

Podstawy ewolucji molekularnej

Ewolucja sekwencji DNA i białek

Zmiany genetyczne w ewolucji

- Mutacje
 - tworzą nowe allele genów
- Inwersje
 - zmieniają układ genów na chromosomach
 - mogą uniemożliwić rekombinację na danym odcinku i doprowadzić do utrwalenia haplotypu
- Duplikacje
 - dotyczą fragmentów DNA, w tym całych genów
 - lub całych chromosomów i całych genomów
 - główne źródło innowacji ewolucyjnej
- Transfer horyzontalny
 - w tym zdarzenia symbiotyczne

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

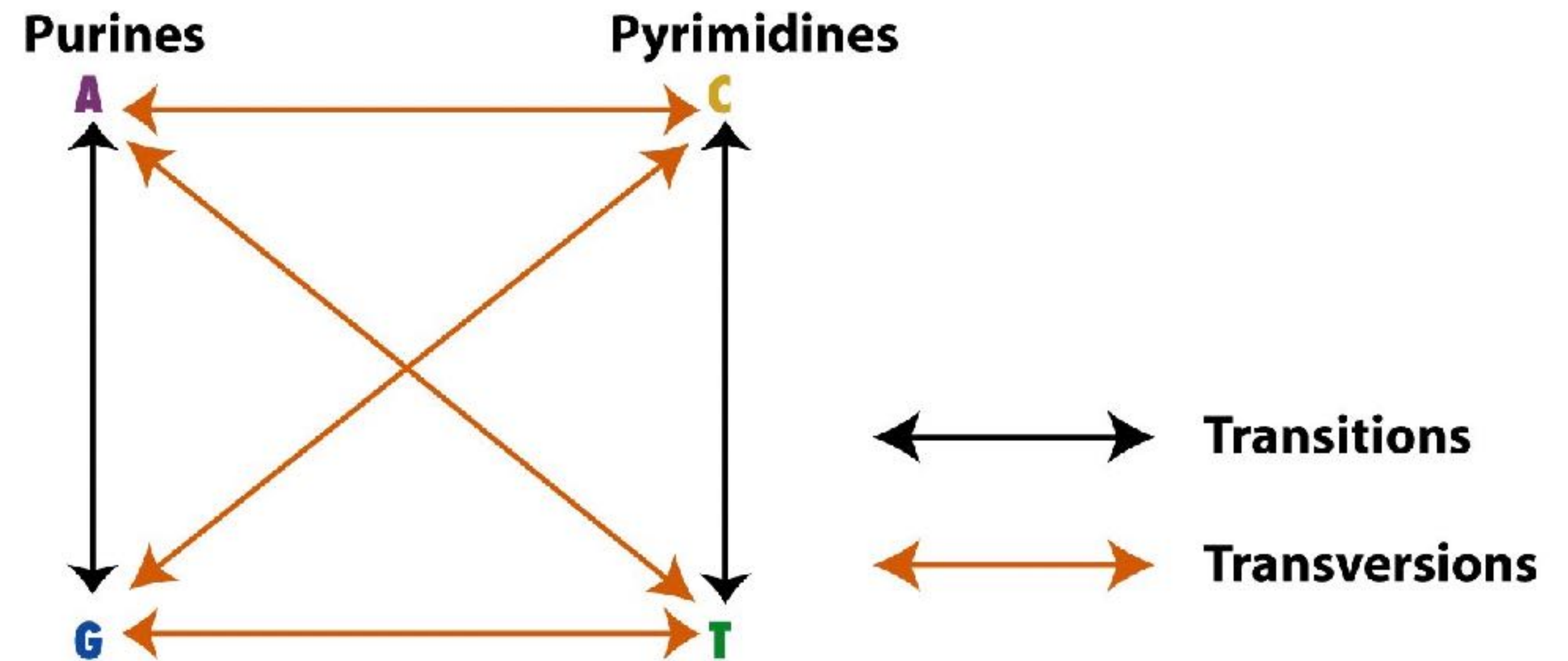


Figure 5-4 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Tranzycje i transwersje

- Dlaczego tranzycje są częstsze?
 - Wyjaśnienia selekcyjne (tranzycje rzadziej zmieniają aminokwas i częściej są neutralne)
- Ale:
 - tranzycje są częstsze też w genach rRNA, pseudogenach i obszarach niekodujących
 - tranzycje są częstsze w pozycjach 4-krotnie zdegenerowanych (kodony typu CUN – Leu)

Tranzycje i transwersje

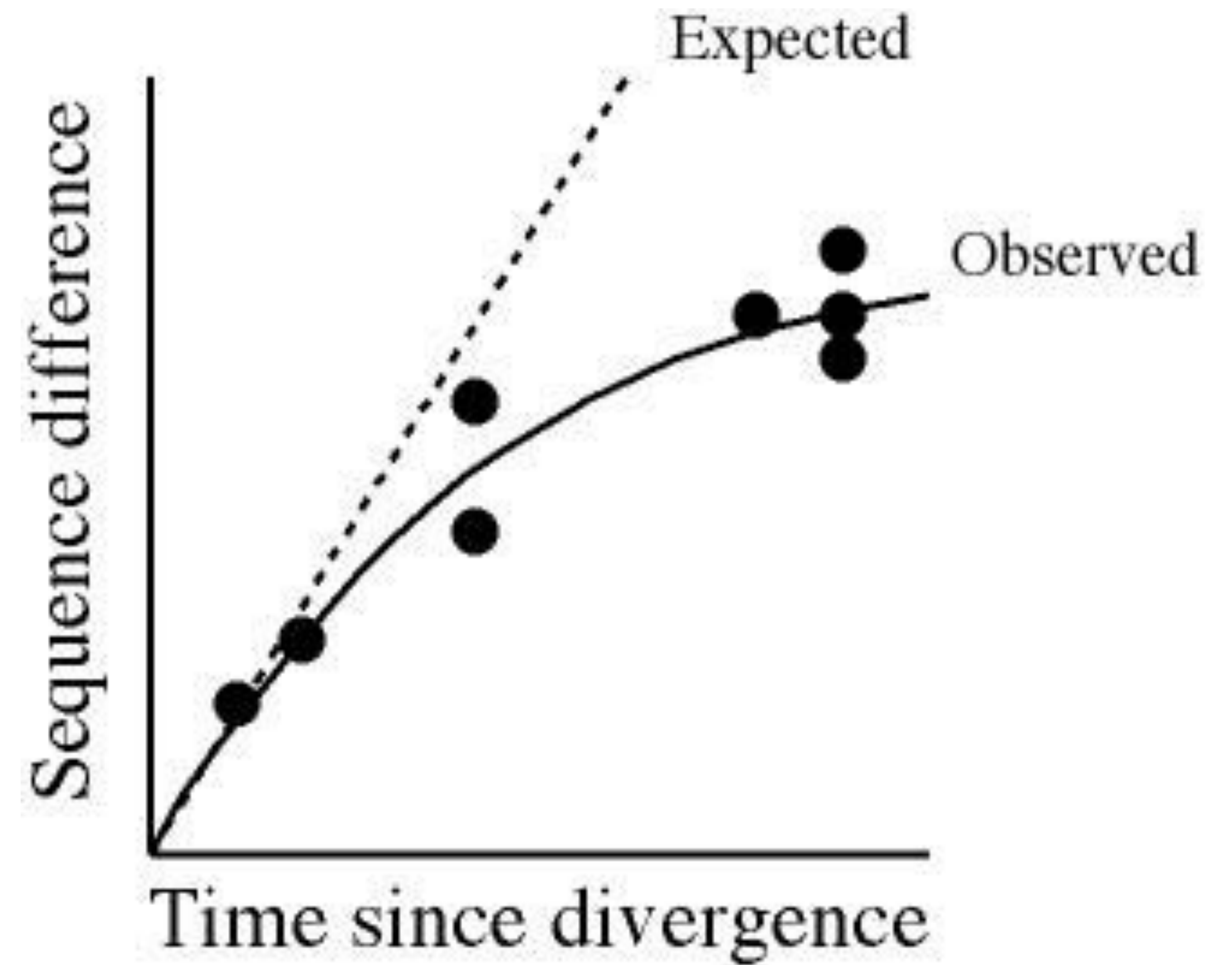
- Dlaczego tranzycje są częstsze?
 - Wyjaśnienia mechanistyczne – mechanizmy powstawania i naprawy mutacji
 - Tranzycje powstają w wyniku m. in.:
 - przejść tautomerycznych
 - deaminacji (np. oksydacyjnej)
 - Tranzycje w mniejszym stopniu zaburzają strukturę podwójnej helisy podczas replikacji
 - mniejsza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

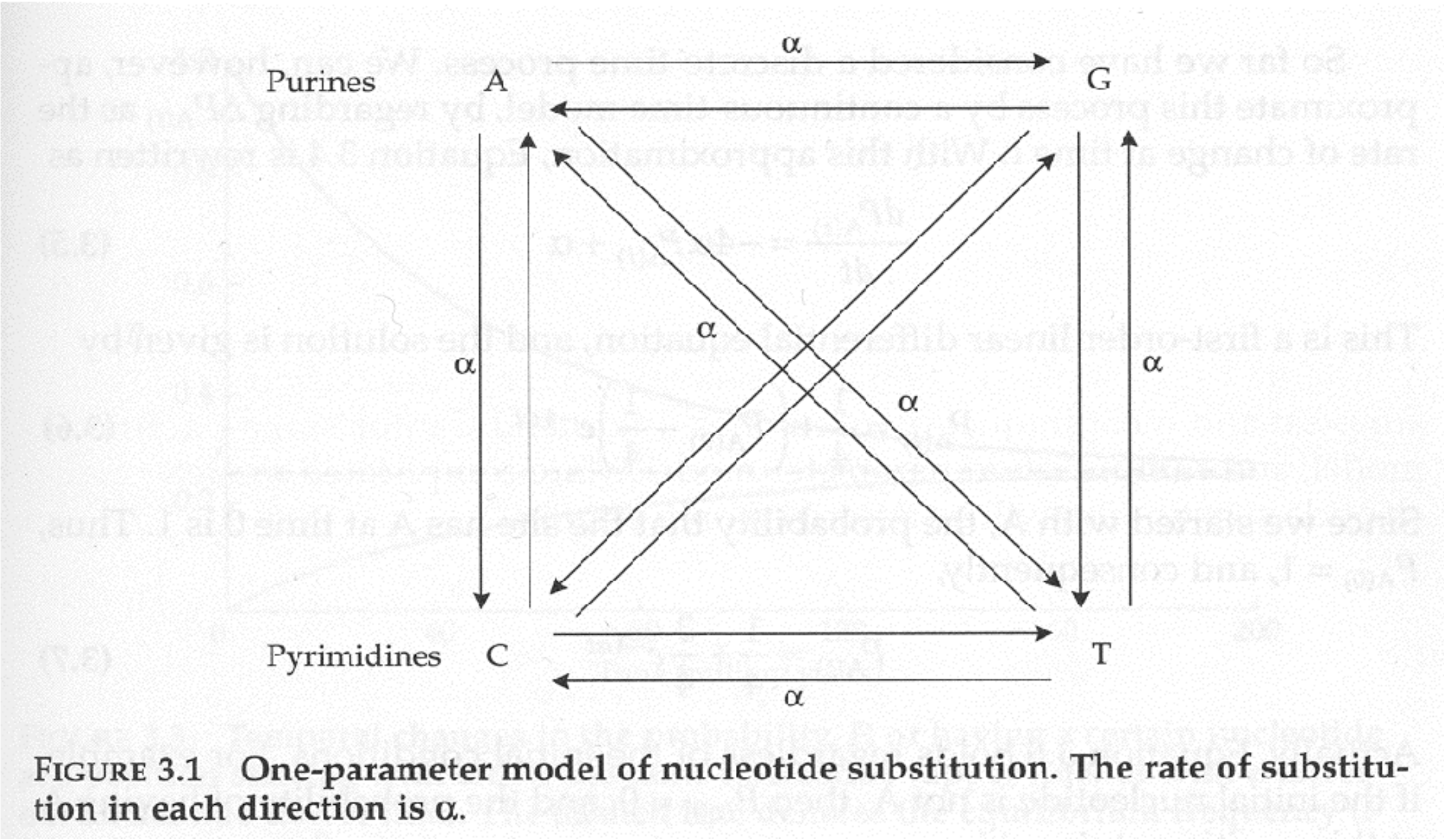
ACGGTGC
↓ ↓
C A
↓ ↓
GCGGTGA



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



	A	C	G	T
A	1-3 α	α	α	α
C	α	1-3 α	α	α
G	α	α	1-3 α	α
T	α	α	α	1-3 α

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

Inne modele

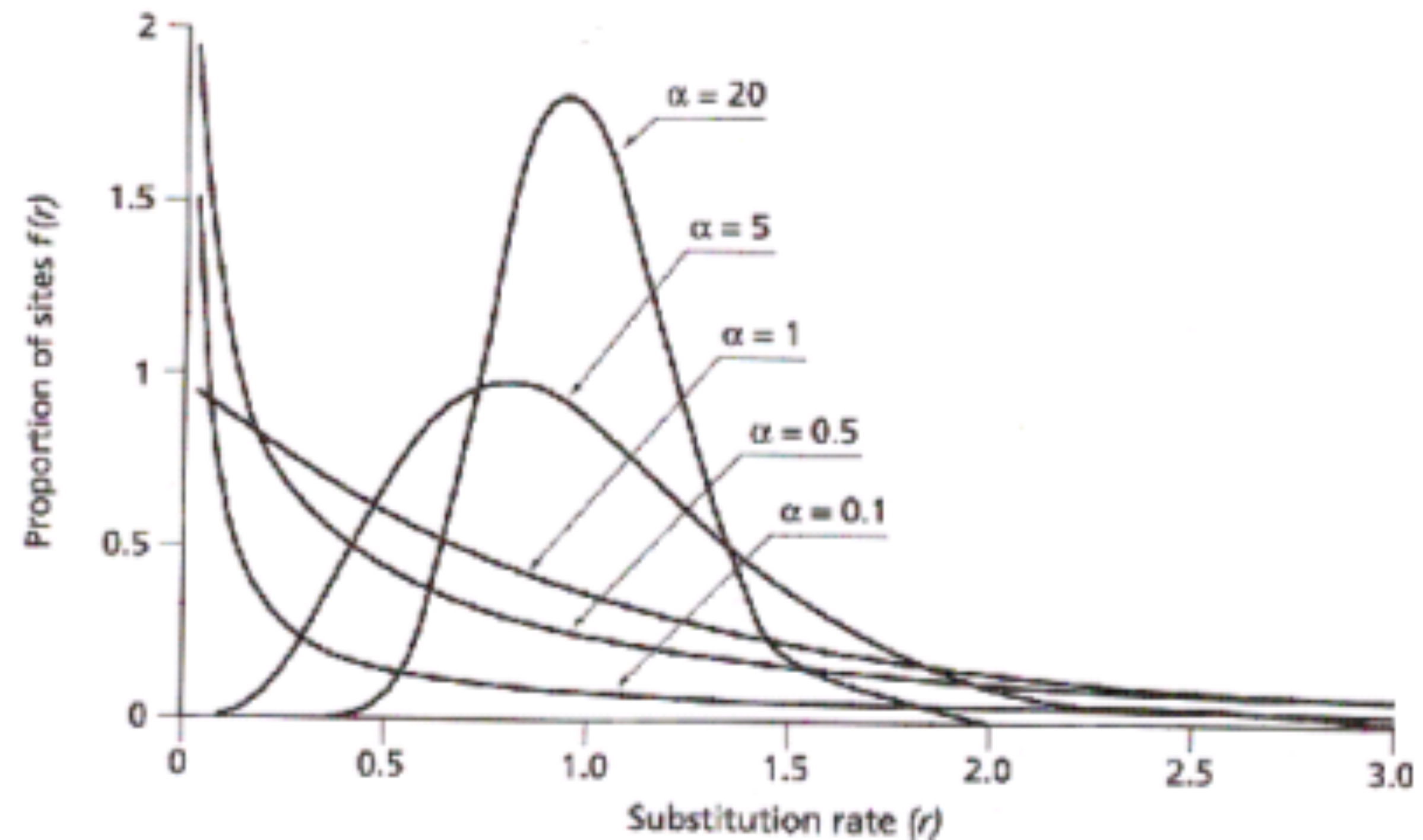
- Kimura (K80, dwuparametrowy) - różne prawdopodobieństwo tranzycji i transwersji
- Felsenstein (F81), Hasegawa-Kishino-Yano (HKY85) - różne częstości nukleotydów (F81) + różne prawd. tranzycji i transwersji (HKY85)
- GTR (General Time Reversible, Tavaré '86)

Model GTR

- Różne prawdopodobieństwo każdej substytucji (ale symetrycznie, czyli np. $A \rightarrow T = T \rightarrow A$) - 6 parametrów
- Różne częstości nukleotydów - 4 parametry

Rozkład gamma

- Proste modele zakładają jednakowe prawdopodobieństwo zmiany w każdej pozycji - nierealistyczne
- Rozkład prawdopodobieństw zmian w różnych pozycjach – rozkład gamma

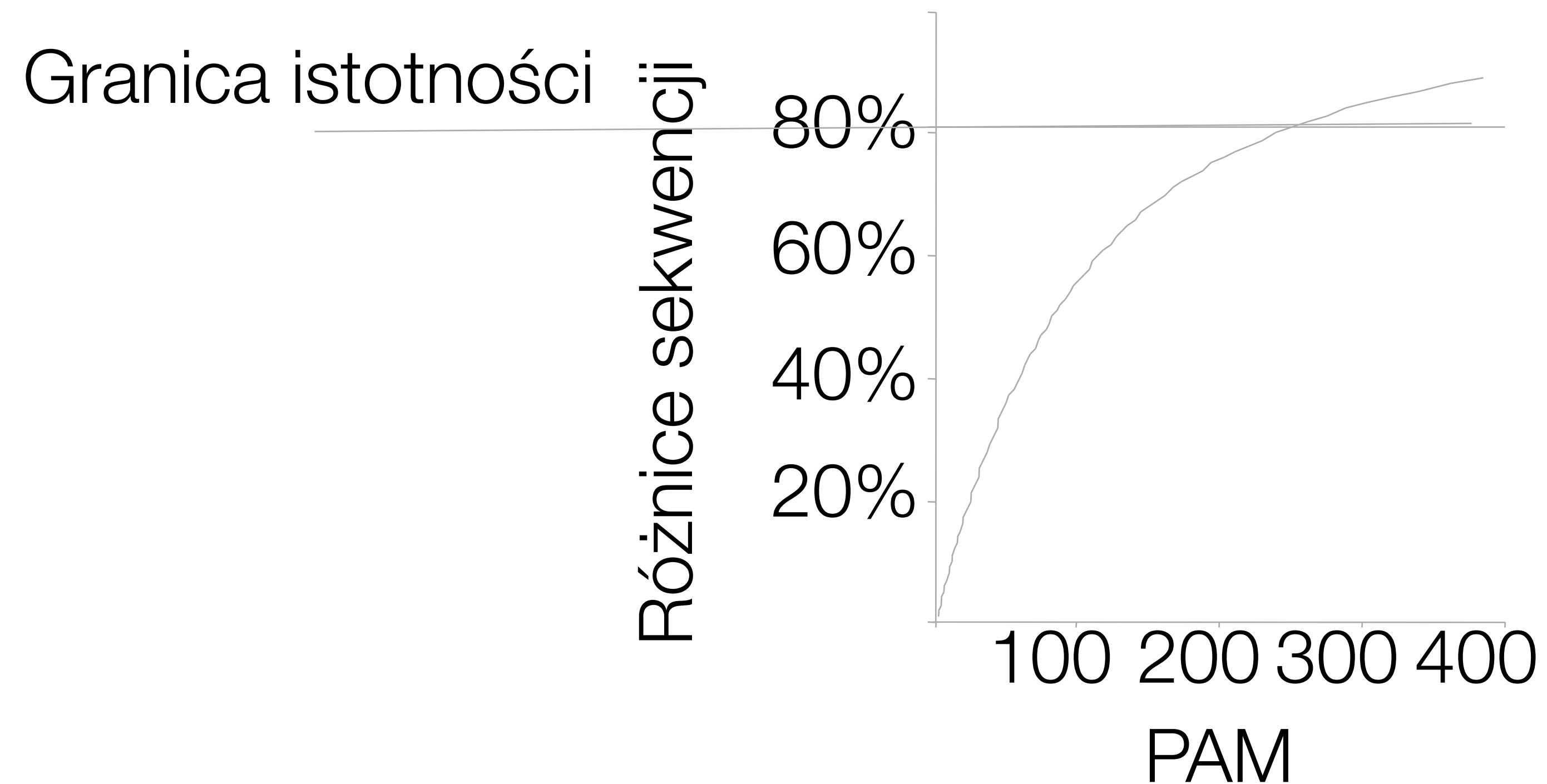


Ewolucja sekwencji aminokwasowych

- 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

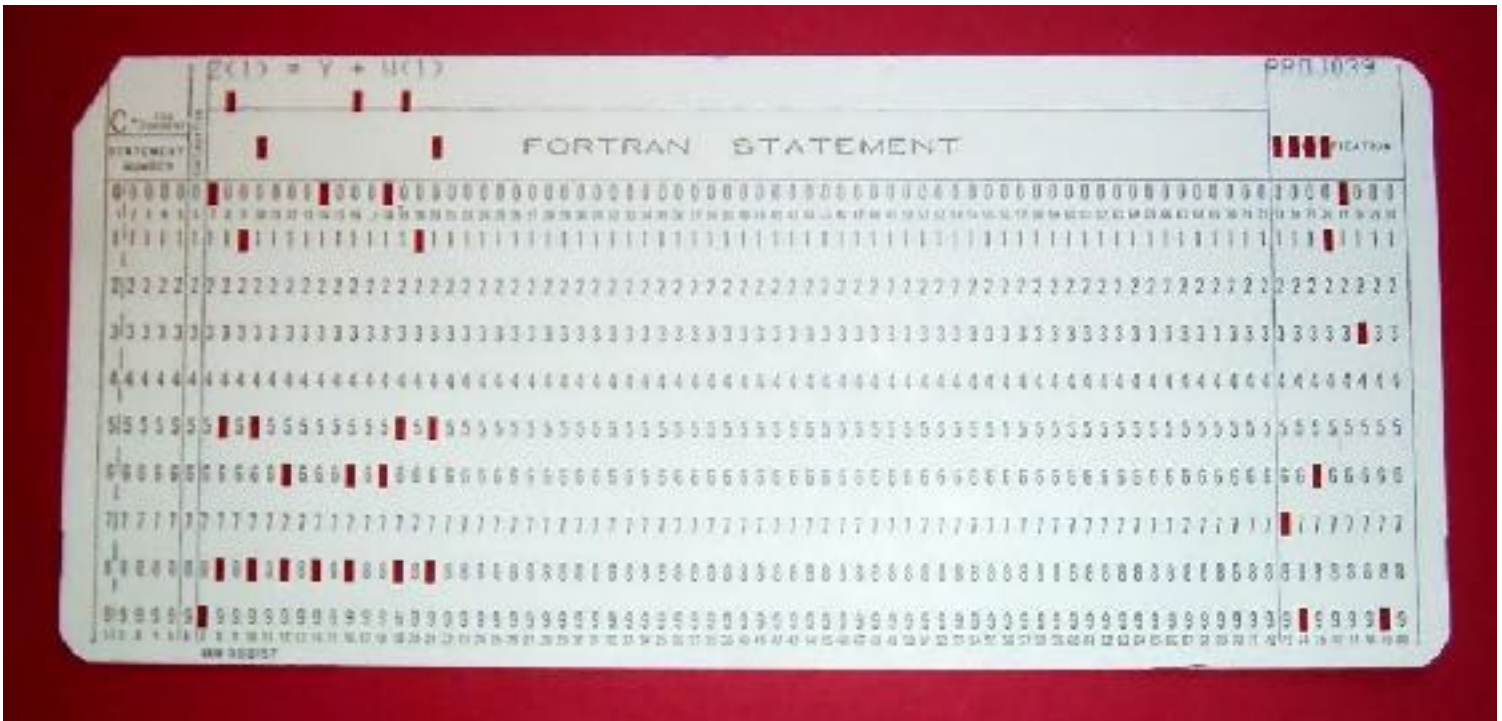
PAM - utrwalone mutacje punktowe/100 pozycji (od *Point Accepted Mutation*)



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)

	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	5															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	-3	-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	



Margaret O. Dayhoff (1925-1983)

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	-1	0	3	

BLOSUM 62

Mutacje na poziomie kodu

- Mutacje mogą:
 - zmienić kodon na inny, ale kodujący ten sam aminokwas
 - mutacje synonimiczne
 - zmienić kodon na kodujący inny aminokwas
 - mutacje niesynonimiczne
 - zmienić kodon na kodon STOP
 - mutacje nonsense
 - spowodować zmianę fazy odczytu
 - wpłynąć na ekspresję genu

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

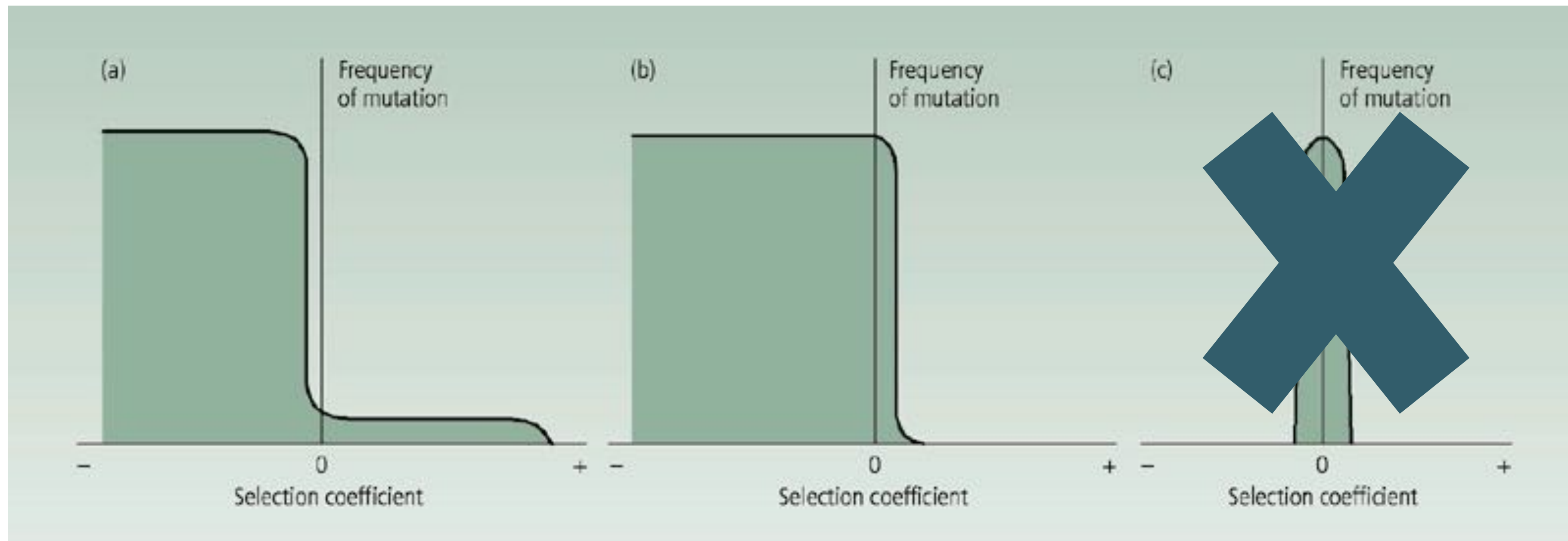
Mutacje i dobór

- 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
 - większość utrwalonych mutacji jest korzystna
 - mutacje neutralne są rzadkie (nie częstsze od korzystnych)
- Neutralizm
 - większość mutacji jest niekorzystna lub neutralna
 - większość utrwalonych mutacji jest neutralna
 - 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

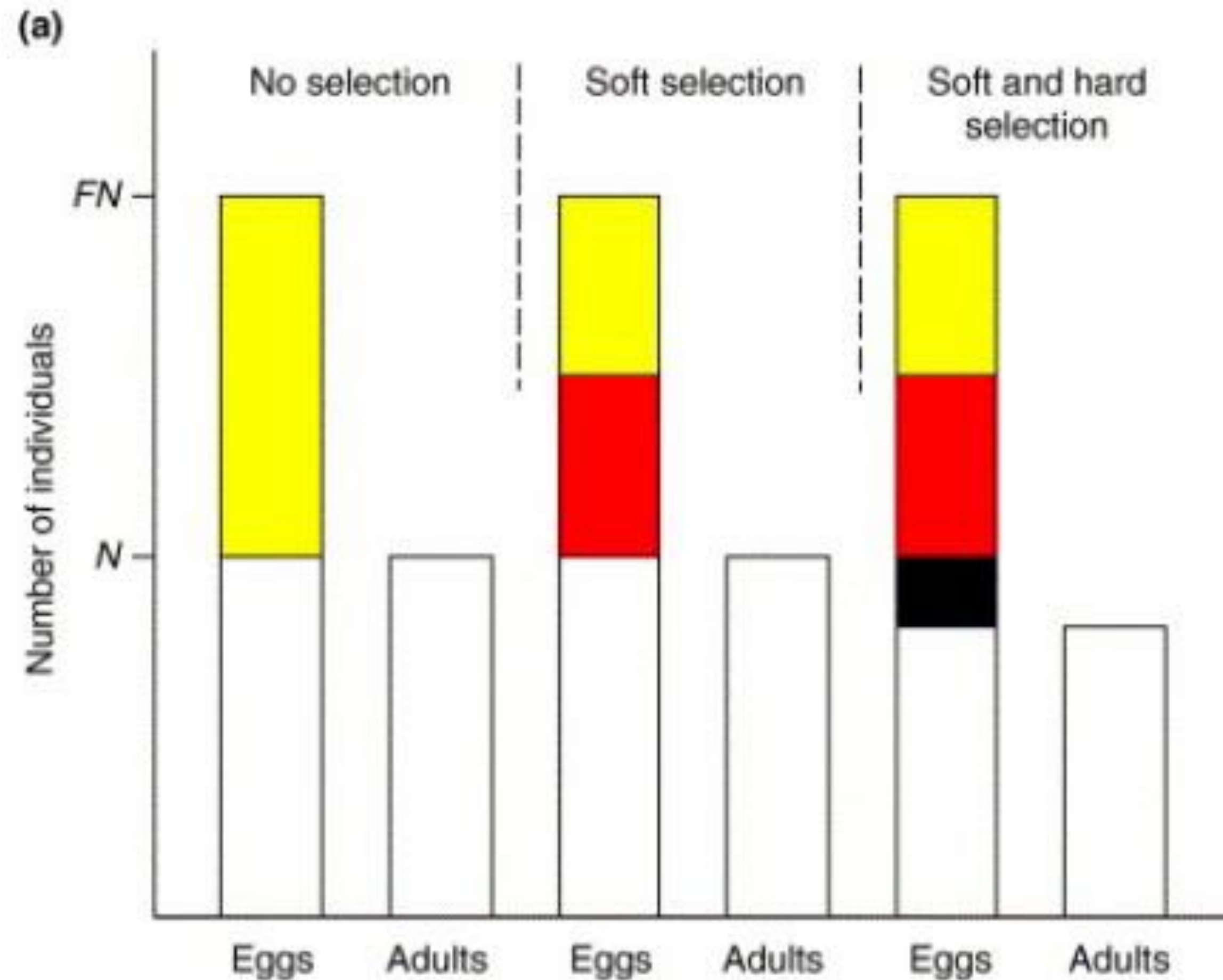
Tempo zmian

- Pojęcie obciążenia genetycznego – spadek średniego dostosowania populacji na skutek działania doboru
- Im silniej selekcjonowany nowy, korzystny allel, tym większy koszt u osobników go pozbawionych
- Silna selekcja to duży koszt dla populacji
- Haldane (1957) szacował maksymalne tempo ewolucji rzędu 1 mutacja na 300 pokoleń

Tempo zmian

- Rzeczywiste tempo zmian jest wyższe niż oszacowane przez Haldane'a
- Ale...
- Obliczenia Haldane'a i Kimury oparte były na założeniu tzw. twardego doboru
 - twardy dobór– zwiększona śmiertelność słabiej przystosowanych osobników, ponad typową ("ekologiczną") śmiertelność w populacji
 - miękki dobór– działa w ramach stałej ("ekologicznej") śmiertelności,
- Tylko dobór twardy znacząco ogranicza tempo mutacji

Dobór twardy i miękki



Dobór twardy –
więcej osobników
ginie (lub nie wydaje
potomstwa)

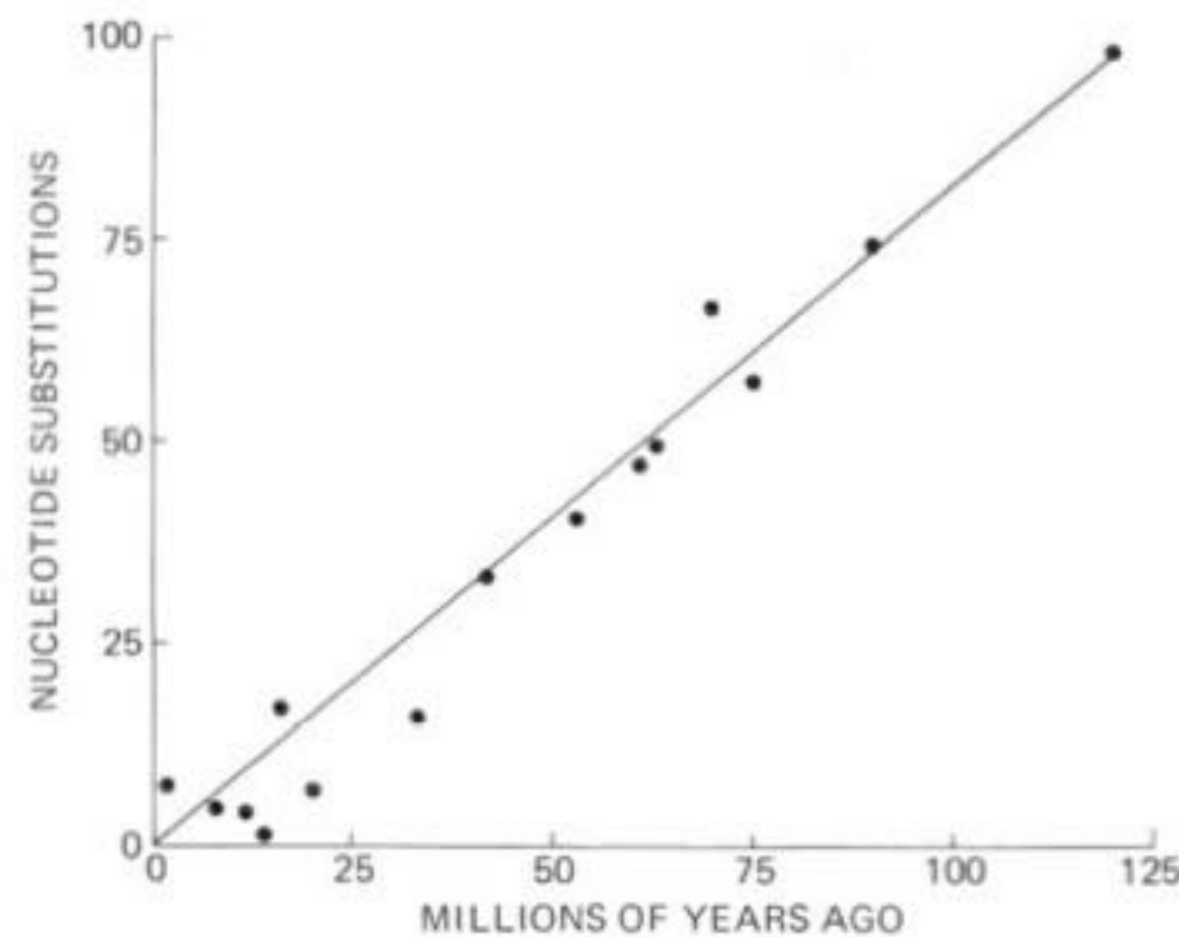
Dobór miękki – ginie
tyle samo
osobników, dobór
wpływa tylko na to,
które giną

Tempo zmian

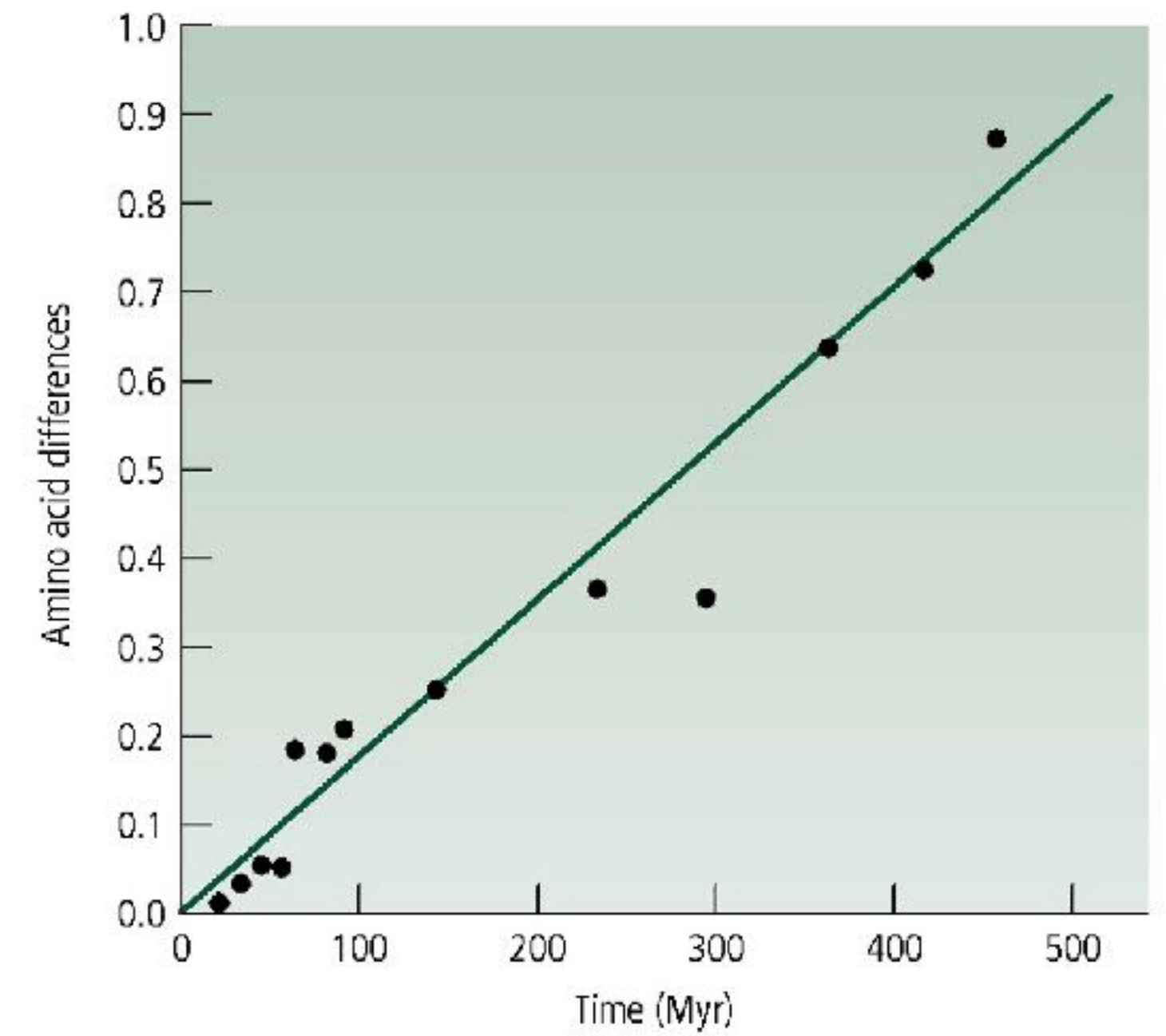
- Obliczenia Haldane'a i Kimury oparte były na założeniu tzw. twardego doboru
 - twardy dobór– zwiększona śmiertelność słabiej przystosowanych osobników, ponad typową (“ekologiczną”) śmiertelność w populacji
 - miękki dobór– działa w ramach stałej (“ekologicznej”) śmiertelności,
- Tylko dobór twardy znacząco ogranicza tempo mutacji
- Dobór równoważący może utrzymywać zróżnicowanie i zwiększać tempo utrwalania mutacji

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

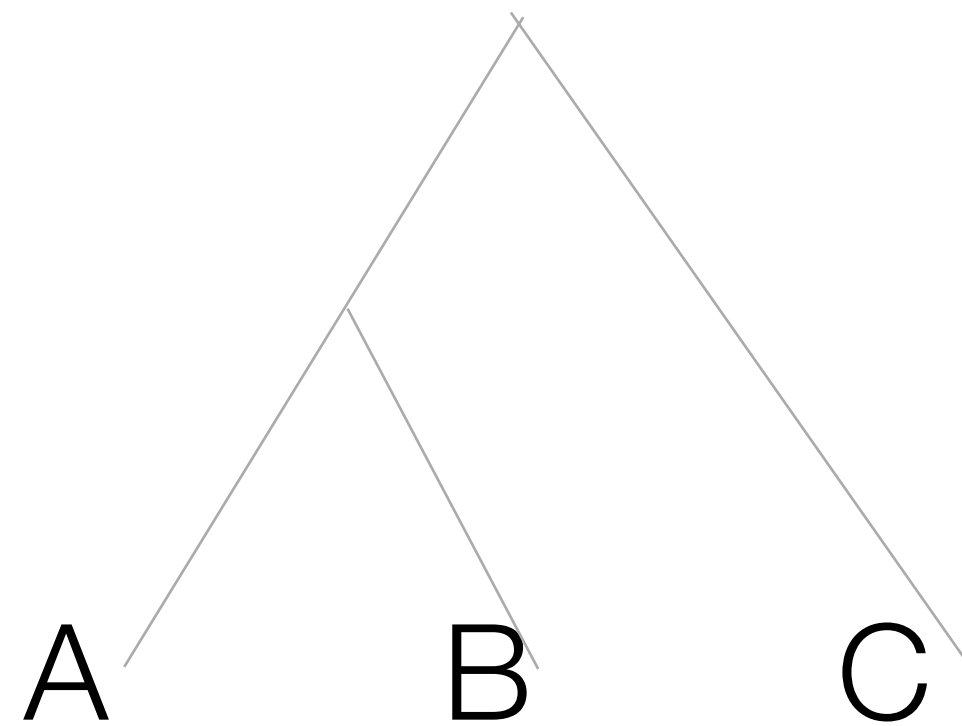
Różnice sekwencji globin kręgowców

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

Zegar molekularny

- Jest konsekwencją neutralnego modelu ewolucji
- Tempo akumulacji zmian w danej sekwencji jest stałe
 - ale różne dla różnych sekwencji
- Weryfikacja – test względnego tempa



$$K_{AC} - K_{BC} = 0$$

- W rzeczywistości testuje stałość tempa pomiędzy gałęziami, ale nie w czasie

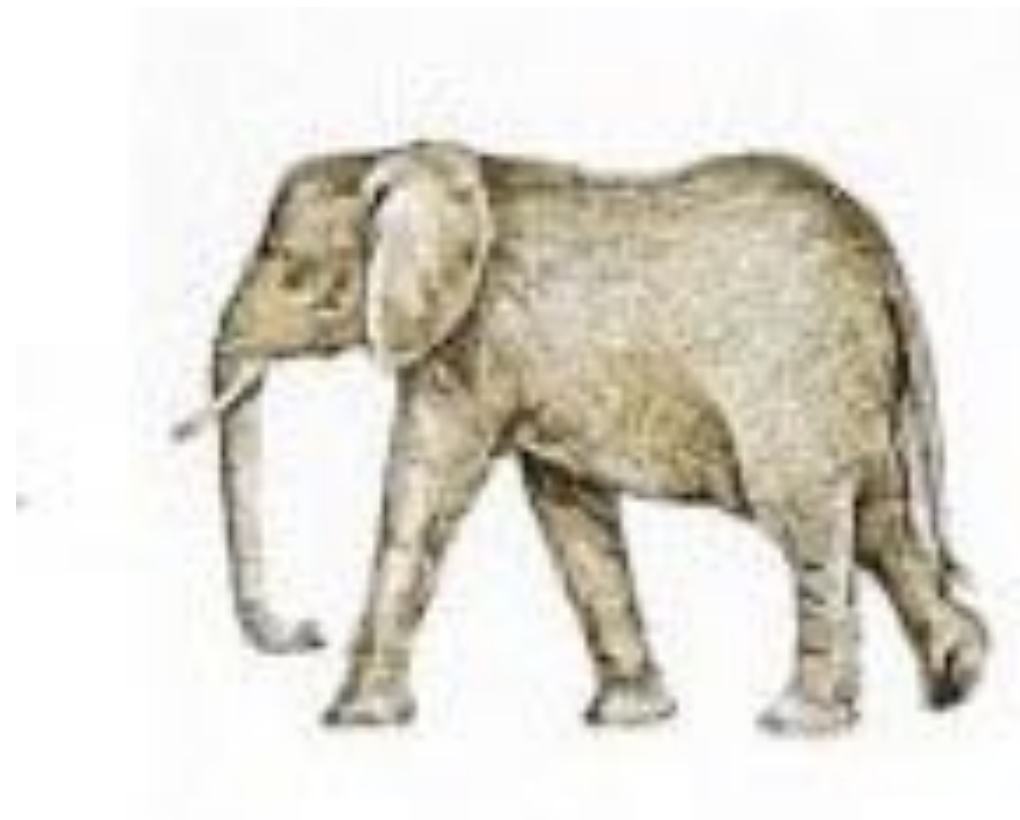
Zegar molekularny - problem

- W modelu neutralnym tempo utrwalania mutacji:
- 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

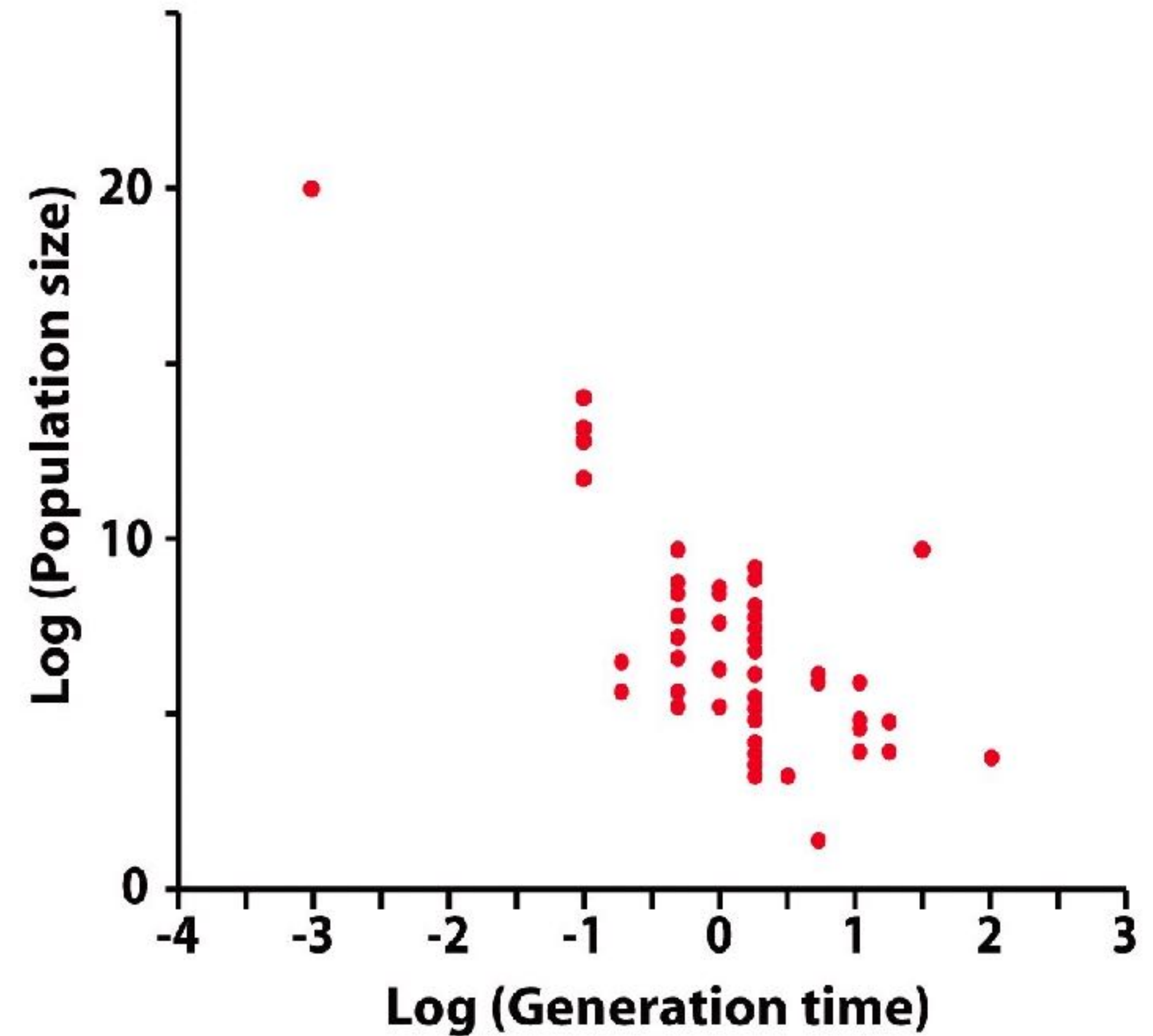
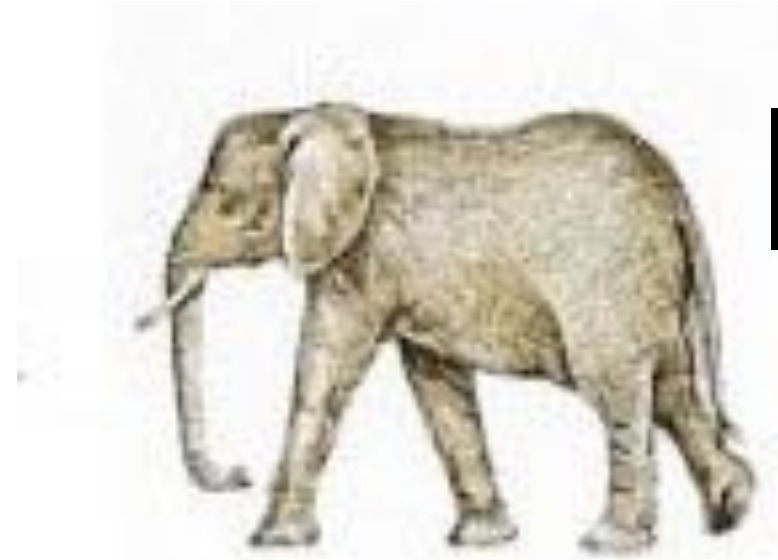


Figure 7-22a Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

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)

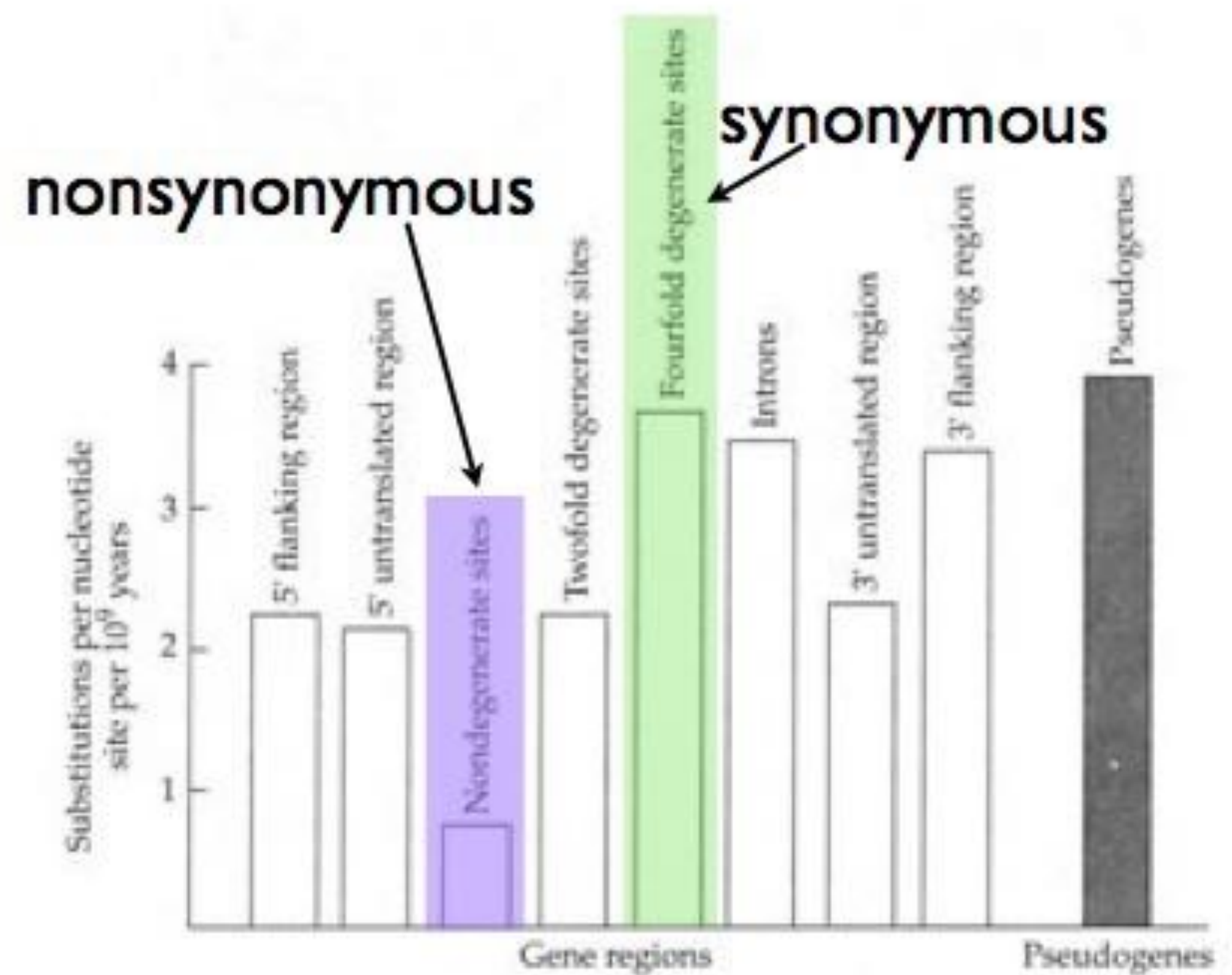
Efekty czasu generacji i wielkości populacji się znosząc, dając
stałe tempo w czasie (Ohta & Kimura, 1971).

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!**



Degeneracja w kodzie

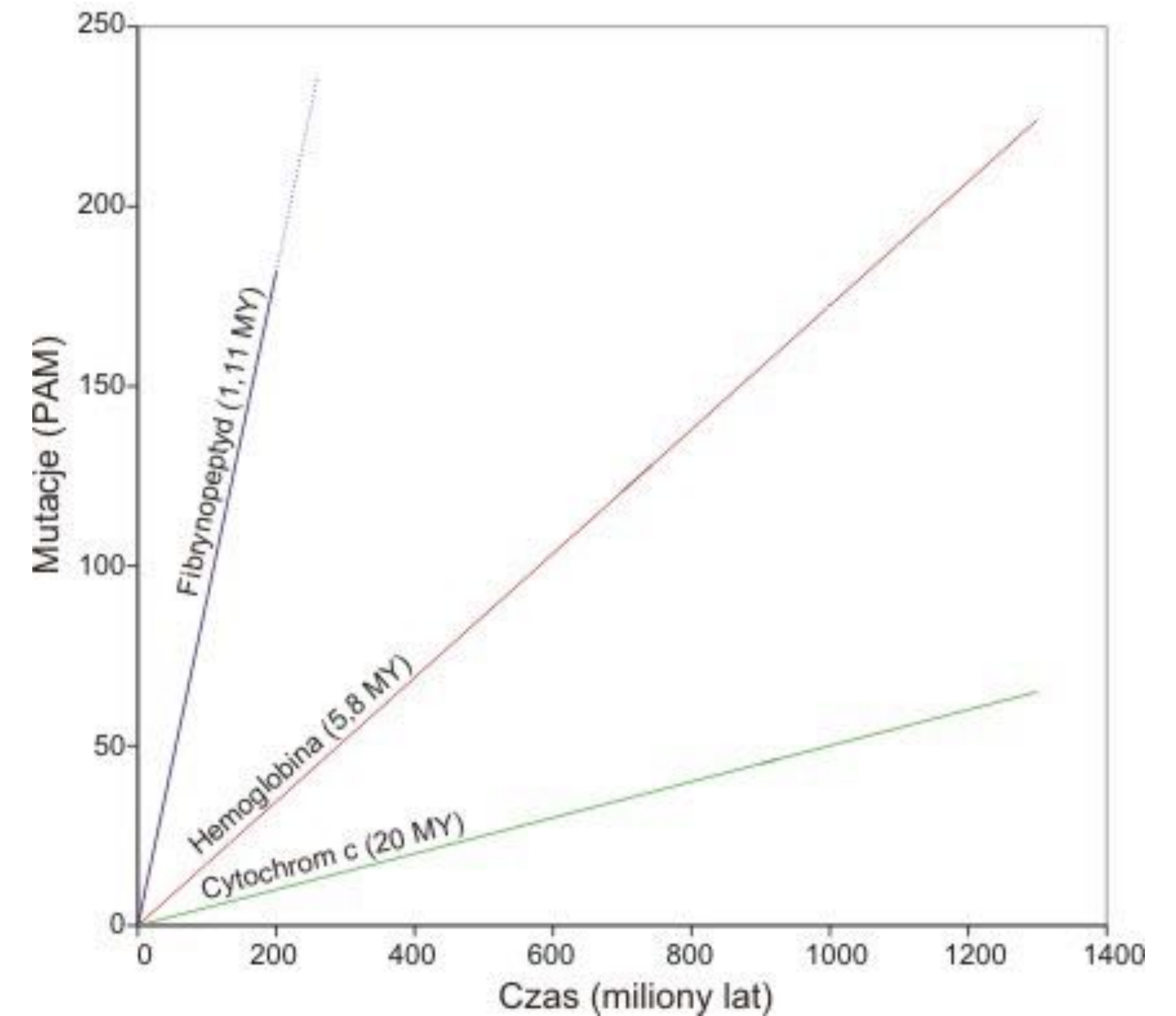
		Second Letter					
		T	C	A	G		
First Letter	T	TTT } Phe TTC } TTA } Leu TTG }	TCT } TCC } Ser TCA } TCG }	TAT } Tyr TAC } TAA Stop TAG Stop	TGT } Cys TGC } TGA Stop TGG Trp	T	C
	C	CTT } CTC } Leu CTA } CTG }	CCT } CCC } Pro CCA } CCG }	CAT } His CAC } CAA Gln CAG }	CGT } CGC } Arg CGA } CGG }	T	C
	A	ATT } ATC } Ile ATA } ATG Met	ACT } ACC } Thr ACA } ACG }	AAT } Asn AAC } AAA Lys AAG }	AGT } Ser AGC } AGA Arg AGG }	T	C
	G	GTT } GTC } Val GTA } GTG }	GCT } GCC } Ala GCA } GCG }	GAT } Asp GAC } GAA Glu GAG }	GGT } GGC } Gly GGA } GGG }	T	C
						A	G

miejsce 4-krotnie zdegenerowane

miejsce 2-krotnie zdegenerowane

Tempo zmian

- Białka zaangażowane w podstawowe funkcje komórki ewoluują wolniej.
- W sekwencji białka obszary kluczowe dla funkcji ewoluują wolniej.
 - Jednostka: PAM/10⁸ lat
 - Jednostka czasu ewolucyjnego: ile lat (w milionach, 10⁶) potrzeba do utrwalenia 1 mutacji/100 aa (1 PAM)



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^{1,*}, Yichen Zheng¹, Nicholas Price¹, Ricardo B.R. Azevedo¹, Rebecca A. Zufall¹, and Eran Elhaik²

¹Department of Biology and Biochemistry, University of Houston

²Department of Mental Health, Johns Hopkins University Bloomberg School of Public Health

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

Accepted: February 16, 2013

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

- Dane molekularne, zwłaszcza genomowe, pozwoliły ocenić zgodność modelu neutralnego z obserwacją zmienności sekwencji
- Kimura: 1968 – nie były wtedy znane metody sekwencjonowania DNA!

Status neutralizmu

- Smith & Eyre-Walker 2002 – 45% podstawień aminokwasowych w ewolucji *Drosophila* sp. utwalonych przez dobór dodatni
- Andolfatto 2005 – pomiędzy *D. melanogaster* i *D. simulans* dobór dodatni odpowiada za utrwalenie:
 - 20% podstawień w DNA w intronach i obszarach międzygenowych
 - 60% podstawień w DNA w sekwencjach UTR

Status neutralizmu

- Głównym i nieprzemijającym osiągnięciem jest stworzenie matematycznego opisu współdziałania dryfu i doboru naturalnego (dodatniego i oczyszczającego) w ewolucji molekularnej
- Dzięki tym modelom opracowano testy poszukujące śladów doboru w sekwencjach (model neutralny jako hipoteza zerowa)
- Istnieje znacząca liczba pozycji i sekwencji ewoluujących według modelu neutralnego
 - można dobrać sekwencje tak, by uzyskać zegar molekularny

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 (ω) jest miarą działania doboru

Badanie doboru

- Wartość ω rzadko przekracza 1 dla całej sekwencji (wyjątek np. geny MHC)
- Średnia wartość ω w porównaniach między naczelnymi a gryzoniami wynosi 0,2, między człowiekiem a szympansem 0,4
- Odchylenie ω od średniej dla konkretnego genu w konkretnej linii ewolucyjnej może świadczyć o działaniu doboru
- W sekwencji mogą występować obszary o różnej wartości ω , wskazując na działanie doboru na poszczególne regiony a nawet pozycje aminokwasowe w białku

Badanie doboru II

- Porównanie zmian synonimicznych i niesynonimicznych w obrębie populacji danego gatunku i pomiędzy gatunkami.

Test McDonald-Kreitmana

- Stosunek mutacji synonimicznych do niesynonimicznych w obrębie populacji vs. taki sam stosunek dla różnic między gatunkami
- Jeżeli zmiany są neutralne, wówczas stosunek ten powinien być w obu przypadkach taki sam

Przykład: gen ADH u trzech gatunków *Drosophila*

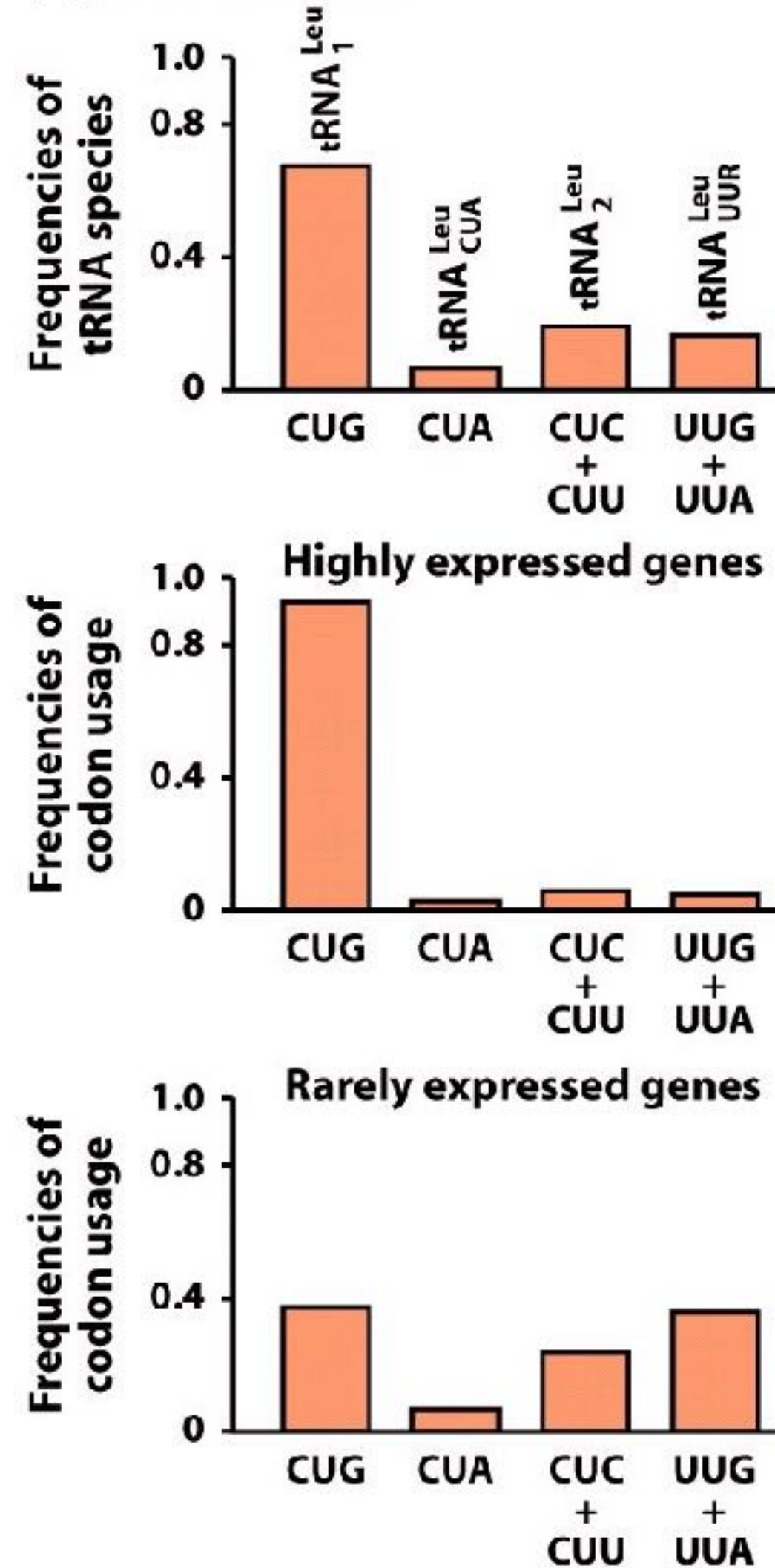
	synonimiczne	niesynonimiczne	stosunek
wewnątrzpopulacyjne	42	2	~0,05
międzygatunkowe	17	7	~0,41

Wniosek: zmiany niesynonimiczne są szybko utrwalane w specjacji – nie są neutralne

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

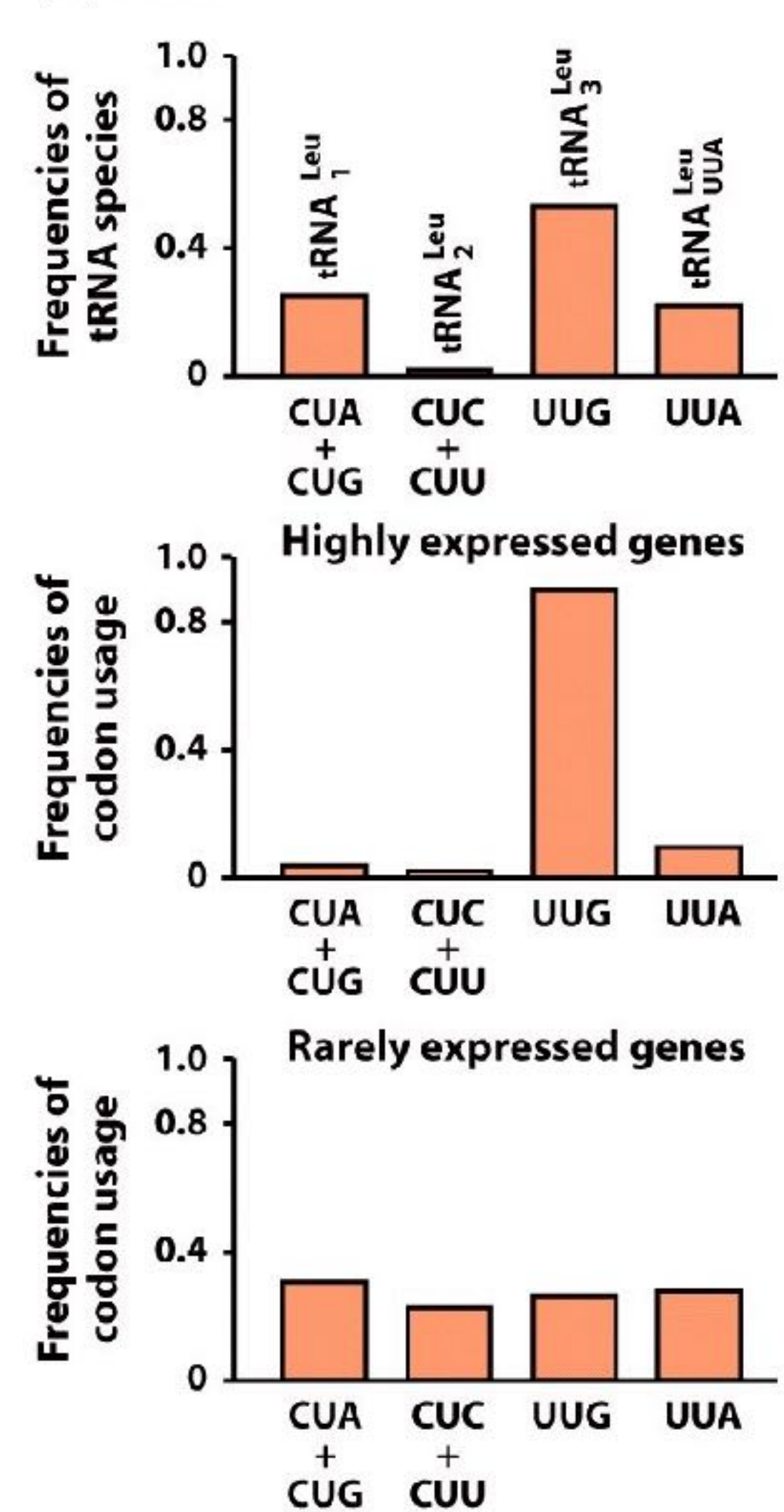


Figure 7-24 Evolutionary Analysis, 4/e
© 2007 Pearson Prentice Hall, Inc.

Czy zmiany synonimiczne są neutralne?

RESEARCH ARTICLE

AMERICAN JOURNAL OF
medical genetics PART
A

A Synonymous Mutation in *TCOF1* Causes Treacher Collins Syndrome Due to Mis-Splicing of a Constitutive Exon

D. Macaya,¹ S.H. Katsanis,¹ T.W. Hefferon,² S. Audlin,¹ N.J. Mendelsohn,³ J. Roggenbuck,³ and G.R. Cutting^{1*}

¹DNA Diagnostic Laboratory, Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland

²Genome Technology Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland

³Children's Hospitals & Clinics of Minnesota, Minneapolis, Minnesota

Received 26 September 2008; Accepted 25 February 2009

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.

The *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). 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 (M.M.G.); jmoh@snu.ac.kr (J.M.O.); kimchi@cber.fda.gov (C.K.-S.)

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