

# Genetyka i ewolucja

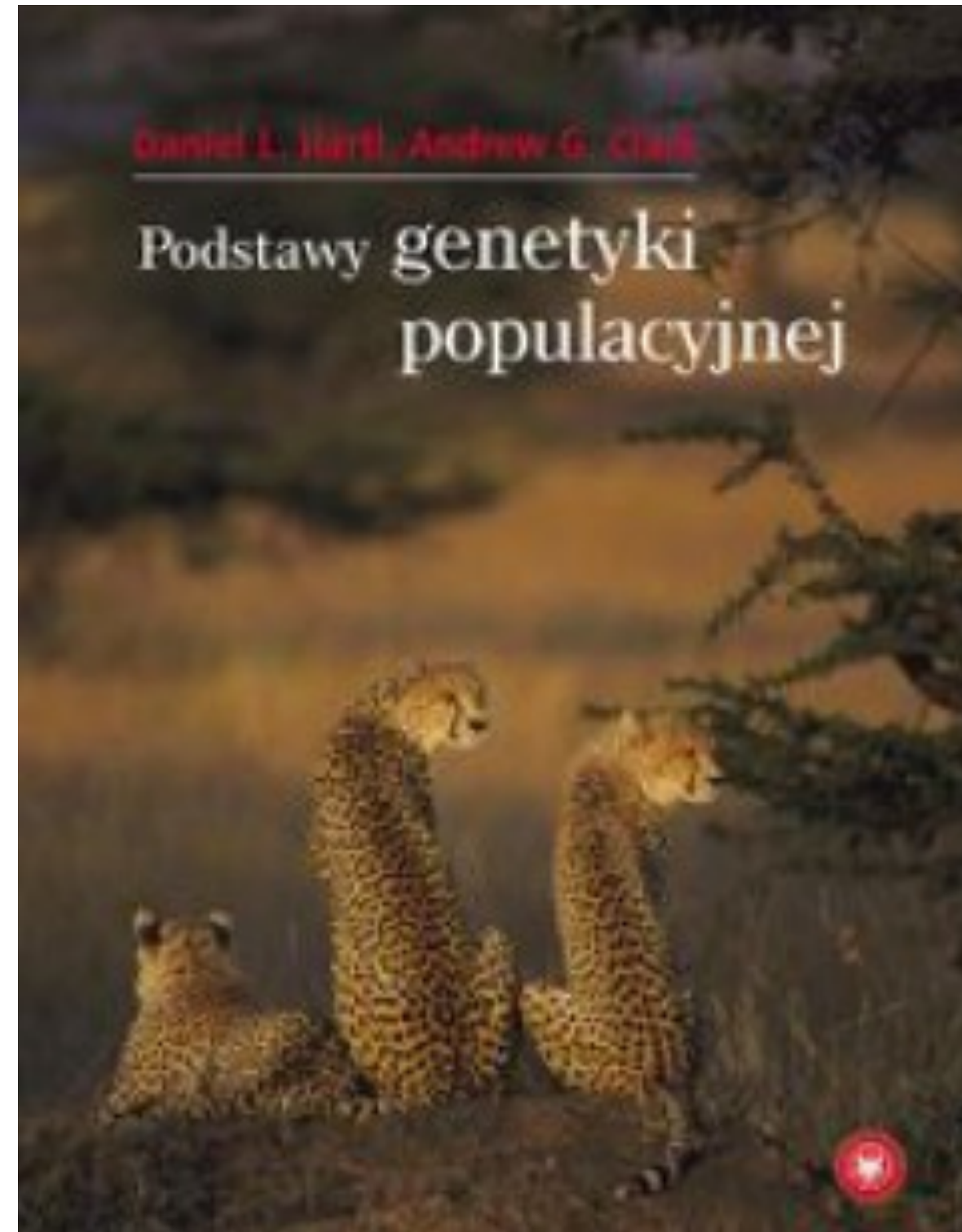
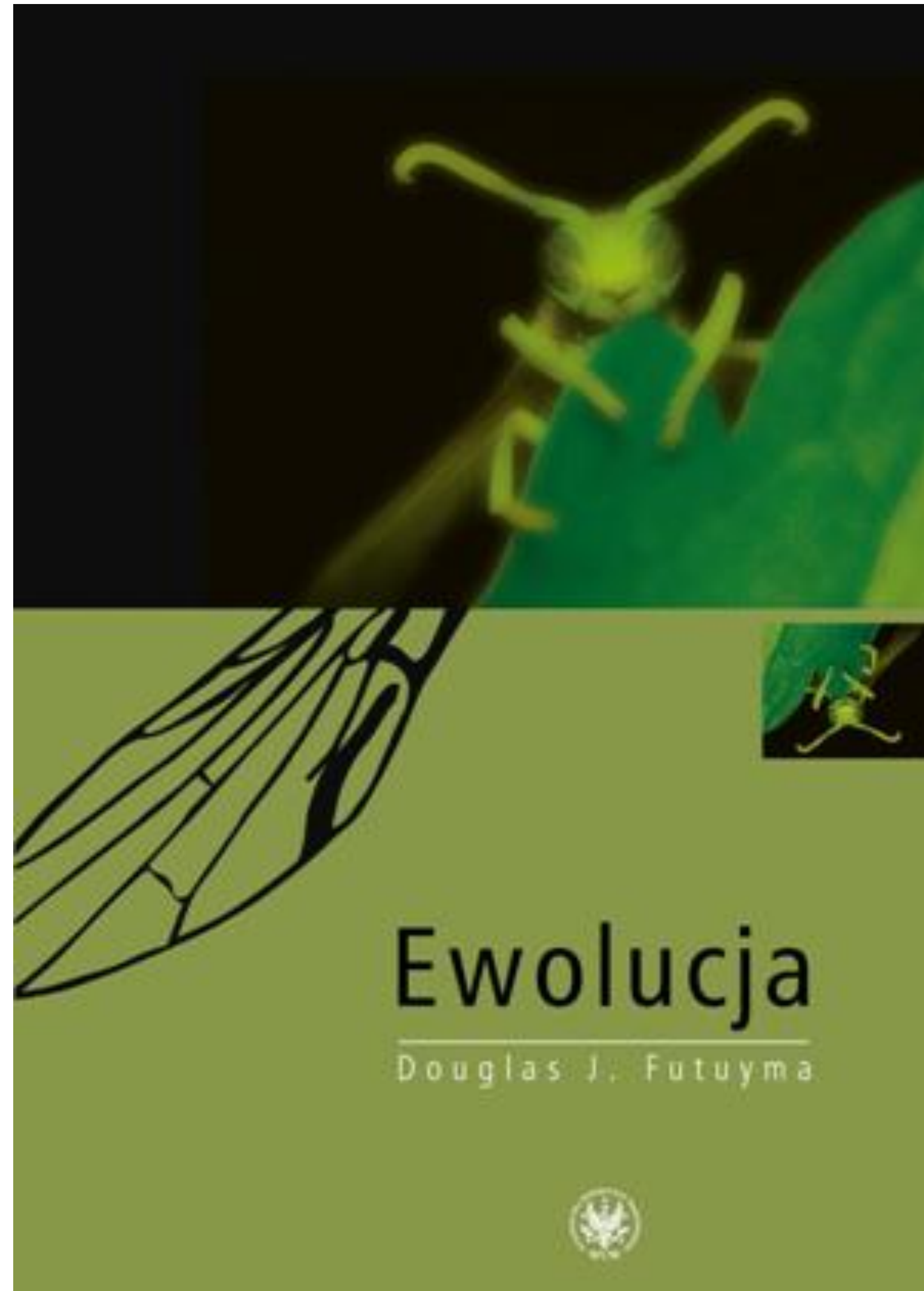
# Podstawy genetyki populacji

---

Genetyka mendlowska i ewolucja.

# Podręczniki

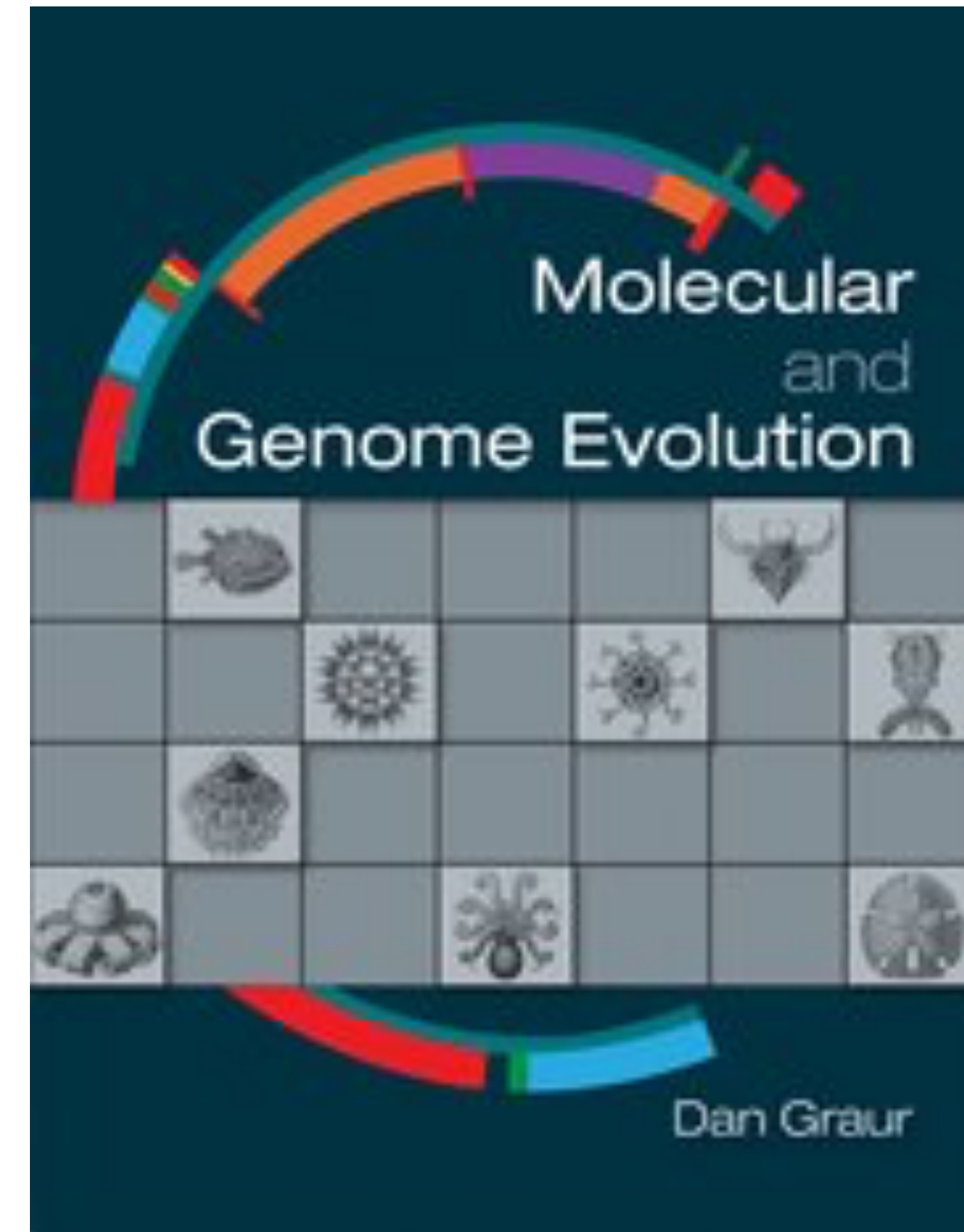
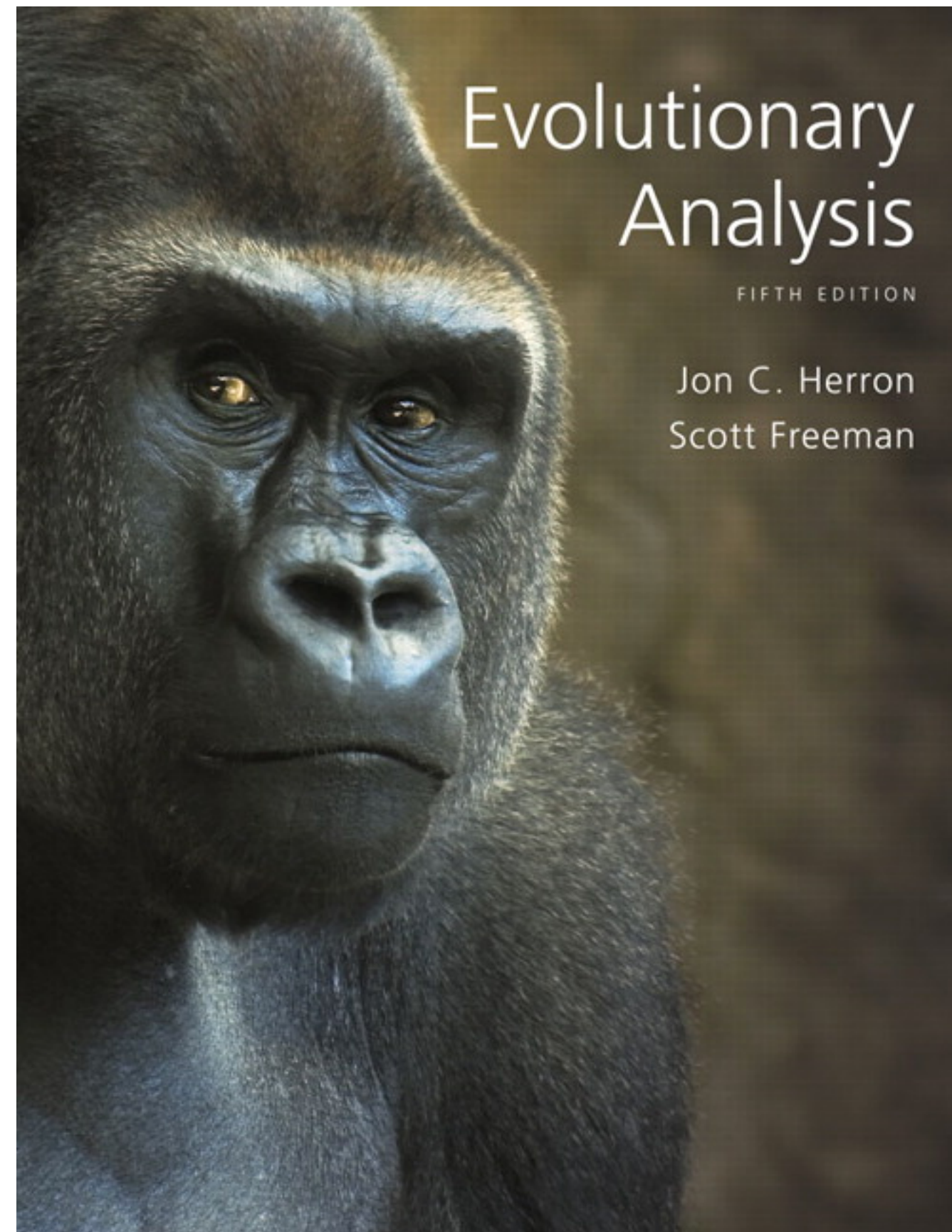
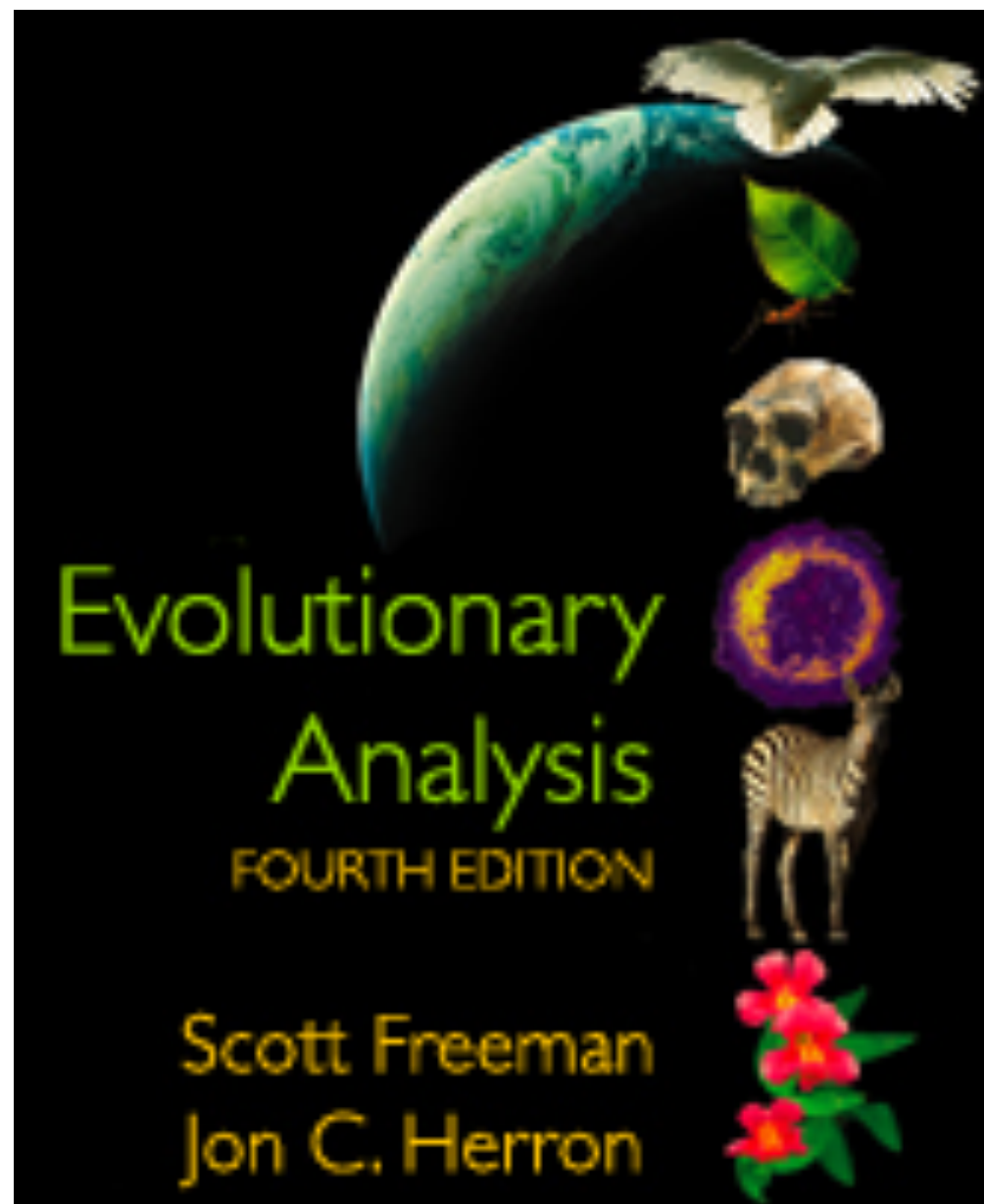
---





# Dla zainteresowanych

---



<https://www.pearsonhighered.com/program/Herron-Evolutionary-Analysis-5th-Edition/PGM296285.html>



# Ewolucja biologiczna

---

- Znaczenie ogólne:
  - proces zmian informacji genetycznej organizmów (częstości i rodzaju alleli w populacji),
  - które to zmiany są przekazywane z pokolenia na pokolenie
    - dotyczy populacji, nie pojedynczego osobnika
    - dotyczy zmian **dziedziczonych**
- Dziedziczenie z modyfikacją (*descent with modification* - Darwin)

# Ewolucja biologiczna

---

- Zjawisko (fakt)
- Teoria ewolucji
- Historia zmian ewolucyjnych



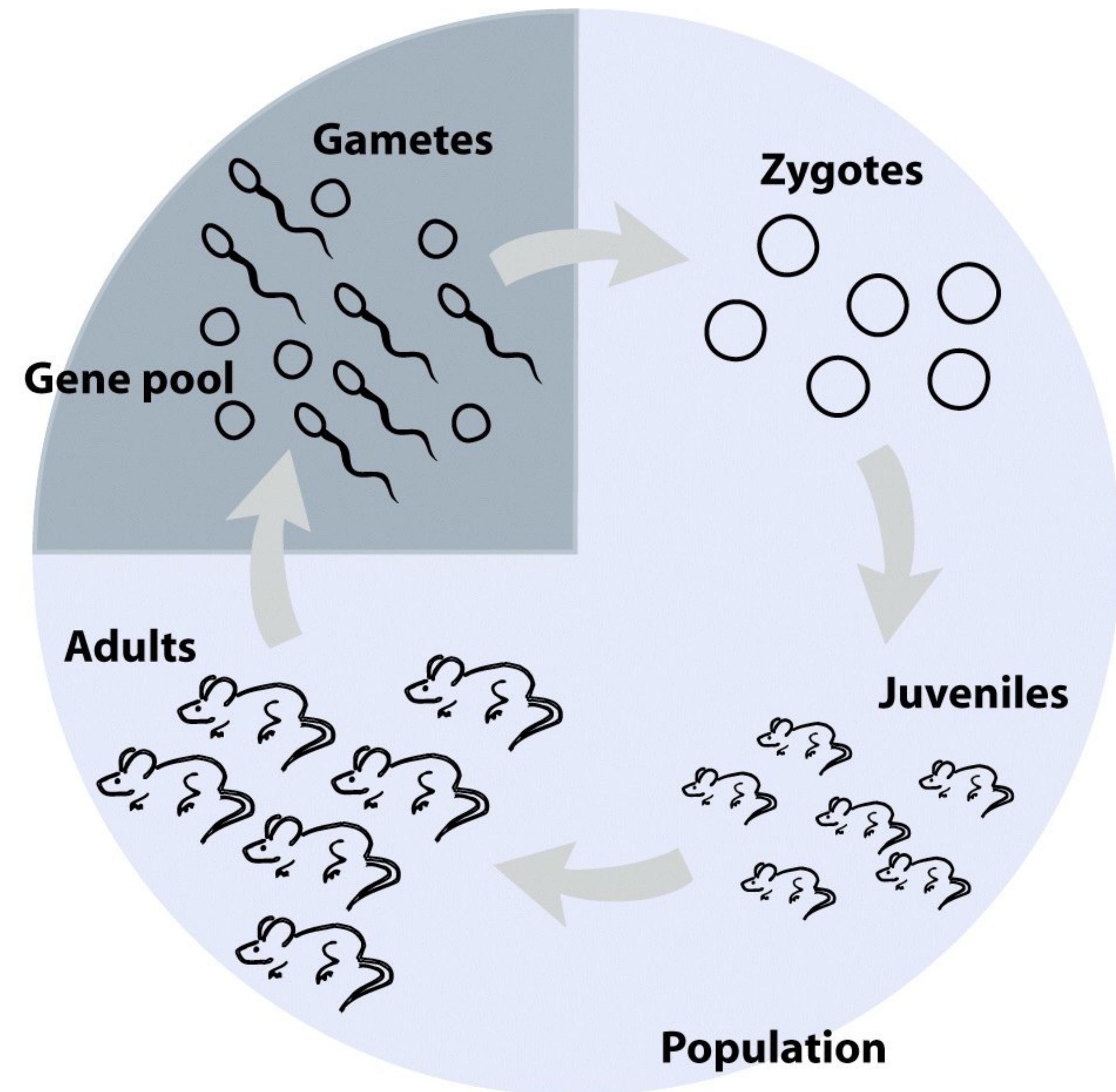
# Syntetyczna teoria ewolucji

---

- Pierwsza synteza: połączenie teorii ewolucji Darwina z genetyką mendlowską na poziomie populacji
- W naturalnych populacjach występują różne allele genów
- Częstość cech fenotypowych w populacji zależy od częstości alleli i genotypów
- Ewolucja jako zmiana częstości alleli w populacji z pokolenia na pokolenie

# Populacja

- Grupa krzyżujących się ze sobą osobników oraz ich potomstwo
- Zbiór wszystkich alleli populacji – pula genowa



# Najprostszy model

---

- Populacja  $N$  organizmów diploidalnych
- Rozważany jeden A gen o dwóch allelach  $A_1$  i  $A_2$
- Częstości alleli, odpowiednio  $p$  i  $q$

$$p + q = 1$$



# Populacja w stanie równowagi

---

- Liczebność populacji bardzo duża ( $N \sim \infty$ )
- Całkowicie losowe krzyżowanie (panmiksja)
- Sukces reprodukcyjny nie zależy od genotypu genu A
- Brak migracji
- Nie zachodzą mutacje zmieniające A1 w A2 i vice versa

# Równowaga Hardy'ego-Weinberga

---

Jeżeli częstości alleli  $A_1$  i  $A_2$  to odpowiednio  $p$  i  $q$

to częstości genotypów

$$A_1A_1 \quad p^2$$

$$A_1A_2 \quad pq + qp = 2pq$$

$$A_2A_2 \quad q^2$$

# Równowaga Hardy'ego-Weinberga

---

Częstości alleli w populacji w stanie równowagi się nie zmieniają

Gamety  $A_1$  - wszystkie gamety homozygot  $A_1A_1$  i połowa gamet heterozygot  $A_1A_2$

W kolejnym pokoleniu:

$$p' = p^2 + \frac{2pq}{2} = p^2 + pq$$

$$q = 1 - p$$

$$p' = p^2 + p \cdot (1 - p) = p^2 + p - p^2 = p$$







# Równowaga Hardy'ego-Weinberga

---

- W populacji będącej w równowadze H-W częstości alleli nie zmieniają się
- Nie przebiega ewolucja
- Mechanizmy zaburzające równowagę H-W mogą być mechanizmami ewolucji

# Wsobność

- Częstsze krzyżowanie osobników spokrewnionych
- Jedna z form krzyżowania asortatywnego – preferencji wobec osobników o zbliżonym fenotypie
- Forma skrajna - samozapłodnienie

**Each individual produces offspring by selfing:**

**$A_1A_1$  individuals produce  $A_1A_1$  offspring**

**$A_1A_2$  individuals produce  $A_1A_1$ ,  $A_1A_2$ , and  $A_2A_2$  offspring in a 1:2:1 ratio**

**$A_2A_2$  individuals produce  $A_2A_2$  offspring**

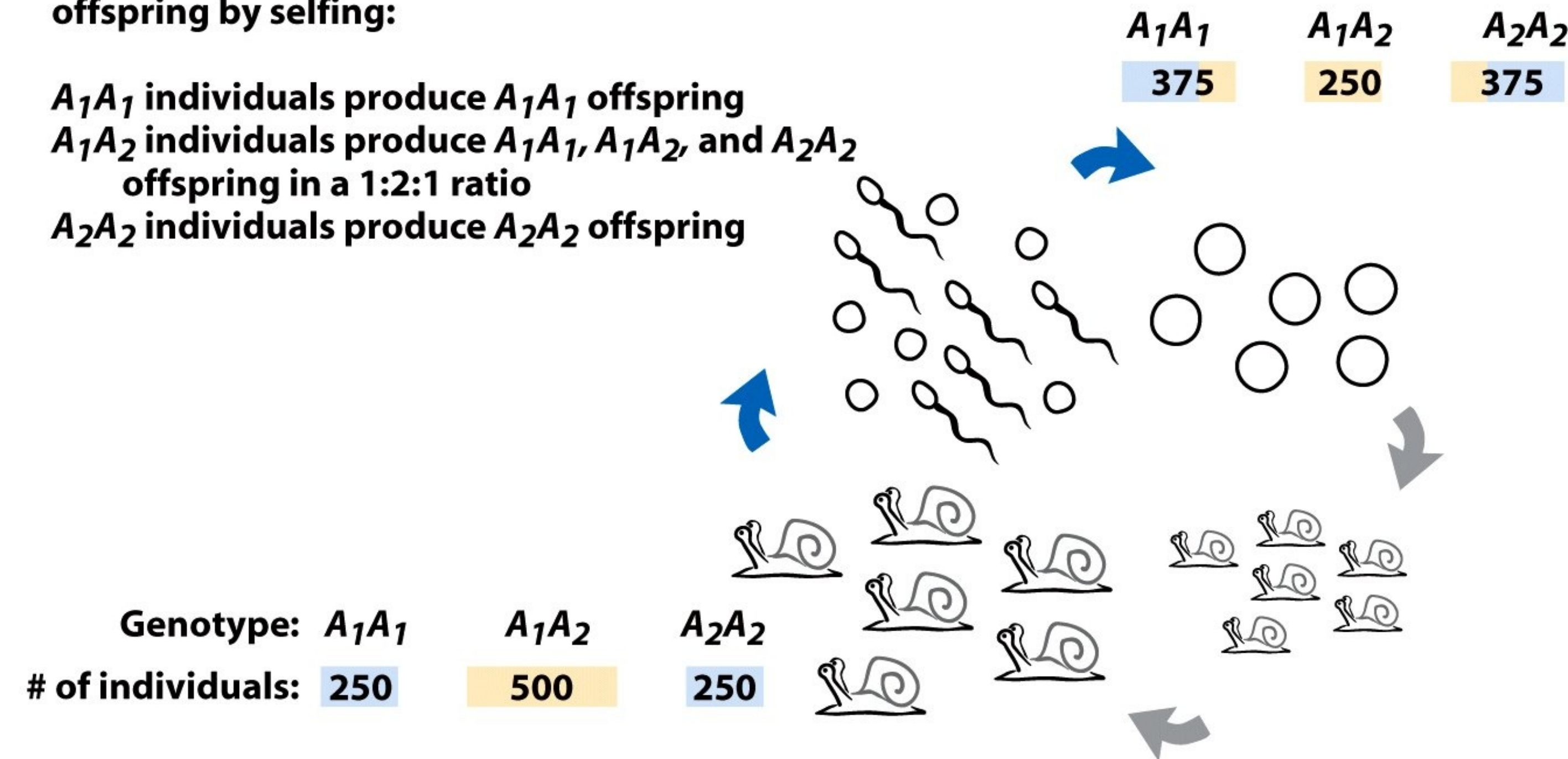


Figure 7-25a Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.

# Wsobność

Krzyżowanie wsobne nie zmienia częstości alleli, ale wpływa na częstość genotypów.

| Genotype:         | $A_1A_1$ | $A_1A_2$ | $A_2A_2$ |              |
|-------------------|----------|----------|----------|--------------|
| # of individuals: | 250      | 500      | 250      | Generation 0 |
|                   | 375      | 250      | 375      | Generation 1 |
|                   | 437.5    | 125      | 437.5    | Generation 2 |
|                   | 468.75   | 62.5     | 468.75   | Generation 3 |

Figure 7-25b Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.

Populacja wsobna – niedobór heterozygot, nadmiar homozygot.



# Współczynnik wsobności

---

- $F$  – prawdopodobieństwo, że oba allele u osobnika są identyczne przez wspólne pochodzenie
- Przy samozapłodnieniu (1 pokolenie)  $F = 1/2$
- Przy krzyżowaniu rodzeństwa  $F=1/4$
- Ogólnie częstości genotypów:

A1A1     $p^2(1-F)+pF = p^2 - p^2F + pF$ , ponieważ  $pF > p^2F$ , to **częstość większa od  $p^2$**

A1A2     $2pq(1-F) = 2pq - 2pqF$ , **częstość mniejsza od  $2pq$**

A2A2     $q^2(1-F)+qF$

- Odchylenie liczby heterozygot od przewidywanej pozwala oszacować wsobność



# Depresja wsobna

- Rzadkie allele recesywne ujawniają się w fenotypach w populacji
- Spada dostosowanie populacji

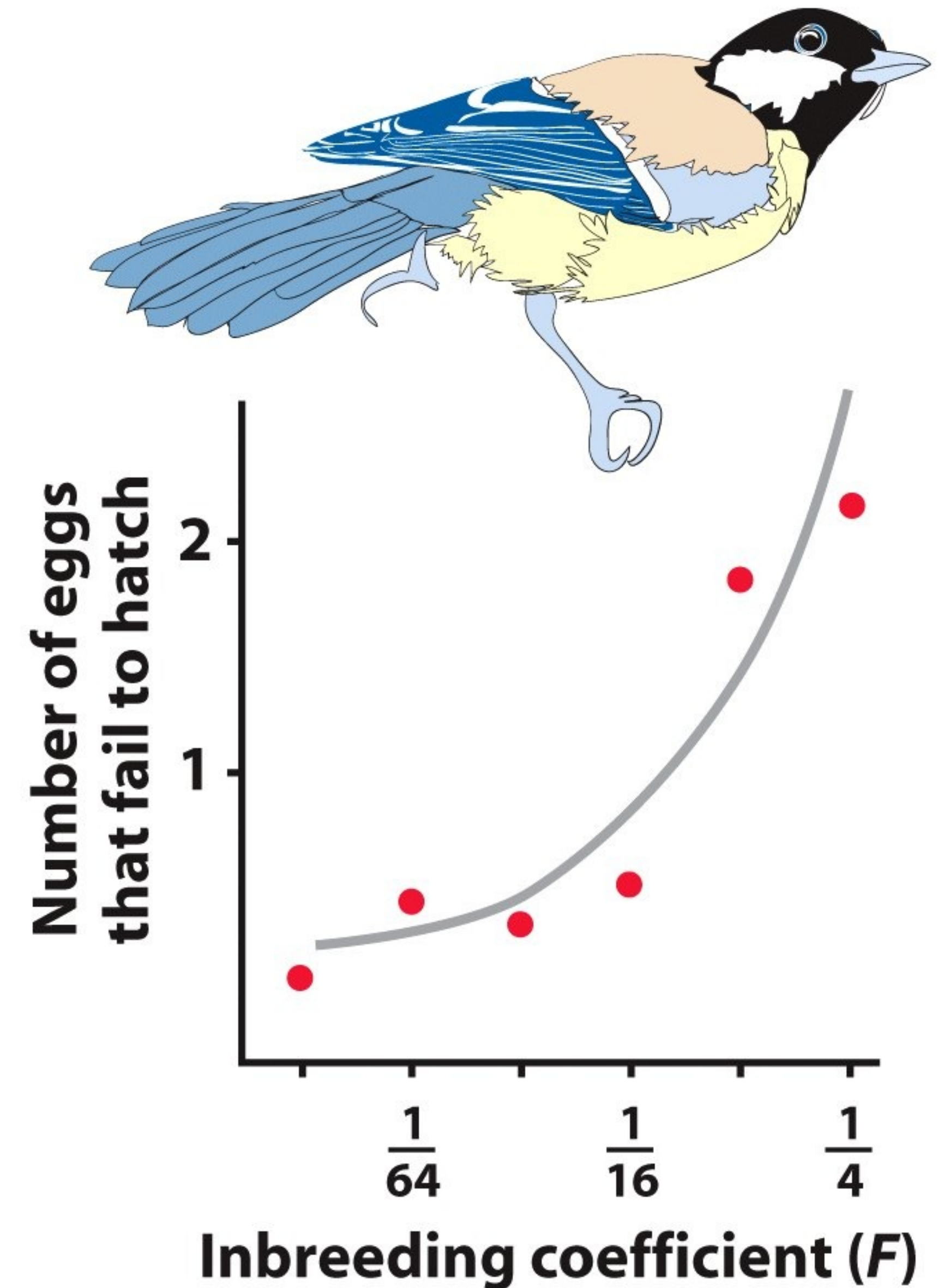
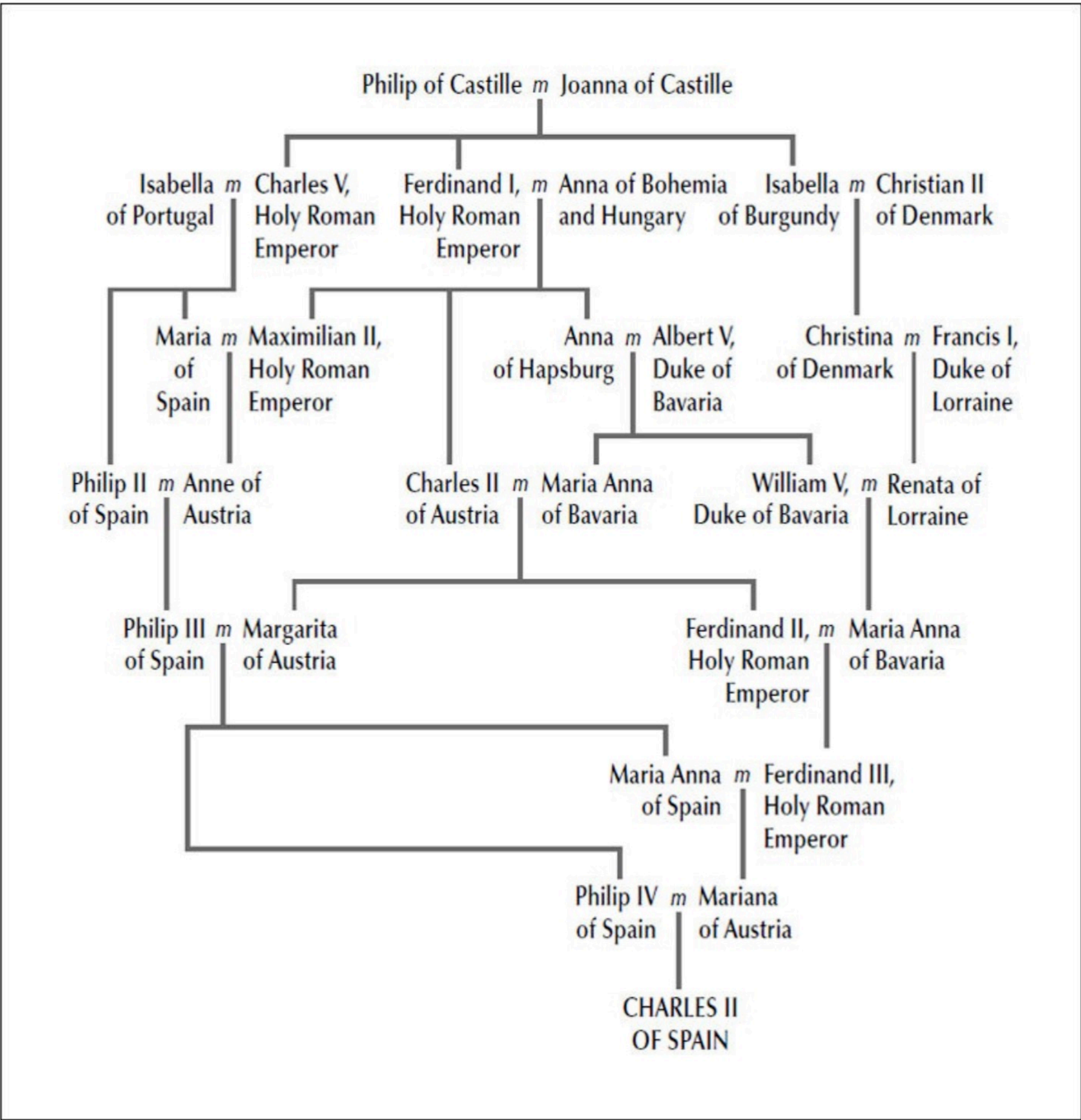


Figure 7-30 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.



# Skutki wsobności



**The family tree of Charles II of Spain.** The Hapsburg family stopped outbreeding more than a century before Charles II's birth in 1661. Six generations should contain 62 different ancestors; Charles II's has 32. Eight generations should contain 254 different ancestors; Charles II's has 82. This is what we might refer to as 'suboptimal'.



Filip IV Habsburg (1605-1655)  
Diego Velazquez



Marianna Habsburżanka (1634-1696)  
Diego Velazquez

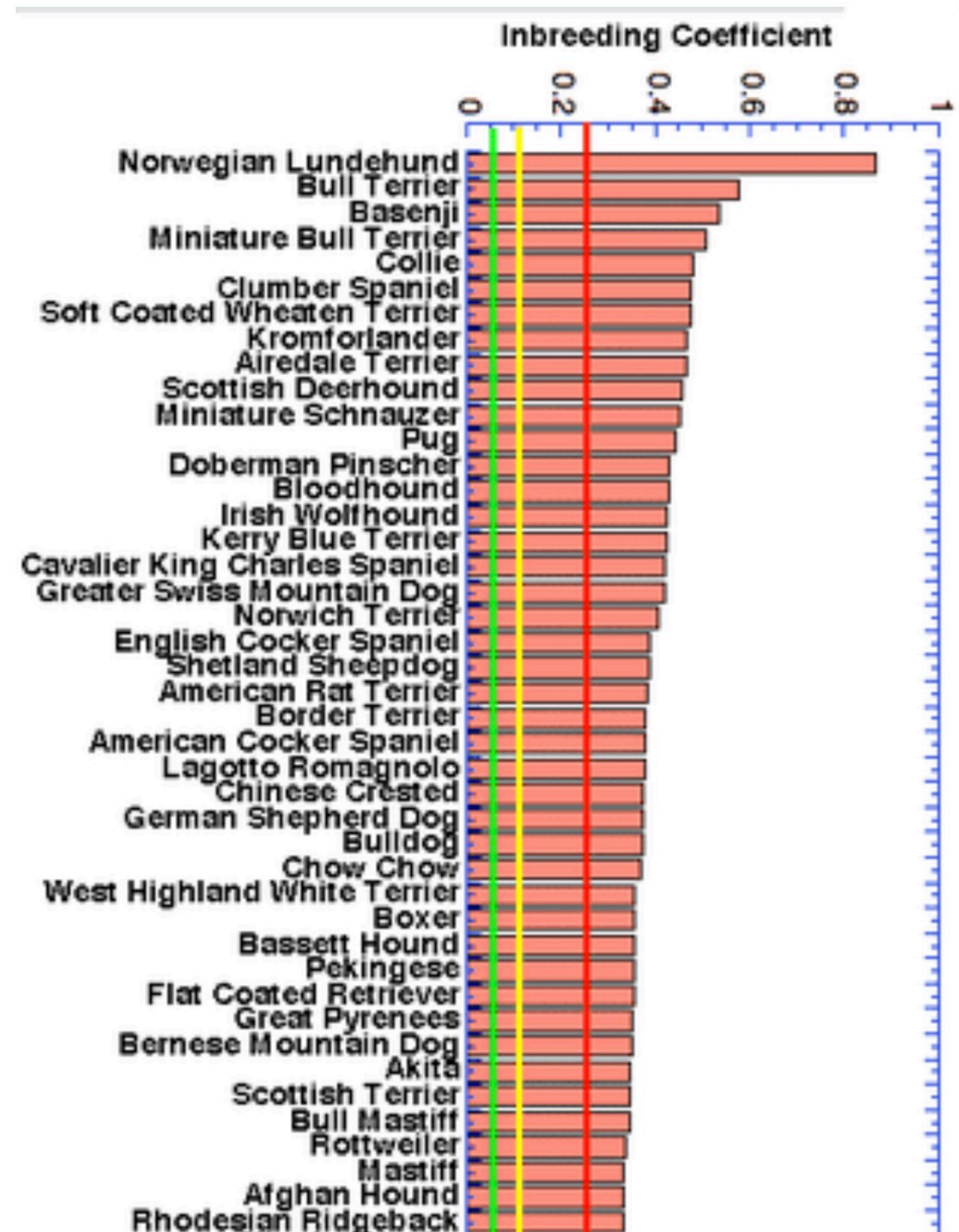


Karol II Habsburg (1661-1700)  
Juan Carreño de Miranda

$$F = 0,254$$



# Wsobność u rasowych zwierząt towarzyszących



Norsk Lundehund

(© wikipedia)

<http://www.instituteofcaninebiology.org/blog/inbreeding-of-purebred-dogs-determined-from-dna>

dane wg.

Whole-genome sequence, SNP chips and pedigree structure: building demographic profiles in domestic dog breeds to optimize genetic-trait mapping

Dayna L. Dreger, Maud Rimbault, Brian W. Davis, Adrienne Bhatnagar, Heidi G. Parker, Elaine A. Ostrander  
Disease Models & Mechanisms 2016 9: 1445-1460; doi: 10.1242/dmm.027037



# Mechanizmy zmieniające częstość alleli

- Mutacje
- Dobór
- Migracje
- Dryf

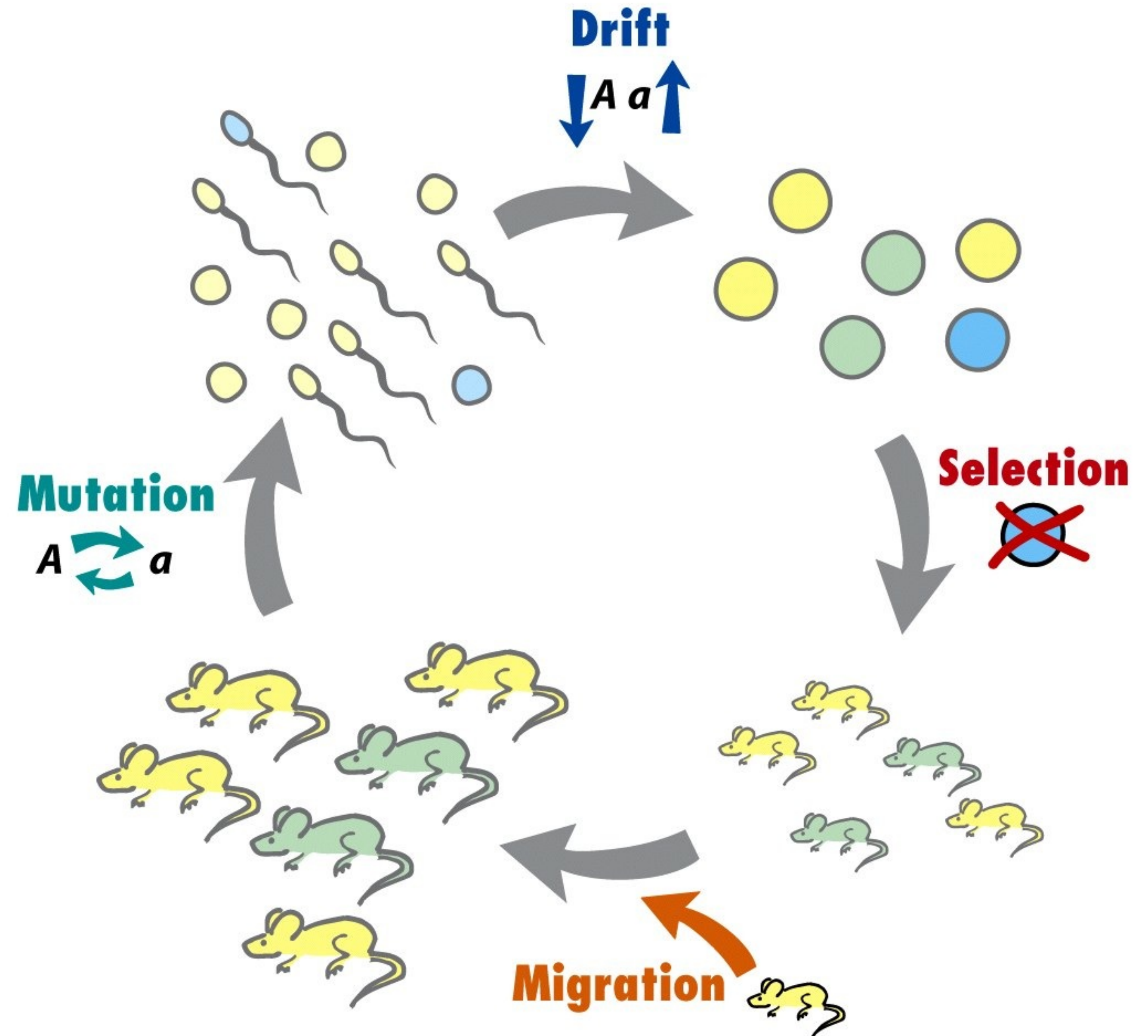


Figure 6-10 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.

# Dobór

---

Dostosowanie ( $w$ ) – prawdopodobieństwo odniesienia sukcesu reprodukcyjnego przez osobnika o danym genotypie

$$\mathbf{A1A1} : w_{11}$$

$$\mathbf{A1A2} : w_{12}$$

$$\mathbf{A2A2} : w_{22}$$

$$w = 1 - s$$

gdzie  $s$  to współczynnik selekcji

Nie ma znaczenia, czy chodzi o prawdopodobieństwo przeżycia, czy o liczbę wyprodukowanych gamet, czy o kondycję potomstwa itp.

“walka o byt” – uproszczona i niekiedy myląca metafora

# Silna i słaba selekcja - symulacje

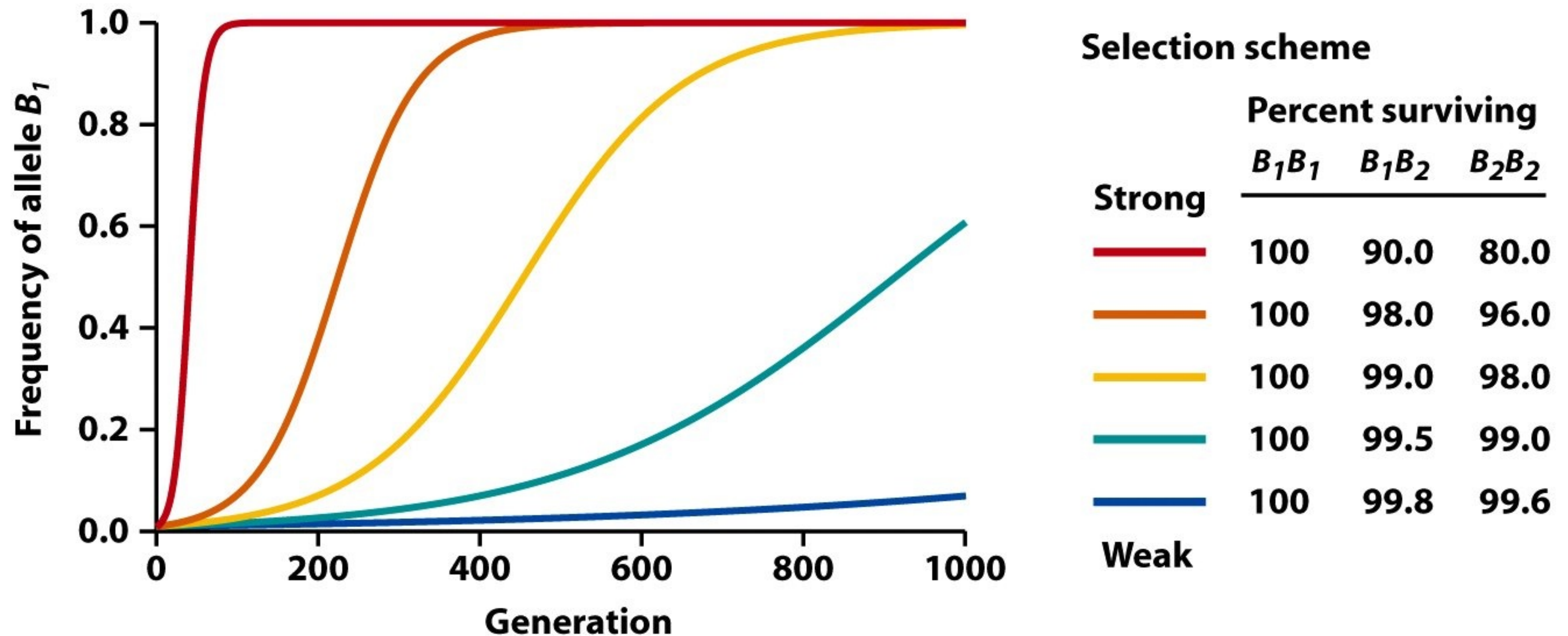


Figure 6-12 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.



# Mutacje i dobór

- Mutacje stają się istotną siłą w ewolucji gdy:
  - działa dobór naturalny
  - działa dryf genetyczny (populacje o skończonym  $N$ )

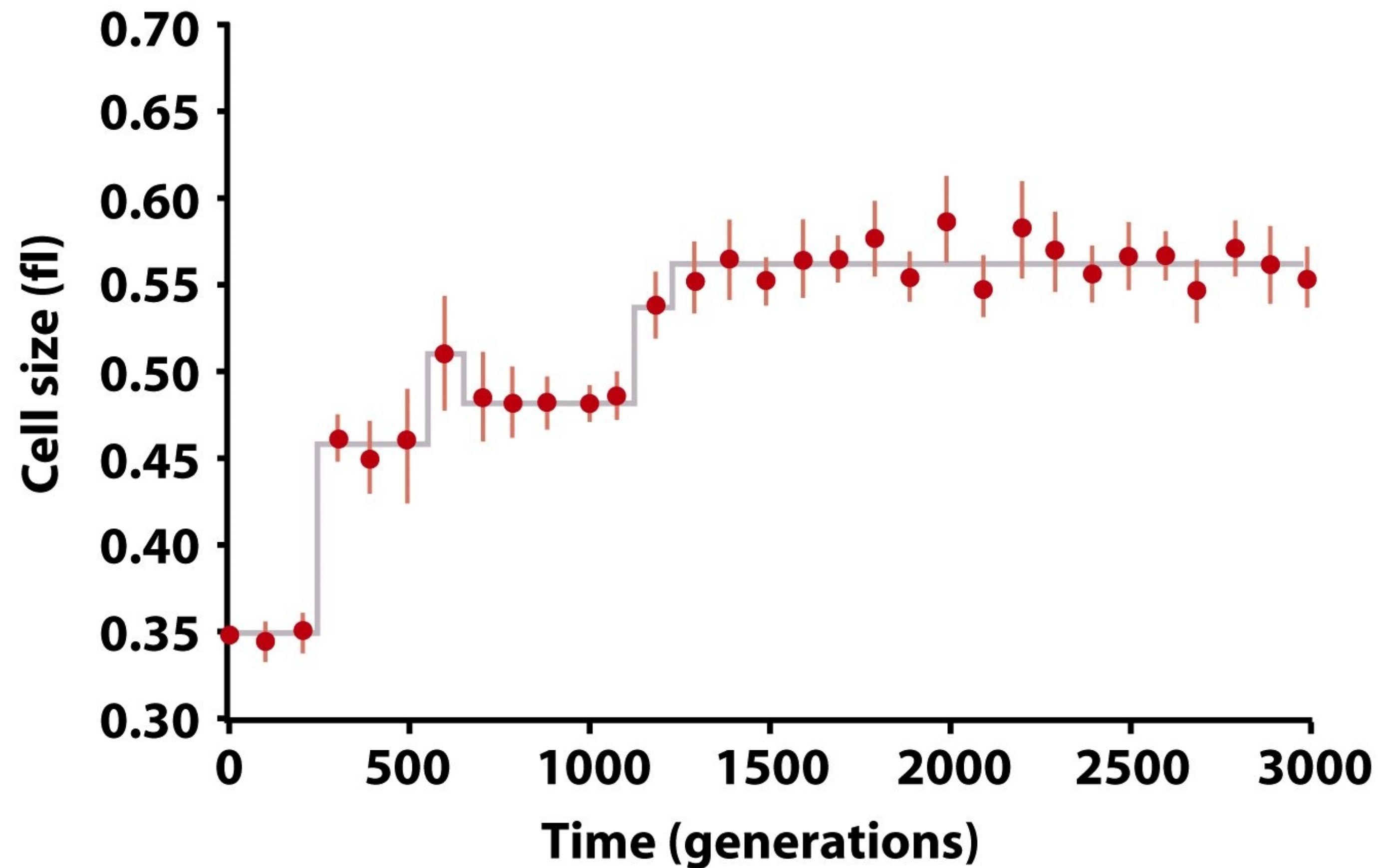
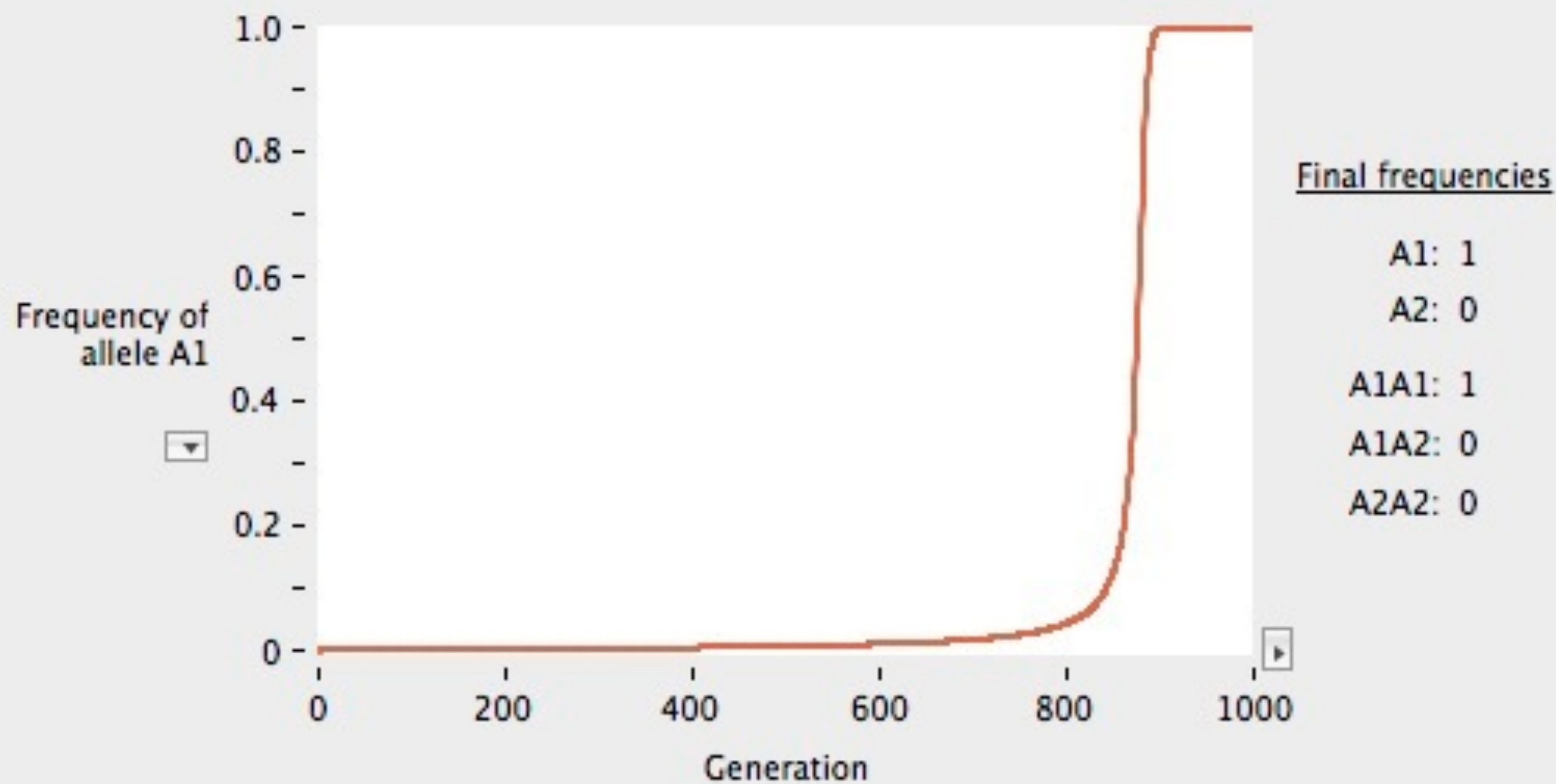


Figure 6-25 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.



Starting frequency of allele A1:

Fraction migrants each generation:

Fitness of genotype A1A1:

Frequency of A1 in source pop'n:

Fitness of genotype A1A2:

Population size:

Fitness of genotype A2A2:

Inbreeding coefficient (F):

Mutation rate from A1 to A2:

Mutation rate from A2 to A1:

Graph lines:

# Równowaga mutacje-selekcja

---

- Większość mutacji obniża dostosowanie, dobór je eliminuje
- Wytwarza się równowaga, utrzymująca w populacji pulę allelu o szkodliwym działaniu

- Dla allelu recesywnego  $\hat{q} = \sqrt{\frac{\mu}{s}}$

- Dla dominującego allelu letalnego  $\hat{q} = \mu$



# Dryf genetyczny a ewolucja

---

- Dobór naturalny nie jest jedynym mechanizmem kształtującym zmiany ewolucyjne
- Losowe procesy w populacjach o skończonej liczebności – dryf genetyczny



# Dryf genetyczny

---

- W populacjach o skończonej liczebności może dochodzić do zmian częstości alleli nawet jeżeli nie działa na nie dobór
- Nowy allel (mutacja) może się utrwalić w populacji nawet bez selekcji
  - częściowo (polimorfizm)
  - całkowicie



# Dryf a wielkość populacji

---

- Efekty dryfu genetycznego są wyraźniejsze w populacjach o mniejszej wielkości
- Z czasem dryf doprowadzi do utraty jednego z alleli i utrwalenia drugiego – utrata heterozygotyczności



# Utrata heterozygotyczności

- Przy braku działania doboru dryf doprowadzi do utraty jednego allelu i utrwalenia (fiksacji) drugiego
- Może powodować powstanie populacji odmiennych genetycznie, bez udziału doboru

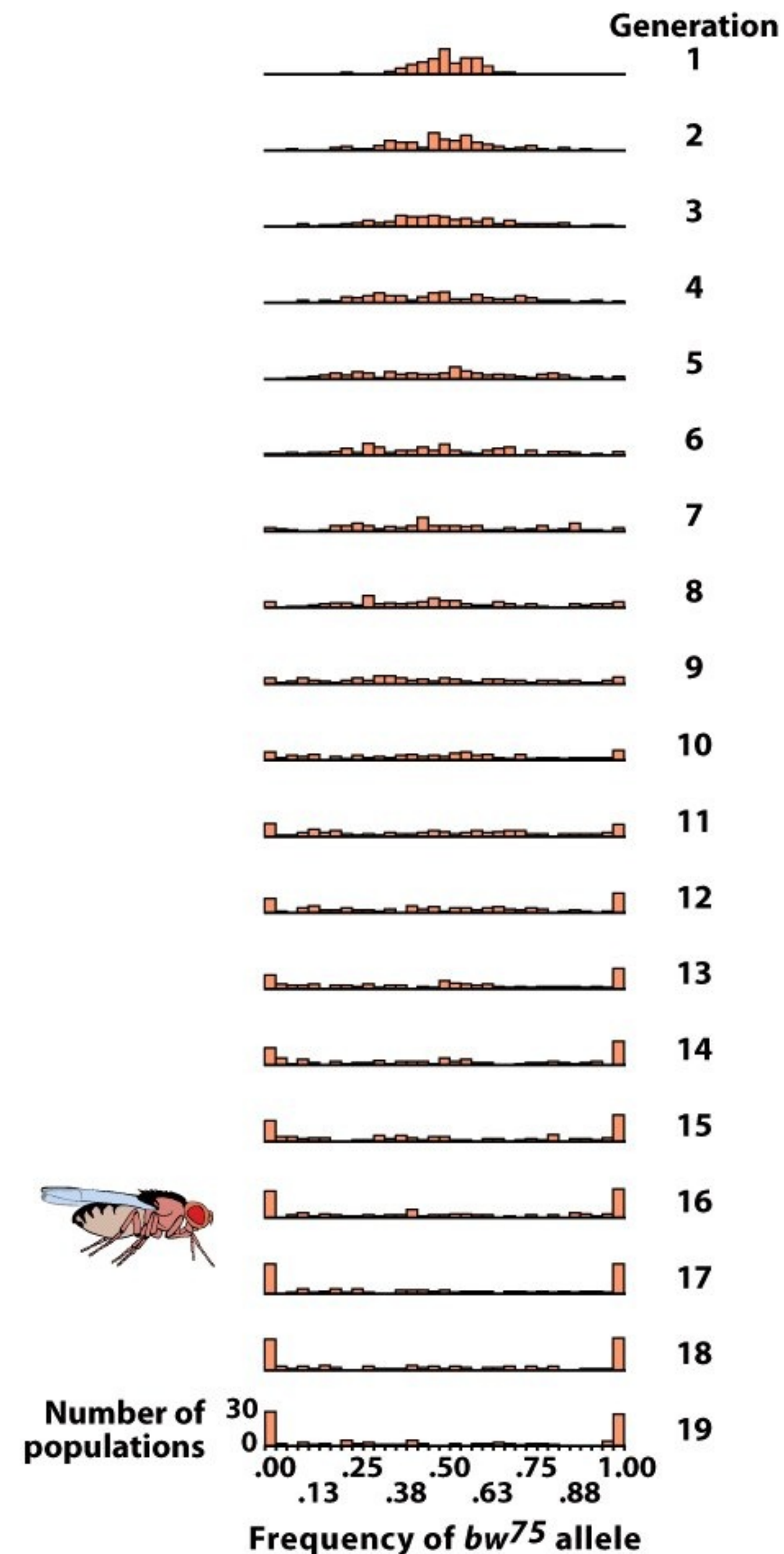


Figure 7-16 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.



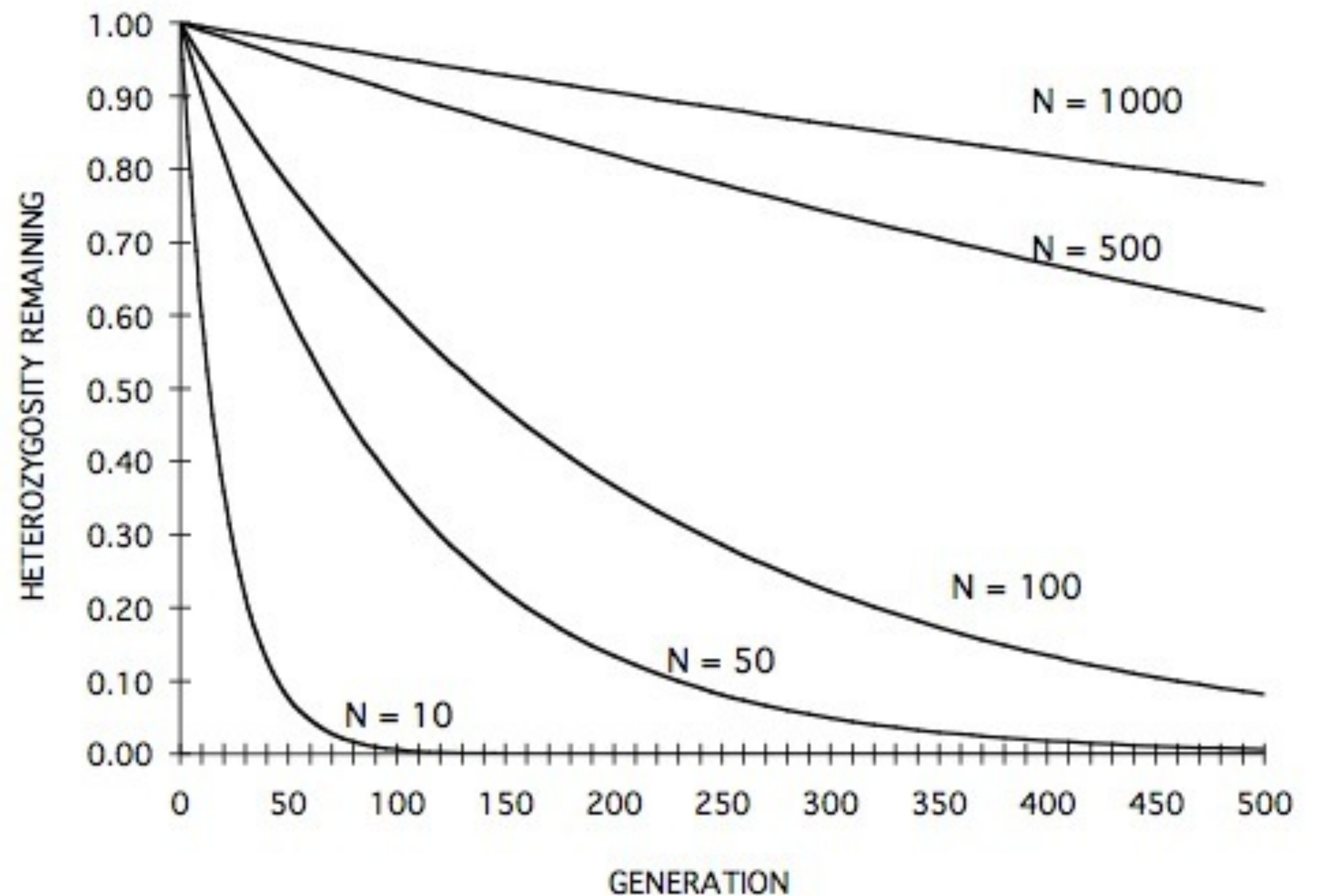
# Utrata heterozygotyczności

$$H_t = H_0 \left( 1 - \frac{1}{2N} \right)^t$$

czas półtrwania heterozygotyczności:

$$H_t = \frac{1}{2} H_0 \text{ przy } t = -2N \ln\left(\frac{1}{2}\right) \approx 1,39N$$

S. Wright, 1931



# Efektywna wielkość populacji

---

- We wszystkich modelach zakładaliśmy panmiksję – jednakowe prawdopodobieństwo wydania potomstwa przez każdego osobnika
- Rzeczywiste populacje nie spełniają tego warunku
  - nierównomierne stosunki płci (haremy)
  - zróżnicowanie sukcesu reprodukcyjnego

# Efektywna wielkość populacji

---

- Efektywna wielkość populacji  $N_e$  jest to liczebność idealnej populacji panmiktycznej, w której tempo dryfu byłoby takie same, jak w badanej populacji o rzeczywistej liczebności  $N$
- We wszystkich dotychczasowych rozważaniach podając  $N$  tak naprawdę mieliśmy na myśli  $N_e$



# Utrwalenie allelu

---

Prawdopodobieństwo utrwalenia konkretnego allelu

- W populacji  $N$  osobników diploidalnych jest  $2N$  alleli
- Utrwalenie oznacza, że wszystkie allele obecne w populacji pochodzą od jednego
- Prawdopodobieństwo tego jest  $1/2N$
- Jeżeli częstość allelu jest  $p$ , to wyjściowo jest  $2Np$  kopii
- Czyli prawdopodobieństwo utrwalenia wynosi:  
 $2Np \times 1/2N = p$



# Dryf i mutacje

- Mutacja powoduje powstanie nowego allelu
- Przy założeniu braku doboru (neutralność)
- Prawdopodobieństwo, że nowy allel się utrwali wynosi  $1/2N$
- Utrwalanie się kolejnych mutacji powoduje ewolucję populacji – ewolucja neutralna

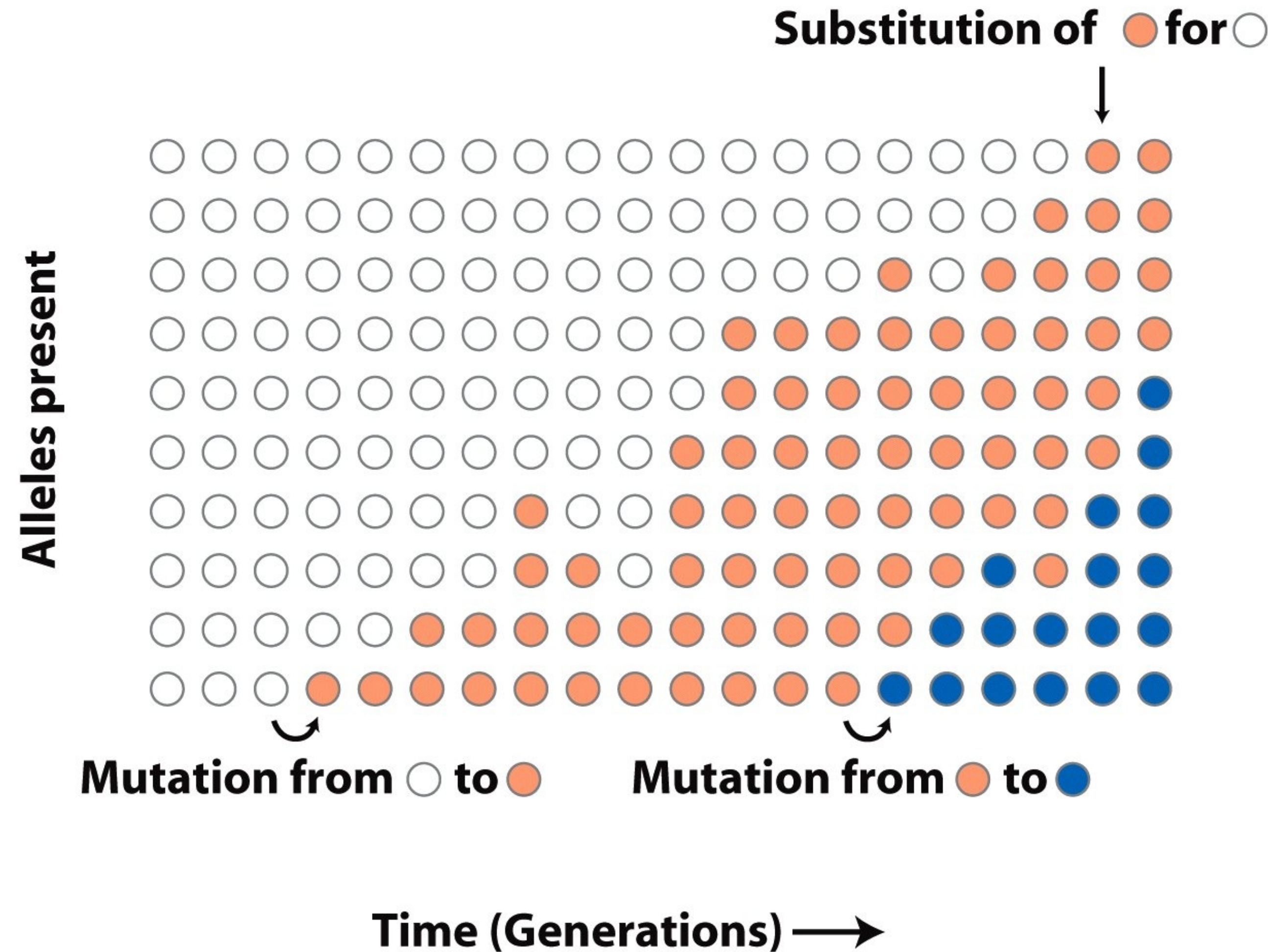


Figure 7-20 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.

# Tempo ewolucji neutralnej

---

Prawdopodobieństwo utrwalenia mutacji neutralnej:  $\frac{1}{2N}$

Prawdopodobieństwo powstania zmutowanego allelu:  $2N\mu$

( $\mu$  - tempo mutacji)

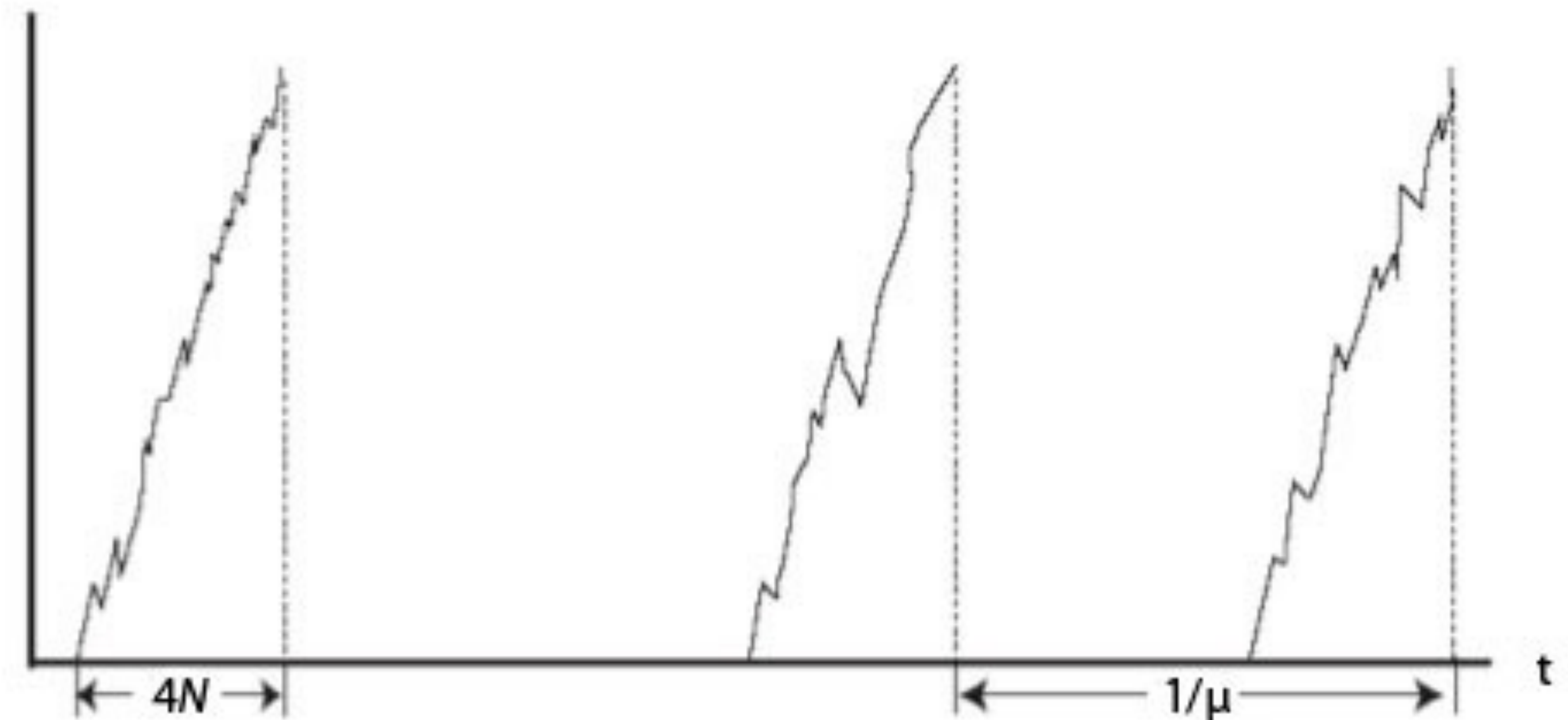
Prawdopodobieństwo powstania i utrwalenia się zmutowanego allelu (tempo ewolucji neutralnej)

$$2N\mu \cdot \frac{1}{2N} = \mu$$

# Czas i częstość utrwalania alleli neutralnych

---

- Tempo ewolucji neutralnej odpowiada częstości mutacji
- Czas od powstania do utrwalenia mutacji średnio  $4N$  ( $2N$  u haploidów)



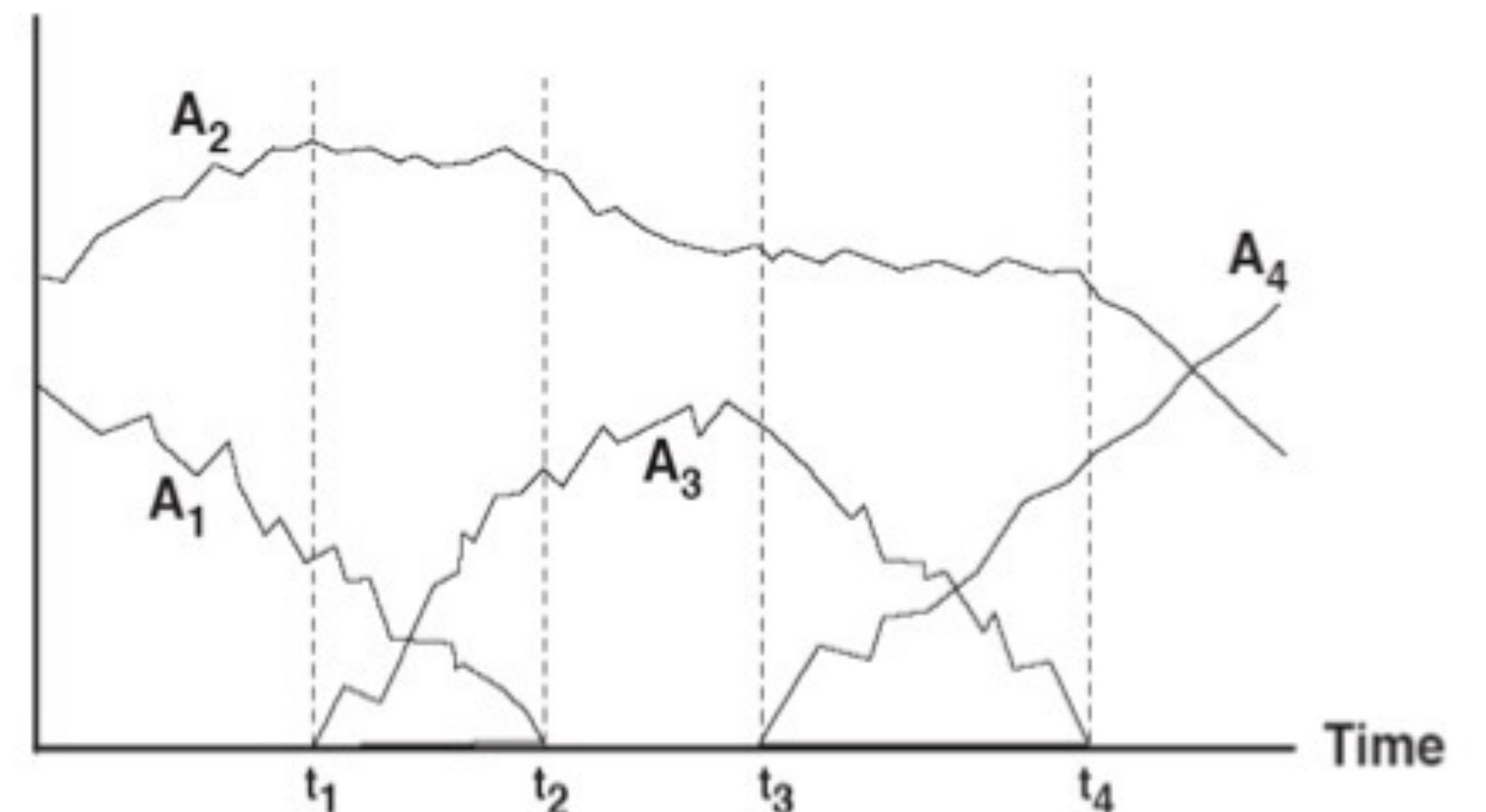


# Ewolucja neutralna

---

- Dryf zmniejsza różnorodność alleli (prowadzi do utrwalania jednego z alleli)
- Mutacje powodują powstawanie nowych alleli
- Dzięki temu różnorodność zostanie zachowana, ale skład konkretnych alleli się będzie zmieniał

Genotype frequency





# Dryf i dobór

---

- Dryf może doprowadzić do utraty allelu korzystnego, albo do utrwalenia allelu niekorzystnego
- Równowaga między dryfem a doborem zależy od wielkości populacji i siły (współczynnika) selekcji
- Prosty model (kodominaacja)

$A_1A_2$

$A_1A_2$

$A_2A_2$

$w \quad 1$

$1+s$

$1+2s$

# Dryf i dobór

---

Prosty model (kodominacja)

|         |       |        |
|---------|-------|--------|
| A1A2    | A1A2  | A2A2   |
| $w = 1$ | $1+s$ | $1+2s$ |

ON THE PROBABILITY OF FIXATION OF  
MUTANT GENES IN A POPULATION<sup>1</sup>

MOTOO KIMURA<sup>2</sup>  
*University of Wisconsin, Madison, Wisconsin*

Received January 29, 1962

*Genetics* **47** : 713–719 June 1962.

Model nie jest trywialny do wyprowadzenia (Kimura 1962)

Rezultat:

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}}$$

# Dryf i dobór

---

Gdy  $s \approx 0$  to  $P \approx q$  (prawdopodobieństwo utrwalenia allelu neutralnego jest równe jego częstości)

(Dla nowej mutacji  $q=1/2N_e$ )

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}}$$



## Dryf i dobór – allele korzystne

---

- Jeżeli  $s > 0$  i  $N$  jest duże to  $P \approx 2s$
- 98% mutacji o  $s = 0,01$  się nie utrwalą

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}}$$

(Dla nowej mutacji  $q=1/2N_e$ )



# Mutacja nieznacznie korzystna

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}}$$

(Dla nowej mutacji  $q=1/2N_e$ )

| S=      | 0,01 |
|---------|------|
| N       | P    |
| 10      | 6%   |
| 25      | 3%   |
| 50      | 2%   |
| 100     | 2%   |
| 1000    | 2%   |
| 10000   | 2%   |
| 100000  | 2%   |
| 1000000 | 2%   |

# Mutacja wyraźnie korzystna

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}}$$

(Dla nowej mutacji  $q=1/2N_e$ )

Efekt praktycznie nie zależy od  $N_e$

|         |     |
|---------|-----|
| s=      | 0,2 |
| N       | P   |
| 10      | 33% |
| 25      | 33% |
| 50      | 33% |
| 100     | 33% |
| 1000    | 33% |
| 10000   | 33% |
| 100000  | 33% |
| 1000000 | 33% |

# Mutacja nieznacznie niekorzystna

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}}$$

(Dla nowej mutacji  $q=1/2N_e$ )

|         |           |
|---------|-----------|
| s=      | -0,01     |
|         |           |
| N       | P         |
| 10      | 4%        |
| 25      | 1%        |
| 50      | 3,16E-03  |
| 100     | 3,77E-04  |
| 1000    | 8,58E-20  |
| 10000   | 3,87E-176 |
| 100000  |           |
| 1000000 |           |



Mutacja bardzo nieznacznie niekorzystna

$$P = \frac{1 - e^{-4N_e s q}}{1 - e^{-4N_e s}}$$

(Dla nowej mutacji  $q=1/2N_e$ )

| s=      |  | -0,001    |
|---------|--|-----------|
| N       |  | P         |
| 10      |  | 5%        |
| 25      |  | 2%        |
| 50      |  | 1%        |
| 100     |  | 4,07E-03  |
| 1000    |  | 3,74E-05  |
| 10000   |  | 8,51E-21  |
| 100000  |  | 3,83E-177 |
| 1000000 |  |           |



# Dlaczego mała populacja to zagrożenie

- Dla małych  $N_e$  bardzo wyraźnie spada skuteczność eliminacji alleli niekorzystnych (doboru oczyszczającego)
- Za to wzrost skuteczności utrwalania alleli korzystnych (doboru dodatniego) nie jest tak istotny
- Szczególnie krytyczne dla alleli w niewielkim stopniu niekorzystnych

| s=      | -0,01     |
|---------|-----------|
| N       | P         |
| 10      | 4%        |
| 25      | 1%        |
| 50      | 3,16E-03  |
| 100     | 3,77E-04  |
| 1000    | 8,58E-20  |
| 10000   | 3,87E-176 |
| 100000  |           |
| 1000000 |           |

| s=      | 0,01 |
|---------|------|
| N       | P    |
| 10      | 6%   |
| 25      | 3%   |
| 50      | 2%   |
| 100     | 2%   |
| 1000    | 2%   |
| 10000   | 2%   |
| 100000  | 2%   |
| 1000000 | 2%   |

| s=      | -0,001    |
|---------|-----------|
| N       | P         |
| 10      | 5%        |
| 25      | 2%        |
| 50      | 1%        |
| 100     | 4,07E-03  |
| 1000    | 3,74E-05  |
| 10000   | 8,51E-21  |
| 100000  | 3,83E-177 |
| 1000000 |           |

# Dryf i dobór

---

- Tempo utrwalania mutacji neutralnych

$$2N\mu \cdot \frac{1}{2N} = \mu$$

- Tempo utrwalania mutacji korzystnych

$$2N\mu \cdot 2s = 4Ns\mu$$

# Dryf i dobór - podsumowanie

---

- Większość mutacji (korzystnych, neutralnych i niekorzystnych) nie utrwalają się w populacji
- Gdy dobór przeciwko allelowi niekorzystnemu jest nieznaczny mutacja szkodliwa jest efektywnie neutralna – zostanie utrwalona z prawdopodobieństwem takim, jak neutralna
- Dobór jest nieznaczny gdy:

$$|s| \leq \frac{1}{4N_e}$$



# Dryf i dobór – równowaga

---

- Gdy  $N_e$  jest duże, mutacje szkodliwe są skutecznie usuwane
- Gdy  $N_e$  jest małe, dryf może prowadzić do akumulacji mutacji szkodliwych!
- Nawet gdy  $N_e$  jest duże, wiele mutacji korzystnych jest traconych, jeżeli  $s$  nie jest bardzo duże



# Druga synteza - ewolucja molekularna

---

- Ewolucja jako zmiany sekwencji DNA (i białek)
- Mechanizmy ewolucji molekularnej
  - źródła zmienności - mutacje i rearanżacje DNA
  - mechanizmy kształtujące zmienność - dobór i dryf
- Początki - lata 60. i 70. XX w. (sekwencjonowanie DNA - 1977, wcześniej sekwencje peptydowe)

# Mutacje

---

- Podstawienia (substytucje)
- Niewielkie delecje i insercje
  - niewielkie tzn. wpływające na sekwencję 1-2 genów

# Substytucje

- Tranzycje zachodzą w naturze częściej od transwersji
- mimo tego, że możliwych transwersji jest więcej
- stosunek ts/tv od ~2 (nDNA) do ~15 (mtDNA człowieka)
- wyjątek – mtDNA roślin
- duże różnice ts/tv u różnych grup organizmów

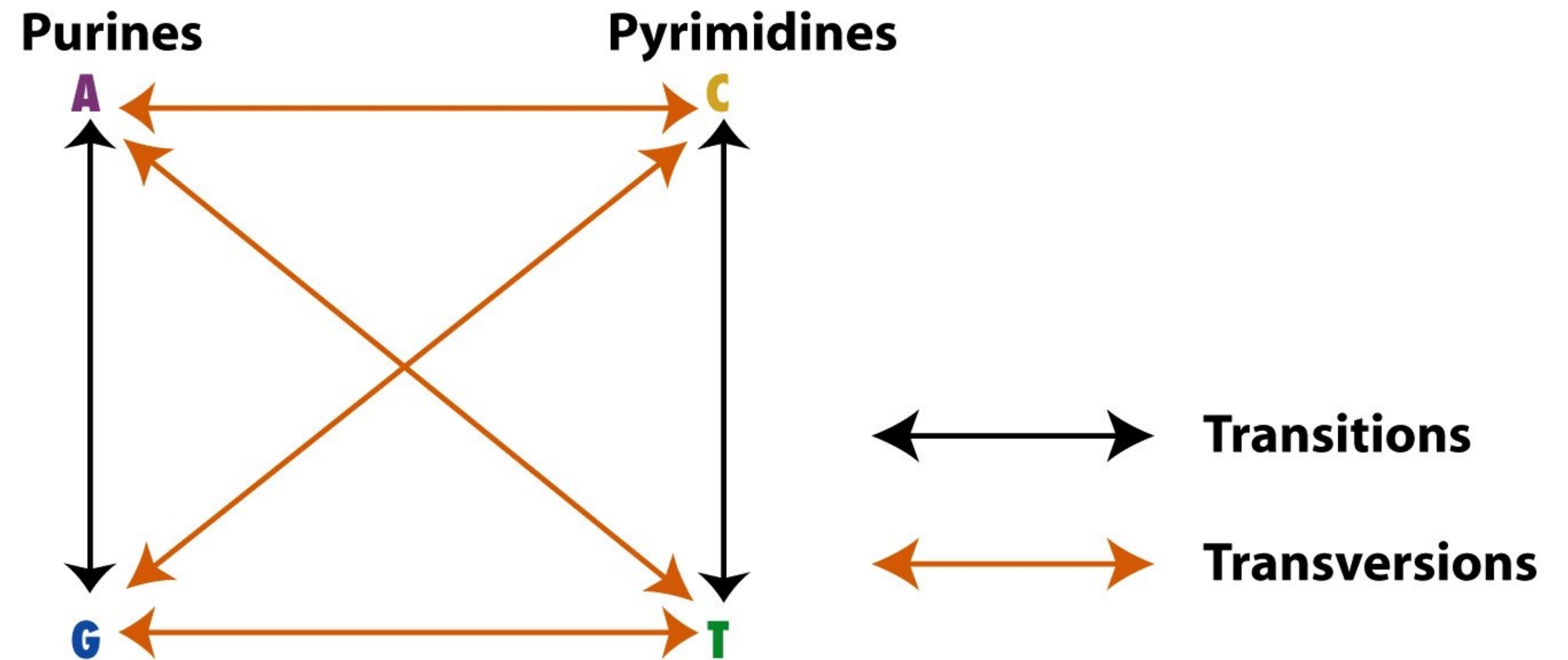


Figure 5-4 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.



# Tranzycje i transwersje

---

- Dlaczego tranzycje są częstsze?
  - Wyjaśnienia mechanistyczne – mechanizmy powstawania i naprawy mutacji
  - Tranzycje powstają w wyniku m. in.:
    - przejść tautomerycznych
    - deaminacji (np. oksydacyjnej)
  - Tranzycje w mniejszym stopniu zaburzają strukturę podwójnej helisy podczas replikacji
    - mniejsza wydajność naprawy przez system MMR



# Modele ewolucji sekwencji

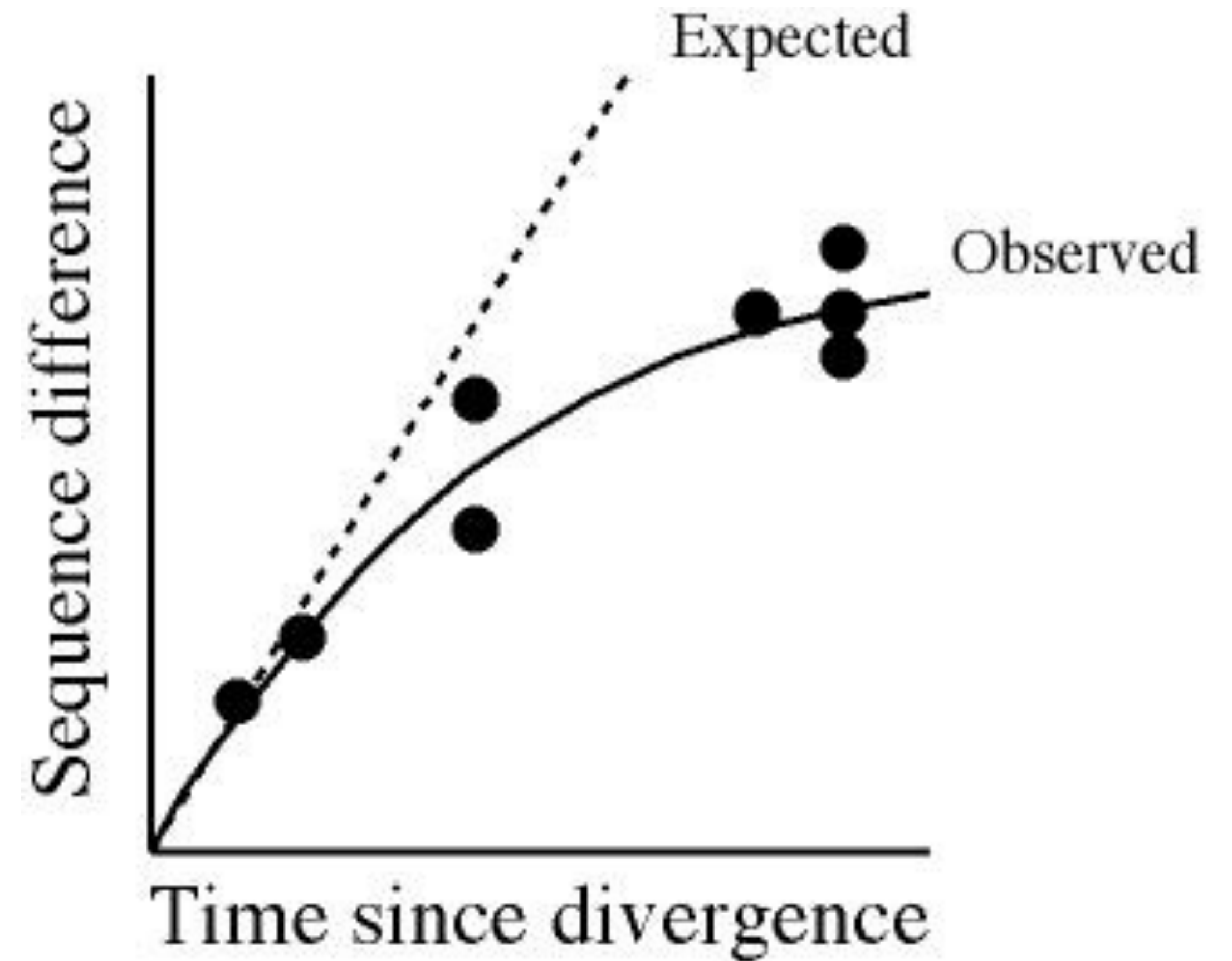
---

- Badając ewolucję nie dysponujemy z reguły sekwencją przodka
- Liczbę mutacji musimy oszacować na podstawie różnic między sekwencjami współczesnymi
- Konieczne jest uwzględnienie wielokrotnych mutacji w tej samej pozycji, zwłaszcza dla bardziej odległych sekwencji

# Problem obliczania odległości (liczby zdarzeń mutacyjnych)

---

ACGGTGC  
↓     ↓  
C     A  
↓     ↓  
GCGGTGA



# Modele ewolucji sekwencji

---

- Modele Markova – stan w pokoleniu  $n + 1$  zależy tylko od stanu w pokoleniu  $n$  i reguł przekształcenia (macierz prawdopodobieństw zmiany stanów)
- Modele o różnym stopniu skomplikowania
- Mogą uwzględniać:
  - mutacje wielokrotne w tej samej pozycji (poprawka Poissona)
  - różne prawdopodobieństwa zmian nukleotydowych (lub białkowych)
  - różne prawdopodobieństwo mutacji w różnych pozycjach sekwencji
  - różne częstości nukleotydów



# Modele ewolucji DNA – model Jukesa-Cantora

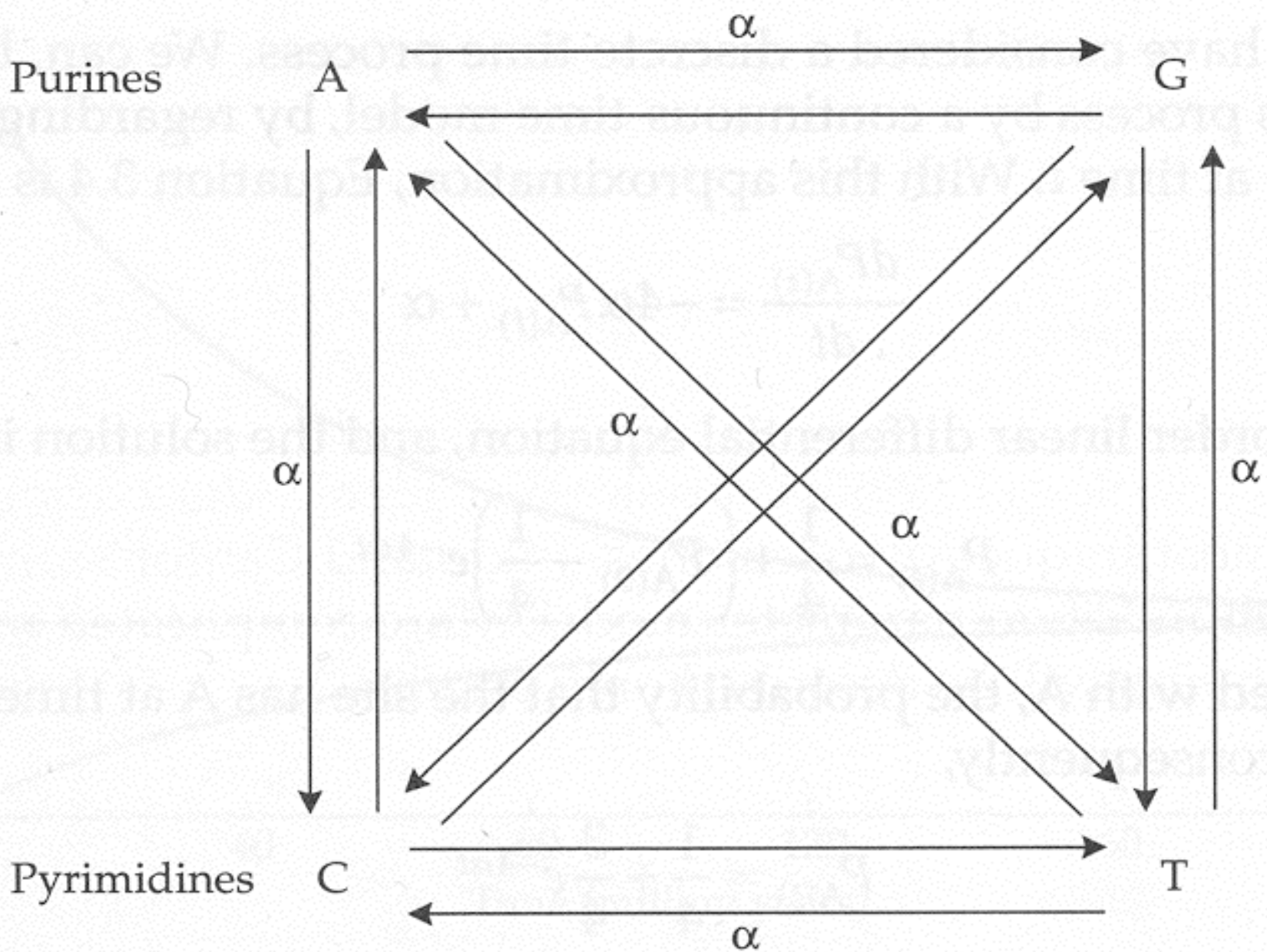


FIGURE 3.1 One-parameter model of nucleotide substitution. The rate of substitution in each direction is  $\alpha$ .

|   | A            | C            | G            | T            |
|---|--------------|--------------|--------------|--------------|
| A | 1-3 $\alpha$ | $\alpha$     | $\alpha$     | $\alpha$     |
| C | $\alpha$     | 1-3 $\alpha$ | $\alpha$     | $\alpha$     |
| G | $\alpha$     | $\alpha$     | 1-3 $\alpha$ | $\alpha$     |
| T | $\alpha$     | $\alpha$     | $\alpha$     | 1-3 $\alpha$ |

$$D_{JC} = -\frac{3}{4}\ln\left(1 - \frac{4}{3}D\right)$$



# Inne modele

---

- Kimura (K80, dwuparametrowy) - różne prawdopodobieństwo tranzycji i transwersji
- Felsenstein (F81), Hasegawa-Kishino-Yano (HKY85) - różne częstości nukleotydów (F81) + różne prawd. tranzycji i transwersji (HKY85)
- GTR (General Time Reversible, Tavaré '86)

# Ewolucja sekwencji aminokwasowych

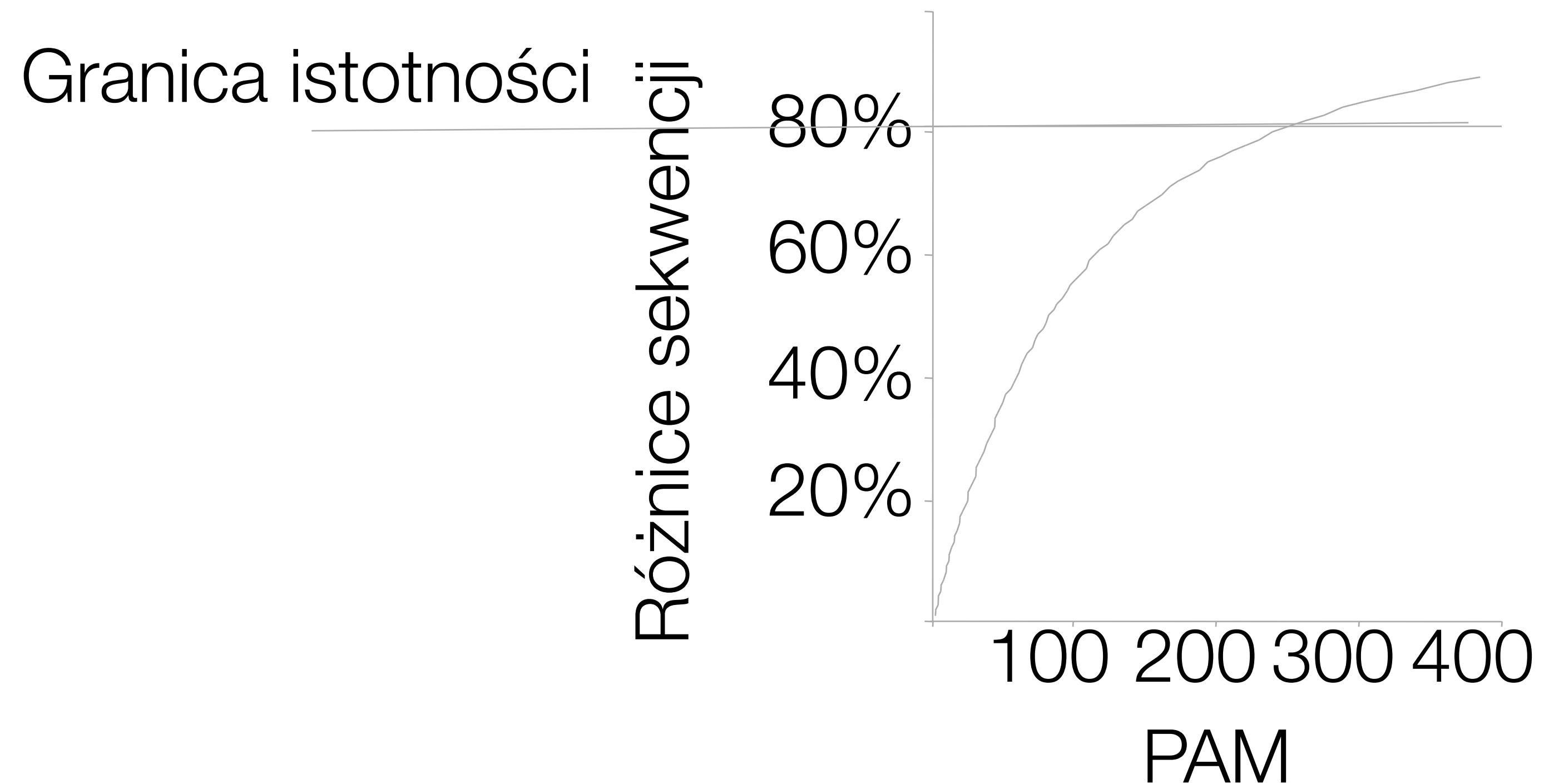
---

- Trudno stworzyć model analityczny
  - złożoność kodu
  - aminokwasy o różnych właściwościach - konieczna miara niepodobieństwa
- Stosuje się empirycznie uzyskiwane macierze prawdopodobieństwa zmiany danego aminokwasu w inny

# Tempo zmian sekwencji białka

---

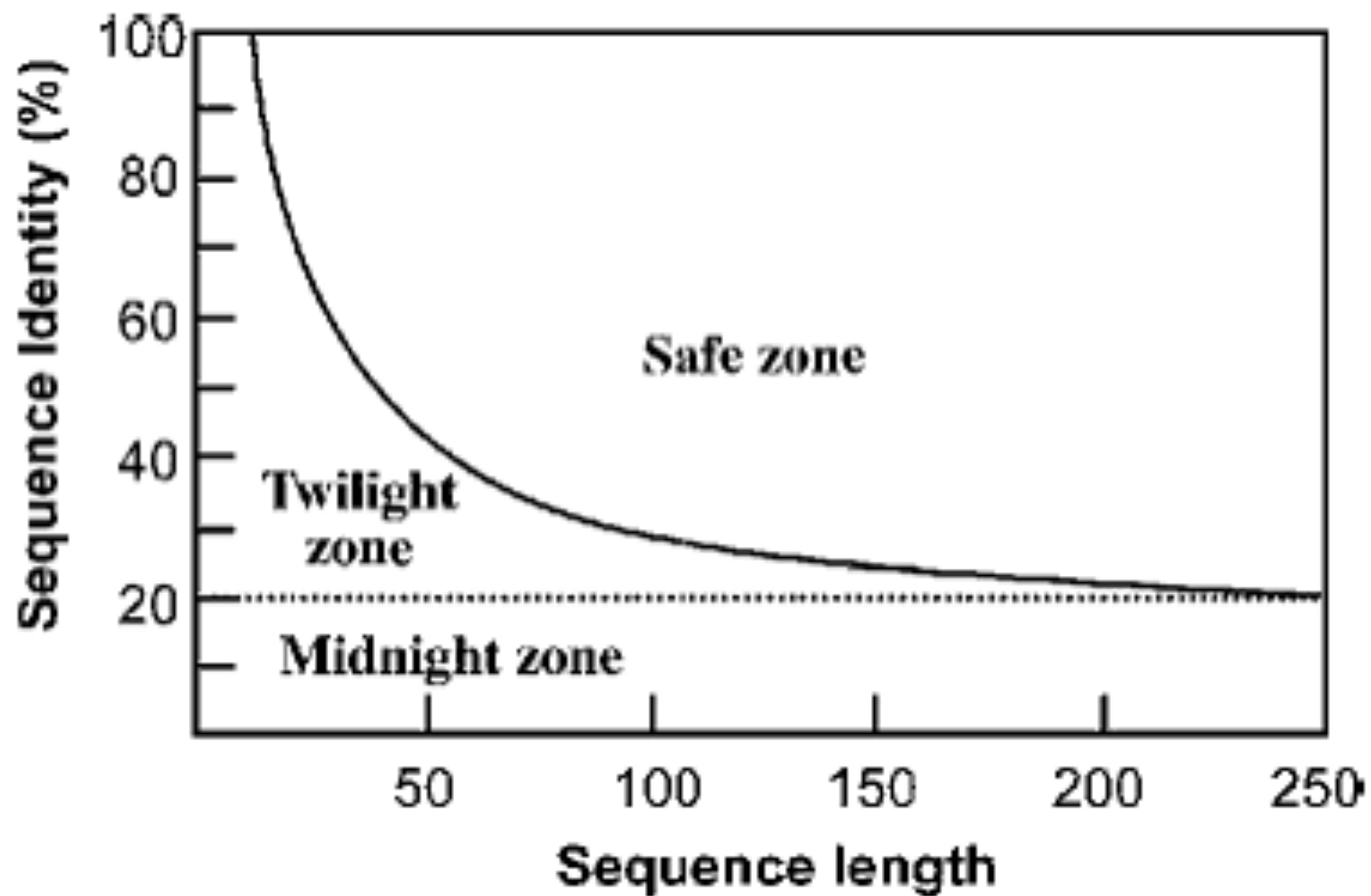
PAM - utrwalone mutacje punktowe/100 pozycji (od *Point Accepted Mutation*)





# Istotność podobieństwa

---





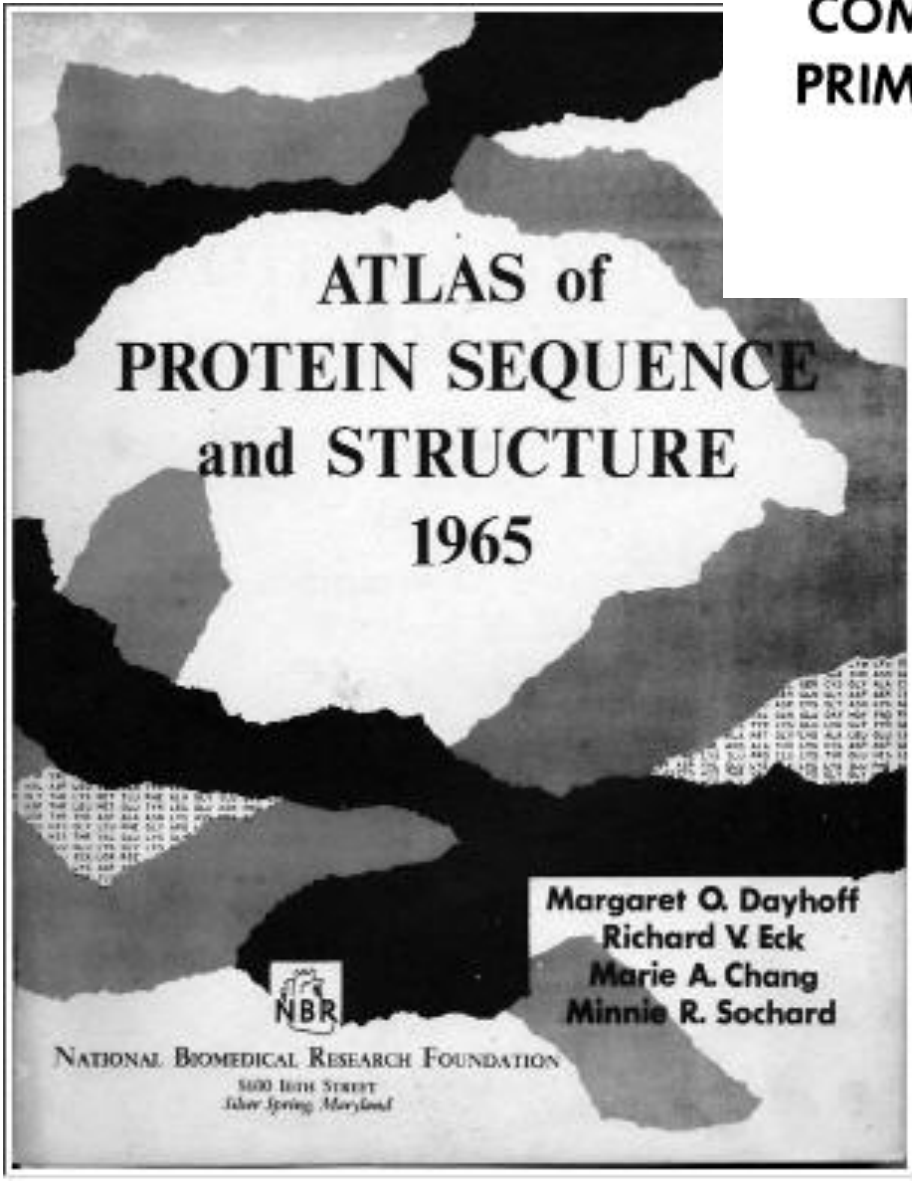
# Porównywanie białek - macierze

- Macierze Dayhoff (PAM)
  - Na podstawie globalnych porównań sekwencji różniących się o 1PAM ustalono prawdopodobieństwo zmiany każdego aminokwasu w inny → macierz PAM-1
  - Ekstrapolacja dla sekwencji bardziej odległych - mnożenie macierzy PAM-1 przez samą siebie odpowiednią liczbę razy → macierze PAM-20, PAM-40, PAM-250 itp. (proces Markova)

Margaret O. Dayhoff (1925-1983)

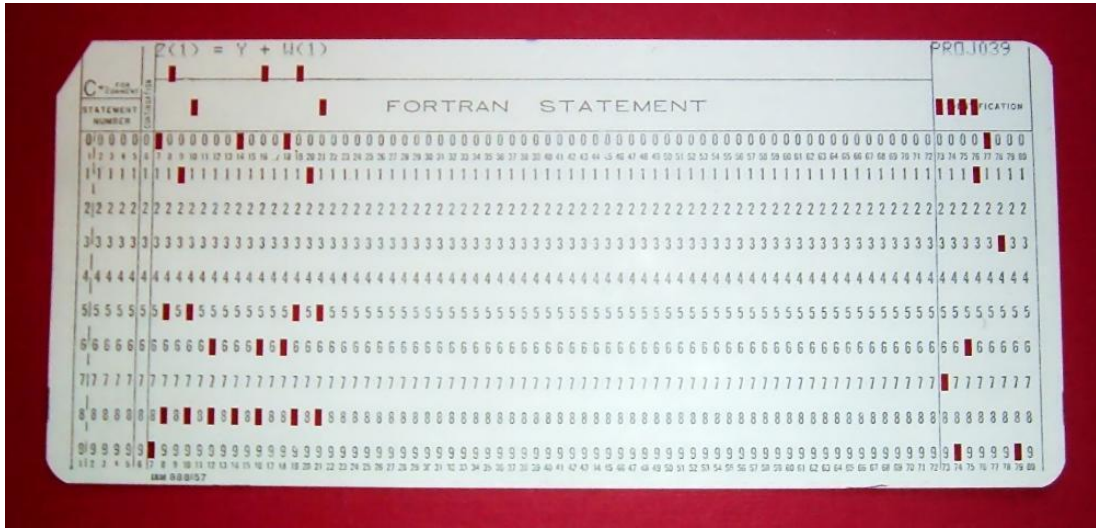


|   | C  | S  | T  | P  | A  | G  | N  | D  | E  | Q  | H  | R  | K  | M  | I  | L  | V  | F | Y  | W  |   |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|---|----|----|---|
| C | 12 |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |   |    |    | C |
| S | 0  | 2  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |    |   |    |    | S |
| T | -2 | 1  | 3  |    |    |    |    |    |    |    |    |    |    |    |    |    |    |   |    |    | T |
| P | -3 | 1  | 0  | 6  |    |    |    |    |    |    |    |    |    |    |    |    |    |   |    |    | P |
| A | -2 | 1  | 1  | 1  | 2  |    |    |    |    |    |    |    |    |    |    |    |    |   |    |    | A |
| G | -3 | 1  | 0  | -1 | 1  | 3  |    |    |    |    |    |    |    |    |    |    |    |   |    |    | G |
| N | -4 | 1  | 0  | -1 | 0  | 0  | 2  |    |    |    |    |    |    |    |    |    |    |   |    |    | N |
| D | -5 | 0  | 0  | -1 | 0  | 1  | 2  | 4  |    |    |    |    |    |    |    |    |    |   |    |    | D |
| E | -5 | 0  | 0  | -1 | 0  | 0  | 1  | 3  | 4  |    |    |    |    |    |    |    |    |   |    |    | E |
| Q | -5 | -1 | -1 | 0  | 0  | -1 | 1  | 2  | 2  | 4  |    |    |    |    |    |    |    |   |    |    | Q |
| H | -3 | -1 | -1 | 0  | -1 | -2 | 2  | 1  | 1  | 3  | 6  |    |    |    |    |    |    |   |    |    | H |
| R | -4 | 0  | -1 | 0  | -2 | -3 | 0  | -1 | -1 | 1  | 2  | 6  |    |    |    |    |    |   |    |    | R |
| K | -5 | 0  | 0  | -1 | -1 | -2 | 1  | 0  | 0  | 1  | 0  | 3  | 5  |    |    |    |    |   |    |    | K |
| M | -5 | -2 | -1 | -2 | -1 | -3 | -2 | -3 | -2 | -1 | -2 | 0  | 0  | 6  |    |    |    |   |    |    | M |
| I | -2 | -1 | 0  | -2 | -1 | -3 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | 2  | 5  |    |    |   |    |    | I |
| L | -6 | -3 | -2 | -3 | -2 | -4 | -3 | -4 | -3 | -2 | -2 | -3 | -3 | 4  | 2  | 6  |    |   |    |    | L |
| V | -2 | -1 | 0  | -1 | 0  | -1 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | 2  | 4  | 2  | 4  |   |    |    | V |
| F | -4 | -3 | -3 | -5 | -4 | -5 | -4 | -6 | -5 | -5 | -2 | -4 | -5 | 0  | 1  | 2  | -1 | 9 |    |    | F |
| Y | 0  | -3 | -3 | -5 | -3 | -5 | -2 | -4 | -4 | -4 | 0  | -4 | -4 | -2 | -1 | -1 | -2 | 7 | 10 |    | Y |
| W | -8 | -2 | -5 | -6 | -6 | -7 | -4 | -7 | -7 | -5 | -3 | 2  | -3 | -4 | -5 | -2 | -6 | 0 | 0  | 17 | W |
|   | C  | S  | T  | P  | A  | G  | N  | D  | E  | Q  | H  | R  | K  | M  | I  | L  | V  | F | Y  | W  |   |



## COMPROTEIN: A COMPUTER PROGRAM TO AID PRIMARY PROTEIN STRUCTURE DETERMINATION\*

Margaret Oakley Dayhoff and Robert S. Ledley  
National Biomedical Research Foundation  
Silver Spring, Maryland





# Porównywanie białek - macierze

- Macierze BLOSUM
  - Na podstawie prawdopodobieństwa zmiany każdego aminokwasu w inny w bloku lokalnego przyrównania sekwencji o n% identycznych aminokwasów (BLOSUM62 - 62% identycznych aa itp.)

|   | A  | C  | D  | E  | F  | G  | H  |  |
|---|----|----|----|----|----|----|----|--|
| A | 4  | 0  | -2 | -1 | -2 | 0  | -2 |  |
| C | 0  | 9  | -3 | -4 | -2 | -3 | -3 |  |
| D | -2 | -3 | 6  | 2  | -3 | -1 | -1 |  |
| E | -1 | -4 | 2  | 5  | -3 | -2 | 0  |  |
| F | -2 | -2 | -3 | -3 | 6  | -3 | -1 |  |
| G | 0  | -3 | -1 | -2 | -3 | 4  | 0  |  |
| H | -2 | -3 | -1 | 0  | -3 | 0  | 3  |  |

*BLOSUM 62*

# Mutacje i dobór naturalny

---

- Efekty działania mutacji obserwujemy pośrednio
  - różnice sekwencji między populacjami (gatunkami)
  - polimorfizm sekwencji w obrębie populacji
- Na allele wytworzone przez mutacje może działać dobór
- Za zmiany częstości powstających alleli może odpowiadać dryf genetyczny
- Obserwujemy mutacje utrwalone całkowicie lub częściowo (polimorfizmy) w puli genowej

# Podstawowe pytanie ewolucji molekularnej

---

- Jaka jest rola dryfu i doboru w wyjaśnieniu obserwowanego zróżnicowania sekwencji?
  - wewnątrpopulacyjnego (polimorfizmy)
  - międzygatunkowego
- Pytanie dotyczy zróżnicowania ilościowego!
  - Nikt nie podaje w wątpliwość tego, że adaptacje w ewolucji powstają dzięki działaniu doboru!



# Dobor czy dryf?

---

- **Selekjonizm**
  - większość utrwalonych mutacji została wyselekcjonowana przed dobór
  - większość polimorfizmów jest utrzymywana przez dobór
    - dobór równoważący, naddominacja, dobór zależny od częstości
- **Neutralizm** (Kimura, 1968)
  - większość utrwalonych mutacji została utrwalona przez dryf
  - za większość polimorfizmów odpowiada dryf
  - mutacje utrwalane przez dobór są rzadkie, nie mają wpływu na ilościową analizę zmienności molekularnej



# Mutacje i dobór

---

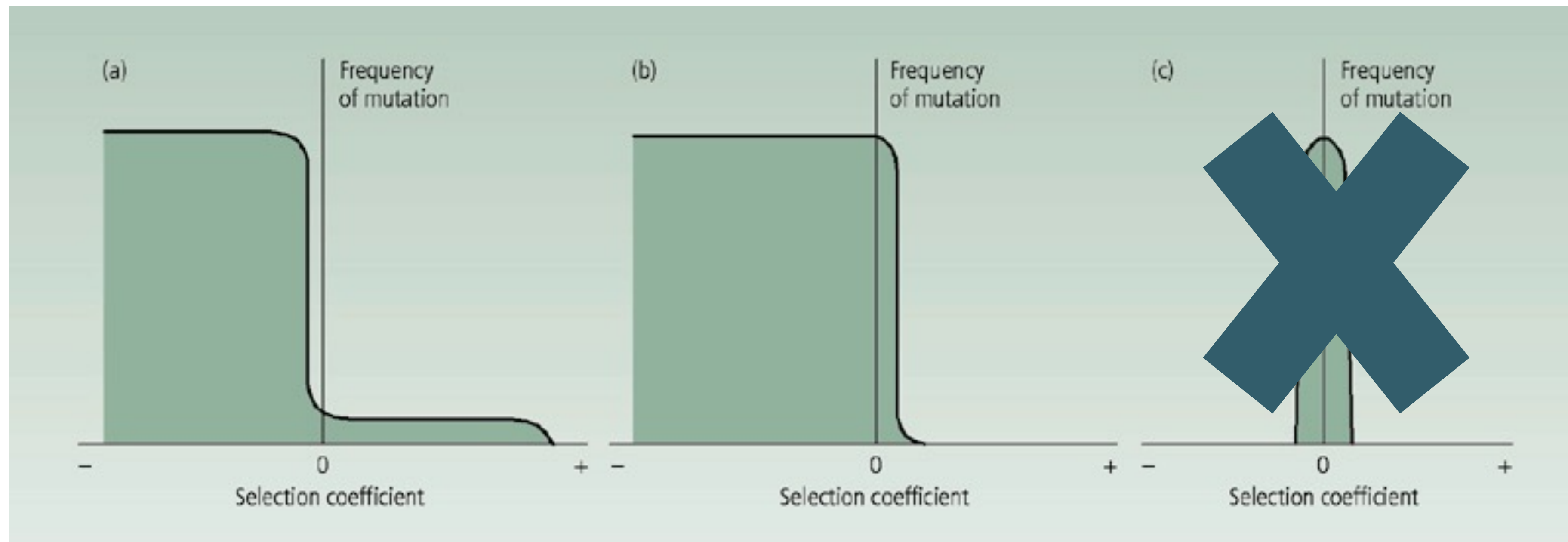
- niekorzystne (szkodliwe)
  - $s < 0$
  - eliminowane przez dobór (oczyszczający/negatywny)
- neutralne
  - $s \approx 0$  (a konkretniej,  $s \leq 1/4N$ )
  - utrwalane przez dryf
- korzystne
  - $s > 0$
  - utrwalane przez dobór (z udziałem dryfu dla niewielkich  $s$ )

# Selekcjonizm i neutralizm

---

- Selekcjonizm:
  - większość mutacji jest niekorzystna lub korzystna
  - większość utrwalonych mutacji jest korzystna
  - mutacje neutralne są rzadkie (nie częstsze od korzystnych)
  - **dobór** jest głównym mechanizmem kształtującym zmienność sekwencji
- Neutralizm
  - większość mutacji jest niekorzystna lub neutralna
  - większość utrwalonych mutacji jest neutralna
  - **dryf** jest głównym mechanizmem kształtującym zmienność sekwencji
  - mutacje korzystne są rzadkie (znacznie rzadsze od neutralnych)

# Selekcjonizm i neutralizm



selekcjonizm

neutralizm

pan-neutralizm

Neutralizm nie oznacza pan-neutralizmu, czyli negowania znaczenia selekcyjnego mutacji!



# Przesłanki teorii neutralnej

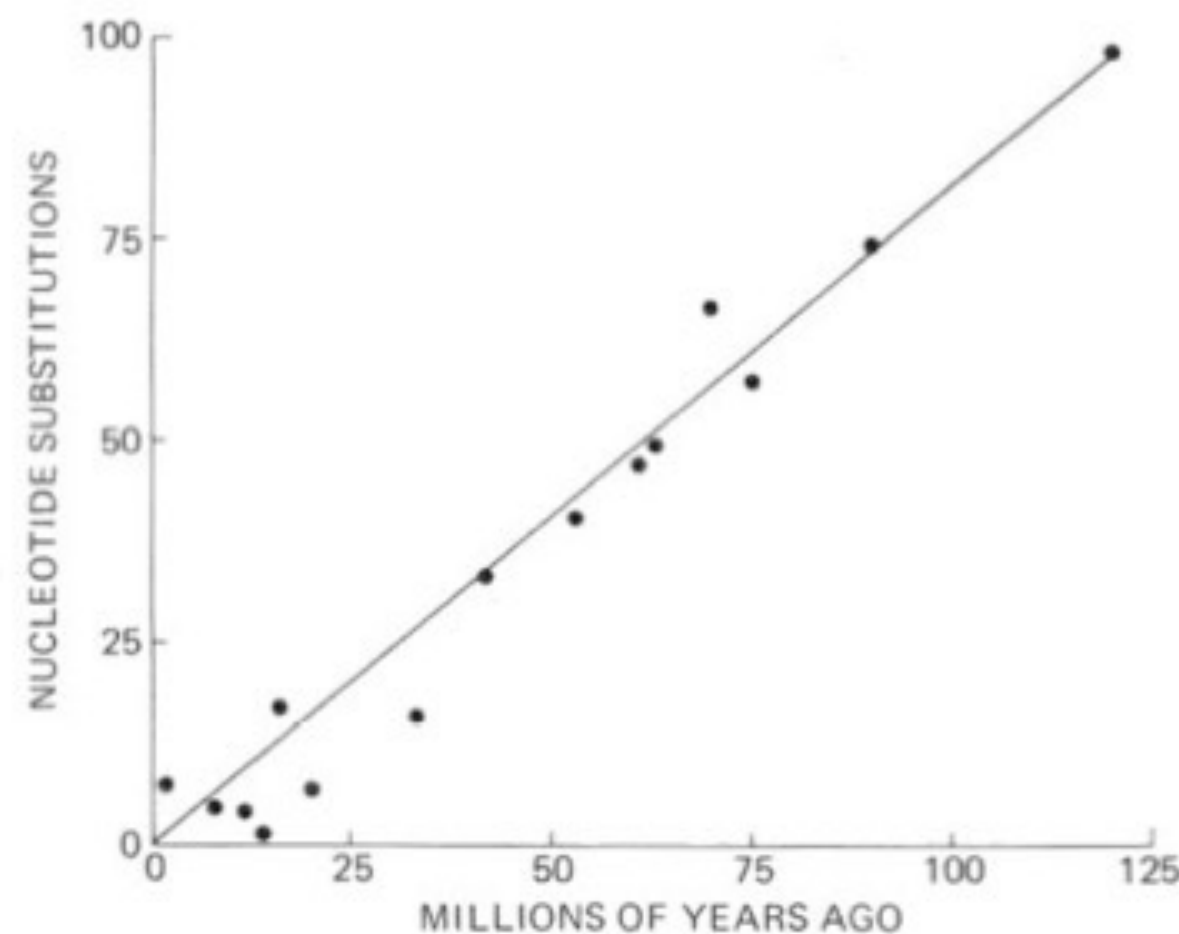
---

- Tempo zmian sekwencji i polimorfizm są zbyt duże, by dały się wyjaśnić doboorem
- Stałe tempo ewolucji molekularnej (zegar molekularny)
- Sekwencje o mniejszym znaczeniu funkcjonalnym (pseudogeny, mniej istotne obszary białek) ewoluują szybciej, niż obszary kluczowe dla funkcji



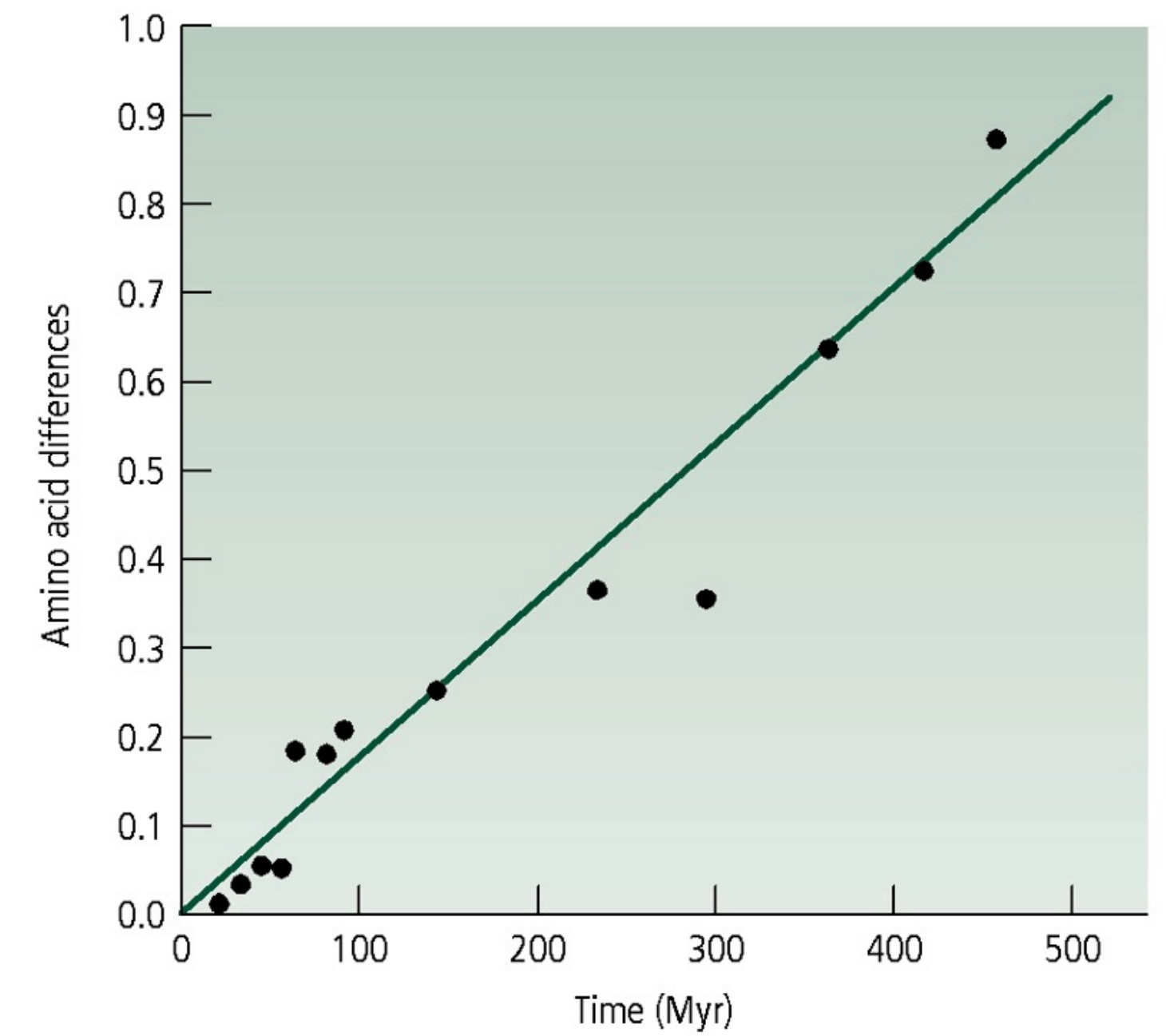
# Stałe tempo ewolucji molekularnej

- Wiele sekwencji ewoluuje w stałym tempie
- Tempo to jest różne dla różnych sekwencji, ale stałe w czasie ewolucji dla danej sekwencji



Pairwise nucleotide differences among 17 mammals from 7 proteins, plotted against date of divergence as estimated from fossil record

- Tzw. zegar molekularny



Różnice sekwencji globin kręgowców

# Tempo ewolucji i dryf

---

- Neutralny dryf jest procesem losowym, ale jego tempo będzie stałe w odpowiednio długim czasie
- Zależy tylko od częstości mutacji (jedna zmiana na  $1/\mu$  pokoleń)

$$2N\mu \cdot \frac{1}{2N} = \mu$$

- Dla doboru stałe tempo zmian oznacza stałe tempo zmian środowiska
- Tempo zmian adaptacyjnych nie wydaje się być stałe

# Zegar molekularny - problem

---

- W modelu neutralnym tempo utrwalania mutacji:

$$2N\mu \frac{1}{2N} = \mu$$

- Powinno być stałe w przeliczeniu na pokolenie
- Czas generacji jest różny u różnych organizmów
- Czyli nie powinna być obserwowana stałość tempa w czasie rzeczywistym
- A często jest (w tych sekwencjach, które zachowują zegar)



# Problem czasu generacji

---

- Czas generacji różnych organizmów jest istotnie różny
- Dlaczego nie wpływa to na tempo utrwalania mutacji?

~0,03 pokolenia/rok



~3 pokolenia/rok





# Zmiany prawie neutralne

---

- Model Kimury dotyczy zmian neutralnych ( $s = 0$ ), takie nie są (w sekwencji białek) częste
- Mutacje zachowują się jak neutralne gdy spełnione jest:

$$|s| \leq \frac{1}{4N_e}$$

- Mutacje o niewielkim współczynniku doboru  $s$  będą zachowywały się jak neutralne w małych populacjach, a w większych populacjach będą podlegały doborowi

# Zmiany prawie neutralne

---

- Istnieje odwrotna korelacja między czasem generacji a wielkością populacji

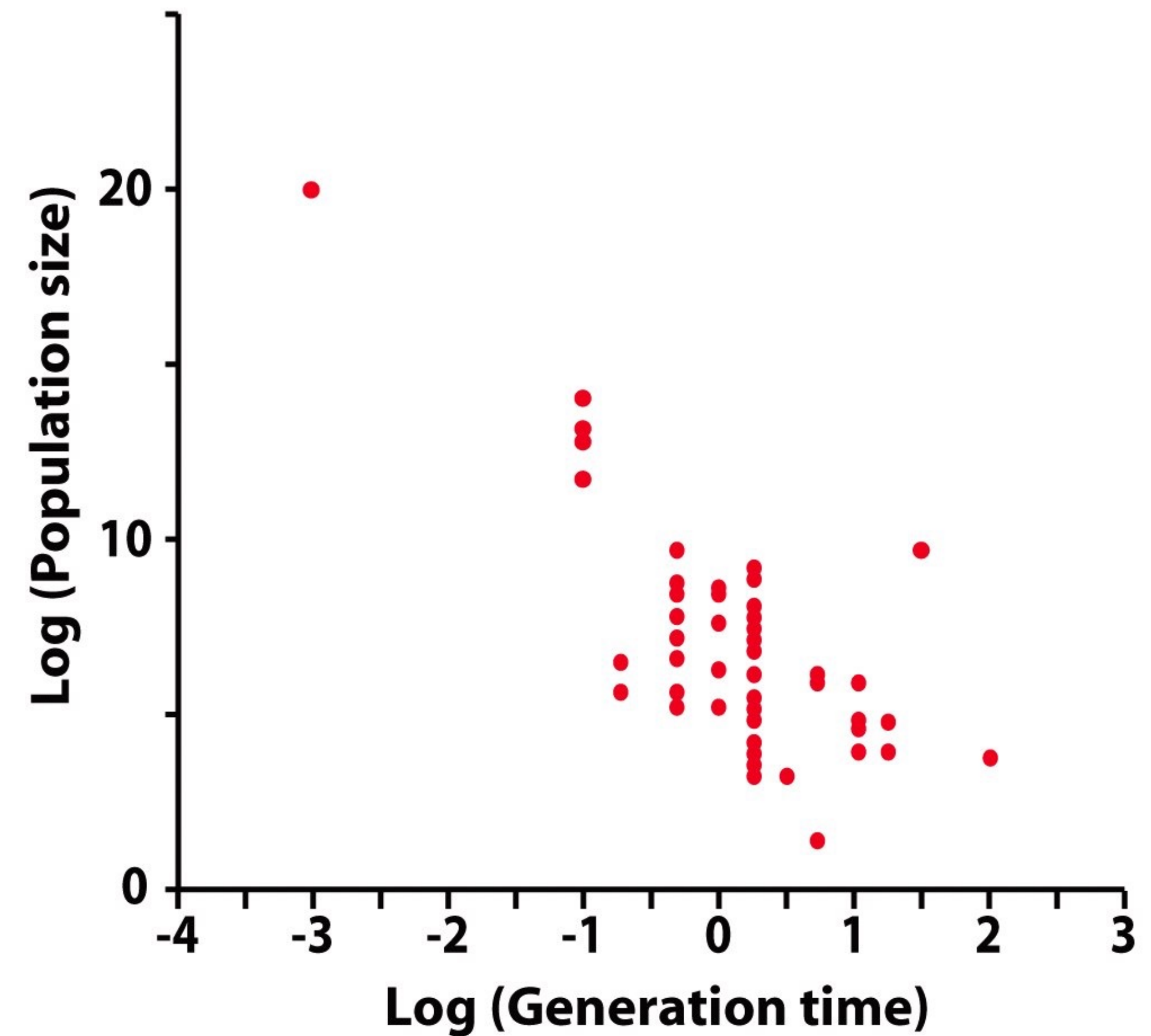


Figure 7-22a Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.

# Zmiany prawie neutralne

---

~0,03 pokolenia/rok



$$|s| \leq \frac{1}{4N_e}$$



~3 pokolenia/rok

Długi czas generacji

Krótki czas generacji

Mniej mutacji na rok

Więcej mutacji na rok

Populacja nieliczna (małe  $N_e$ )

Populacja liczna (duże  $N_e$ )

Więcej mutacji zachowuje się jak  
neutralne i utrwała przez dryf

Więcej mutacji podlega doborowi (i jest  
eliminowane przez dobór oczyszczający)

Efekty czasu generacji i wielkości populacji się znosząc, dając  
stałe tempo w czasie (Ohta & Kimura, 1971).



# Zegar molekularny

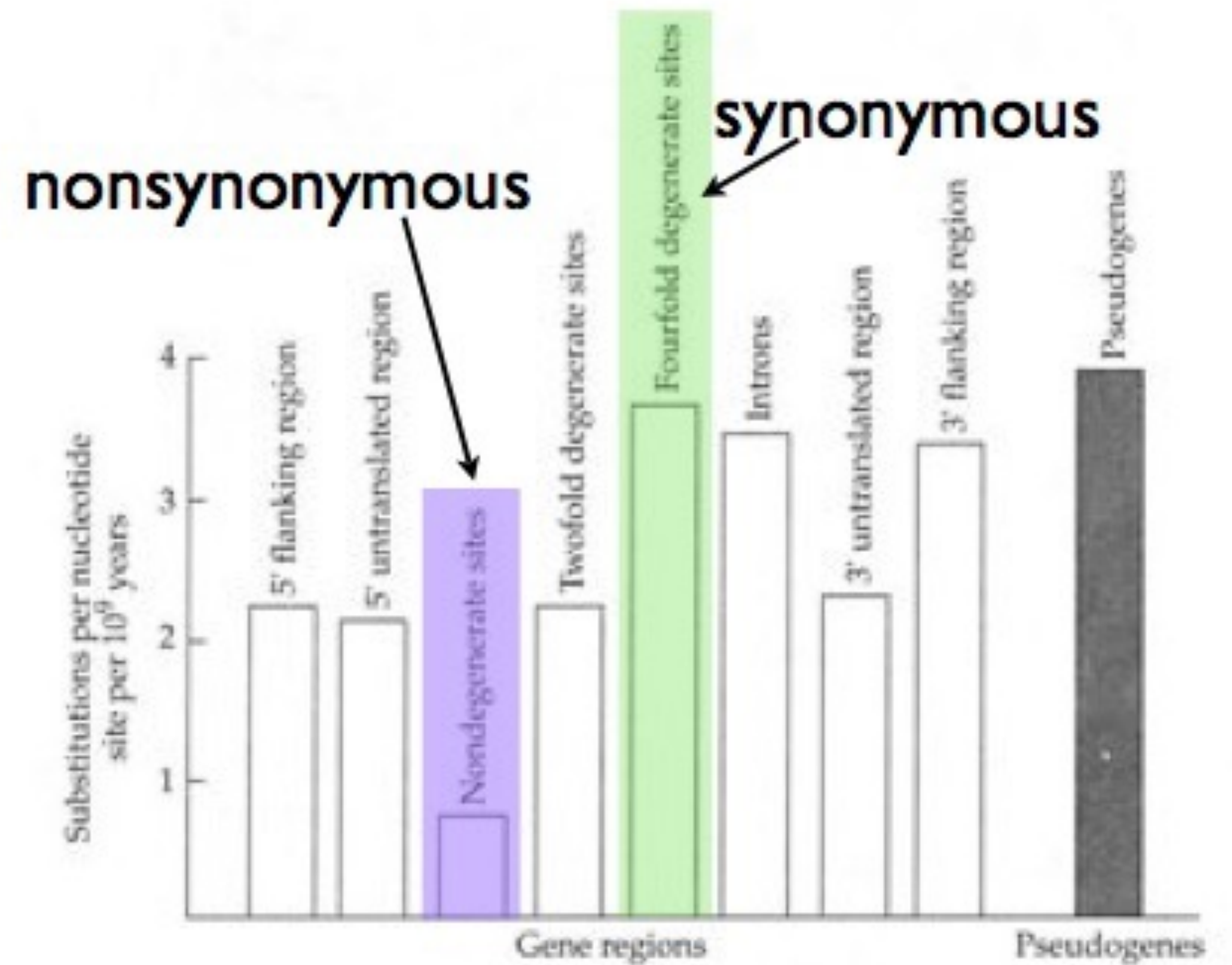
---

- Dla sekwencji białek i zmian niesynonimicznych w DNA zmiany jednostajne w czasie
- Na poziomie DNA,
  - dla mutacji synonimicznych
  - pseudogenów
  - niektórych sekwencji niekodujących
- tempo ewolucji zależy od czasu generacji



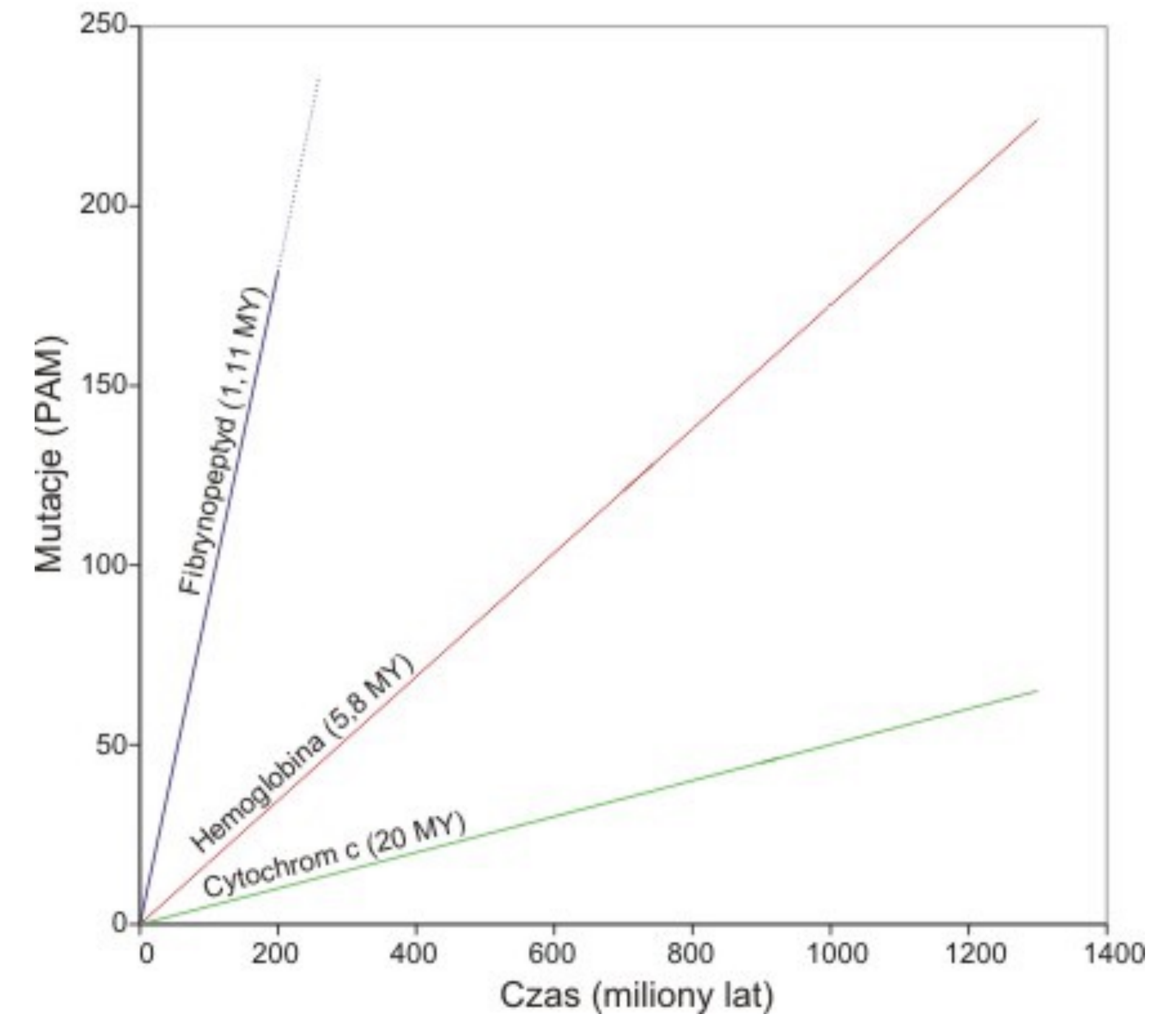
# Tempo ewolucji sekwencji a funkcja

- Głównym czynnikiem determinującym ilościową zmienność sekwencji jest dobór negatywny (oczyszczający)
- Sekwencje o mniejszym znaczeniu funkcjonalnym (pseudogeny, mniej istotne obszary białek) ewoluują szybciej, niż obszary kluczowe dla funkcji
- **Konserwacja sekwencji świadczy o jej funkcji!**



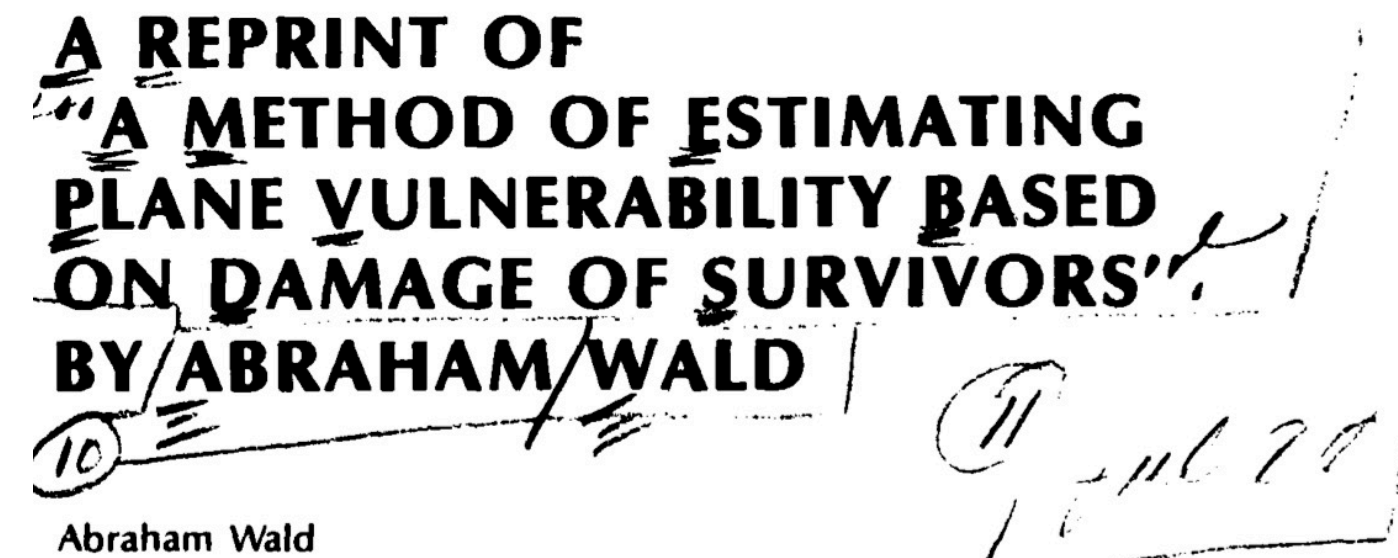
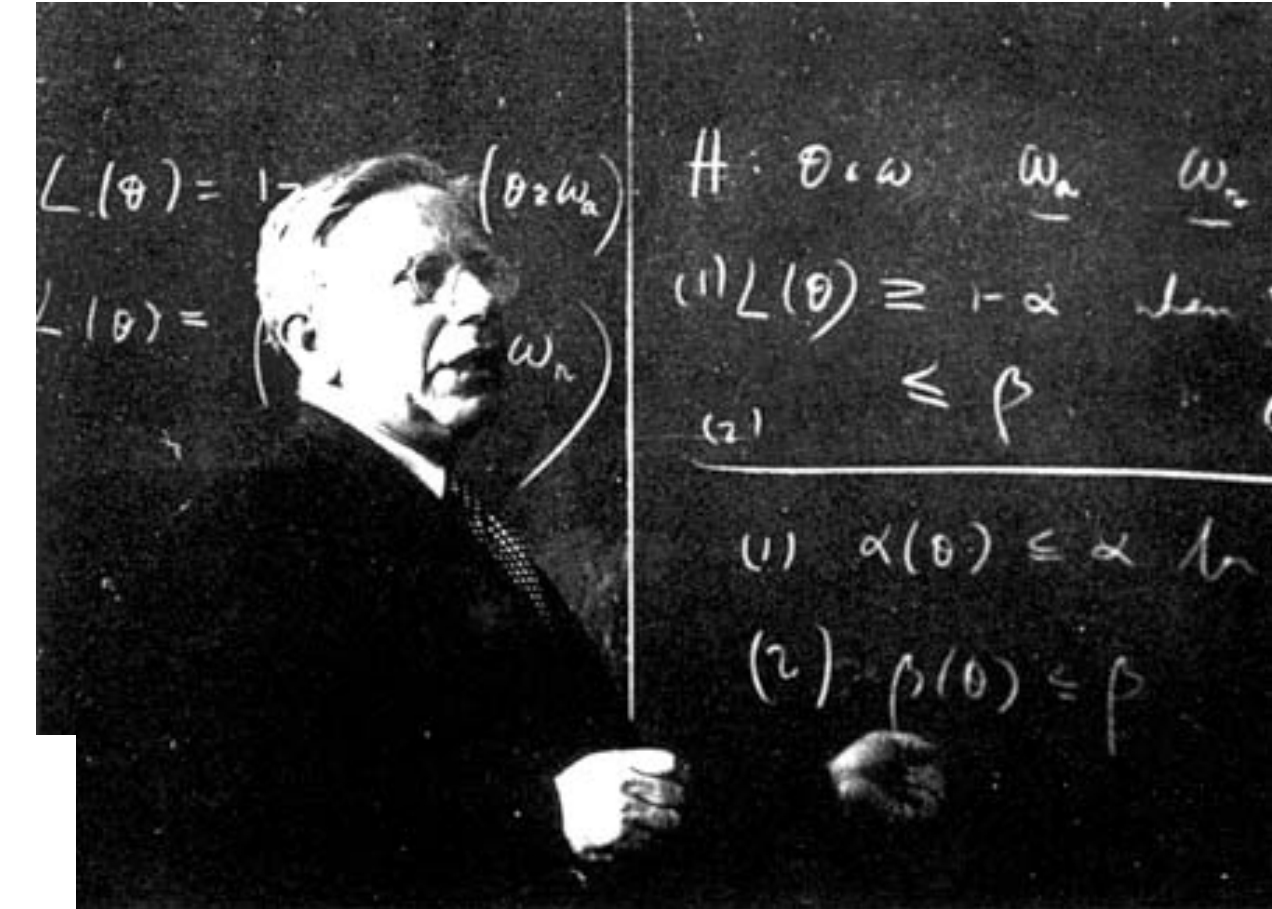
# Tempo zmian

- Białka zaangażowane w podstawowe funkcje komórki ewoluują wolniej.
- W sekwencji białka obszary kluczowe dla funkcji ewoluują wolniej.
  - Jednostka: PAM/ $10^8$  lat
  - Jednostka czasu ewolucyjnego: ile lat (w milionach,  $10^6$ ) potrzeba do utrwalenia 1 mutacji/100 aa (1 PAM)



# Abraham Wald

- Pionier badań operacyjnych (teoria decyzji)
- Prace dla *Center for Naval Analyses* podczas II w. ś.
- Analiza rozmieszczenia przestrzelin w uszkodzonych samolotach
  - oryginalny plan: wzmocnić pancerz w miejscach, gdzie obserwuje się najwięcej przestrzelin
  - analiza Walda: wzmocnić tam, gdzie nie obserwuje się przestrzelin (samoloty tam trafione nie wróciły)



<http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Wald.html>



 **Operations Evaluation Group**  
**CENTER FOR NAVAL ANALYSES**  
2000 North Beauregard Street, Alexandria, Virginia 22311

<http://oai.dtic.mil/>



# Zachowawczość sekwencji

- Obszary sekwencji białkowych najwolniej się zmieniające są zwykle kluczowe dla funkcji

H\_sapiens LFP--FFLRHAKQIFP-----VLDCKDDLRLKIS-----DLRIPPWNWYPDARAMQRKIIF  
 C\_elegans LYP--FFLDHAKQVFP-----HLEAMDDLRIIS-----DLTRPHNWNYPEARSVTRKIFF  
 S\_cerevisiae ITTLDKFIHVLQQTDFDHVYEQEI LPMMTNTDDTDGAHNV DITNPAEWFPEARKIRRHIIIM  
 A\_thaliana\_2 LFP--VFVEFCIEEFPD-----EIKRFSKIVDTA-----DLTKPATWEPFARAMKRKIVY  
 A\_thaliana\_1 VLTRNYRKNSGSPKFD-----TGTGTTSKFDFT-----DLTCPHTWYPIARKKKRKVIL  
  
 I II  
 H\_sapiens HSGPTNSGKTYHAIQKYFS AKSGVYCGPLTSLAHEIFEKSNAAGVPCDLVTGEE--RVTV  
 C\_elegans HAGPTNSGKTYHALKRFGEAKSAVECGPLKLLAAEVFHRNTNELGIPCDLVTGEE--RRFA  
 S\_cerevisiae HIGPTNSGKTYRALQKLKSVDRGYAGPLRLLAREVYDRFHAEKIRCNLLTGEEVIRDLD  
 A\_thaliana\_2 HCGPTNSGKTYNALQRFMEAKNGLYCSPLRLLAMEVFDKVNALGIYCSLLTGQE--KKYV  
 A\_thaliana\_1 HVGPTNSGKTYSALKHLEQSSSGVYCGPLRLLAWEVAKRLNKANVPCDLITGQE--KDLV  
  
 V VI\*  
 H\_sapiens QPNGKQASHVSC TVEMCSVTTPYEVAVIDEIQM-----IRDPAR-----GWAWTRALLGL  
 C\_elegans KDNHHPSQHLSSSTVEMLSSTQMRVEVAVIDEIQM-----LRDEQR-----GWAWTRALLGA  
 S\_cerevisiae DR-GNSAGLTS GTVEMVPI NQKF DVVVLDEIQM-----MSDGDR-----GWAWTNALLGV  
 A\_thaliana\_2 P----FANHVSCTVEMVSTDELYEVAVLDEIQM-----MADPSR-----GHAWTKALLGL  
 A\_thaliana\_1 EG----ATHKAVTVEMADVTSVYDCAIIDEIQASLARLWLKKSTRTFCLGEAFTRALLGI  
  
 H\_sapiens CAEEVHLCGEPAAIDLVME LMYTTGEEVEVRDYKRLTPI SVLDHALE-SLDNLRPGDCIV  
 C\_elegans AADEIHLCGEPAAIDIVKKLLEPIGETVEVRYYERKSPLAIADKAIE-SYSNIEPGDCIV  
 S\_cerevisiae VSKEVHLCGEKSVLPLVKSIVKMTGDKLTINEYERLGLKLSVEEKPIKDGIKGLRKGD CVV  
 A\_thaliana\_2 KADEIHLCGDPVSLDIVRKMCADTGDELVEEHYERFKPLVVEAKTLLGELKNVKS GDCVV  
 A\_thaliana\_1 AADELHLCGDPAVVPLVEDILKV TGDDVEVHTYERLSPLVPLKVPVS-SVSSIKTGDCLV  
  
 VII  
 H\_sapiens CFSKNDIYSVSRQIEIR-GLESA VIYGS LPPG TKLAQA AKKFNDPNDPCKILVATDAI GMG  
 C\_elegans CFSKRSIFFNSKKLEEN-GIKPAVIYGD LPPG TKLAQA AKKFNDPDDEC NVLVATDAI GMG  
 S\_cerevisiae AFSKKKILDLKLKIEKDTNLKVAVIYGS LPPPETRVQQAALFNN--GEYD IMVASDAI GMG  
 A\_thaliana\_2 AFSRREIFEVKMAIEKHTNHRCCVIYGA LPPPETRRQQAALFNDQENEYDVLVASDAI GMG  
 A\_thaliana\_1 TFSRKDIYAYKKTIERAGKHLCSVYVYGS LPPPETRTAQATRFNDETND EYDVLVASDAI GMG  
  
 VIII  
 H\_sapiens LNL SIRRIIFYSLIKPSINEKGERELEPITTSQALQIAGRAGRFSSR-----FKEGEVTT  
 C\_elegans LNLNIRRVIFNSCTR-----QTELLP--TYAALQIAGRAGRFGTA-----YANGVATT  
 S\_cerevisiae LNL SIDRVVFTTNMKYN-----GEELMEMTSSQIKQIGRAGRFKRSRSASGGVPQGFITS  
 A\_thaliana\_2 LNLNIRRVVFYSLNKYN-----GDKIVPVAASQVKQIAGRAGRRGSR-----YPDGLTTT  
 A\_thaliana\_1 LNLNISRIIFSTLQKYD-----GSETRDLTVSEIKQIAGRAGRFQSK-----FPIGEVTC  
  
 H\_sapiens MNHEDLSLLKEILKRFPVDPIRAAGLHPTAEQIEMFAYHLPDAT-LSNLIIDIFVDFSQ--V  
 C\_elegans MRKEDLGT LKAILSEKIEPIANVGIAPTYDQIETFSFHL PQAS-FVRLLDLFVSVCS--V  
 S\_cerevisiae FESKVLKSVRKAIEAPVEYLKTA VTP TDEICAQLMTQFP PGTPTSVLLQTISDELEKSS  
 A\_thaliana\_2 LHLEDLNYLIECLQPPFDEVTKVGLFPFFFEQIELFAAQVPDMA-FSNLLEHFGKHCR--L  
 A\_thaliana\_1 LHKEDLPL LHSSLKSPS-PILEANLWH-----



# Tempo zmian

---

- Czynnikiem decydującym o tempie zmian jest dobór oczyszczający (negatywny)
  - w “ważniejszych” sekwencjach więcej zmian będzie niekorzystnych (eliminacja przez dobór)
  - w mniej istotnych sekwencjach więcej zmian będzie neutralnych (utrwalanie przez dryf)
- zmiany bez znaczenia dla funkcji będą neutralne
  - pseudogeny
  - niekodujące obszary międzygenowe?
  - podstawienia synonimiczne?



# Spór wokół ENCODE

- ENCODE - projekt opisujący sekwencje w genomie (Encyclopedia of DNA Elements)
- Wiele sekwencji międzygenowych, niekodujących ulega transkrypcji
  - 80% genomu funkcjonalne
  - czy istnieje “śmieciowy DNA”?
- Czy to znaczy, że są funkcjonalne?
- **Jeżeli nie ma śladów działania doboru - nie ma funkcji!**
- Ślady działania doboru: 2-15% całego genomu

## ARTICLE

doi:10.1038/nature11247

### An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium\*

The human genome encodes the blueprint of life, but the function of the vast majority of its nearly three billion bases is unknown. The Encyclopedia of DNA Elements (ENCODE) project has systematically mapped regions of transcription, transcription factor association, chromatin structure and histone modification. These data enabled us to assign biochemical functions for 80% of the genome, in particular outside of the well-studied protein-coding regions. Many

GBE

#### On the Immortality of Television Sets: “Function” in the Human Genome According to the Evolution-Free Gospel of ENCODE

Dan Graur<sup>1,\*</sup>, Yichen Zheng<sup>1</sup>, Nicholas Price<sup>1</sup>, Ricardo B.R. Azevedo<sup>1</sup>, Rebecca A. Zufall<sup>1</sup>, and Eran Elhaik<sup>2</sup>

<sup>1</sup>Department of Biology and Biochemistry, University of Houston

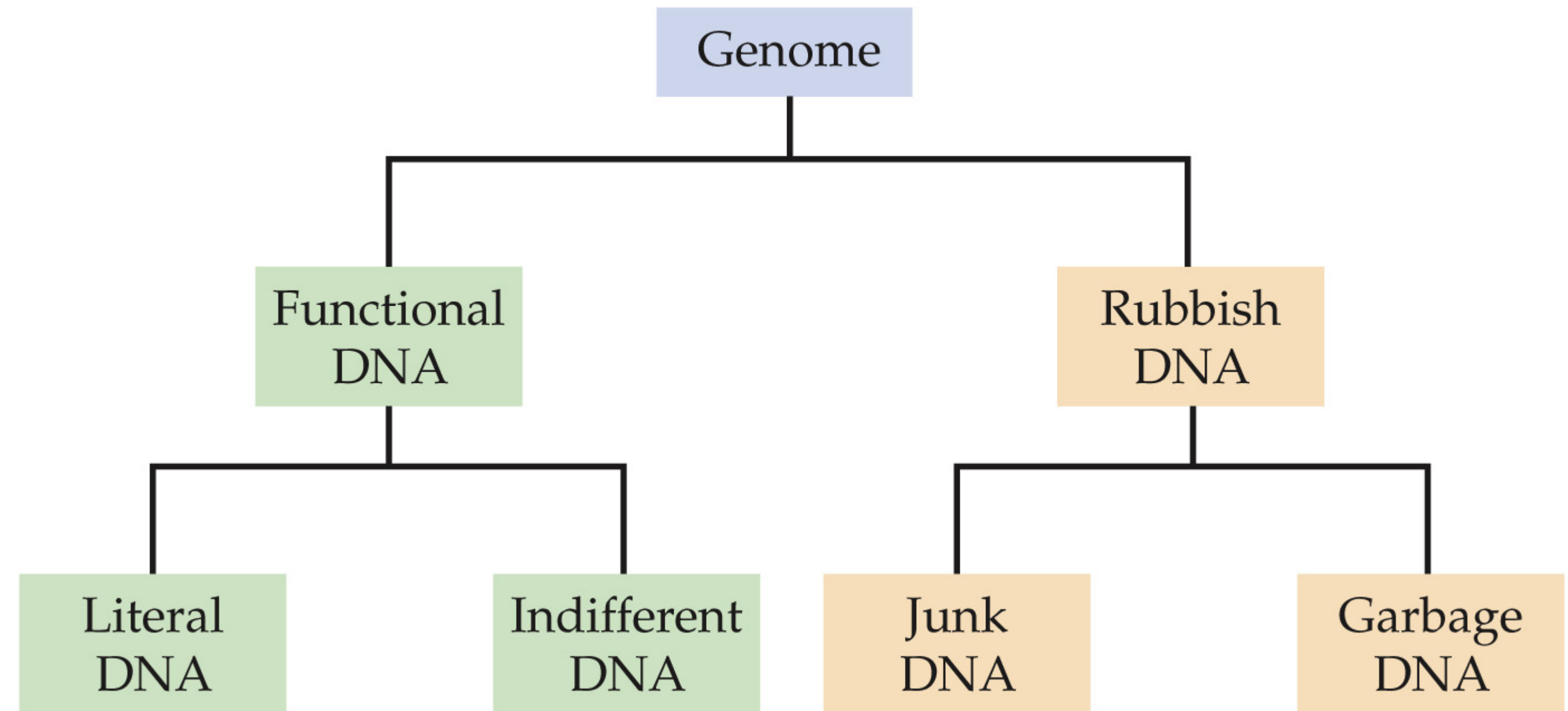
<sup>2</sup>Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health

\*Corresponding author: E-mail: dgraur@uh.edu.

Accepted: February 16, 2013

# Elementy genomu - klasyfikacja ewolucyjna

- *Literal DNA* - selekcji podlega sekwencja nukleotydowa
- *Indifferent DNA* - fragmenty potrzebne, ale sekwencja nieistotna (np. strukturalne)
- *Junk DNA* - neutralne
- *Garbage DNA* - niekorzystne, ale utrzymują się mimo kontrselekcji, samolubne geny



MOLECULAR AND GENOME EVOLUTION 1e, Figure 11.1  
© 2016 Sinauer Associates, Inc.

# Status neutralizmu

---

- Wyjaśnia wiele zjawisk obserwowanych w ewolucji molekularnej
  - wysoki polimorfizm sekwencji DNA i białek
  - zegar molekularny
    - ale jest wiele odstępstw, nie istnieje globalny zegar prawdziwy dla wszystkich gałęzi drzewa życia
  - wolniejsza ewolucja sekwencji o kluczowym znaczeniu
    - to też można wyjaśnić modelem, w którym większość mutacji jest albo niekorzystna, albo korzystna, ale niekorzystnych jest więcej
- Jest bardzo przydatny jako hipoteza zerowa do badania doboru naturalnego na poziomie sekwencji!



# Status neutralizmu

---

- Dryf genetyczny ma w ewolucji molekularnej bardzo znaczącą, ale nie wyłączną rolę
  - znaczne obszary genomu ewoluują w sposób bliski neutralnemu



# Badanie doboru

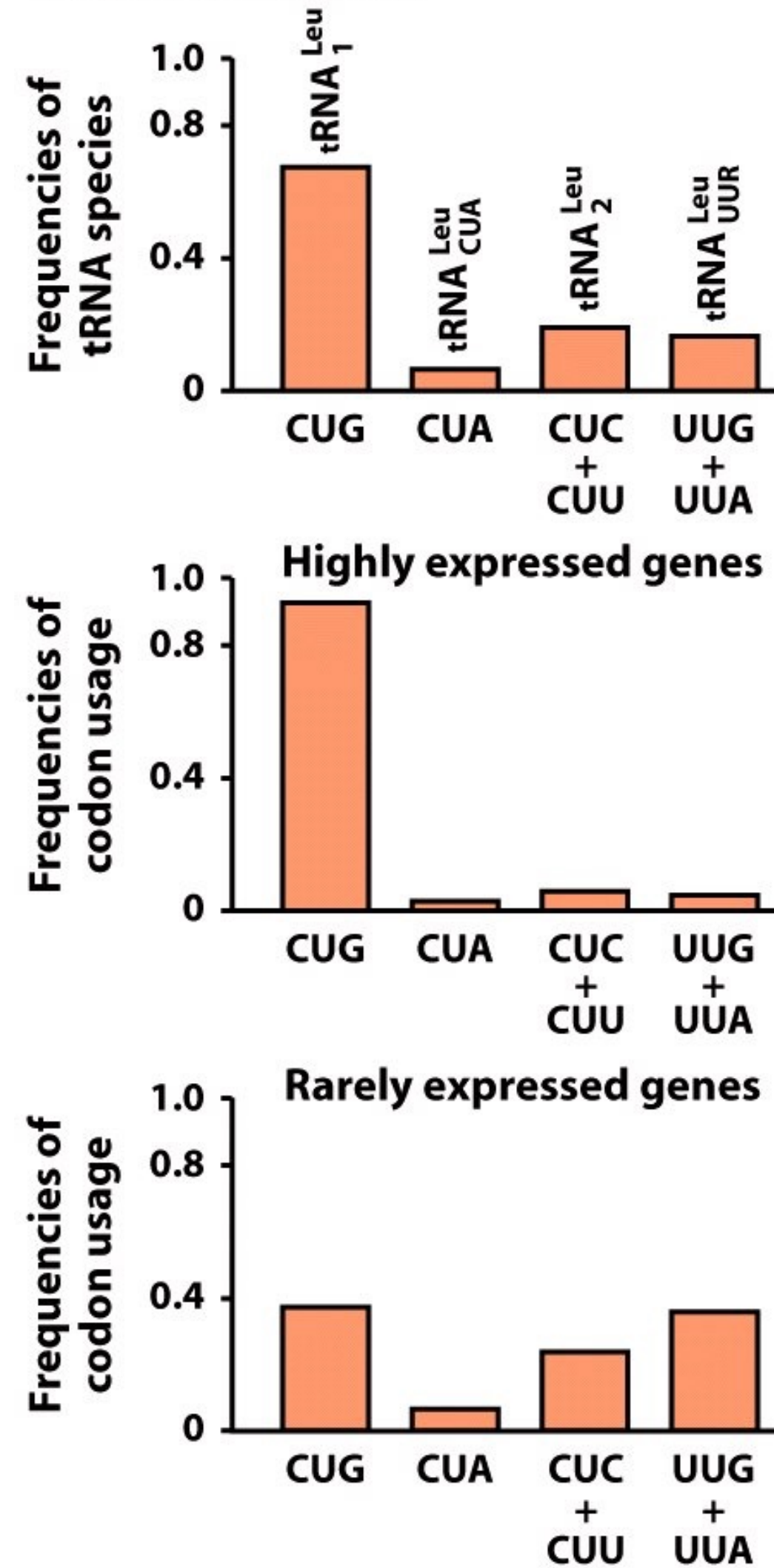
---

- Założenie: mutacje synonimiczne są neutralne, sekwencje porównywane są parami
- $K_a$  (dN) – liczba mutacji niesynonimicznych na liczbę możliwych miejsc niesynonimicznych
- $K_s$  (dS) – liczba mutacji synonimicznych na liczbę możliwych miejsc synonimicznych
- Stosunek  $K_a/K_s$  ( $\omega$ ) jest miarą działania doboru

# Czy zmiany synonimiczne są neutralne

- Kodony synonimiczne nie są równocenne
- Zmiana kodonu częstego na rzadki może wpłynąć na poziom ekspresji i kinetykę translacji

(a) *Escherichia coli*



(b) Yeast

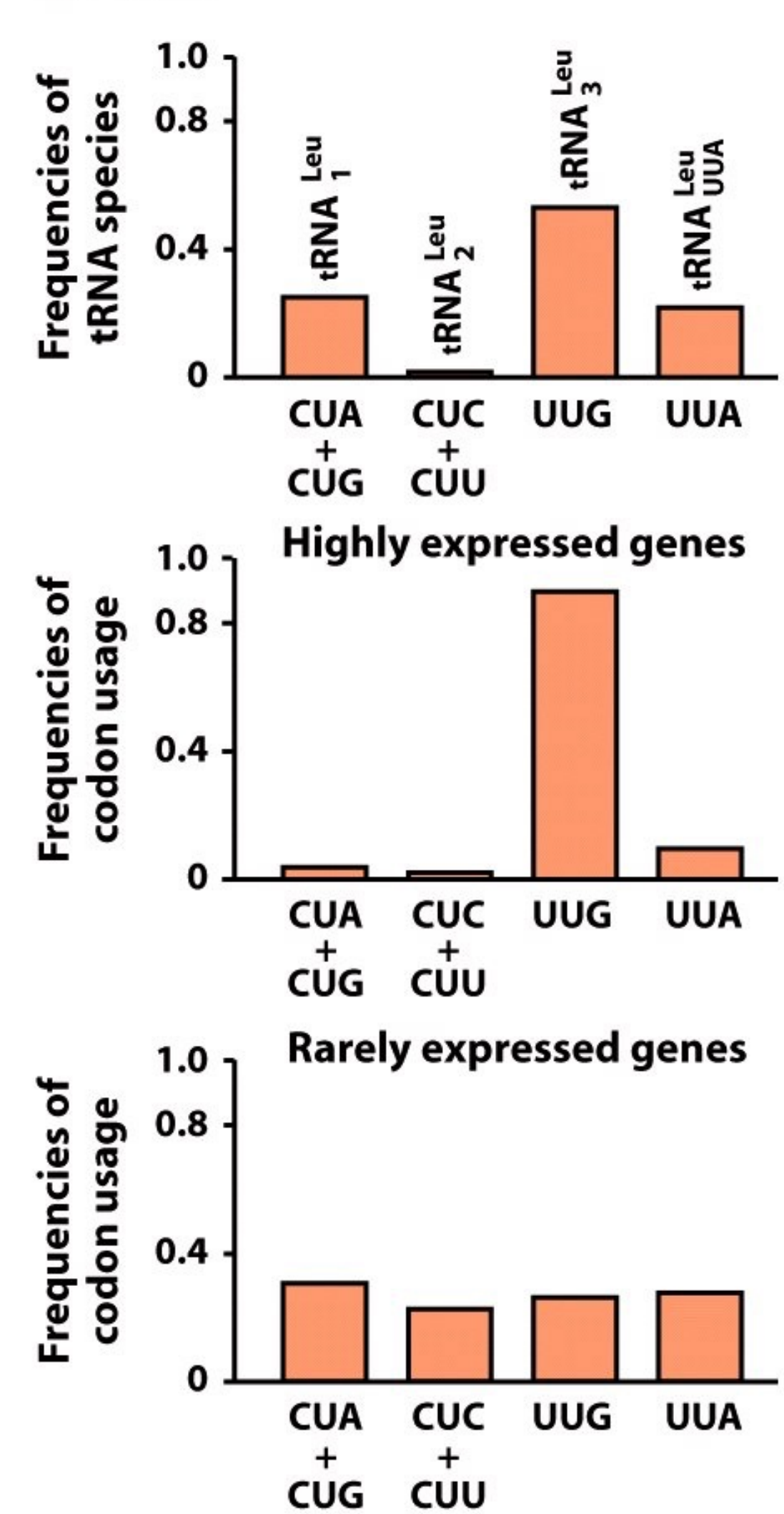


Figure 7-24 Evolutionary Analysis, 4/e  
© 2007 Pearson Prentice Hall, Inc.



# Czy zmiany synonimiczne są neutralne?

## A “Silent” Polymorphism in the *MDR1* Gene Changes Substrate Specificity

Chava Kimchi-Sarfaty,\*† Jung Mi Oh,†‡ In-Wha Kim, Zuben E. Sauna, Anna Maria Calcagno, Suresh V. Ambudkar, Michael M. Gottesman†

Synonymous single-nucleotide polymorphisms (SNPs) do not produce altered coding sequences, and therefore they are not expected to change the function of the protein in which they occur. We report that a synonymous SNP in the *Multidrug Resistance 1* (*MDR1*) gene, part of a haplotype previously linked to altered function of the *MDR1* gene product P-glycoprotein (P-gp), nonetheless results in P-gp with altered drug and inhibitor interactions. Similar mRNA and protein levels, but altered conformations, were found for wild-type and polymorphic P-gp. We hypothesize that the presence of a rare codon, marked by the synonymous polymorphism, affects the timing of cotranslational folding and insertion of P-gp into the membrane, thereby altering the structure of substrate and inhibitor interaction sites.

**T**he *MDR1* gene product, the adenosine triphosphate (ATP)-binding cassette (ABC) transporter ABCB1 or P-gp, is an ATP-driven efflux pump contributing to the pharmacokinetics of drugs that are P-gp substrates and to the multidrug resistance of cancer cells (1, 2). To date, more than 50 single-nucleotide polymorphisms (SNPs) have been reported for *MDR1* ([www.ncbi.nlm.nih.gov/SNP/GeneGt.cgi?geneID=5243](http://www.ncbi.nlm.nih.gov/SNP/GeneGt.cgi?geneID=5243)). One of these, a synonymous SNP in exon 26 (C3435T), was

sometimes found to be associated with altered P-gp activity (3–6) and, when it appears in a haplotype, with reduced functionality (7). This association may be explained in different ways. Perhaps it is because C3435T is in linkage disequilibrium with other common functional non-synonymous polymorphisms such as G2677T. In fact, the C1236T (a synonymous SNP), G2677T, and C3435T polymorphisms are part of a common haplotype (8, 9). Another possible explanation is that allele-specific differences in

mRNA folding could influence splicing, processing, or translational control and regulation (10, 11). A third possibility is that the effect of the C3435T polymorphism on the levels of cell surface P-gp activity or its function is rather modest or drug-specific. Finally, numerous environmental factors are known to affect the expression and phenotypic activity of P-gp (12).

To determine whether the C3435T polymorphism actually does affect P-gp activity, we expressed wild-type and polymorphic P-gps in HeLa cells with the use of a transient expression system (13). The same experiments were carried out on BSC-1 (epithelial cells of African green monkey kidney origin), Vero-76 (monkey kidney cells), and 12E1 (CEM human cells) cell lines (14), with similar results, indicating that this phenomenon is not specific to HeLa cells.

Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, Bethesda, MD 20892, USA.

\*Present address: Center for Biologics Evaluation and Research, Food and Drug Administration, 29 Lincoln Drive, Room 316, Bethesda, MD 20892, USA.

†To whom correspondence should be addressed. E-mail: [mgottesman@nih.gov](mailto:mgottesman@nih.gov) (M.M.G.); [jmoh@snu.ac.kr](mailto:jmoh@snu.ac.kr) (J.M.O.); [kimchi@cber.fda.gov](mailto:kimchi@cber.fda.gov) (C.K.-S.)

‡Present address: College of Pharmacy, Seoul National University, Seoul 151-742, South Korea.



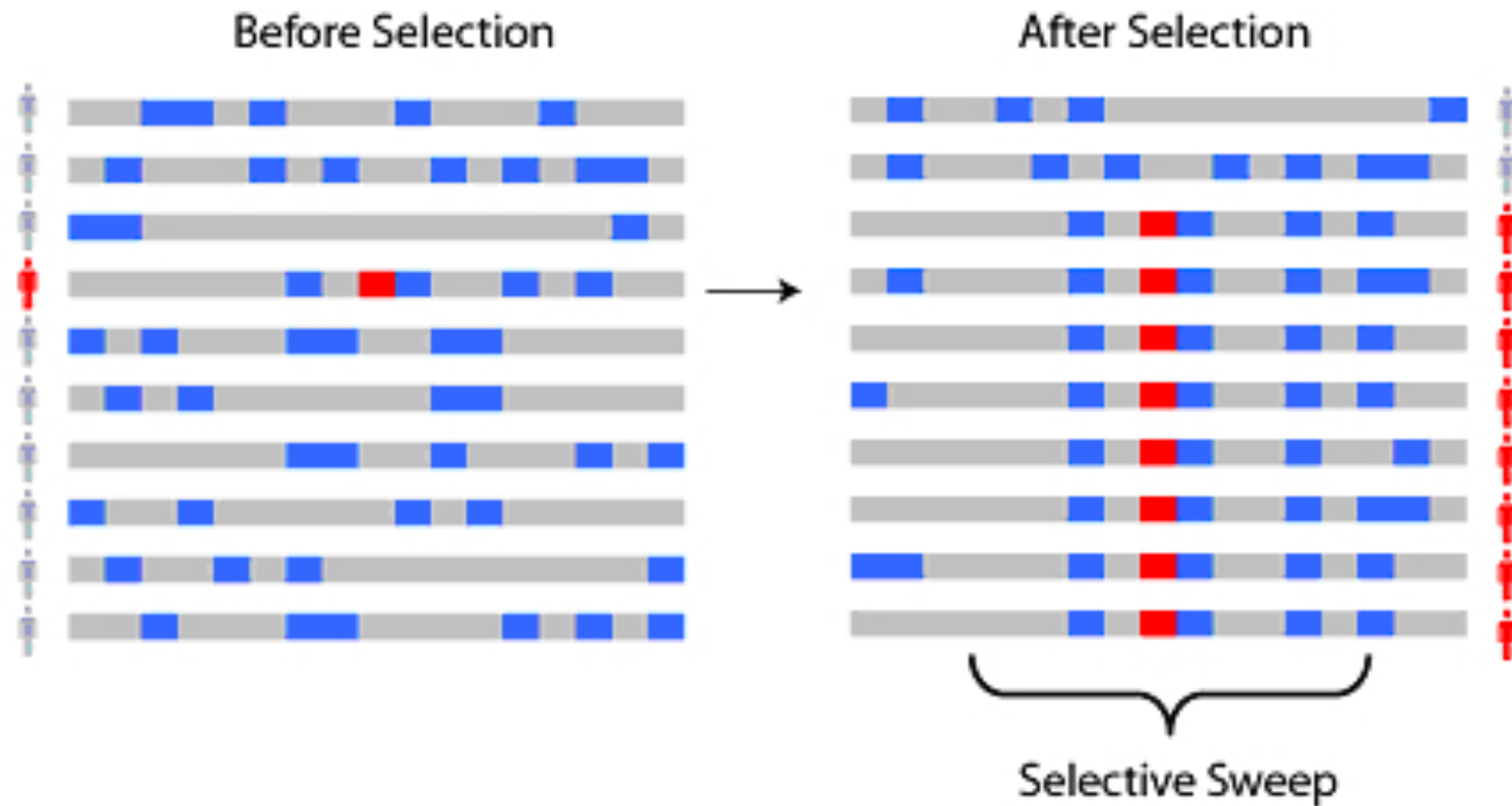
## *Hitch-hiking* i zmiatanie selekcyjne

---

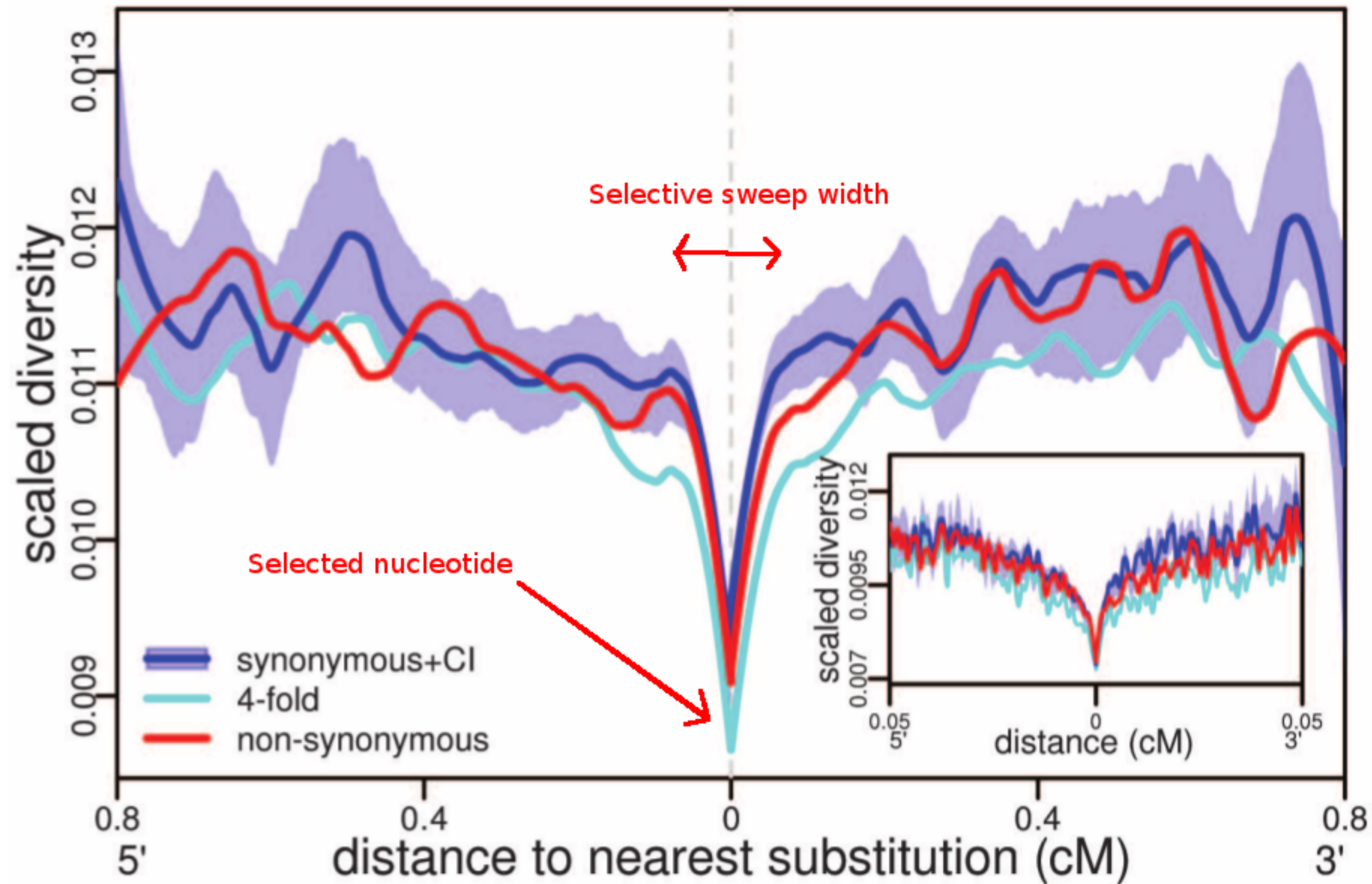
- W jednym *locus* pojawia się korzystna mutacja
- Dobór naturalny szybko utrwała ten korzystny allel
- Wraz z nim utrwalają się neutralne (a nawet niekorzystne) allele *loci* blisko sprzężonych – *genetic hitch-hiking*
- W sąsiedztwie niedawno utrwalonego korzystnego allelu obserwuje się zmniejszoną różnorodność alleliczną – zmiatanie selekcyjne (*selective sweep*)



# Zmiananie selekcyjne

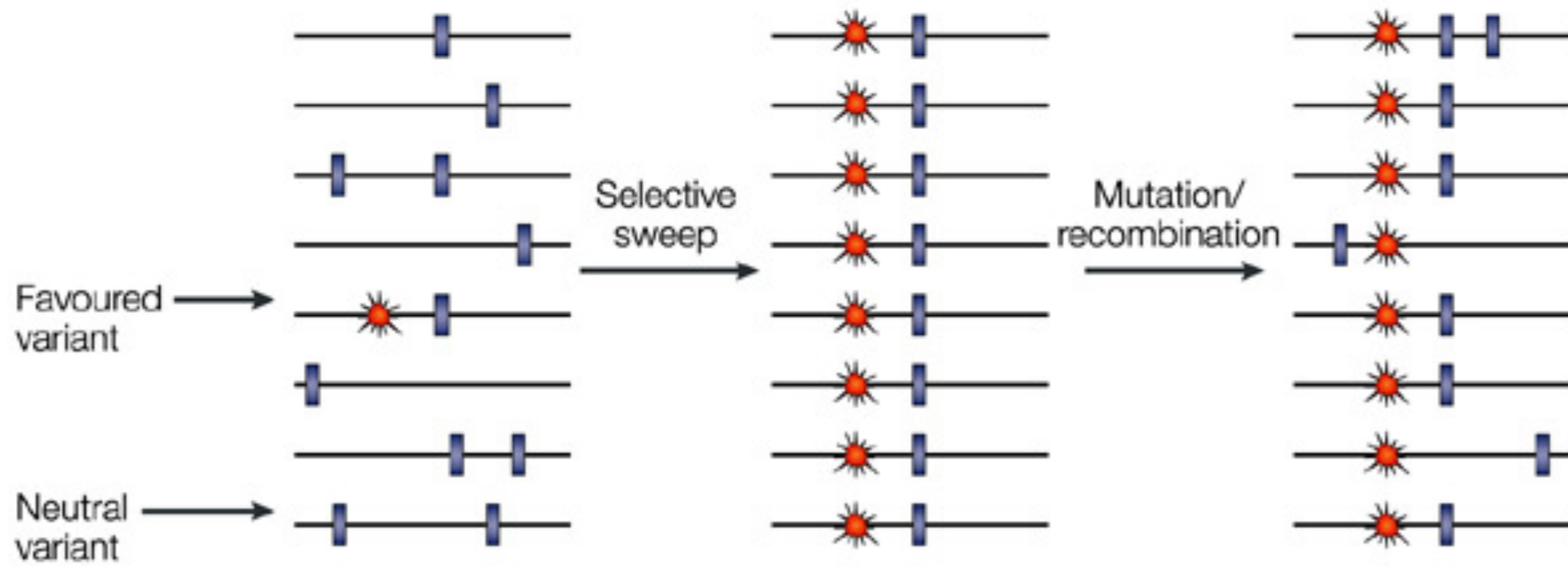


# Ślady zmiatania selekcyjnego



# Zmiatanie selekcyjne jest zjawiskiem krótkotrwałym

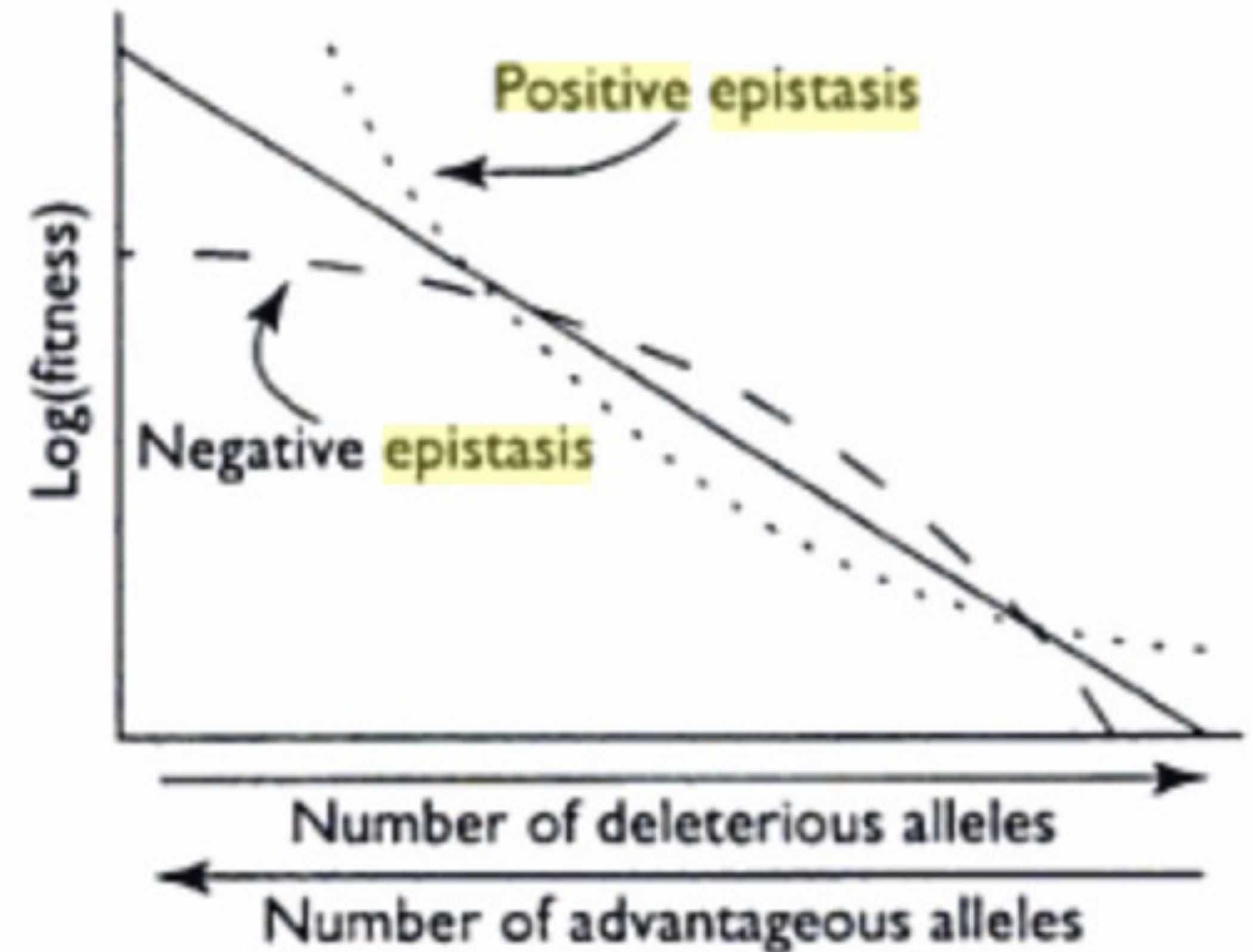
Powstała w wyniku zmiatania selekcyjnego nierównowaga sprzężeń z czasem zanika na skutek rekombinacji i kolejnych mutacji





# Epistaza

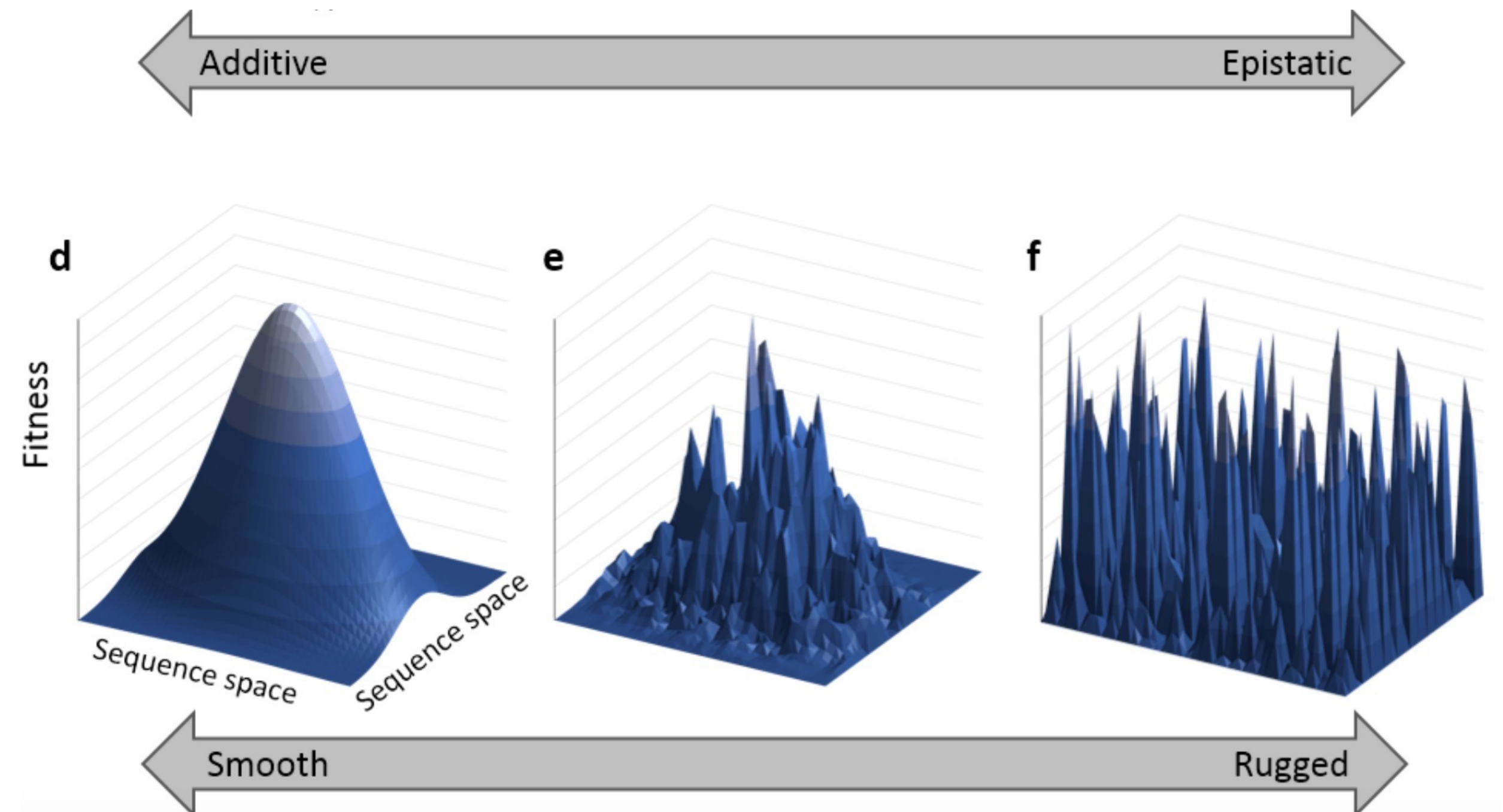
- W genetyce populacji termin oznacza dowolne oddziaływania genetyczne (w genetyce klasycznej tylko jeden rodzaj oddziaływania)
- Epistaza
  - pozytywna - dostosowanie podwójnego mutantu większe, niż oczekiwane
  - negatywna - dostosowanie podwójnego mutantu mniejsze, niż oczekiwane
  - epistaza znaku - np. dwie mutacje szkodliwe (pojedynczo) razem są korzystne
  - model wartości oczekiwanej: addytywny lub (częściej) multiplikatywny





# Epistaza i ewolucja

- Przy braku epistazy nie ma znaczenia kolejność nabywania mutacji
- Przy epistazie efekt danej mutacji zależy od tła genetycznego (innych mutacji) - trajektoria ewolucyjna zależy od kolejności zdarzeń

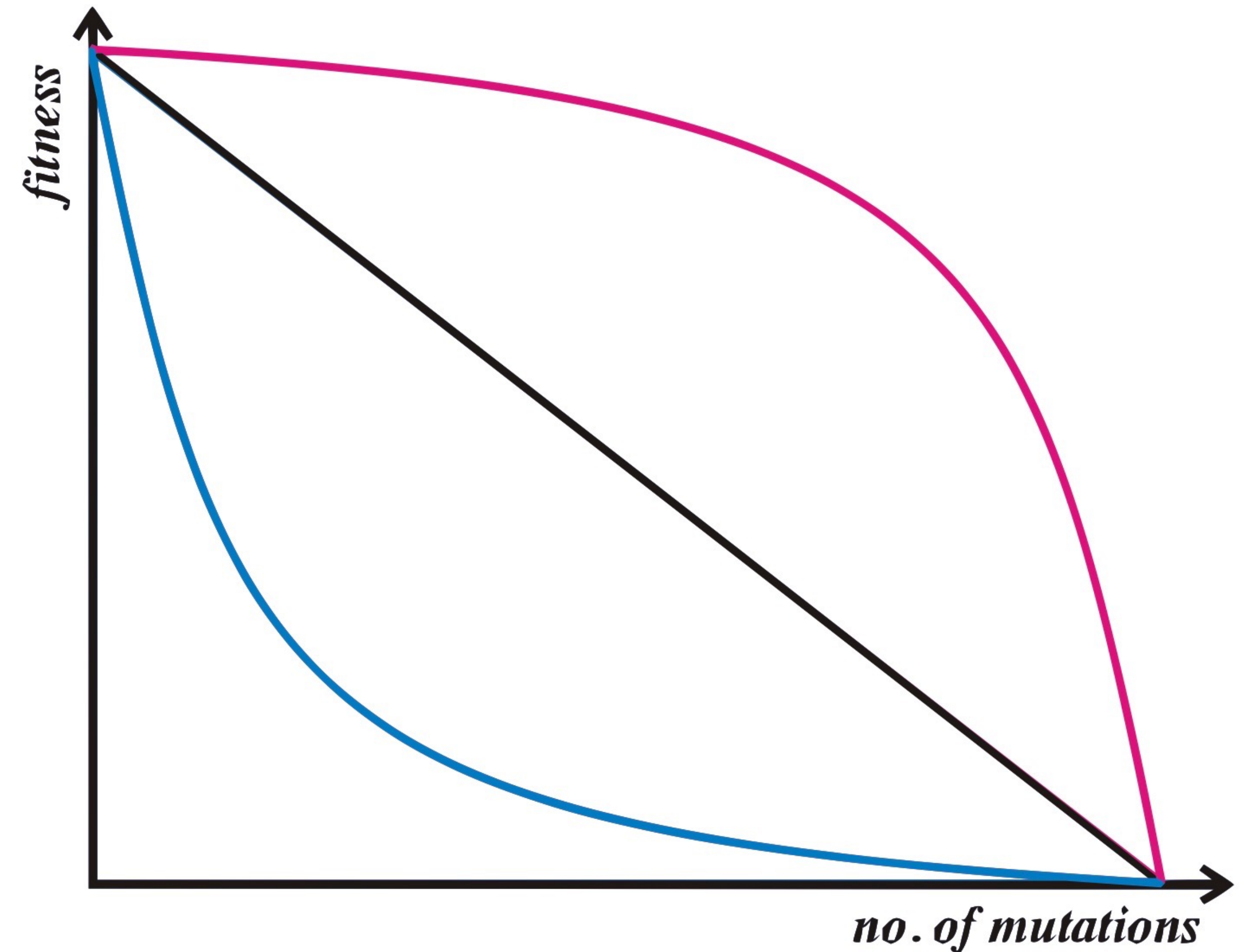


T. Shafee, wikimedia commons

# Epistaza i płęć

---

- W populacjach rozmnażających się płciowo epistaza negatywna pomaga usuwać mutacje szkodliwe dzięki ich synergii:
- hipoteza deterministyczna mutacyjna (A. Kondrashov, 1988)
- inna hipoteza - płęć ułatwia powstanie korzystnych kombinacji niezależnych mutacji



# Genetyka specjacji a epistaza

---

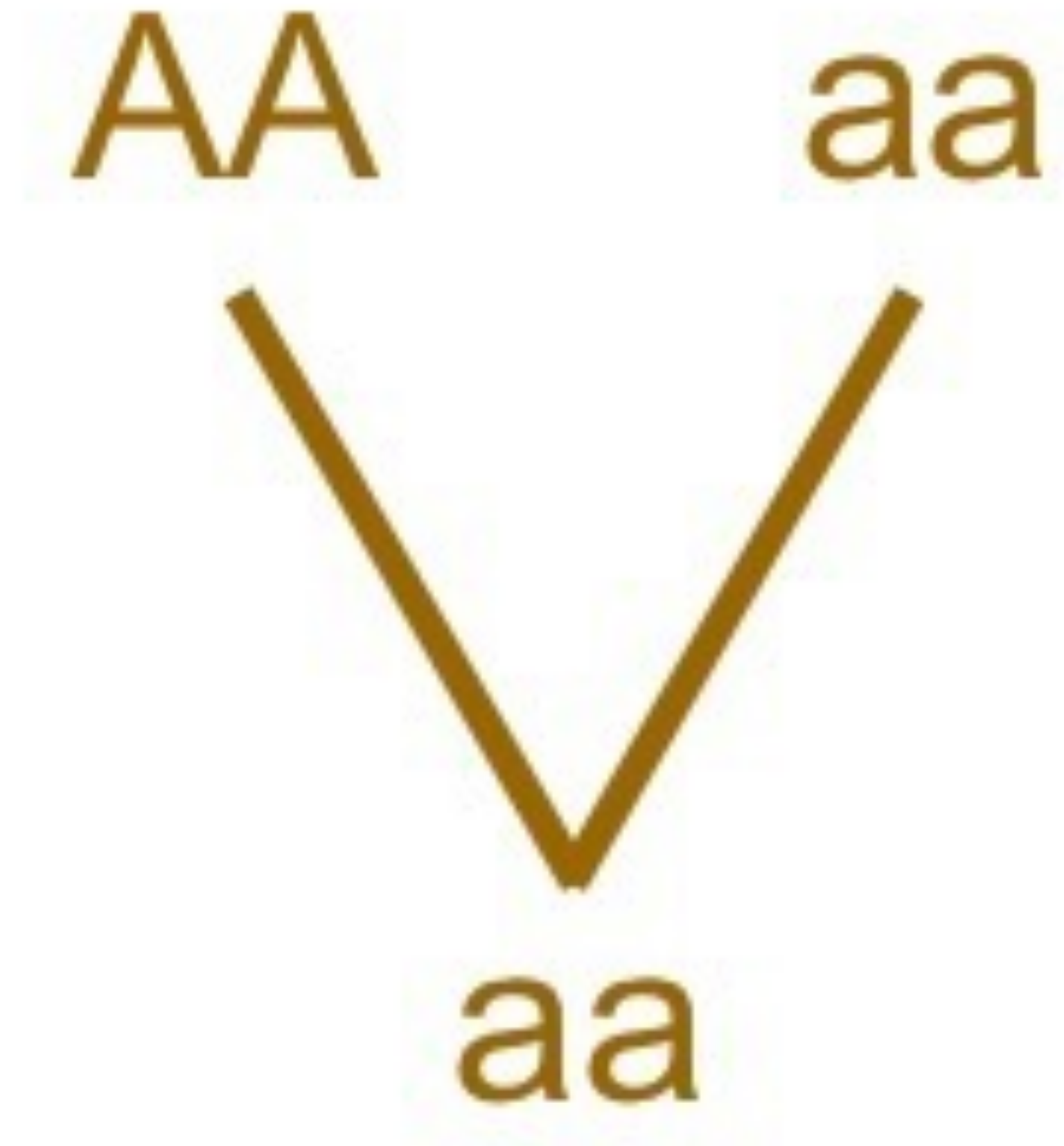
- Podstawą specjacji jest pojawienie się bariery reprodukcyjnej
- Sama bariera reprodukcyjna nie ma wartości dostosowawczej, jej ewolucja jest skutkiem innych procesów
  - doboru
  - dryfu
- Jak powstają bariery rozrodcze?



# Genetyka barier rozrodczych

---

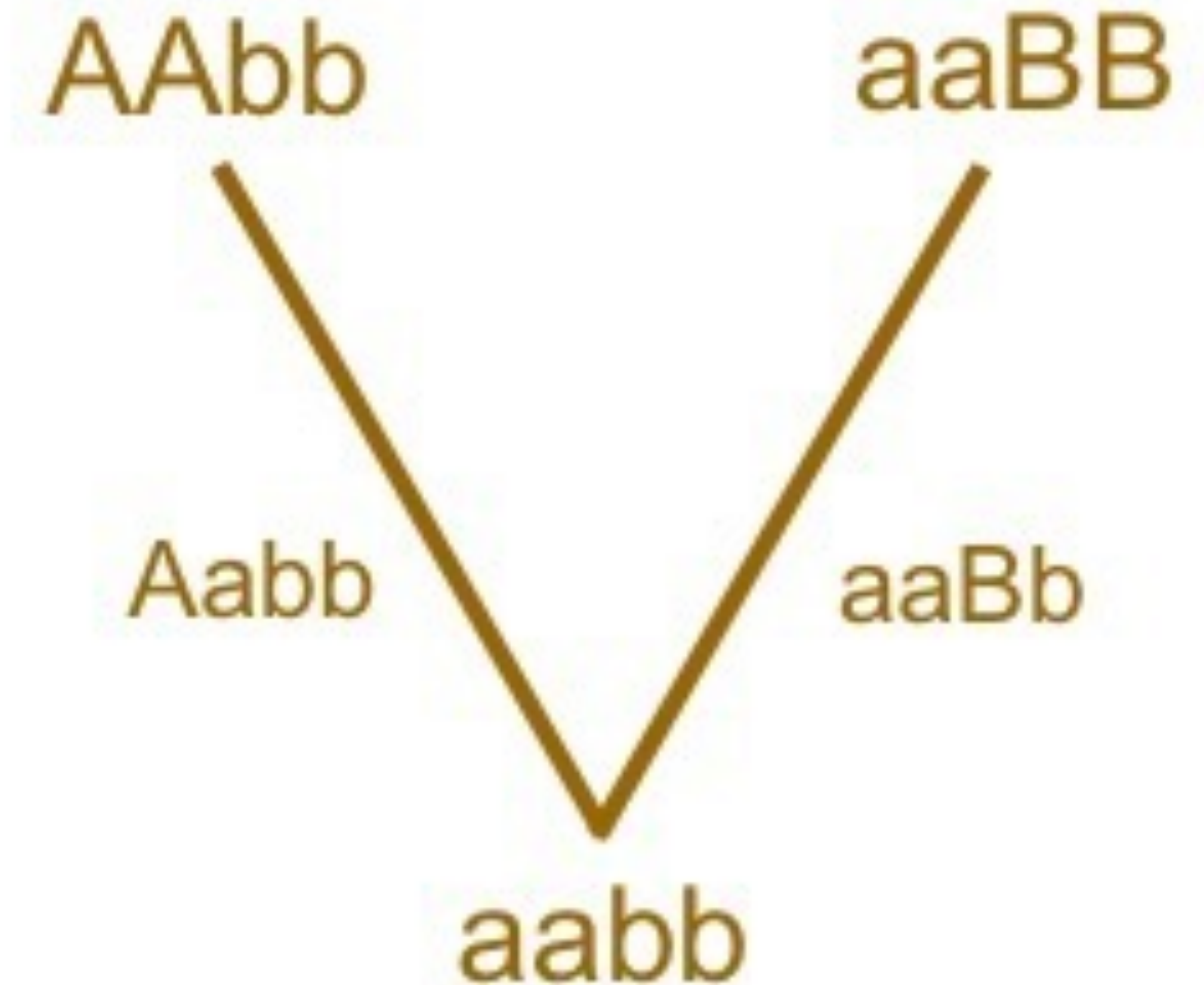
- Pojedynczy *locus* nie wystarczy
- Taki układ nie ma jak wyewoluować bez przejściowego stadium *Aa*
- Jeżeli allele *A* i *a* są niekompatybilne, to nie może zająć





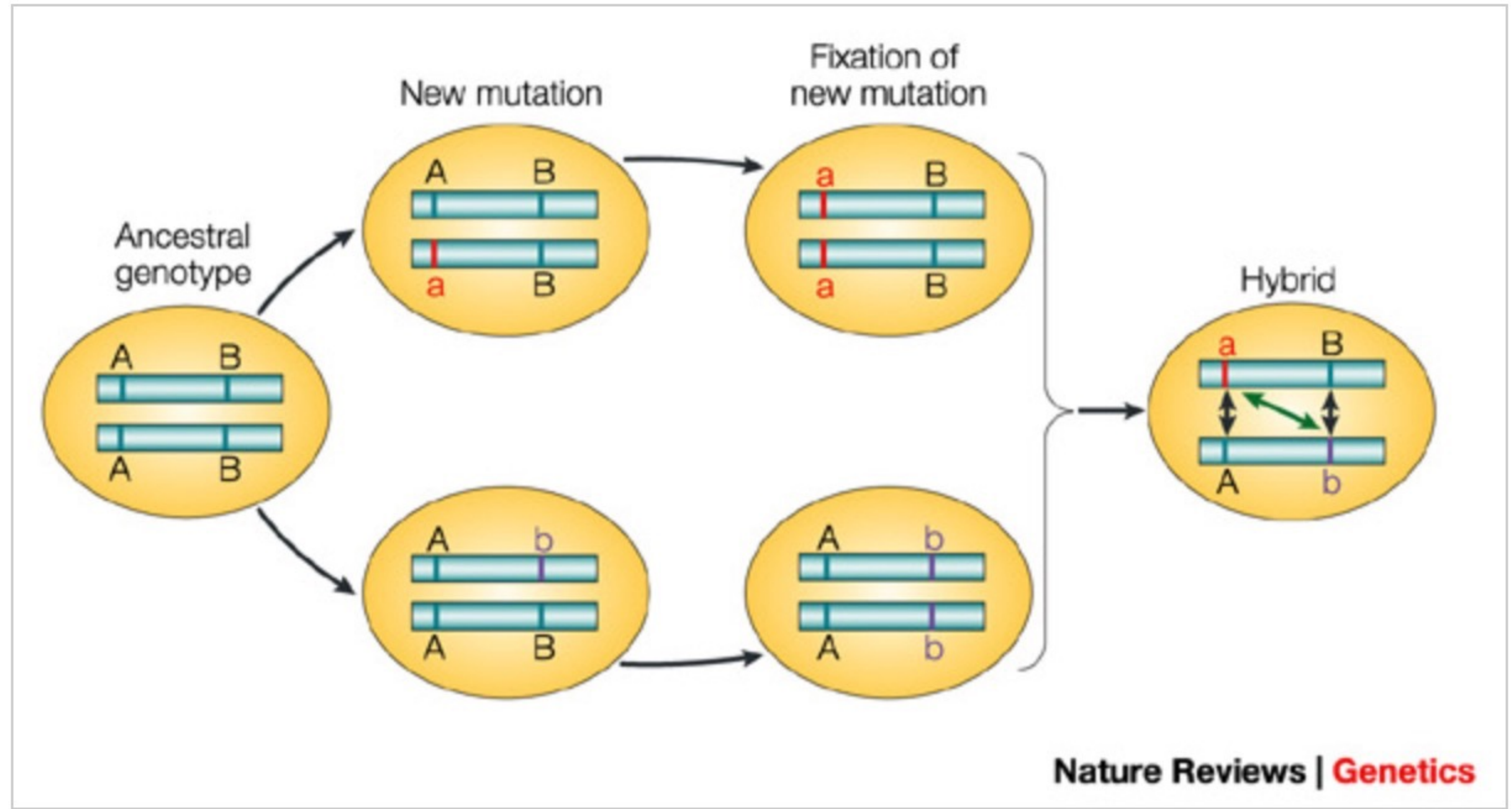
# Model Dobzhanskiego-Müllera

- Oparty na interakcji genetycznej 2 różnych *loci*, w których niezależnie powstają niekompatybilne allele
- Allele *a* i *b* są kompatybilne
- Allel *A* jest kompatybilny z *b*
- Allel *B* jest kompatybilny z *a*
- Allele *A* i *B* nie są kompatybilne
- Może wyewoluować nie przechodząc przez stadium o obniżonym dostosowaniu



# Model Dobzhanskiego-Müllera

- Najprostszy model ewolucji niekompatybilności genetycznej
- Bardzo uproszczony, ale przykłady są znane



Informacja genetyczna a teoria informacji



# Teoria informacji

---

- Zmienna losowa  $X$  – obiekt mogący przyjąć skończoną liczbę stanów  $x_1, \dots, x_N$ , z określonymi prawdopodobieństwami  $p_1, \dots, p_N$
- Przykład



$x_1; p_1=0,5$



$x_2; p_2=0,5$



# Informacja – ujęcie matematyczne

---

- Entropia – miara niepewności dotyczącej stanu zmiennej losowej

$$H(X) = - \sum_{i=1}^N p_i \log p_i$$

- Podstawa logarytmu definiuje jednostkę
  - $\log_2$  – bity
  - $\ln$  – naty
  - $\log_N$  – mery

# Informacja – ujęcie matematyczne

---

- W przypadku monety



$x_1; p_1 = 0,5$



$x_2; p_2 = 0,5$

$$H(X) = - \sum_{i=1}^N p_i \log p_i$$

$$H(X) = - [0,5 * \log_2(0,5) + 0,5 * \log_2(0,5)] = \\ - [0,5 * (-1) + 0,5 * (-1)] = 1 \text{ bit}$$

- Informacja – zmniejszenie niepewności
  - Znajomość wyniku rzutu monetą to 1 bit informacji
  - Maksymalna informacja, którą może nieść układ jest równa jego entropii

# Teoria informacji

---

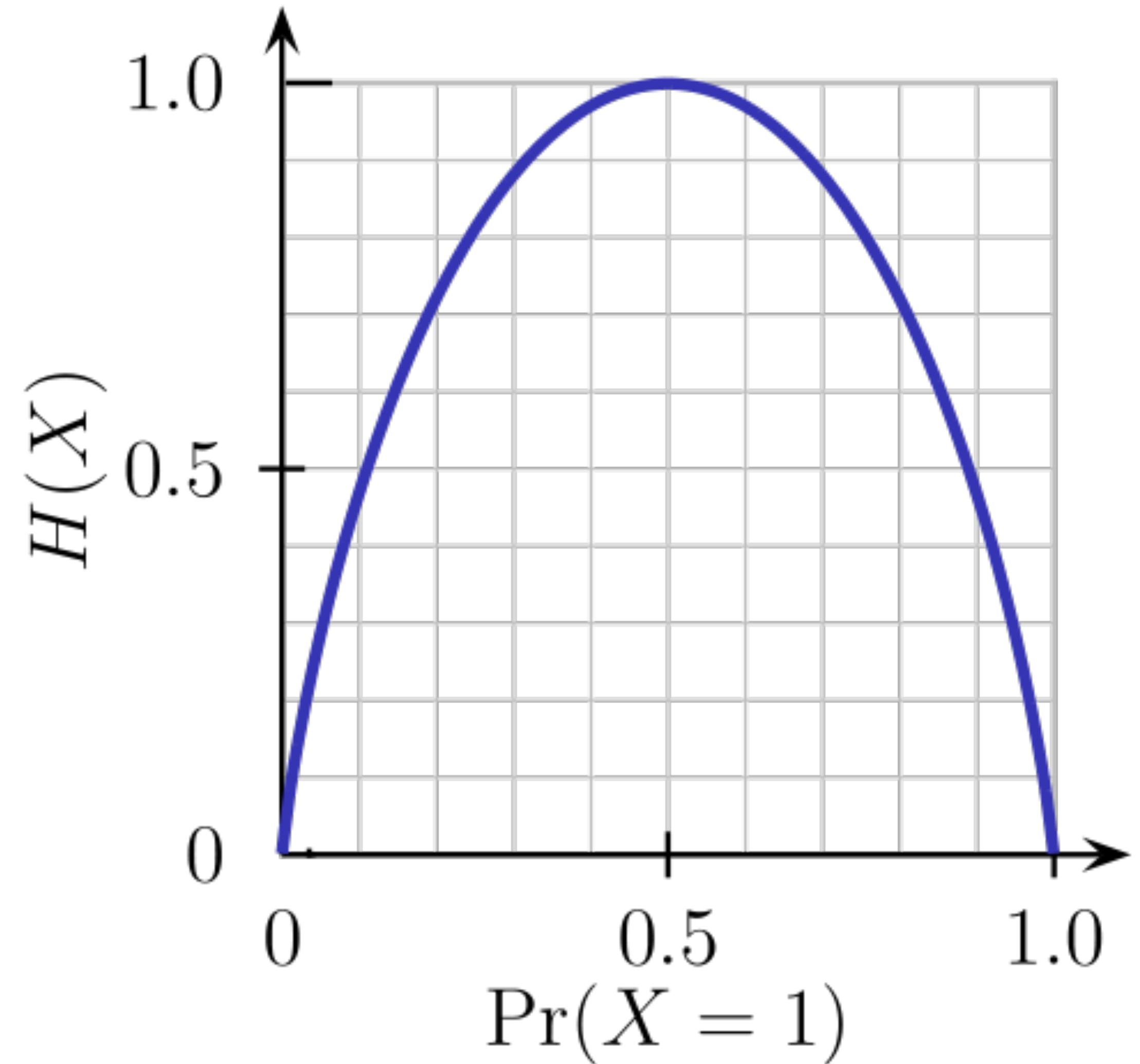
- Informacja musi mieć reprezentację fizyczną (moneta, bity w pamięci komputerowej, nukleotydy w DNA, neurony)
- Informacja musi mieć kontekst
- Dowolny układ symboli jest potencjalną informacją (posiada entropię), ale dopiero związek ze światem fizycznym nadaje mu status informacji



# Maksimum entropii

---

- Entropia jest największa wtedy, gdy  $p_1=p_2=\dots p_N$  – największa niepewność





# Teoria informacji w biologii

---

- W układach biologicznych informacja zapisana jest w DNA
- Jedna pozycja nukleotydowa, gdzie każdy z nukleotydów może występować z jednakowym prawdopodobieństwem:  $H(X) = 2$  bity (1 mer)
- Jest to maksymalna entropia dla tej pozycji
- W rzeczywistości prawdopodobieństwo wystąpienia danego nukleotydu w danej pozycji nie zawsze wynosi 0,25, jeżeli na sekwencję działa dobór

# O czym ta informacja?

---

- Żeby określić właściwą zawartość informacji i odróżnić ją od redundancji, trzeba wiedzieć czego dotyczy
- Każda populacja żyje w określonym środowisku
- Mutacje wywołują zmienność
- Środowisko dokonuje selekcji
- Informacja płynie ze środowiska do genomu (genomów)

# Ewolucja a informacja

---

Informacja w biologii jest więc ściśle powiązana z ewolucją.

Tylko analiza ewolucyjna pozwala na określenie informacji w biologii.



# Informacja a ewolucja

---

- W rzeczywistości prawdopodobieństwo wystąpienia danego nukleotydu w danej pozycji nie zawsze wynosi 0,25, jeżeli na sekwencję działa dobór
- W danym środowisku prawdopodobieństwa wystąpienia w sekwencji określonego nukleotydu są różne (np. C w 78% w danej pozycji, itp.)
- Na tej podstawie obliczamy  $H(X)$
- Różnica między  $H_{max}$  a  $H(X)$  to miara informacji o środowisku zawartej w tej pozycji



# Informacja w sekwencji białka

---

Maksymalna entropia pozycji jeżeli nie ma żadnych ograniczeń:

$$H(X_i) = H_{max} = \log_2(20) \approx 4,32 \text{ bity}$$

W rzeczywistości aminokwasy w danej pozycji nie występują losowo – ogranicza to dobór naturalny, czyli środowisko

Sekwencje DNA i białek zawierają informację o środowisku (szeroko pojmowanym)

# Informacja w sekwencji białka

---

Stan środowiska to zmienna losowa  $E$  o stanach  $e_j$  o określonym prawdopodobieństwie.

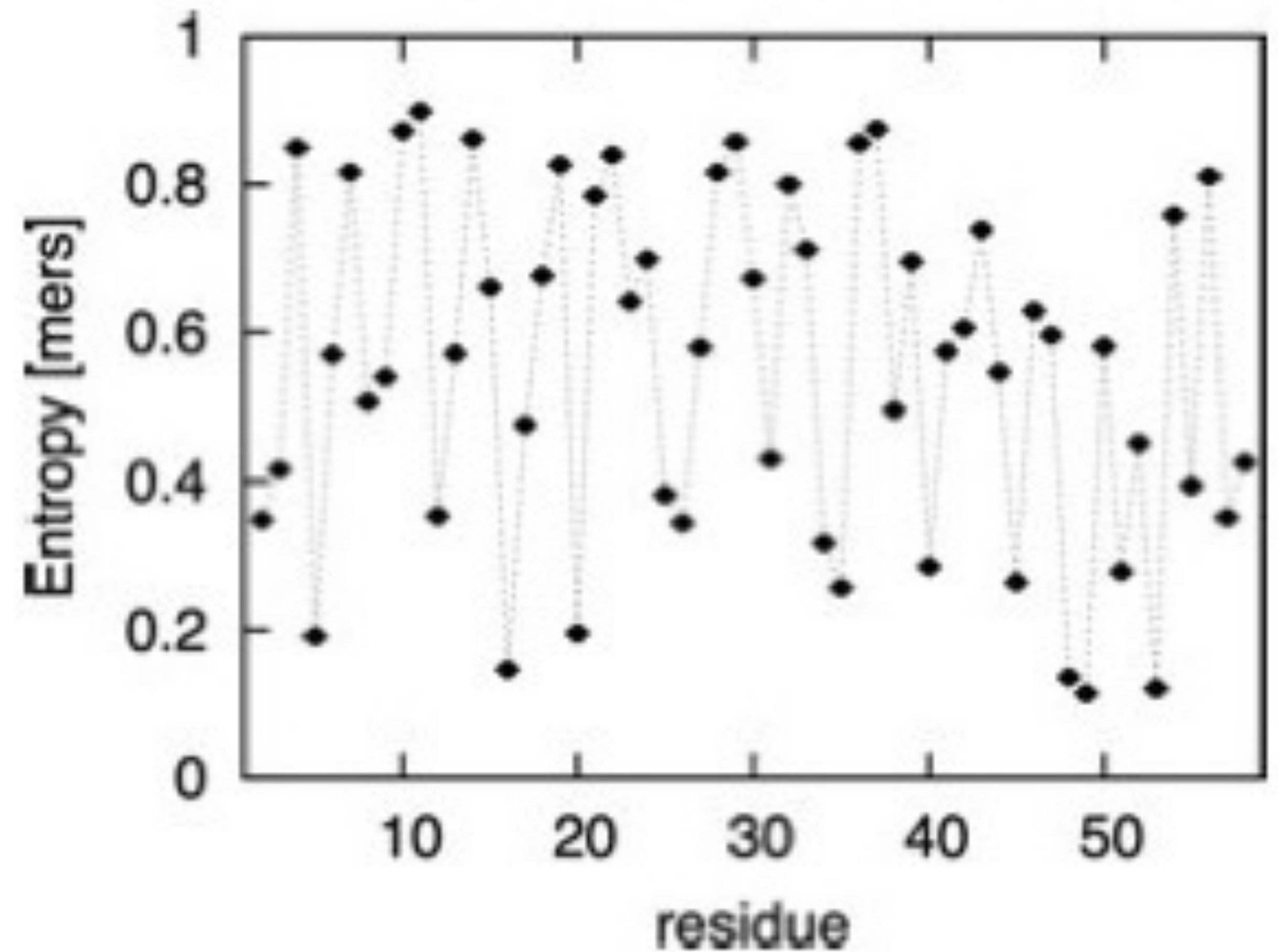
Środowisko ogranicza występowanie w danej pozycji konkretnych aminokwasów (np. w danej grupie w 80% w danej pozycji Trp, itp.).

Na tej podstawie można obliczyć obserwowaną entropię danej pozycji, kształtowaną przez środowisko (poprzez dobór):  $H(X_i|E = e_j)$

Różnica  $H_{max} - H(X_i|E = e_j)$  to informacja o środowisku zapisana w pozycji i sekwencji.

# Profil entropii białka

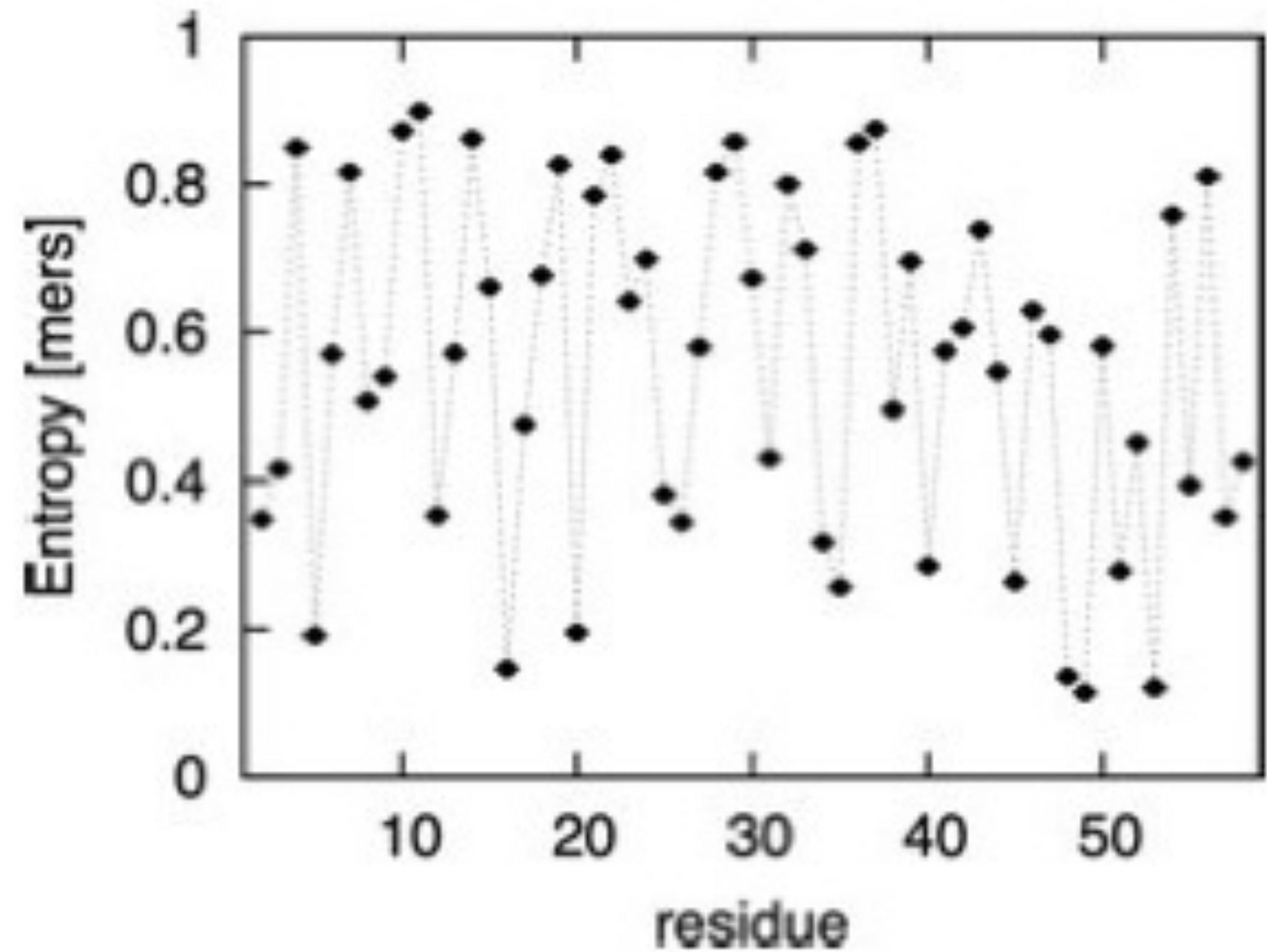
- Przykład: 57 aminokwasów homeodomeny gryzoni, na podstawie porównania 810 sekwencji





# Całkowita zawartość informacji

- Całkowitą entropię (i całkowitą zawartość informacji) uzyskamy sumując entropię dla wszystkich pozycji

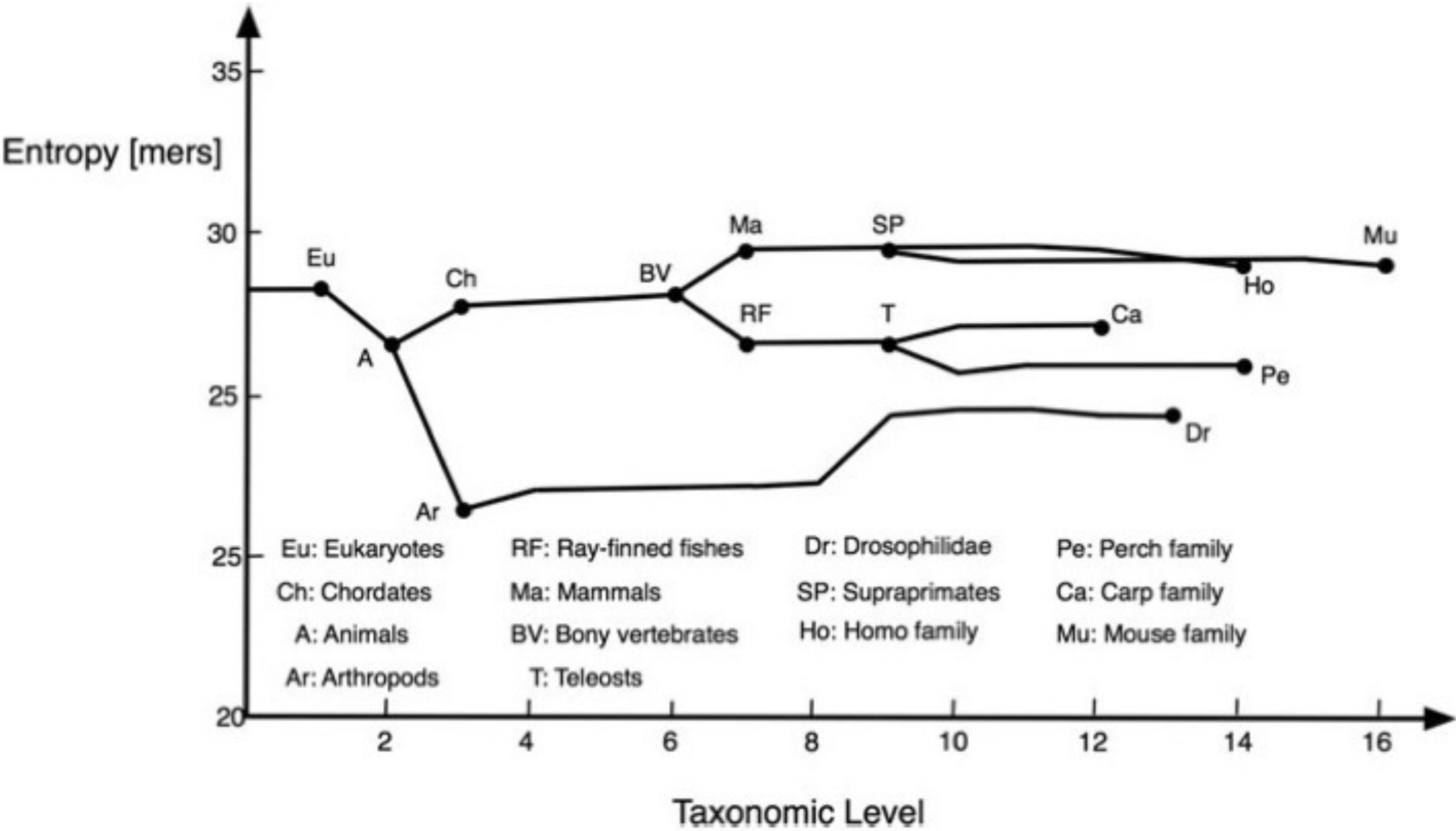


$$I_{\text{gryzonie}} = 57 - \sum_{i=1}^{57} H(X_i)$$

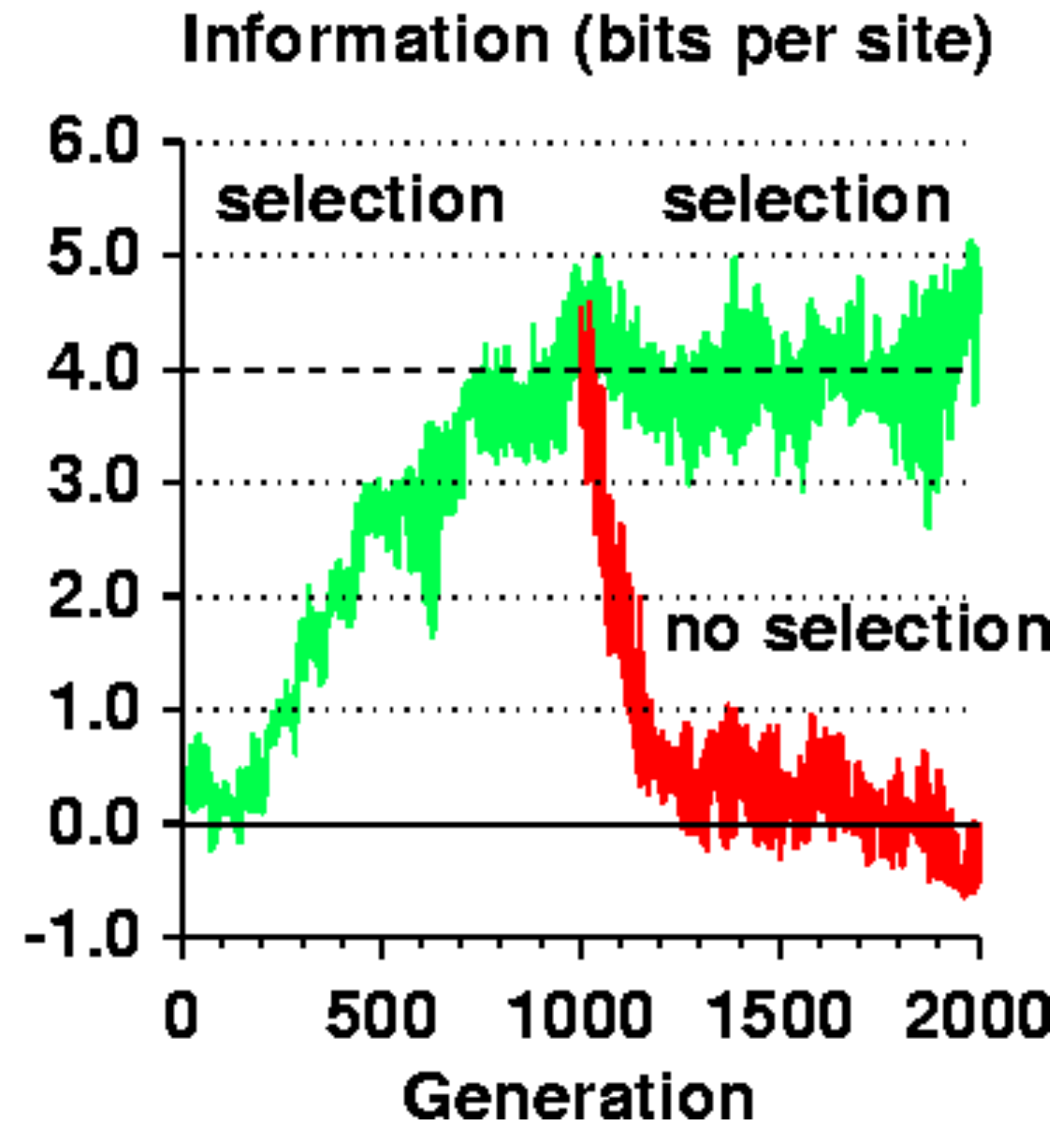
$$I_{\text{gryzonie}} = 25,29 \pm 0,09 \text{ merów} \approx 109 \text{ bitów}$$



# Entropia homeodomeny w ewolucji



# Gdyby nie było doboru...



# Skąd pochodzi informacja genetyczna

---

Informację genetyczną “zapisuje” środowisko poprzez proces doboru naturalnego, równoważąc utratę informacji na skutek błędów replikacji