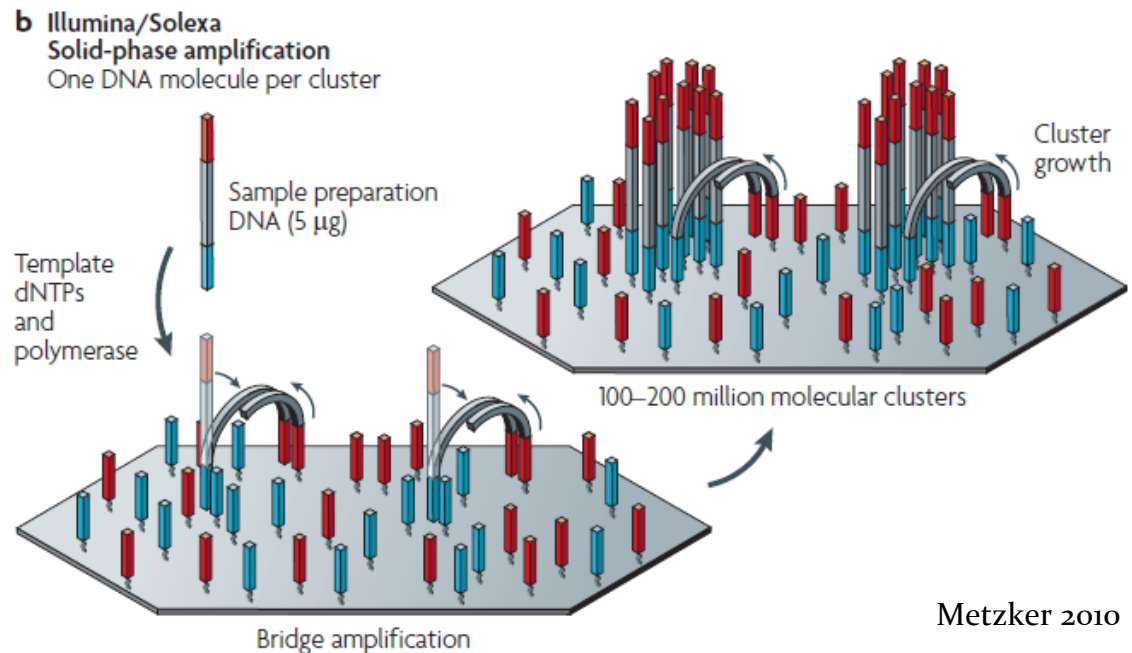


Illumina

Amplifikacja „mostkowa” na szkiełku



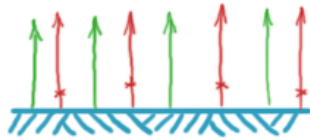
www.illumina.com



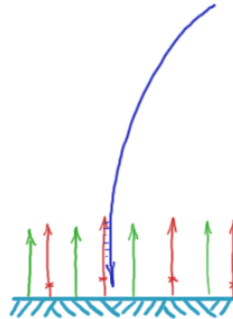
Metzker 2010

Setki milionów skupień, każde $>10^3$ identycznych (pochodzących z jednej cząsteczki) cząsteczek DNA

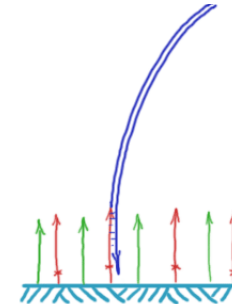
Ilumina – amplifikacja mostkowa



Startery
przytwierdzone do
szkiełka



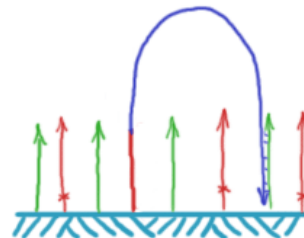
Dodanie jednoniciowej
matrycy z adapterami na
końcach



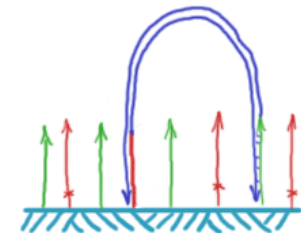
Replikacja matrycy



Denaturacja – matryca
przytwierdzona do
szkiełka

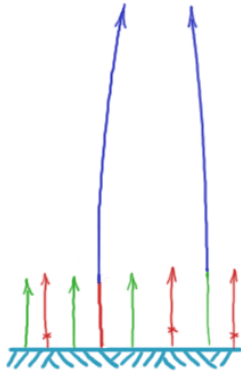


Przyłączenie do
startera – zmiana
buforu

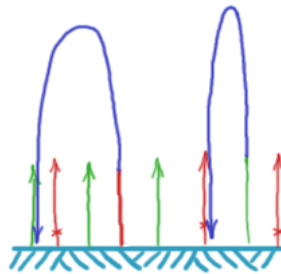


Elongacja

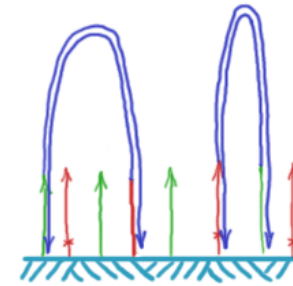
Illumina – amplifikacja mostkowa



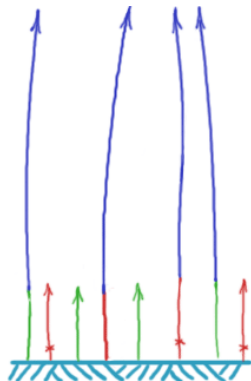
Denaturacja



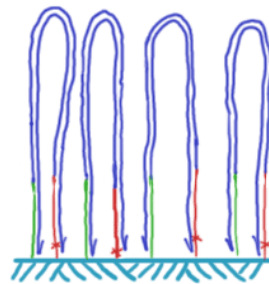
Przyłączenie do startera – zmiana buforu



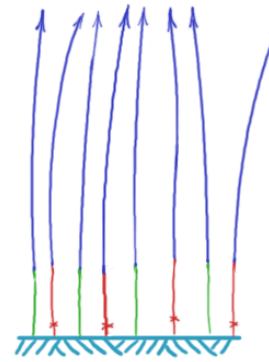
Elongacja



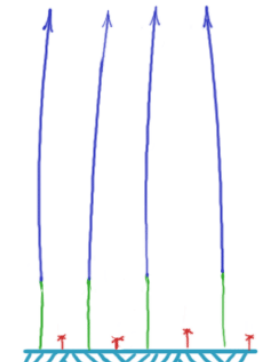
Denaturacja



Elongacja



Po 35 cyklach $>10^3$ cząsteczek



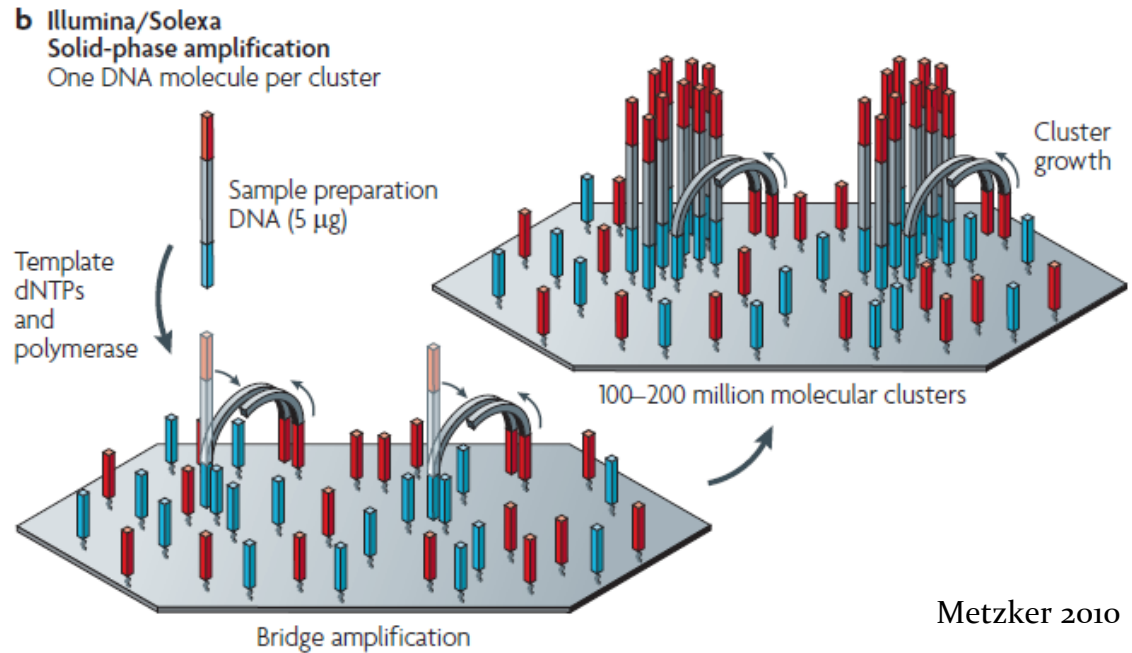
Odcięcie jednego startera – skupienie klonalnych jednoniciowych fragmentów DNA

Illumina

Amplifikacja „mostkowa” na szkiełku



www.illumina.com

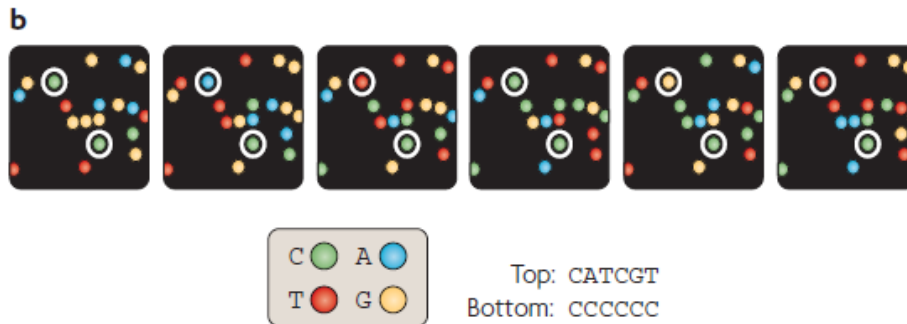


Metzker 2010

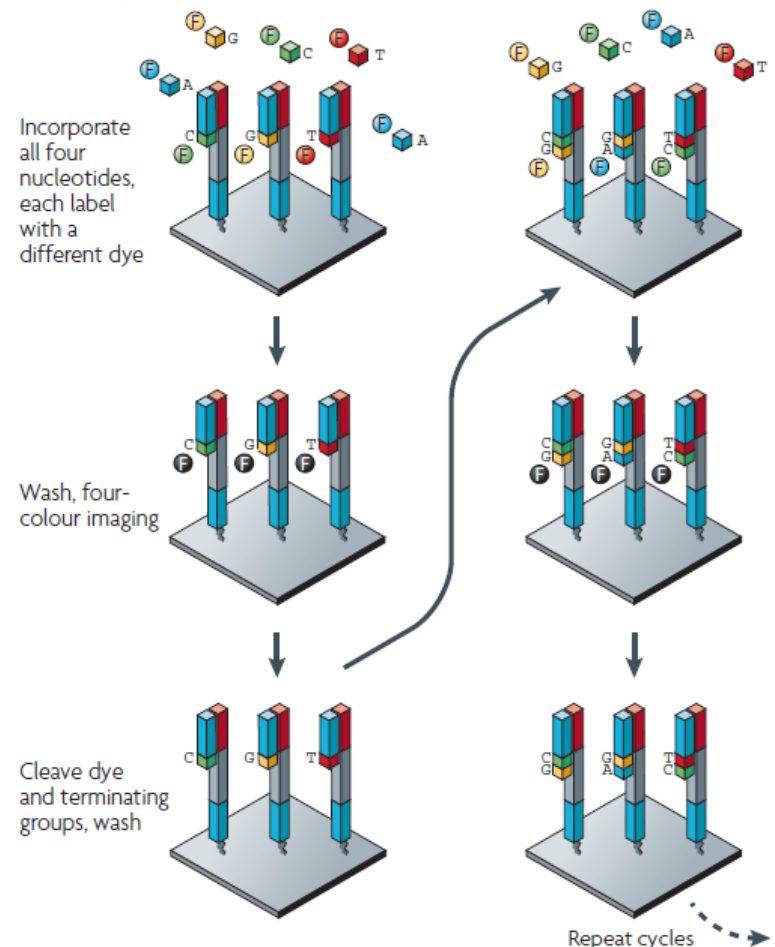
Setki milionów skupień, każdy $>10^3$ identycznych (pochodzących z jednej cząsteczki) cząsteczek DNA

Illumina – sekwencjonowanie z odwracalnymi terminatorami

- Cztery dNTP w każdym cyklu, działają jako terminator, każdy wyznakowany innym barwnikiem– elongacja o 1 pz
- Odplukanie niewykorzystanych terminatorów i obrazowanie
- Odcięcie terminatora wraz z barwnikiem – łańcuch gotowy na kolejnego kroku wydłużania

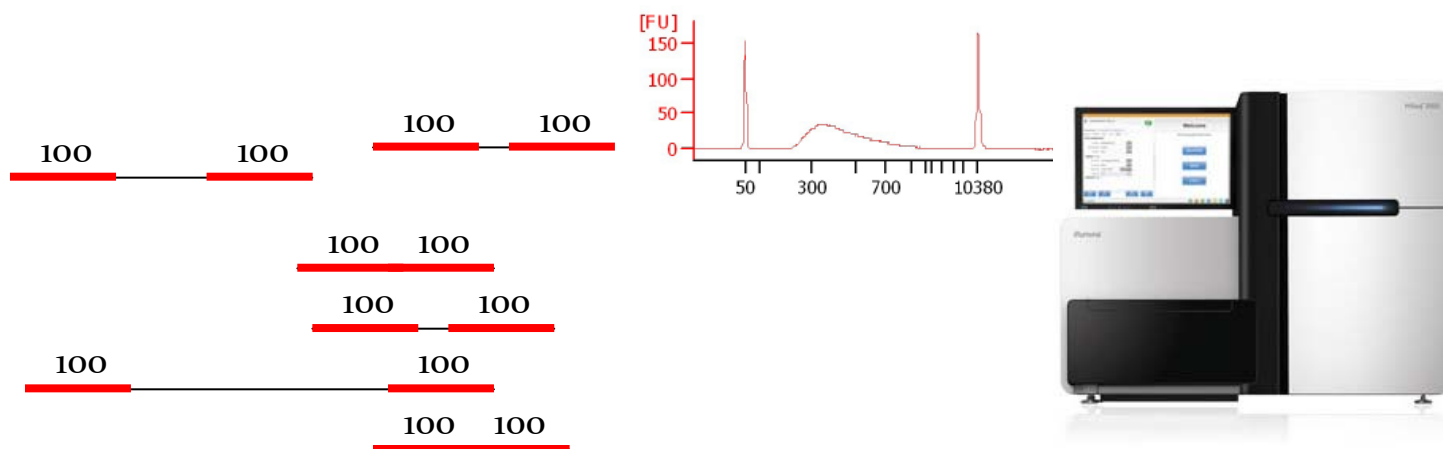


a Illumina/Solexa — Reversible terminators



Illumina

- Do 600 mld (6×10^{11}) pz w jednej analizie
- Możliwość odczytu z obu stron każdego fragmentu DNA – częściowa kompensacja krótkich odczytów, np. 2×100 pz
- 1.5 -11 dni w zależności o długości odczytu: 36 – 100 (150) pz i od tego czy sekwencjonowanie z jednego czy oby końców
- Najczęstsze błędy - podstawienia
- Najważniejsza technika NGS

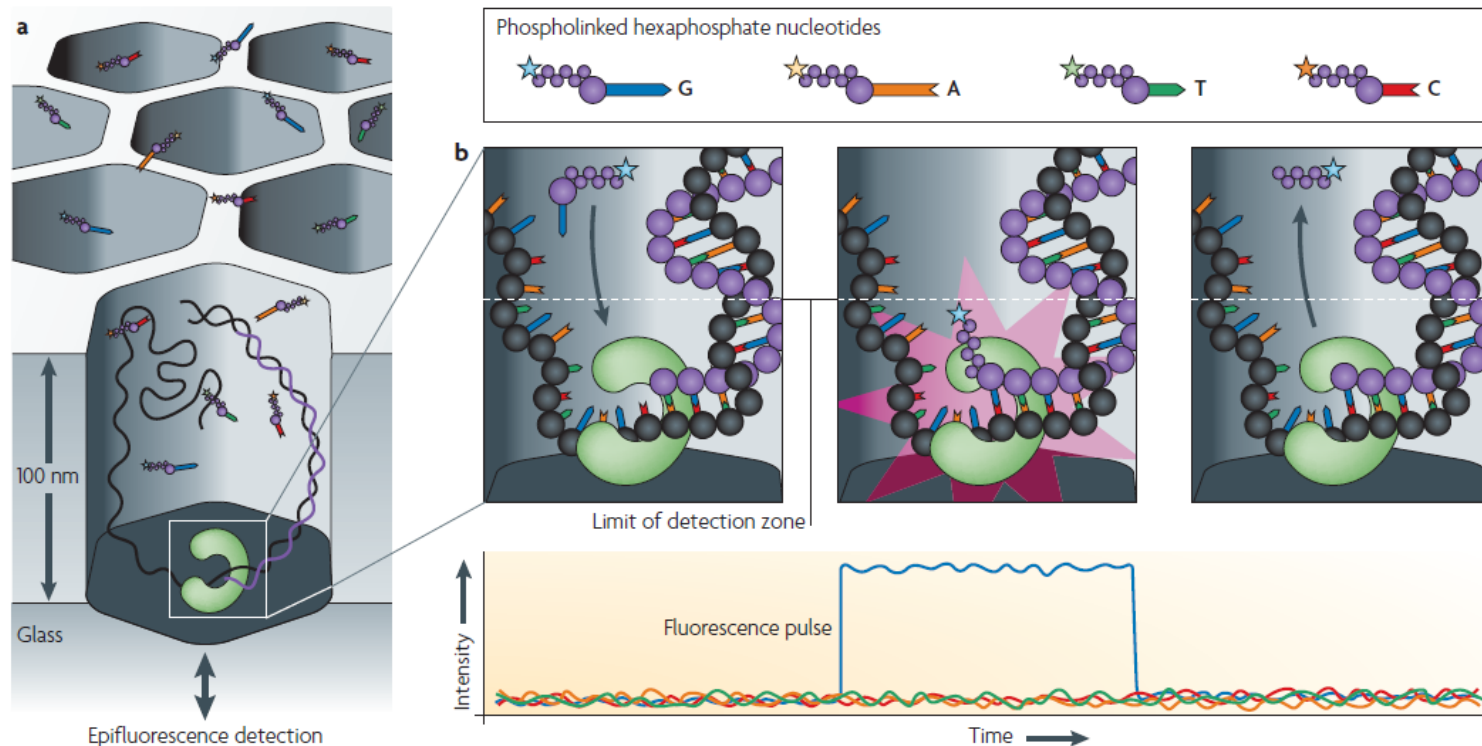


Pacific Biosciences

- Nie ma amplifikacji matrycy przed sekwencjonowaniem – **sygnał** odczytywany **z pojedynczych cząsteczek DNA**
- cząsteczki polimerazy DNA immobilizowane w studzienkach 100 nm – zero-mode waveguides (ZMV)
- Przyłączenie nukleotydu do łańcucha DNA produkuje błysk światła o określonej długości fali, który jest rejestrowany w czasie rzeczywistym – detekcja w objętości 10^{-21} l

Pacific Biosciences

Pacific Biosciences — Real-time sequencing



Pacific Biosciences

- **Długie odczyty** > 1000 (10000) pz
- Bardzo szybkie (minuty),
lecz ograniczona wydajność
- ZMV na jednorazowych chipach
(SMRT cells)
- **Wysoka częstość błędów** (ok. 15% dla pojedynczych
odczytów) – złożona analiza,
wielokrotne sekwencjonowanie,
ciągle bardzo wczesne stadium
- Złożone, wielkie i drogie
urządzenia



„Osobiste” sekwenatory

- tańsze, mniejsze wersje urządzeń dostosowane do potrzeb pojedynczych laboratoriów
- mniejsza przepustowość
- tańsze
- szybsze



GS Junior
Roche (454)



MiSeq
Illumina



Ion Torrent
Life Technologies

Porównanie wydajności

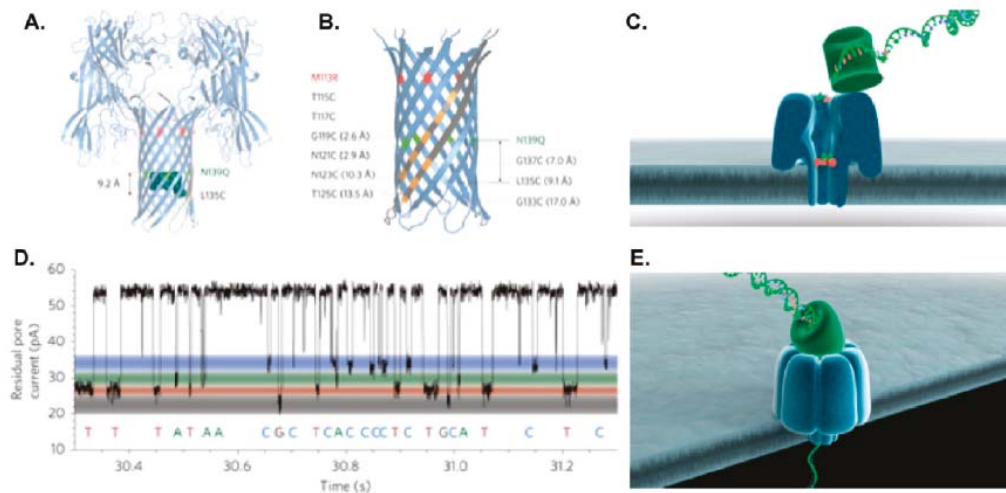
Instrument	Czas analizy [h]	Odczytów [mln]	Odczyt [bp]	Gpz na analizę
454 FLX+	20	1	650	0.65
GS Junior	10	0.1	400	0.04
Illumina HS2000	260	3000	100+100	600
MiSeq	40	15	250+250	8
SOLiD 5500XL	190	1400	75+35	155
Ion Torrent	3	4	200	0.8
Ion Proton	4	60	200	10
Pacific	0.5-2	0.01	>1000	0.01

Porównanie kosztów (USD)

Instrument	Odczynniki/ analizę	\$\$\$/Mpz	Koszt jednostkowy (% pełnej analizy)
454 FLX+	6200	7	2000 (10%)
GS Junior	1100	22	1500 (100%)
Illumina HS2000	23500	0.04	3000 (6%)
MiSeq	1160	1	1400 (100%)
SOLiD 5500XL	10500	0.07	2000 (12%)
Ion Torrent	750	1	1200 (100%)
Pacific	300-1700	7-38	500 (100%)

Przyszłość NGS?

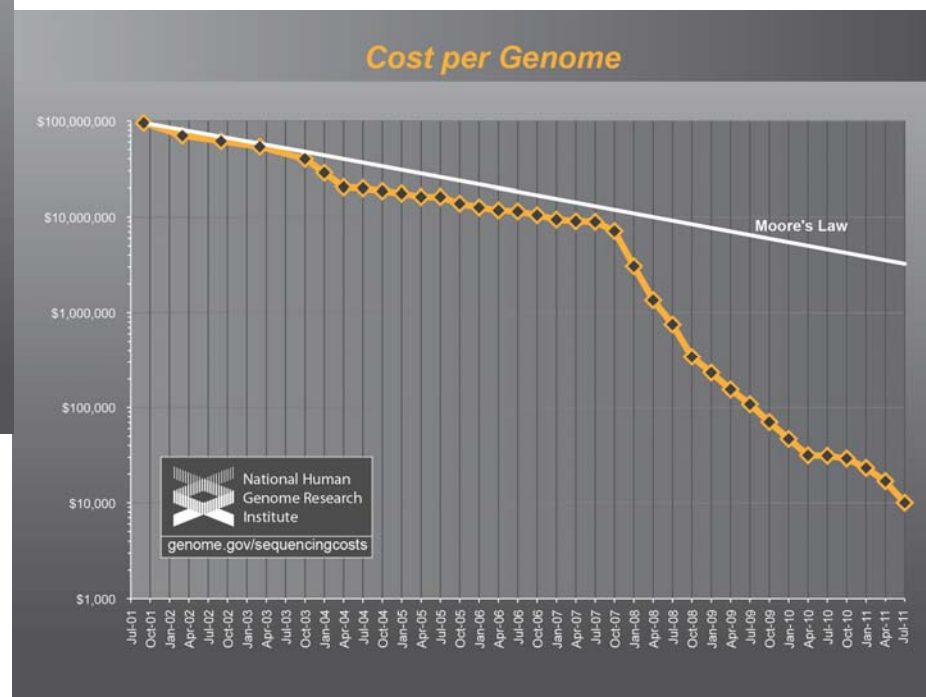
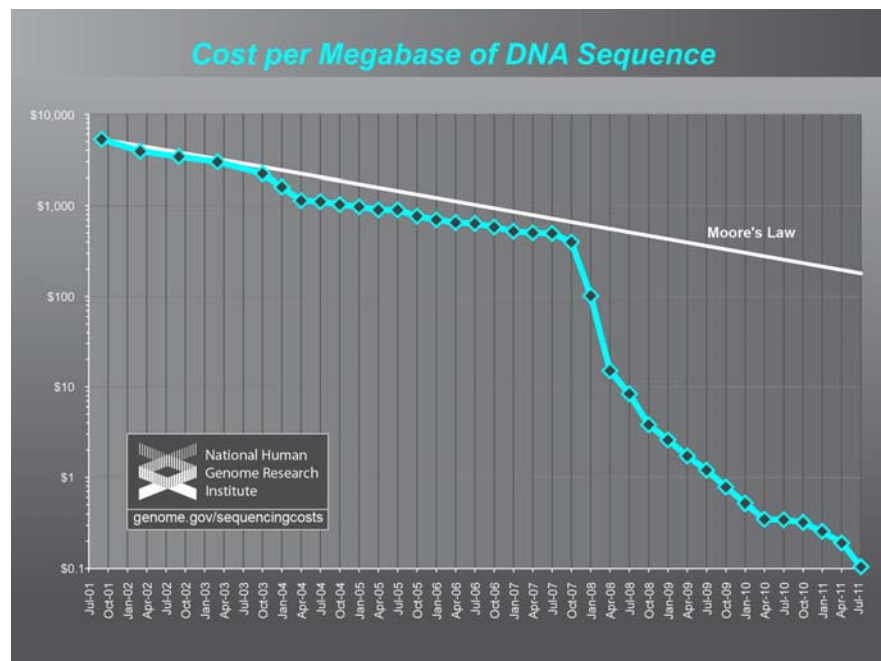
- Sekwencjonowanie długich fragmentów ułatwiłoby assembling
- Nanopory
- Mapowanie optyczne
- Wiele technologii może i będzie koegzystować



Niedringhaus i in. 2011

Koszty spadają dramatycznie

koszty sekwencjonowania
ludzkiego genomu



Powódź danych

- Ogromne ilości danych sekwencyjnych można uzyskać szybko i tanio
- Wąskim gardłem jest analiza oraz koszt składowania danych
- Świat jest pełen w minimalnym stopniu przeanalizowanych danych sekwencyjnych



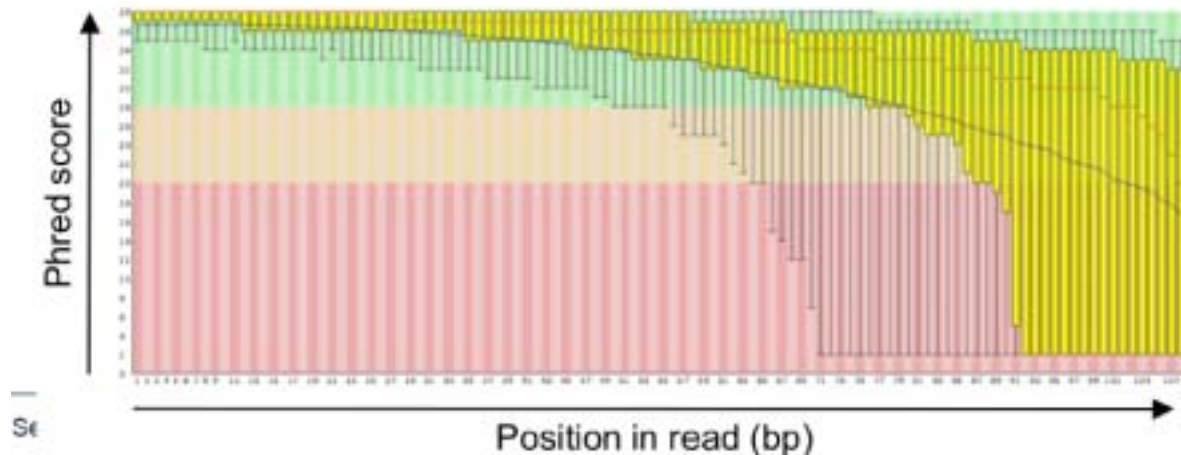
Analiza danych z NGS?

Tylko bioinformatyka

- **Sekwencjonowanie de novo** – chcemy ustalić sekwencję nukleotydów w genomie dotychczas niezsekwencjonowanego gatunku – **SKŁADANIE** (assembling)
- **Resekwencjonowanie** – mamy już sekwencję genomu danego gatunku (genom referencyjny, referencja) – **MAPOWANIE** (mapping)

Danych nie da się zobaczyć - kontrola

- Konieczne automatyczne sposoby analizy jakości sekwencji
- Basecallers – „nazywają” zasady i przypisują do nich jakość
- Phred-scores $QV = -10 \log_{10} P$; P – prawdopodobieństwo, że zasada została odczytana błędnie; 10 – 1/10 że błędnie, 20 – 1/100 że błędnie, 40 – 1/10000 że błędnie
- obcinanie fragmentów o niskich QV



Polimorfizm Pojedynczych Nukleotydów (Punktowe)

SNP – ang. Single Nucleotide Polymorphism

Polymorphism

"Poly" *many* "morpho" *form*



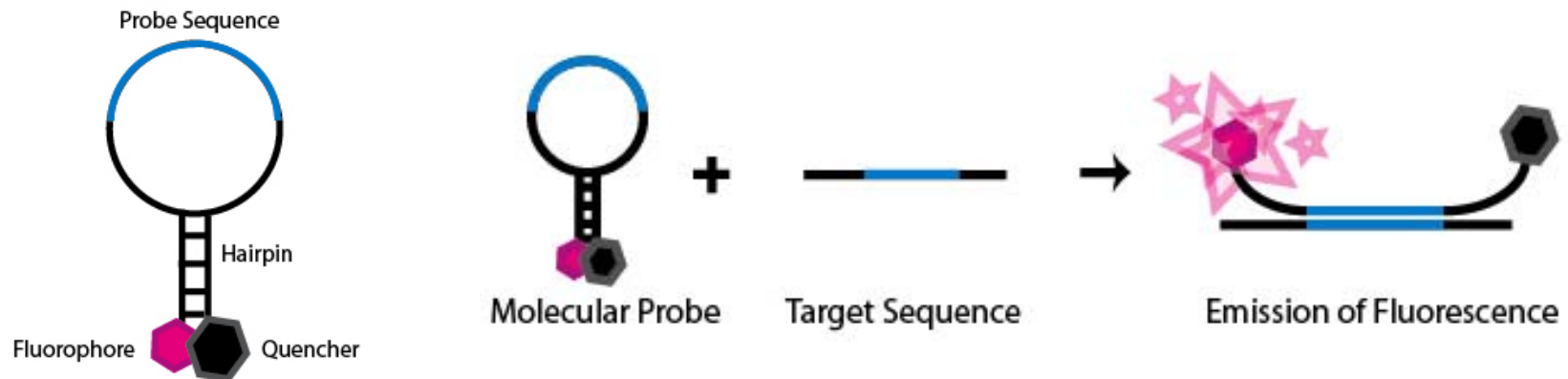
Polimorfizm Pojedynczych Nukleotydów

SNP – ang. Single Nucleotide Polymorphism

- W genomie człowieka około 10-30 mln SNPs o częstości $> 1\%$
- Mogą występować w rejonach kodujących i nie kodujących
- Wiele z nich nie ma wpływu na zdrowie ludzi, ale
- Istnieją SNP, które decydują o wystąpieniu pewnych chorób czy predyspozycji
- Wiele metod badawczych w tym mikromacierze pozwalające na jednorazowe genotypowanie nawet **milionów** SNP
- Do efektywnego genotypowania SNP niezbędna jest **znajomość kontekstu genomowego**

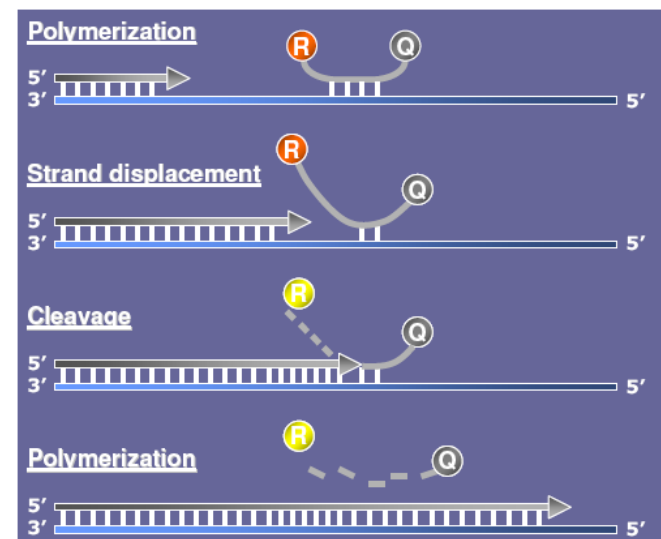
Polimorfizm Pojedynczych Nukleotydów SNP – ang. Single Nucleotide Polymorphism

Latarnie molekularne (Molecular Beacons)



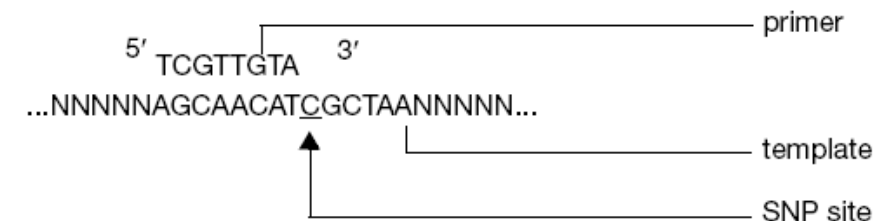
5'-nukleaza – TaqMan

Detekcja w czasie rzeczywistym
(real time PCR)

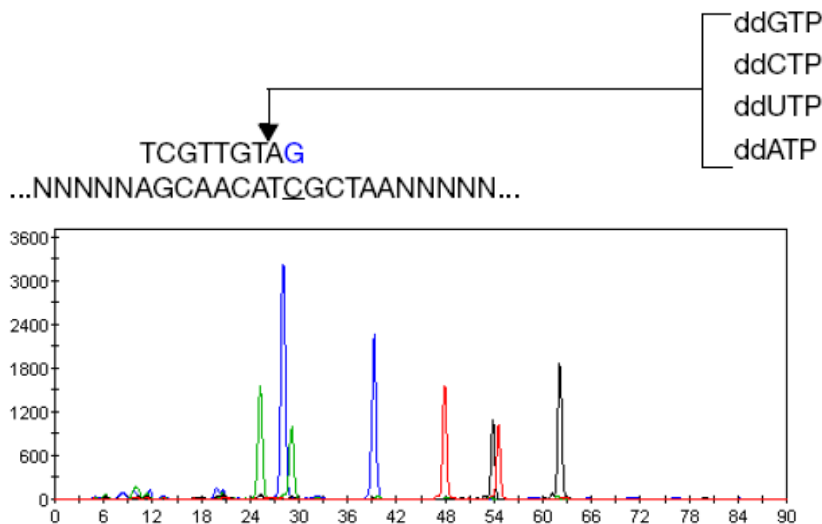


Polimorfizm Pojedynczych Nukleotydów SNP – ang. Single Nucleotide Polymorphism

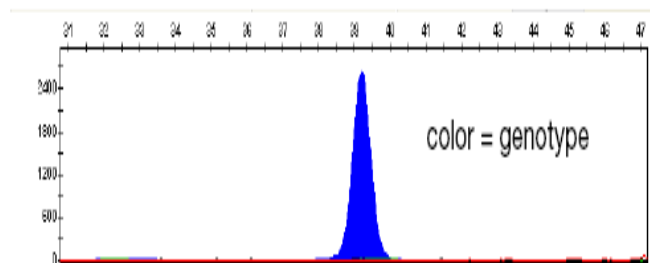
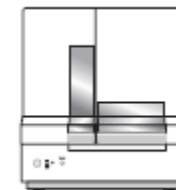
Mikrosekwencjonowanie (np. SNaPshot)



Wszystkie dodawane nukleotydy
to terminatory
Wydłużanie tylko o 1 zasadę
Elektroforeza znakowanego
produktu mikrosekwencjonowania



Multipleks

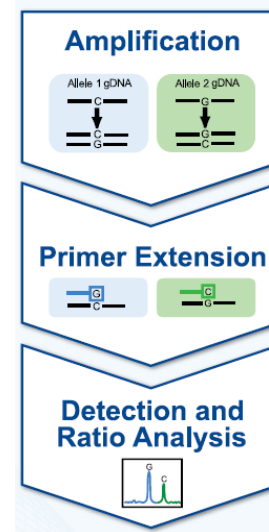


**Genotyp
homozygota G**

Polimorfizm Pojedynczych Nukleotydów SNP – ang. Single Nucleotide Polymorphism

Mikrosekwencjonowanie (np. MassArray)

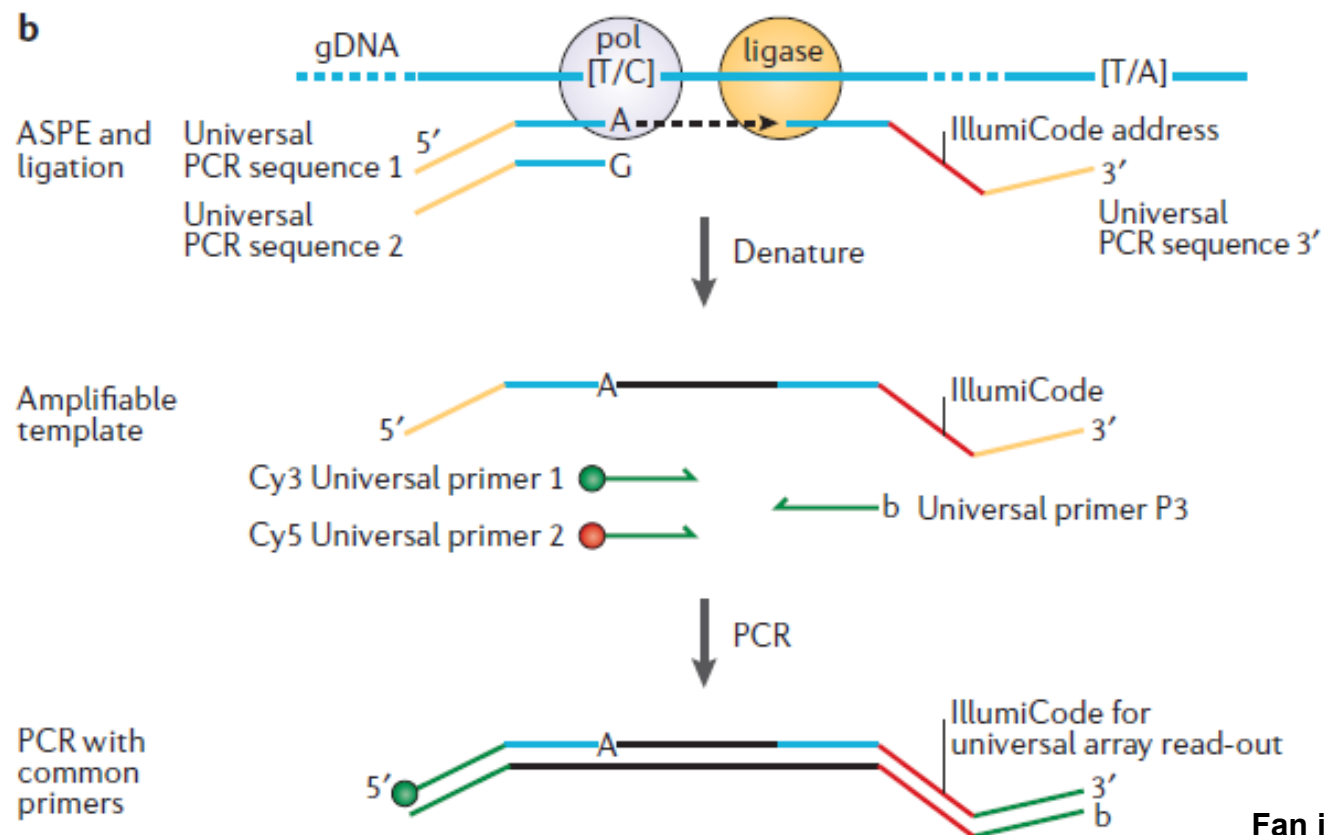
- wszystkie dodawane nukleotydy to terminatory
- wydłużanie tylko o 1 zasadę
- w jednej reakcji namnażane i mikrosekwencjonowane do 36 SNP
- masa produktów mikrosekwencjonowania oznaczana spektrometrią masową
 - dla każdego SNP znamy dokładnie masę możliwych produktów reakcji
 - sprawdzamy jakie masy mamy w naszej mieszaninie -> odczytujemy genotypy



Polimorfizm Pojedynczych Nukleotydów

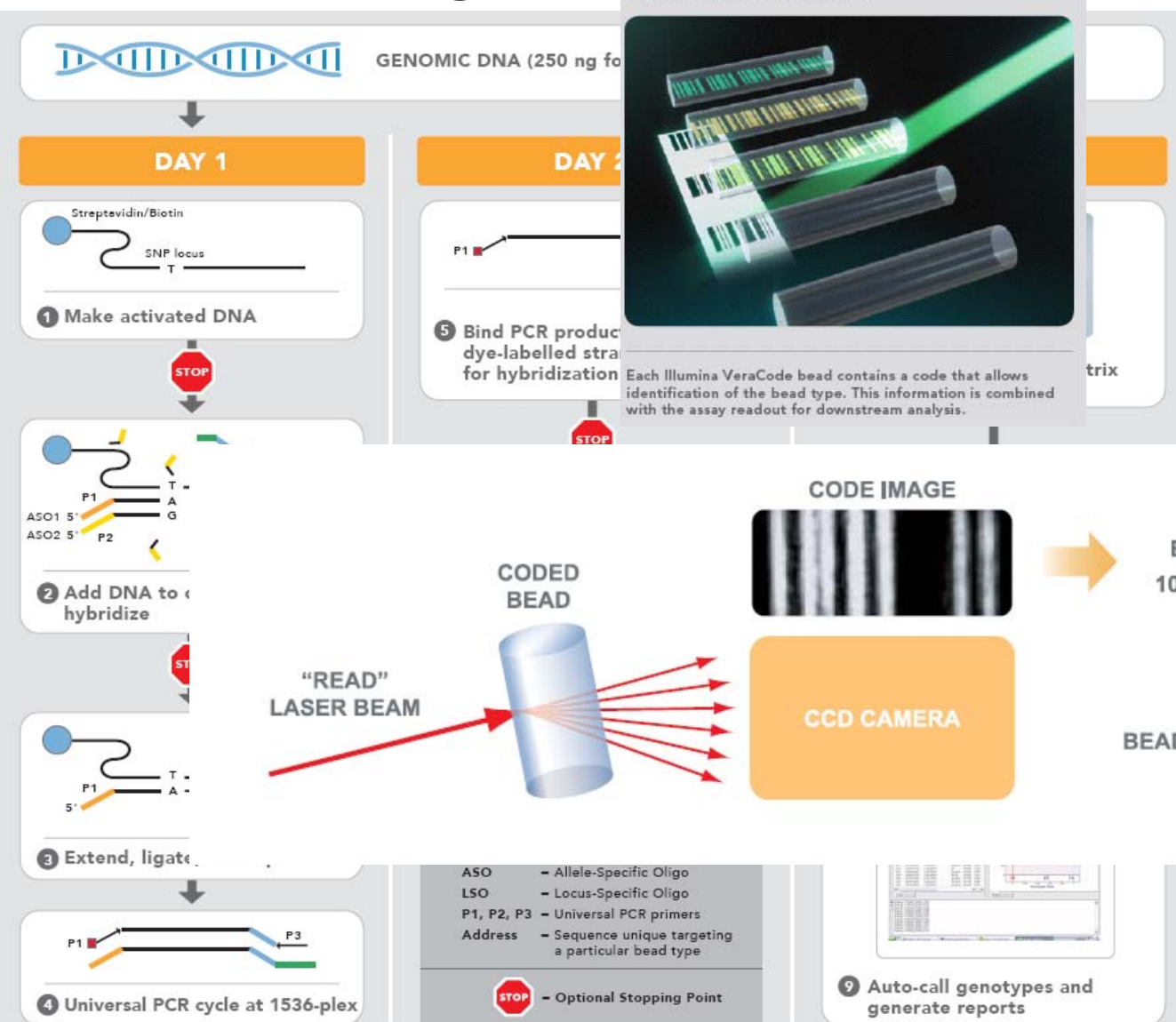
SNP – ang. Single Nucleotide Polymorphism

Golden Gate - genotypowanie setek SNP w jednej reakcji



Polimorfizm Pojedynczych Nukleotydów

SNP – ang. Single Nucleotide Polymorphism

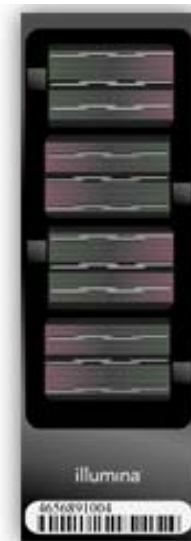
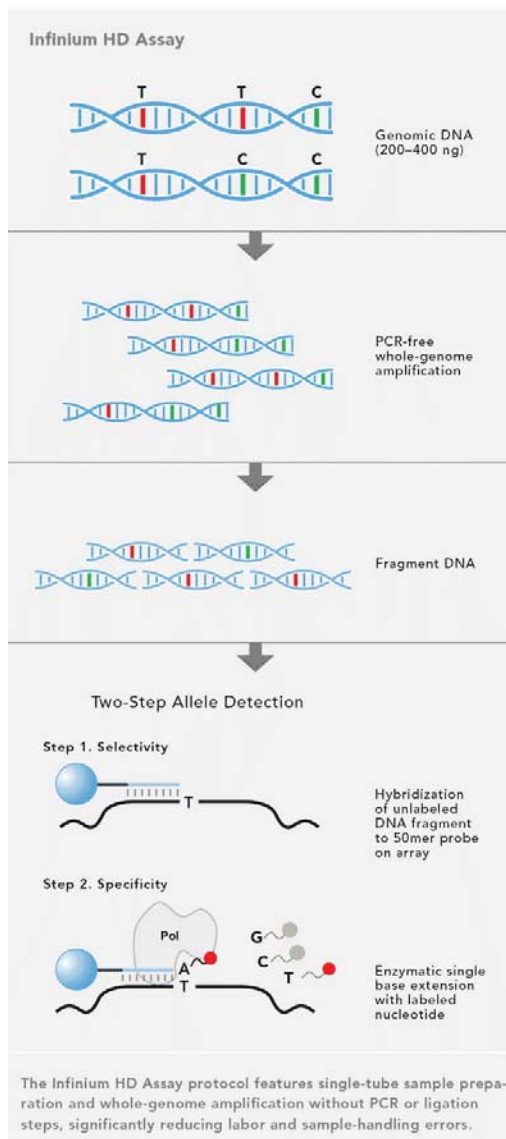


Golden Gate
genotypowanie setek
SNP w jednej reakcji

Polimorfizm Pojedynczych Nukleotydów

SNP – ang. Single Nucleotide Polymorphism

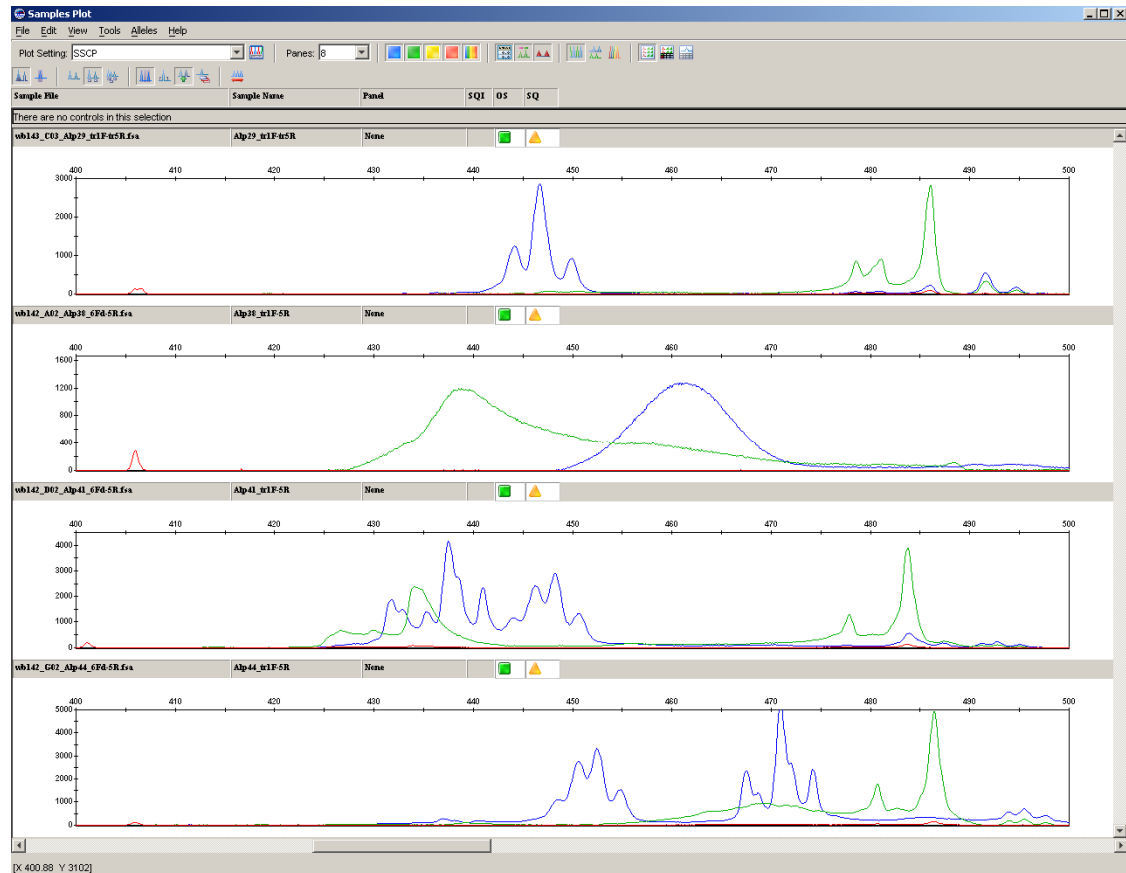
Dostępne komercyjnie mikromacierze (np. Affymetrix, Illumina) pozwalają na genotypowanie milionów SNP u organizmów modelowych i człowieka



SSCP (Single Strand Conformational Polymorphism)

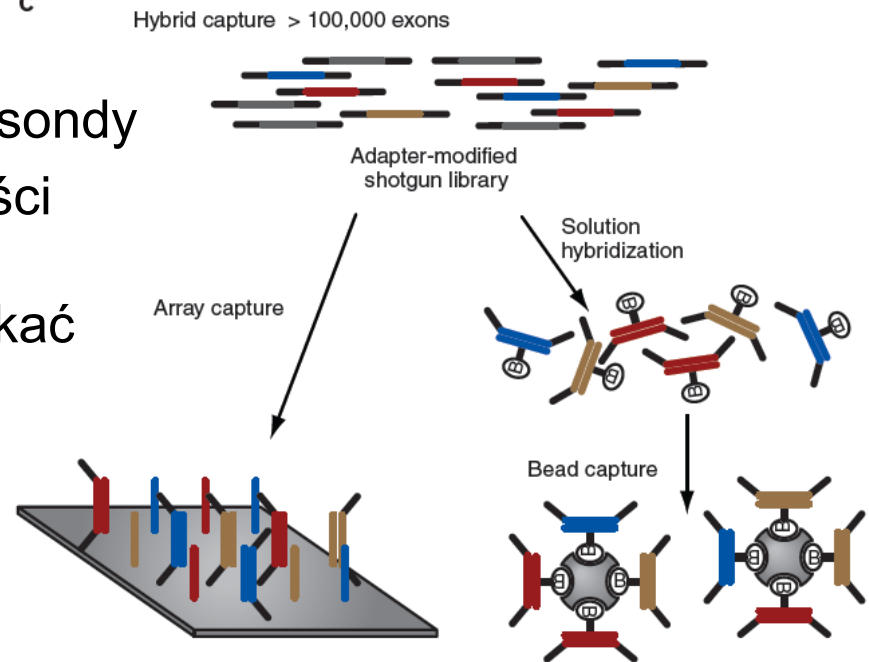
SSCP wykorzystuje się do wykrywania mutacji i genotypowania fragmentów określonych genów (namnożonych PCR) o jednakowej długości lecz różniących się sekwencją

Jednoniciowe fragmenty DNA
100-400 pz zwijają się w
sposób zależny od sekwencji, a
od sposobu zwinięcia zależy
tempo ich wędrówki w żelu
Można wykryć różnicę 1 pz



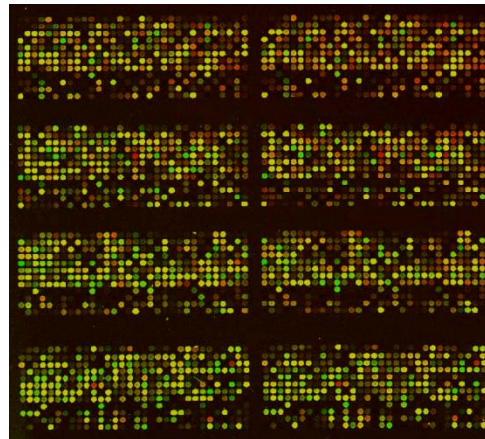
Wyciąganie z genomu interesujących nas fragmentów

- Sekwencjonowanie pełnych genomów ciągle zbyt drogie
- Analiza i składowanie danych drogie
- Często interesują nas tylko niektóre geny, lub części genomu, np. wszystkie egzony stanowią ok. 1.2% genomu człowieka
- Można je wszystkie wyciągnąć, c „odsiać” od reszty genomu stosując macierze zawierające sondy
- Te techniki wymagają znajomości sekwencji genomu lub tej jego części, którą chcemy uzyskać
- Następnie uzyskane fragmenty można zsekwencjonować



Mikromacierze DNA (microarrays, DNA chips)

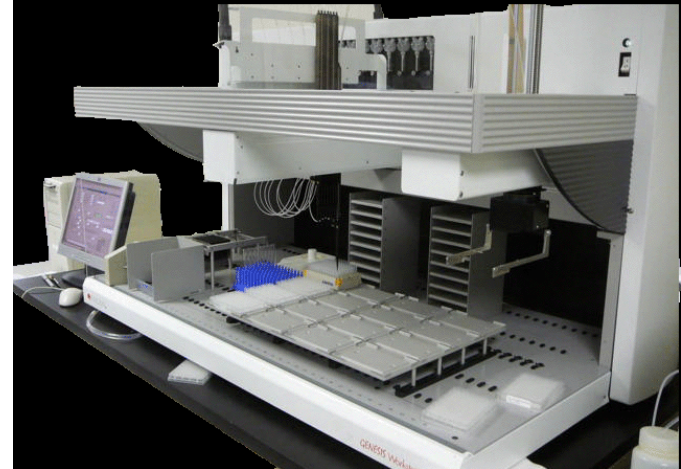
- Opierają się na hybrydyzacji znakowanego fluorescencyjnie DNA lub RNA do fragmentów (sond) unieruchomionych na szklanej płytce
- Bardzo duża gęstość, nawet miliony różnych sond na jednym szkiełku
- Początkowo stosowane do analizy ekspresji genów
- Obecnie wiele innych zastosowań, np. genotypowanie SNP



animacja

Techniki biologii molekularnej są niemal w pełni zautomatyzowane

Możliwe jest uzyskanie w stosunkowo krótkim czasie wyników dla setek osobników i kilkuset markerów molekularnych.



Użyteczność markerów molekularnych

	allozymy	AFLP	mikrosatelity	SNP	sekwencje genów jądrowych	sekwencje cpDNA i mtDNA
Identyfikacja osobników	+	++	+++	+	-	+
Zmienność genetyczna	++	+	+++	++	+	-
Zróżnicowanie genetyczne	++	+	+++	+++	++	++
Przepływ genów (migracje)	+	+	+++	++	+	+
Filogeografia	+	+	+	++	+++	+++
Hybrydyzacja	++	++	++	+++	+++	+++

U eukariotów występują różne genomy

	nDNA	cpDNA	mtDNA(Z)	mtDNA(R)
Dziedz.	♀	♀ (O), ♂ (N)	♀	♀
Struktura	liniowy	kolisty	Kolisty	Kolisty
Tempo mutacji	8	2	130	1

W zależności od problemu badawczego wykorzystuje się markery molekularne ze wszystkich genomów

Zagadnienia (1)

- Identyfikacja
 - płci
 - osobników (także ze śladów)
 - rodzicielstwa
 - gatunków
- Zmienność genetyczna i czynniki na nią wpływające
 - miary zmienności
 - zmienność a procesy populacyjne
- Przepływ genów i struktura genetyczna populacji
 - metody genetyczne badania dyspersji
 - miary struktury genetycznej i ich znaczenie biologiczne

Zagadnienia (2)

- Genealogie genów i wnioskowanie z genealogii
 - historia demograficzna
 - przepływ genów między populacjami
- Filogeografia
- Genetyka konserwatorska – wykorzystanie markerów molekularnych dla ochrony gatunków
- Wpływ organizmów modyfikowanych genetycznie na ekosystemy
- Badanie adaptacji na poziomie molekularnym

Identyfikacja płci

Konieczna w przypadku:

- braku dymorfizmu
- do określenia proporcji płci w potomstwie
- osobników młodocianych, larw, etc.
- analiz z odchodów i pozostałości
- u zwierząt prowadzących skryty tryb życia

Markery specyficzne dla płci nie są uniwersalne

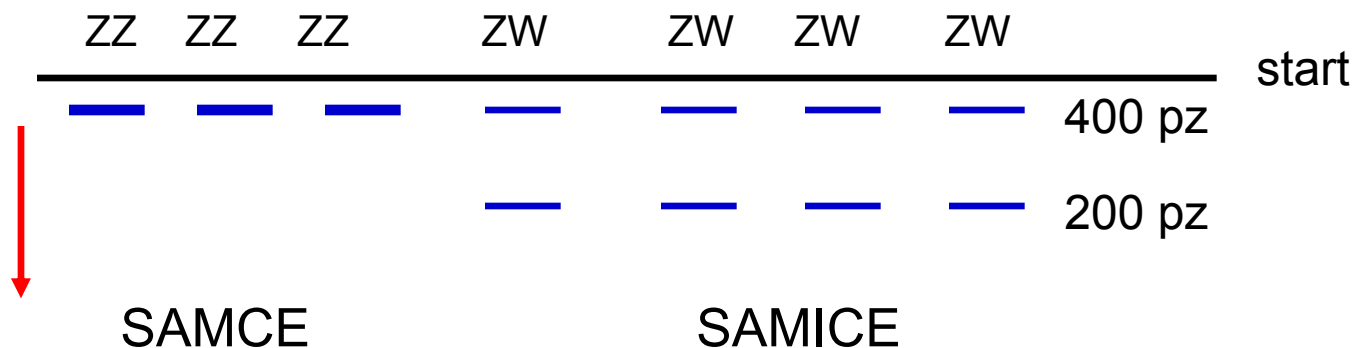
Identyfikacja płci

Ptaki

- genetyczna identyfikacja płci osobników przy braku dymorfizmu płciowego
- wyznaczanie pierwotnej proporcji płci (dla zarodków w jajach)

Geny CHD

Fragment Intronu A genu CHD1-Z i CHD1-W



Identyfikacja płci

Ssaki

- genetyczna identyfikacja płci osobników z pozostawionych śladów biologicznych (np. odchody)
- przyżyciowe określanie płci osobników młodocianych

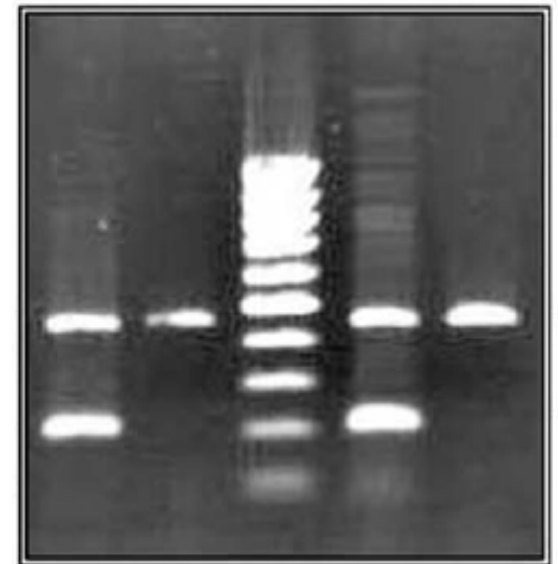
Gen SRY

Mikrosatelity
na chromosomie Y

gen występujący u obu płci
służy jako kontrola pozytywna
(że sam PCR się udał)

Zfy/Zfx
Sry

Bryja i Konecny 2003



♂

♀

♂

♀

Identyfikacja osobników

Nawet **mała liczba polimorficznych loci** tworzy **dużą liczbę możliwych genotypów**

$$G = \frac{n^2(n+1)^2}{4} \quad (\text{dla 2 loci})$$

n = liczba alleli; dla $n = 4$ alleli, $G = 100$

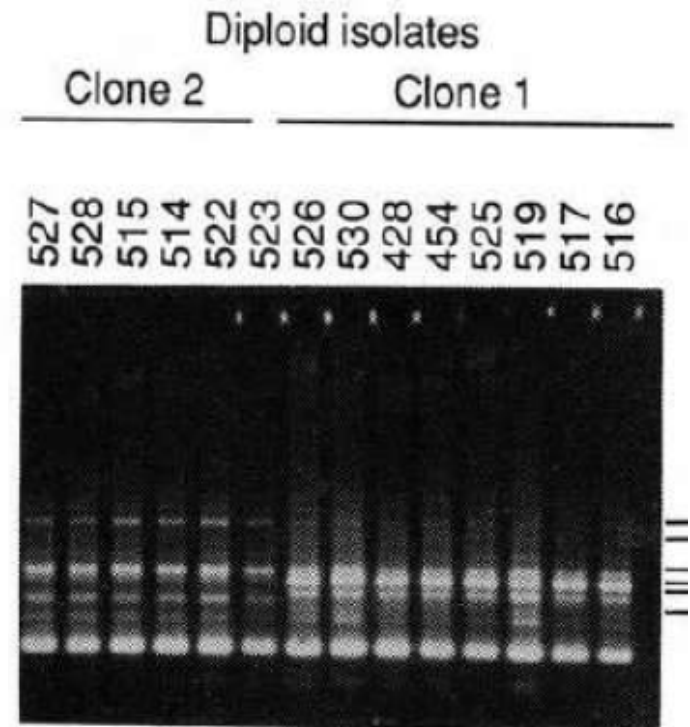
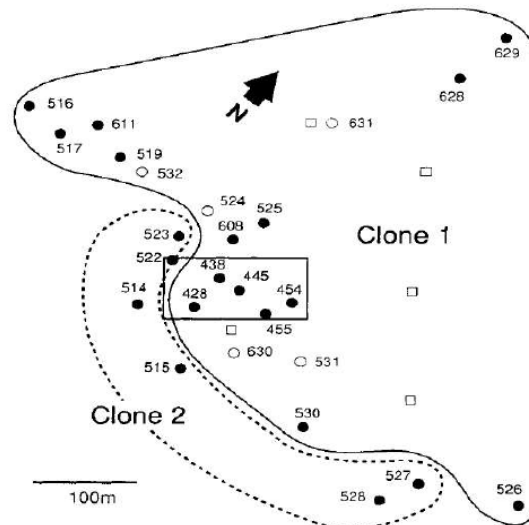
liczba możliwych genotypów dla n
loci z dwoma allelami każdy to 3^n

**polimorficzne markery molekularne są idealne
dla identyfikacji osobników**

Identyfikacja osobników

Rozróżnianie osobników u organizmów klonalnych

opieńka *Armillaria bulbosa*,
jeden klon może
zajmować 15 ha, ważyć
10 ton i mieć 1500 lat;
analiza oparta na
markerach RAPD



Smith i in. 1992

Behawioralny a genetyczny system kojarzeń

Behawioralny system kojarzeń – opis na podstawie obserwacji, które osobniki łączą się w pary i kojarzą się

Genetyczny system kojarzeń – opis na podstawie danych genetycznych, które osobniki się rozmnażają i z kim wydają potomstwo

?

Behawioralny system kojarzeń = Genetyczny system kojarzeń

Niezbędna jest genetyczna identyfikacja rodziców

Analiza rodzicielstwa

Kilka podejść, najprostsze i najczęściej używane

- Wykluczanie

opiera się na wykorzystaniu zasad dziedziczenia mendlowskiego – u diploidów rodzic ma dokładnie jeden wspólny allel z potomkiem w każdym locus autosomalnym

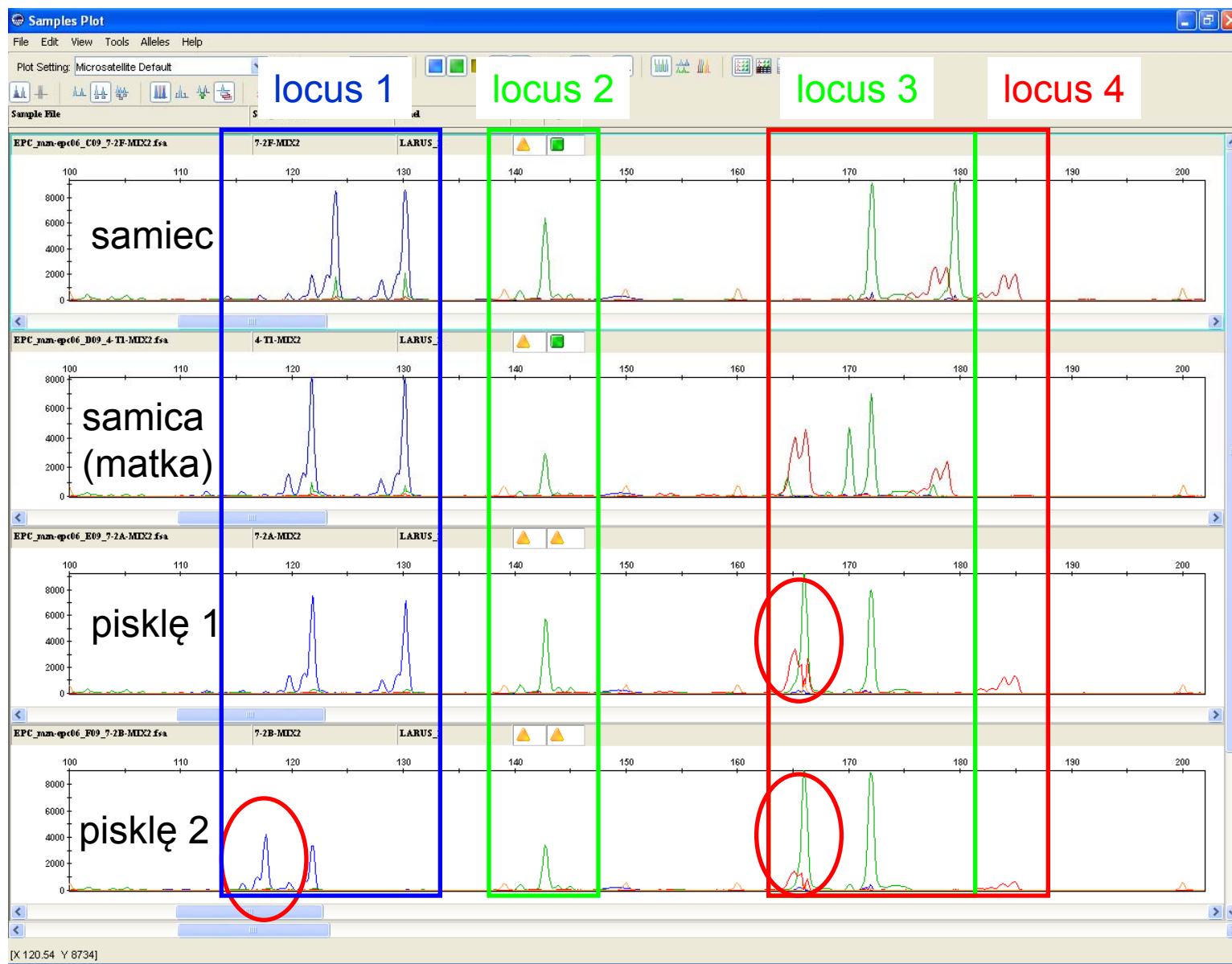
- Identyfikacja kategoryczna

Dla każdego potencjalnego rodzica (z puli niewykluczonych) oblicza się prawdopodobieństwo że jest on rodzicem i za rodzica uznaje się tego o najwyższym prawdopodobieństwie; prawdopodobieństwo rodzicielstwa zależy od genotypu

Identyfikacja najłatwiejsza gdy znamy genotyp jednego z rodziców (np. owoce zebrane z rośliny)

Można rekonstruować genotypy rodziców dysponując tylko potomstwem

Wykluczenie rodzicielstwa



Analiza rodzicielstwa

- **najlepsze** markery to wysoce zmienne **mikrosatelity**, czasem SNP, wyjątkowo AFLP lub allozymy
- markery powinny być wysoce zmienne
- ich liczba powinna być znaczna i zależy od poziomu zmienności poszczególnych markerów, 5-30 mikrosatelitów ale > 50 SNP
- genotypowanie musi być wiarygodne, należy oszacować błąd genotypowania i dążyć aby był on niski
- loci powinny być niesprężone
- kluczowe jest odpowiednie pobranie prób w terenie, np. jak największa część potencjalnych rodziców
- trzeba brać pod uwagę właściwości biologiczne organizmów (np. wielkość miotu)
- liczne programy komputerowe: Cervus, FAP, Pasos, Colony itd.

Rodzicielstwo poza parą

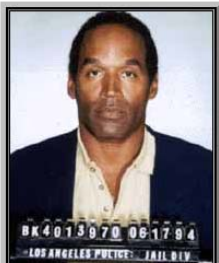
- występuje u 90% socjalnie monogamicznych (tworzących pary) gatunków ptaków
- u tych gatunków średnio 11% potomstwa pochodzi z kojarzeń poza parą – duża zmienność, różne czynniki odpowiedzialne
- potrzos (*Emberiza schoeniclus*)
 - socjalny partner samicy nie był ojcem średnio 55 % piskląt,
 - 86% gniazd zawierało młode z kojarzenia poza parą
- od połowy lat 80tych minisatelity
- obecnie mikrosatelity: kilka, kilkanaście
- wysoka skuteczność



Identyfikacja osobnicza u człowieka

- Sprawy kryminalne
- Identyfikacja szczątków ludzkich (osoby zaginione, ofiary katastrof)
- Analiza spornego ojcostwa

The O.J. Simpson Case



Sprawdź, kto jest biologicznym ojcem!
Może ojcem jesteś ty?!
A może właśnie nim nie jesteś?!



Wybór markera

- Markery stosowane do identyfikacji człowieka
 - markery DNA jądrowego
 - mikrosatelity (STR) – chromosomy autosomalne
 - mikrosatelity (STR) – chromosomy płci (Y-STR, X-STR)
 - markery DNA mitochondrialnego (analiza włosów i kości)
 - HV1, HV2
- Materiał porównawczy w sprawie decyduje o wyborze markerów i wartości dowodu z badania DNA
 - próbka pobrana od podejrzanego/ofiary
 - przedmiot osobisty należący do zaginionego (np. szczoteczka do zębów)
 - próbki od ojca i matki (wszystkie markery)
 - tylko od matki (DNA jądrowy + mtDNA)
 - tylko od ojca (DNA jądrowy w tym dodatkowo Y-STR)
 - od dalszych krewnych - w lini matczynej, ojcowskiej
 - od syna, córki
- Badania materiału pochodzenia zwierzęcego
 - cytochrom b, mikrosatelity

STR = mikrosatelity

Identyfikacja grobu Mikołaja Kopernika

- W katedrze we Fromborku znaleziono kości
- Na podstawie czaszki zrekonstruowano twarz
- Uzyskano profil DNA z kości i zębów
- Profil był identyczny z profilem włosów znalezionych w książce która była własnością Kopernika



Identyfikacja gatunków i ich mieszańców

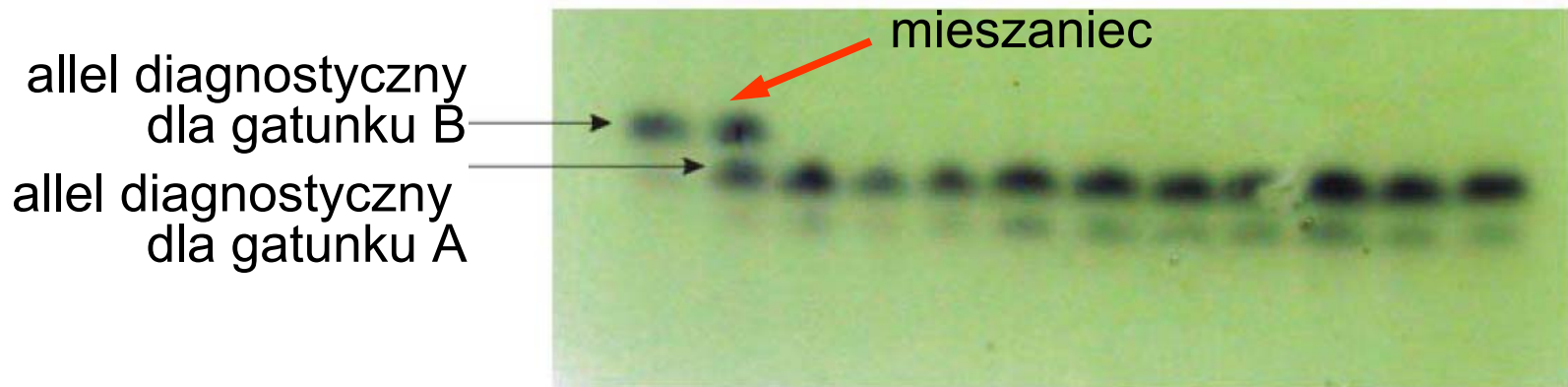
**Nie ma (i nie będzie?)
uniwersalnej definicji gatunku**

Biologiczna koncepcja gatunku vs. inne definicje
np. genetyczna koncepcja gatunku
(odrębność pul genowych)

Istnieje spora grupa gatunków nierozróżnialnych
morfologicznie (ang. sibling, cryptic species), a w
przypadku mikroorganizmów jest to problem
zasadniczy

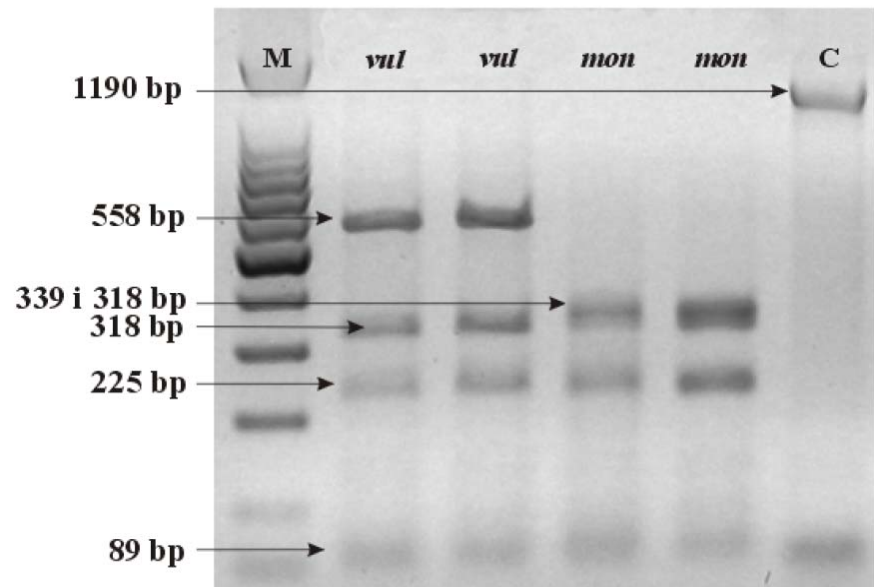
Identyfikacja gatunków i ich mieszańców

- poszczególne gatunki wykazują unikatowe cechy, które mają podstawy genetyczne -> **gatunki różnią się genetycznie**
- **markery diagnostyczne** – takie, w których dany gatunek ma unikatowy allel (allele)
- markery diagnostyczne pozwalają na łatwą identyfikację mieszańców międzygatunkowych



Identyfikacja gatunków i ich mieszańców

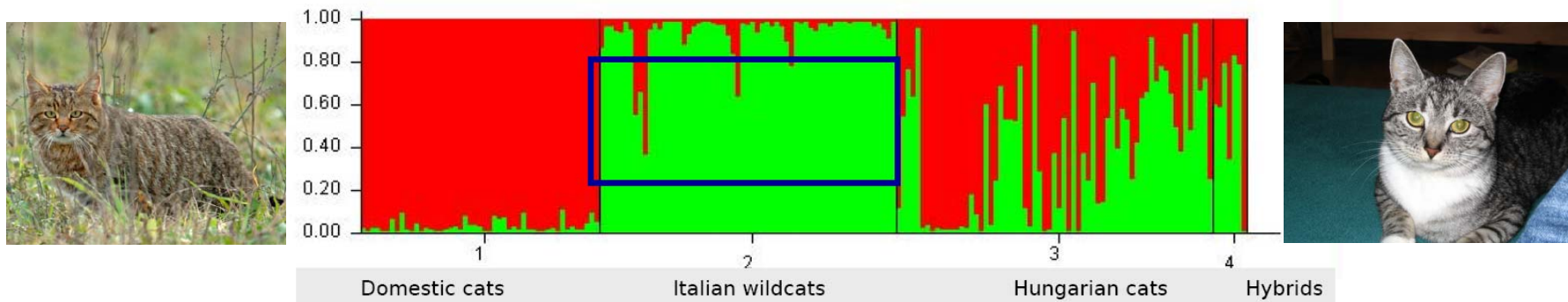
- jeżeli mieszańce są płodne i krzyżują się z gatunkami rodzicielskimi, mogą być heterozygotyczne tylko w niektórych genach markerowych
- **im więcej markerów diagnostycznych tym lepiej**; tym dalsze pokolenia mieszańców = osobniki zawierające mniejszą proporcję genów jednego z gatunków jesteśmy w stanie wykryć
- markery przekazywane tylko przez jedną płęć pozwalają określić **kierunek hybrydyzacji**, np. ♀A x ♂B ale nie ♀B x ♂A
mtDNA



Identyfikacja gatunków i ich mieszańców

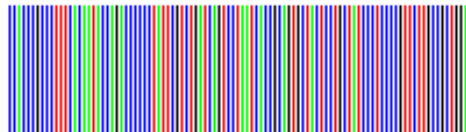
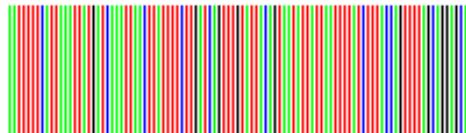
Metody pozwalające na szacowanie, na podstawie genotypu określonego dla wielu części genomu (np. wielu mikrosatelitów), jaki procent genomu danego osobnika pochodzi od każdego z gatunków rodzicielskich

Oparte na modelach probabilistycznych i twierdzeniu Bayesa – np. STRUCTURE



Kody kreskowe DNA

- Standaryzowana procedura identyfikacji gatunków poprzez krótkie ($100 < \text{dł.} < 1000$ pz), unikatowe sekwencje DNA
- Odpowiedni gen
 - konserwatywne startery
 - wystarczająca zmienność
 - występuje u wszystkich gatunków w dużej grupie taksonomicznej (bakterie, archeowce, rośliny, zwierzęta, grzyby)



Kody kreskowe DNA

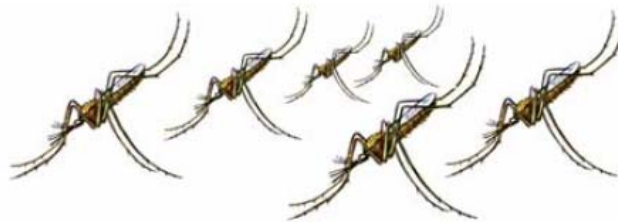
Gene ^a	Genomic location	Number of sequences			
		Animals	Plants	Protists	Fungi
COI-barcode ^b	Mitochondria	195 777	520	1931	410
16S-rDNA	Mitochondria	41 381	221	2059	285
<i>cytb</i>	Mitochondria	88 324	165	1920	1084
ITS1-rDNA	Nucleus	12 175	57 693	68 839	56 675
ITS2-rDNA	Nucleus	13 923	58 065	67 332	56 349
18S-rDNA	Nucleus	21 063	17 121	32 290	33 327
<i>rbcL</i>	Plastid	NA ^c	30 663	37 328	NA

^aGene abbreviations: COI, cytochrome *c* oxidase I; *cytb*, cytochrome *b*; ITS, internal transcribed spacer; *rbcL*, large subunit of ribulose 1,5-bisphosphate carboxylase/oxygenase.

^bCOI-barcode statistics are retrieved from Barcode of Life Data systems (<http://www.barcodinglife.org>). Statistics for other loci are retrieved from GenBank.

^cNA, not applicable.

Hajibabei i in. 2007



- Tworzy się bazę sekwencji z rozpoznanych taksonomicznie gatunków
- Porównuje się sekwencje z próbek z bazą
- W perspektywie automatyczne „czytniki”



Illustration: a person using a handheld DNA sequencer device to analyze a sample.

Kody kreskowe DNA

CONSORTIUM FOR THE BARCODE OF LIFE

<http://barcoding.si.edu>

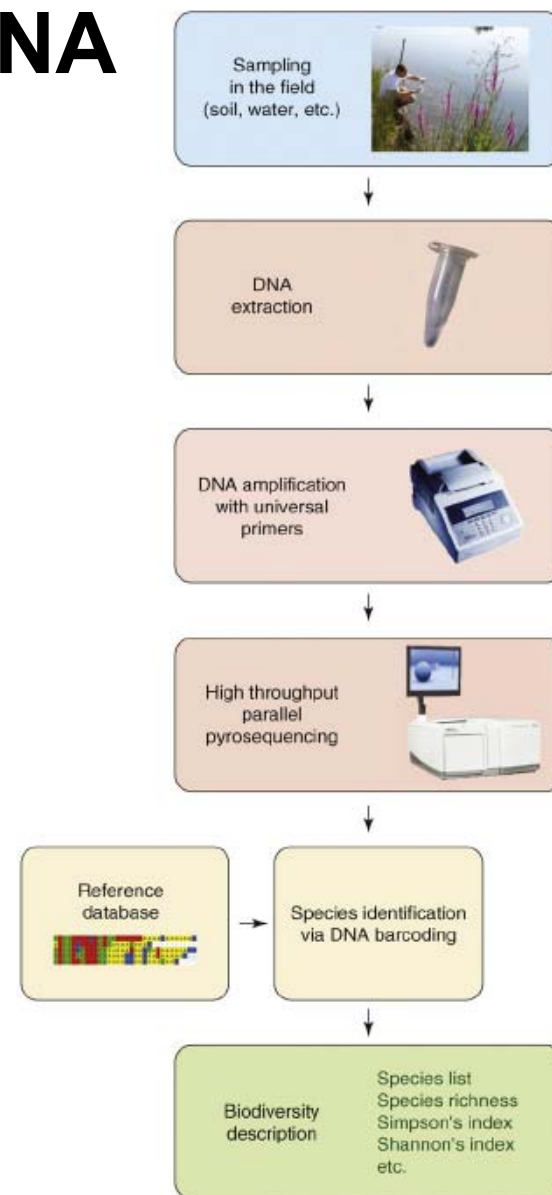
<http://www.barcodinglife.org>



- Mimo początkowych wątpliwości udowodniono dużą przydatność w wielu grupach taksonomicznych (skuteczność rzędu 95 %)
- Ograniczeniem jest konieczność tworzenia baz
- Ale można również bez baz...

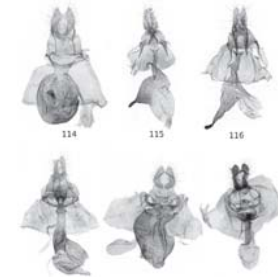
Sekwencjonowanie nowej generacji i kody kreskowe DNA

- tradycyjne sekwencjonowanie Sangera jest drogie i pracochłonne
- gdy uzyskujemy DNA z prób środowiskowych i wykorzystujemy konserwatywne startery PCR konieczne klonowanie – jeszcze droższe i bardziej pracochłonne
- NexGenSeq dostarcza w pojedynczej analizie milionów sekwencji – możliwa bardzo szybka analiza bioróżnorodności całych ekosystemów



Zastosowania kodów kreskowych DNA

- identyfikacja gatunków, których morfologiczne rozróżnienie jest trudne, np. muchówki, niektóre motyle, nicienie (mniej niż 1% gatunków znanych na podstawie opisów morfologicznych)
- identyfikacja gatunków na podstawie włosów, odchodów itp. – monitoring zagrożonych ssaków: tygrys syberyjski, lampart amurski
- identyfikacja organizmów z jakich wytworzono przetworzone produkty – mięso, skóry, kawior - monitoring handlu gatunkami chronionymi



Zastosowania kodów kreskowych DNA

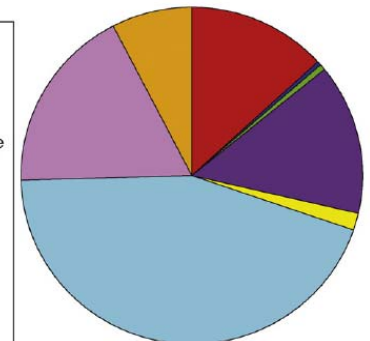
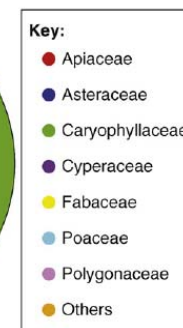
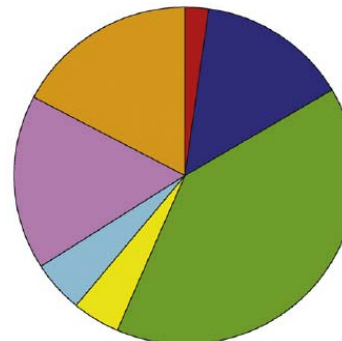
- paleoekologia: analiza DNA z rdzenia lodowego sprzed 450 tys lat. wykazała istnienie lasów iglastych na Grenlandii
- skład diety (a przy okazji zmiany w wegetacji) wymarłego leniwca lądowego (10 – 30 tys. lat temu)
- skład diety współczesnych zwierząt – analiza zawartości żołądków lub odchodów



Golden marmot



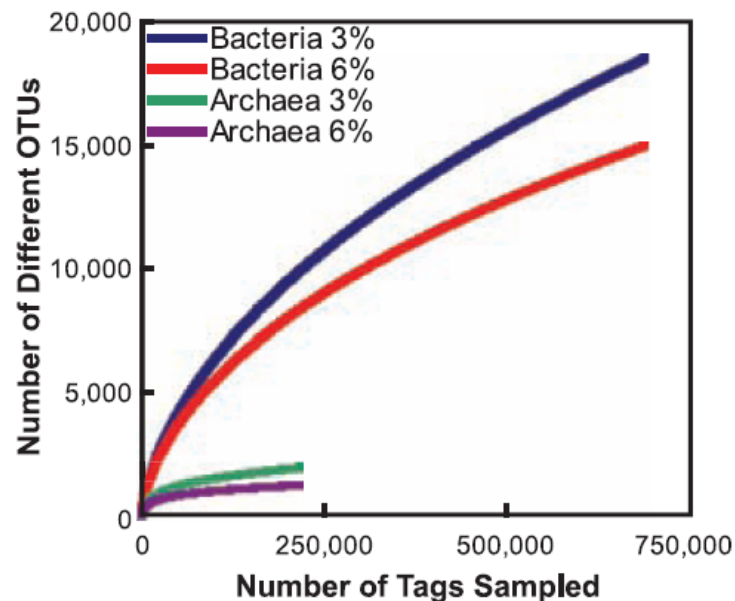
Brown bear



Genomika środowiskowa

taksonomia DNA

- Geny kodujące rybosomalny RNA (rRNA) pozwalają na identyfikację mikroorganizmów
- Ogromnej większości mikroorganizmów nie da się hodować na pożywkach a więc nie wiadomo o nich praktycznie nic, ich liczba jest nieznana
- Sekwencjonuje się glebę, wodę morską, osady morskie itd.
- Na podstawie sekwencji określonych fragmentów rRNA szacuje się liczbę taksonów w próbach środowiskowych MOTU (Molekularne Jednostki Taksonomiczne) – progi różnicowania sekwencji
- „Rzadka biosfera”



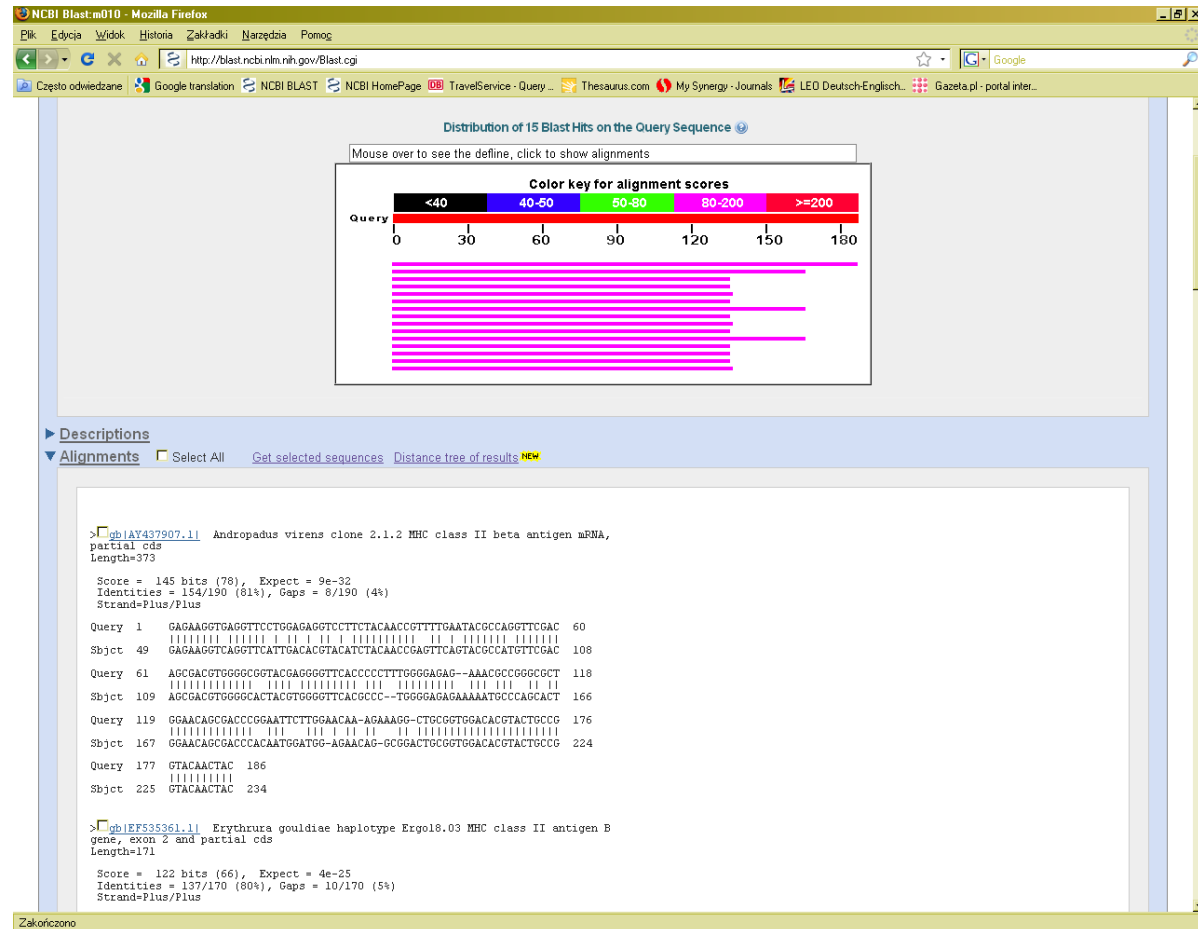
Problemy z kodami kreskowymi i taksonomią DNA

- oparte na jednym fragmencie genomu – możliwa niepoprawna identyfikacja gatunków, bo hybrydyzacja, introgresja, niekompletne sortowanie linii
- używanie uniwersalnych progów dywergencji do wyznaczania MOTU może nie odpowiadać rzeczywistości
- zazwyczaj nieznany stosunek zmienności wewnątrzgatunkowej do zmienności międzygatunkowej
- długość fragmentów niezbędnych dla identyfikacji, tradycyjne kody kreskowe zazwyczaj > 500 bp, a amplifikacja fragmentów zdegradowanego DNA > 150 bp często niemożliwa

Bazy sekwencji i ich przeszukiwanie

- NCBI – GenBank i inne bazy
- BLAST – szukanie podobnych sekwencji
- przeszukiwanie genomów
- przeszukiwanie baz sekwencji białek

Darmowy dostęp do światowych zasobów genetycznych



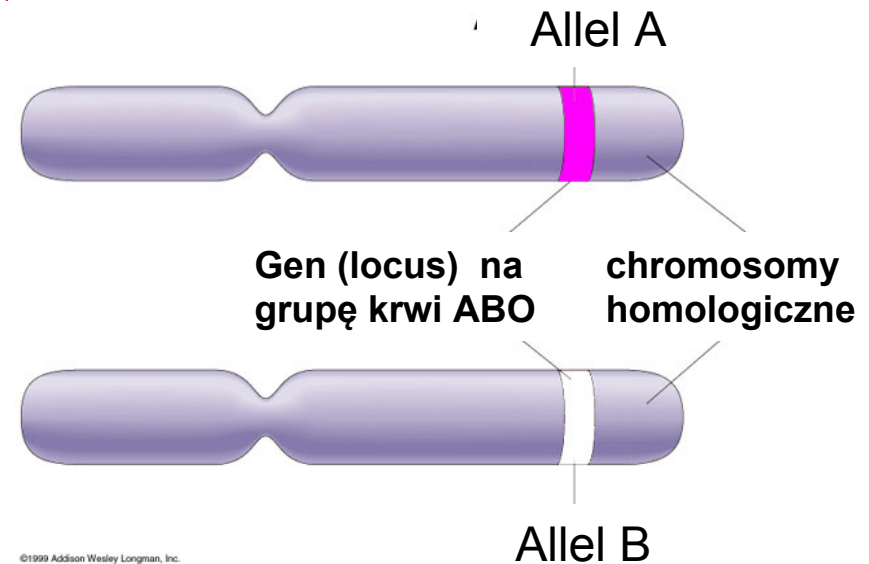
Tanio produkuje się mnóstwo danych genetycznych –
spora szansa że informacje których potrzebujemy są już w bazach –
od nich trzeba zacząć

Pojęcia

- **locus** – miejsce na chromosomie gdzie znajduje się segment DNA, np. dany gen, termin często używany zamiennie z gen
- **allel** – forma genu rozróżnialna od innych form tego samego genu, czasem używane też na określenie kopii genu – **wynika z kontekstu**
 - w populacji może występować wiele różnych alleli w danym locus
 - diploidalny osobnik ma maksymalnie dwa różne allele
- **kopia genu** – termin używany przy liczeniu genów, nie interesuje nas czy kopie genów są takie same czy różne,
 - diploidany osobnik ma dwie kopie każdego genu autosomalnego
 - w populacji N diploidalnych osobników jest $2N$ kopii każdego genu autosomalnego
- **fenotyp** – właściwość organizmu lub ich grupy
 - kolor oczu, włosów, grupa krwi
- **genotyp** – typ genetyczny w jednym lub więcej genów
 - w locus A osobnik może być homozygotą A_1A_1 lub heterozygotą A_1A_2
- **haplotyp** – segment DNA (zazwyczaj nierekombinujący)

Gen (locus) i allel

- miejsce na chromosomie które zajmuje dany gen to **locus**
- **forma genu** która znajduje się w danym locus na konkretnym chromosomie to **allel**
- w każdym locus dostajemy **jeden allel (kopię genu) od każdego z rodziców** –
wyjątki to mitochondrialny DNA i chromosomy płci
- te allele (kopie genu) mogą być **takie same** (homozygota) **lub różne** (heterozygota)
- skład alleli w danym locus to **genotyp**



heterozygota AB – grupa krwi AB

Pojęcia

- **częstość genotypu** – **proporcja** danego genotypu wśród badanych osobników (w badanej populacji)

- dwa allele A_1 i A_2 , gatunek diploidalny, gen autosomalny

$$P = N_{A_1A_1}/N, \quad H = N_{A_1A_2}/N, \quad R = N_{A_2A_2}/N$$

- **częstość allelu** – **proporcja** danego allelu wśród wszystkich badanych kopii genów

- dwa allele A_1 i A_2 , gatunek diploidalny, gen autosomalny
częstość homozygot plus połowa częstości heterozygot (mają tylko jeden allel):

$$p = P + 1/2 H,$$

$$q = Q + 1/2 H, \quad q = 1 - p$$

liczba alleli (kopii genu) danego typu podzielona przez całkowitą liczbę alleli (kopii genu) w populacji

$$p = N_{A_1}/2N,$$

$$q = N_{A_2}/2N$$

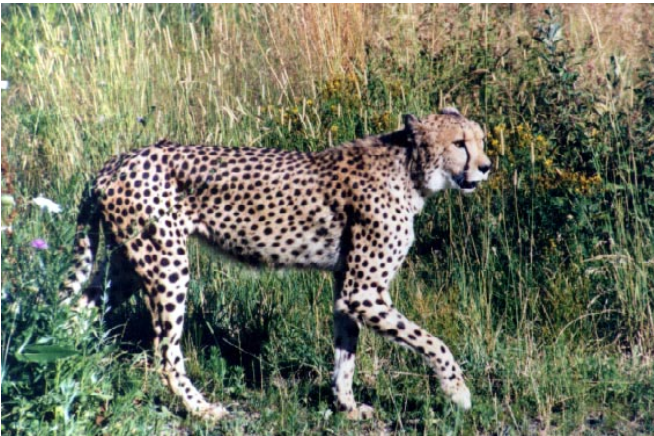
Parametry opisujące zmienność genetyczną i procesy ewolucyjne w populacji

- częstości alleli w badanych loci
- procent loci polimorficznych (P)
- średnia liczba alleli na locus (A), bogactwo alleli (R)
- średnia heterozygotyczność obserwowana i oczekiwana (H_o i H_e)
- odchylenia od równowagi Hardy'ego-Weinberga (HWE)
- nierównowaga sprzężeń (LD)

Procent loci polimorficznych (P)

P = liczba loci polimorficznych/całkowita liczba zbadanych loci

- parametr zależy od kryterium polimorfizmu, jakie przyjmiemy, np. 95%



Brak zmienności genetycznej w 47 loci u geparda ($P = 0.0$)

Średnia liczba alleli na locus (A)

A = całkowita liczba stwierdzonych alleli / liczba zbadanych loci

parametr jest podatny na wielkość próby

bogactwo alleli R to średnia liczba alleli na locus dostosowana do liczebności najmniejszej próby

Najbardziej zmienne loci mogą mieć nawet kilkaset alleli:

- mikrosatelity
- geny zgodności tkankowej

Prawo Hardy'ego-Weinberga

Zależność częstości genotypów od częstości alleli.

Dwa allele o częstościach p (A_1) i q (A_2)

Częstości genotypów:

$$\begin{array}{c} A_1A_1 \\ p^2 \end{array}$$

$$\begin{array}{c} A_1A_2 \\ 2pq \end{array}$$

$$\begin{array}{c} A_2A_2 \\ q^2 \end{array}$$

Opisuje stan braku ewolucji, częstości będą stałe

- nieskończenie duża populacja
- losowe kojarzenie
- brak mutacji
- brak migracji
- brak doboru

Stan równowagi populacja osiągnie po jednym pokoleniu losowego kojarzenia

Populacje naturalne zazwyczaj bliskie równowagi H-W

Hipoteza zerowa w genetyce populacji

Średnia heterozygotyczność obserwowana i oczekiwana (H_o i H_e)

H_o = proporcja heterozygotycznych osobników w badanej próbie uśredniona dla wszystkich loci

H_e = oczekiwana zgodnie z prawem Hardy'ego-Weinberga proporcja heterozygotycznych osobników uśredniona dla wszystkich loci

A_1A_1 : 4 osobniki $H_o = 1/5 = 0.20$

A_1A_2 : 1 osobnik $H_e = [2(0.9 \times 0.1)] = 0.18$

A_2A_2 : 0 osobników

$f(A_1) = p = 0.9$

$f(A_2) = q = 0.1$

Różnorodność nukleotydów

Różnorodność nukleotydów (π) – proporcja pozycji nukleotydowych różniących się między parą sekwencji losowo wybranych z populacji – czyli średnia dla wszystkich możliwych porównań; heterozygotyczność na poziomie nukleotydów

$$\pi = \frac{\sum_{i < j} \pi_{ij}}{\binom{n}{2}} = \frac{\sum_{i=1}^n \sum_{j=i+1}^n \pi_{ij}}{\frac{n(n-1)}{2}}$$

2	0,13							
3	0,59	0,55						
4	0,67	0,63	0,25					
5	0,80	0,84	0,55	0,46				
6	0,80	0,67	0,38	0,46	0,59			
7	0,84	0,71	0,50	0,59	0,63	0,21		
8, 9, 10	1,13	1,10	0,88	0,97	0,59	0,59	0,38	
11	1,12	1,18	0,97	1,05	0,84	0,67	0,46	0,42

π_{ij} proporcja różnych pozycji nukleotydowych w porównaniu i z j

$$\pi = 0.0065 = 0.65\%$$

Różnorodność haplotypów

- Oczekiwana heterozygotyczność (H_E) jest **obliczana z częstości alleli**

$$H_E = 1 - \sum_{i=1}^n p_i^2$$

- Dlatego, **organizm/locus nie musi być diploidalny** aby obliczyć H_E
- Dla gatunków lub markerów haploidalnych (mtDNA, cpDNA, chromosom Y) H_E często nazywa się **różnorodnością haplotypów (h)** i może być **traktowana jako miara zmienności genetycznej**

Odchylenia od równowagi Hardy'ego-Weinberga (HWE)

Obserwowane

Oczekiwane z H-W

A_1A_1 : 4 osobniki

$$p^2 = 0.9^2 = 0.81 \rightarrow 4.05$$

A_1A_2 : 1 osobnik

$$2pq = 2 \times 0.1 \times 0.9 = 0.18 \rightarrow 0.9$$

A_2A_2 : 0 osobników

$$q^2 = 0.1^2 = 0.01 \rightarrow 0.05$$

$$f(A_1) = p = 0.9$$

$$f(A_2) = q = 0.1$$

Bardzo dobra zgodność z oczekiwaniami

Odchylenia od prawa H-W testujemy testem χ^2 , testem dokładnym Fishera a najczęściej testami randomizacyjnymi

Potencjalne przyczyny odchyień od HWE

- nielosowe kojarzenia
- dobór
- samozapłodnienie
- klonalność
- efekt Wahlunda (traktowanie dwu populacji o różnych częstościach alleli jako jednej populacji, niedobór heterozygot)
- strefa kontaktu dwóch zróżnicowanych form
- istnienie zerowych alleli (takich których nie potrafimy zaobserwować)
- inne...

Efekt Wahlunda

Mamy dwie izolowane populacje w równowadze H-W różniące się częstościami alleli, ale myślimy że pobieramy próbę z jednej niepodzielonej populacji

populacja I

$$p = 0.9 \quad q = 0.1$$

Proporcja heterozygot

$$2pq = 2 \times 0.9 \times 0.1 = 0.18$$

populacja II

$$p = 0.1 \quad q = 0.9$$

Proporcja heterozygot

$$2pq = 2 \times 0.1 \times 0.9 = 0.18$$

Nasza próba

$$p_o = 0.5 \quad q_o = 0.5$$

Oczekiwana proporcja heterozygot

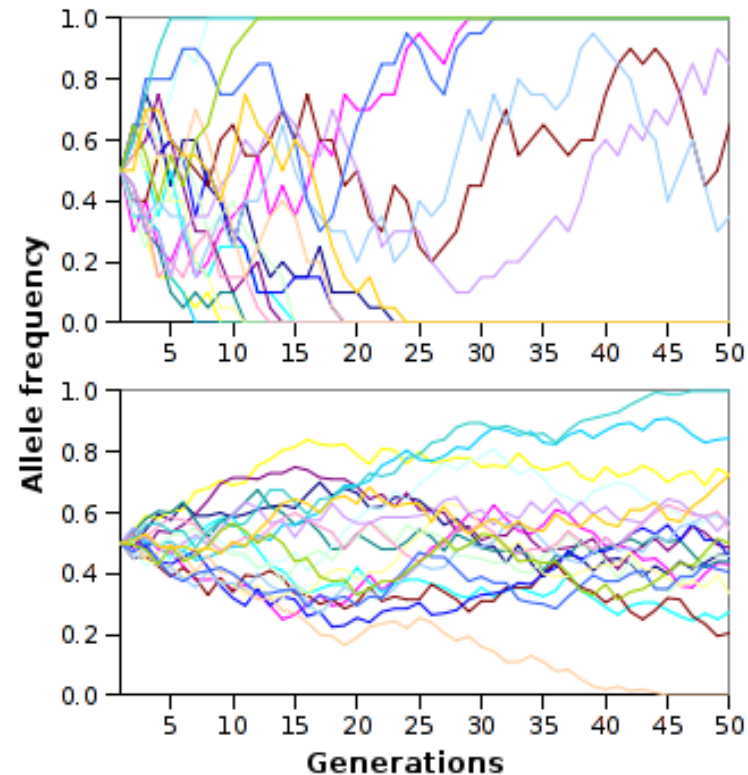
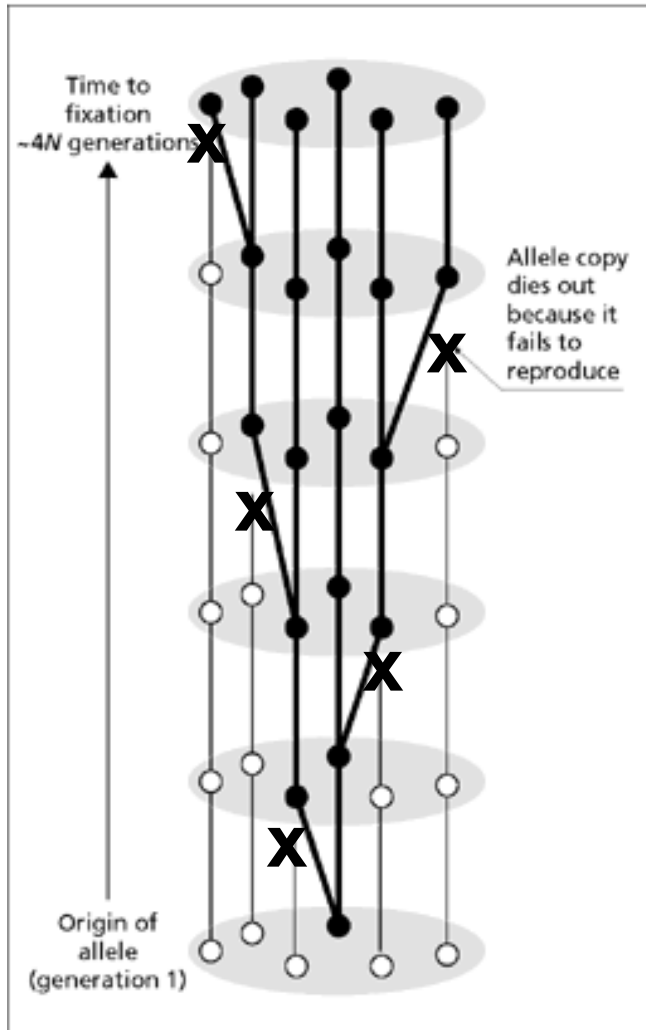
$$2p_o q_o = 2 \times 0.5 \times 0.5 = 0.5$$

Obserwowana proporcja heterozygot 0.18

Pozorny niedobór heterozygot wynika z podziału populacji

Mechanizm dryfu genetycznego

czas

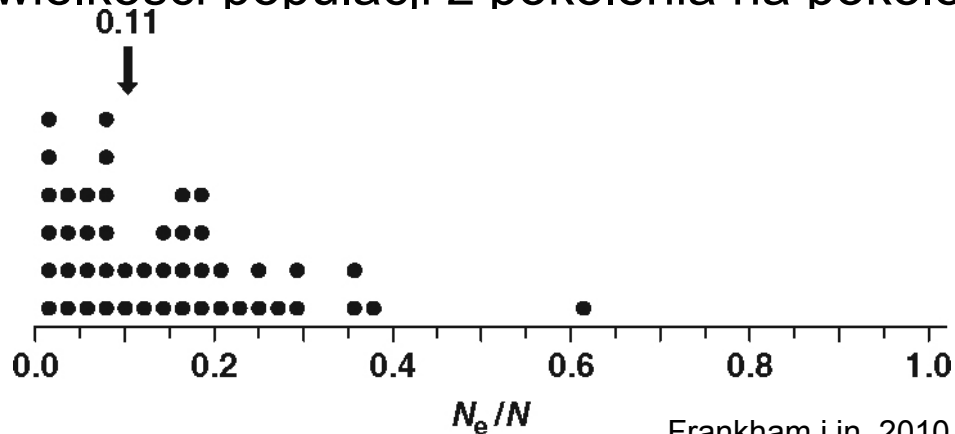


Dryf silniejszy w małych populacjach

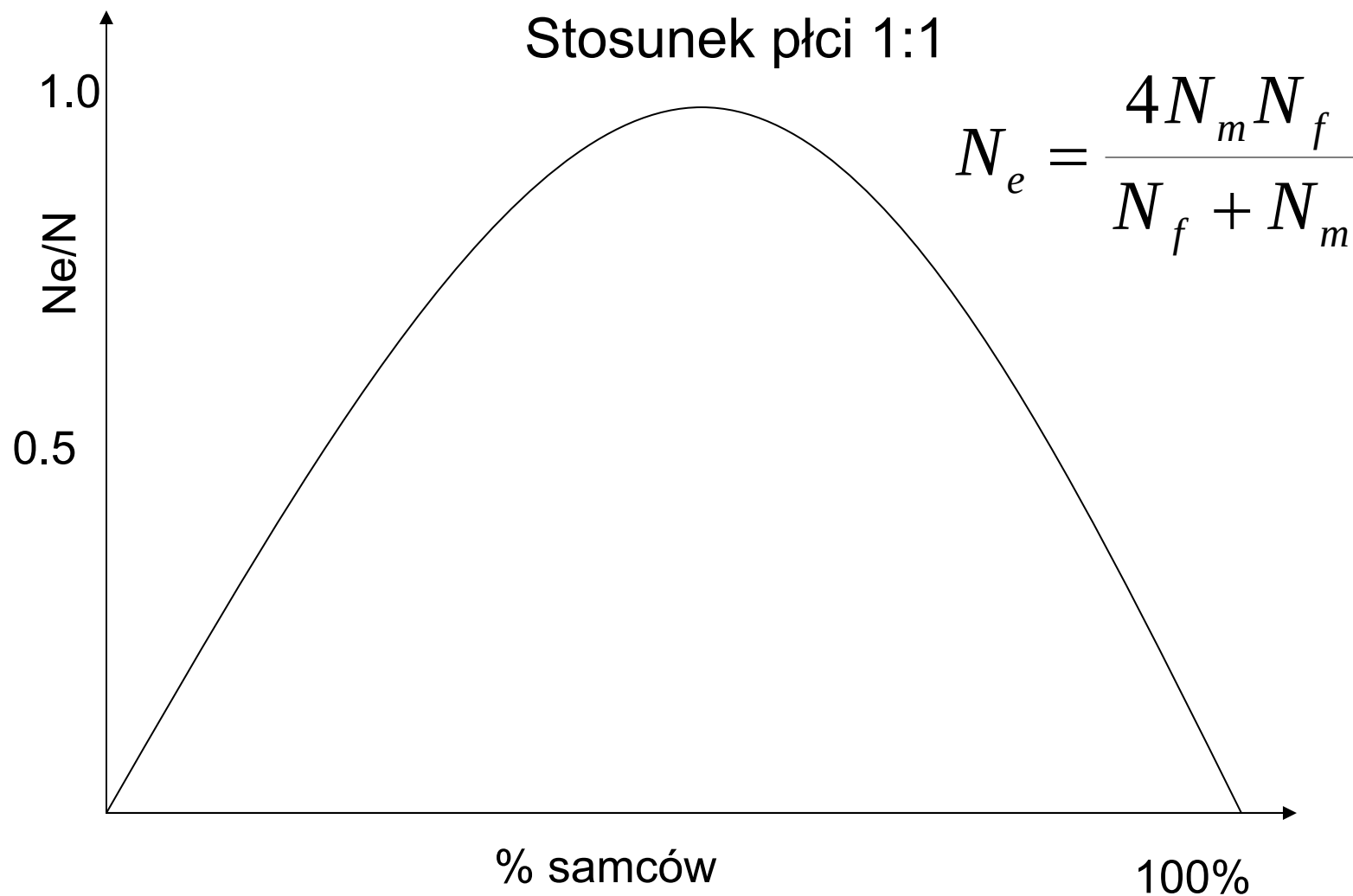
$$\sim 1/N_e$$

Efektywna wielkość populacji (N_e)

- N_e jest to wielkość populacji idealnej (stała wielkość, niezachodzące pokolenia, obupłciowość, średni sukces rozrodczy na osobnika taki sam dla wszystkich klas osobników), przy której populacja traciłaby zmienność genetyczną w identycznym tempie, co populacja badana
- Jest ona zazwyczaj dużo mniejsza od rzeczywistej wielkości populacji (N) bo:
 - nie wszystkie osobniki biorą udział w rozrodzie
 - duża zmienność w sukcesie rozrodczym między osobnikami
 - stosunek płci wśród osobników biorących udział w rozrodzie często odbiega od 1:1
 - zmiany wielkości populacji z pokolenia na pokolenie



Ne zależy od proporcji płci



***Ne* zależy od proporcji płci**

- Północny słoń morski *Mirounga angustirostris*; haremowy system rozrodu, został prawie wytępiony, możliwe że wtedy wszystkie samice były zapładniane przez jednego lub parę samców
- Gdyby w rozrodzie brało udział 100 samic i 1 samiec to efektywna wielkość populacji wynosiłaby $Ne = 4 \times 1 \times 100 / (100 + 1) \approx 4$
- *Ne* zależy również od siły doboru płciowego
- Gdy dobór płciowy działa silniej na jedną z płci to efektywna wielkość populacji części genomu przekazywanej przez tą płć jest niższa – w tym przypadku *Ne* chromosomu Y będzie niższa niż *Ne* mtDNA



Ne zależy od systemu kojarzeń

<u>Kot domowy – wolnożyjące populacje</u>	<u>Ne/N</u>
---	-------------

Populacja promiskuityczna	0.42
---------------------------	------

Populacja polygyniczna	0.33
------------------------	------



Kaeuffer et al. 2004

N_e zależy od zmian wielkości populacji

- zmiany wielkości populacji są najważniejszym czynnikiem wpływającym na długotermiową N_e

$$N_e = \frac{t}{\sum_i \frac{1}{N_i}}$$

- średnia harmoniczna wielkości populacji w kolejnych pokoleniach
- średnia harmoniczna mniejsza od średniej arytmetycznej

Dla liczebności w kolejnych pokoleniach: 1000, 700, 200, 15, 100

średnia arytmetyczna to 403 a N_e (średnia harmoniczna to 59)

Szacowanie N_e metodami molekularnymi

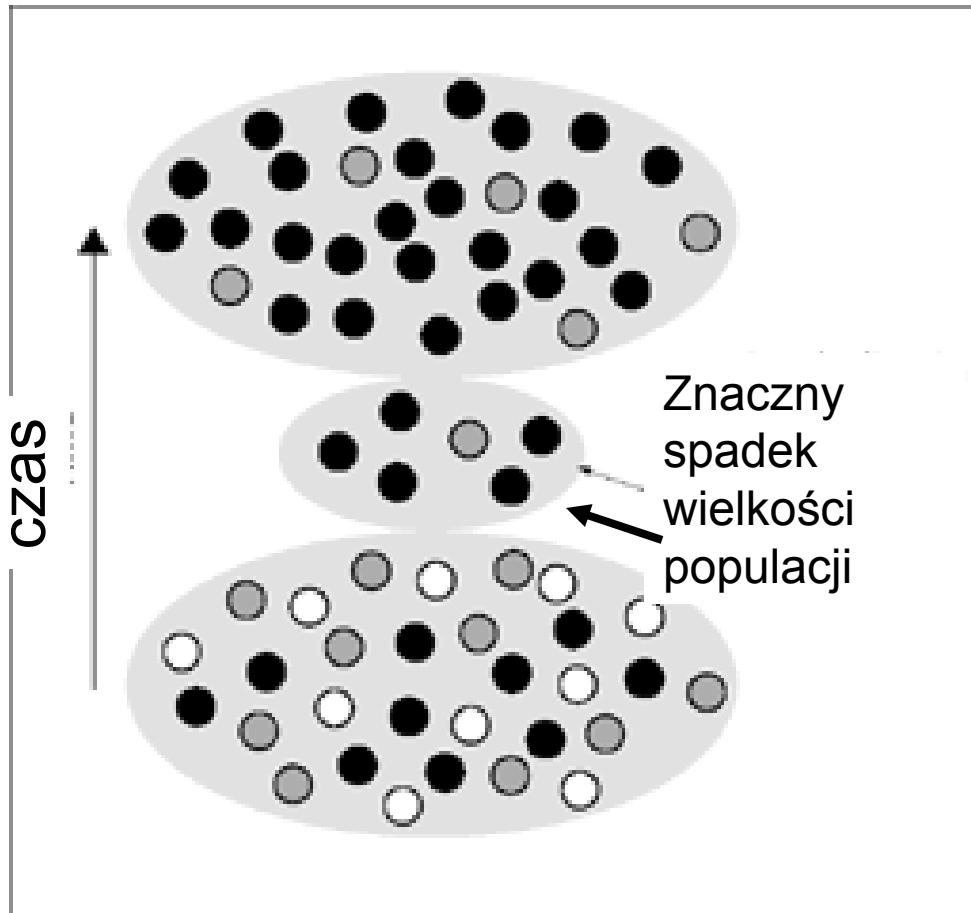
- ze spadku heterozygotyczności markerów molekularnych w kolejnych pokoleniach
- ze zmian częstości alleli z pokolenia na pokolenie w wyniku działania dryfu
- spadku różnorodności allelicznej
- długoterminową wielkość populacji można szacować ze zmienności sekwencji DNA

Estimation of effective population sizes from data on genetic markers

Jinliang Wang*

Institute of Zoology, Zoological Society of London, Regent's Park, London NW1 4RY, UK

Efekt wąskiego gardła (bottleneck)



Niska zmienność
mikrosatelitów u traszki
górskiej w izolowanych
populacjach w Górach
Świętokrzyskich



Efekt wąskiego gardła (bottleneck)

Silny dobór naturalny może utrzymywać zmienność nawet w bardzo małych populacjach, przeciwdziała dryfowi



Lis *Urocyon littoralis* z wyspy San Nicolas u wybrzeży Kalifornii
Najbardziej monomorficzne zwierzę – całkowity brak zmienności w mikrosatelitach i odciskach palców DNA

Polimorficzne geny MHC – odporność na patogeny – zmienność utrzymywana przez dobór

Aguilar et al. 2004

Poziom spadku zmienności genetycznej zależy od:

- stopnia redukcji liczebności
- czasu trwania redukcji liczebności
- ewentualnej imigracji osobników

Żubr – mimo że odtworzono go jedynie z 12 (7) założycieli zachowała się znacząca zmienność mikrosatelitów i MHC, wąskie gardło trwało tylko jedno pokolenie



Znaczna liczebność populacji jest niezbędna do zachowania zmienności genetycznej

w krótkiej skali czasowej pozwoli to uniknąć silnego działania dryfu genetycznego

w długiej skali czasu zapewni utrzymanie potencjału ewolucyjnego - możliwości adaptacji do zmieniających się warunków środowiska

**do działania doboru naturalnego
niezbędna jest zmienność**

Metody badania dyspersji w populacjach naturalnych

- Metody bezpośrednie (terenowe)
 - Bezpośrednia obserwacja przemieszczania się osobników w terenie
 - Metody oparte na znakowaniu osobników: obrączkowanie, znakowanie i powtórny odłów (CMR)
 - Telemetry
 - Wykorzystanie ogrodzonych powierzchni badawczych z monitorowanymi wyjściami
- Metody pośrednie wykorzystujące techniki molekularne

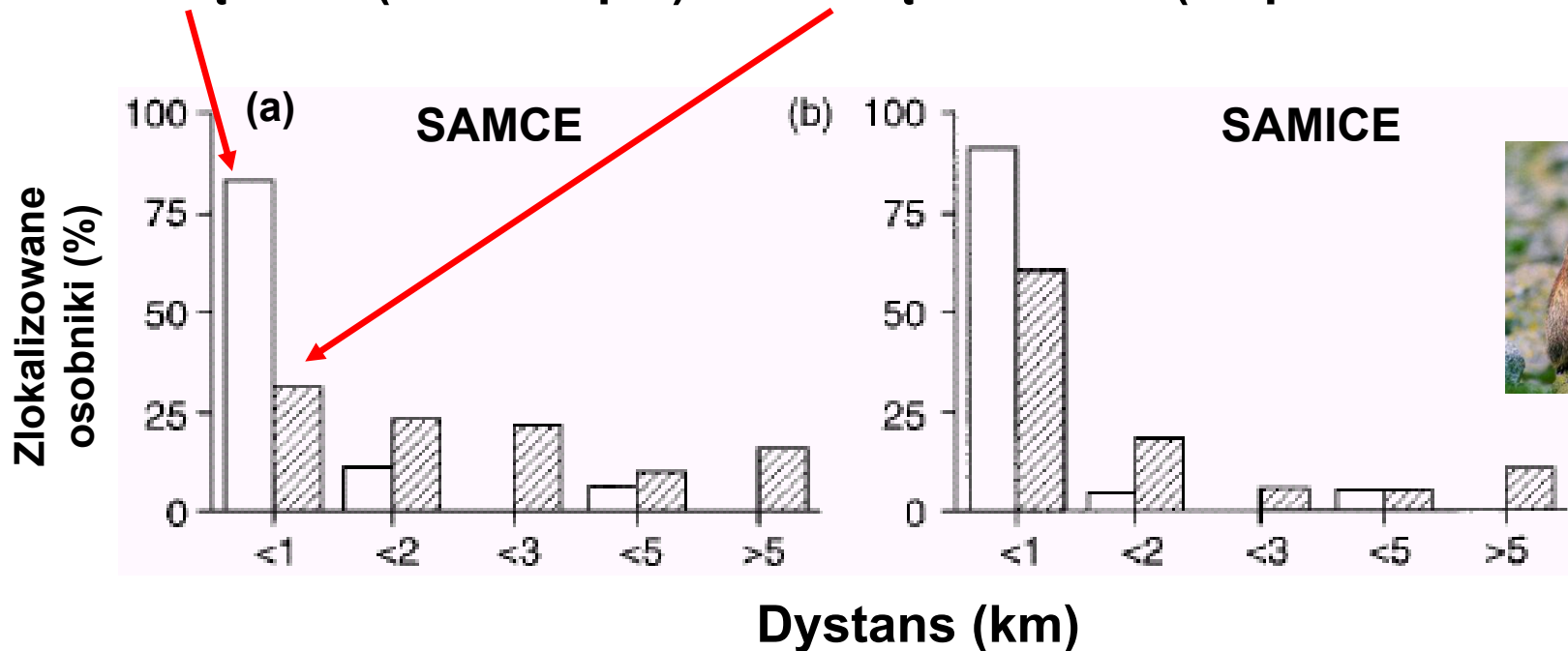
Problemy z metodami bezpośrednimi

- Ograniczona wielkość obszaru objętego badaniami
- Osobników wielu gatunków nie da się znakować
- Niemożność oznakowania wszystkich zwierząt na badanym obszarze
- Trudności w ponownym odłowieniu osobników – emigracja czy śmierć
- Metody bezpośrednie mierzą przemieszczanie się osobników; nie wiadomo czy się rozmnożył
- Różne wyniki przy zastosowaniu różnych metod

Która z metod bezpośrednich jest lepsza?

różne wyniki przy zastosowaniu różnych metod

Procent zlokalizowanych świstaków (*Marmota flaviventris*) metodą CMR (białe słupki) i metodą telemetrii (słupki zakreskowane)



SAMCE - średnia dyspersja: 0,77 km (CMR); 2,56 km (telemetria)

SAMICE - średnia dyspersja: 0,51 km (CMR); 1,44 km (telemetria)

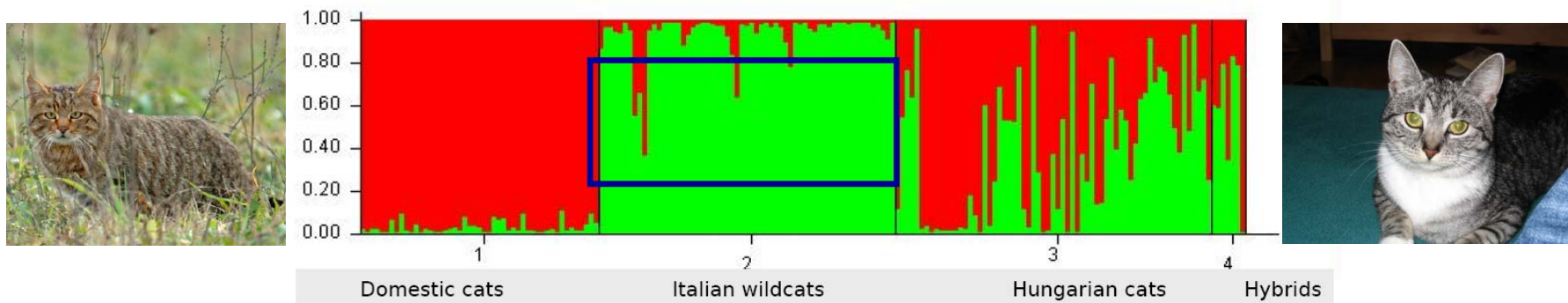
Pośrednie metody badania dyspersji

- Identyfikacja migrantów i ich potomków na podstawie genotypów w wielu loci (poligenotypów)
- Wymaga, aby populacje różniły się genetycznie, lecz te różnice mogą być subtelne
- Wyniki interpretuje się probabilistycznie
- Możliwa identyfikacja potomków migrantów w kolejnych pokoleniach
- Zdolność rozdzielcza zależy od różnic między populacjami oraz liczby dostępnych markerów
- Markery: allozymy, mikrosatelity, AFLP, SNP

Identyfikacja imigrantów

Metody pozwalające na szacownie, na podstawie genotypu określonego dla wielu części genomu (np. wielu mikrosatelitów), jaki procent genomu danego osobnika pochodzi z każdej populacji

Oparte na modelach probabilistycznych i twierdzeniu Bayesa – np. STRUCTURE – stosuje się do badania zarówno hybrydyzacji między gatunkami jak i migracji między słabo zróżnicowanymi genetycznie populacjami



Dyspersja sezonowa

Zróżnicowanie genetyczne między populacjami (F_{ST}) w różnych sezonach → kiedy zaszła dyspersja?

Genetyczne zróżnicowanie (F_{ST}) i wskaźnik inbredu (F_{IS}) w populacjach nornicy rudej



Sezon	F_{ST}	F_{IS}
Wiosna	0,012	0,011
Jesień	0,027	0,058

Borkowska & Ratkiewicz 2004

Efektywna dyspersja osobników nornicy rudej zachodzi wczesną wiosną - przed rozpoczęciem sezonu rozrodczego

Dyspersja zależna od płci

- występuje powszechnie
- u ptaków zazwyczaj wywędrowują samice
- u ssaków płcią migrującą bardziej są samce
- badanie dyspersji zależnej od płci jest ważne z różnych powodów
 - wykorzystanie przez gatunki zasobów
 - zrozumienie czynników kształtujących strukturę genetyczną populacji
 - badania systemów rozrodu

Dyspersja zależna od płci

- analiza markerów dziedziczonych po obu rodzicach (allozymy, mikrosatelity autosomalne, AFLP, SNP autosomalne)
 - porównanie samców i samic po dyspersji ale przed rozmnażaniem
 - sygnał znika u potomstwa
- porównanie markerów o różnym sposobie dziedziczenia
 - porównanie markerów przekazywanych tylko przez jedną płć z markerami autosomalnymi
 - porównanie markerów przekazywanych tylko przez samice z tymi przekazywanymi tylko przez samce

Dyspersja zależna od płci a pokrewieństwo między osobnikami

Porównanie średniego pokrewieństwa (r)
w grupie samic i w grupie samców

Gatunek	Samice	Samce	Źródło
Wydra kanadyjska <i>Lontra canadensis</i>	0,143	0,086	Blundell i in. 2002
Królik <i>Oryctolagus cuniculus</i>	0,185	-0,166	Surridge i in. 1999
Zębiełek <i>Crocidura russula</i>	0,049	0,260	Balloux i in. 1998

 Płeć przemieszczająca się
(osobniki są mniej spokrewnione)

Dyspersja zależna od płci a F_{ST}

Porównanie zróżnicowania genetycznego między populacjami (F_{ST}) wyliczonego dla markerów autosomalnych oddzielnie dla samic i samców

Gatunek	Samice	Samce	Źródło
Zębielek <i>Crocidura russula</i>	0,025	0,145	Balloux i in. 1998
Wydra kanadyjska <i>Lontra canadensis</i>	0,131	0,064	Blundell i in. 2002
Legwan morski <i>Amblyrhynchus cristatus</i>	0,120	0,090	Rassmann i in. 1997

 Płeć przemieszczająca się
(odzwierciedla to małe zróżnicowanie
genetyczne)



Dyspersja zależna od płci markery o różnym sposobie dziedziczenia

MARKERY:

- mtDNA
- geny (np. *SRY*) lub mikrosatelity na chromosomie Y

PORÓWNIANIA:

zmienność markerów specyficznych dla płci (mtDNA, gen *SRY*)



zmienność markerów autosomalnych (allozymy, mikrosatelity)

zmienność markerów specyficznych dla samic (mtDNA)



zmienność markerów specyficznych dla samców (np. gen *SRY*)

Markery przekazywane przez jedną płęć a markery autosomalne

Porównanie F_{ST} dla mtDNA i mikrosatelitów

Gatunek	mtDNA	mikrosat. DNA	Źródło
Nocek Bechsteina <i>Myotis bechsteini</i>	0,961	0,015 (samce)	Kerth i in. 2002
Jaguar <i>Pantera onca</i>	0,342	0,045 (samce)	Eizirik i in. 2001
Ryjówka aksamitna <i>Sorex araneus</i>	0,030 (samice)	0,103	Balloux i in. 2000



Płęć przemieszczająca się

Porównanie bezpośrednich i pośrednich metod badania dyspersji

Metody bezpośrednie

mierzą **mobilność** organizmów w czasie trwania badań

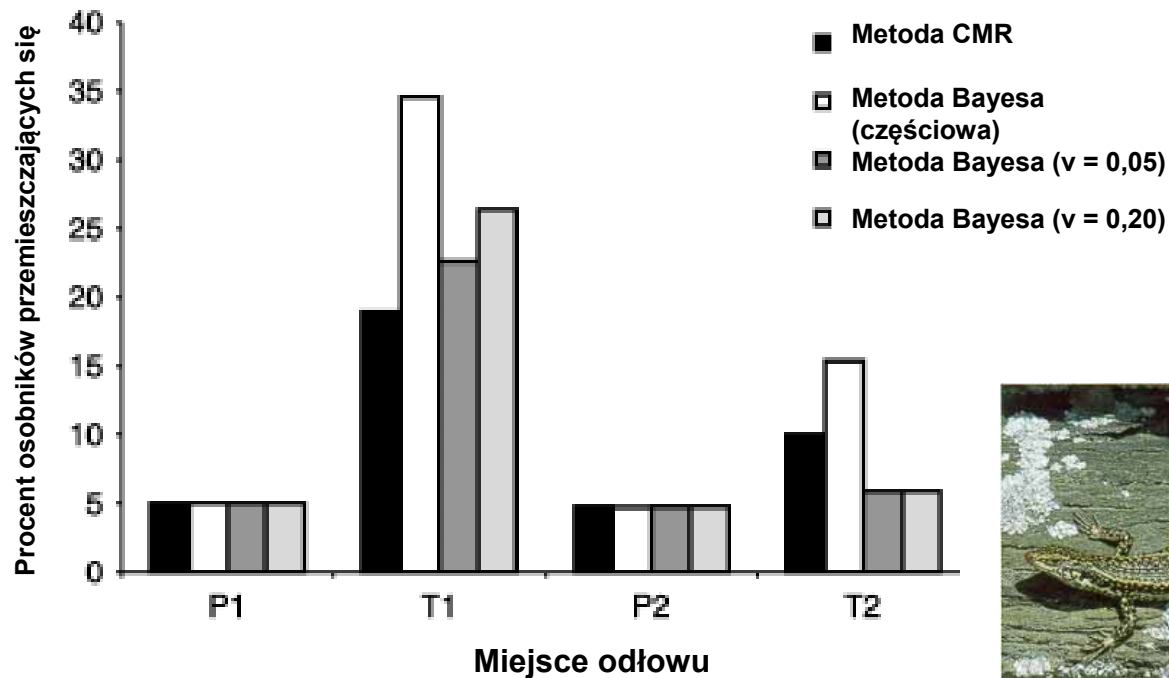
Metody pośrednie

mierzą **efektywną dyspersję**, tj. imigrację zakończoną rozrodem, która zaszła przed rozpoczęciem i w czasie trwania badań



Porównanie bezpośrednich i pośrednich metod badania dyspersji

Procent przemieszczających się scynków *Oligosoma grande* między populacjami w różnych miejscach odłowu określony różnymi metodami

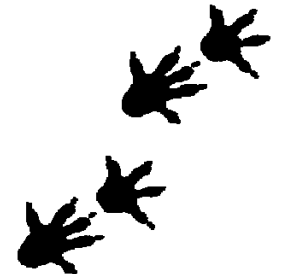


Berry et al. 2004

Metoda CMR > 7 lat badań Metoda genetyczna < 3 miesiące badań

Dyspersja - podsumowanie

- Bezpośrednie metody wyznaczania poziomu dyspersji mają wiele ograniczeń
- Markery molekularne pozwalają identyfikować imigrantów oraz charakter migracji - mierzą efektywną dyspersję
- Dyspersję zależną od płci można badać na wiele sposobów
- Zastosowanie markerów genetycznych oszczędza środki i czas pracy w terenie



Populacje niemal wszystkich gatunków wykazują pewien stopień zróżnicowania w skali lokalnej

Najprostszy powód to osiedlanie się potomstwa w pobliżu miejsca urodzenia

Pionierskie badania Selandera i in. z lat 70. wykazały różnice genetyczne między subpopulacjami myszy z różnych stodół na tej samej farmie



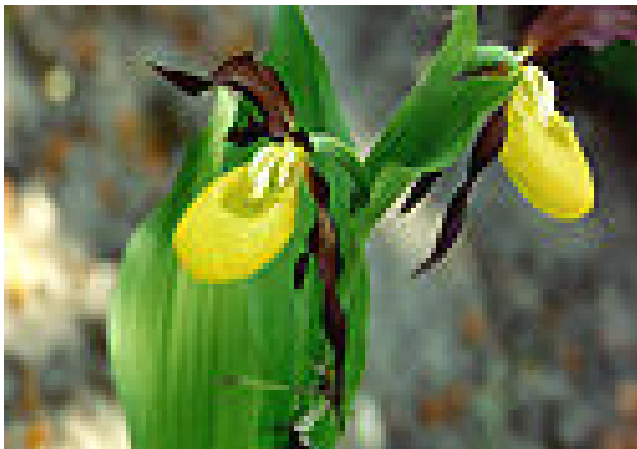
Czynniki kształtujące strukturę genetyczną populacji

- przepływ genów (migracje) powoduje zmniejszenie zróżnicowania, wyrównuje częstości alleli
- dryf genetyczny powoduje wzrost zróżnicowania między populacjami
- dobór naturalny może powodować wzrost jak i spadek zróżnicowania
- mutacje powodują wzrost zróżnicowania

Różny poziom zróżnicowania oczekiwany dla różnych grup organizmów

Np. dla roślin:

- zapylanych przez osiadłe owady powinniśmy obserwować większe zróżnicowanie między populacjami i małą zmienność wewnątrzpopulacyjną
- rośliny wiatropylne powinny wykazywać małe różnice międzypopulacyjne





Zdolne do dalekich dyspersji ptaki i niektóre owady
powinny wykazywać małe zróżnicowanie w
porównaniu z np. płazami.



Statystyki F i podział populacji

Heterozygotyczność na trzech poziomach
dla locus z dwoma allelami:

H_I – obserwowana heterozygotyczność osobnika

H_S – oczekiwana heterozygotyczność w subpopulacji

$$H_S = 2p_i q_i$$

H_T – oczekiwana heterozygotyczność w całej populacji

$$H_T = 2p_o q_o$$

Statystyki F i podział populacji

współczynnik wsobności
spadek heterozygotyczności z powodu nielosowych
kojarzeń w obrębie subpopulacji

$$F_{IS} = 1 - \frac{H_I}{H_S} = \frac{H_S - H_I}{H_S}$$

Statystyki F i podział populacji

spadek heterozygotyczności subpopulacji
z powodu działania dryfu genetycznego

miara zróżnicowania genetycznego
między subpopulacjami

wariancja częstości alleli
między subpopulacjami

$$F_{ST} = 1 - \frac{H_S}{H_T} = \frac{H_T - H_S}{H_T}$$

Statystyki F i podział populacji

Całkowity współczynnik wsobności

Spadek heterozygotyczności z powodu nieelosowych kojarzeń w subpopulacjach **oraz** podziału populacji na subpopulacje (zróżnicowania będącego wynikiem dryfu)

$$F_{IT} = 1 - \frac{H_I}{H_T} = \frac{H_T - H_I}{H_T}$$

Skala Wrighta (1978) dla F_{ST}

0.00 - 0.05: małe zróżnicowanie genetyczne

0.05 - 0.15: umiarkowane

0.15 - 0.25: duże

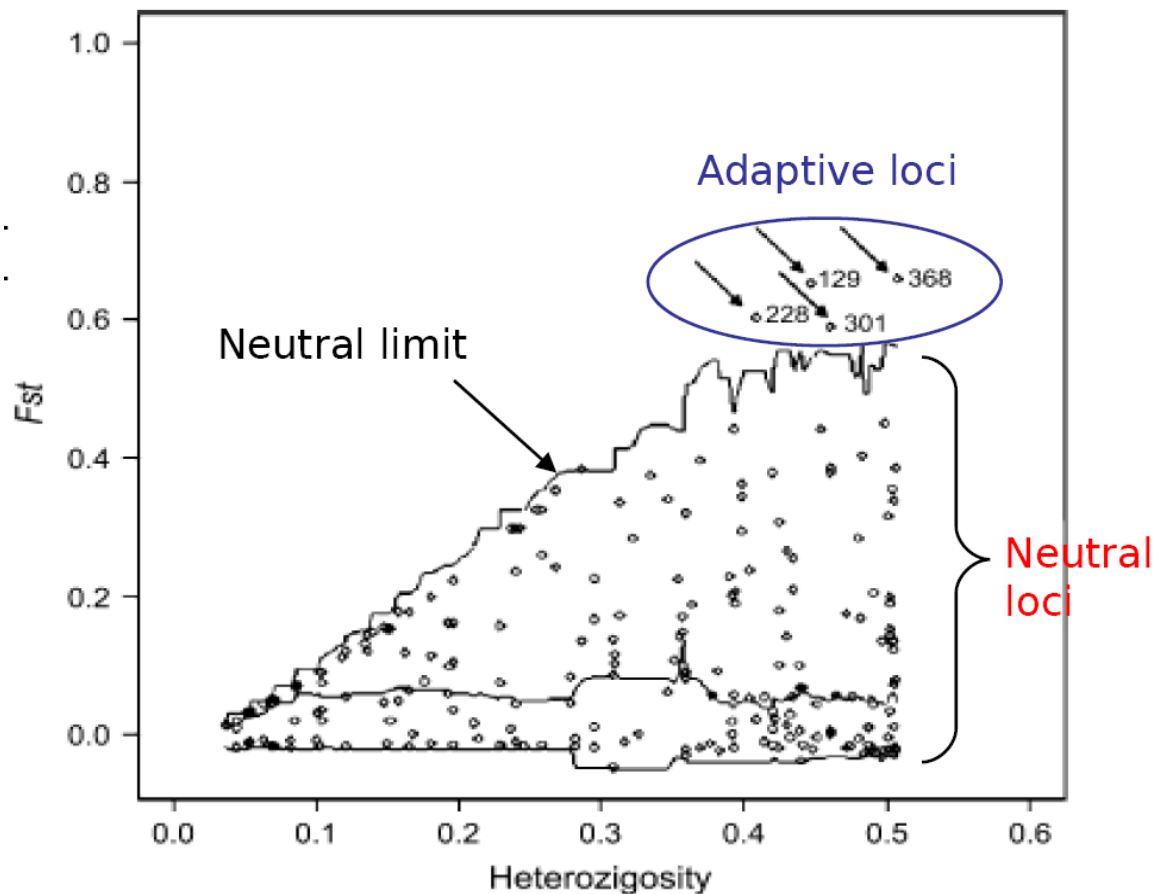
0.25 - 1.00: bardzo duże

Ta skala ma sens tylko dla markerów biallelicznych, dla innych jest myląca, bo maksymalne możliwe F_{ST} zależy od zmienności markera

F_{ST} a wykrywanie genów odpowiedzialnych za lokalne adaptacje

Dryf działa na wszystkie loci tak samo,
a więc neutralne winny mieć podobne F_{ST}

Spodziewamy się
że loci odpowiedzialne
za lokalne adaptacje
będą miały wyższe F_{ST}



***F_{st}* jako miara przepływu genów**

$$Nm \approx (1 - F_{ST}) / 4F_{ST}$$

Nm = liczba migrantów na pokolenie między dwoma populacjami w równowadze

Jeżeli *Nm* < 1, to między populacjami mogą utrwalić się różnice genetyczne

Nm wyliczone w ten sposób nie pozwala odróżnić dawnej dywergencji od współczesnej bariery dla przepływu genów

lub

ekspansji populacji z 1 źródła od swobodnego przepływu genów między populacjami.

Inny sposób obliczania *Nm* wykorzystuje częstości rzadkich alleli

Zróżnicowanie genetyczne populacji ryb

wzrasta od ryb słonowodnych, przez anadromiczne do słodkowodnych - odpowiednio: 1.6, 3.7 i 29.4% - jest to efekt zasiedlania nieciągłego środowiska

Fragmentacja środowisk może prowadzić do spadku zmienności wewnątrzpopulacyjnej i wzrostu zróżnicowania genetycznego między populacjami.



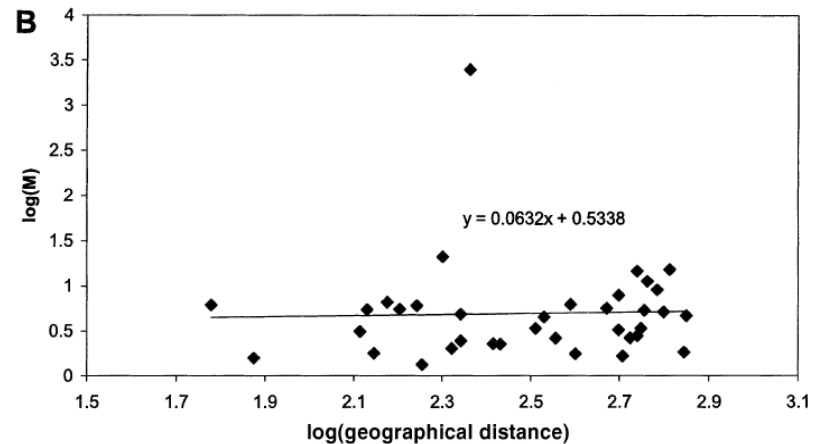
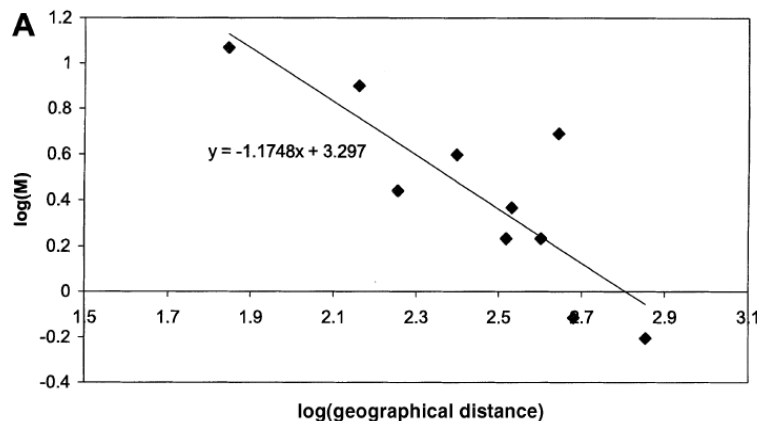
Model izolacji przez odległość (IBD)

Odległość genetyczna między populacjami jest dodatnio skorelowana z odległością geograficzną między nimi

- zasięg gatunku zwykle większy niż możliwości dyspersji pojedynczego osobnika
- populacje bliższe geograficznie są również genetycznie zbliżone wskutek np. filopatrii
- model ten ma zastosowanie dla populacji charakteryzujących się równowagą między dryfem a dyspersją

Model izolacji przez odległość (IBD)

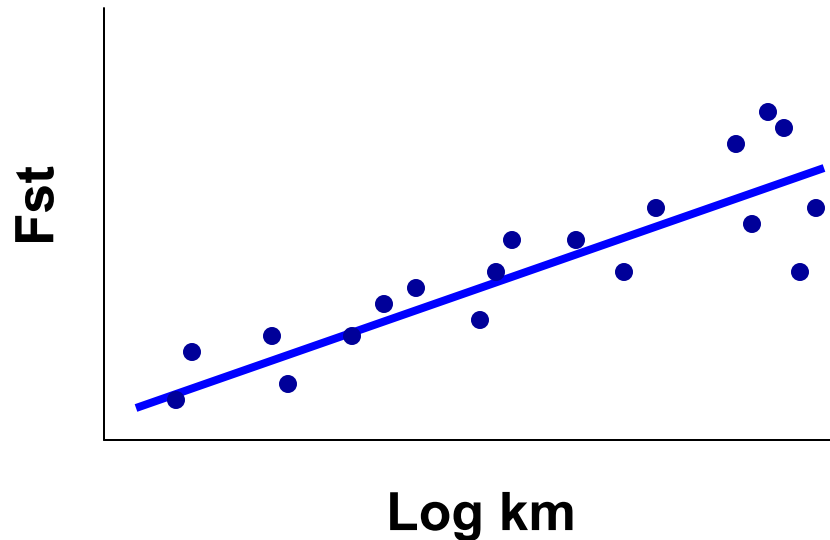
Korelacja między przepływem genów a odległością geograficzną
u żaby moczarowej na pd. od Karpat,
brak takiej zależności dla na pn. od Karpat



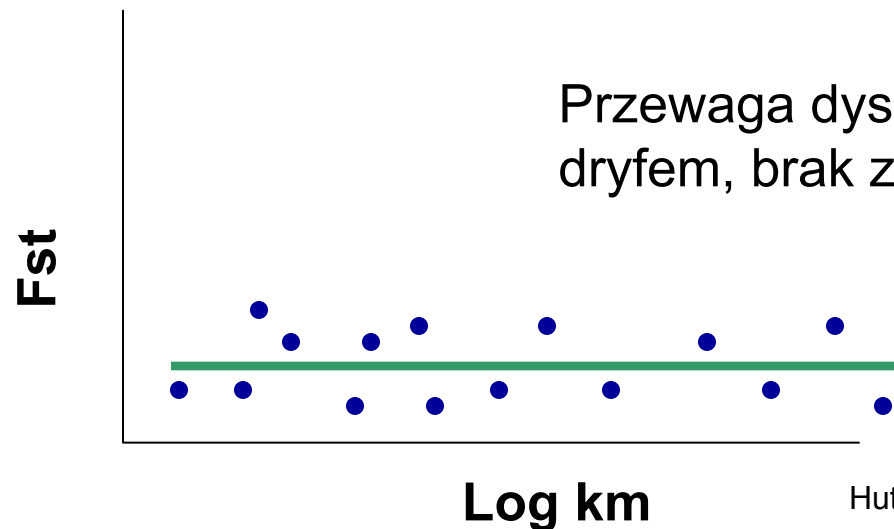
Brak zależności między zróżnicowaniem genetycznym a odległością geograficzną

- małe F_{st} między wszystkimi populacjami - niedawna ekspansja lub intensywny przepływ genów w chwili obecnej
- duże F_{st} między populacjami - przewaga dryfu nad dyspersją

Testowanie modelu izolacji przez odległość

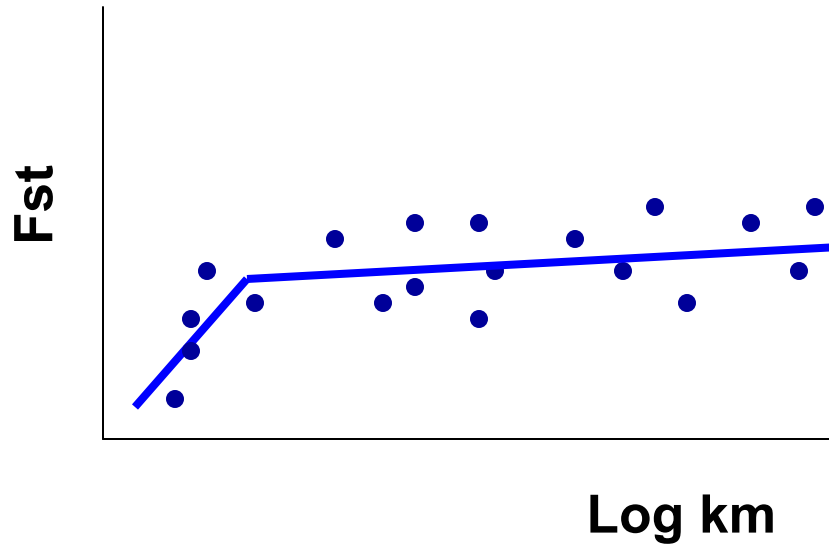


Równowaga między
dryfem a dyspersją,
zależność IBD

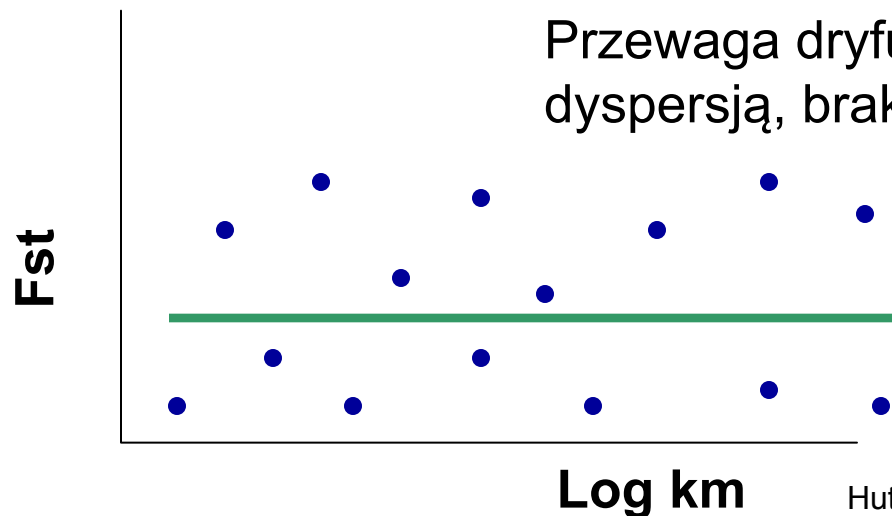


Przewaga dyspersji nad
dryfem, brak zależności IBD

Testowanie modelu izolacji przez odległość



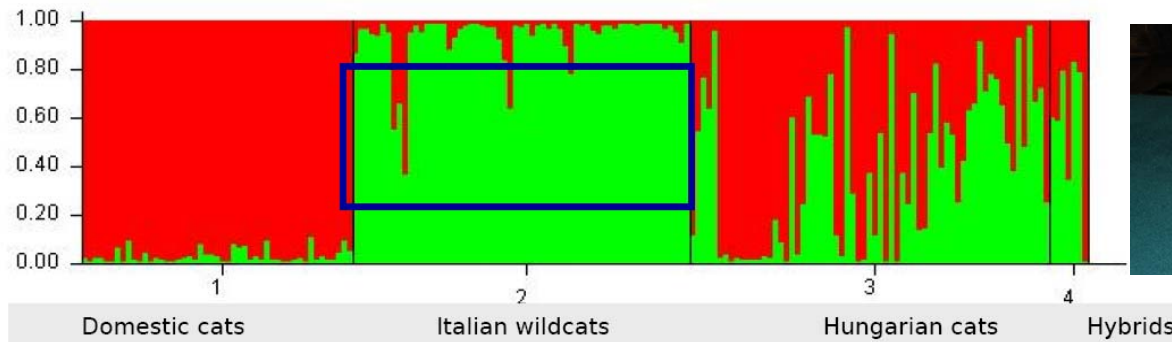
Przewaga dyspersji w skali lokalnej, przewaga dryfu na większych odległościach



Przewaga dryfu nad dyspersją, brak zależności IBD

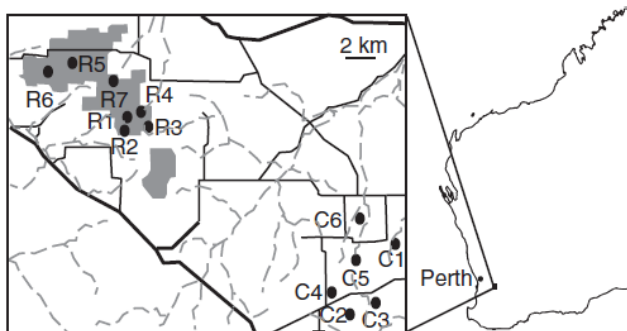
Structure

- często trudno zdecydować ile odrębnych genetycznie populacji występuje na danym obszarze
- K to założona liczba różnych genetycznie grup
- algorytm szacuje częstości alleli w wielu loci charakterystyczne dla każdej z K grup, i jednocześnie szacuje dla każdego osobnika jaki procent jego genomu należy do każdej z K grup (populacji)
- sprawdzamy różne wartości K , oceniając statystycznie dopasowanie modelu do danych
- w ten sposób można nie tylko określić liczbę populacji, zaklasyfikować osobniki do populacji, ale też zidentyfikować potomków migrantów czy mieszańce międzygatunkowe



Rozmieszczenie środowisk a dyspersja

- Jaszczurka *Ctenophorus ornatus* żyje na **wychodniach granitowych** w zach. Australii W
- Niektóre wychodnie otoczone buszem (Reserve), inne terenami uprawnymi (Cleared land)
- Nie ma redukcji środowiska – wychodnie nieprzydatne dla rolnictwa
- 22 mikrosatelity
- Porównanie **przepływu genów** między wychodniami w **Reserve** oraz w **Cleared land**



Rozmieszczenie środowisk a dyspersja

Outcrop	Sample size	Allelic richness
Reserve outcrops		
R1	41	7.8 ± 0.4
R2	13	6.9 ± 0.5
R3	14	6.8 ± 0.5
R4	12	7.6 ± 0.5
R5	29	6.6 ± 0.5
R6	19	7.1 ± 0.4
R7	20	6.8 ± 0.4
Mean	21.1 ± 4.0	7.1 ± 0.2
Cleared land outcrops		
C1	16	5.1 ± 0.4
C2	24	5.1 ± 0.3
C3	15	4.3 ± 0.3
C4	19	5.6 ± 0.3
C5	12	5.6 ± 0.4
C6	19	5.0 ± 0.3
Mean	17.5 ± 1.7	5.1 ± 0.2

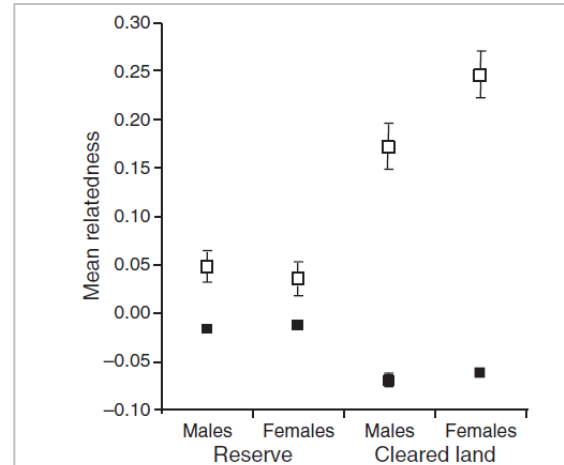


Fig. 2 Mean pairwise relatedness and 95% confidence intervals (as determined by bootstrapping) between individuals within outcrops (open squares) and between outcrops (closed squares).

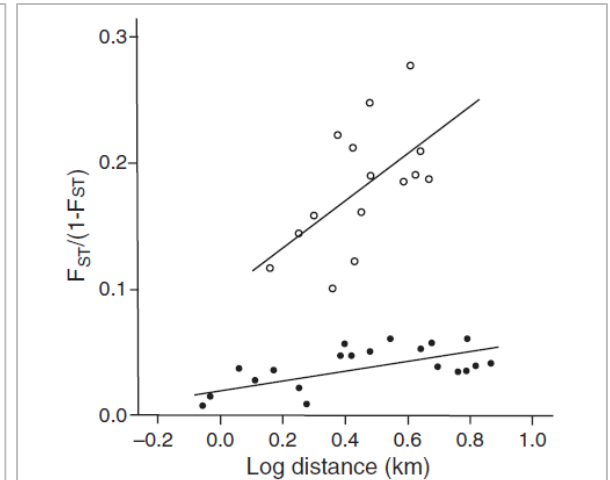
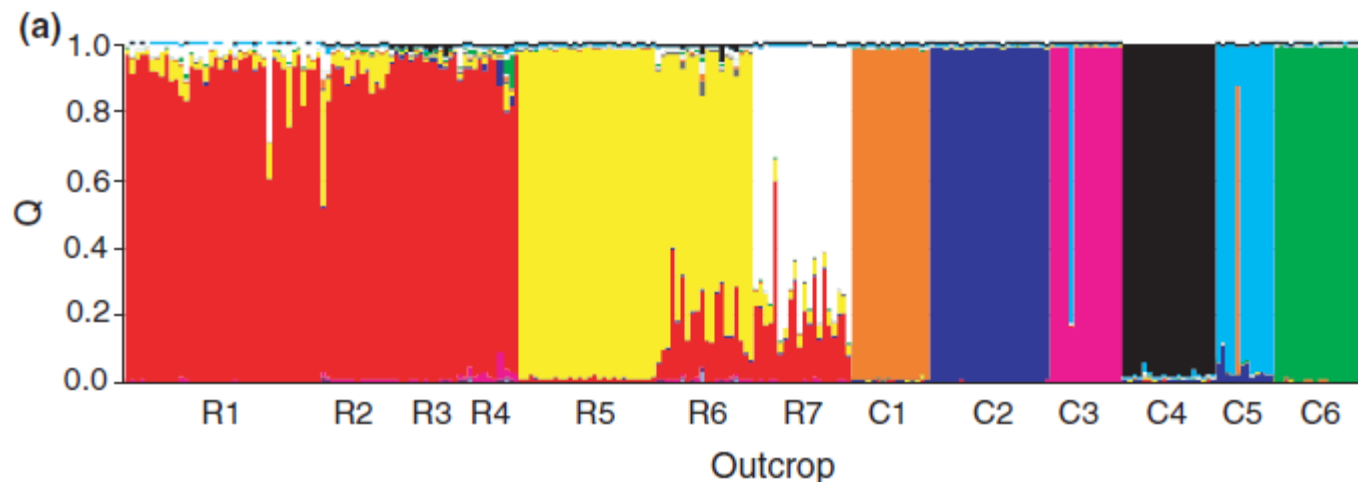


Fig. 3 Relationships between genetic divergence and geographical distance for reserve (closed circles) and cleared land (open circles) outcrops (a log of 0.1 = 100 m and log of 1 = 10 km).



Genomy organizmów zawierają informacje historyczne o:

- wieku gatunków
- wielkości populacji i jej zmianach
- demograficznych aspektach specjacji
- o historycznym i współczesnym przepływie genów między populacjami i gatunkami
- tempie rekombinacji

Te informacje można „odczytać”

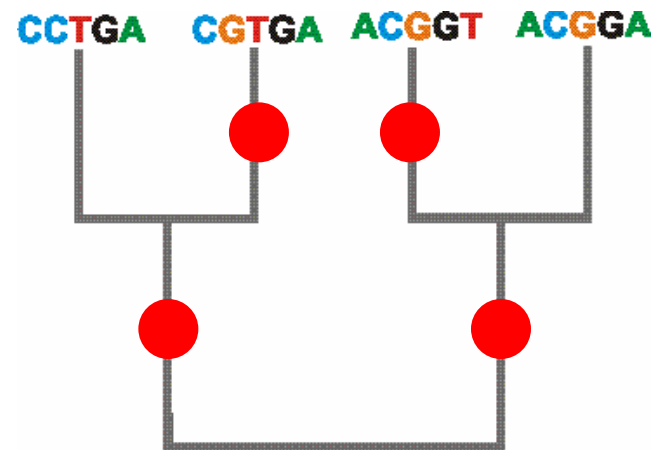
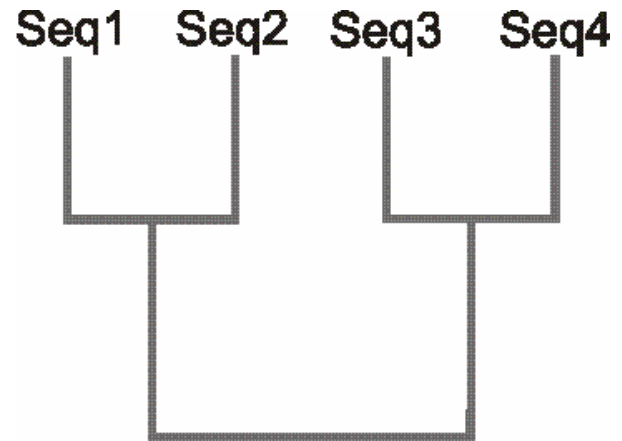
Genealogie genów

genealogie genów = genealogie =
drzewa genów = drzewa

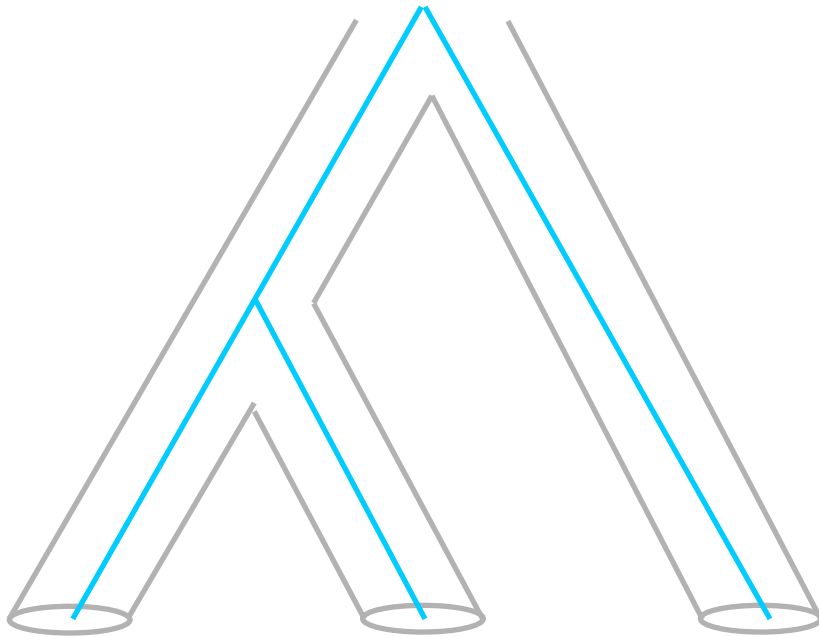
Pokrewieństwa między
sekwencjami w próbie pobranej ze
współczesnych populacji lub
blisko spokrewnionych gatunków

Genealogie istnieją zawsze ale
potrafimy je zrekonstruować tylko
gdy jest zmienność

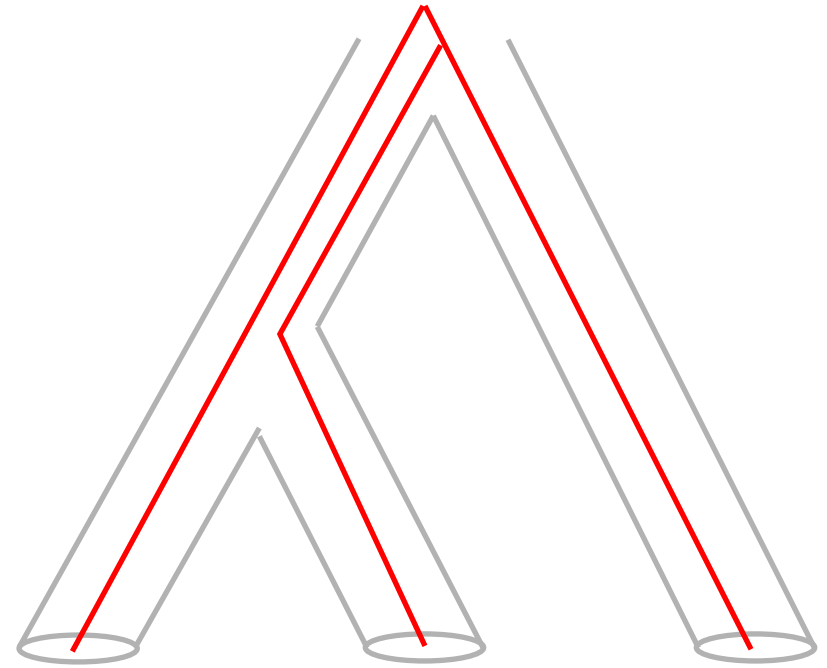
Próba ma genealogię nawet gdy
wszystkie sekwencje są
identyczne!



Drzewa genów nie są drzewami gatunków

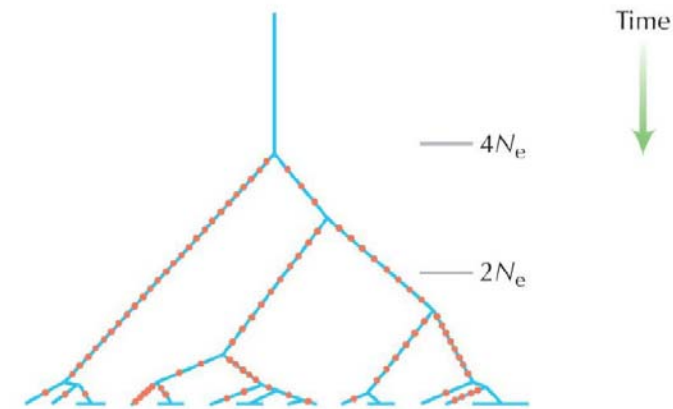


drzewo genu
identyczne jak drzewo
gatunków



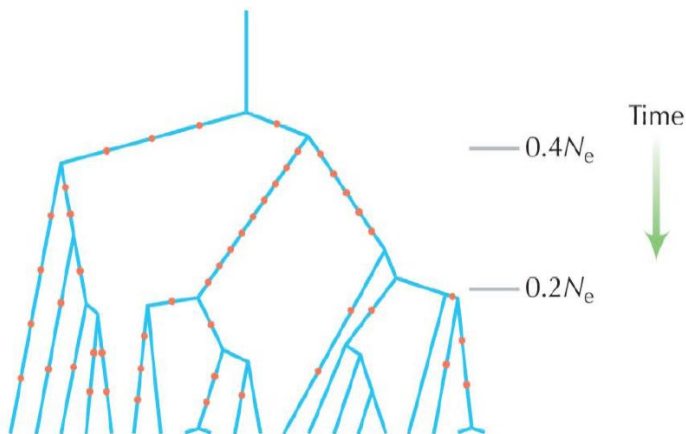
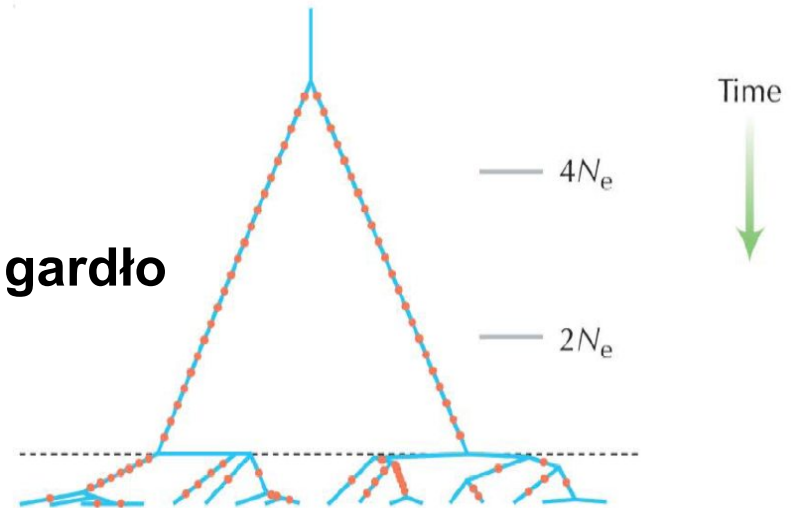
drzewo genu różne od
drzewa gatunków,
polimorfizm w gatunku
ancestralnym

Kształt genealogii genów zależy od demografii populacji



Stała wielkość

Wąskie gardło



Wzrost

Barton et al. 2007

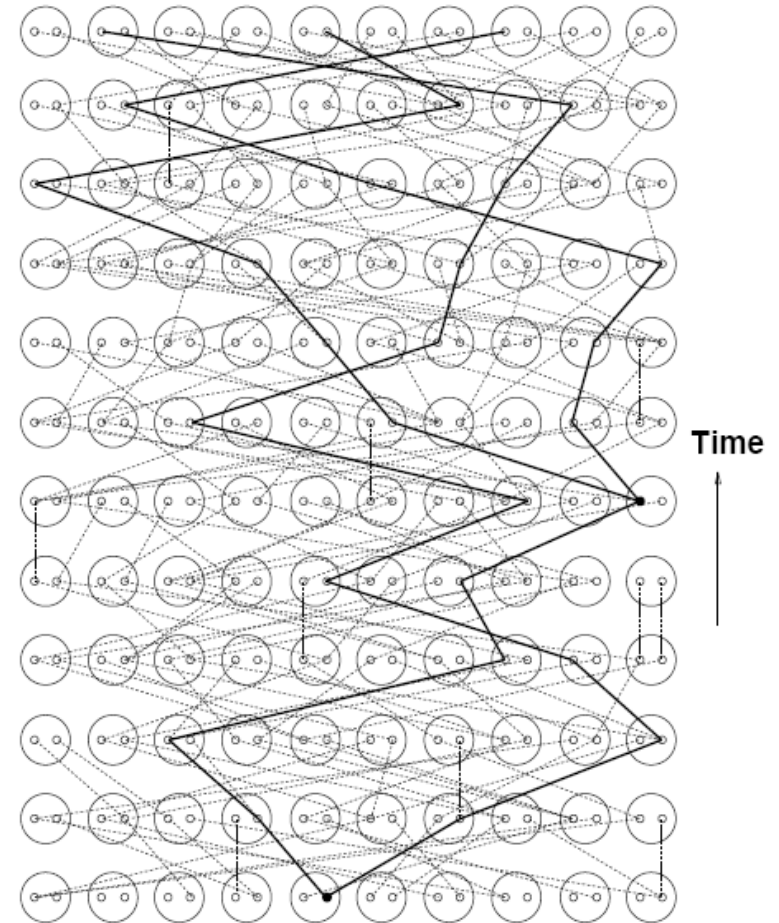
Sekwencje DNA są informatywne, uniwersalne i wygodne

- dostępne dla wszystkich organizmów
- dostarczają informacji genealogicznych, to znaczy można ustalić pokrewieństwa między allelami
- gdy brak danych kopalnych jedyne źródło informacji genealogicznej
- gdy dostępne dane kopalne informacje z sekwencji i skamieniałości można łączyć a także wykorzystywać do formułowania i testowania hipotez

**Ale trzeba wiedzieć jak „odczytać”
informacje z sekwencji**

Teoria koalescencji

- Rozpatrujemy genealogię próby sekwencji, nie osobników
- U diploidnego gatunku liczba przodków wzrasta wstecz, każdy osobnik ma dwu rodziców
- Kopia genu (sekwencja) ma zawsze jednego przodka
- Czasem dwie sekwencje mają wspólnego przodka w poprzednim pokoleniu – **dochodzi do koalescencji**



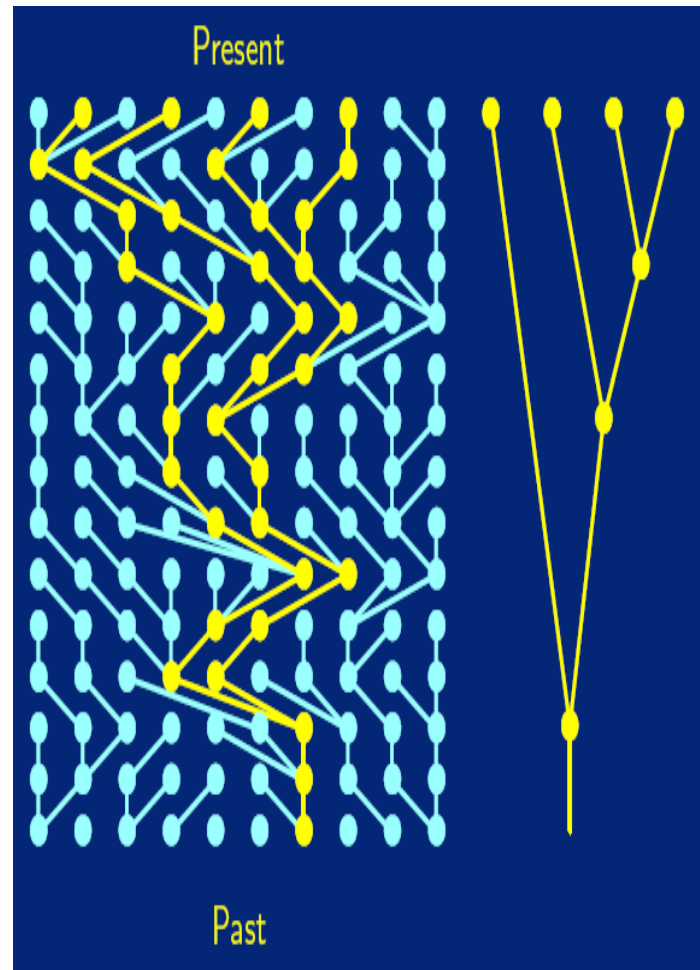
Teoria koalescencji

- Zajmuje się przeszłością próby sekwencji określonego locus
- Czas biegnie „do tyłu” – do ostatniego wspólnego przodka próby

Koalescencja zajmuje się przeszłością próby

- Śledzimy historię naszej próby w miarę jak linie zlewają się ze sobą w przeszłości (ulegają koalescencji)
- tempo koalescencji zależy od aktualnej liczby linii (i) oraz wielkości populacji
- w małych populacjach koalescencja jest szybsza!

$$\frac{i(i-1)}{4N}$$



Teoria koalescencji

- Zajmuje się przeszłością próby sekwencji określonego locus
- Czas biegnie „do tyłu” – do ostatniego wspólnego przodka próby
- Można obliczyć prawdopodobieństwo danej genealogii dla danej wielkości populacji (N)

Można obliczyć prawdopodobieństwo genealogii

dla danego N

ale musimy znać czasy u_i

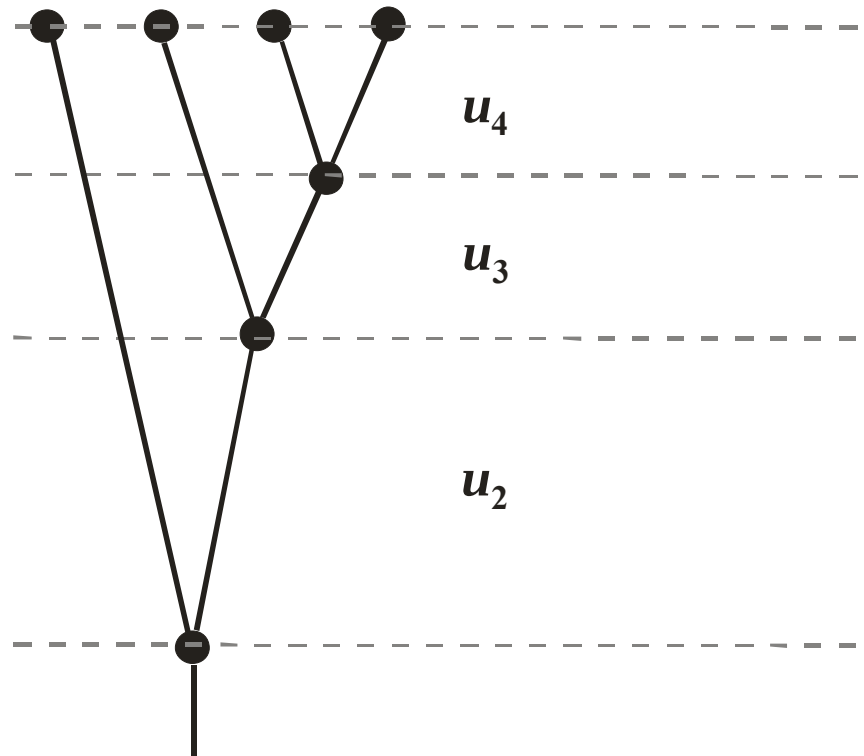
nie potrafimy bezpośrednio
mierzyć czasu

ale potrafimy mierzyć
dywergencję za pomocą
mutacji

a więc zamiast N mamy
złożony parametr

$$\Theta = 4\mu N$$

$$P(G | N) = \frac{1}{2N} \prod_i \exp\left(-u_i \frac{i(i-1)}{4N}\right)$$

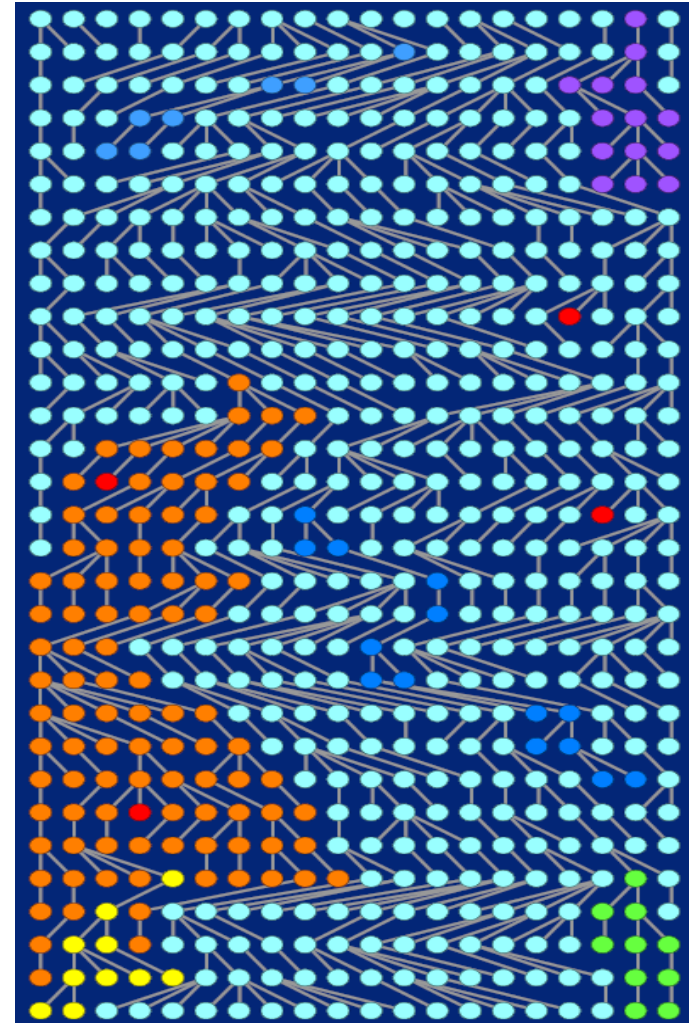


Teoria koalescencji

- Zajmuje się przeszłością próby sekwencji określonego locus
- Czas biegnie „do tyłu” – do ostatniego wspólnego przodka próby
- Można obliczyć prawdopodobieństwo danej genealogii dla danej wielkości populacji (N)
- Teorię koalescencji wyprowadzono dla modelu Wrighta-Fishera...

Model Wrighta-Fishera

- bardzo prosty lecz użyteczny
- stała wielkość populacji N
- niezachodzące na siebie pokolenia
- kolejne pokolenie tworzy się łącząc losowo allele z efektywnie nieskończonej puli
- wszystkie mutacje są neutralne



Teoria koalescencji

- Zajmuje się przeszłością próby sekwencji określonego locus
- Czas biegnie „do tyłu” – do ostatniego wspólnego przodka próby
- Można obliczyć prawdopodobieństwo danej genealogii dla danej wielkości populacji (N)
- Teorię koalescencji wyprowadzono dla modelu Wrighta-Fishera...
- Ale koalescencja „obowiązuje” dla wielu innych modeli populacji jeżeli zamiast N podstawimy N_e

Koalescencja jest niezastąpiona w symulacjach

- symuluje się tylko genealogię próby, nie zaś całej populacji
- wystarczy mała próba (5-10 sekwencji na populacje)
- demograficznych aspektach specjacji
- o historycznym i współczesnym przepływie genów między populacjami i gatunkami
- tempie rekombinacji

Te informacje można „odczytać”

Koalescencja jest niezastąpiona w symulacjach

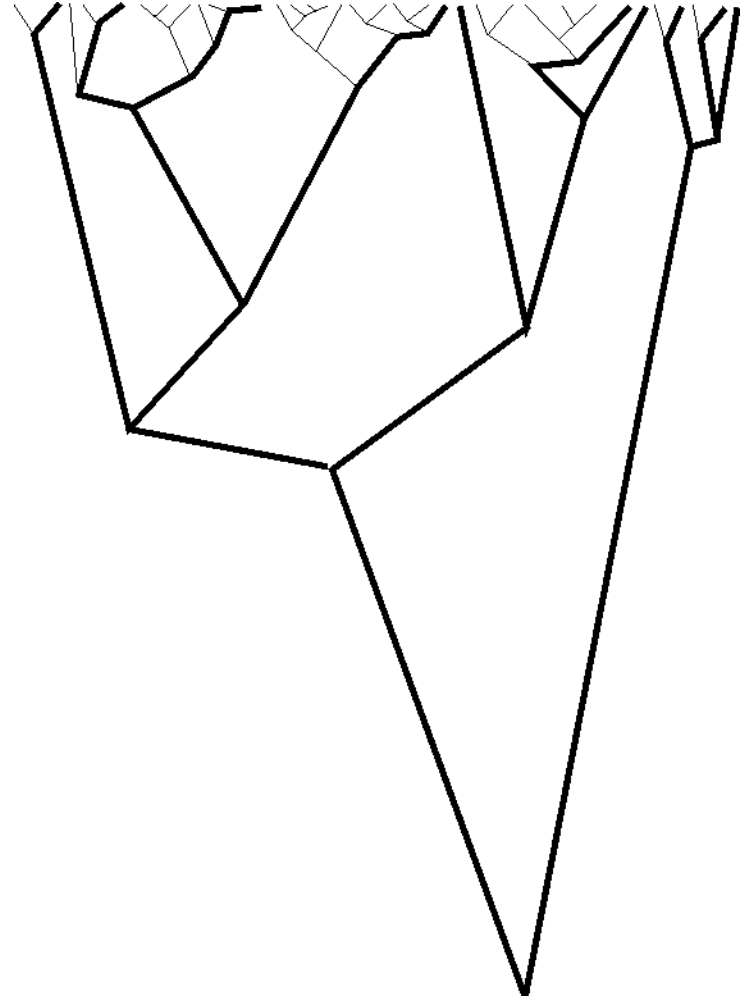
- symuluje się tylko genealogię próby, nie zaś całej populacji
- wystarczy mała próba (5-10 sekwencji na populację)

Zwiększanie wielkości próby niewiele daje

- kilka sekwencji zazwyczaj reprezentuje całą głębokość genealogii

$$P = \frac{n-1}{n+1}$$

- kolejne sekwencje znajdują się zazwyczaj blisko wierzchołków, a więc dają niewiele dodatkowych informacji



Koalescencja to proces stochastyczny

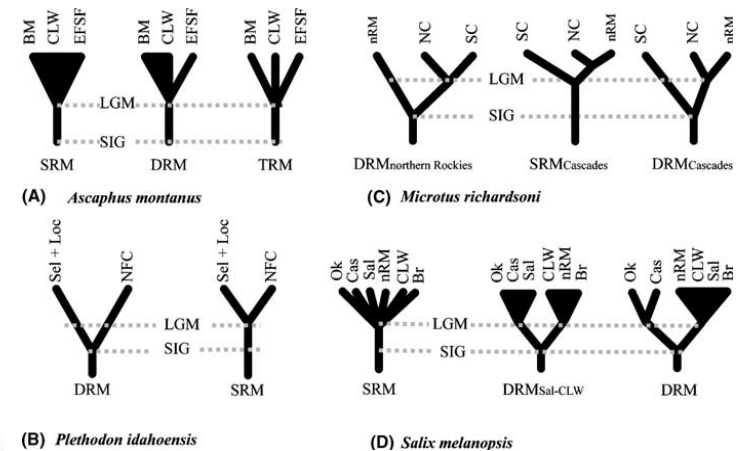
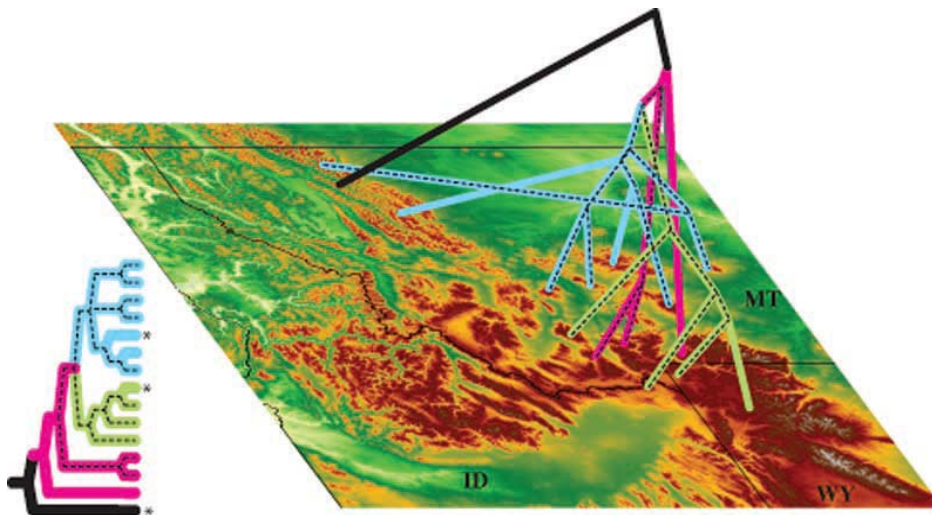
- Różne części genomu u osobników rozmnażających się płciowo mają różne historie
- Genealogie różnych genów z tej samej populacji będą się bardzo różnić z powodów losowych



Dlatego, ponieważ zazwyczaj interesują nas parametry **dla populacji** a nie poszczególnych genów, **lepiej dodać więcej genów niż więcej osobników**

Koalescencja jest niezastąpiona w symulacjach

- symuluje się tylko genealogię próby, nie zaś całej populacji
- wystarczy mała próba (5-10 sekwencji na populację)
- można symulować wiele sytuacji i sprawdzać która jest bardziej prawdopodobna – testowanie hipotez



Uzyskanie wartości parametrów (np. N_e) byłoby łatwe

- gdybyśmy znali prawdziwą genealogię
- moglibyśmy sprawdzić różne wartości parametrów i wybrać ten dla którego nasza genealogia największe prawdopodobieństwo

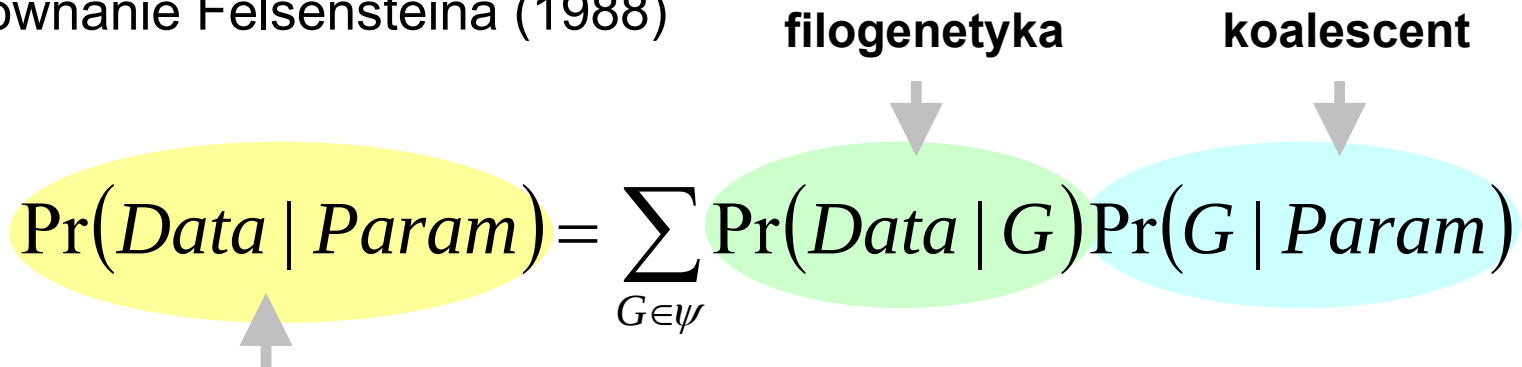
$$P(G \mid N) = \frac{1}{2N} \prod_i \exp\left(-u_i \frac{k(k-1)}{4N}\right)$$

- analogicznie do wielkości populacji moglibyśmy poznać tempo migracji czy rekombinacji

Ale nie znamy prawdziwej genealogii

Jak sobie poradzić z niepewnością genealogii

- Liczba możliwych drzew jest ogromna, dla 10 taksonów ponad 10 milionów, dla 20 $> 2.2 \times 10^{20}$
- Zmienność sekwencji w danych populacyjnych ograniczona
- A więc znaczna niepewność co do genealogii
- Rozwiązanie: weźmy pod uwagę wszystkie drzewa, każde w proporcji odpowiadającej ich prawdopodobieństwu
- Równanie Felsensteina (1988)

$$\Pr(Data | Param) = \sum_{G \in \psi} \Pr(Data | G) \Pr(G | Param)$$


Wiarygodność danych

Jak to zrobić?

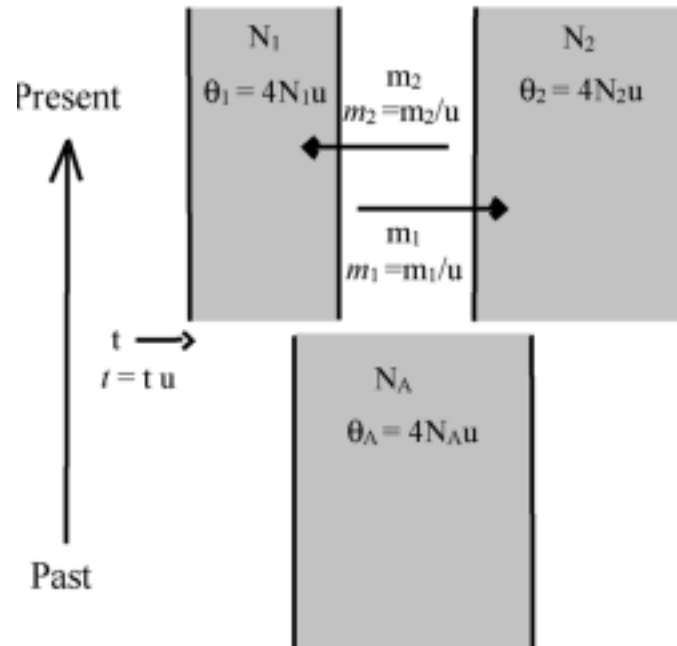
Łańcuchy Markova Monte Carlo

- większość drzew jest bardzo mało prawdopodobna, prawie nie będą uwzględnione
- skupiamy uwagę na najbardziej prawdopodobnych drzewach
- można przeprowadzić taką symulację komputerową która pobiera genealogie i wartości parametrów z łącznego rozkładu genealogii i parametrów warunkowanego danymi – algorytm Metropolis-Hastings
- intensywne obliczeniowo: dni, tygodnie, miesiące...

Genetyka dywergencji populacji (DPG)

- blisko spokrewnione gatunki lub populacje jednego gatunku
- kiedy się rozdzieliły?
- jakie mają wielkości populacji?
- jaka była wielkość populacji-przodka z której powstały?
- czy istniał między nimi przepływ genów od czasu gdy się rozeszły od wspólnego przodka? w którym kierunku?
- efekt sortowania linii może przypominać hybrydyzację

Model DPG

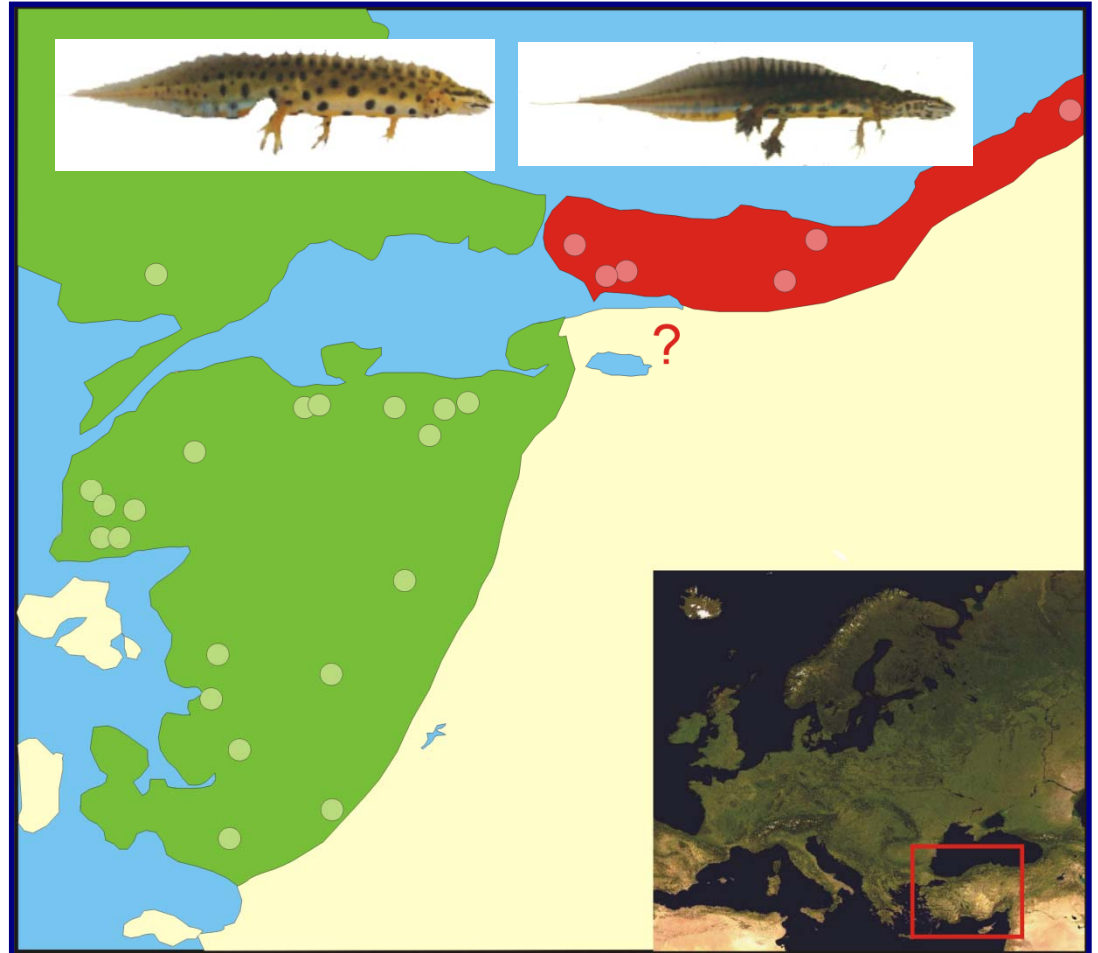


6 parametrów: t , Θ_1 , Θ_2 , Θ_A , m_1 , m_2

MCMC pobiera próby z rozkładu:

$$f(\text{Param} \mid X_1, X_2, \dots, X_n) = cf(\text{Param}) \prod_{i=1}^n \int_{G_i} f(X_i \mid G_i) f(G_i \mid \text{Param}) dG_i$$

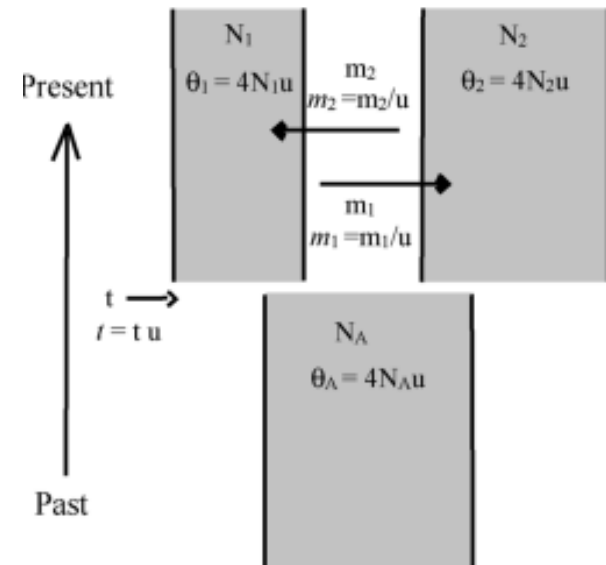
Historia ewolucyjna traszki zwyczajnej w Turcji



Historia ewolucyjna traszki zwyczajnej w Turcji

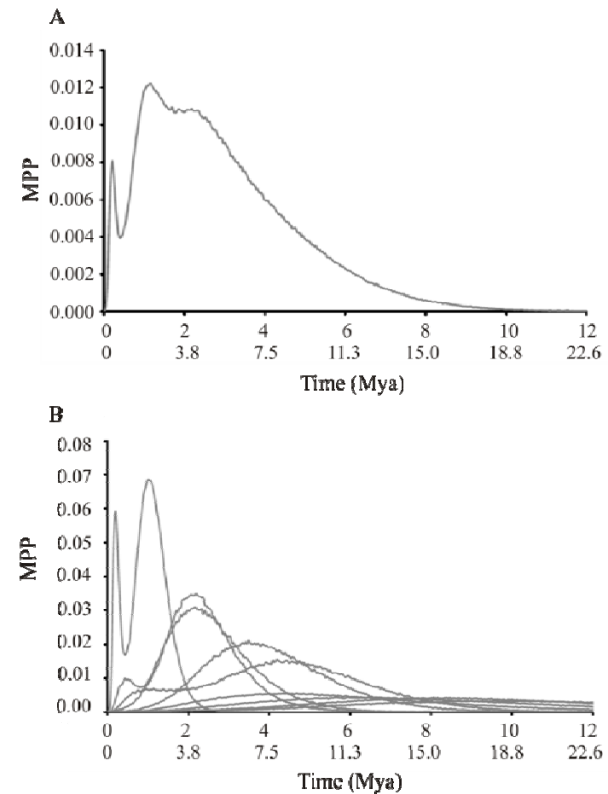
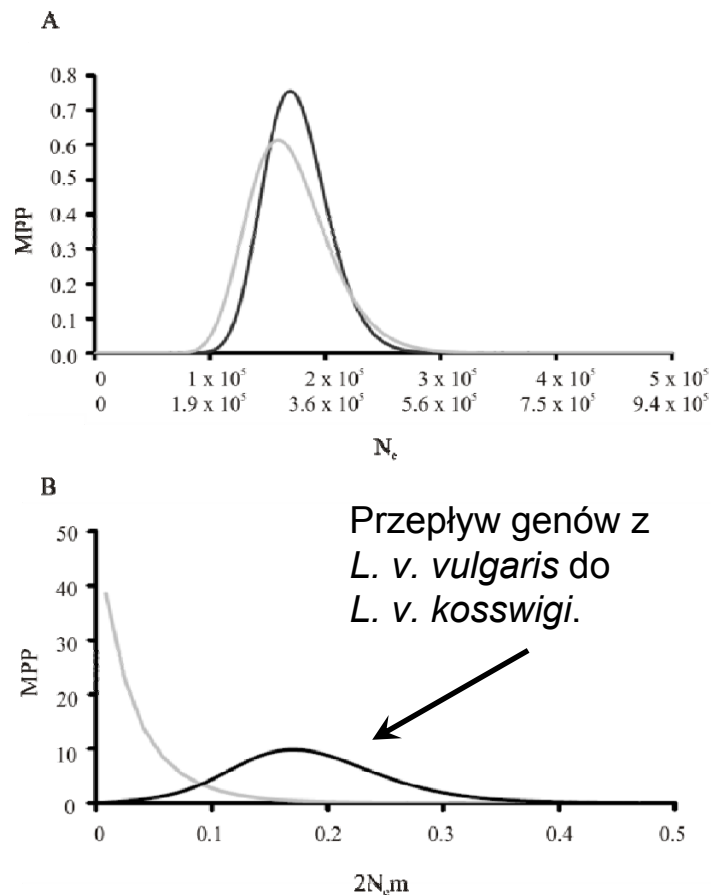
- Jakie są wielkości obu populacji (podgatunków)?
- Jaka była wielkość populacji ancestralnej?
- Czy dochodziło (nadal dochodzi) do przepływu genów między podgatunkami?
- W którym kierunku?

- 9 genów
- model IM



Historia ewolucyjna traszki zwyczajnej w Turcji

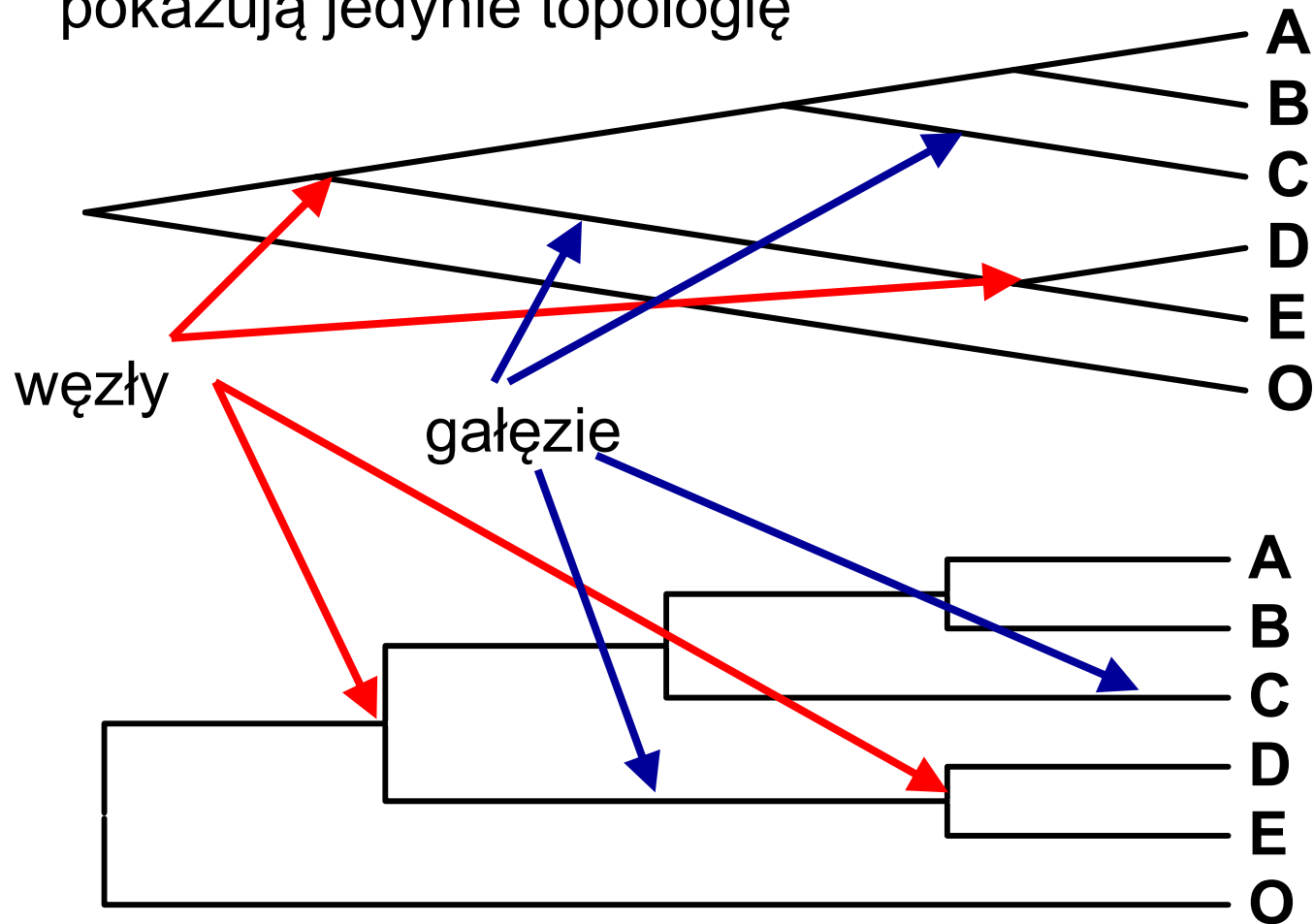
- Zbliżone długoterminowe wielkości populacji
- Asymetryczna introgresja



- Zarówno dawny jak i niedawny (współczesny?) przepływ genów

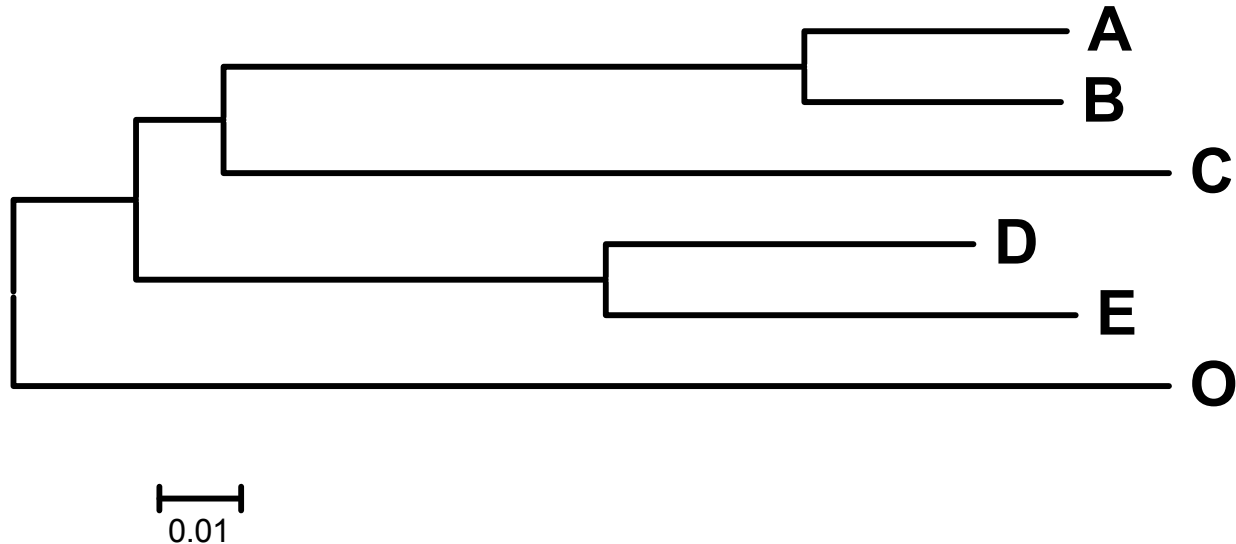
Drzewa filogenetyczne

oba drzewa przedstawiają dokładnie to samo
pokazują jedynie topologię



Drzewa filogenetyczne

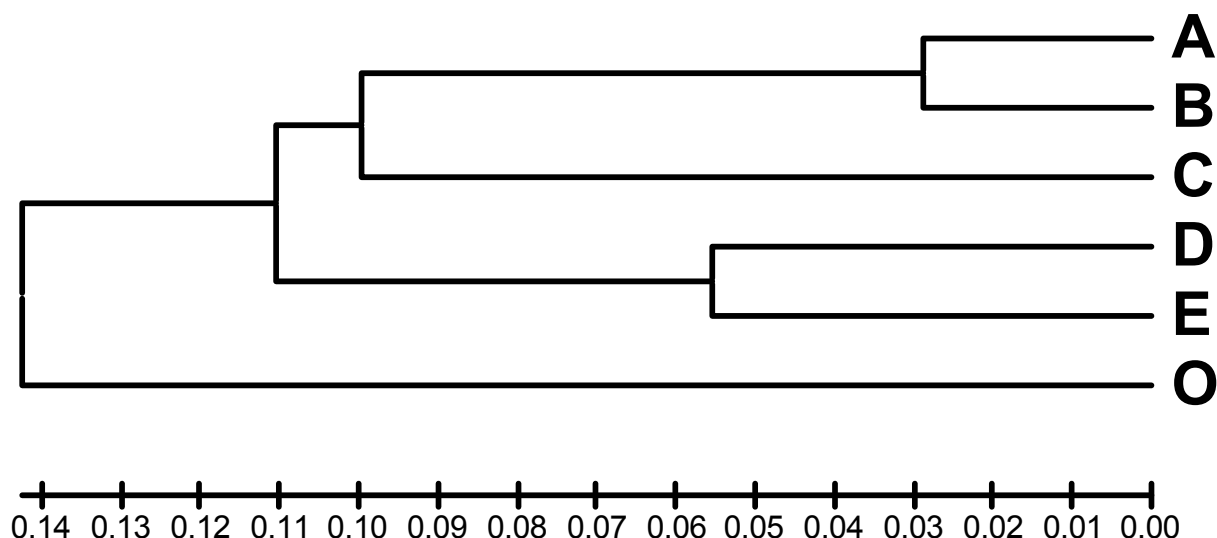
na tym drzewie zaznaczono **długości gałęzi**
czyli „**ilość ewolucji**” wzdłuż każdej gałęzi



drzewo addytywne

Drzewa filogenetyczne

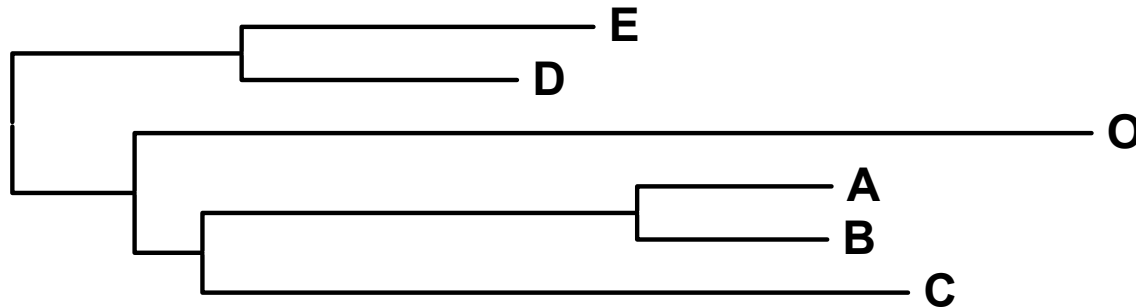
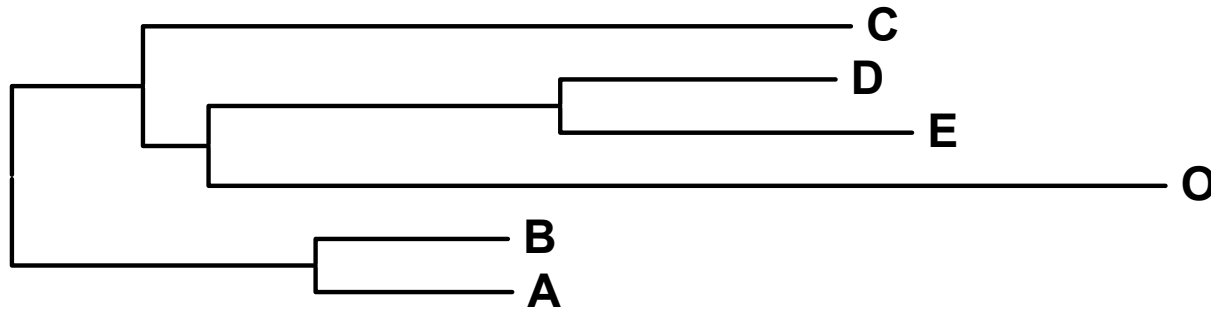
na tym drzewie też zaznaczono długości gałęzi
lecz założono że każdy takson jest równie odległy
od wspólnego przodka – **długości gałęzi pokazują czas**



drzewo ultrametryczne

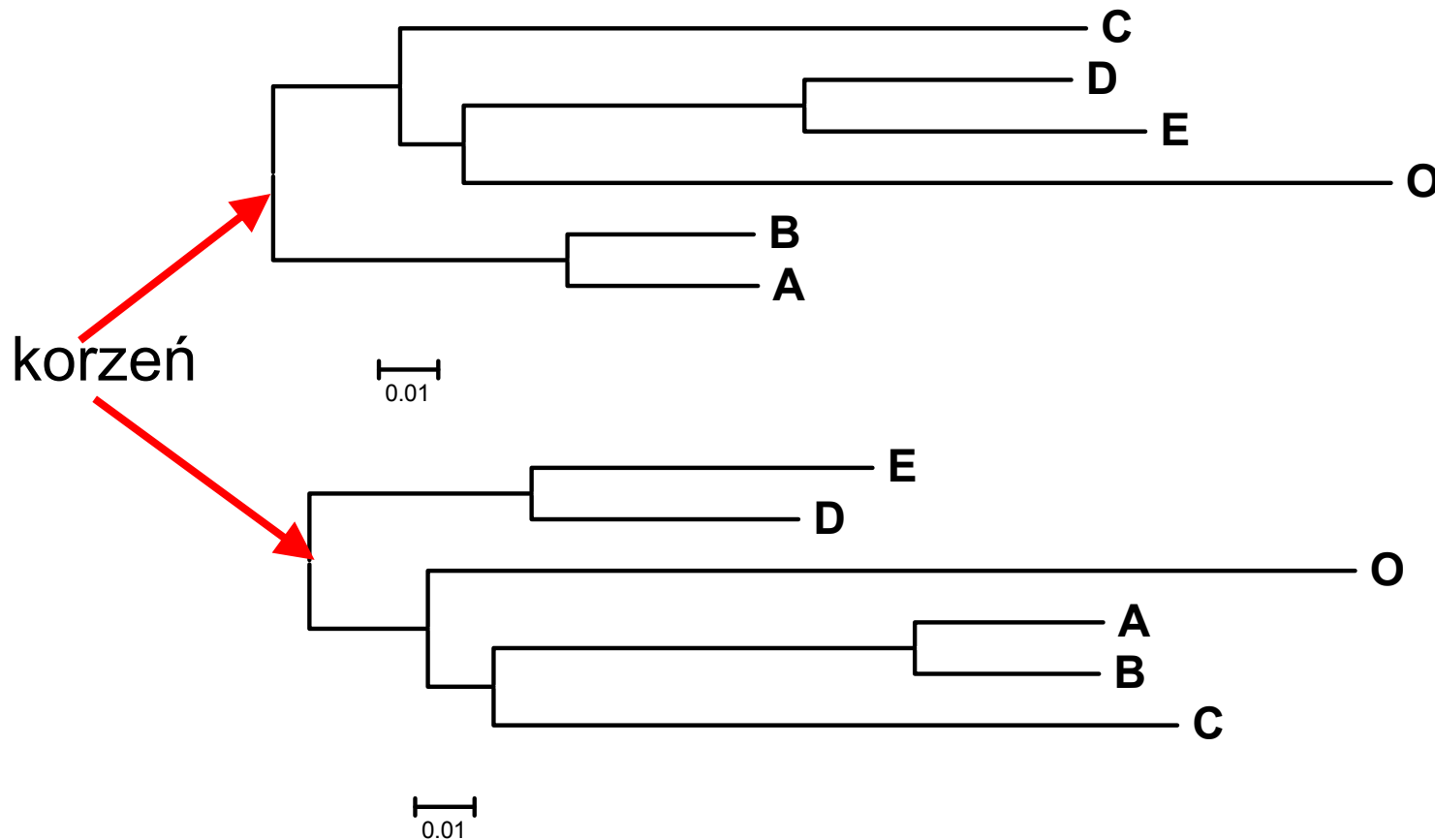
Drzewa filogenetyczne

drzewa **nieukorzenione**, oba są identyczne
„początek” lub korzeń można umieścić
w dowolnym miejscu



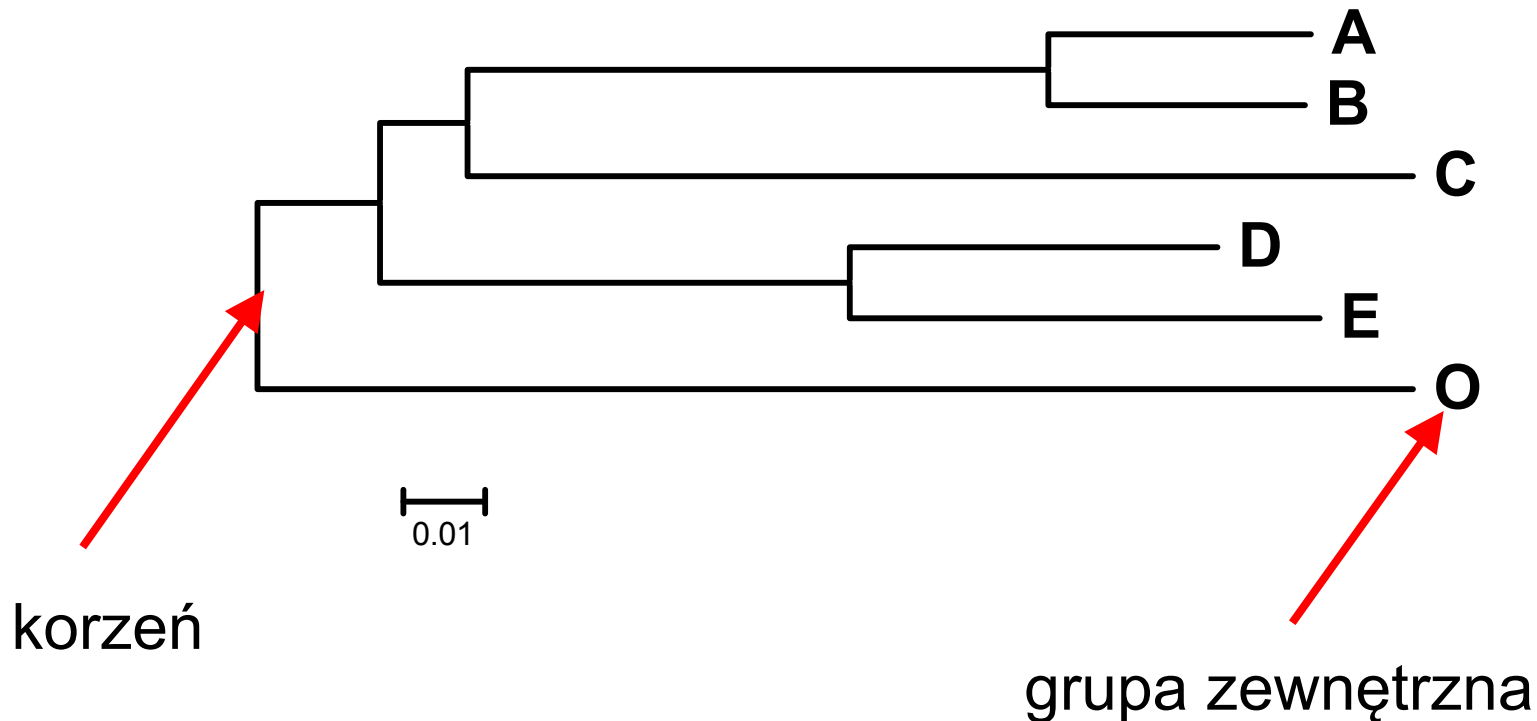
Drzewa filogenetyczne

drzewa **ukorzenione** –
drzewa poniżej są bardzo różne,
bo mają korzeń w innym miejscu



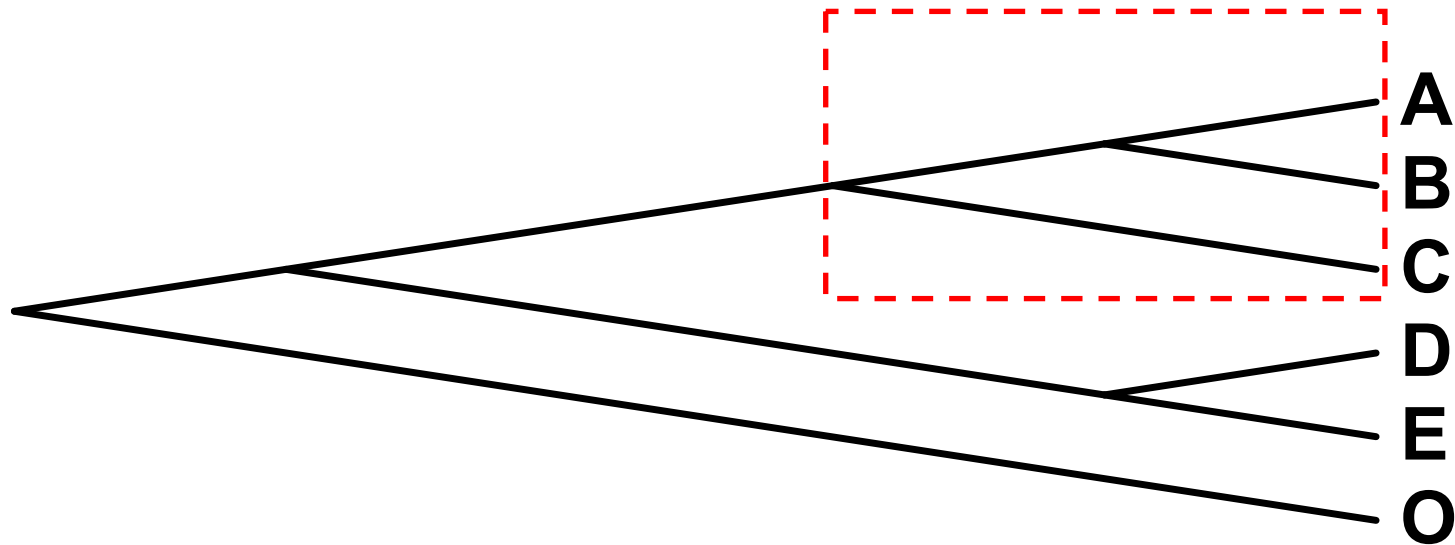
Drzewa filogenetyczne

Najczęściej drzewo ukorzenia się
grupą zewnętrzną



Drzewa filogenetyczne

Mono-, para- i polifiletyzm

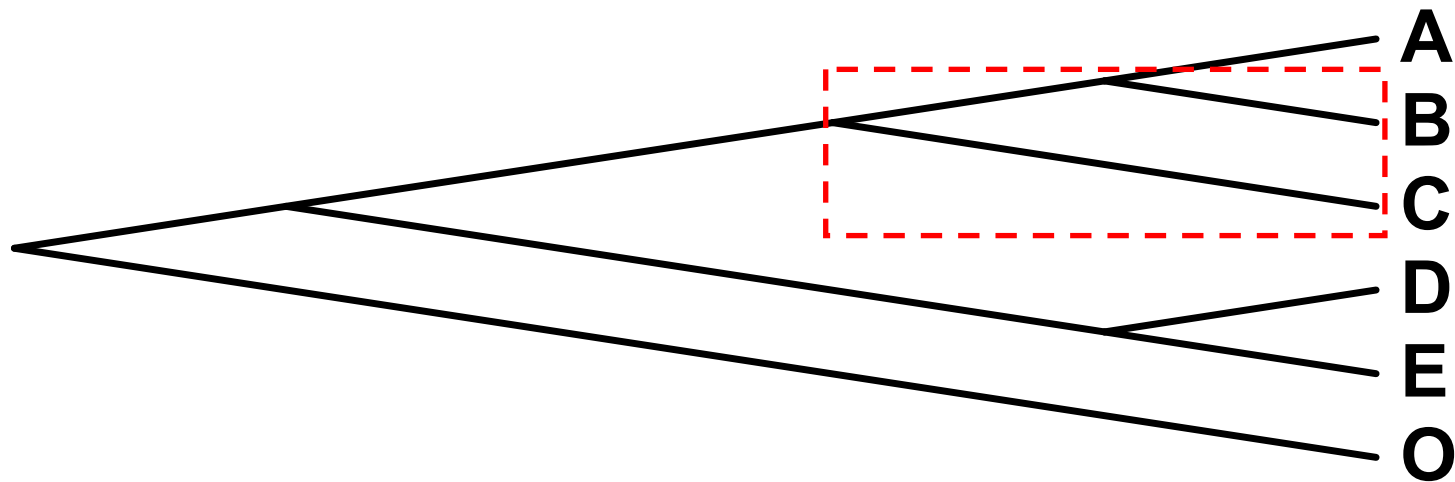


grupa monofiletyczna – obejmuje wspólnego przodka i **wszystkich** jego potomków

jedynie grupy monofiletyczne są **naturalne**

Drzewa filogenetyczne

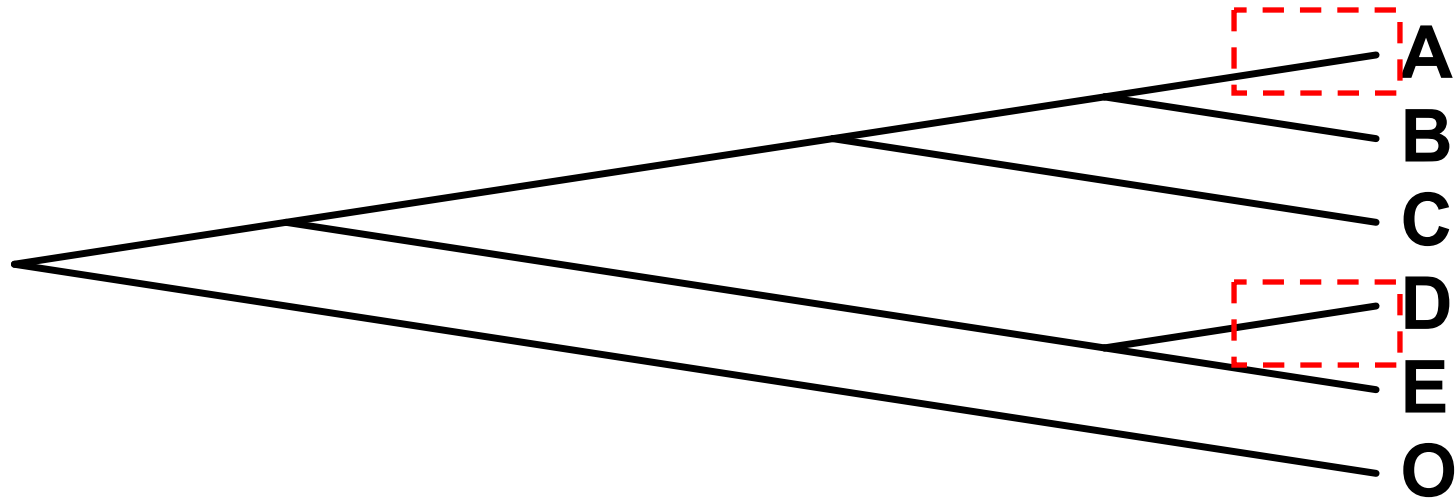
Mono-, para- i polifiletyzm



grupa parafiletyczna – obejmuje wspólnego przodka i tylko niektórych jego potomków

Drzewa filogenetyczne

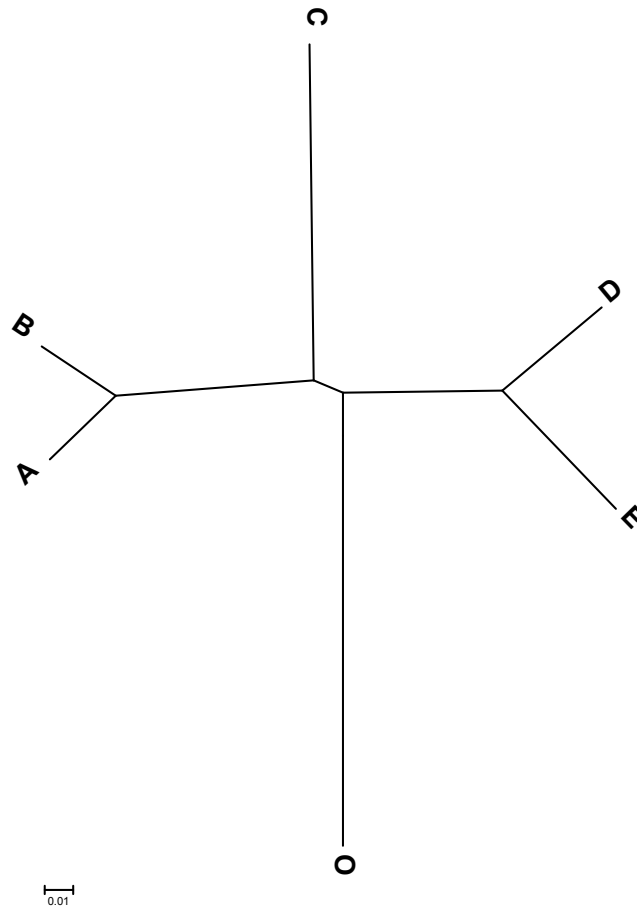
Mono-, para- i polifiletyzm



grupa polifiletyczna – gdy grupuje się razem taksony które miały różnych przodków

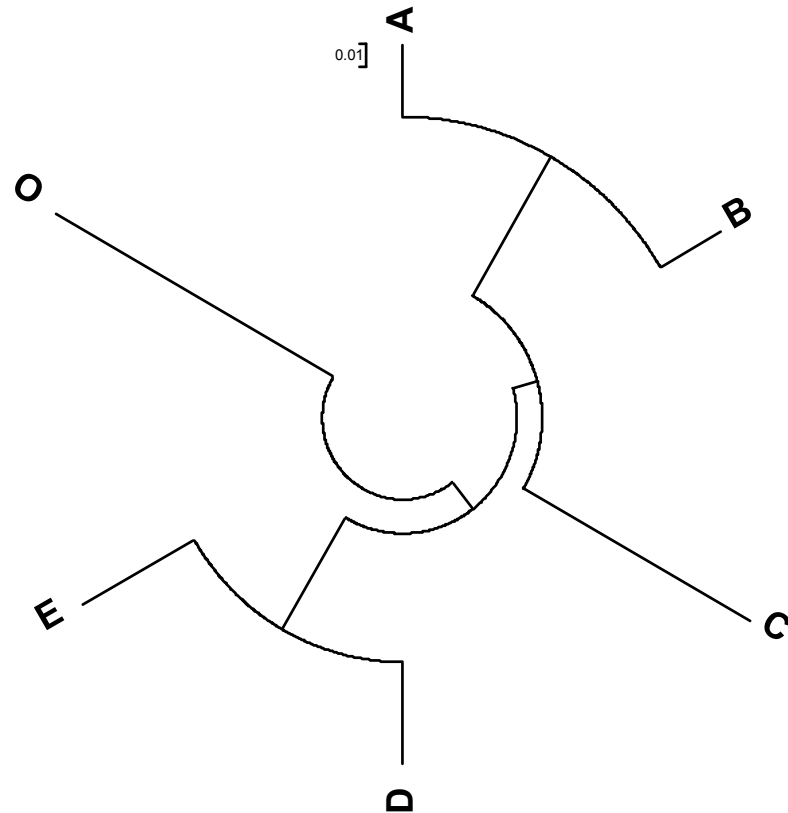
Drzewa filogenetyczne

to samo drzewo filogenetyczne można narysować
na wiele sposobów



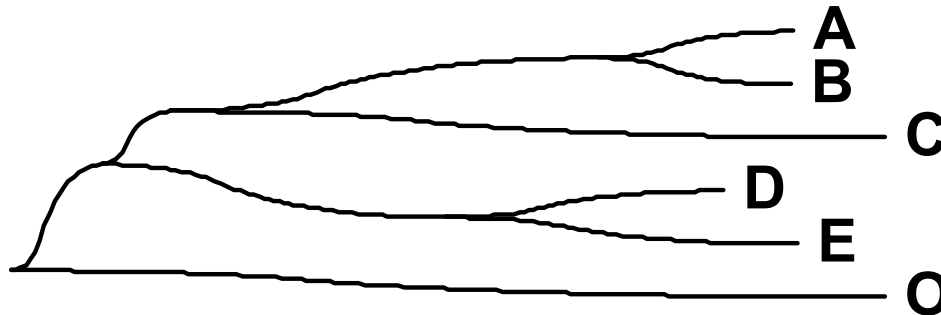
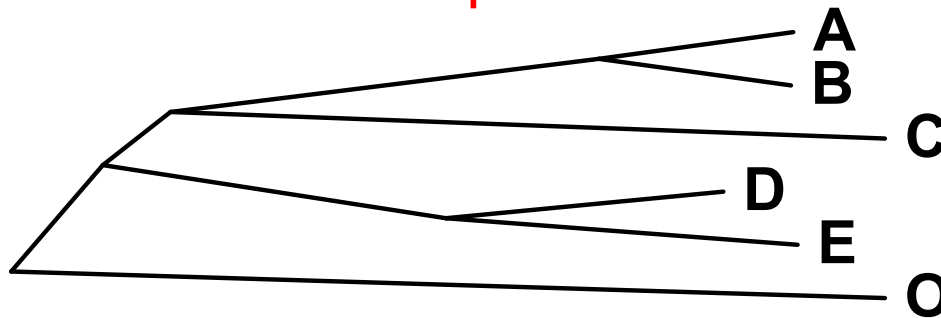
Drzewa filogenetyczne

to samo drzewo filogenetyczne można narysować
na wiele sposobów



Drzewa filogenetyczne

to samo drzewo filogenetyczne można narysować
na wiele sposobów



Drzewa filogenetyczne

- obrazują pokrewieństwa taksonów, opierają się na podobieństwie
- drzewa mogą być ukorzenione lub nieukorzenione
- to samo drzewo można przedstawić (narysować na różne sposoby)
- układ rozgałęzień na drzewie to topologia
- istnieje gigantyczna liczba możliwych topologii, dla 10 taksonów ponad 10 milionów, dla 20 – 2×10^{20}

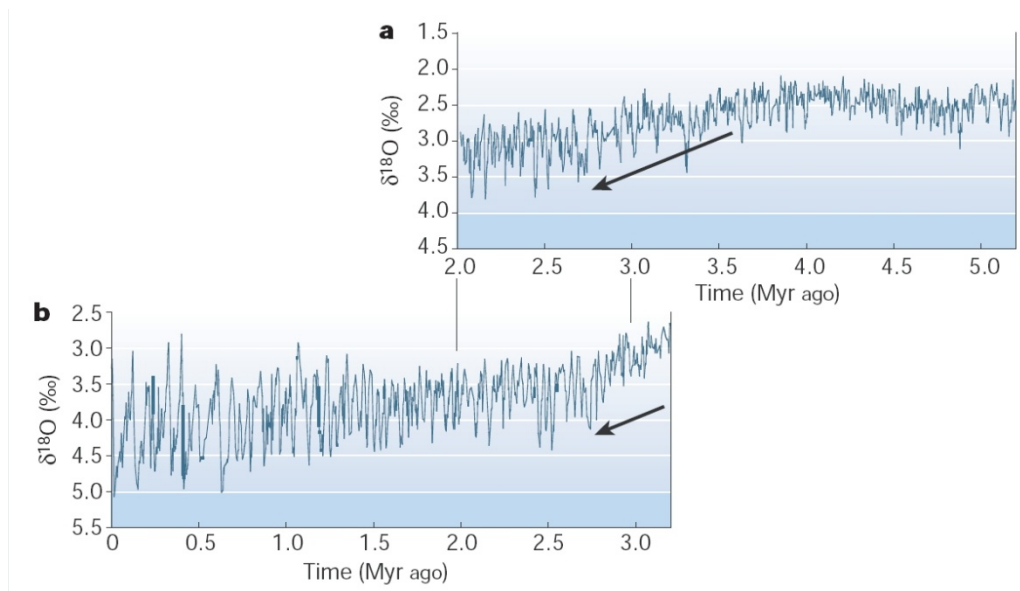
Drzewa filogenetyczne konstruuje się z różnych typów danych

- cechy morfologiczne
- cechy behawioralne
- cechy historii życiowych
- częstości alleli (np. allozymy, mikrosatelity)
- wzorce cięcia enzymów restrykcyjnych
- sekwencje białek
- sekwencje DNA/RNA

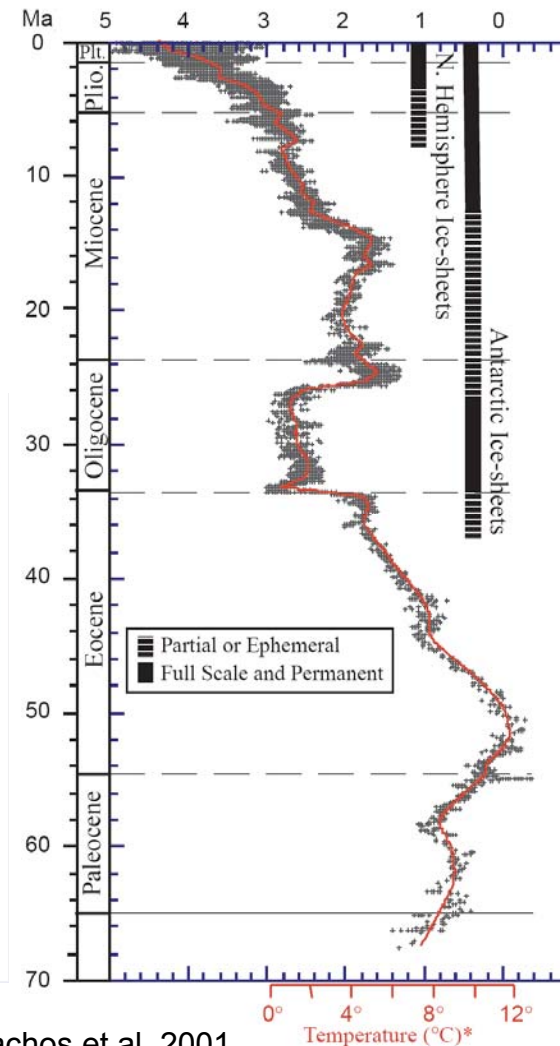
Obecnie zdecydowanie najważniejsze są sekwencje białek i DNA

Zmiany klimatyczne w przeszłości

- stopniowe ochładzanie klimatu w Kenozoiku
- początek zlodowaceń Plejstoceńskich ok 2.7 - 2.4 mln lat temu



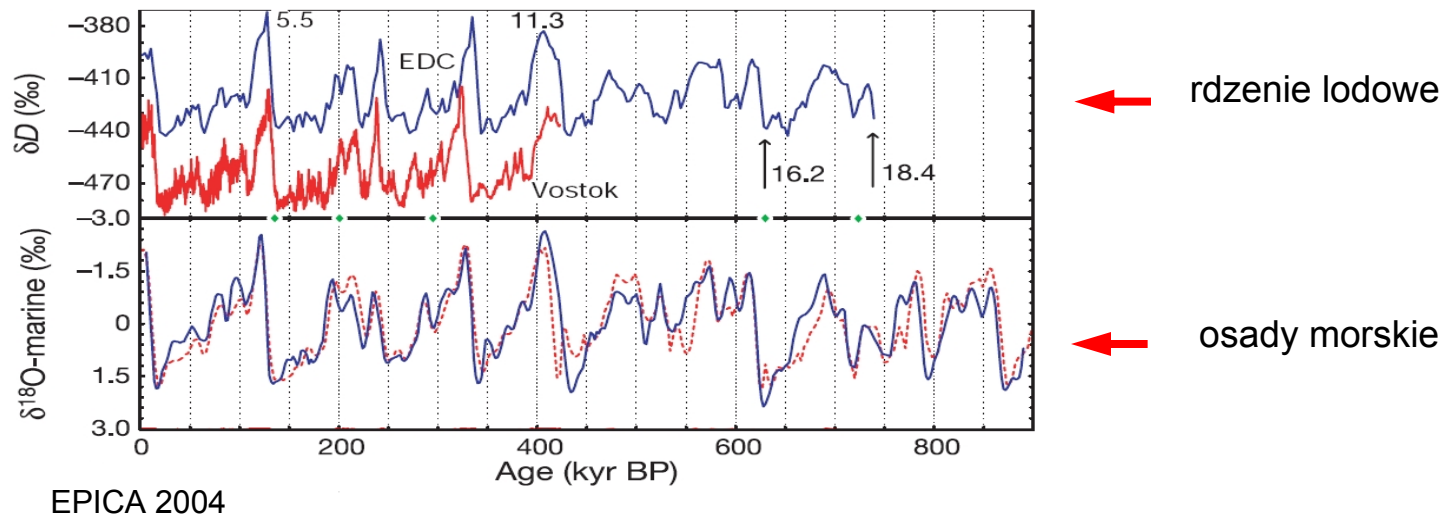
Lambeck et al. 2002



Zachos et al. 2001

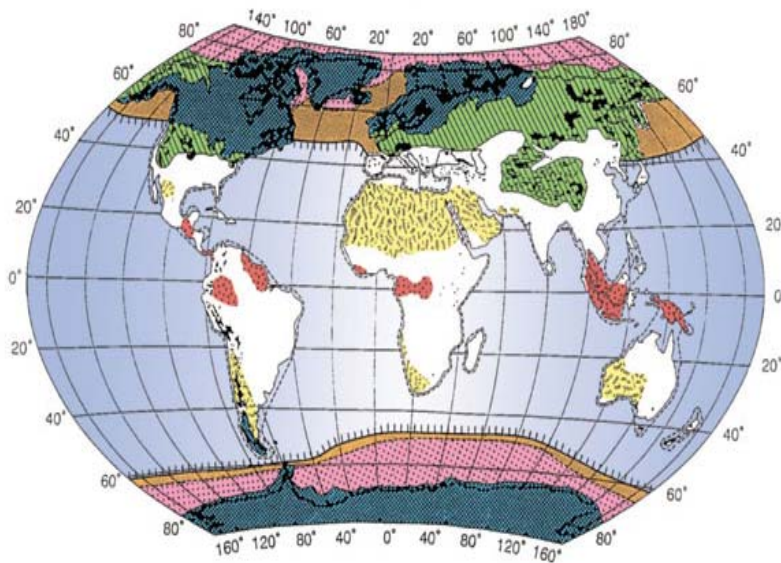
Zmiany klimatyczne w przeszłości

- Cykliczne zmiany klimatu – długie glacjały i krótkie interglacjały rozpoczęły się ok. 2.7 -2.4 mln lat temu
- rekonstrukcja paleotemperatury z rdzeni lodowych Antarktyki i Grenlandii oraz z osadów morskich

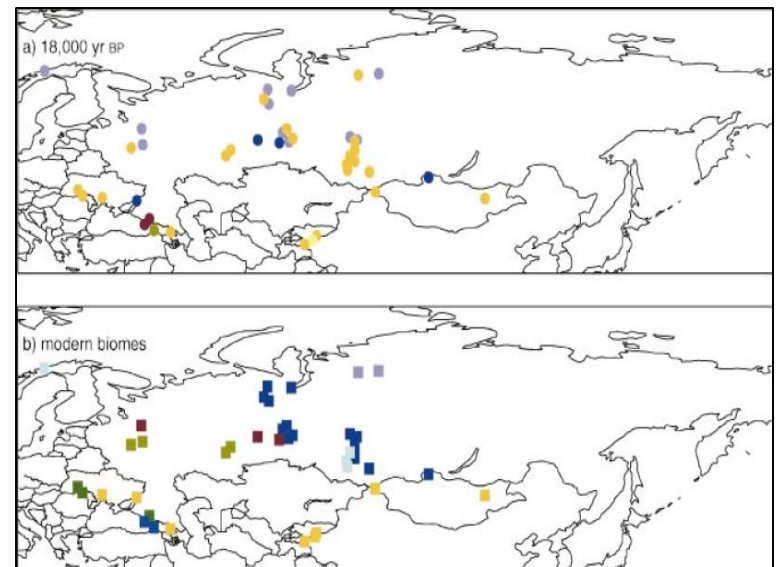


Zmiany klimatyczne w przeszłości

- Ostatnie zlodowacenie zaczęło się ok. 115-110 tys. lat temu
- Jego najzimniejsza część, ostatnie maksimum lodowcowe (LGM) ok. 25-15 ky temu



Hewitt 2000



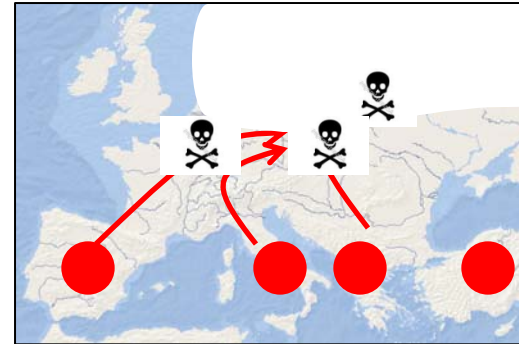
Tarasov et al. 2000

Organizmy wobec zmian klimatycznych

- dramatyczne zmiany klimatu mogą zajść szybko, nawet w okresach poniżej 100 lat
- zmiany mogą być zbyt szybkie aby organizmy mogły się do nich przystosować
- organizmy muszą zmieniać zasięg wraz podążając za swoim środowiskiem lub wymrzeć
- wahania klimatu w plejstocenie dotknęły szczególnie gatunki klimatu umiarkowanego
- ich historię w plejstocenie charakteryzowało cykliczne rozszerzanie i kurczenie się zasięgów

Organizmy wobec zmian klimatycznych

Populacje z południowych półwyspów rozszerzały swój zasięg w miarę poprawy klimatu



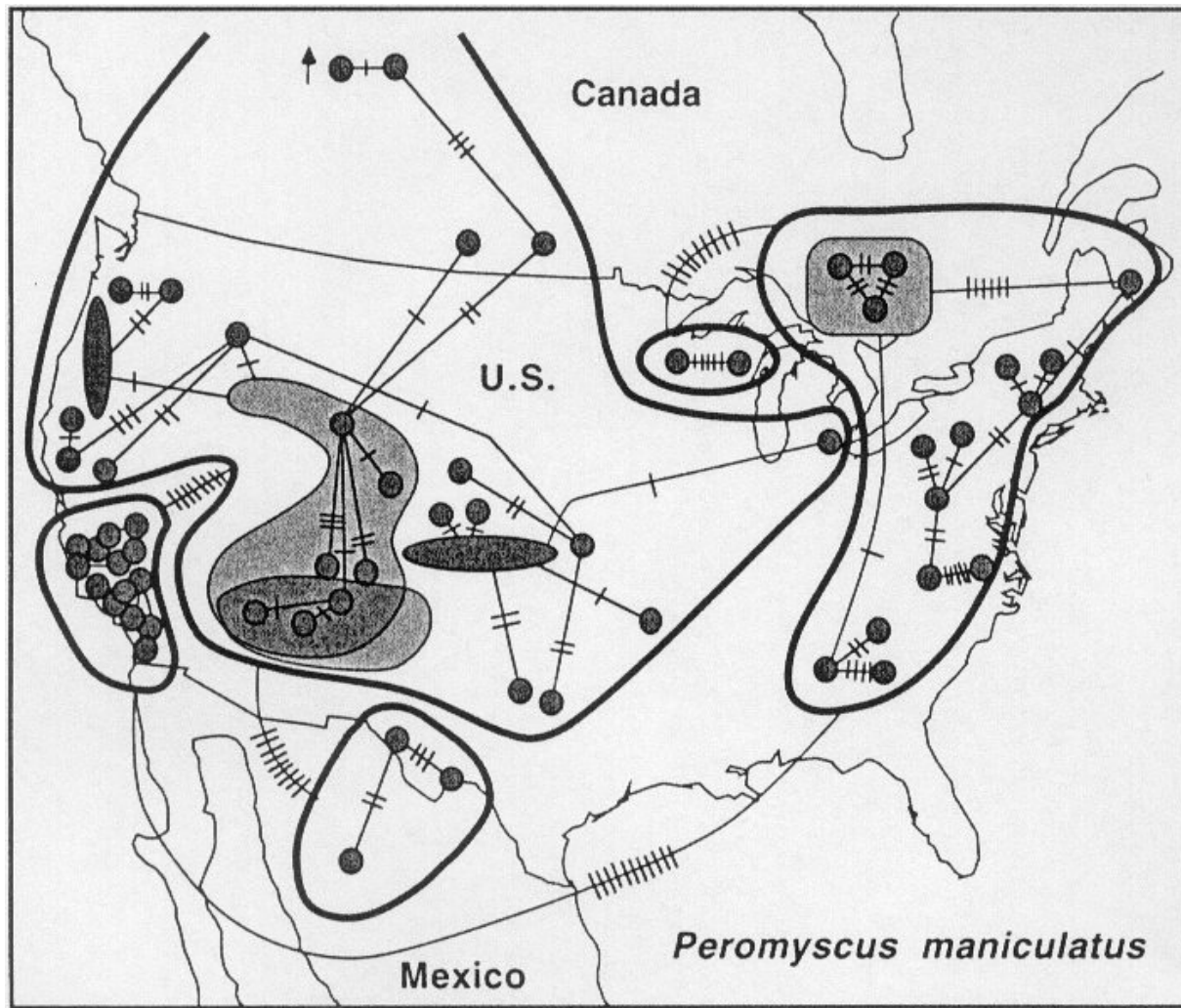
A w następnym cyklu lodowcowym populacje na północy wymierały

W efekcie populacje stały się obecne jedynie w refugiach – tam najwięcej zmienności genetycznej

Dlaczego zwierzęcy mtDNA?

- przed wprowadzeniem PCR łatwo dostępny do badań, tworzono mapy restrykcyjne całego mtDNA,
- PCR – uniwersalne startery i sekwencjonowanie,
- brak rekombinacji,
- dziedziczony tylko po matce – genealogie linii matczynych,
- szybkie tempo ewolucji sekwencji,
- efektywna wielkość populacji (N_e) nawet 4 razy mniejsza niż dla genów jądrowych – szybkie sortowanie linii,
- większość mutacji neutralna

Rozmieszczenie linii mtDNA a geografia

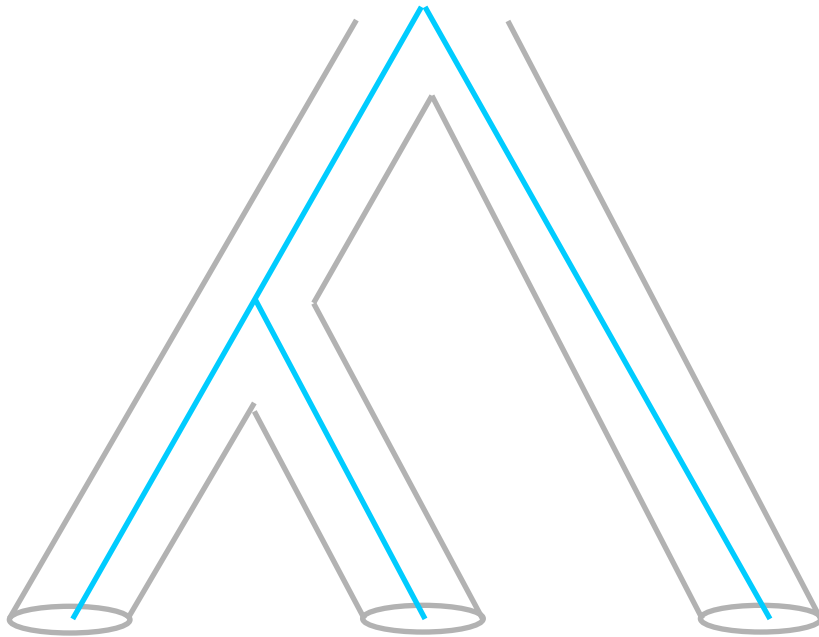


Lansman et al. 1983

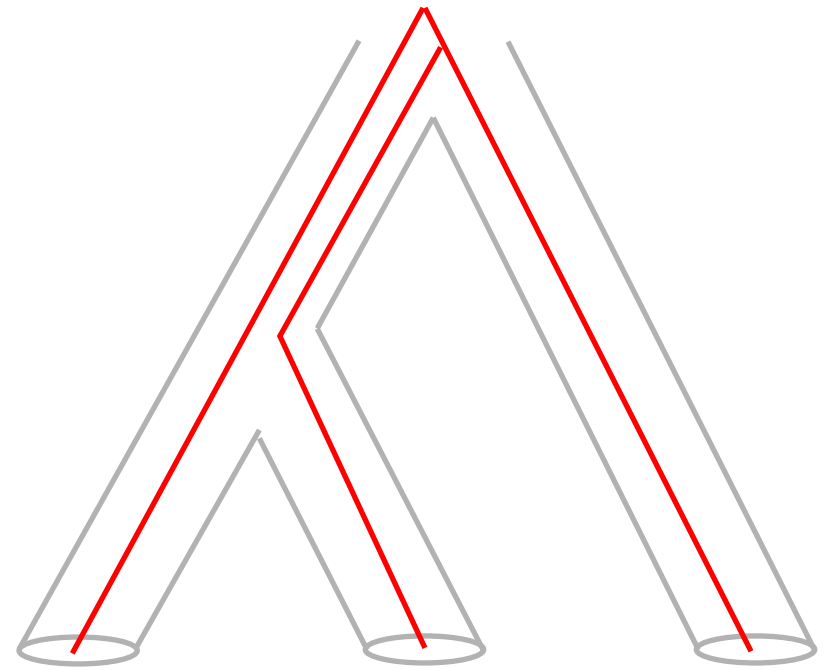
Kłopoty z mtDNA

- cały mtDNA to jeden „locus”,
- filogeneza genów nie jest tym samym co filogeneza organizmów – badanie jednego genu nie pozwala na wnioskowanie o historii organizmów,
- dane jedynie o liniach matczynych,
- działanie doboru – utrwalenie się korzystnej mutacji całkowicie wymazuje z mtDNA informacje dotyczące historii populacji przed utrwaleniem – *selective sweeps*,
- pseudogeny jądrowe – numtDNA, numts.

Drzewa genów nie są drzewami gatunków



drzewo genu
identyczne jak drzewo
gatunków



drzewo genu różne od
drzewa gatunków,
polimorfizm w gatunku
ancestralnym

Kłopoty z mtDNA

- cały mtDNA to jeden „locus”,
- filogeneza genów nie jest tym samym co filogeneza organizmów – badanie jednego genu nie pozwala na wnioskowanie o historii organizmów,
- dane jedynie o liniach matczynych,
- działanie doboru – utrwalenie się korzystnej mutacji całkowicie wymazuje z mtDNA informacje dotyczące historii populacji przed utrwaleniem – wymiatanie zmienności przez dobór (*selective sweeps*),
- pseudogeny jądrowe – numtDNA, numts

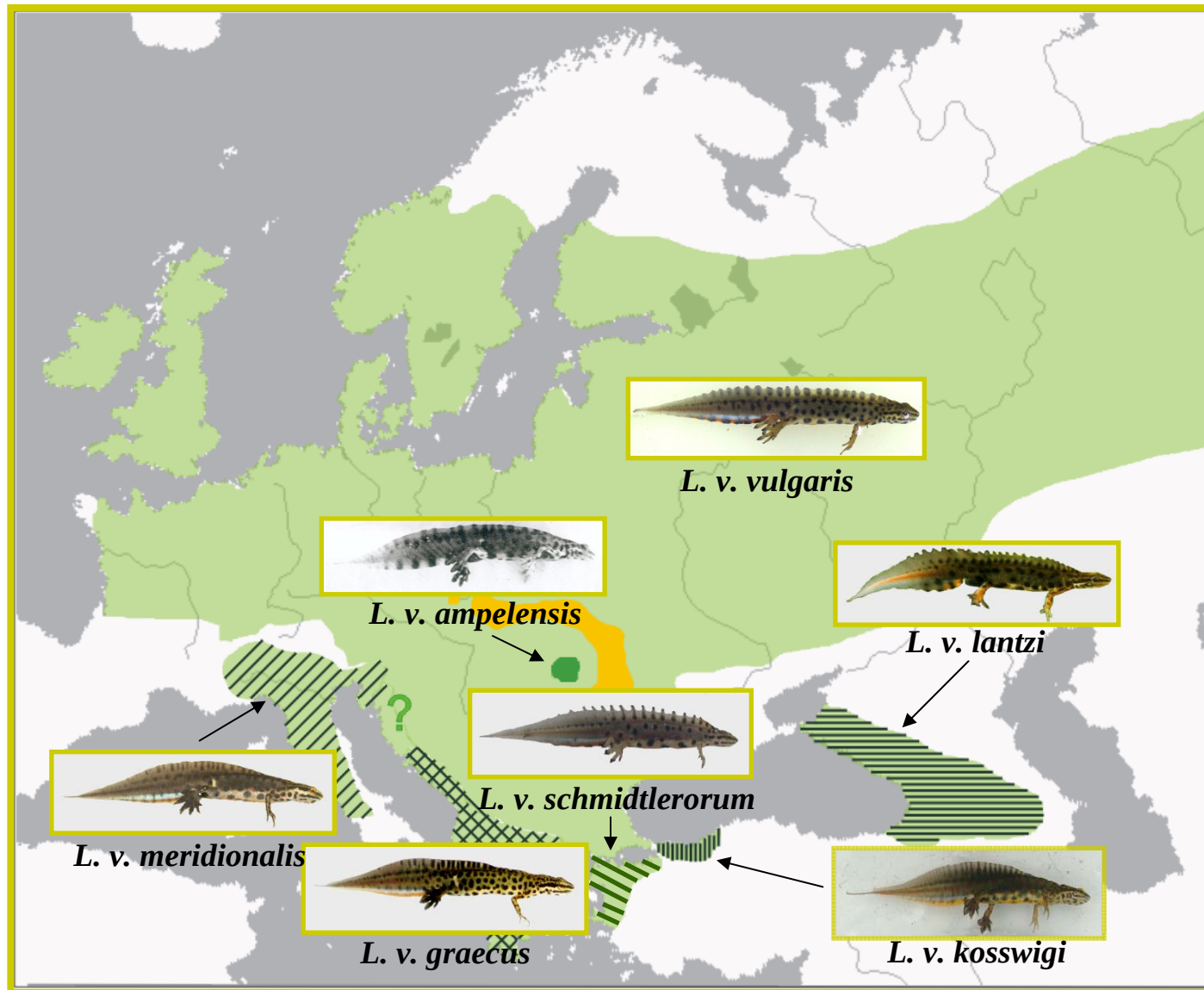
Sekwencje genów jądrowych w filogeografii

Są niezbędne i szeroko stosowane, ale:

- trudności techniczne, występowanie rodzin genów,
- rekombinacja – trudności z analizą,
- wolniejsze tempo ewolucji – niewielka ilość informacji,
- trudno ocenić wpływ doboru,
- większa efektywna wielkość populacji – wolniejsze sortowanie linii.

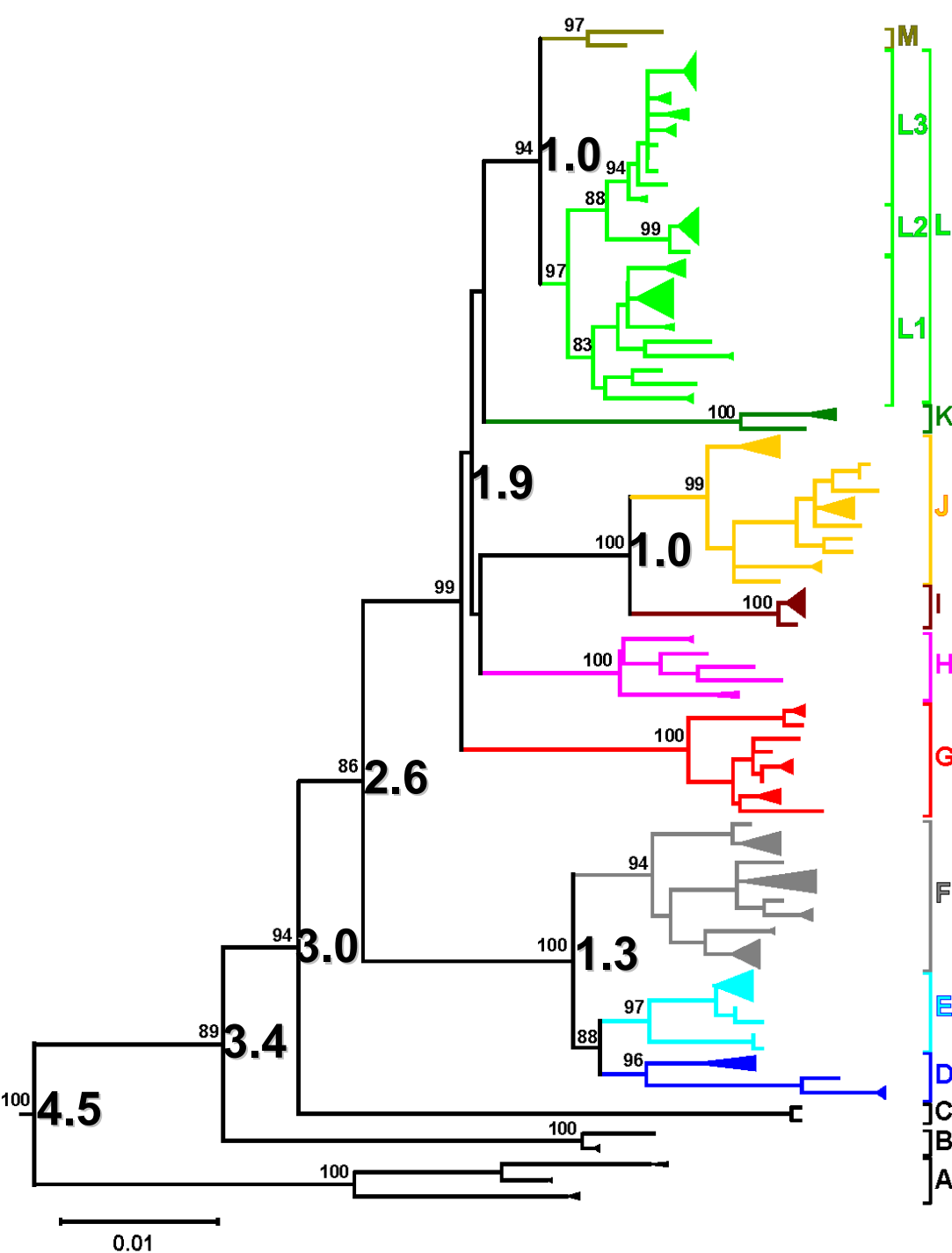
Dlatego mtDNA jest nadal podstawowym markerem
dla **wstępnej** charakterystyki
wzorców filogeograficznych u zwierząt

Filogeografia traszek zwyczajnej i karpackiej



Filogeografia traszek zwyczajnej i karpackiej

Duże zróżnicowanie
mtDNA, wiele starych
grup

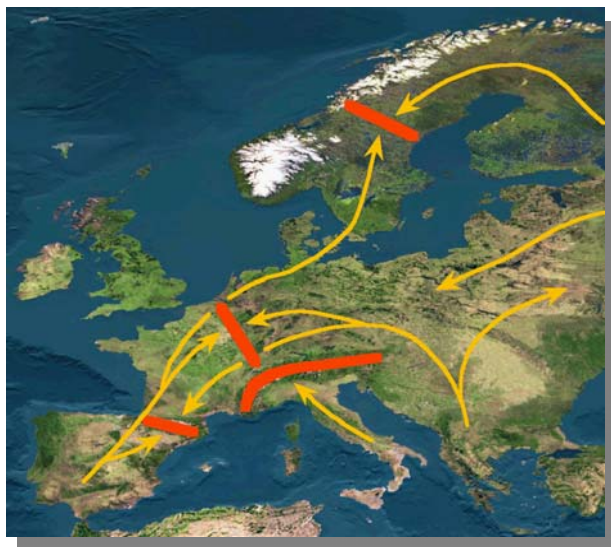


Filogeografia traszki zwyczajnej i karpackiej

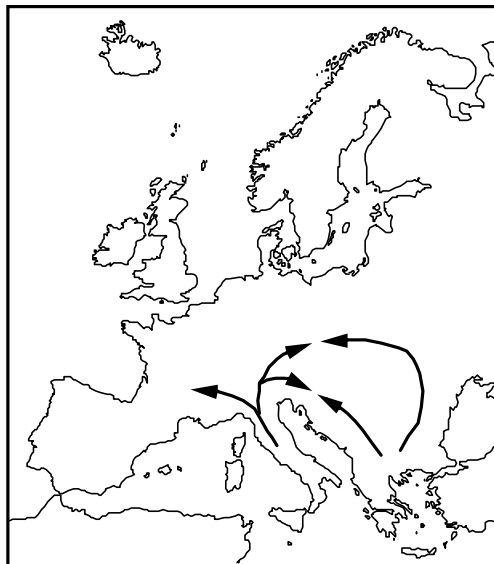


Rozmieszczenie linii mtDNA wskazuje że najważniejsze refugia znajdowały się w Europie Środkowej

Filogeografia porównawcza

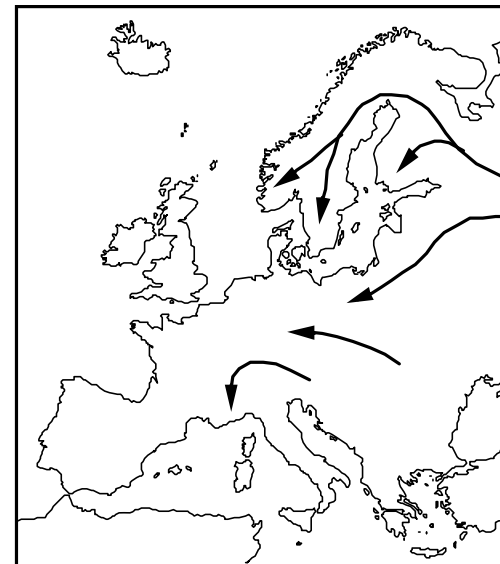


Abies alba (silver fir)



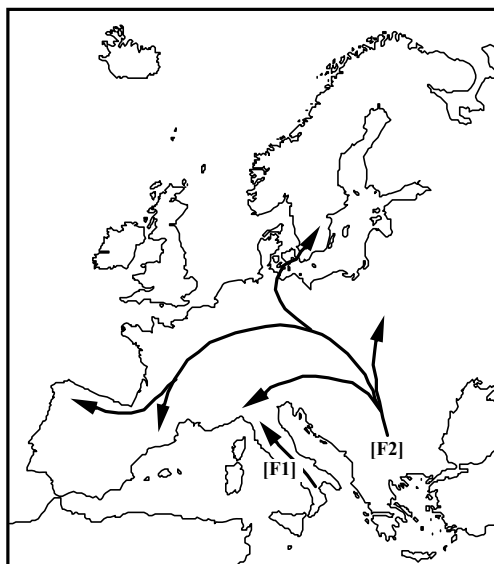
Ograniczona kolonizacja
z 2 refugium

Picea abies (Norway spruce)



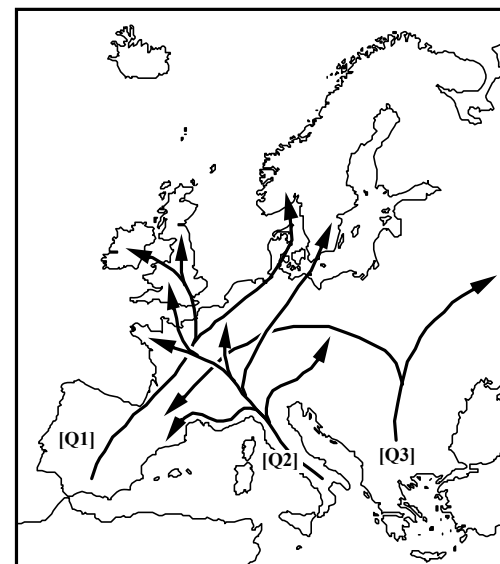
Kolonizacja ze Wschodu
(gat. zimnolubne)

Fagus sylvatica (common beech)



Kolonizacja z 1 refugium

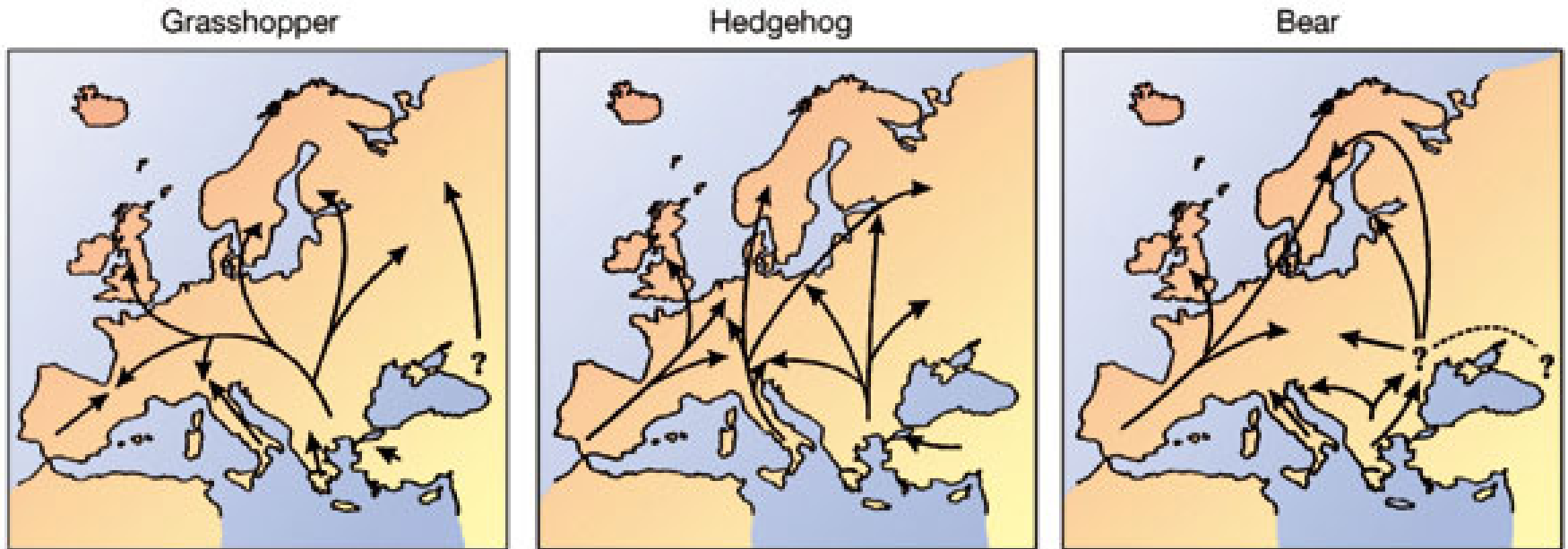
Quercus sp. (white oaks)



Kolonizacja z 3 refugium

Filogeografia porównawcza

Wzorce kolonizacji z różnych refugium
Gatunki reagowały na zmiany klimatyczne indywidualnie

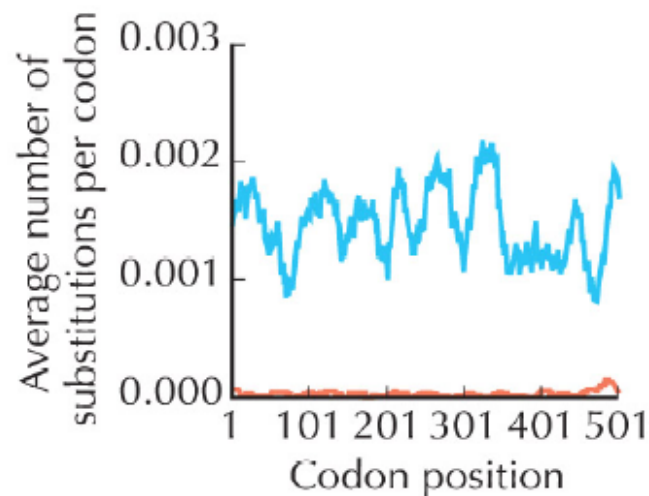


Dobór naturalny

- Zróżnicowane przeżywanie i reprodukcja osobników zależne od ich dziedzicznych (a więc warunkowanych genetycznie) właściwości
- Miarą „jakości” osobnika z punktu widzenia doboru naturalnego jest jego dostosowanie – mierzone całowyciowym sukcesem rozrodczym, zdolność do produkcji jak największej ilości jak najlepszej jakości płodnego potomstwa
- Tak jak mówimy o dostosowaniu osobnika możemy mówić o dostosowaniu genu
- **Dobór naturalny to jedyny znany mechanizm wytwarzający adaptacje**, właściwości organizmów umożliwiające im przystosowanie się do środowiska życia
- Dobór z pewnością ma zasadnicze znaczenie na poziomie organizmu, a jak jest na poziomie molekularnym?

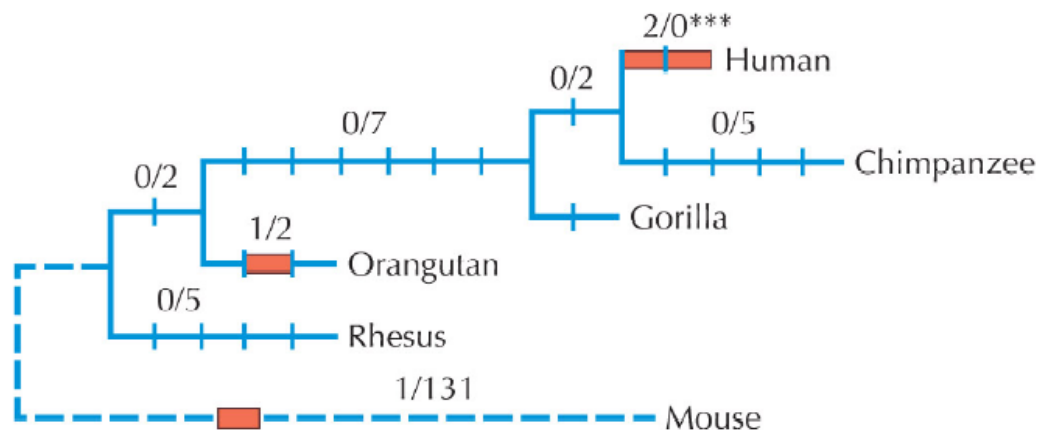
Wykrywanie doboru w genach kodujących białka

- Z powodu degeneracji kodu genetycznego mamy miejsca synonimowe i niesynonimowe
- Można obliczyć liczbę podstawień na miejsce synonimowe (dS) i na miejsce niesynonimowe (dN): dN/dS to ω
- Przy **neutralności** oczekujemy $\omega = 1$
- $\omega < 1$ czyli niedobór podstawień niesynonimowych wskazuje na **dobór oczyszczający**
- $\omega > 1$ czyli nadmiar podstawień niesynonimowych wskazuje na **dobór pozytywny**
- W zdecydowanej większości białek obserwujemy **dobór oczyszczający**, którego siła zależy od stopnia ograniczeń funkcjonalnych jakim podlega białko, takie ograniczenia są niezwykle silne np. w histonach



Wykrywanie doboru w genach kodujących białka

- Ale nawet w białkach ewoluujących pod działaniem doboru pozytywnego większość zmian aminokwasów będzie szkodliwych – dlatego trudno się spodziewać że średnia w będzie często > 1 choć znane są takie przypadki
- Można identyfikować, na podstawie analiz filogenetycznych, poszczególne kodony pod działaniem doboru pozytywnego
- Dobór pozytywny może działać na wiele genów lecz epizodycznie, przez krótki czas
- Przykładem może być gen *FOXP2* u człowieka

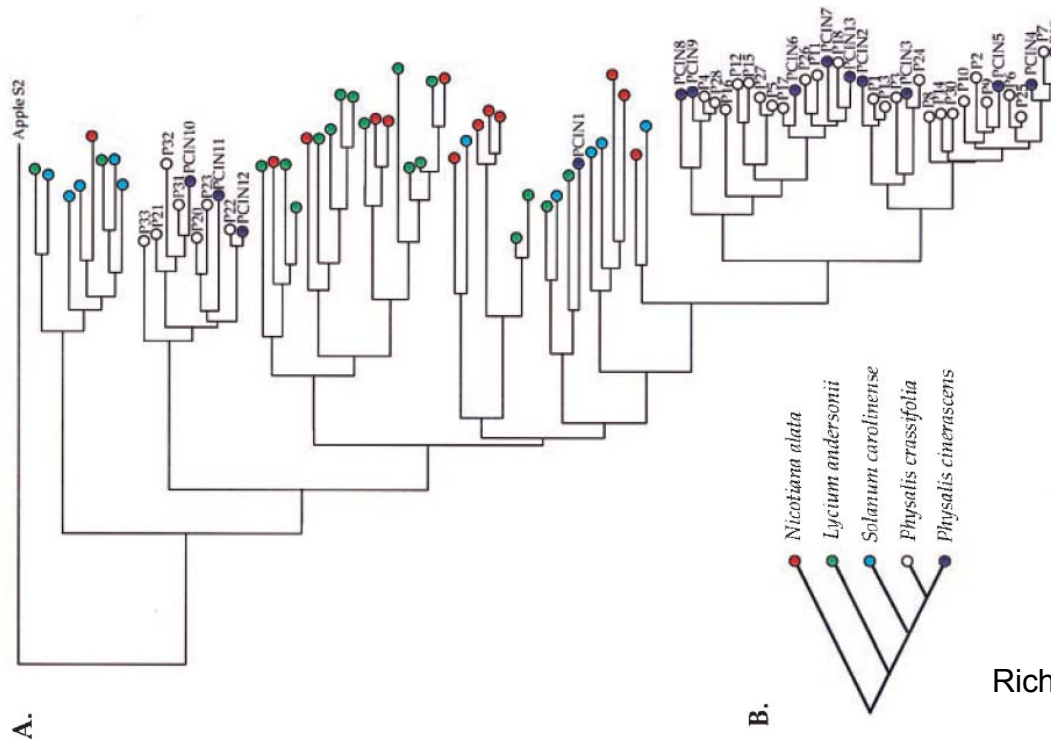


Wykrywanie doboru w genach kodujących białka

- Istnieją geny ewoluujące przede wszystkim pod działaniem doboru pozytywnego, związane są one często z:
 - unikaniem kojarzeń krewniaczych – systemy samoniezgodności u roślin
 - walką z patogenami – geny zgodności tkankowej (MHC) kręgowców
 - rozrodem – gatunkowo specyficzne interakcje plemnika i komórki jajowej, białka płynu nasiennego *Drosophila*
- Allele genów ewoluujących pod działaniem doboru pozytywnego są bardzo często również pod działaniem doboru równoważącego, który utrzymuje w populacjach wiele bardzo starych i bardzo różniących się od siebie alleli
- Takie linie alleli są często starsze niż gatunki

Dobór równoważący i polimorfizm transspecyficzny

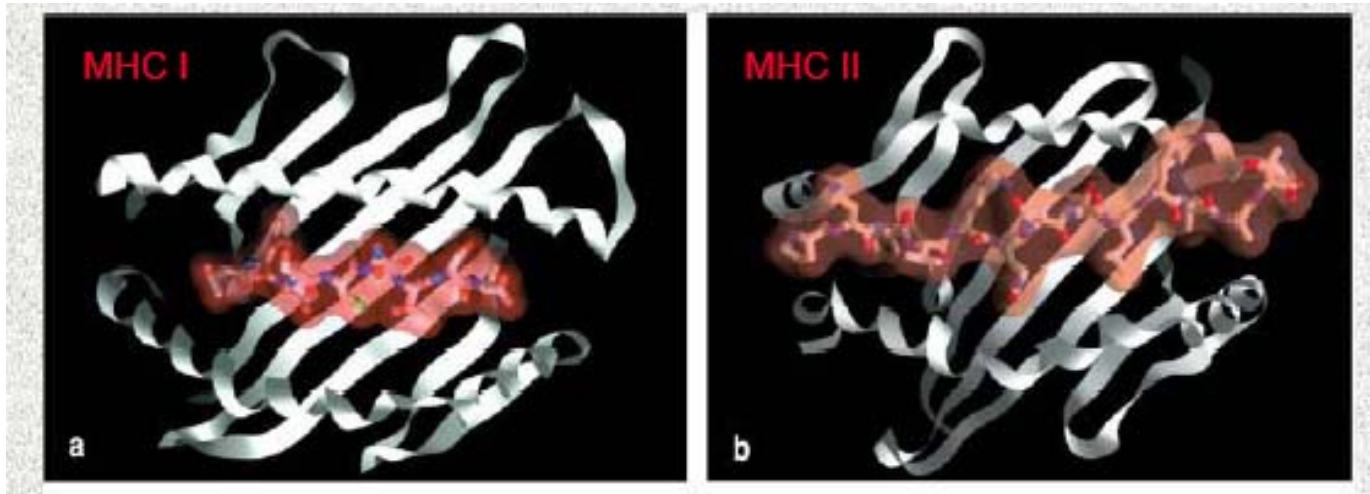
- Geny samoniezgodności u roślin zapobiegają kojarzeniom krewniaczym
- Rzadkie allele są faworyzowane, wielka zmienność w populacjach
- Linie alleliczne bardzo długo utrzymywane przez dobór
- Filogeneza alleli różni się od filogenezy gatunków



Richman & Kohn 1999

MHC

- u kręgowców odpowiadają za prezentację limfocytom antygenów wewnątrzkomórkowych, np. wirusów (MHC I) i zewnątrzkomórkowych, jak bakterie, pierwotniaki, grzyby (MHC II)
- w wiązanie antygenów zaangażowane są niektóre pozycje aminokwasowe (PBR, ABS), allele różniące się w tych miejscach mogą prezentować różnorodne antygeny

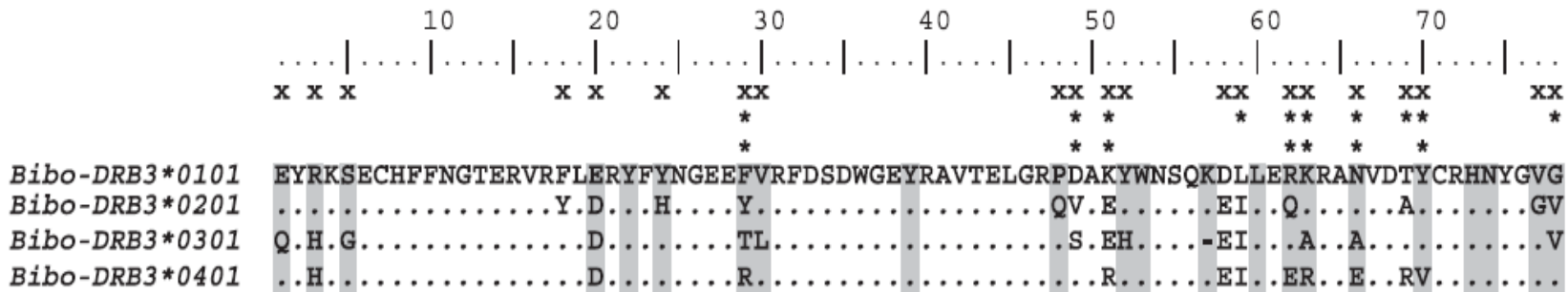
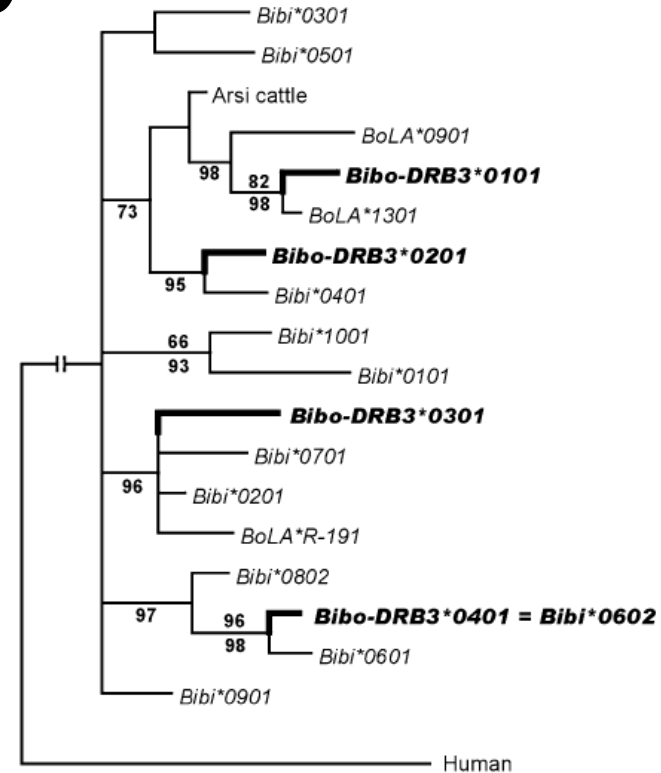


MHC

- Geny MHC są wysoce zmienne w większości populacji naturalnych
- Również w nich częsty jest polimorfizm transspecyficzny
- Za główne mechanizmy utrzymujące zmienność uważa się dobór zależny od częstości i przewagę heterozygot
- Niektóre genotypy MHC związane są z podatnością a inne z odpornością na choroby (np. malaria, HIV, SIV)
- Zmienność MHC może być utrzymywana przez dobór nawet gdy w wyniku działania dryfu tracona jest w różnych częściach genomu (lis *Urocyon littoralis*)
- Gatunki o niskiej zmienności MHC często wykazują zwiększoną podatność na choroby (gepardy, tamaryny)
- Zachowanie zmienności MHC celem wielu programów konserwatorskich

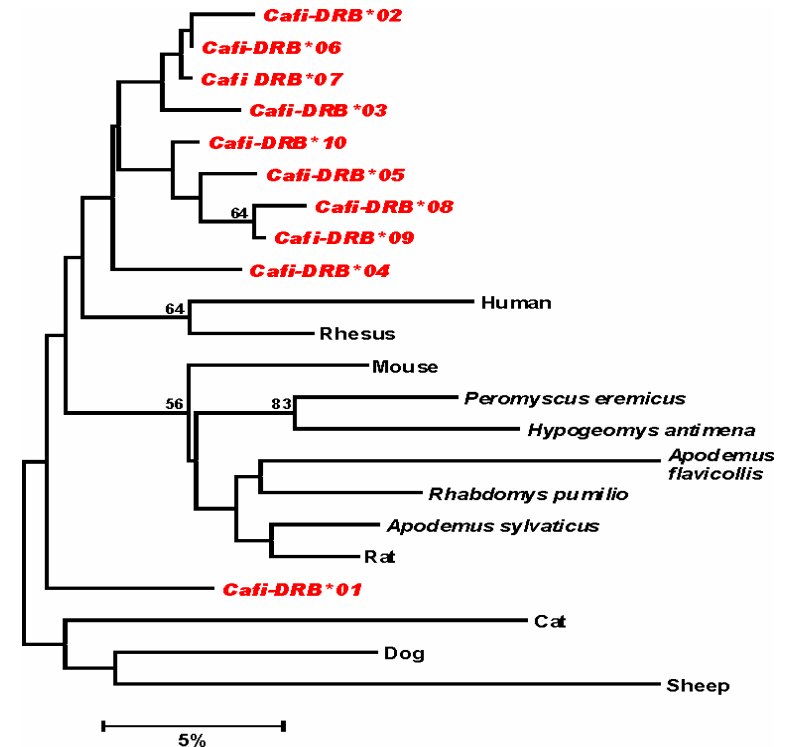
MHC

MHC w Białowieskiej populacji żubra, odtworzonej z 7 osobników, przetrwały 4 allele MHC II, bardzo różne od siebie, wyraźne ślady doboru pozytywnego, głównie w miejscach ABS



MHC

MHC u bobra, również bardzo różne allele zachowane, lecz w poszczególnych populacjach brak zmienności, historyczny dobór pozytywny, ostatnio efekt wąskiego gardła



Sites	d_N	d_S	Z	P
All	8.0 (2.2)	2.3 (1.2)	2.792	0.003
ABS	23.8 (7.9)	4.5 (3.0)	2.525	0.006
Non-ABS	3.3 (1.7)	1.6 (1.3)	0.894	0.187



Test McDonald-Kreitmana

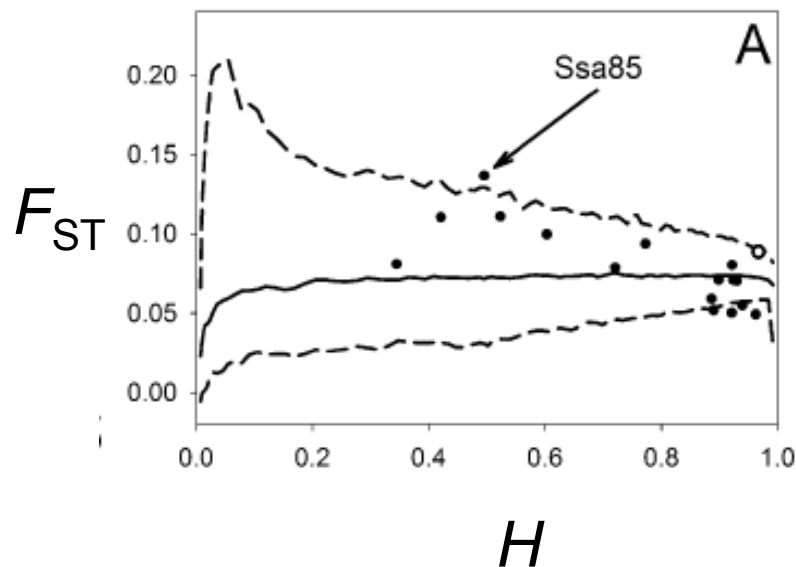
- Jeżeli ewolucja jest neutralna to te same procesy odpowiadają za zmienność w obrębie gatunku i między gatunkami
- Dywergencja między gatunkami i zmienność w ich obrębie powinny być proporcjonalne do tempa mutacji neutralnych
- Dane z *Adh* u *Drosophila melanogaster*, *D. simulans* i *D. yakuba*

	Utrwalone	Polimorficzne
Synonimowe	17	42
Niesynonimowe	7	2

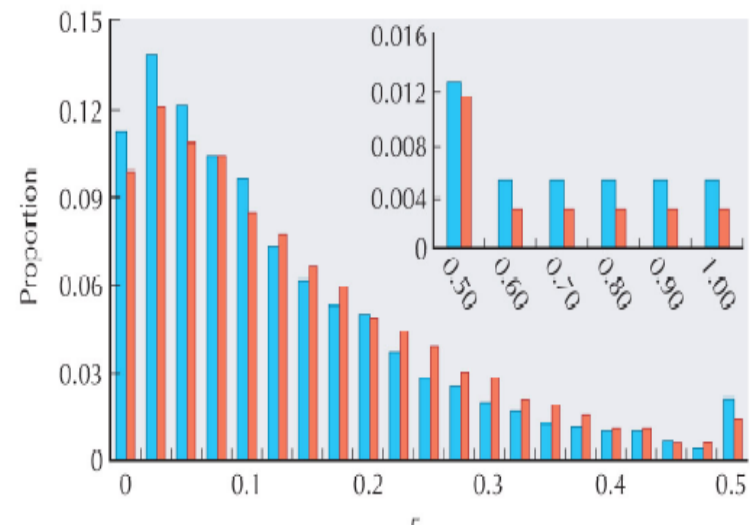
- O wiele większy procent zmian niesynonimowych utrwalony między gatunkami, prawdopodobnie utrwalił je dobór
- Nadmiar niesynonimowych polimorfizmów sugerowałby dobór równoważący lub oczyszczający, można je rozróżnić patrząc na częstości alleli

Zróżnicowanie genetyczne a dobór

- Geny odpowiedzialne za lokalne adaptacje będą bardziej zróżnicowane między populacjami
- Geny pod działaniem np. doboru równoważącego będą mniej zróżnicowane



Aguilar & Garza 2006

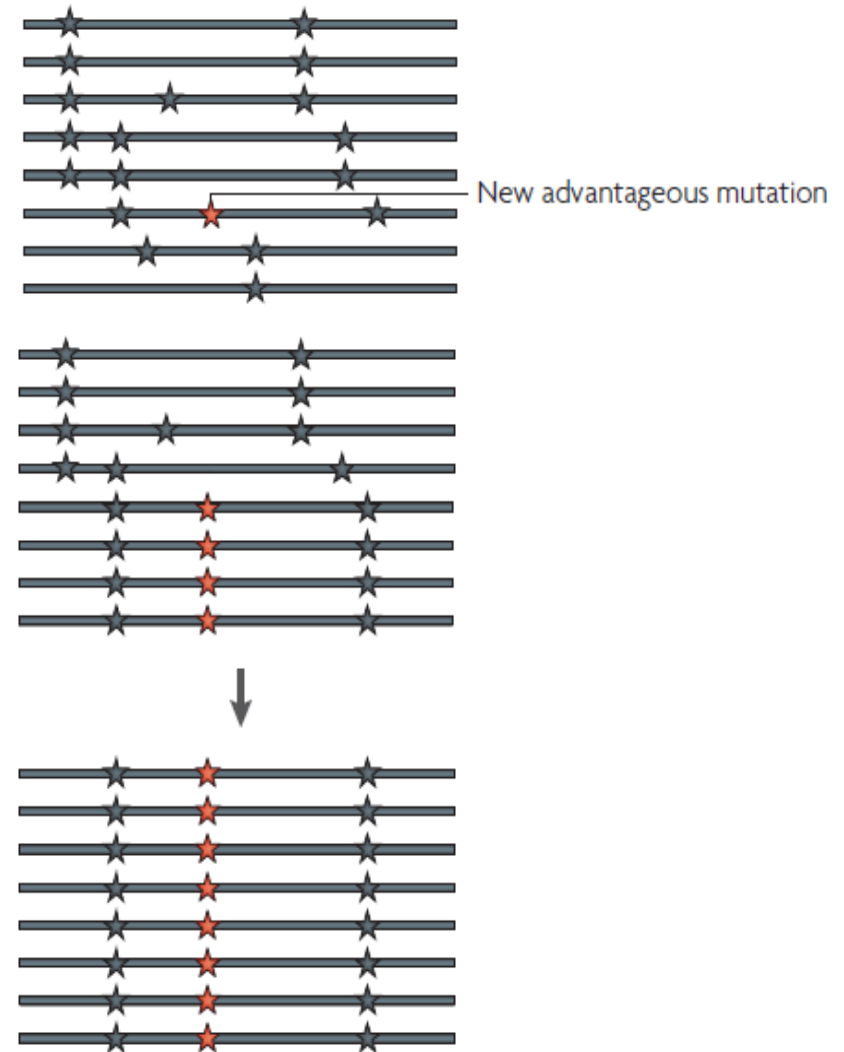


F_{ST} w ludzkich genach

Barton et al. 2007

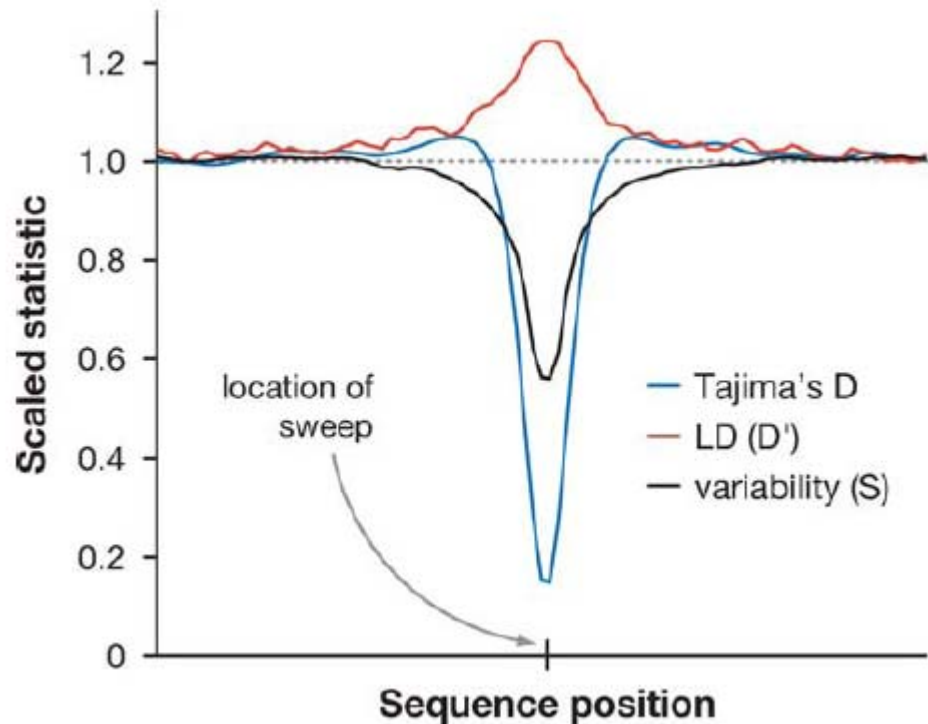
Wymiatanie selekcyjne

- Gdy pojawi się nowa mutacja korzystna, region chromosomu na którym wystąpiła utrwali się w populacji
- Zmienność która była w tym regionie na innych chromosomach ulegnie zostanie stracona
- Allele blisko allelu faworyzowanego zostaną podwiezione (*hitchhiking*) do wysokiej częstości
- Długość regionu wymiatania zależy od siły doboru i tempa rekombinacji



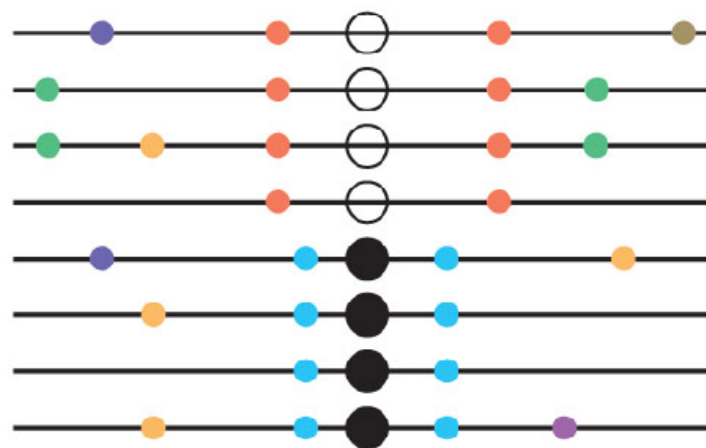
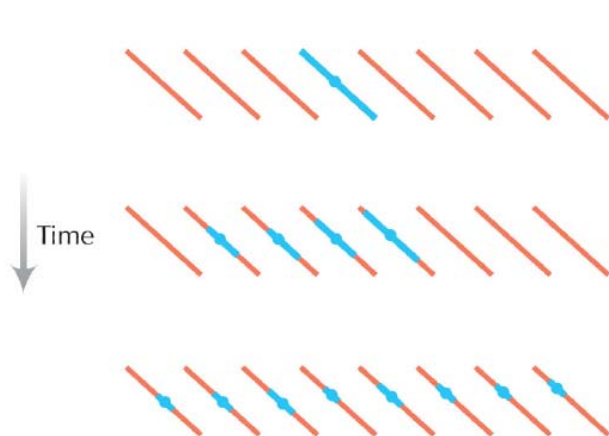
Wymiatanie selekcyjne

- Sygnaturą wymiatania jest lokalnie obniżona zmienność i podwyższona nierównowaga sprzężeń



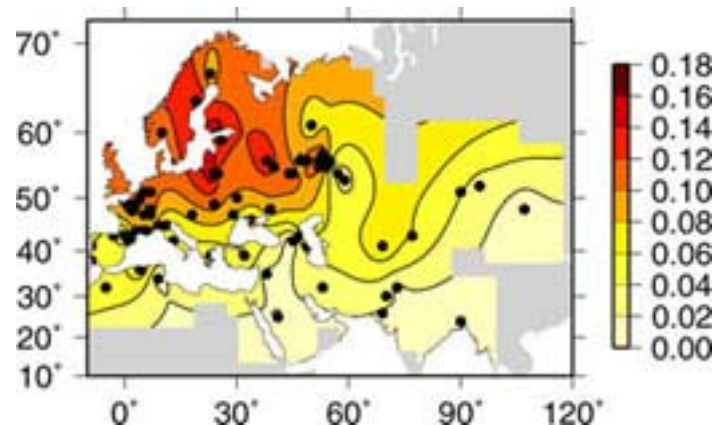
Wiek korzystnych mutacji można określić

- gdy pojawia się korzystna mutacja pojawia się w określonym „kontekście genomowym”
- pewien obszar genomu będzie w nierównowadze sprzężeń (LD)
- znając zakres LD i tempo rekombinacji możemy obliczyć wiek mutacji



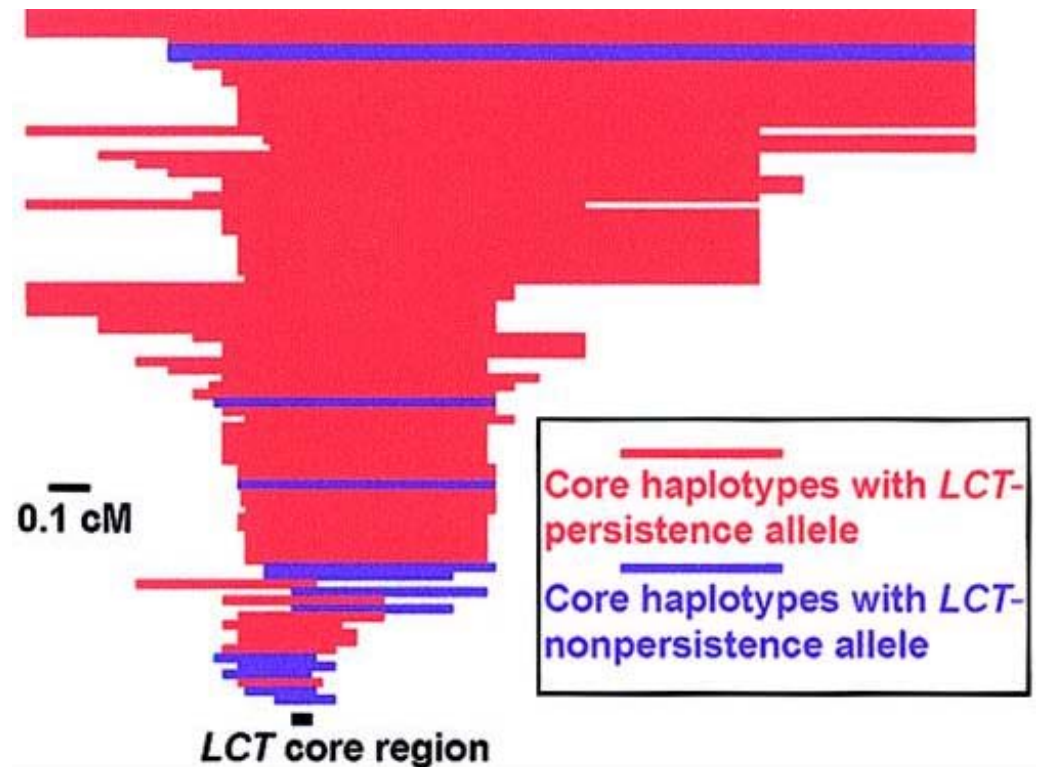
Wiek korzystnych mutacji można określić CCR5 a odporność na HIV

- CCR Δ 5 to mutacja receptora na powierzchni limfocytów który wirus HIV używa by wejść d komórki, delecja czyni białko nieaktywnym
- wysoka częstość w pn. Europie, brak gdzie indziej
- być może zapewniała odporność na inne infekcje (dżuma)
- początkowo oszacowano wiek z LD na 700 lat
- nowe badania wskazują że jest starsza, ok. 3 000 lat



Wiek korzystnych mutacji można określić tolerancja na laktozę

- normalnie zdolność trawienia mleka zanika u dorosłych ludzi
- wysoka częstość mutacji umożliwiającej tolerancję laktozy w populacjach, do 77% w Skandynawii,
- bloki wysokiego LD znacznie dłuższe w allelach z mutacją
- współczynnik doboru kilka %



Wykrywanie doboru pozytywnego w genomie - podsumowanie

- Silny dobór pozytywny pozostawia ślady w genomie
- Różne rodzaje śladów widoczne są przez różny czas:
 - Dawno (mln lat): wysoka proporcja zmian funkcjonalnych
 - Średnio dawno (setki tys lat) – spadek zmienności wywołany wymiataniem selekcyjnym
 - Stosunkowo niedawno (dziesiątki tys lat) – wysoka częstość nowopowstałych (zmienionych alleli)
 - Całkiem niedawno (dziesiątki tys lat) – zróżnicowanie międzypopulacyjne
 - Bardzo niedawno (kilkanaście tys lat – teraz) – długie haplotypy

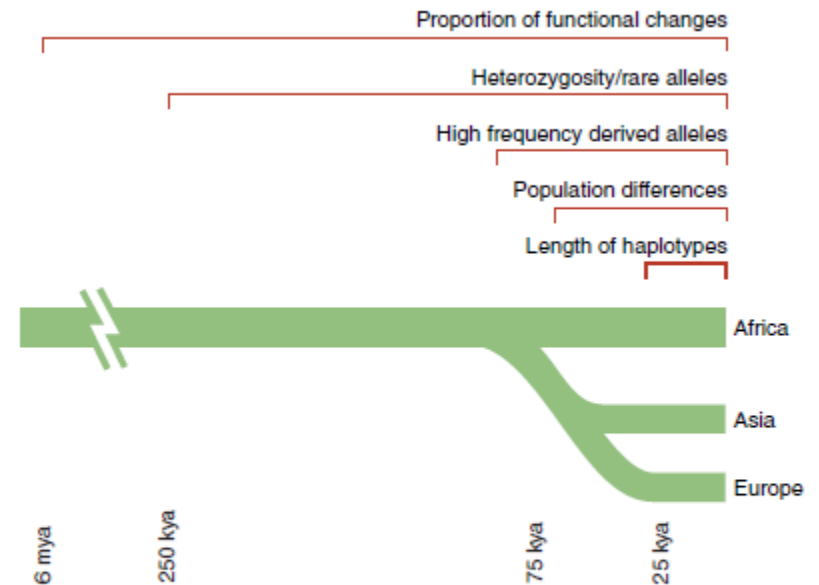
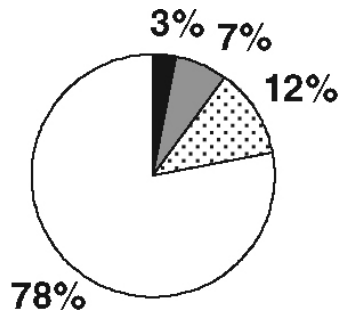
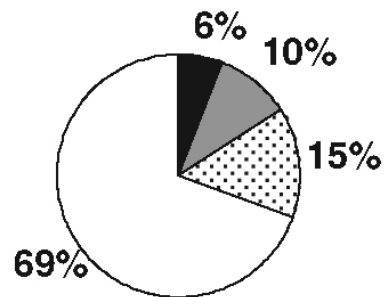


Fig. 1. Time scales for the signatures of selection. The five signatures of selection persist over varying time scales. A rough estimate is shown of how long each is useful for detecting selection in humans. (See fig. S1 for details on how the approximate time scales were estimated).

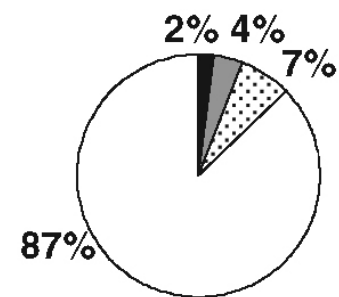
Zagrozone gatunki



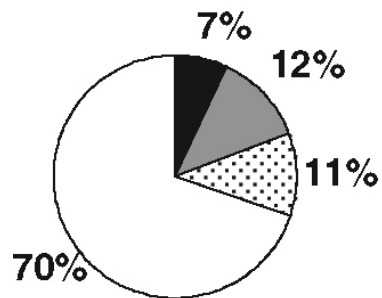
Mammals



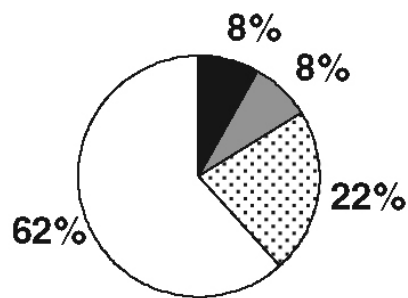
Reptiles



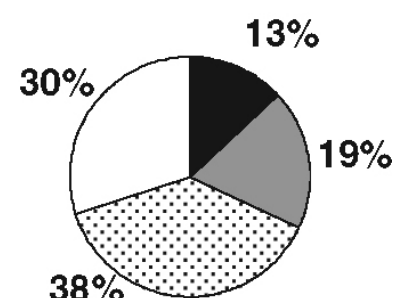
Birds



Amphibians



Fishes



Plants

Cr. endangered



Endangered



Vulnerable



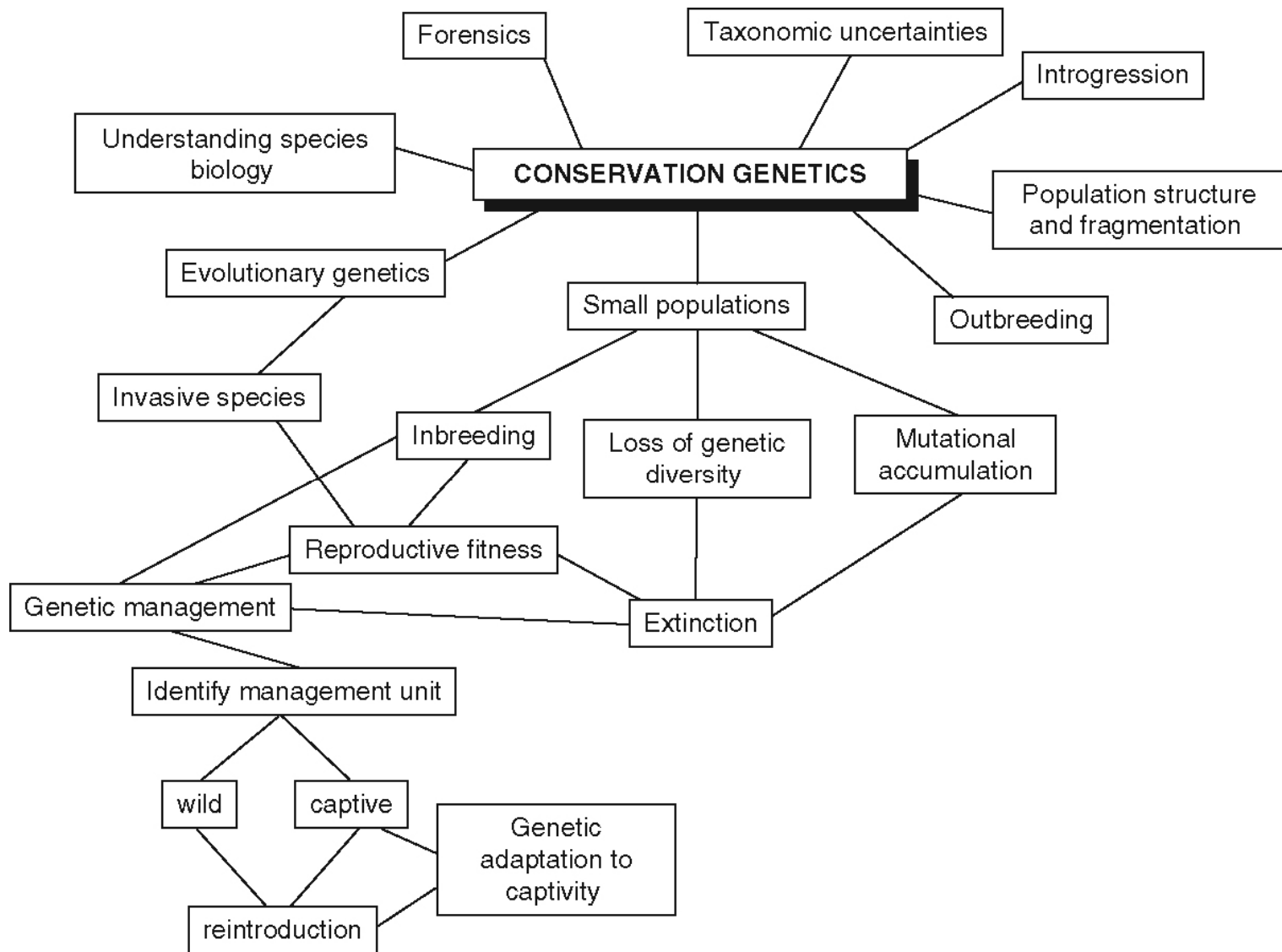
Low risk, etc.



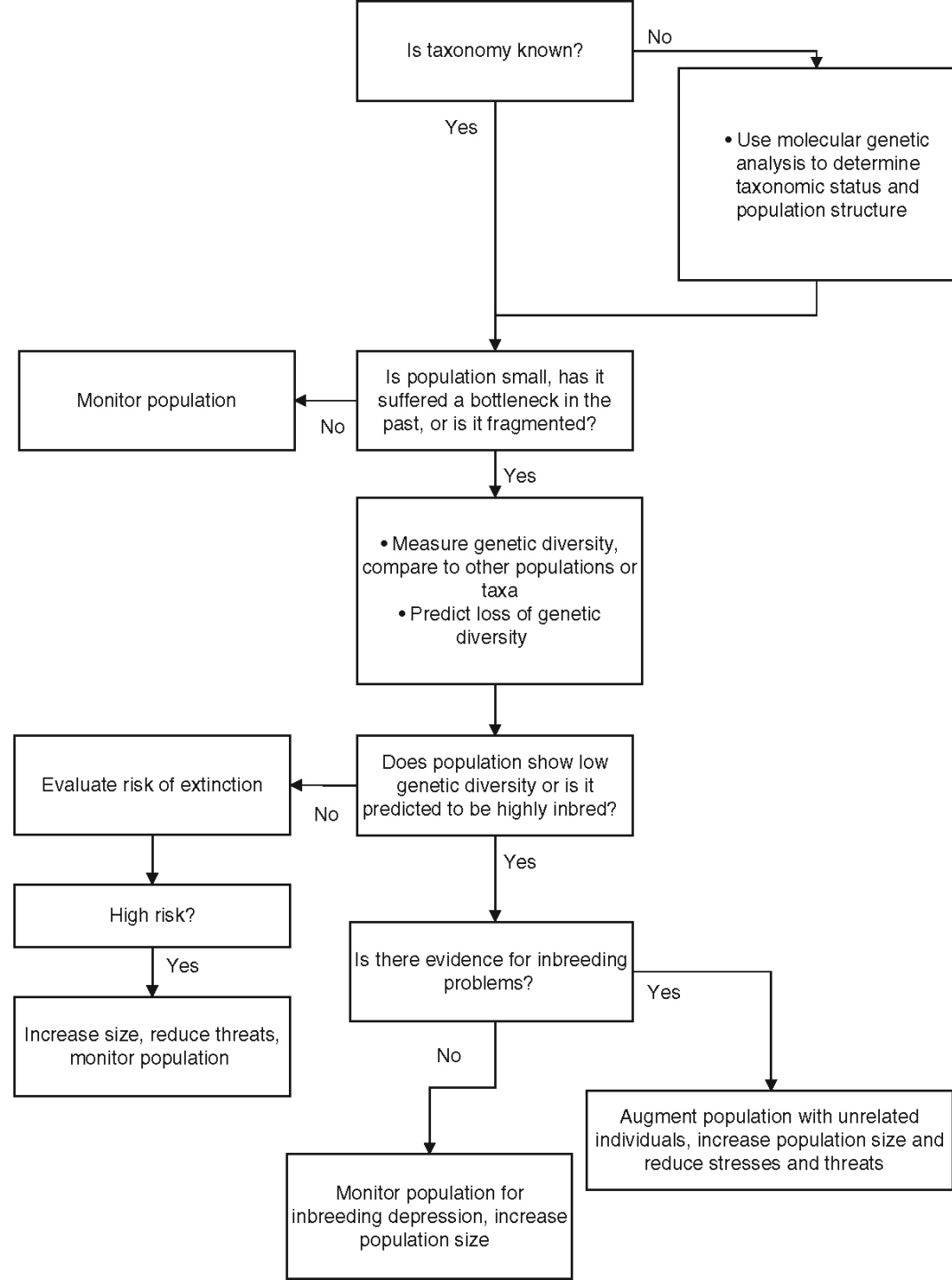
Genetyka konserwatorska

- Zastosowanie metod genetycznych, a szczególnie genetyczno-molekularnych w ochronie gatunków, biocenoz i biomów
- Dostarczenie informacji jak najlepiej wykorzystać ograniczone środki dostępne dla działań ochronnych
- Heterozygotyczność a żywotność populacji, depresja wsobna
- Identyfikacja pokrewieństwa w programach hodowlanych
- Identyfikacja centrów różnorodności genetycznej oraz jednostek zasługujących na ochronę, MU, ESU
- Identyfikacja produktów z zagrożonych gatunków – wymuszanie przestrzegania prawa

Genetyka konserwatorska

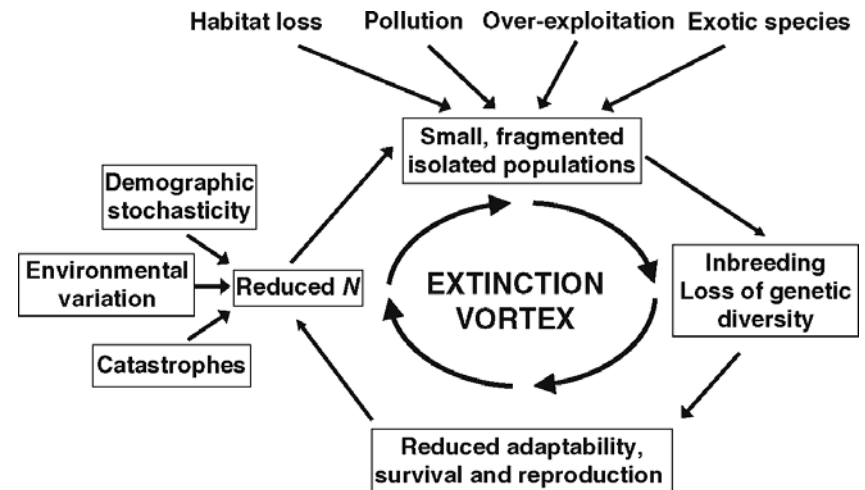
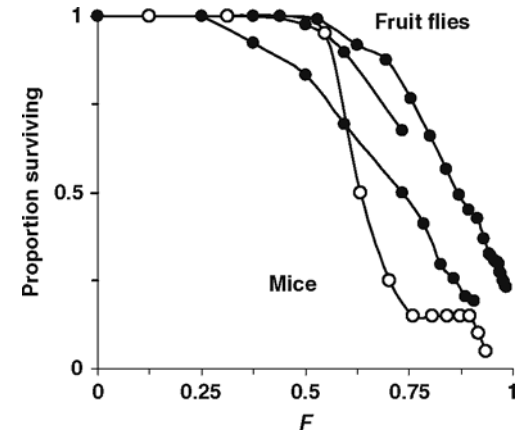


Jakie pytania się zadaje?



Zmienność, wsobność i ryzyko wymarcia

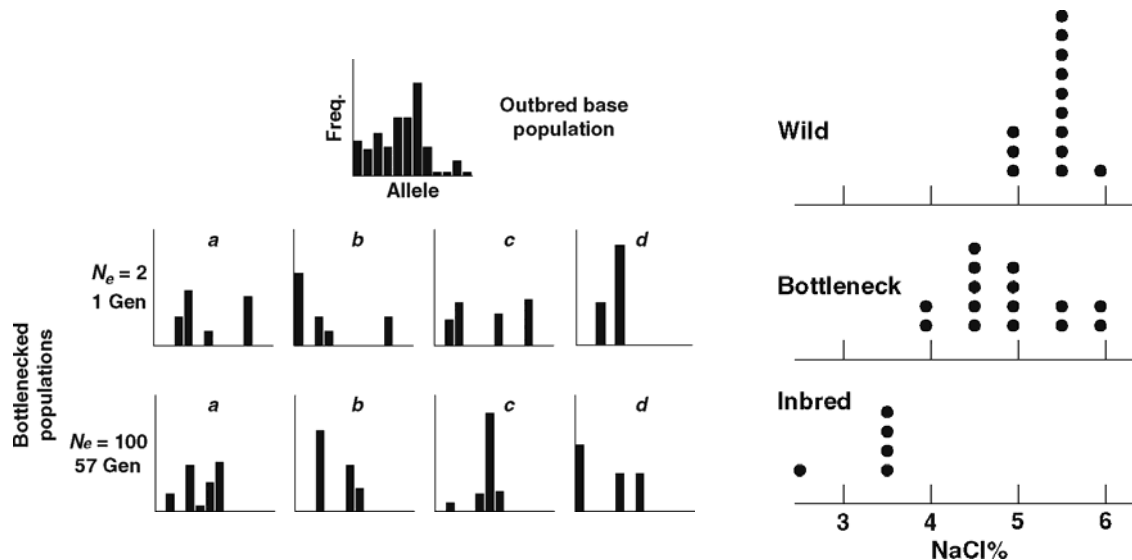
- zmienność genetyczna jest niezbędna dla adaptacji
- większość (80%) zagrożonych gatunków ma mniejszą zmienność genetyczną niż spokrewnione z nimi gatunki niezagrożone
- wzrost wsobności powoduje większe ryzyko wymarcia
- w małych populacjach wzrasta wsobność i spada zmienność genetyczna, są one bardziej narażone na działanie czynników stochastycznych – extinction vortex



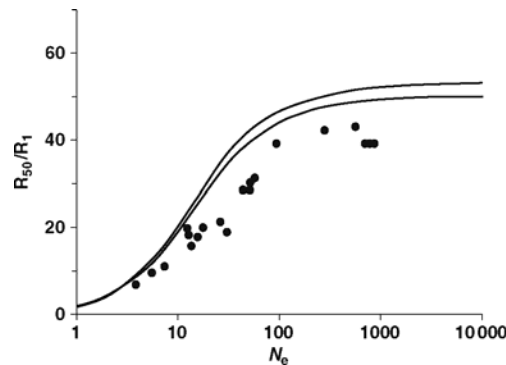
Niska heterozygotyczność (często) szkodzi

- Markery molekularne to najlepszy sposób szacowania heterozygotyczności (H) na poziomie całego genomu
- Niskie H może być konsekwencją spadków liczebności – wąskich gardeł
- Niezwykle niska zmienność u geparda
- Wirus FIP spowodował 50-60% śmiertelność u gepardów w niewoli w porównaniu do 1% u kota domowego
- Wysoka śmiertelność młodych, liczne zaburzenia rozwojowe
- Lwy z Gir Forest: niski testosteron, nienormalne plemniki
- Zagrożony gołąb z Puerto Rico: mniejsza liczba jaj, piskląt, wylotów
- U rybki *Poeciliopsis occidentalis* wszystkie badane składniki dostosowania (przeżywalność, wzrost, płodność i stabilność rozwojowa) okazały się dodatnio skorelowane ze średnią heterozygotycznością allozymów

Wąskie gardła populacyjne powodują zarówno spadek zmienności jak i spadek potencjału ewolucyjnego



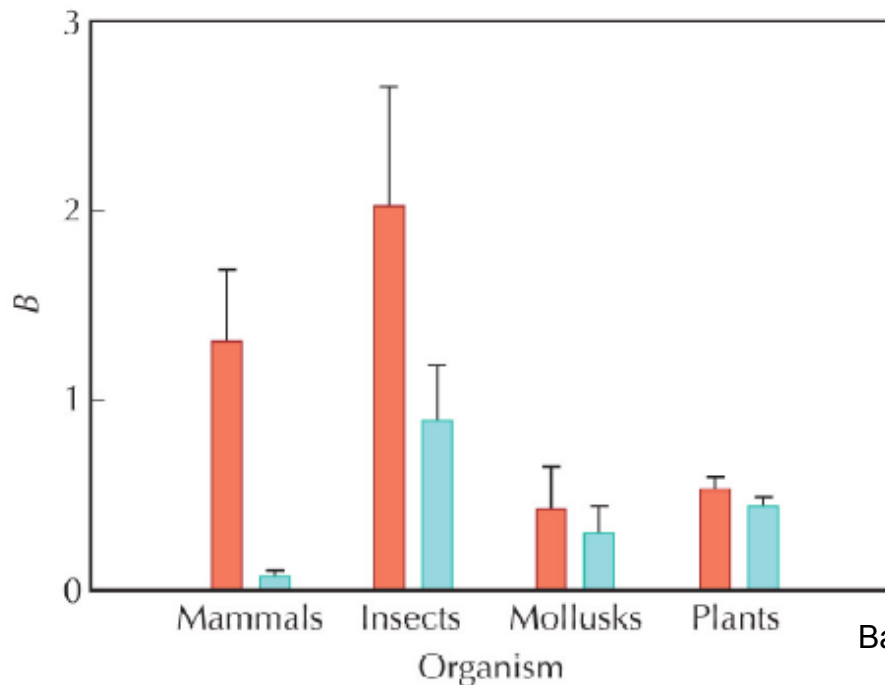
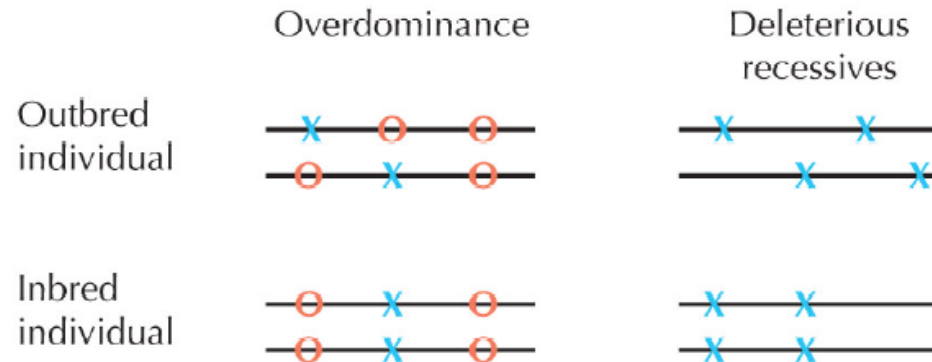
Frankham i in. 2010



Depresja wsobna

- Depresja wsobna (DW) to spadek przeżywalności, płodności lub tempa wzrostu, obserwowany często w następstwie kojarzeń krewniaczych
- DW jest szczególnie ważna w genetyce konserwatorskiej gdyż poziom wsobności w małych populacjach jest często znaczny
- Genetycznie, populacje wsobne mają obniżoną heterozygotyczność (większą homozygotyczność)
- Dwie konkurencyjne hipotezy:
 - Dominacji: spadek dostosowania wywołany ujawnianiem się rzadkich szkodliwych alleli w stanie homozygotycznym, gatunki o długiej historii wsobności powinny radzić sobie lepiej
 - Naddominacji: sama heterozygotyczność w wielu loci podnosi dostosowanie
 - Wydaje się że **dominacja** ważniejsza

Depresja wsobna



Z teorii dominacji wynika oczekiwanie że powinno działać ‘czyszczenie ze szkodliwych mutacji’

Depresja wsobna

- Gatunki różnią się bardzo wrażliwością na kojarzenia krewniacze (wsobność)

Gatunek	Koszt wsobności
Tygrys sumatrzański (<i>Panthera tigris sumatrae</i>)	<0.01
Pies leśny (<i>Speothos venaticus</i>)	0.06
Opos krótkoogonowy (<i>Monodelphis domestica</i>)	0.10
Gaur (<i>Bos gaurus</i>)	0.12
Hipopotam karłowaty (<i>Choeropsis liberensis</i>)	0.33
Galago większy (<i>Galago g. crassicaudatus</i>)	0.34
Gazela dorkas (<i>Gazella dorcas</i>)	0.37
Ryjokoczek (<i>Elephantulus rufescens</i>)	0.41
Marmozeta lwia (<i>Leontopithecus r. rosalia</i>)	0.42
Lemur rudy (<i>Lemur fulvus</i>)	0.90

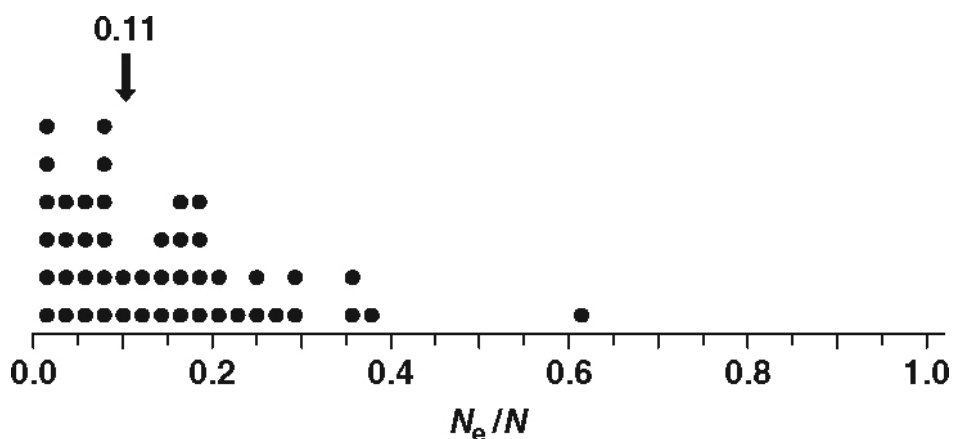
Depresja wsobna

- jest **powszechna** wśród organizmów nie rozmnażających się przez samozapłodnienie
- ma znaczący **komponent stochastyczny**
- zazwyczaj jest silniejsza w warunkach stresowych
- zazwyczaj jest silniejsza w populacjach dzikich niż w niewoli
- nie występuje u organizmów haploidalnych i w genach w których brak dominacji lub naddominacji
- jej skutki w małych populacjach można usunąć wprowadzając osobniki z innych populacji (pod warunkiem że będą się kojarzyć i produkować potomstwo z rezydentami) – **genetic rescue**

Zasada 50/500 i MVP

Minimalna wielkość żywotnej populacji (Minimum Viable Population, **MVP**) - minimalna wielkość populacji mającej możliwość przetrwania i adaptacji do zmieniających się warunków środowiska. MVP powinna odpowiadać efektywnej wielkości populacji (N_e) = 50/500

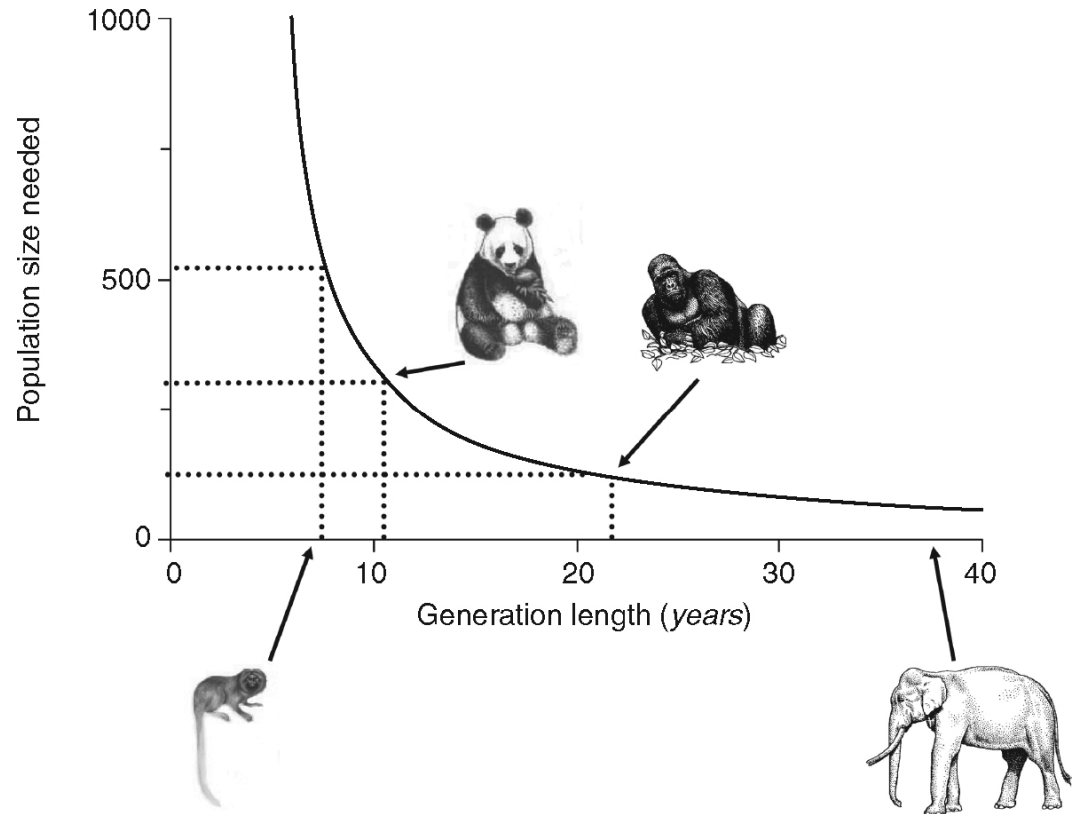
Jeżeli jednak $N_e = 1/10N$, to N (MVP) powinna wynosić 500/5000



Zarządzanie populacjami w niewoli

zazwyczaj celem zarządzania jest utrzymanie 90% początkowej zmienności genetycznej przez 100 lat

Jest to kompromis wynikający z ograniczonych zasobów



Niejasności taksonomiczne

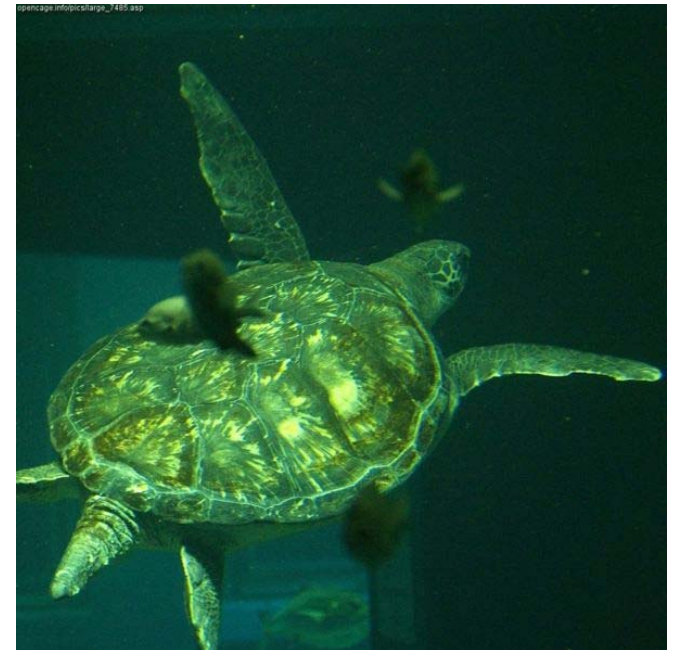
- w planowaniu działań ochronnych bardzo pomaga znajomość pokrewieństw populacji i ich grup w obrębie gatunków
- czasem nieznajomość taksonomii może prowadzić do wymarcia gatunków -> gatunki kryptyczne
- czasem wiele zasobów poświęca się ochronie populacji bardzo zbliżonych genetycznie do innych, niezagrożonych populacji
- rozwikłanie ewentualnych niejasności taksonomicznych powinno być pierwszym krokiem w działaniach ochronnych – metody molekularne

MU i ESU

- **Jednostki o znaczeniu ewolucyjnym** (ang. Evolutionary Significant Units, **ESU**) to populacje charakteryzujące się **historią ewolucyjną w znacznym stopniu odrębną** od innych takich jednostek (Ryder 1986). ESU są głównym źródłem historycznej (a być może również adaptacyjnej) różnorodności genetycznej w obrębie gatunku i jako takie zasługują na szczególną uwagę w kontekście konserwatorskim
- Jeżeli populacja wymienia z innymi na tyle mało migrantów, by zachować odrębność genetyczną, będzie niezależna od innych populacji także demograficznie. Takie populacje to **Jednostki Zarządzania** (Management Units, **MU**), jednostki niższego rzędu niż ESU
- Gdy, jak to się często zdarza, brak danych z natury na temat stopnia łączności demograficznej populacji, lub dane takie są niewystarczające, MU można wstępnie zdefiniować na podstawie **istotnych różnic częstości alleli** w neutralnych loci markerowych, np. mtDNA

ESU

- ESU wyróżnia się przede wszystkim metodami molekularnym, jedna z najpopularniejszych definicji uznaje za ESU takie grupy populacji, które są „wzajemnie monofiletyczne pod względem alleli mtDNA oraz różnią się istotnie częstościami alleli w loci jądrowych” (Moritz 1994)
- MU i ESU to różne poziomy zróżnicowania genetycznego w obrębie gatunku
- Przykład: żółw zielony
- Poszczególne kolonie lęgowe stanowią MU
- Populacje z różnych basenów oceanicznych stanowią odrębne ESU



Przestrzeganie przepisów ochronnych

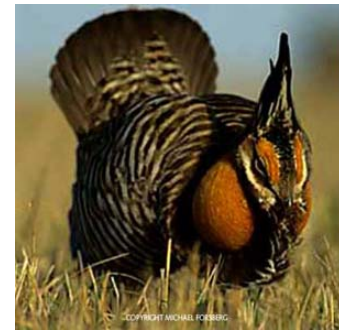
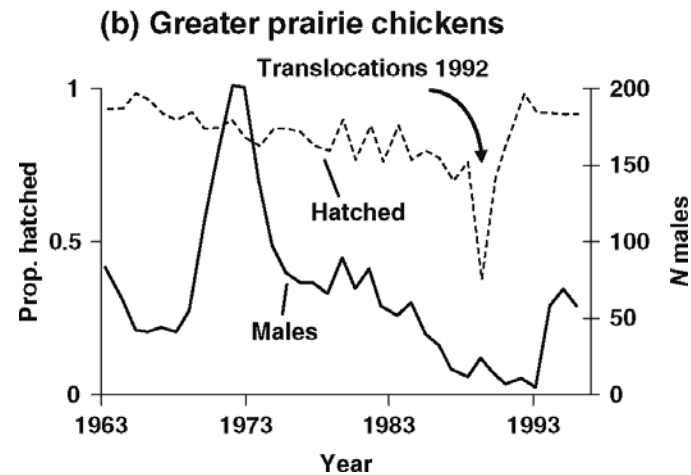
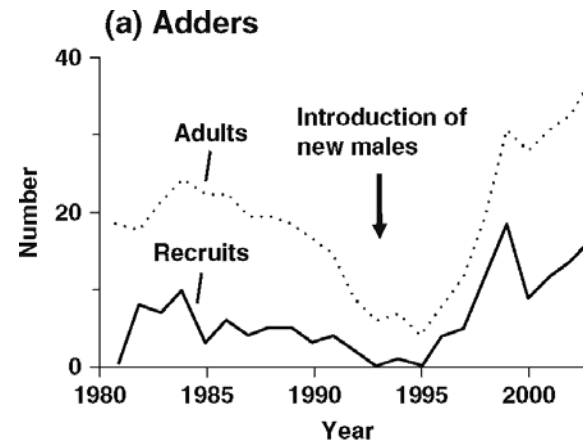
- W egzekwowaniu prawa o ochronie dzikich organizmów często wyzwaniem jest identyfikacja biologicznego źródła krwi, tusz, mięsa, piór lub przetworzonych produktów wyprodukowanych z zagrożonych lub nielegalnie pozyskanych gatunków.
- Markery molekularne często umożliwiają identyfikację gatunku nawet w trudnych sytuacjach
- Od lat 70 tych allozymy służyły monitorowaniu okresów ochronnych i obszarów chronionych w rybołówstwie
- 1986 moratorium na połowy wielu walen
- W oparciu o bazę sekwencji mtDNA monitorowano przestrzeganie prawa sprawdzając mięso wielorybie z targów Azji Pd-Wsch, Japonii i Skandynawii – stwierdzono nielegalne pozyskiwanie np. humbaków i finwali
- Internetowy program nadzoru DNA walen
- 1995 Ashford (Oregon) - laboratorium U.S. Fish and Wildlife Service które zajmuje się głównie tego typu badaniami

Diagnozowanie problemów genetycznych

- Jak duża jest populacja (N_e)?
- Czy przechodziła przez wąskie gardło(a)?
- Czy nastąpiła utrata zmienności genetycznej?
- Czy jest widoczna depresja wsobna?
- Jak wygląda rozmieszczenie geograficzne?
- Czy populacja jest pofragmentowana w takim stopniu że utrudniony/niemożliwy jest przepływ genów między fragmentami?

Genetic rescue

gdy populacja jest mała i cierpi na depresję wsobną, translokacje osobników z innych populacji mogą bardzo pomóc



Genetic rescue

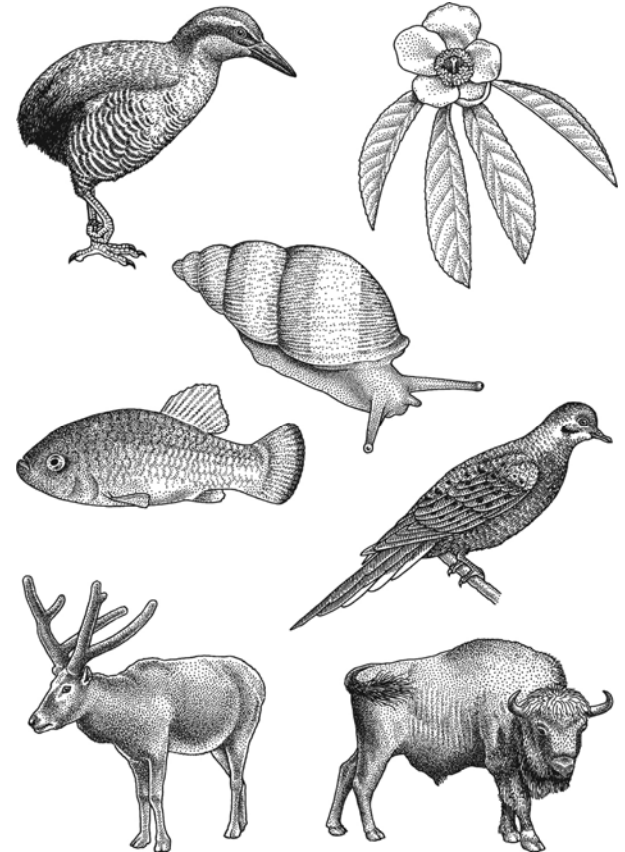
Frankham i in. 2010

Translokacje

- translokacje są bardzo pomocne w ratowaniu populacji ginących gatunków
- potrzeba stałego monitoringu
- najpierw trzeba sprawdzić:
 - które osobniki translokować
 - ile?
 - jak często?
 - skąd?
 - jak często?
 - kiedy przestać
- monitorowanie translokacji zarówno przez obserwacje jak i z wykorzystaniem markerów molekularnych

Programy hodowlane (PH)

- Prowadzi się je w niewoli dla gatunków szczególnie zagrożonych stosując specjalne schematy kojarzeń aby zachować maksymalną ilość zmienności
- Ostatecznym celem PH jest ponowne wprowadzenie gatunku do środowiska naturalnego
- Stadia PH:
 - stwierdzenie zagrożenia
 - założenie populacji w niewoli
 - doprowadzenie jej (ich) do „bezpiecznych” rozmiarów
 - zarządzanie w niewoli
 - wybór osobników do reintrodukcji
 - zarządzanie populacjami w naturze



Programy hodowlane (PH)

- Markery molekularne pozwalają ustalić pokrewieństwo osobników gdy brak ksiąg rodowodowych a także skorygować błędy w rodowodach: ibis grzywiasty (*Geronticus eremita*), oryks arabski (*Oryx leucoryx*)
- Identyfikacja ojcostwa w PH żurawia krzykliwego (*Grus americana*)
- Identyfikacja płci szczudłaka czarnego (*Himantopus novaezelandiae*) w PH rozpoczętym z zaledwie 12 ptaków
- Nieinwazyjne oznaczenie płci u ostatniego osobnika ary modrej (*Cyanopsitta spixii*). Okazał się on samcem a więc jako potencjalną partnerkę do rozrodu wypuszczono samicę z niewoli



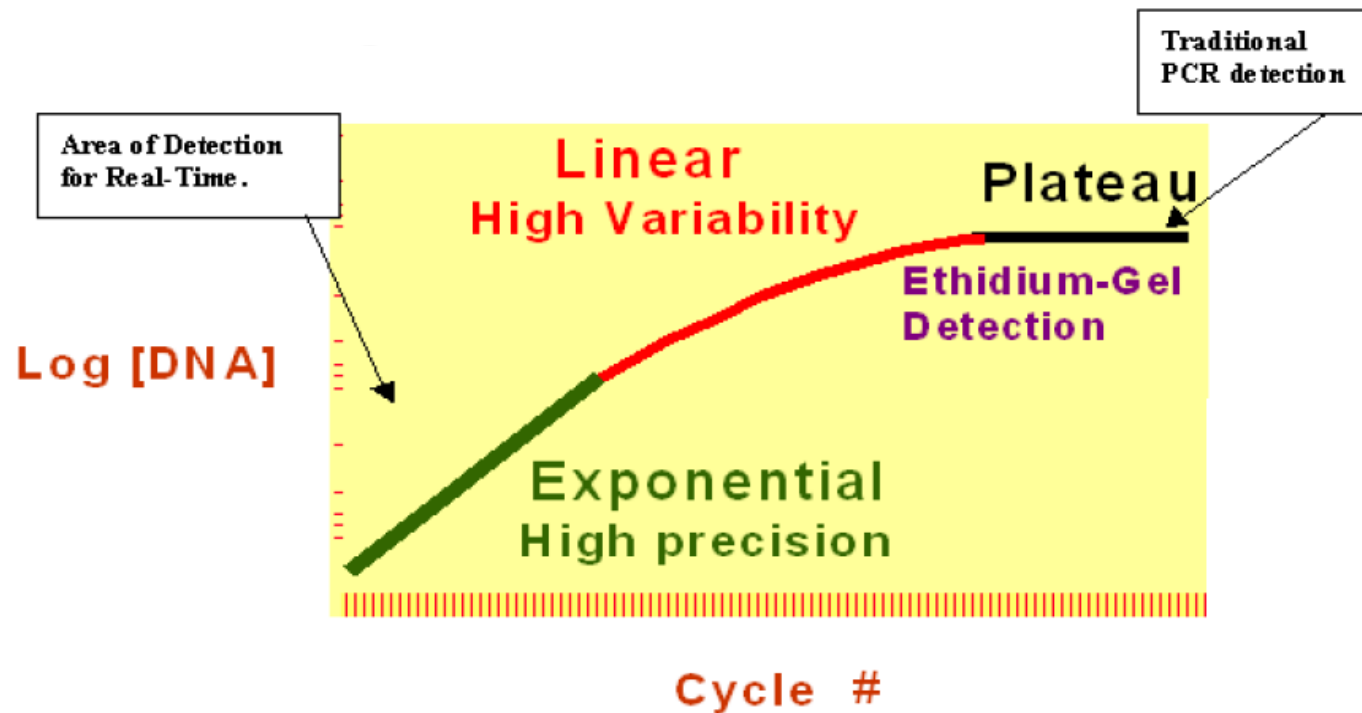
Ekologia molekularna a organizmy zmodyfikowane genetycznie (GMO)

Potencjalny wpływ organizmów transgenicznych na ekosystemy naturalne może być badany jedynie przy wykorzystaniu technik molekularnych

Organizmy transgeniczne zawierają markery pozwalające na szybkie i proste wykrywanie transgenów (zazwyczaj real-time PCR), ale też zwykły PCR lub ELISA

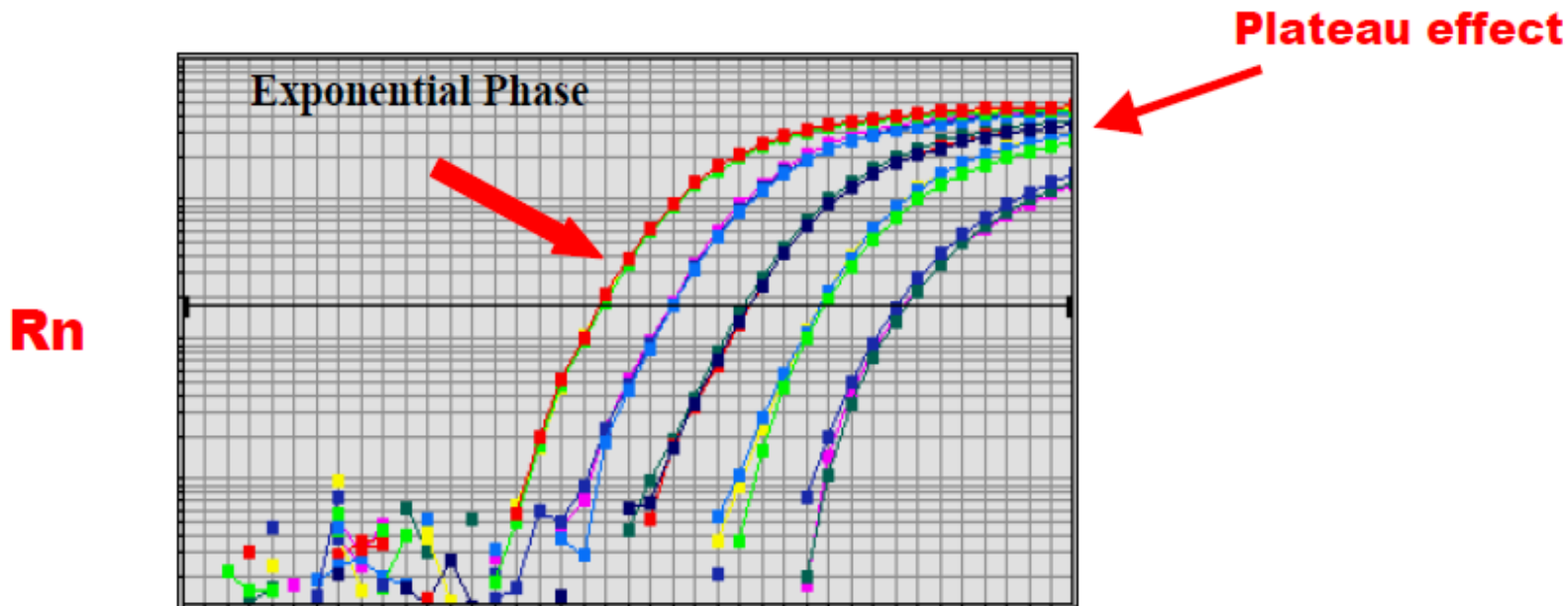


Real Time PCR



Etapy amplifikacji DNA w reakcji PCR

Real Time PCR



Pomiar stężenia DNA w wykładniczej fazie amplifikacji pozwala precyzyjnie zmierzyć początkową ilość matrycy; SYBR Green łączy się z dwuniciowym DNA

Zalety Real Time PCR

- Zwykły PCR mierzy DNA po reakcji, RT PCR w fazie wykładniczej amplifikacji
- Wzrost fluorescencji jest wprost proporcjonalny do wzrostu ilości produktu
- Duży zakres dynamiczny detekcji
- Nie ma potrzeby dalszych analiz po PCR
- Precyzyjny pomiar początkowego stężenia, nawet dwukrotne różnice

Ekologia molekularna a organizmy zmodyfikowane genetycznie (GMO)

- odporne na szkodniki i choroby
- szybko rosnące zwierzęta hodowlane
- szczepionki w pokarmie
- rośliny bez toksyn (np. kawa bezkofeinowa)
- lepszy smak i jakość?
- gleba i wody wolne od pestycydów
- w ponad 1 000 000 km² pól obsadzonych roślinami GMO
- w USA > 80% kukurydzy, soi i bawełny jest GM



Potencjalne zagrożenia środowiskowe GMO

- powstanie bardziej niebezpiecznych szkodników i patogenów
- “udoskonalenie” istniejących szkodników w wyniku hybrydyzacji ze spokrewnionymi organizmami transgenicznymi
- szkodliwość dla gatunków innych niż zamierzone: organizmów glebowych, nieszkodliwych owadów, ptaków i innych zwierząt
- zaburzenia w funkcjonowaniu ekosystemów przez wpływ bezpośredni lub pośredni na wiele ich składników
- nieodwracalne zmiany w bioróżnorodności lub różnorodności genetycznej gatunków

Mapowanie molekularne cech ilościowych

- wiele cech o podstawowym znaczeniu dla dostosowania to cechy ilościowe: masa ciała, szybkość wzrostu, płodność, agresywność itd.
- zazwyczaj odpowiedzialne za nie jest wiele genów
- analiza cech ilościowych wykorzystuje techniki statystyczne jak analiza wariancji, korelacja, kowariancja itd. i opiera się o pomiar korelacji cechy między krewnymi – znajomość podstaw genetycznych cechy nie jest potrzebna
- analiza loci ilościowych (QTL) pozwala na identyfikację fragmentów genomu w znacznym stopniu warunkujących cechę ilościową

Mapowanie molekularne cech ilościowych

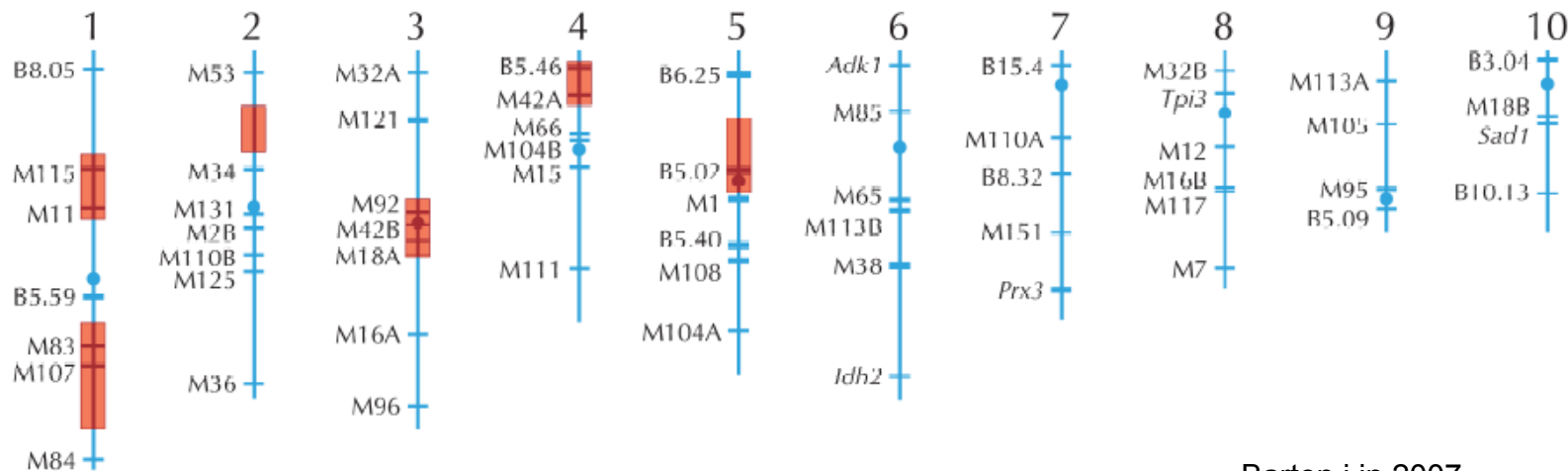
- Interesujące jest ile regionów genom wpływa na wykształcenie cechy
- Czy udział któregoś z regionów jest większy niż innych?



Doebley i in. 2006

Mapowanie molekularne cech ilościowych

- Musimy dysponować gęstą mapą genetyczną, najczęściej opartą na mikrosatelitach

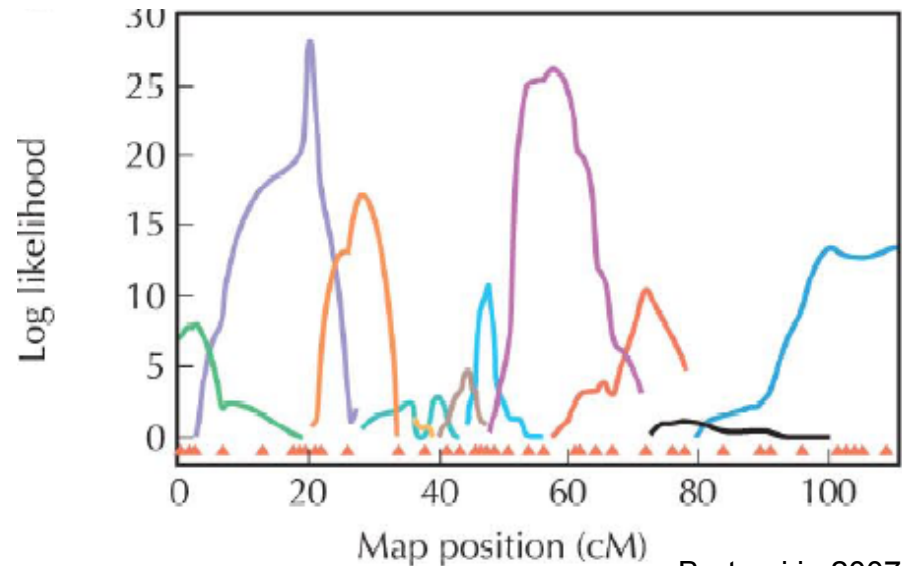
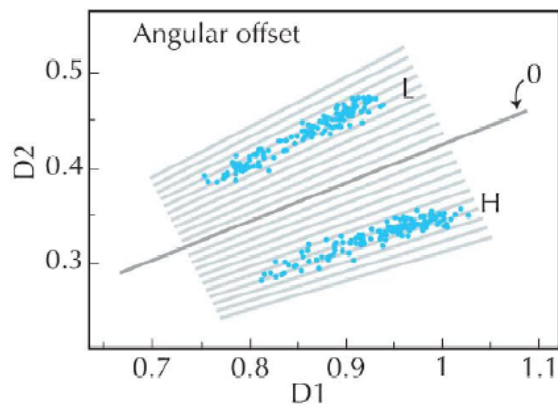
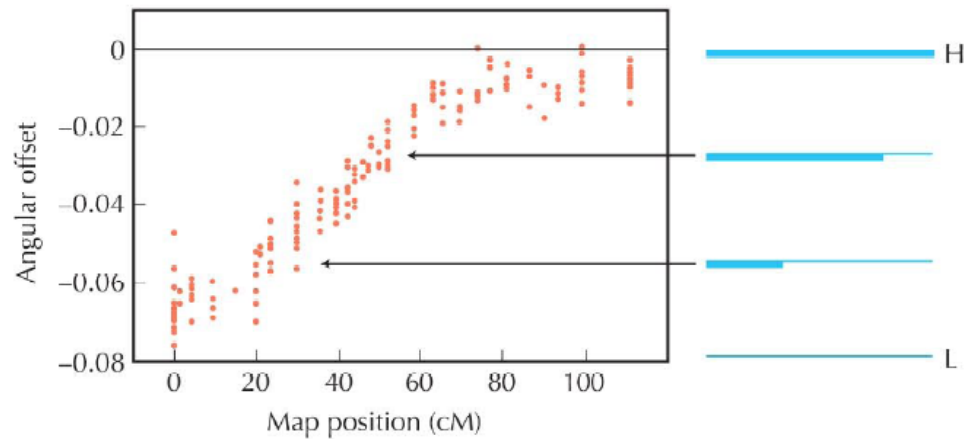
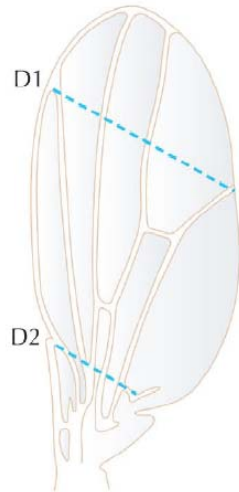


Barton i in 2007

- Prowadzimy dobór na określoną cechę

Mapowanie molekularne cech ilościowych

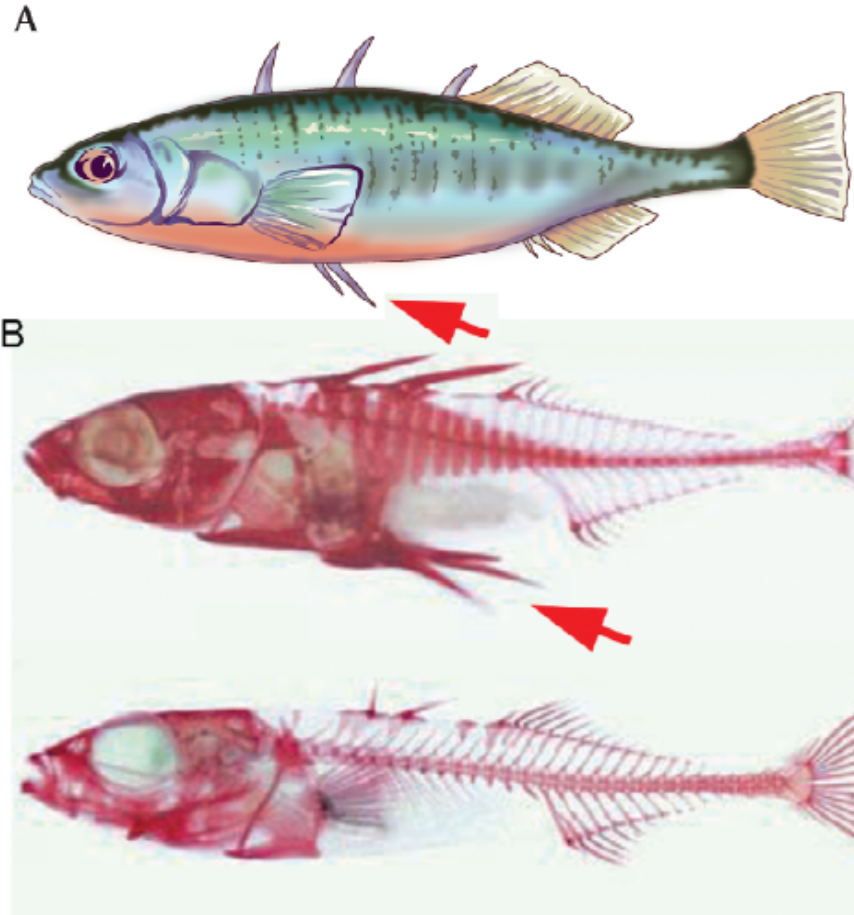
kształt skrzydła u *Drosophila*



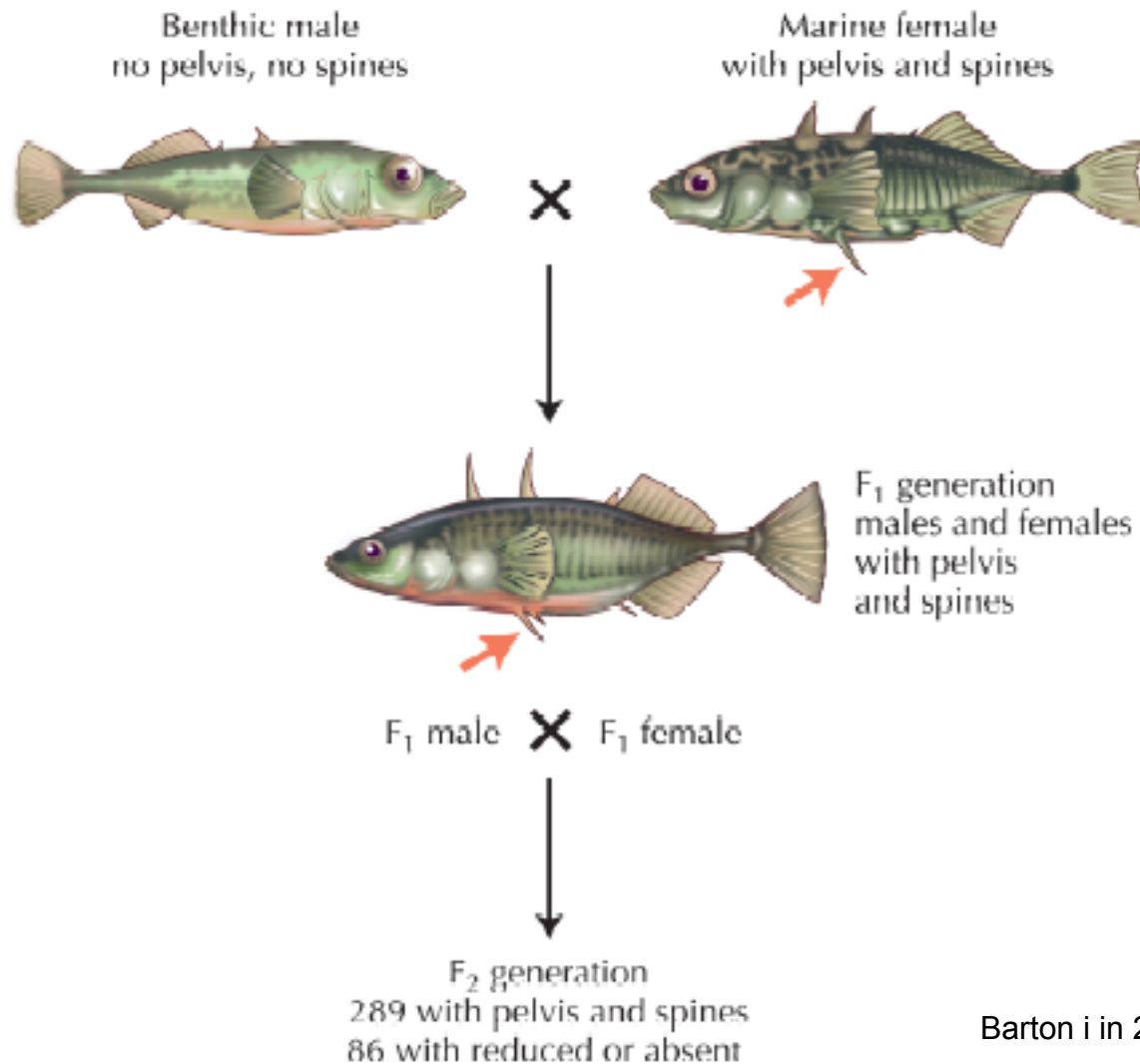
Geny warunkujące złożone cechy – QTL i geny-kandydaci

- Cierniki tworzą populacje słono i słodkowodne < 15 tys. lat
- W słonowodnych silnie rozwinięta miednica jest podstawą dla kolców brzusznych - adaptacja

Geny warunkujące złożone cechy – QTL i geny-kandydaci



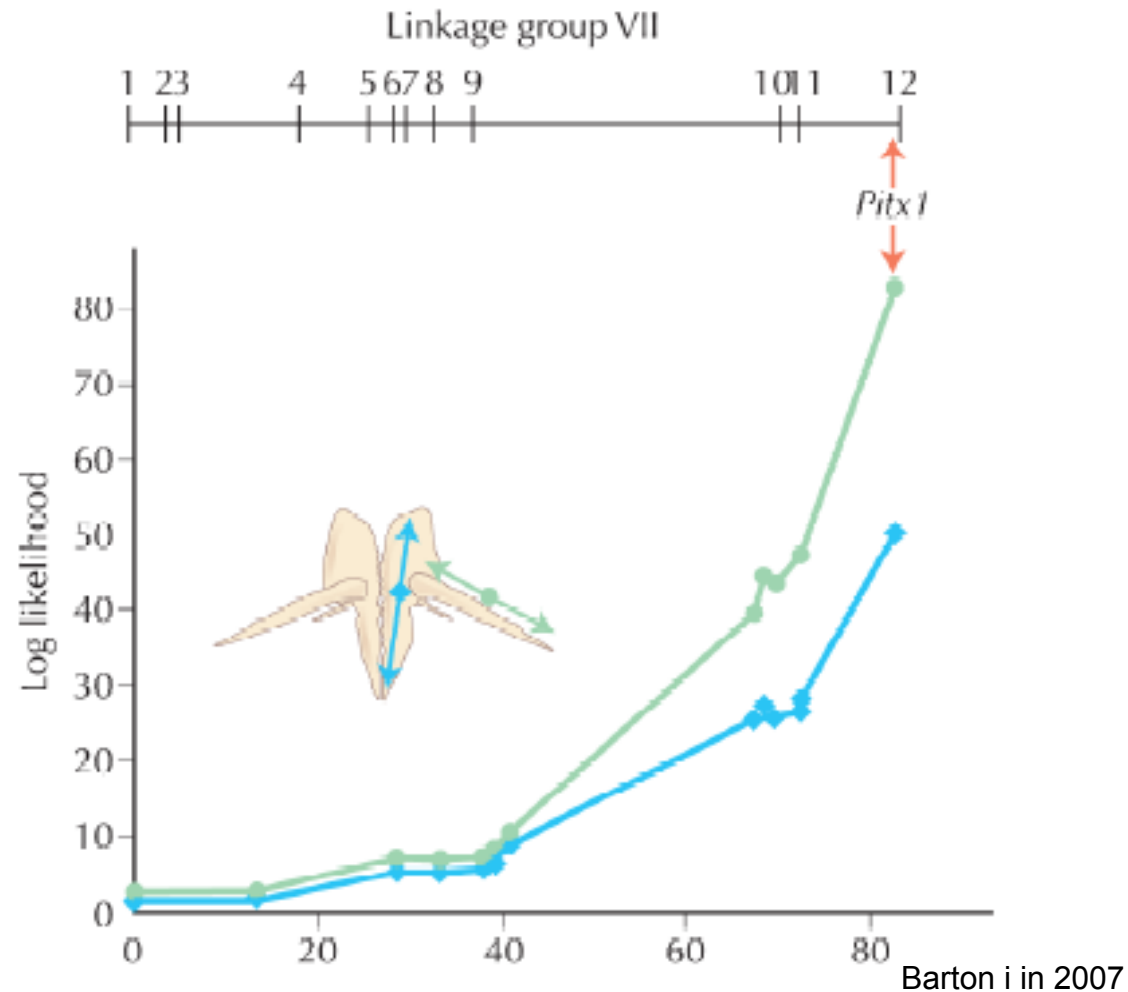
Geny warunkujące złożone cechy – QTL i geny-kandydaci



Geny warunkujące złożone cechy – QTL i geny-kandydaci

- Cierniki tworzą populacje słono i słodkowodne < 15 tys. lat
- W słonowodnych silnie rozwinięta miednica jest podstawą dla kolców brzusznych - adaptacja
- Analiza QTL wskazała region chromosomalny wyjaśniający 13 – 44% zmienności miednicy i kolców
- W tym regionie znajduje się gen *Pitx1* wpływający na rozwój miednicy i tylnych kończyn u myszy

Geny warunkujące złożone cechy – QTL i geny-kandydaci



Geny warunkujące złożone cechy – QTL i geny-kandydaci

- Cierniki tworzą populacje słono i słodkowodne < 15 tys. lat
- W słonowodnych silnie rozwinięta miednica jest podstawą dla kolców brzusznych - adaptacja
- Analiza QTL wskazała region chromosomalny wyjaśniający 13 – 44% zmienności miednicy i kolców
- W tym regionie znajduje się gen *Pitx1* wpływający na rozwój miednicy i tylnych kończyn u myszy
- Potwierdzono jego rolę stosując hybrydyzację in situ
- Brak różnic w części kodującej – jedynie w regulacji miejsca i poziomu ekspresji

Geny warunkujące złożone cechy – QTL i geny-kandydaci



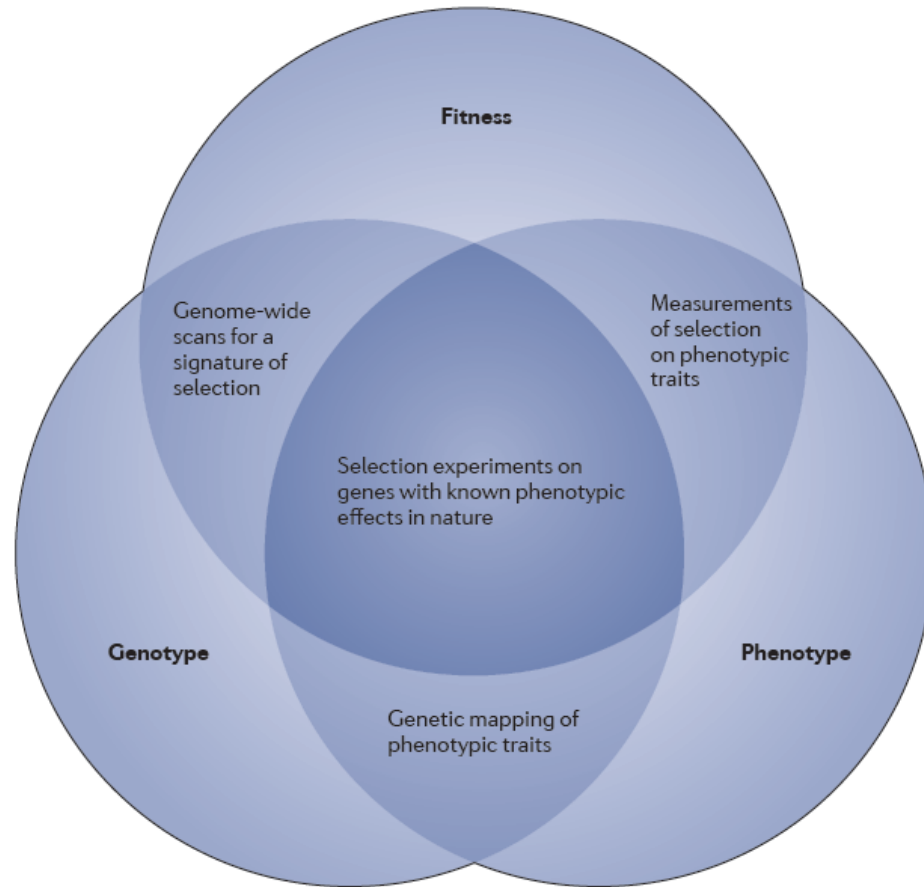
D Barton i in 2007

Mapowanie molekularne cech ilościowych

- Analiza QTL ma liczne ograniczenia
- nie wykrywa loci o małych efektach
- zawyża efekty loci o dużych efektach
- Ma ograniczoną rozdzielczość ca 20 cM ~ 20 mln bp, samo tempo rekombinacji stanowi ograniczenie, w znalezionym regionie może być wiele genów
- Badanie korelacji między cechą a markerami genetycznymi w populacji może być pomocne

Genetyczne podstawy adaptacji

- przekonująca identyfikacja genetycznych podstaw adaptacji jest trudna
- należy wykazać związek między wariantem genetycznym (allele), genotypem, fenotypem i dostosowaniem



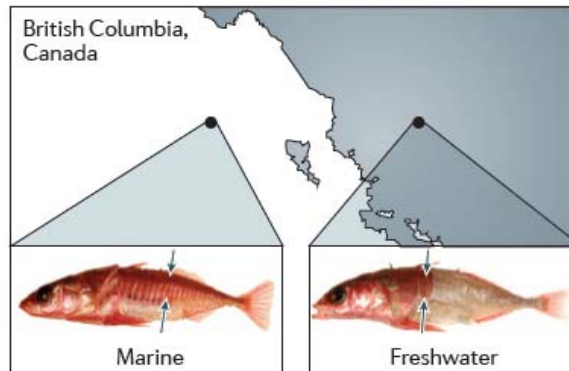
Barrett & Hoekstra 2011

Wykrycie genów warunkujących adaptację

Environment-phenotype
association

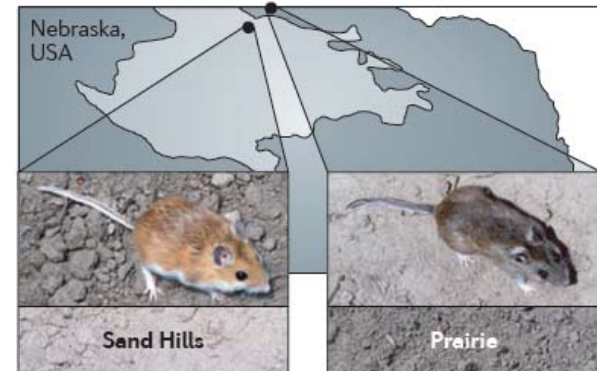
A Sticklebacks

Aa Armour plating in stickleback fish



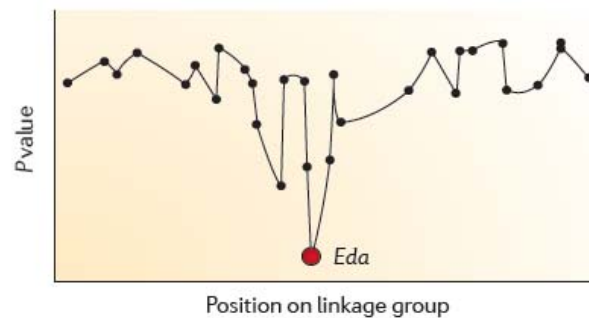
B Deer mice

Ba Coat colour in deer mice

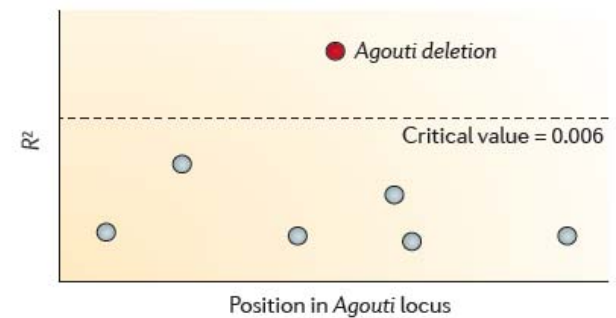


Identifying the gene
or mutation

Ab LD mapping



Bb Association mapping



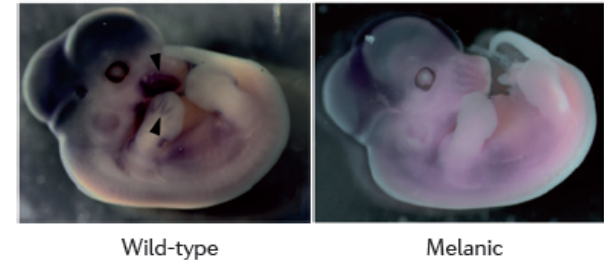
Wykrycie genów warunkujących adaptację

Test of gene function

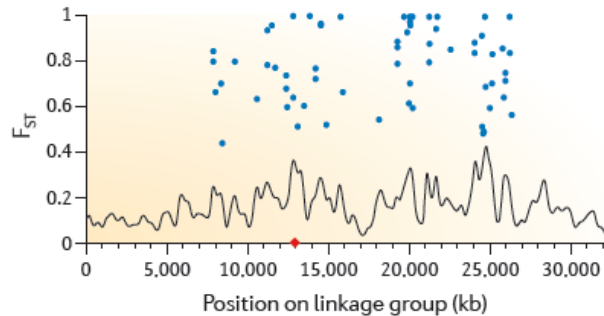
Ac Gain-of-function



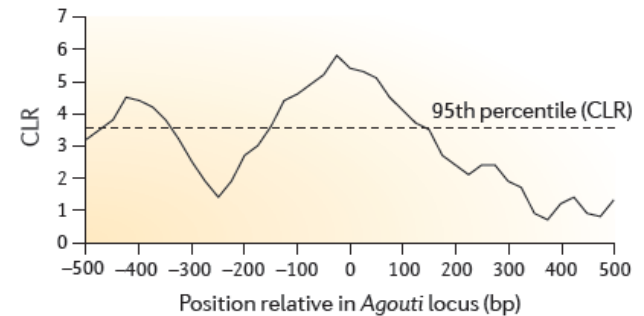
Bc Loss-of-function



Ad Genetic differentiation

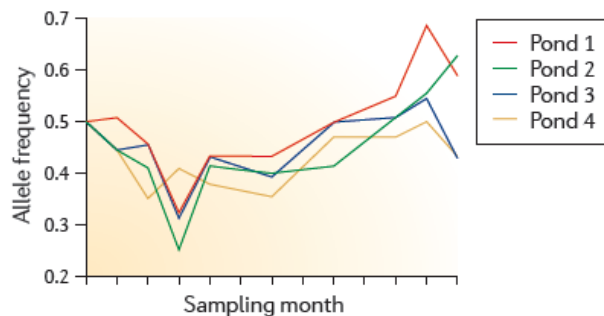


Bd Composite likelihood ratio test

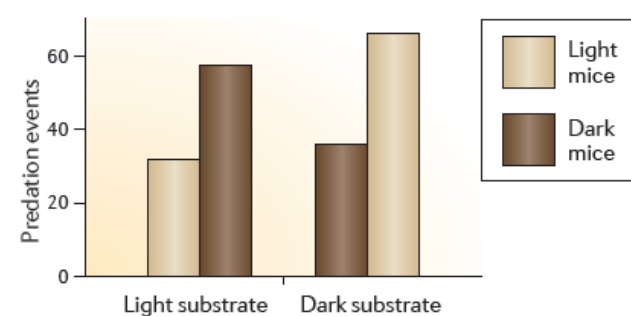


Signatures of selection

Ae Colonization experiment



Be Predation experiment

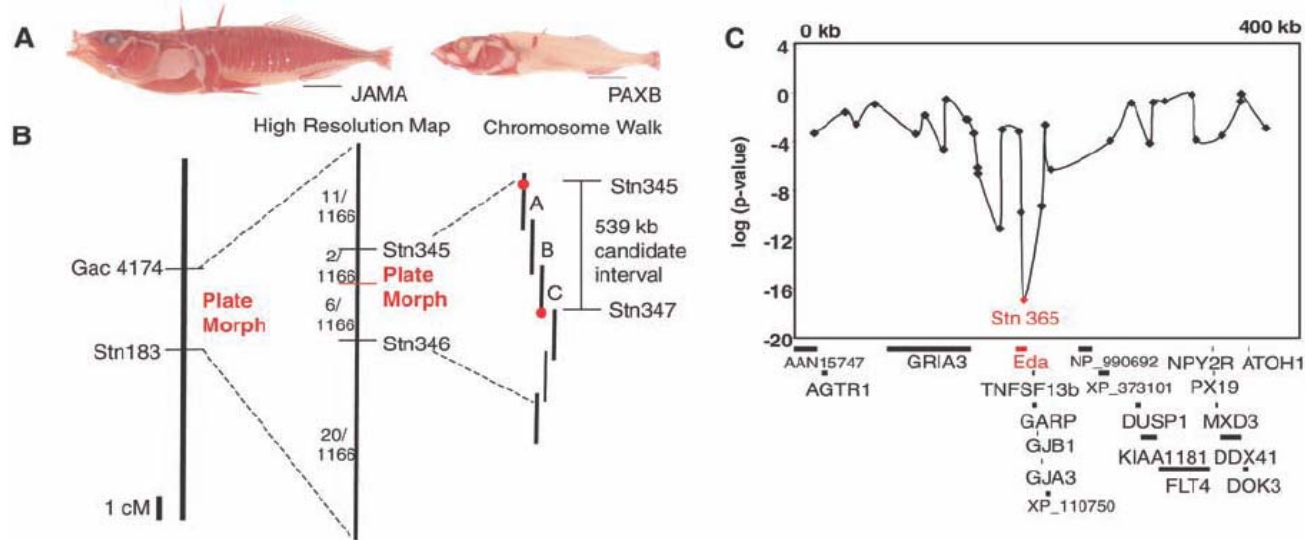


Experimental tests of selection

Eda u cierników – powtarzalne adaptacje

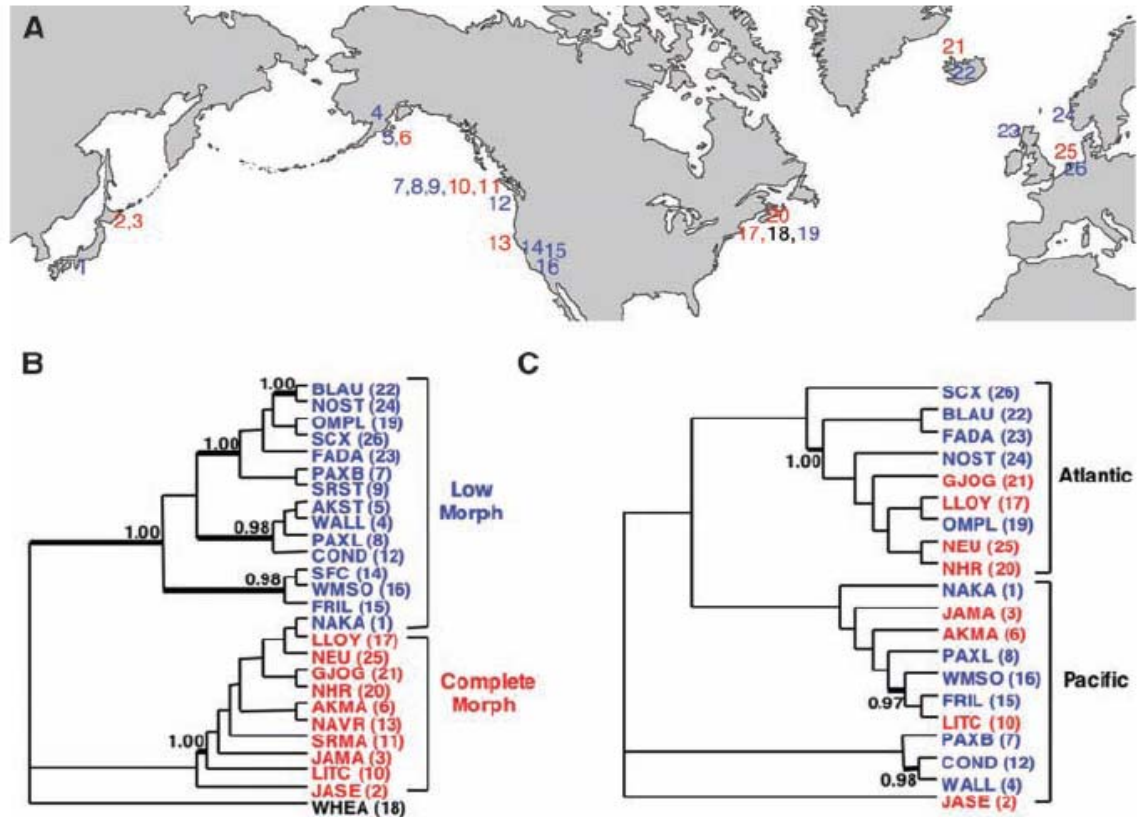
Gen *Eda* koduje białko ektodysplazynę, które jest częsteczką sygnałową zaangażowaną w tworzenie rozwój struktur ektodermalnych

u ciernika stwierdzono że różne linie alleliczne *Eda* determinują obecność lub brak płytek kostnych



Eda u cierników – powtarzalne adaptacje

każda z linii allelicznych jest monofiletyczna, allele warunkujące niską liczbę płytek występują z niską częstością w populacjach morskich podczas kolonizacji wód słodkich ich częstość wzrasta – nie powstają w wyniku nowych mutacji – adaptacja z istniejącej zmienności (standing genetic variation)



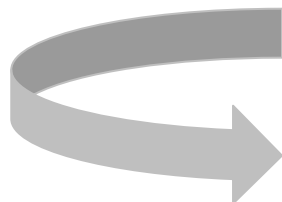
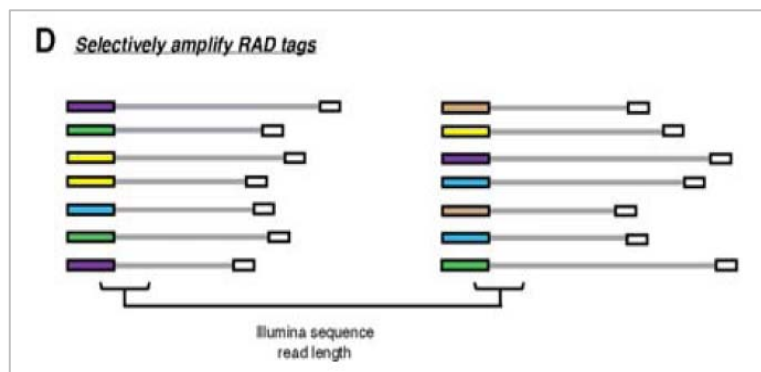
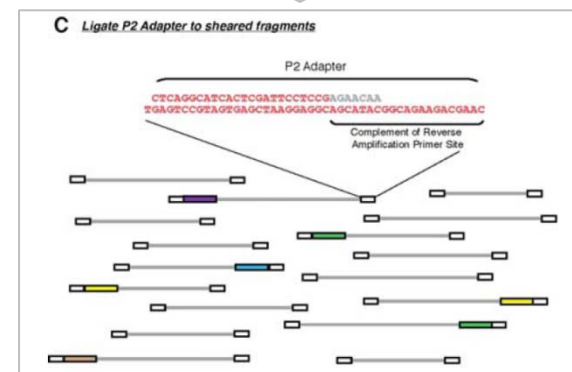
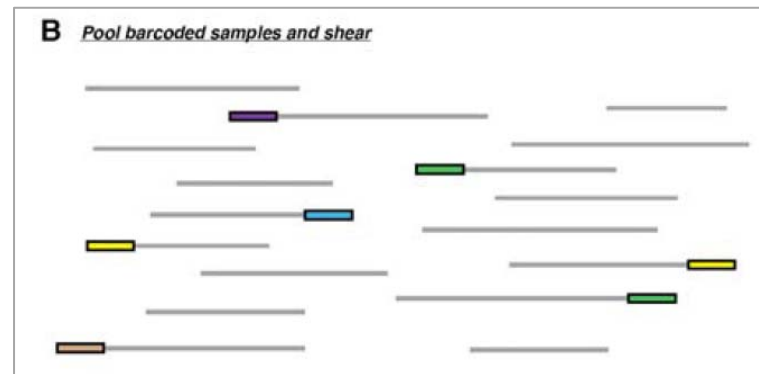
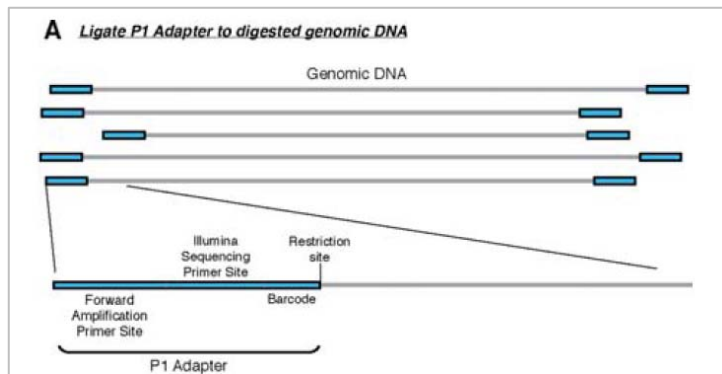
Genomika w poszukiwaniu adaptacji

- podejścia oparte na QTL i genach kandydatach dostarczyły wielu informacji, ale mają ograniczenia
- badanie ogólnogenomowego zróżnicowania między populacjami z różnych środowisk/różniących się cechami fenotypowymi
- nie musimy mieć kandydatów
- wykrywane regiony genomów często będą niewielkie – dużo rekombinacji w historii populacji – zawężenie do ograniczonej liczby kandydatów
- regiony wykryte metodami genomiki to jedynie – kandydaci – potrzebna weryfikacja eksperymentalna

RAD-tag (Restriction-site Associated DNA)

- Określona część genomu, fragmenty sąsiadujące z miejscami restrykcyjnymi danego enzymu, jest sekwencjonowana u wielu osobników
- W zsekwencjonowanych fragmentach wykrywa się SNP
- Osobniki można znakować za pomocą krótkich sekwencji (tagów) i sekwencjonować pule
- Liczba zsekwencjonowanych fragmentów zależy od wielkości genomu, zawartości GC i użytego enzymu
 - *NsiI* 342 miejsc cięcia / Mbp
 - *NotI* 1 miejsce cięcia / Mbp i wiele innych enzymów
- Najwięcej informacji gdy dostępny genom referencyjny

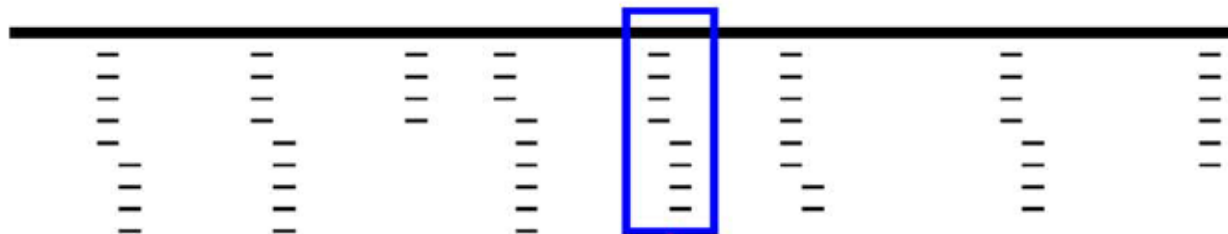
RAD-tagi



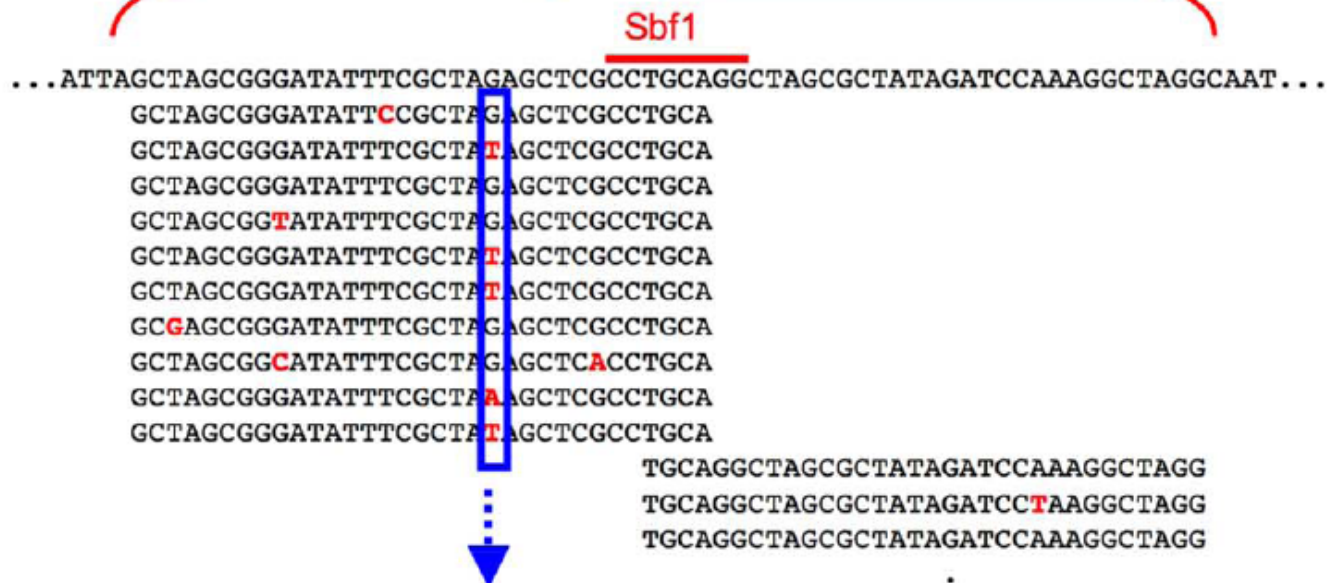
sekwencjonowanie
Illumina

RAD-tagi

A



B



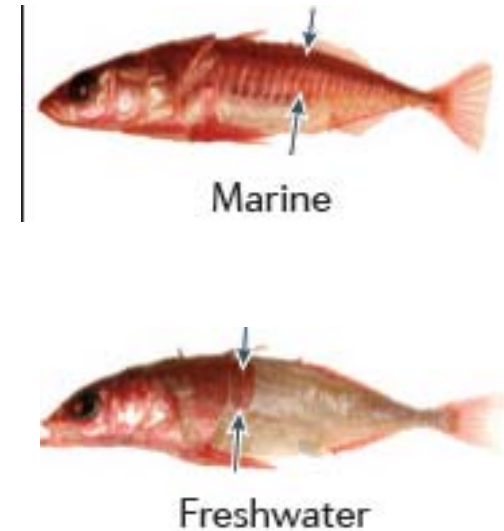
RAD-tagi (Restriction-site Associated DNA)

- Określona część genomu, fragmenty sąsiadujące z miejscami restrykcyjnymi danego enzymu, jest sekwencjonowana u wielu osobników
- W zsekwencjonowanych fragmentach wykrywa się SNP
- Osobniki można znakować za pomocą krótkich sekwencji (tagów) i sekwencjonować pule
- Liczba zsekwencjonowanych fragmentów zależy od wielkości genomu, zawartości GC i użytego enzymu
 - *NsiI* 342 miejsc cięcia / Mbp
 - *NotI* 1 miejsce cięcia / Mbp i wiele innych enzymów
- Najwięcej informacji gdy dostępny genom referencyjny

Adaptacje do słodkowodności u cierników



Hohenlohe et al. 2010

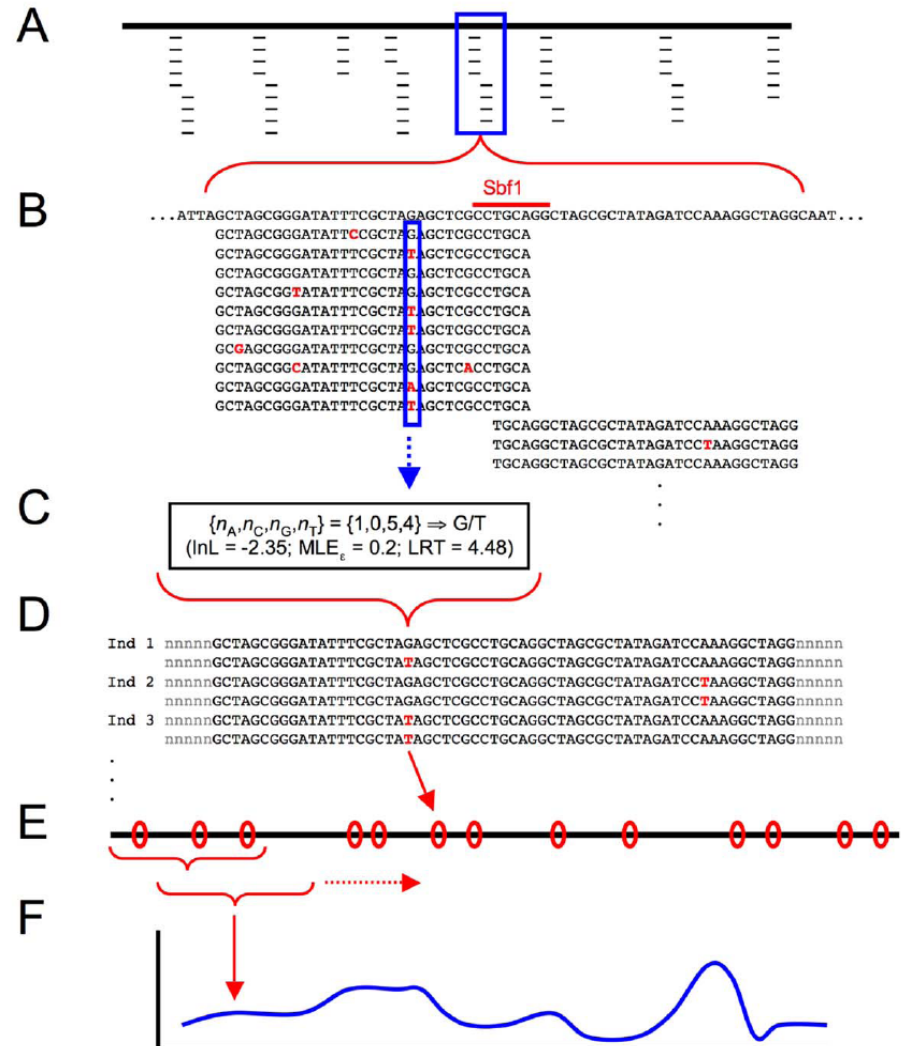


Barrett & Hoekstra 2011

Jakie różnice w genomie?
Czy są powtarzalne?
Zsekwencjonowany
i złożony genom ok. 400 Mpz

Adaptacje do słodkowodności u cierników

- 2 populacje oceaniczne
- 3 populacje słodkowodne
- 20 osobników na populację
- > 40 000 RAD-tagów
- statystyczna ocena genotypu w każdej pozycji nukleotydów
- > 45 000 SNP
- RAD-tagii zmapowane do genomu



Adaptacje do słodkowodności u cierników

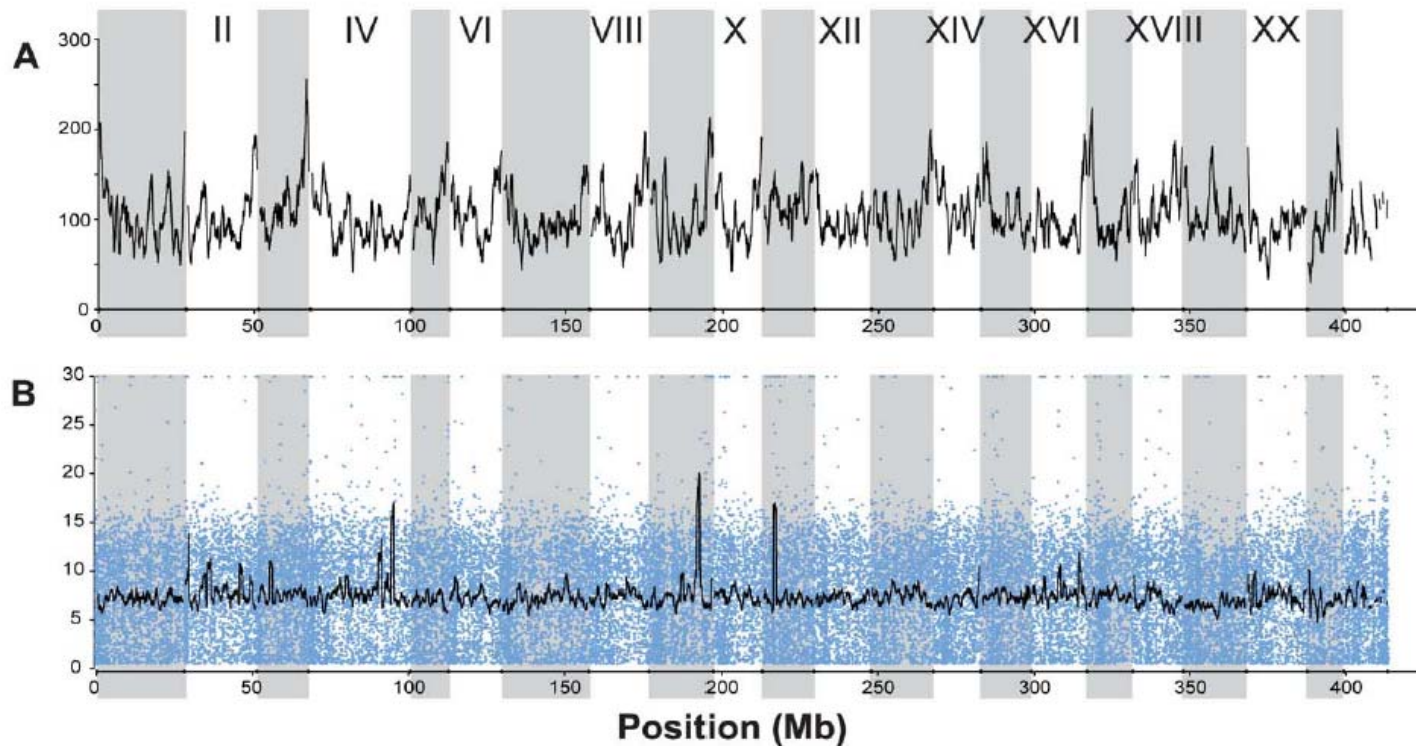


Figure 3. Depth of RAD sequencing coverage. (A) Number of RAD tags sequenced per 1-Mb sliding window across the genome. Each RAD tag represents either 30 or 47 bp of sequence data (see Table S1). Vertical gray shading indicates Linkage Groups I through XXI, followed by all unassembled scaffolds greater than 1 Mb in length. Not all RAD tags were sequenced in all individuals, because of both random sampling in the sequencing process and polymorphism in the restriction enzyme recognition site. (B) Sequencing depth per RAD tag per individual from one sample run (22 May 2009, lane 7; see Table S1). Blue dots represent the average number of reads per individual across 16 individuals sampled for each RAD tag. The black line shows the mean depth per individual in a 1-Mb sliding window. A total of 5,597,895 barcoded and aligned sequence reads from 16 individuals were generated from this run.
doi:10.1371/journal.pgen.1000862.g003

Adaptacje do słodkowodności u cierników

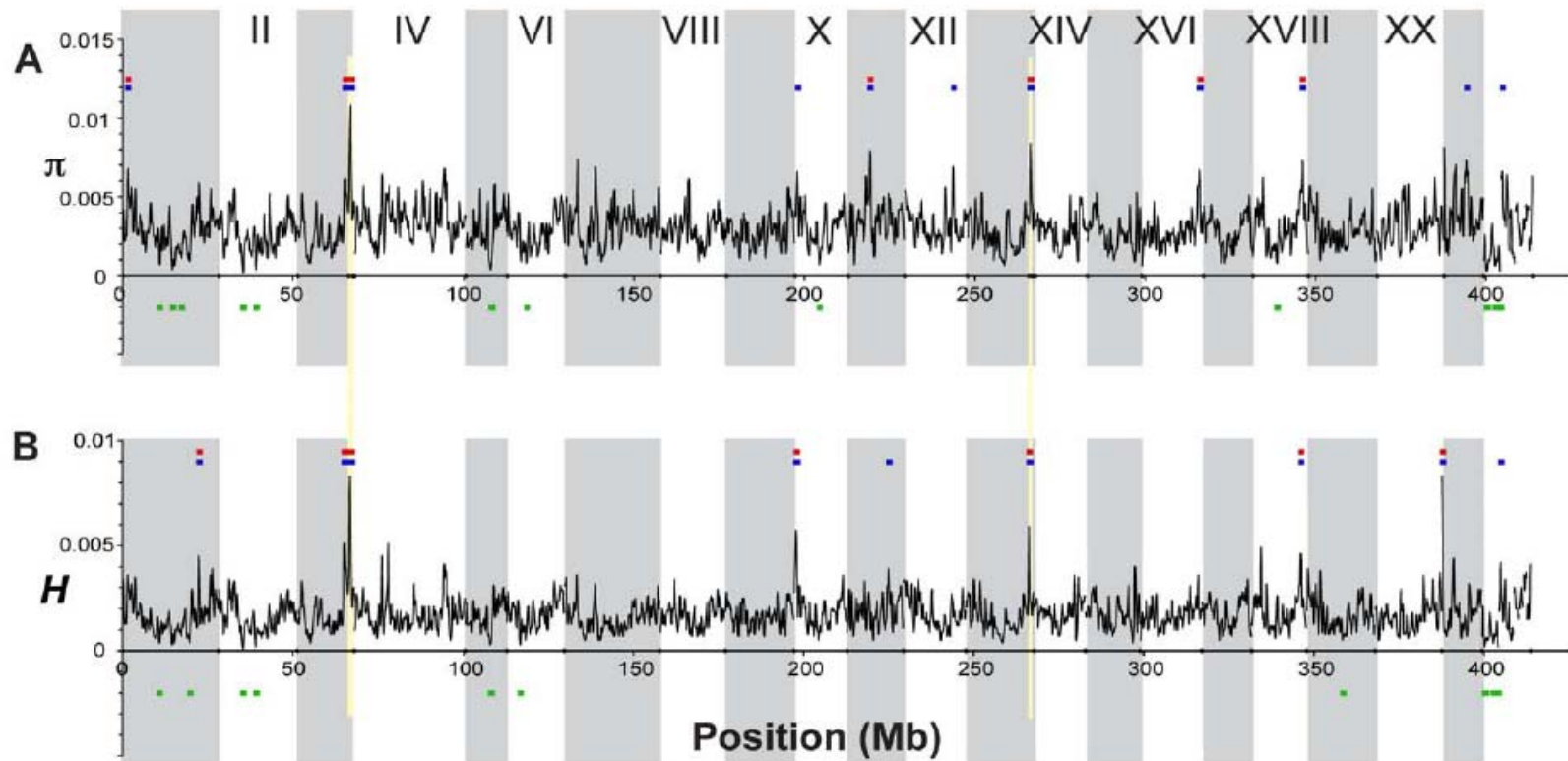


Figure 4. Genome-wide patterns of nucleotide diversity. Each plot shows a smoothed distribution of the statistical measure across the genome (black lines). Colored bars above and below the distributions indicate regions of significantly elevated ($p \leq 10^{-5}$, blue; $p \leq 10^{-7}$, red) and reduced ($p \leq 10^{-5}$, green) values, assessed by bootstrap resampling. Vertical shading indicates the 21 linkage groups and the unassembled scaffolds greater than 1 Mb in length, and gold shading indicates two regions showing evidence of balancing selection as discussed in the text. (A) Nucleotide diversity (π) across all five stickleback populations sampled. (B) Heterozygosity (H) across all five populations.

doi:10.1371/journal.pgen.1000862.g004

Adaptacje do słodkowodności u cierników

Wysoka heterozygotyczność,
wysoka różnorodność
nukleotydów i małe
zróżnicowanie
międzypopulacyjne sugerują
działanie doboru
równoważącego w grupie
sprzężeń III,
znajduje się tam wiele genów
zaangażowanych w walkę z
patogenami

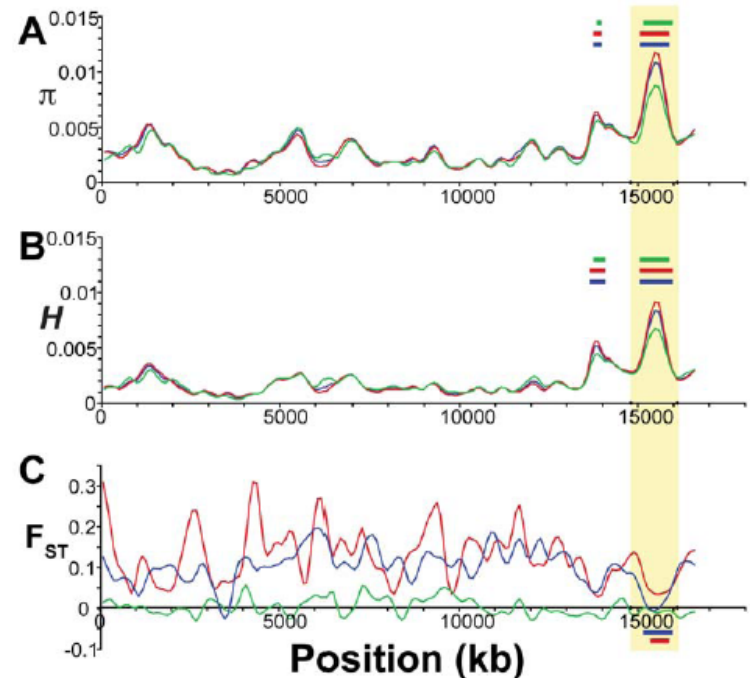
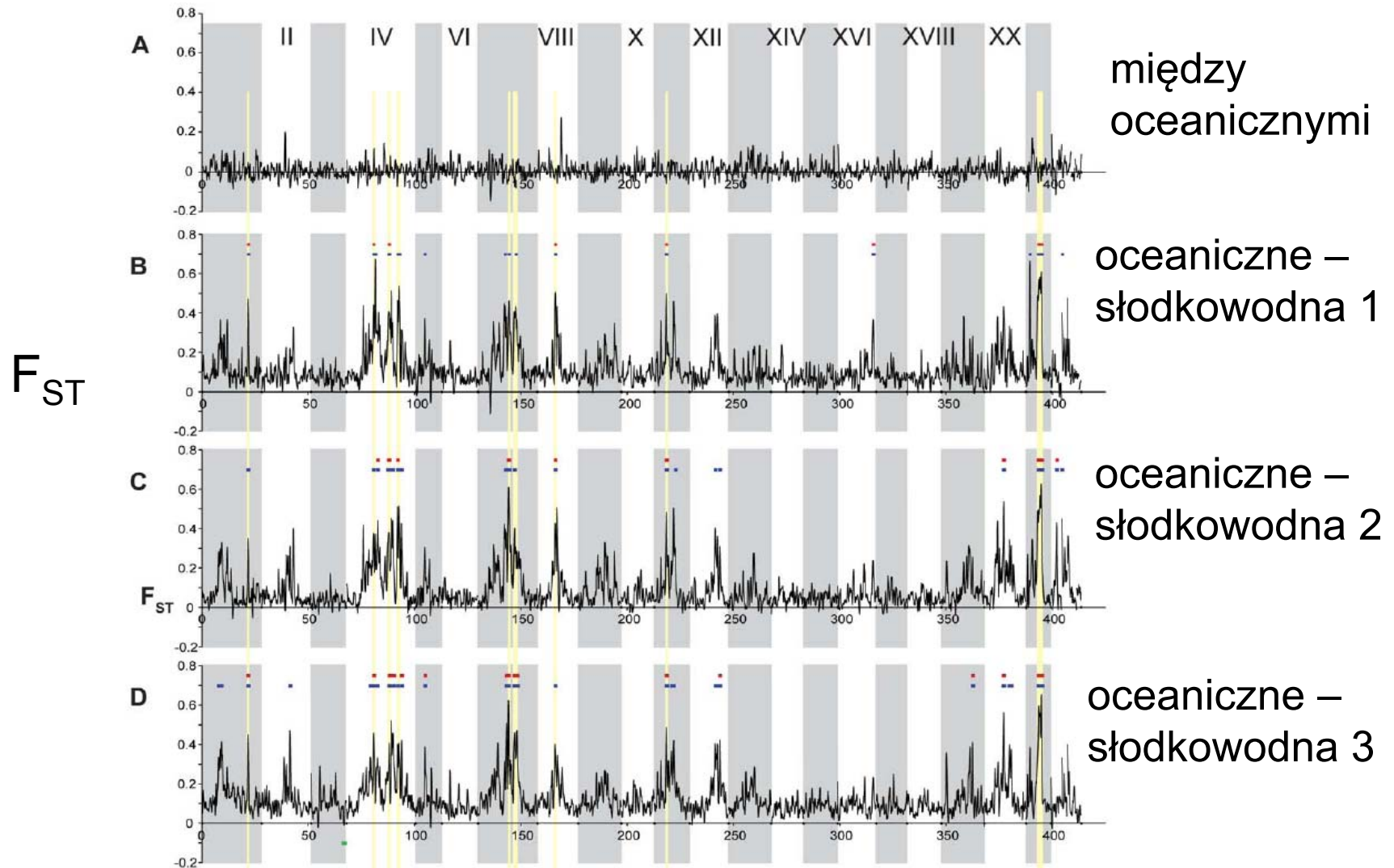


Figure 5. Evidence for balancing selection on Linkage Group III. Population genetic measures plotted along Linkage Group III. (A) Nucleotide diversity (π) and (B) heterozygosity (H) across all five (blue), the three freshwater (red), and the two oceanic (green) populations. (C) Population differentiation (F_{ST}) between oceanic and freshwater (blue), among freshwater (red), and between oceanic (green) populations. Colored bars indicate significant ($p \leq 10^{-5}$) regions of elevated (above the plots) or reduced (below the plots) values of each statistic for the corresponding set of populations. Vertical yellow shading indicates the region of putative balancing selection used for candidate gene annotation.

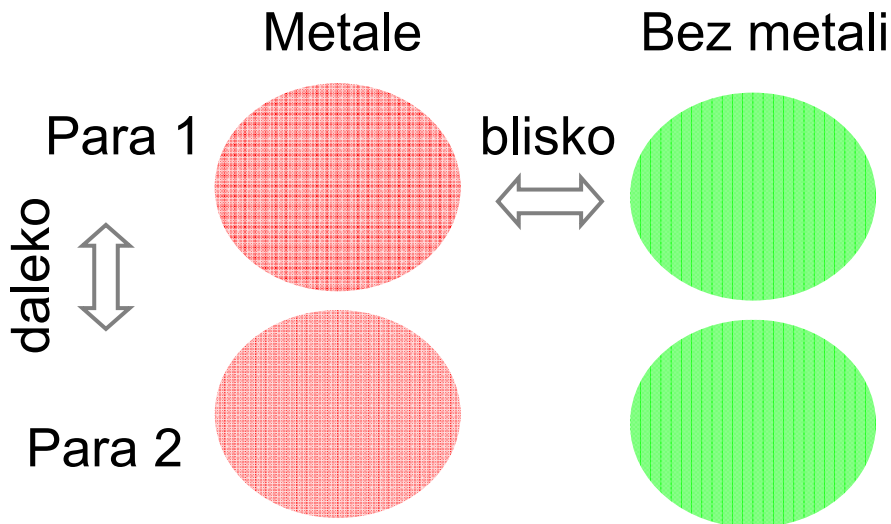
doi:10.1371/journal.pgen.1000862.g005

Adaptacje do słodkowodności u cierników



Adaptacje *Arabidopsis* do metali ciężkich

- *Arabidopsis lyrata* – populacje na glebach normalnych i z wysoką zawartością metali ciężkich
- Które geny odpowiadają za adaptacje do wysokiego stężenia metali ciężkich?
- Znany, niewielki (ok. 200 Mb) genom



Turner et al. 2010

Adaptacje *Arabidopsis* do metali ciężkich

- Z każdej populacji po 25 roślin, z każdej taka sama ilość DNA
- Tę mieszaninę zsekwencjonowano techniką Illumina dla każdej populacji osobno do pokrycia śr. 10x w każdej populacji
- Takie pokrycie pozwoli wykryć allele (SNP) których częstości różnią się mocno między populacjami
- Jeżeli te same SNP wystąpią w obu parach to będą dobrymi kandydatami na związek z adaptacją
- znaleziono ok. 80 genów-kandydatów, wśród nich wiele zaangażowanych w detoksyfikację metali ciężkich oraz transport wapnia i magnezu
- zmienność trzech genów kandydatów zbadano w populacjach szkockich pokazując podobne zróżnicowanie

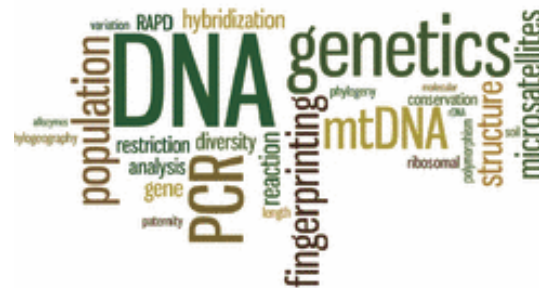


Turner et al. 2010

Przyszłość ekologii molekularnej?

- dostępność markerów będzie coraz łatwiejsza
- mniej ograniczeń technicznych
- szybsze analizy laboratoryjne
- trudniejsza analiza i składowanie danych

1995



2000



2005



2010



Czas angażowany w etapy projektu

