

Metagenomika, filogenetyka i ewolucja molekularna drobnoustrojów

– dotychczasowe i aktualne wyzwania matematyki

Dr inż. Marcin Olszewski

Katedra Biotechnologii Molekularnej i Mikrobiologii
Wydział Chemiczny
Politechnika Gdańska

Gdańsk 2014

Program wykładu

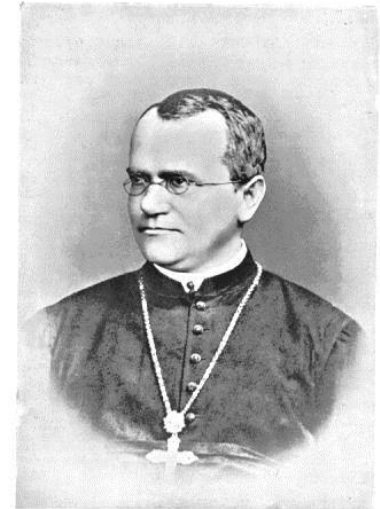
1. Zagadnienia wstępne
2. Metagenom
3. Filogenetyka molekularna
4. Ewolucja molekularna

Elementy graficzne nie podpisane pochodzą z książki „Genomy” T. A. Brown 2009.

Genetyka

Genetyka – nauka zajmująca się badaniem zjawisk **dziedziczności i zmienności organizmów**.

1866 r. - Grzegorz Mendel opublikował prawa dziedziczenia cech w pracy pt. „*Badania nad mieszańcami roślin*”. znany jest obecnie jako człowiek, który odkrył **podstawowe zasady dziedziczności**. Za życia był nieznanym austriackim mnichem i naukowcem z zamiłowania, a jego badania nie zyskały uznania ówczesnego świata nauki.



Podstawowe terminy używane w genetyce

Gen – odcinek DNA kodujący jeden polipeptyd albo rodzinę podobnych polipeptydów lub jeden rodzaj rRNA lub tRNA.

Niewłaściwe definicje genu:

- jeden gen = jedno białko
- jeden gen = jedna cecha

Genotyp – zestaw genów danego osobnika warunkujący jego właściwości dziedziczne.

Fenotyp – (gr. *phainomai* - przejawiać; *typos* - wzór, norma) zespół ujawnionych cech (morfologicznych, anatomicznych, fizjologicznych i biochemicznych) danego osobnika warunkowanych przez genotyp i czynniki środowiska.

Podstawowe terminy używane w genetyce

Genom - podstawowy komplet informacji genetycznej, pojęcie to oznacza haploidalny zespół chromosomów.

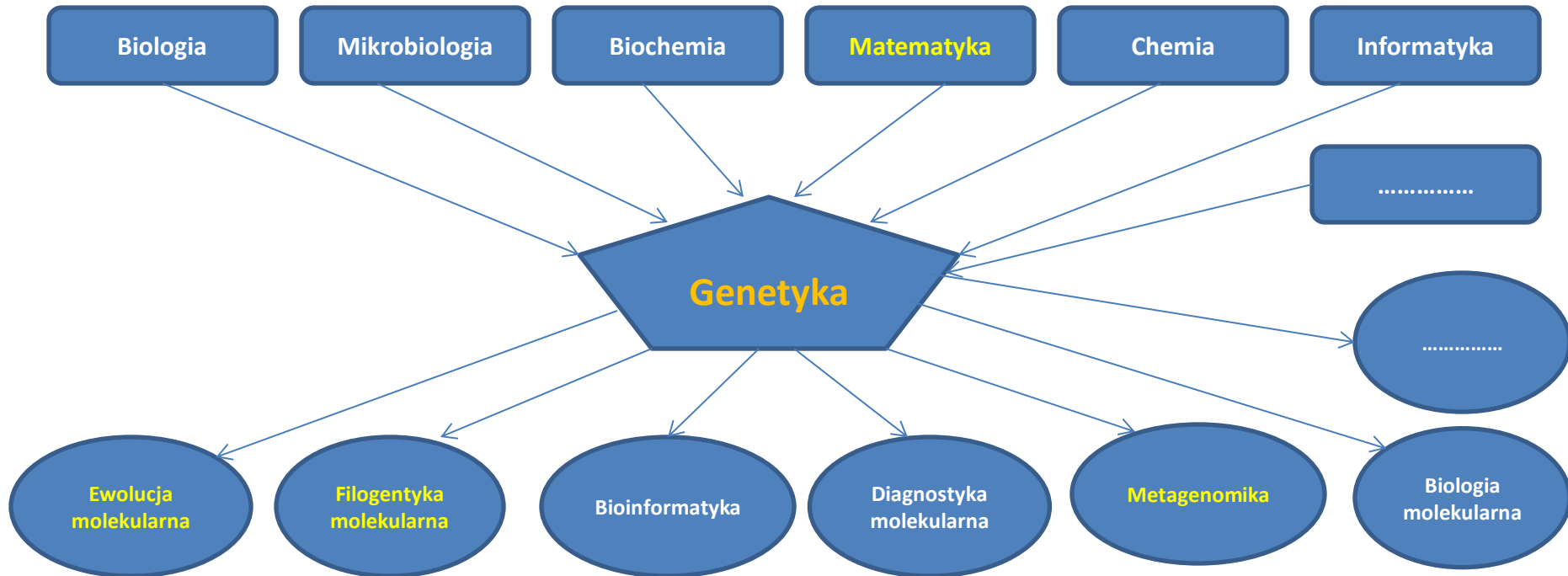
Genomy wirusów i Prokariota są **haploidalne**, natomiast większość Eukariota to organizmy **diploidalne**.

Haploid (z gr. *απλος* - niezłożony, pojedynczy) komórka zawierająca po jednym chromosomie każdego typu. Komórki płciowe (gamety) są haploidalne.

Diploid (z gr. *διπλος* - podwójny) komórka zawierająca po dwa chromosomy danego typu. Komórki somatyczne większości roślin i zwierząt są diploidalne.

Powiązania genetyki

Genetyka korzysta z różnych naukowych dyscyplin sama dając podwaliny pod nowe dziedziny nauki



Wszystkie komórki pochodzą od prąkomórki

Prąkomórka = Archeon?

W czym wszystkie organizmy są do siebie podobne???

Rosną, mnożą się, przekształcają energię, ..., umierają.

We wszystkich organizmach:

- ✓ geny są „zmagazynowane” w cząsteczkach DNA
- ✓ geny są zapisane tym samym kodem
- ✓ geny są zbudowane z tych samych podjednostek
- ✓ replikowane w ten sam sposób, pozwalający organizmom rozmnażać się
- ✓ białka powstają z połączenia tych samych 20 aminokwasów w różnych kombinacjach

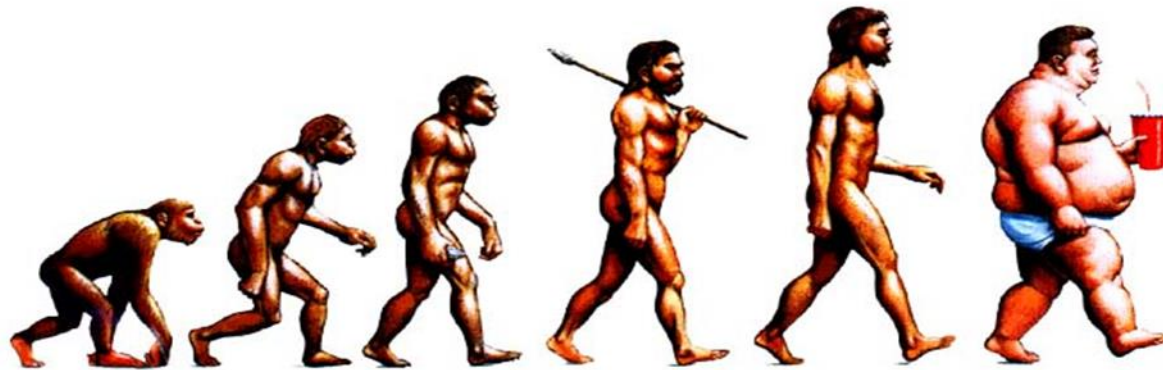


Wpływ matematyki na genetykę

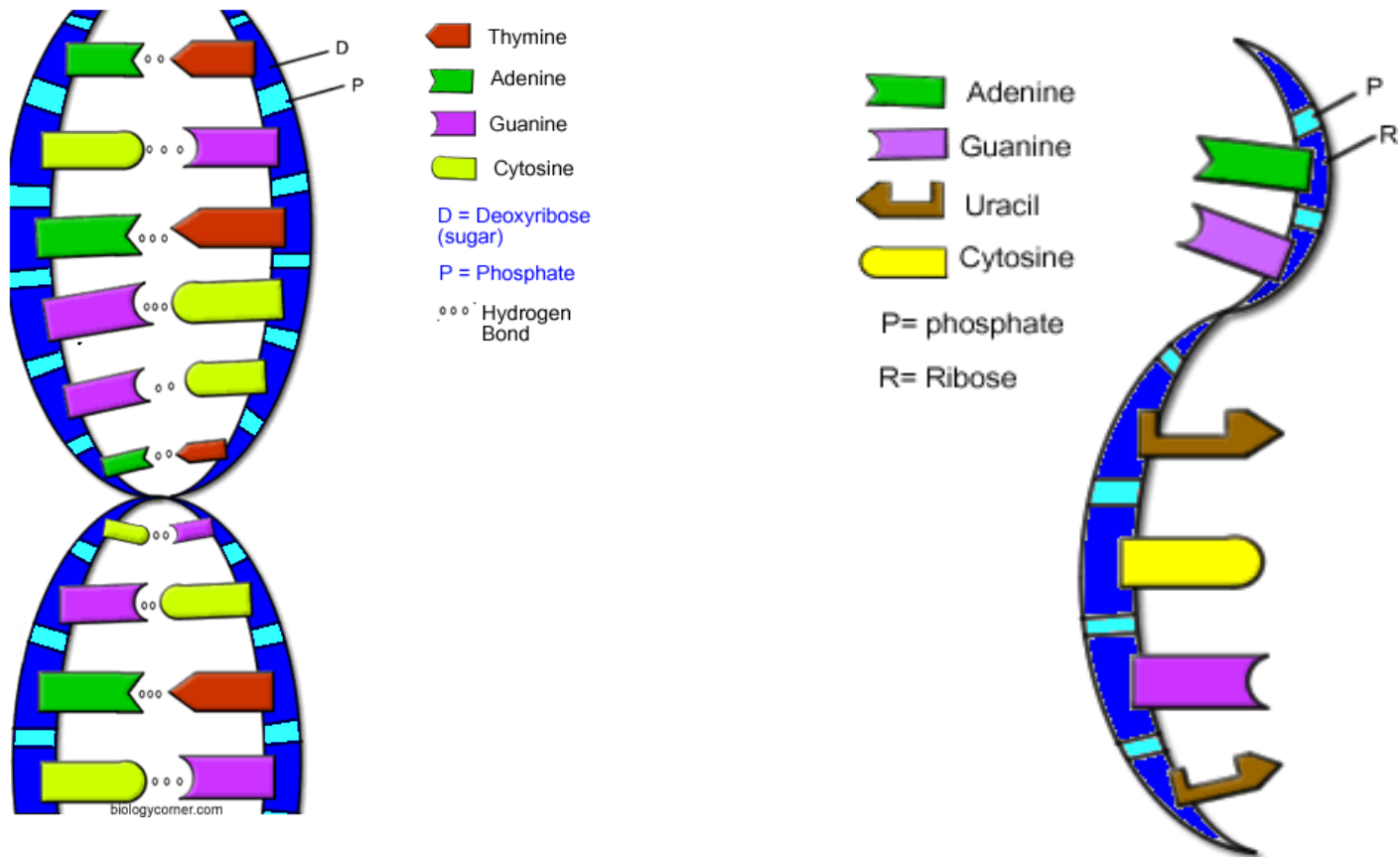
W klasycznej genetyce **rachunek prawdopodobieństwa** i **statystyka matematyczna**

okazują się niezbędne np. aby opisać:

- występowanie zjawisk masowych
- szanse na powstanie określonego fenotypu
- odległość między genami
- położenie genu na chromosomie
-

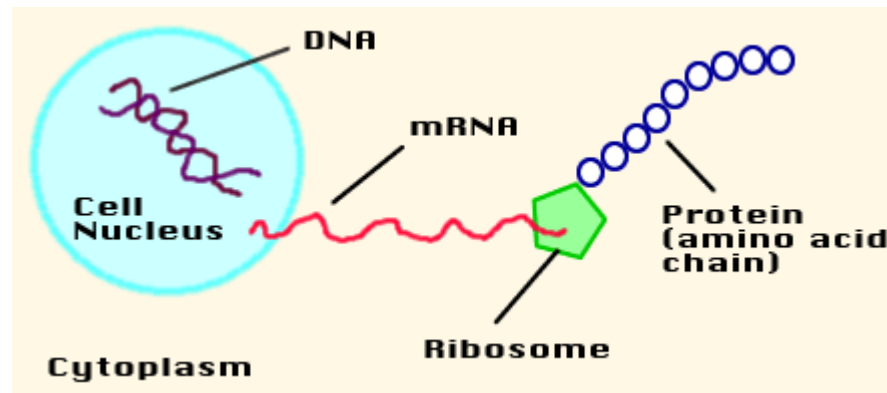


DNA a RNA



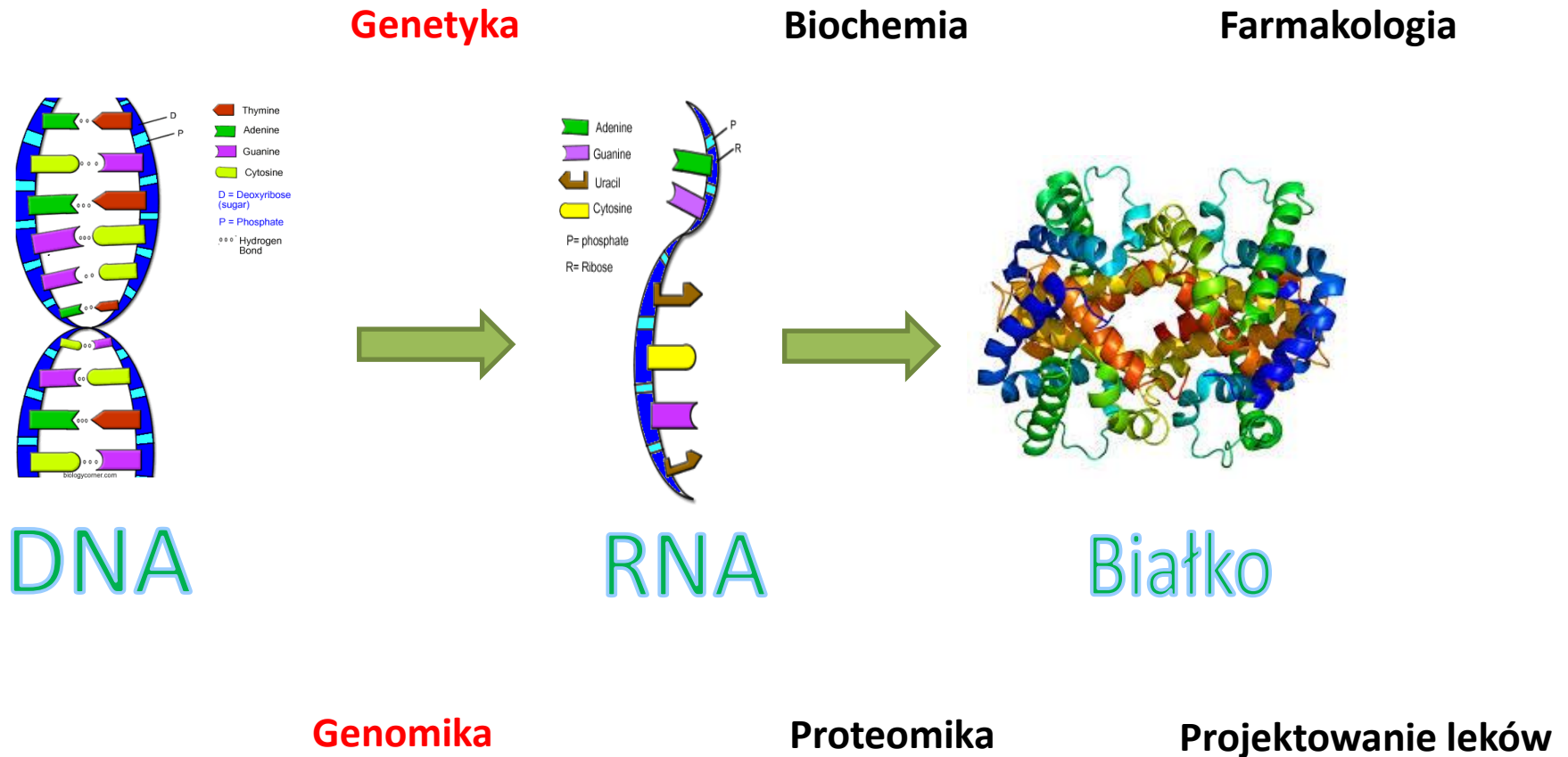
Źródło: www.biologycorner.com

Przepływ informacji – podstawowy dogmat



Źródło: www.biologycorner.com

Miejsce genetyki w świecie molekularnym



Źródło: www.biologycorner.com

Metagenom

1. Metagenom
2. Genom
3. Mikrobion



Metagenom

Genom człowieka o wielkości $2,84 \times 10^9$ pz zawiera około 23 tys. genów kodujących białka i RNA

Organizm człowieka zbudowany jest z ok. 10^{13} komórek somatycznych i rozrodczych, ale zasiedla go **10 razy większa** populacja komórek mikroorganizmów (10^{14}), należących do ponad 500 gatunków zwanych mikrobiotą.

Zakłada się, że liczba genów bakteryjnych w organizmie człowieka jest nawet **100 razy większa**, niż w jego własnym genomie !!!

Genom + Mikrobiom = Metagenom



Genom Eu- i Prokaryota

Genom to całkowity DNA komórki, obejmujący zarówno wszystkie geny, jak i odcinki międzygenowe

Genom – materiał genetyczny zawarty w podstawowym (**monoploidalnym**) zespole chromosomów. Termin mylony jest z **genotypem**, czyli całością informacji genetycznej zawartej w chromosomach organizmu

Eukaryota

- genom jądrowy
- genom mitochondrialny
- genom chloroplastowy (rośliny i organizmy fotosyntetyzujące)

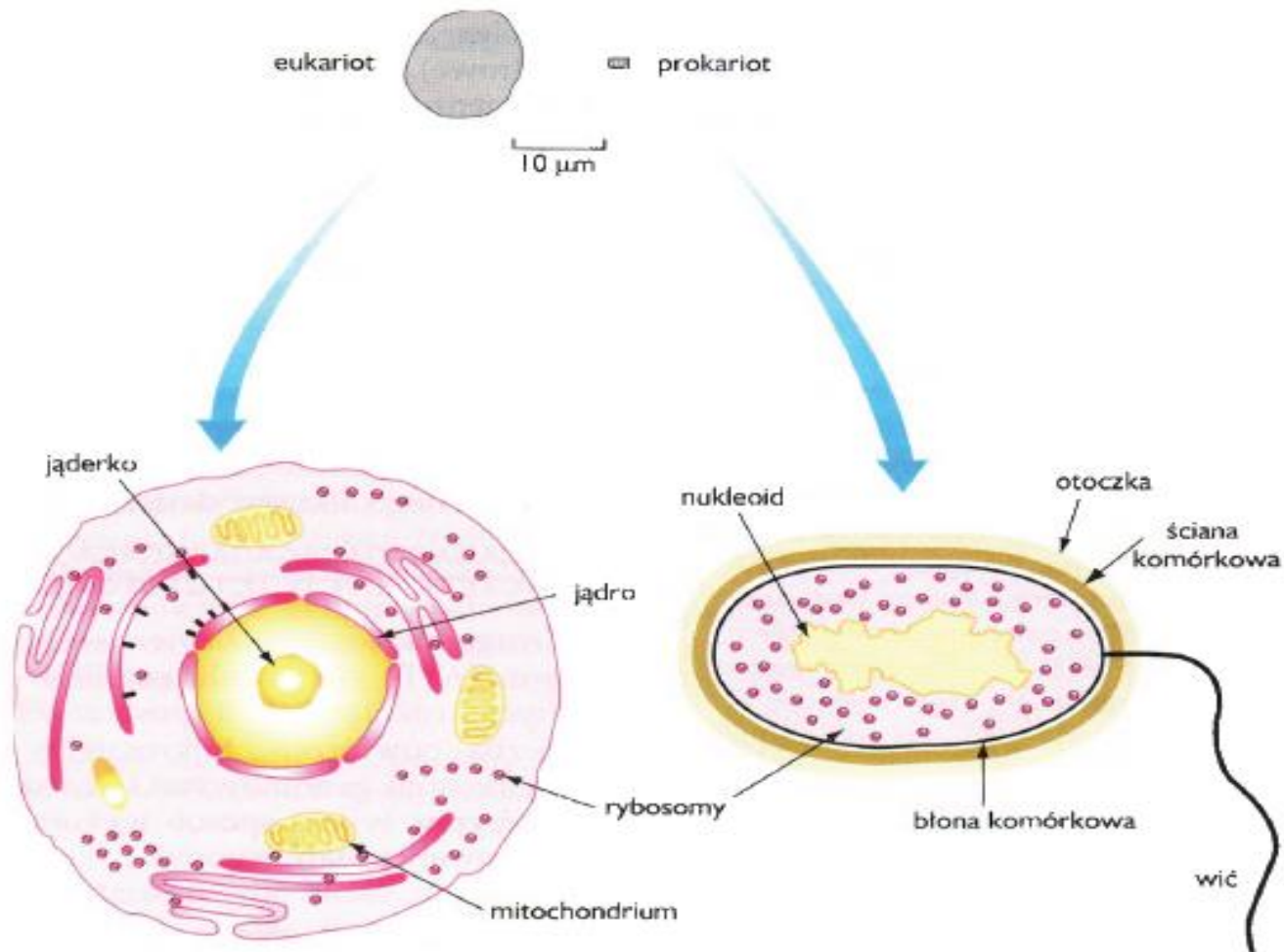
W komórkach Eukaryota materiał genetyczny oddzielony błonami - przedziały (jądro i organelle takie jak mitochondria, a u roślin chloroplasty)

Prokaryota

- genom to kolistą cząsteczką DNA
- dodatkowy materiał genetyczny to liniowy lub kolisty plazmid

W komórkach Prokaryota materiał genetyczny w cytoplazmie - brak wewnętrznych przedziałów

Porównanie genomów



Porównanie genomów

GENOM BAKTERYJNY

- ▶ od jednego do kilku mln par zasad
- ▶ pojedyncza, kolista cząsteczka DNA
- ▶ od 1000 do ponad 4000 różnych genów
- ▶ najbardziej złożony *P. aeruginosa*
- ▶ bardzo ściśle upakowanie informacji genetycznej – niewiele niekodującego DNA
- ▶ geny są ciągłe
- ▶ rzadko występują sekwencje powtarzające się

GENOM EUKARIOTYCZNY

- ▶ zawarty w otoczonym błonami jądrze kom
- ▶ podzielony na liniowe cz. DNA, każda zawarta w innym chromosomie
- ▶ genomy różnią się wielkością
- ▶ geny są mniej zwarte, nieciągłe (introny i eksony)
- ▶ zawierają rozproszone sekwencje repetytywne (powtarzające się)

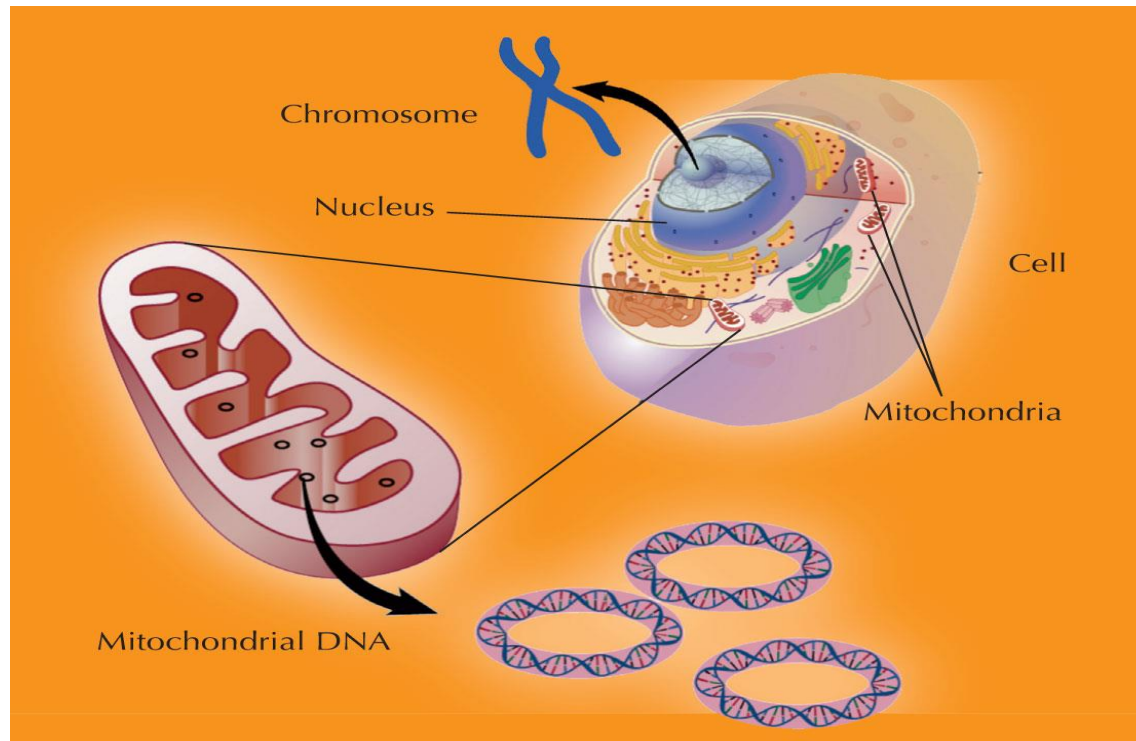
Rozmiary genomów

Organizm	Domena	Przybliżona wielkość genomu (tys. par zasad)	Liczba genów	Rok opublikowani a sekwencji
<i>Mycoplasma genitalium</i>	Bacteria	580	~500	1995
<i>Haemophilus influenzae</i>	Bacteria	1 800	~1 700	1995
<i>E. coli</i>	Bacteria	4 800	~4 300	1997
<i>Pseudomonas aeruginosa</i>	Bacteria	6 300	~5 600	2000
<i>S. cerevisiae</i>	Eukarya	12 000	~6 300	1996
<i>C. elegans</i>	Eukarya	97 000	~19 000	1998
<i>D. melanogaster</i>	Eukarya	185 000	~12 000-14 000	2000
<i>Arabidopsis thaliana</i>	Eukarya	140 000	~25 000	2000
Mysz	Eukarya	2 500 000	~30 000	2002
Szympons	Eukarya	3 000 000	~30 000	2005

Genom człowieka

Genom człowieka:

1. **genom jądrowy** – ok. 3 mld pz. Liniowe cząsteczki DNA (od 55 Mpz do 250 Mpz). W sumie występują 24 rodzaje chromosomów.
2. **genom mitochondrialny** – zaledwie 16 569 pz. Kłosa cząsteczka występująca w wielu kopiach znajduje się w mitochondriach.



www.scienceinschool.org

Mikrobiom człowieka

Joshua Lederberg (laureat nagrody Nobla, 1958), wprowadził do literatury naukowej termin „mikrobiom” w celu określenia informacji genetycznej zawartej w populacji mikroorganizmów zasiedlających organizm człowieka.

Z genetycznego punktu widzenia na supermetagenom człowieka składają się ludzkie geny (1%) i geny populacji mikroorganizmów (99%) .



Mikrobiom człowieka

Obecnie na świecie żyje 7 miliardów ludzi.

Zasiedla nas 10^{24} komórek różnych gatunków bakterii (Bacteria i Archea), w tym w ogromnej większości dotąd niehodowlalnych.

Z danych Metagenomiki oraz różnych kalkulacji ekologicznych wynika, że Ludzki Metabakteriom jest drugim, po Metabakteriomie Oceanów i Mórz (10^{29}), największym rezerwuarem bakterii na naszej planecie Ziemi.

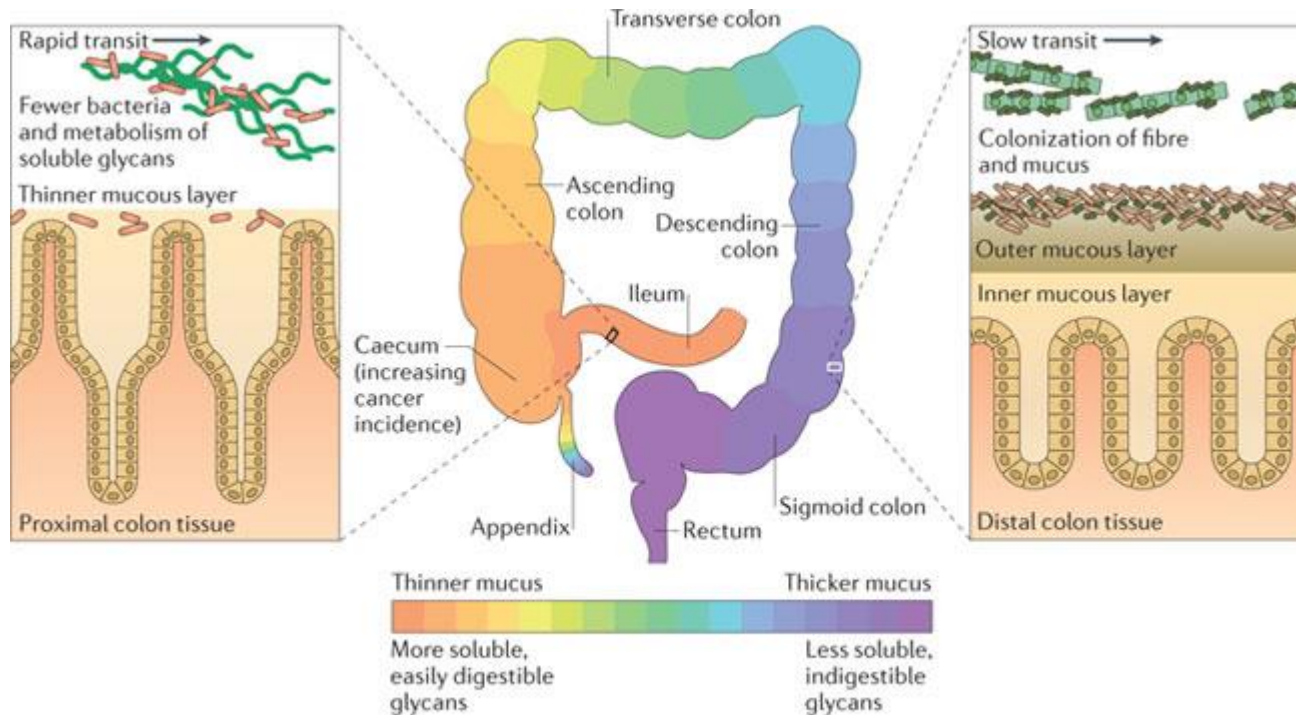
Whitman W. B. et al., 2006



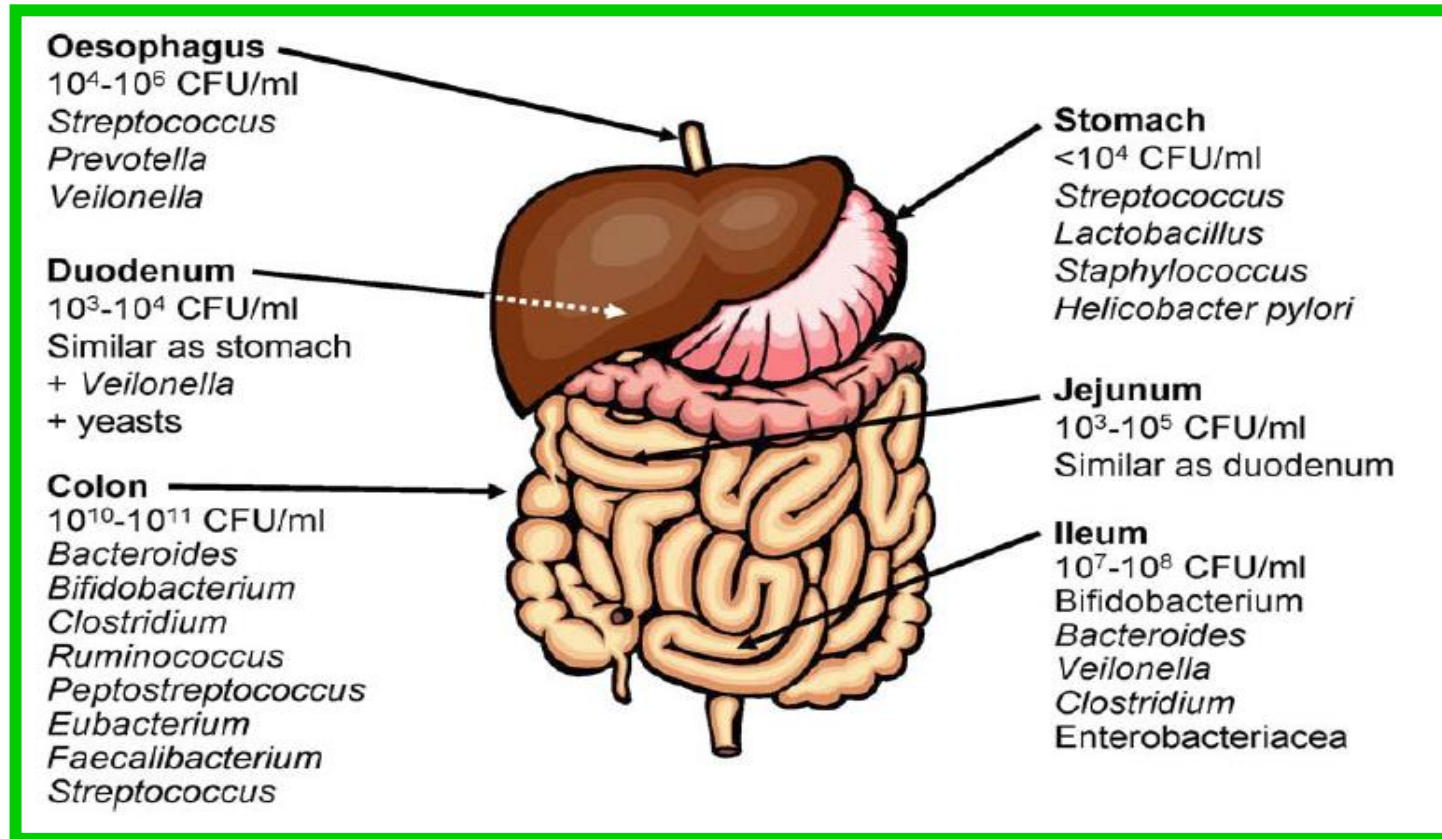
Linda Vu, Berkeley Lab

Mikrobiom przewodu pokarmowego człowieka (GI)

W przewodzie pokarmowym człowieka bytuje 10^{11} - 10^{12} komórek mikroorganizmów, co stanowi **0,5-1,5 kg** masy przeciętnego człowieka.



Mikrobiom przewodu pokarmowego człowieka (GI)



Krstii Tiionen et al., 2010 ; Ouwehand and Vesterlund , 2003

Human Microbion Project

Kierunkowym celem powstałego w 2007 roku projektu HMP jest scharakteryzowanie ludzkiego MIKROBIOMU na poziomie:

METAGENOMU - sekwencji nukleotydowej całego genomowego DNA populacji hodowalnych i niehodowalnych mikroorganizmów człowieka

METATRANSKRYPTOMU - sekwencji nukleotydowej całkowitego mRNA

METAPROTEOMU - syntetyzowanych białek bakteryjnych

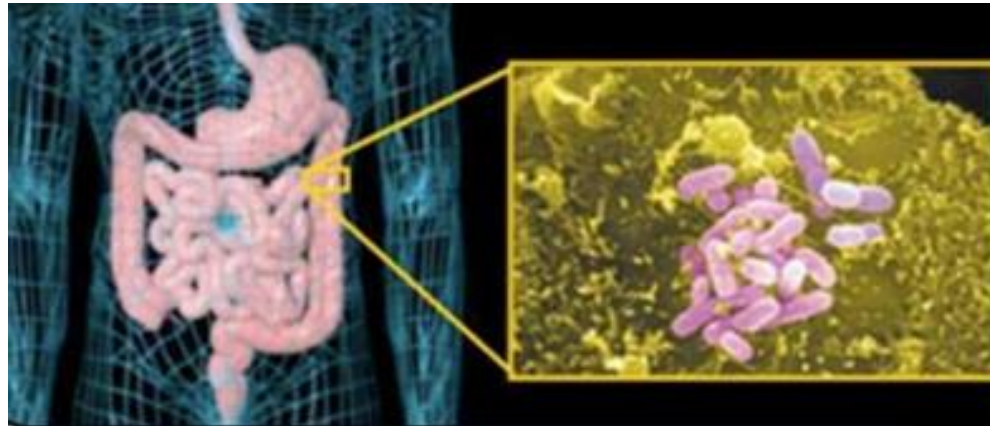
METABOLOMU - produktów metabolizmu



<http://www.reproduccionasistida.org/zona/embarazo/>

Cele projektu HMP

1. Lepsze zrozumienie zróżnicowania genetycznego i fizjologicznego ludzi oraz roli Mikrobiomu w ewolucji człowieka.
2. Identyfikacja określonych warunków, czynników i produktów Mikrobiomu mających bezpośredni lub pośredni wpływ na funkcjonowanie zdrowego organizmu, a także predysponujących do chorób.
3. Opracowanie strategii monitorowania oraz modyfikowania składu oraz funkcji Mikrobiomu indywidualnego człowieka – jako ważnego elementu diagnostyki medycznej.



Prof Dr Tom Van de Wiele

Pytania stawiane HMP

Jaka jest **dynamika zmian indywidualnego mikrobiomu** człowieka w czasie całego, osobniczego życia?

Jakie **podobieństwa i różnice jakościowe, ilościowe oraz funkcjonalne** istnieją w mikrobiomach: członków jednej rodziny, ludzi żyjących w podobnych oraz różnych społeczeństwach i środowiskach, strefach klimatycznych, a także pomiędzy ludźmi różnych grup etnicznych?

W jakim stopniu **styl życia człowieka**, jego fizjologiczny stan, a także patofizjologiczny, wpływają na skład mikrobiomu i jego zróżnicowanie?

Czy wszyscy ludzie posiadają **identyczny „rdzeń” mikrobiomu**, jeśli tak - to w jaki sposób zachodzi jego horyzontalna transmisja?

Metagenomika a matematyka

[Home](#) [About](#) [Science](#) [Events](#) [Participate](#) [Support INI](#) [Search](#)

Mathematical, Statistical and Computational Aspects of the New Science of Metagenomics

Isaac Newton Institute for Mathematical Sciences

Science

Scientific Programmes

Mathematical, Statistical and Computational Aspects of the New Science of Metagenomics

- > Overview
- > [Participants](#)
- > [Preprints](#)
- > [Seminars](#)
- > [Workshops and Other Events](#)

Current Programme

24th March 2014 to 17th April 2014

Organisers: [Wally Gilks](#) (Leeds), [Daniel Huson](#) (Tübingen), [Elisa Loza](#) (Rothamsted Research), [Simon Tavaré](#) (Cambridge), [Gabriel Valiente](#) (Technical University of Catalonia) and [Tandy Warnow](#) (Texas at Austin)

Scientific Advisory Committee: Vincent Moulton (East Anglia), Mihai Pop (Maryland)

Programme Theme

Metagenomics is the study of the total genomic content of microbial communities. In metagenomic studies, DNA material is sampled collectively from the microorganisms that populate the environment of interest (e.g. agricultural soil, ocean water, or the human gut). The extracted DNA sequences are subsequently used to profile the environment and its biodiversity, its dominant microbial classes or biological functions, and whether and how this profile differs from those of other environments.

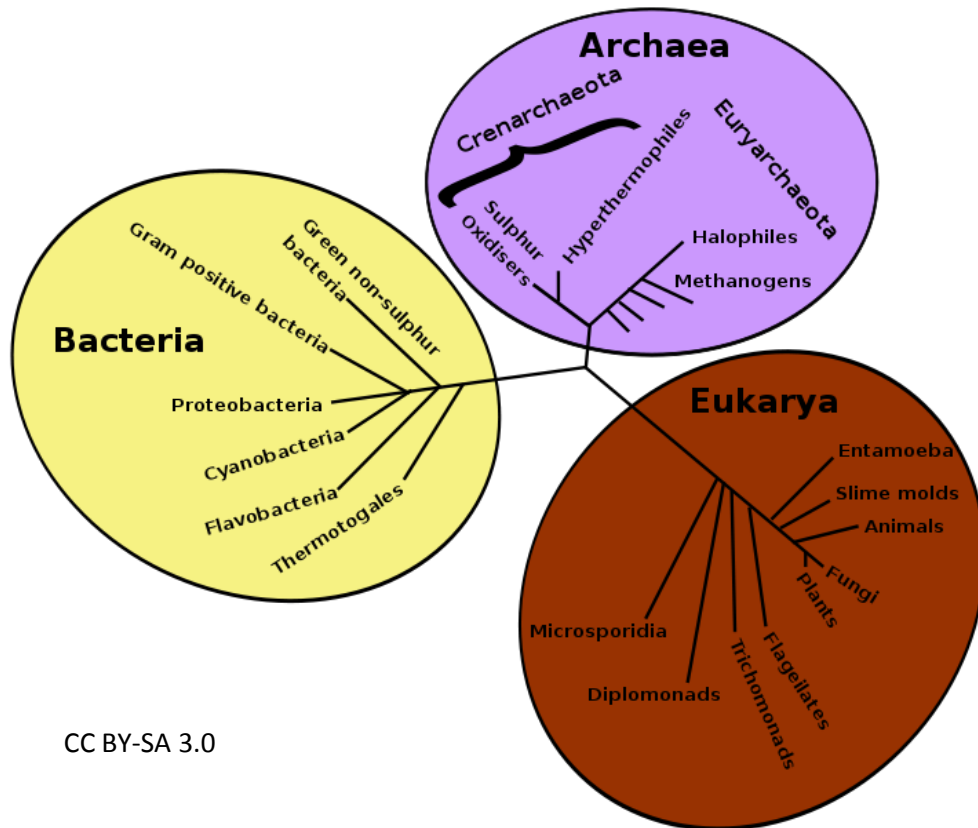
The impact of metagenomics in our understanding of the natural world has been, and will continue to be, revolutionary and profound. Insights derived from metagenomic studies have become increasingly relevant in areas as diverse as human health, bioenergy, environmental sciences and paleontology.

This research programme will bring together leading expertise in the multiple disciplines involved in metagenomics including mathematics,



Filogenetyka molekularna

1. Początki filogenetyki molekularnej.
2. Konstruowanie drzew filogenetycznych na podstawie analizy DNA.



CC BY-SA 3.0

Początki filogenetyki molekularnej

<u>Linnaeus</u> 1735 2 królestwa	<u>Haeckel</u> 1866 3 królestwa	<u>Chatton</u> 1925 (bez opisu)	<u>Copeland</u> 1938, 1956 4 królestwa	<u>Whittaker/Margulis</u> 1969 5 królestw	<u>Woese i in.</u> 1977 6 królestw	<u>Woese i in.</u> 1990 3 domeny
	<u>Protista</u>	<u>Prokaryota</u>	<u>Monera</u>	<u>Monera</u>	<u>Eubacteria</u>	<u>Bacteria</u>
					<u>Archaeobacteria</u>	<u>Archaea</u>
<u>Vegetabilia</u>	<u>Plantae</u>	<u>Plantae</u>	<u>Protoctista</u>	<u>Protista</u>	<u>Protista</u>	<u>Eucarya</u>
			<u>Plantae</u>	<u>Fungi</u>	<u>Fungi</u>	
			<u>Plantae</u>	<u>Plantae</u>	<u>Plantae</u>	
<u>Animalia</u>	<u>Animalia</u>		<u>Animalia</u>	<u>Animalia</u>	<u>Animalia</u>	

Monera – archeony, bakterie, sinice

Protista – pierwotniaki

Eubacteria – bakterie właściwe

Początki filogenetyki molekularnej

Ze względu na **ewolucję genomu** (stopniowe gromadzenie mutacji) liczba różnic pomiędzy sekwencjami nukleotydowymi genomów 2 gatunków powinna wskazywać jak dawno temu istniał wspólny przodek tych genomów.

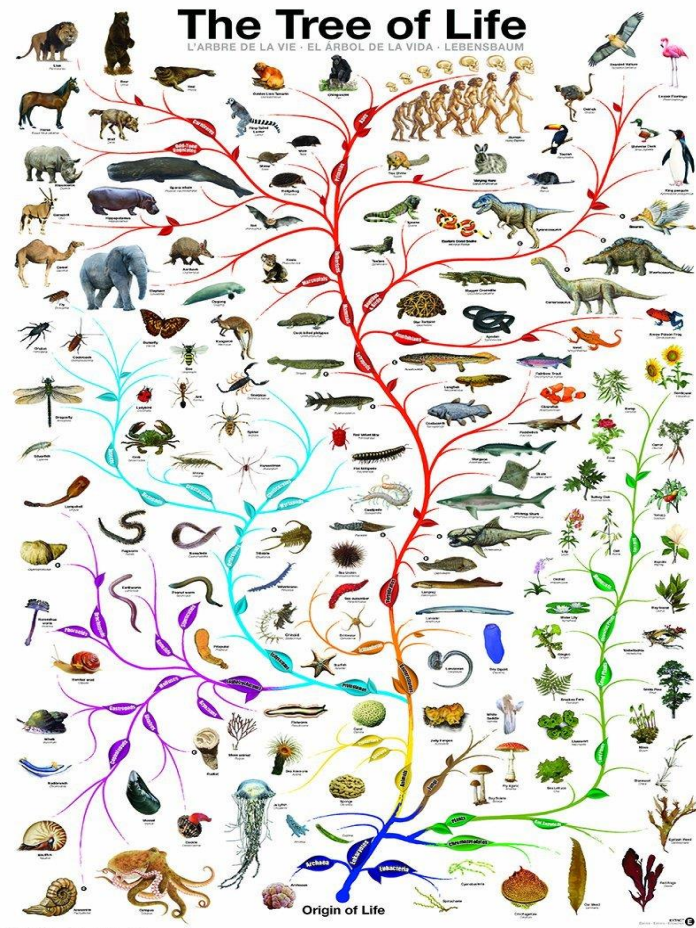
Filogenetyka molekularna wyprzedziła sekwencjonowanie DNA o parę dekad.

XVIII w. **Linneusz** – systematyk (a nie ewolucjonista), klasyfikacja organizmów *Systema Naturae*, **nieświadomie** stworzył podstawy późniejszych schematów ewolucyjnych
Przyrodnicy XVIII i początku XIX w. porównywali hierarchię Linneusza do „**Drzewa życia**”

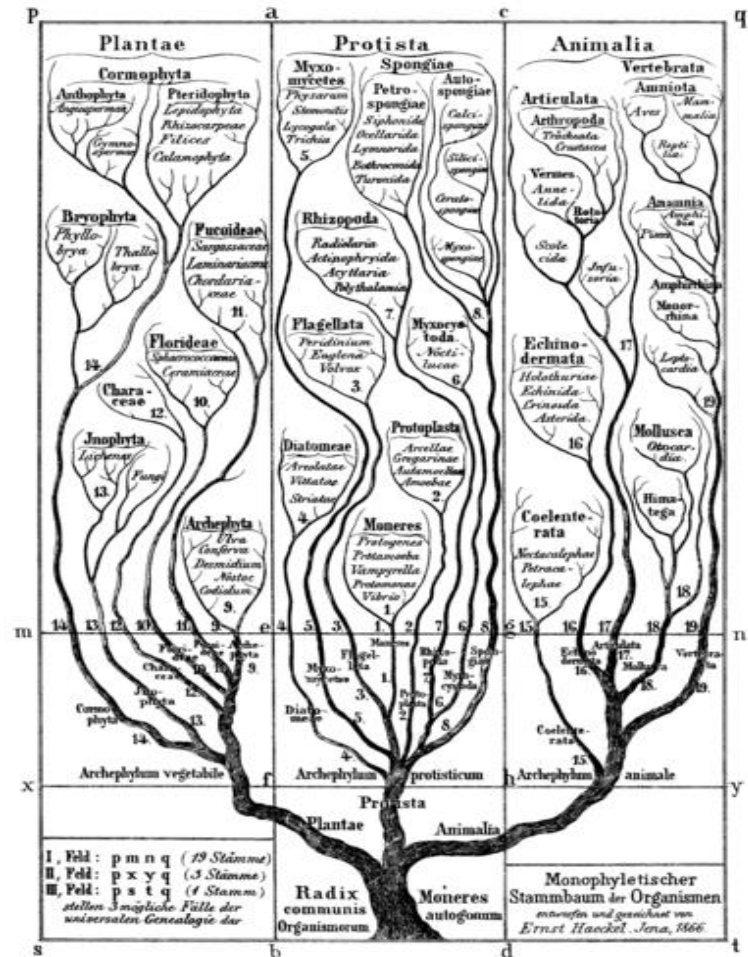
1859r. **Darwin**, twórca teorii ewolucji

Schemat klasyfikacji Linneusza zrozumiano w nowy sposób jako **filogenezę**, przedstawiającą nie tylko podobieństwa między gatunkami ale też ich pokrewieństwo ewolucyjne.

Drzewo życia



by Generic



Haeckel (1866)

Podejście klasyczne do filogenetyki

Fenetyka – klasyfikację organizmów opiera na **porównaniu ograniczonej liczby cech**, bezwzględnie przyjętych przez taksonomów za **istotne**. Porównanie powinno obejmować jak największą liczbę cech, które po przedstawieniu w postaci numerycznej należy poddać analizie ścisłymi metodami **matematycznymi**.

Kladystyka – nie wszystkim cechom przypisuje jednakową wagę. Kladystyka „pyta” o **znaczenie ewolucyjne** poszczególnej cechy.

Błędy w ustalaniu wzoru rozgałęzienia można zmniejszać odróżniając 2 rodzaje nieprawidłowych danych:

- **ewolucja zbieżna (homoplazja)** – taka sama cecha zdarza się niezależnie w 2 oddzielnych liniach (ptaki i nietoperze mają skrzydła). Cecha „posiadania skrzydeł” może być myląca w ewolucji kręgowców
- **odróżnienie cech amorficznych od plezjomorficznych** – **c. plezjomorficzne** dziedziczone są po dalekim przodku (np. 5 palców), **c. apomorficzna** wyewoluowała u niedawnego wspólnego przodka. Np. posiadanie 1 palca u konia jest c. apomorficzną (a jaszczurka ma 5 palców)

Wg PWN:

fenetyka - metoda klasyfikacji organizmów uwzględniająca wyłącznie **ogólne podobieństwa** ich cech anatomicznych, embriologicznych, biochem., behawioralnych itp.; bywa nazywana **taksonomią numeryczną**, gdyż posługuje się komputerową analizą wielkiej liczby cech.

kladystyka - metoda klasyfikacji organizmów na podstawie **wspólnego pochodzenia ewolucyjnego**, a nie ich podobieństw; poszczególne taksony wyodrębnia się na podstawie **cech apomorficznych**, tj. zjawiających się po raz pierwszy w ewolucji badanej grupy i dziedziczonych.

Podjęcie molekularne do filogenetyki

Niezależnie od celu – stworzenia klasyfikacji czy wnioskowania o filogenezie – dane potrzebne do analizy otrzymuje się badając **zmienne cechy** porównywanych organizmów.

Początkowo posługiwano się **cechami morfologicznymi**, ale zaskakująco wcześniej zaczęto stosować **dane molekularne!!!**

W **1904 r. George Nuttall** zastosował **testy immunologiczne** do wnioskowania o pokrewieństwie rozmaitych zwierząt.

W/w testy nie były stosowane do lat 50-tych ze względu na ograniczenia techniczne i brak ich docenienia. Zmiany zaszły wraz z wprowadzeniem **fenetyki** i **kladystyki**. Obie opierają się na analizie dużych zestawów danych za pomocą ścisłych procedur **matematycznych**.

Trudność jaką przedstawiało zgromadzenie takich danych w przypadku cech morfologicznych, znikła, gdy stopniowo zaczęto korzystać **z białek** i **DNA**, mającymi w porównaniu z innymi rodzajami danych następujące zalety:

- wiele cech molekularnych można **poznać jednocześnie**; każda pozycja nukleotydowa w sekwencji DNA jest cechą mającą **4 stany** **cech A, G, C i T**.
- molekularne stany cech są **jednoznaczne**: A, G, C i T są łatwo odróżnialne. Stany cech morfologicznych czasami nie są do odróżnienia
- dane molekularne można **łatwo sprowadzić do postaci numerycznej**, a zatem bardzo dobrze nadają się do analizy **matematycznej** i **statystycznej**.

Podjęcie molekularne do filogenetyki

Najdokładniejszych i najbardziej jednoznacznych wyników filogenetyce molekularnej dostarczają **sekwencje DNA i białek**.

Białka zaczęto sekwencjonować **w latach 60-tych** a DNA **w 70-tych**.

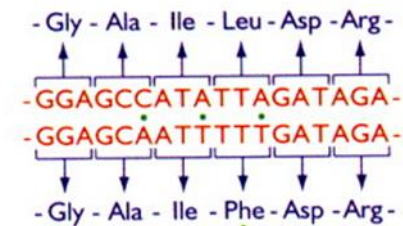
Początkowo badania polegały na określaniu zmienności DNA i białek przy użyciu 3 metod:

- **dane immunologiczne** (od 1904r), reakcja krzyżowa podczas mieszania p/ciała specyficznego w stosunku do określonego białka z 1 organizmu z analogicznym białkiem z 2 organizmu,
- **elektroforez białek** - porównywanie blisko spokrewnionych gatunków,
- **dane o hybrydyzacji DNA-DNA** – porównuje się hybrydyzację DNA 2 organizmów.

W latach 80-tych zaczęto przeprowadzać na dużą skalę sekwencjonowanie DNA (sekw. DNA „wyparło” sekw. białka).

Zalety stosowania DNA do analiz filogenetycznych:

- DNA dostarcza **więcej informacji filogenetycznej** niż b
- analiza DNA pozwala uzyskać zupełnie nową informację kodujących jak i niekodujących
- łatwość przygotowania próbek do analizy sek. DNA (P



szarów

Dwie sekwencje DNA różnią się w 3 pozycjach, choć sekwencje białkowe różnią się w 1 pozycji

Konstruowanie drzew filogenetycznych na podstawie analizy DNA

Celem większości badań filogenetycznych jest konstruowanie **drzew** opisujących związki ewolucyjne między badanymi organizmami lub genami.

Główne cechy drzew filogenetycznych zbudowanych na podstawie DNA

Topologia drzewa

Węzły wewnętrzne – geny przodków

Węzły zewnętrzne – porównywane geny

Gałęzie – długość decyduje w jakim stopniu różnią się geny

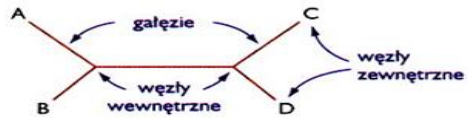
W przypadku **braku uкорzenia drzewa** do analizy trzeba włączyć przynajmniej 1 gen homologiczny mniej spokrewniony (niż pozostałe geny między sobą), by uкорzenić drzewo.

Drzewo otrzymane dzięki analizie jest **drzewem proponowanym** – opisuje ciąg zdarzeń ewolucyjnych, wywnioskowanych z analizowanych danych i nie musi być identyczne z **drzewem rzeczywistym**.

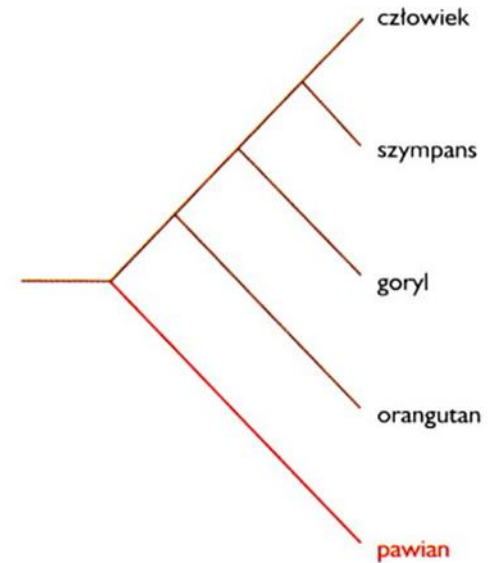
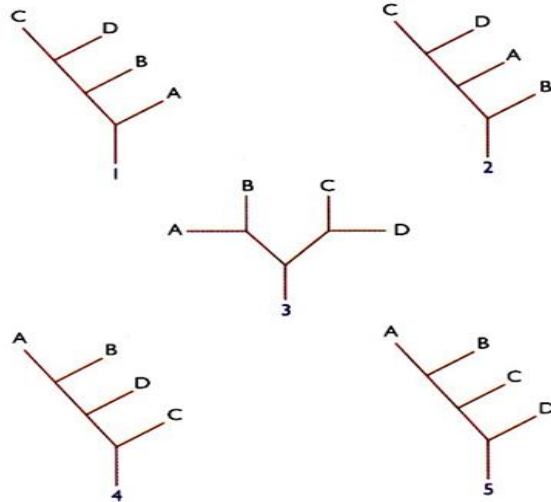
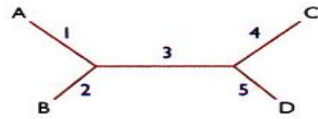
Większość analiz filogenetycznych **nie dostarcza w pełni jednoznacznych wyników**, dlatego mogą wynikać różnice między drzewem proponowanym i rzeczywistym!

Drzewa filogenetyczne

(A) Drzewo nieukorzenione



(B)



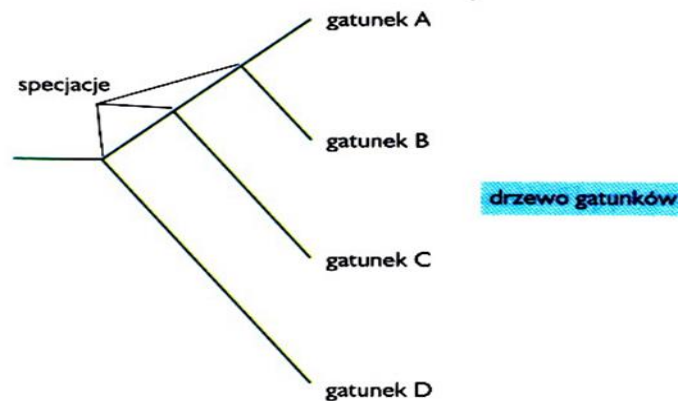
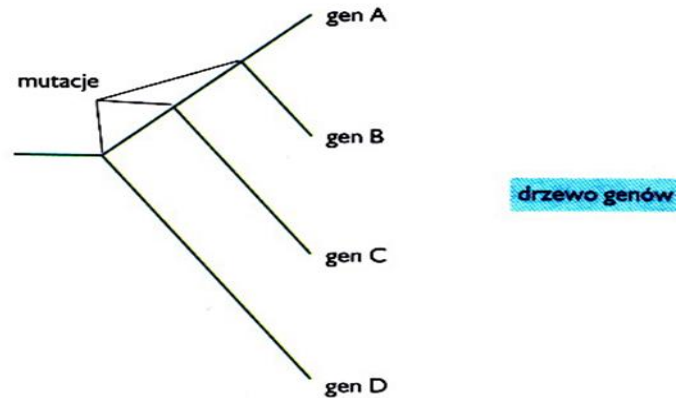
Wykorzystanie grupy zewnętrznej do ukorzeniania drzewa filogenetycznego

Drzewa genów nie są tożsame z drzewami gatunków

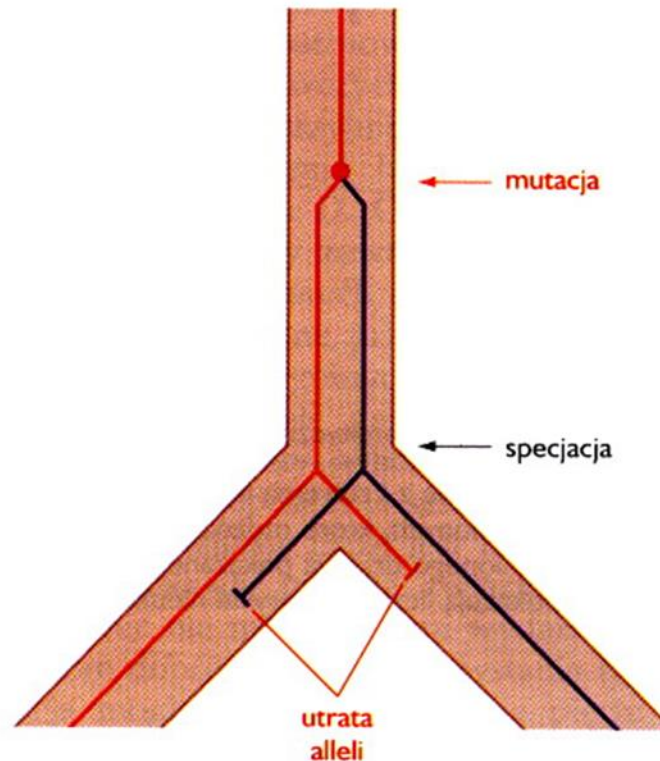
Aby drzewa genów i gatunków były tożsame, ich węzły wewnętrzne musiałyby być dokładnie równoważne, co zwykle nie następuje:

- węzeł wewnętrzny **w drzewie genów** odpowiada **dywergencji na skutek mutacji** genu wyjściowego, w wyniku czego powstają 2 geny o różnych sekwencjach
- węzeł wewnętrzny **w drzewie gatunków** odzwierciedla **specjację**: następuje ona przez podzielenie się populacji na 2 grupy, które nie są w stanie krzyżować się ze sobą (np. w wyniku izolacji geograficznej).

Różnica między drzewem genów a drzewem gatunków

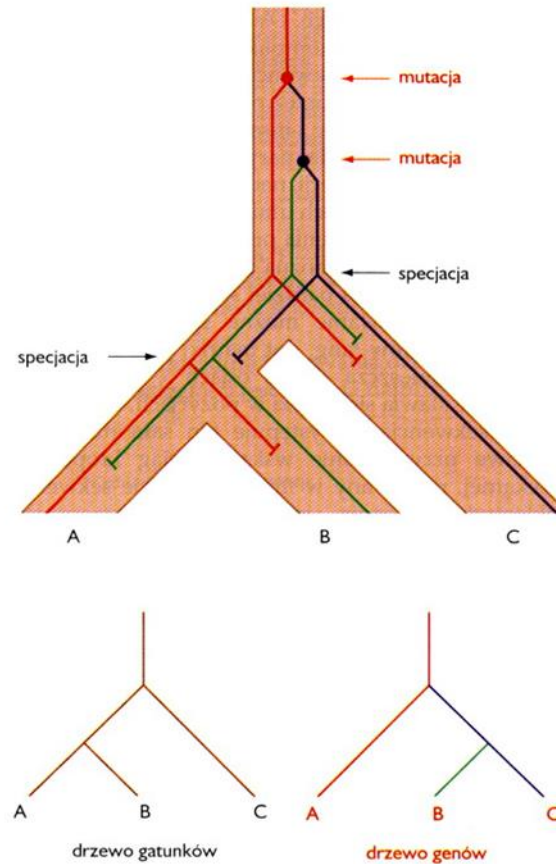


Mutacja i specjacja nie muszą występować w tym samym momencie



Mutacja może poprzedzać specjację

Drzewo genów może się rozgałęziać w inny sposób niż drzewo gatunków



Budowanie drzew

Etapy budowania drzew

1. **Porównanie** sekwencji DNA i **otrzymanie danych** do budowy drzewa.
2. Kilka różnych metod stosowanych do **przekształcenia danych** porównawczych i budowania drzewa.
3. Testy służące ustaleniu stopnia **wiarygodności** drzewa i jego odcinków.
4. Stosowanie **zegara molekularnego** do **ustalenia dat** rozgałęzień drzewa.

Adn. 1

Porównanie sekwencji DNA i otrzymanie danych do budowy drzewa

Porównania dokonuje się dopasowując sekwencje tak, że można zliczyć różnice w sekwencjach nukleotydów – **krytyczny punkt** przedsięwzięcia – jeśli przyrównanie jest niepoprawne, drzewo zbudowane na jego podstawie będzie nieprawdziwe.

Przyrównanie sekwencji

Problemy do rozwiązania

- Ustalenie czy sekwencje są homologiczne
- Dopasowanie sekwencji w sposób pozwalający na porównanie nukleotydów homologicznych

Podczas porównania nie da się stwierdzić czy jest delecja czy insercja (=indel)

(A) Proste przyrównanie sekwencji

```
AGCAATGGCCAGACAATAATG
AGCTATGGACAGACATTAATG
*** * * * * * * * * *
```

(B) Trudniejsze przyrównanie sekwencji

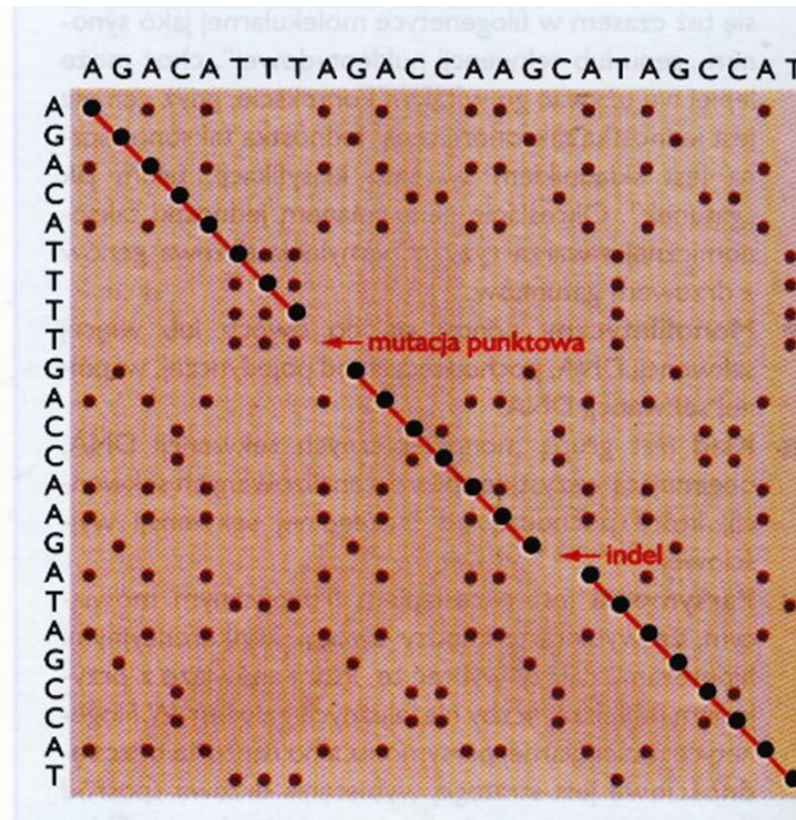
```
GACGACCATAGACCAGCATAG
GACTACCATAGA - CTGCAAAG
*** * * * * * * * * *

GACGACCATAGACCAGCATAG
GACTACCATAGACT - GCAAAG
*** * * * * * * * * *
```

dwa możliwe
położenia indeli

Przyrównanie sekwencji metodą macierzy punktów

Dla bardziej skomplikowanych par sekwencji dopasowanie można sporządzić metodą **macierzy punktów**.



Bardziej sformalizowane **metody matematyczne** to:

- **Metoda podobieństwa** (1970) – maksymalizacja liczby dopasowanych, identycznych w obu sekwencjach nukleotydów
- **Metoda odległościowa** (1976) - celem jest minimalizacja dopasowań

Często zastosowanie w/w metod daje ten sam wynik.

Najczęściej stosowany program to **Clustal (1998)**.

Przekształcenie danych porównawczych

Do dziś nikt nie stworzył doskonałego sposobu budowania drzew i zwykle stosuje się kilka metod.

Macierz odległości – najprostsza metoda – tabela pokazująca **odległości ewolucyjne** dla wszystkich par sekwencji w zbiorze danych. Odległości oblicza się na podstawie ilości różnic nukleotydowych między parą sekwencji i stosuje się do ustalenia długości gałęzi.

równanie wielu sekwencji

```
1 AGGCCAAGCCATAGCTGTCC
2 AGGCAAAGACATACCTGACC
3 AGGCCAAGACATAGCTGTCC
4 AGGCAAAGACATACCTGTCC
```

macierz odległości

	1	2	3	4
1	–	0,20	0,05	0,15
2		–	0,15	0,05
3			–	0,10
4				–

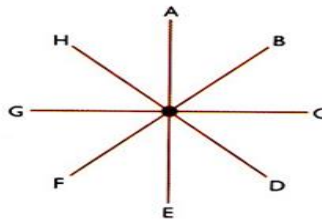
Macierz pokazuje odległości ewolucyjne dla wszystkich par sekwencji przyrównania. W tym przykładzie odległość ewolucyjną wyraża ilość różnic nukleotydowych przypadająca na jeden nukleotyd dla każdej pary. Na przykład, sekwencje 1 i 2 mają po 20 nukleotydów i różnią się 4 nukleotydami co odpowiada różnicy ewolucyjnej $4/20 = 0,2$.

Etapy przeprowadzane podczas budowy drzewa metodą NS

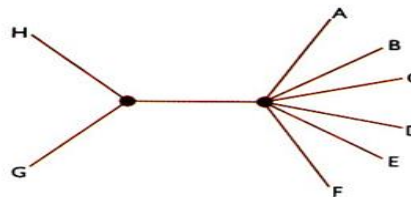
Metoda najbliższego sąsiada (NS) – popularna, zasady:

1. Budowę rozpoczyna się od 1 węzła łączącego wszystkie sekwencje DNA
2. Wybiera się losowo parę sekwencji i usuwa ją z drzewa „gwiazdy” i przywiązuje się do drugiego węzła wewnętrznego
3. W nowym drzewie oblicza się całkowitą długość gałęzi
4. Umieszcza się sekwencje w ich początkowych pozycjach i do drugiego węzła przyczepia się inną sekwencję
5. Czynności powtarza się do wyczerpania wszystkich możliwych par sekwencji, pozwalając na wybór pary sekwencji dającej najkrótszą długość gałęzi.
6. Znalezione 2 sekwencje traktuje się jak jedną.
7. Procedurę powtarza się do określenia położenia wszystkich sekwencji i uzyskania pełnego drzewa.

(A) Punkt wyjścia dla metody najbliższych sąsiadów



(B) Usunięcie dwóch sekwencji z gwiazdy



Zaletą metody NS jest stosunkowa łatwość manipulowania danymi.

Wadą metody NS jest utrata części informacji, zwłaszcza odnoszących się do identyczności nukleotydów sekwencji wyjściowych i sekwencji pochodnych.

Informacji tej nie traci się w **metodzie największej oszczędności** – zakłada się (prawopodobnie błędnie), że ewolucja przebiega **najkrótszą możliwą drogą** i poprawnym drzewem jest to, które do powstawania różnic obserwowanych między sekwencjami wymaga najmniejszej możliwej liczby podstawień nukleotydowych.

Metoda największej oszczędności zmusza do manipulacji większą liczbą danych – **PROBLEM matematyczny!**

Przy 5 sekwencjach jest możliwych tylko **15** nieukorzenionych drzew

Przy 10 sekwencjach jest możliwych **2 027 025** nieukorzenionych drzew

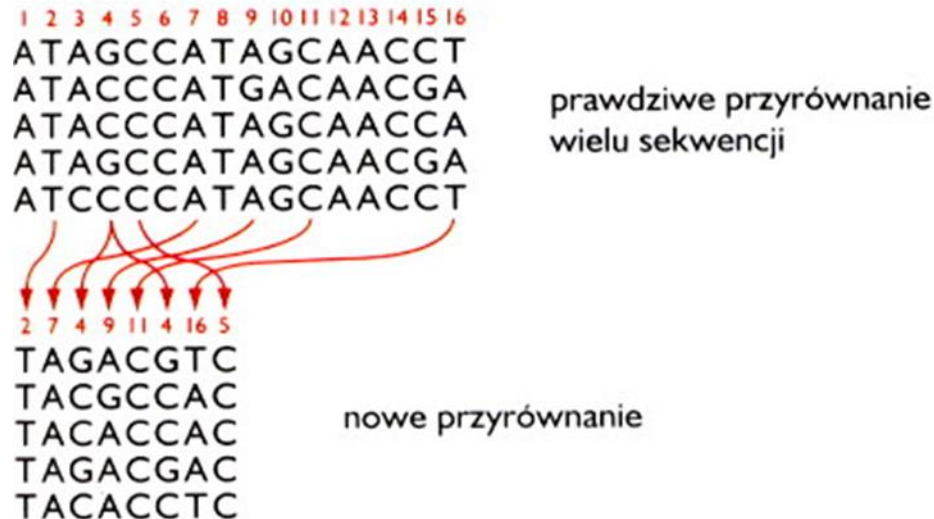
Przy 50 sekwencjach liczba nieukorzenionych drzew **przekracza liczbę atomów we wszechświecie!**

Określanie dokładności skonstruowanego drzewa

Znane są testy **statystyczne** dokładności skonstruowanych drzew, ale ze względu na geometryczny a nie numeryczny charakter drzew testy te są złożone.

W praktyce wykorzystuje się **metodę BOOTSTRAP**.

Trzeba utworzyć dodatkowe przyrównania wielu sekwencji, odmienne od prawdziwego, ale mu **równoważne**. Nowe przyrównania tworzy się losowo zmieniając miejsce kolumny z prawdziwego przyrównania.



Nowe przyrównania dotyczą sekwencji różniących się od wyjściowych, ale o **podobnej zmienności** – powinno się **nie tylko powtórzyć pierwotną analizę, ale i otrzymać to samo drzewo!**

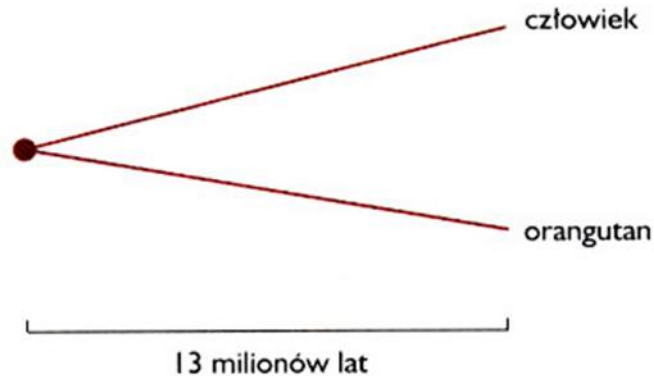
Buduje się 1000 przyrównań i konstruuje się dla nich 1000 drzew. Na ich podstawie można przypisać **wartość bootstrap** każdemu wewnętrznemu węzłowi początkowego drzewa. Wartość bootstrap jest równa liczbie wystąpień (w drzewach wytworzonych metodą Bootstrap) węzłów identycznych z węzłem z drzewa początkowego.

Wartość bootstrap dla węzła $>700/1000$ oznacza, że można węzłowi przypisać rozsądne prawdopodobieństwo.

Zegar molekularny

Pozwala oszacować czas, jaki upłynął od rozdzielenia się sekwencji – wymaga to **kalibracji**, czyli określa się jakiej liczby **podstawień nukleotydowych** można się spodziewać dla okresu **milion lat**.

Znowu **matematyka** 😊



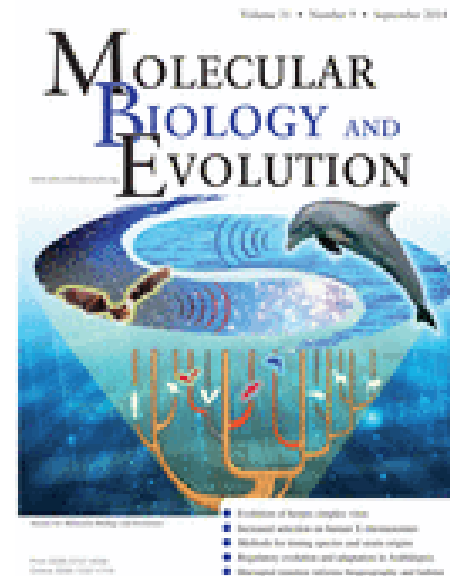
Zegar molekularny dla człowieka

liczba podstawień	$= x$
liczba podstawień na linię	$= \frac{x}{2}$
liczba podstawień na linię na milion lat	$= \frac{x}{2 \times 13}$

Liczbę podstawień (x) określa się dla pary homologicznych genów człowieka i orangutana.

Ewolucja molekularna – droga ewolucji genomów

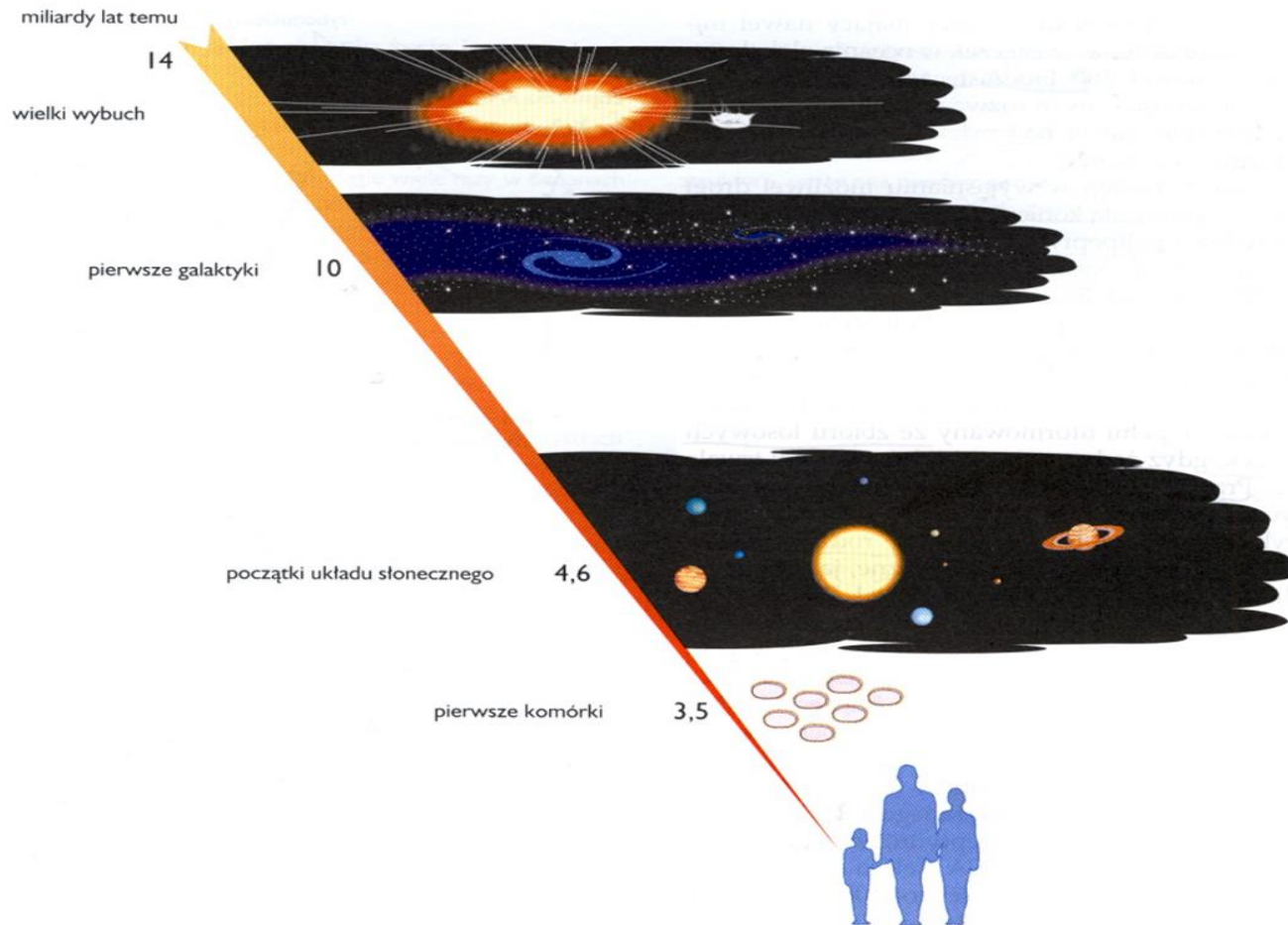
1. Pochodzenie genomów.
2. Powstawanie nowych genów.
3. Nabywanie genów od innych gatunków



Genomy: pierwsze 10 mld lat

1. Kosmolodzy uważają, że wszechświat powstał ok. **14 mld** lat temu - **Wielki Wybuch**
2. Wg modeli **matematycznych** **10 mld** lat temu powstały pierwsze **galaktyki**
3. W naszej galaktyce mgławica słoneczna zaczęła kondensować tworząc **słońce i planety** **4,6 mld** lat temu
4. Początkowo ziemia pokrywała woda, w której pojawiły się pierwsze **systemy biochemiczne**
5. **Życie komórkowe** powstało **3,5 mld** lat temu, gdy zaczęły pojawiać się pierwsze lądy
6. Życie komórkowe było dość późnym stadium w ewolucji biochemicznej - najpierw poprzedziły ją **samoreplikujące się polinukleotydy** (przodkowie pierwszych genów)

Ewolucja molekularna – droga ewolucji genomów



Początki wszechświata, galaktyk, układu słonecznego i życia komórkowego

Powstanie prostych związków organicznych

1. Uważa się, że pierwotne oceany posiadały sole o podobnym składzie do współczesnych.
2. Atmosfera ziemska, a zatem i gaz rozpuszczone w oceanach, znacznie różniły się od dzisiejszych
3. Udział tlenu stał się znaczący dopiero po pojawieniu się **fotosyntezy**.
4. Początkowo najobficiej występującymi gazami były najprawdopodobniej **metan** i **amoniak**.

Doświadczenia próbujące odtworzyć warunki występujące **w prądawnej atmosferze** wykazały, że wyładowania elektryczne (w mieszaninie amoniaku i metanu) powodują syntezę **wielu aminokwasów** (m.in. alaniny, glicyny, waliny). Powstawał też cyjanowodór i formaldehyd, które w dodatkowych reakcjach przyczyniły się do powstania innych **aminokwasów**, **puryn** i **pirymidyn**, a w mniejszym stopniu też **cukrów**.

Pierwsze systemy biochemiczne były oparte na RNA!

Polimeryzacja podstawowych składników w biocząsteczki mogła nastąpić w oceanach albo mogła spowodować ją wielokrotna kondensacja i parowanie kropel wody w chmurach

Polimeryzacja mogła też następować na powierzchniach stałych (np. ił)

Jak przejść od zbioru losowych biocząsteczek do ich uporządkowanej grupy?!

Ten etap nie został nigdy odtworzony doświadczalnie.

Przyjmując 1 mld lat i zagęszczenie cząsteczek w oceanie 10^{10} biocząsteczek/litr to daje nieograniczone rozważania... i pole do popisu dla **matematyków?**s

W latach 80-tych nastąpił przełom – odkrycie **katalitycznej cząsteczki RNA**.

RNA przeprowadza:

1. Cięcie autokatalityczne (samowycinające się introny grupy I, II i III)
2. Cięcie innych cząsteczek RNA (Rnaza P)
3. Synteza wiązań peptydowych
4. Synteza rybonukleotydów (*in vitro*)
5. Synteza cząsteczek RNA (*in vitro*)

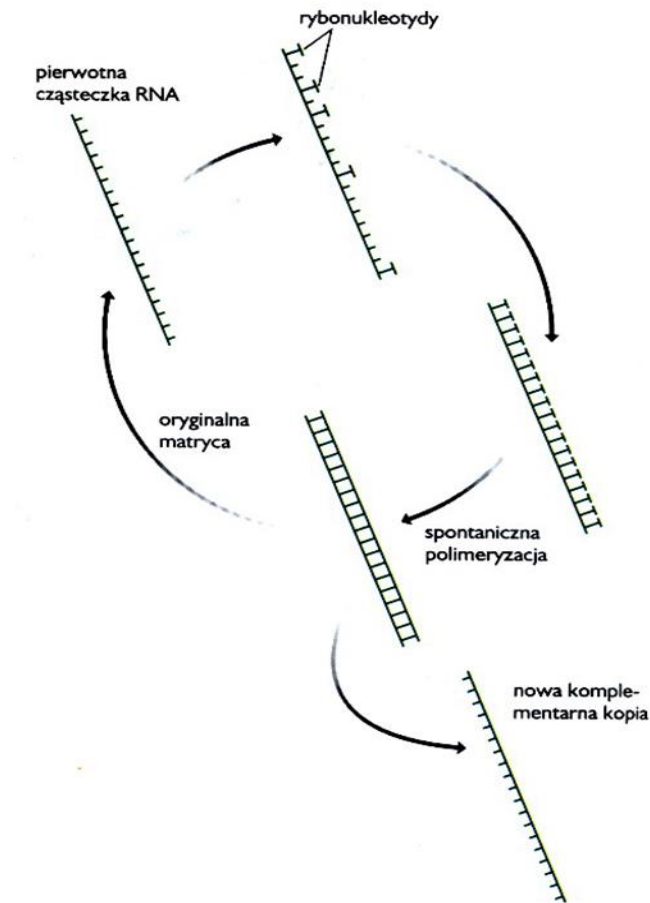
Pierwsze systemy biochemiczne mogły opierać się w całości na RNA!

Prawdopodobnie cząsteczki RNA replikowały powoli i niedokładnie.

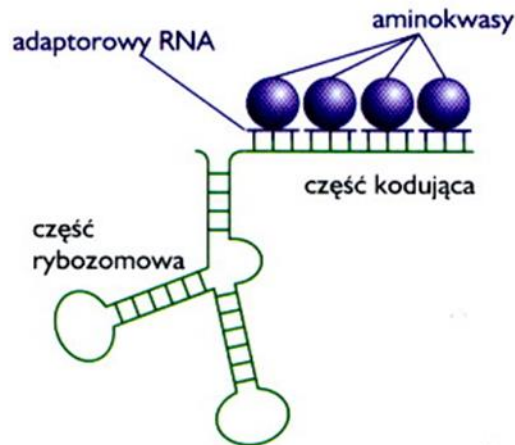
W/w cząsteczek RNA nie można nazywać genomami ale **proteomami** - zdolność cząsteczek do samoreplikacji i przeprowadzania prostych reakcji biochemicznych.

Po zamknięciu cząsteczek RNA wewnątrz przedziałów lipidowych spowodowało powstanie pierwszych **struktur komórkopodobnych**.

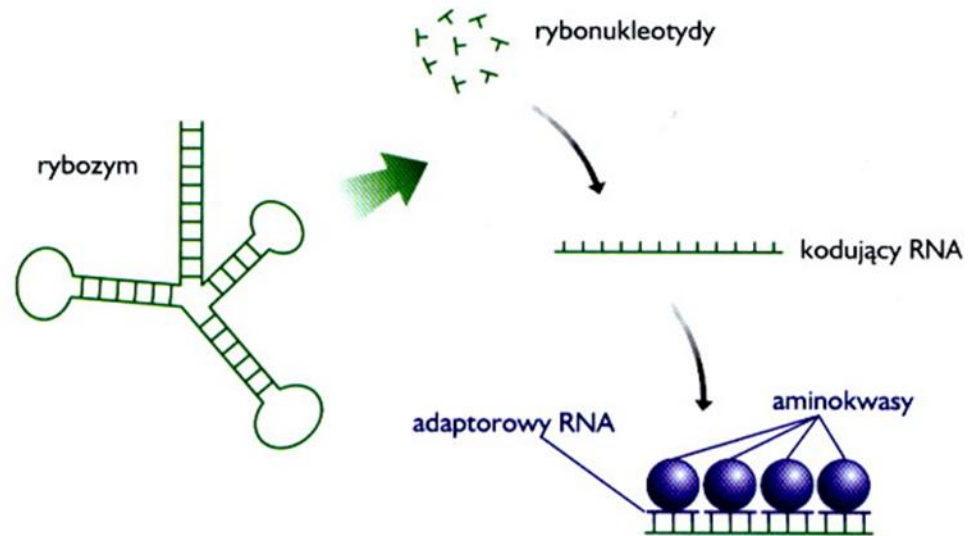
Kopiowanie cząsteczek RNA we wczesnym świecie RNA



Dwie możliwości przebiegu ewolucji pierwszego kodującego RNA



Rybozym będący jednocześnie cząsteczką kodującą



Rybozym syntetyzujący cząsteczkę kodującą

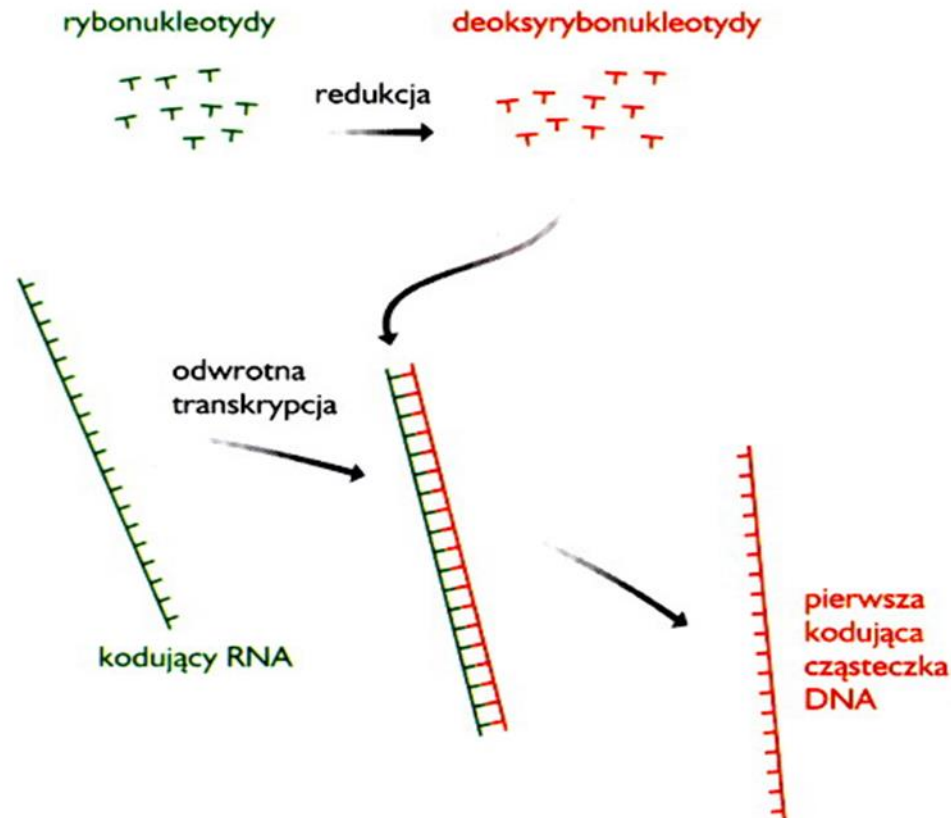
Pierwsze genomy zbudowane z DNA

Jak świat RNA przekształcił się w DNA?

1. Pierwszą ważną zmianą był prawdopodobnie rozwój **enzymów białkowych**, które uzupełniły a później zastąpiły aktywność katalityczną rybozymów.
2. Protogenomy RNA utraciwszy swoją rolę enzymatyczną zachowały funkcję kodowania.
3. Przeniesienie funkcji kodowania na **stabilny DNA** wydaje się nieuniknione i dosyć łatwe (redukcja grupy 2'OH, U w T).
4. **Odwrotna transkrypcja** pozwoliła polimeryzować DNA z RNA.
5. Przyjęcie przez kodujący DNA **formy dsDNA** spowodowane możliwością naprawy uszkodzeń DNA przez kopiowanie siostrzanej nici. DNA.

Pierwsze genomy zbudowane z DNA składały się z wielu oddzielnych cząsteczek, z których każda odpowiadała innemu białku i była odpowiednikiem pojedynczego genu.

Przekształcenie kodującej cząsteczki RNA w przodka pierwszego genomu DNA



Kalendarium powstawania organizmów

3,5 mld lat temu powstały komórki podobne do bakterii

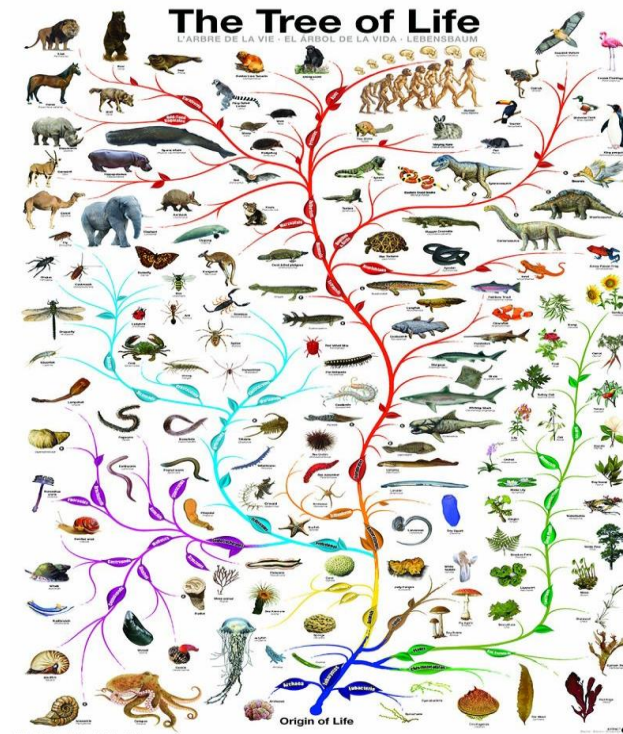
1,4 mld lat temu powstały komórki eukariotyczne jednokomórkowe glony

0,9 mld lat temu powstały komórki eukariotyczne wielokomórkowe glony

0,53 mld lat temu powstały bezkręgowce

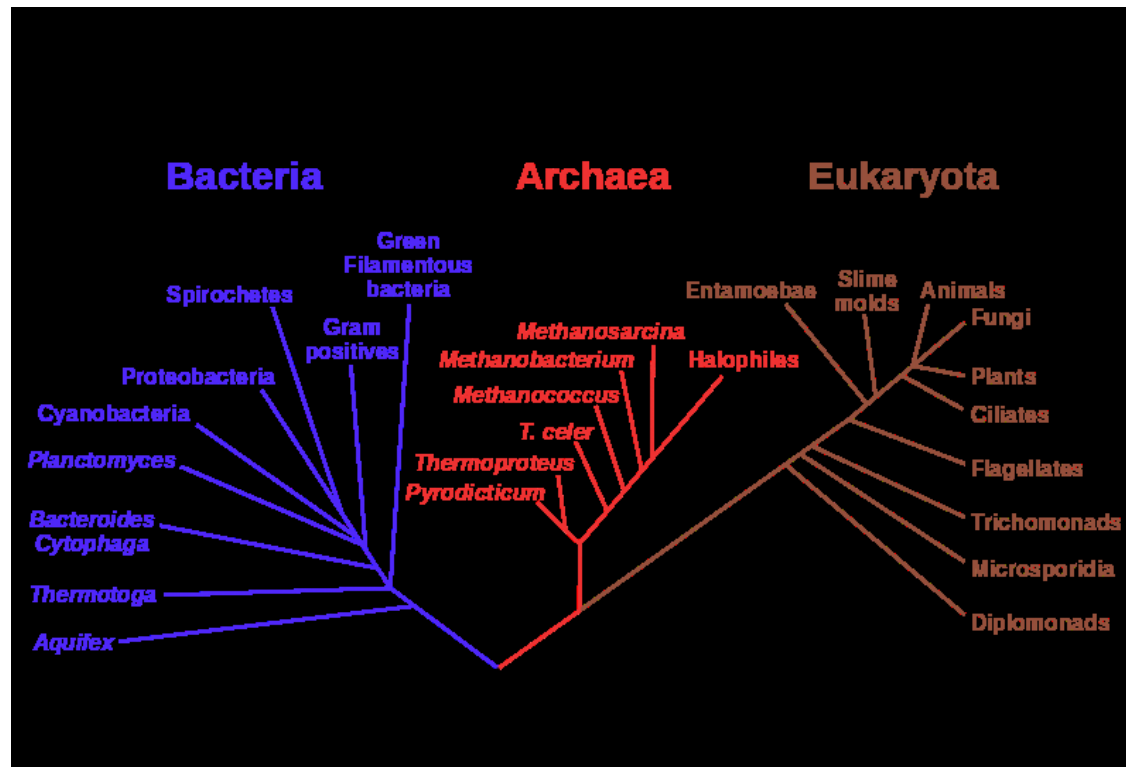
0,3 mld lat temu powstały rośliny lądowe

4,5 mln lat temu powstały pierwsze hominidy



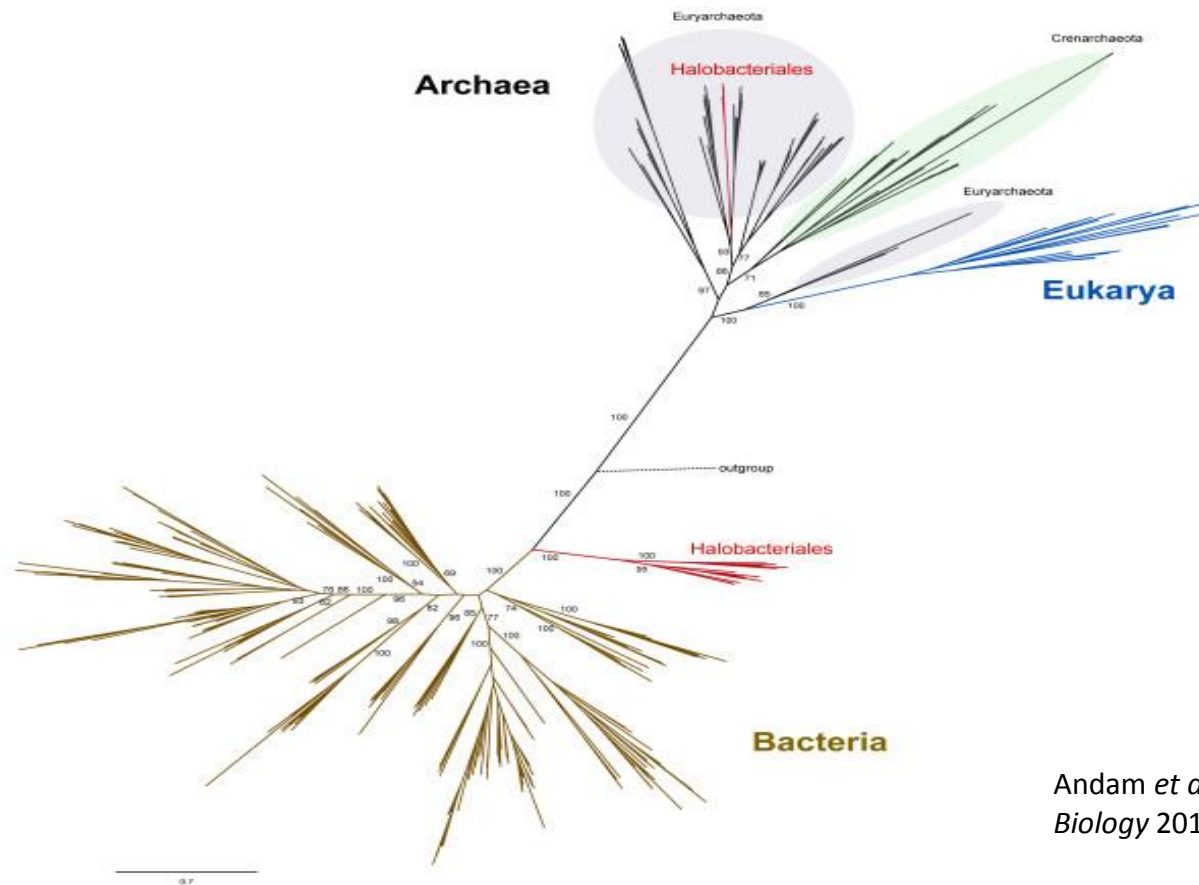
by Generic

Drzewo filogenetyczne oparte na badaniach rRNA



Eric Gaba - NASA Astrobiology Institute

Drzewo filiogenetyczne uzyskane na podstawie sekwencji aminokwasowych białka LeuRS



Andam *et al.* *BMC Evolutionary Biology* 2012 **12**:85

Ewolucji morfologicznej towarzyszyła ewolucja genomu!

Wraz z poruszaniem się po drzewie ewolucji wzrasta złożoność genomów - wskaźnikiem jest liczba kodujących genów od ok. 1000 dla bakterii do 80 000 dla niektórych kręgowców.

Istnieją **2 drogi pojawiania się nowych genów w genomie**:

1. Przez **duplikację** niektórych lub wszystkich genów w genomie
2. Przez **nabywanie** genów od innych gatunków

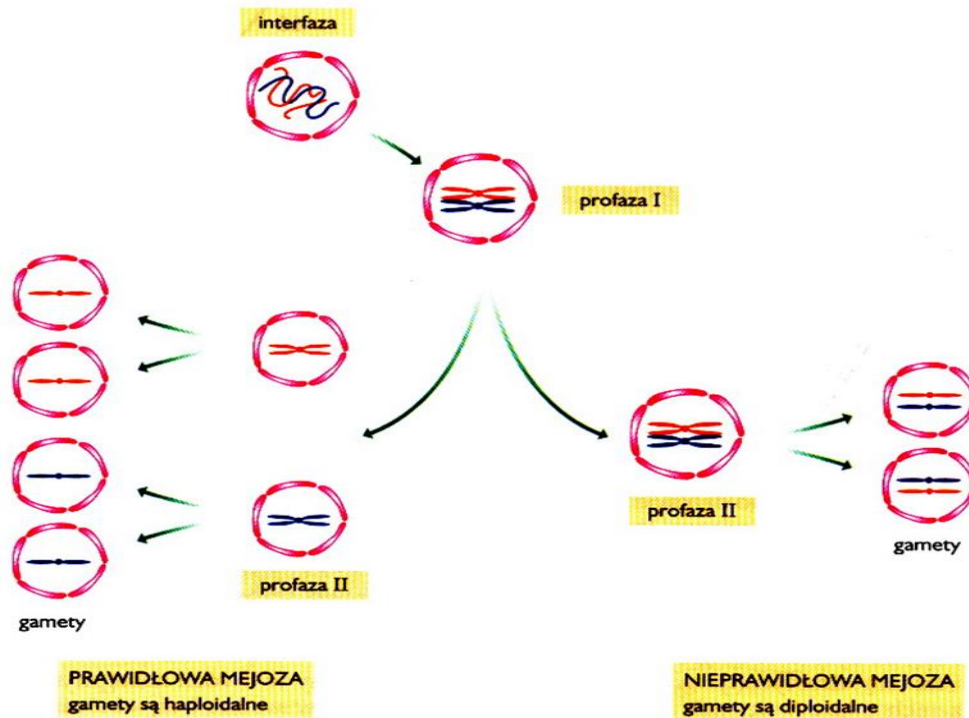
Najważniejszym mechanizmem powstawania nowych genów podczas ewolucji genomów jest duplikacja genów już istniejących:

1. Przez duplikację całego genomu
2. Przez duplikację pojedynczego chromosomu (trisomia 21)
3. Przez duplikację pojedynczego genu lub grupy genów

Duplikacja pojedynczego chromosomu

Przykłady

- a) trisomia 21(zespół Downa) - zwielokrotnienie jednego chromosomu
- b) autopoliploidyzacja – zwielokrotnienie całego garnituru chromosomów (rośliny)



Podłoże autopoliploidyzacji

Autopoliploidyzacja

Autopoliploidyzacja nie prowadzi bezpośrednio do zwiększenia liczby genów, gdyż początkowo powstaje organizm mający po prostu dodatkowe kopie każdego z genów, a nie nowe geny.

Autopoliploidy najczęściej spotykane są w świecie roślin, rzadko w świecie zwierząt.

Rośliny autopoliploidalne są najczęściej większe, bujniejsze, wykazują podwyższoną żywotność niż organizmy diploidalne, co wykorzystuje się w rolnictwie.



Rzepak poliploidalny

www.macroevolution.net

Allopoliploidyzacja

Allopoliploidyzacja - krzyżowanie międzygatunkowe

Allopoliploidyzacja jest jednym z ważniejszych procesów ewolucyjnych prowadzących do powstania **nowego gatunku**. Doskonałym modelem do śledzenia zmian towarzyszących temu procesowi są sztucznie tworzone gatunki allopoliploidalne. Zapewniają one wgląd w procesy towarzyszące formowaniu allopoliploidów podczas najwcześniejszych etapów ich ewolucji.

np. **Pszenżyto** = pszenica + żyto



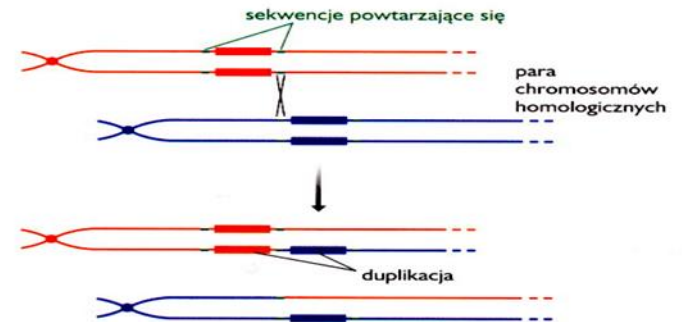
Pszenżyto

www.strube.pl

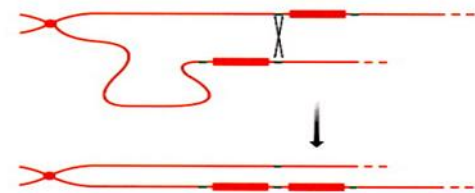
Duplikacja pojedynczego genu lub grupy genów

1. Nierównomierny crossing-over
2. Nierównomierna wymiana chromatyd siostrzanych
3. Replikacja DNA u bakterii
4. Poślizg replikacyjny

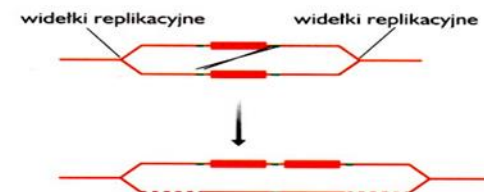
(A) Nierównomierny crossing-over



(B) Nierównomierna wymiana chromatyd siostrzanych

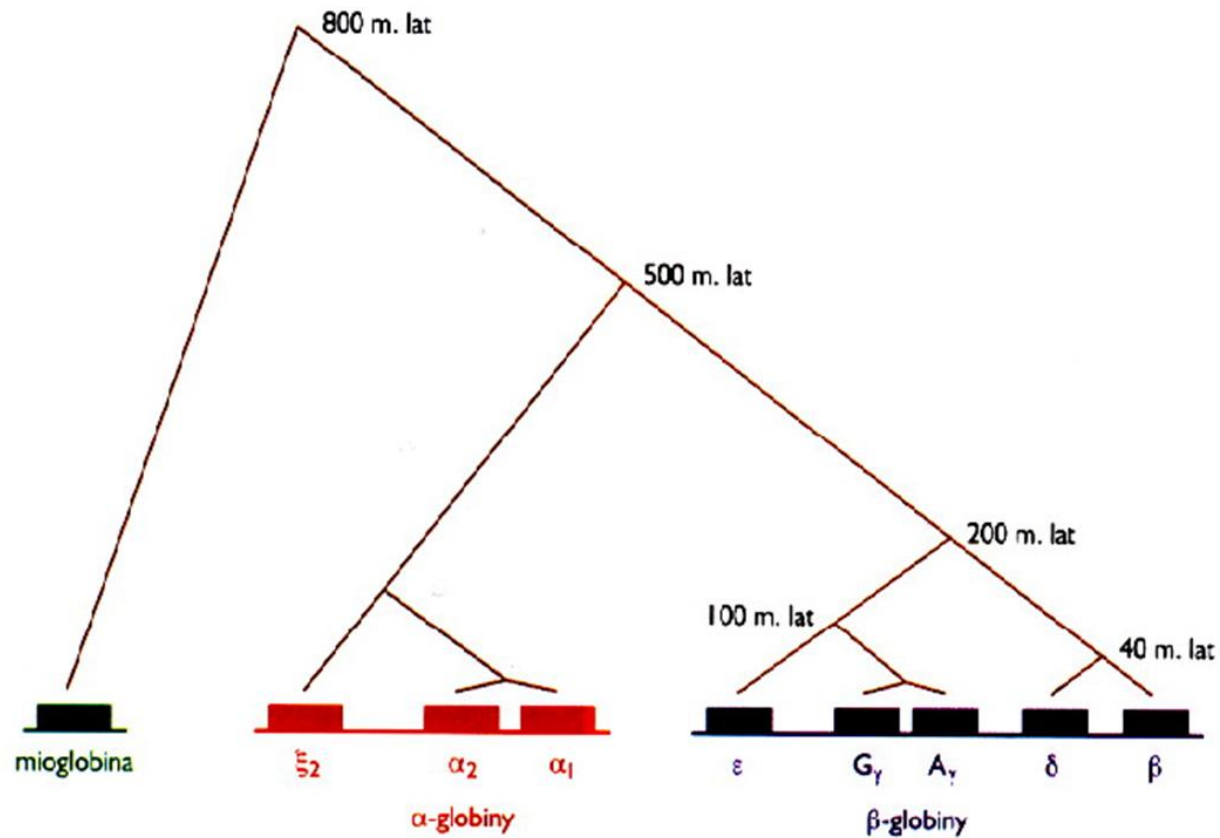


(C) Podczas replikacji DNA



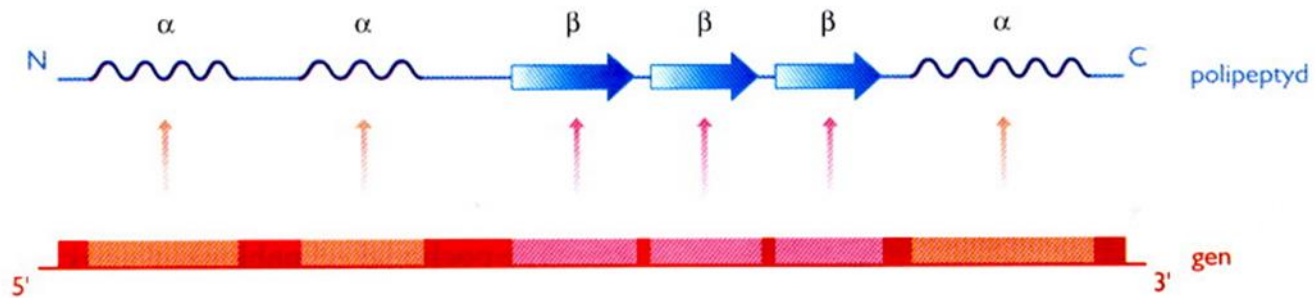
Modele duplikacji genów

Duplikacje genów w ewolucji rodzin ludzkich genów globin



W ewolucji genomu następują też rearanżacje istniejących genów

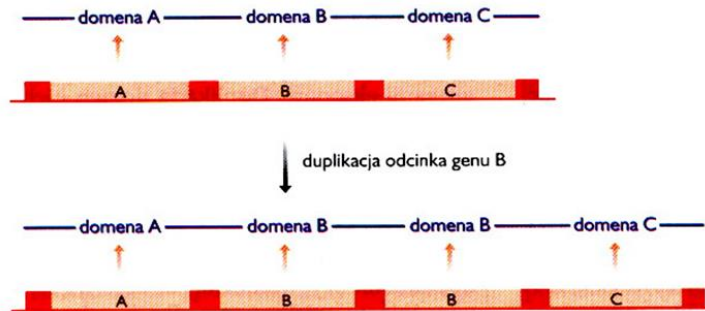
1. **Duplikacja domen** (nierównomierny crossing-over) – skutkiem większa stabilność białka
2. **Tasowanie domen** – skutkiem inna funkcja biochemiczna białka



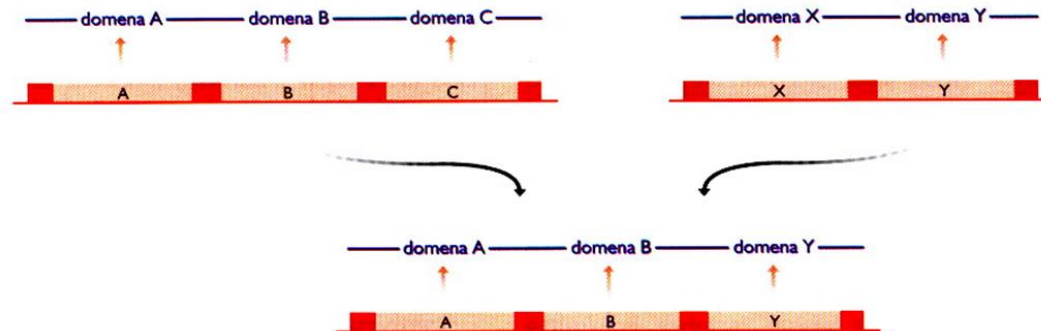
Domeny strukturalne są jednostkami w łańcuchu polipeptydowym kodowanym przez poszczególne fragmenty genu

Powstawanie nowych genów przez duplikację i tasowanie domen

(A) Duplikacja domen



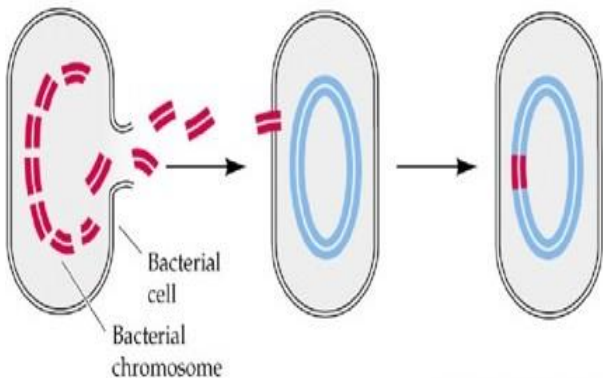
(B) Tasowanie domen



Nabywanie genów od innych gatunków

1. **Koniugacja** (bakterie) – proces odwracalny, mało trwały, głównie plazmidy.
2. **Transformacja** (bakterie) - wszystkie gatunki nabyły tą drogą min. kilka genów
3. **Poliploidyzacja** (rośliny)
4. **Transdukcja** (bakterie) – za pośrednictwem wirusów

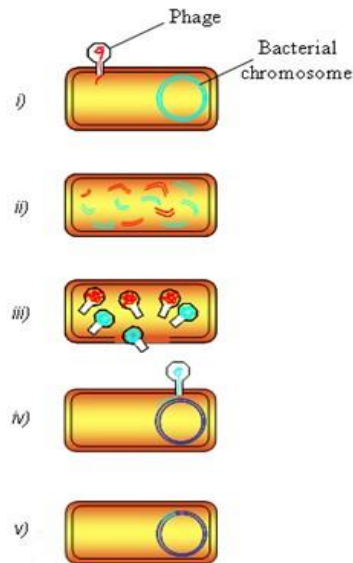
Transformacja



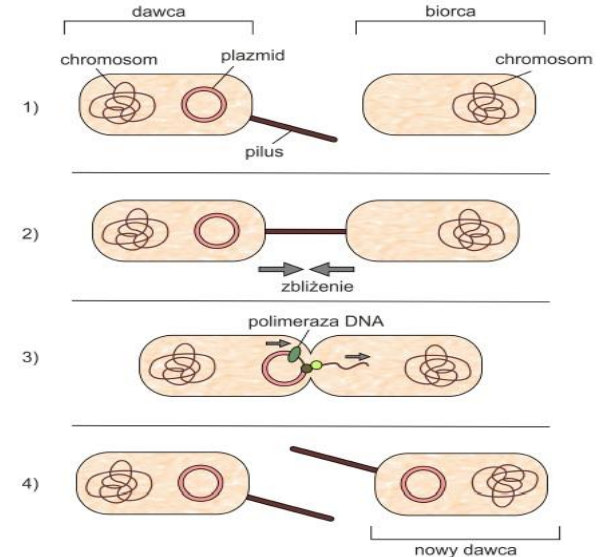
Permission pending from Sinauer Associates, Inc.

www.slideplayer.pl

Transdukcja

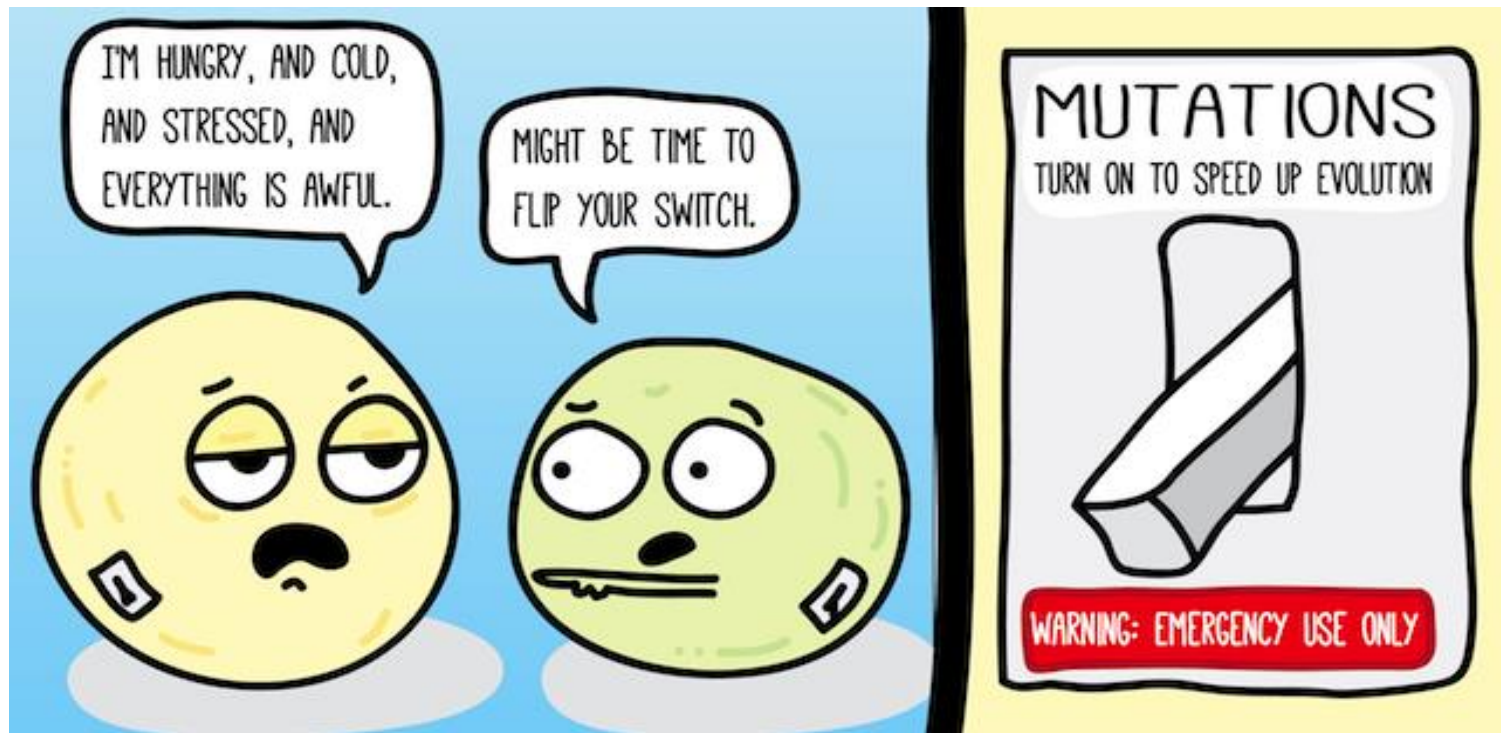


Koniugacja



www.e-biotechnologia.pl

Dziękuję za uwagę



Beatrice the Biologist