

**Mutacje** są to zmiany sekwencji nukleotydów w cząsteczce dna, mogące powstać na skutek różnych czynników:

- błędów replikacji nie naprawionych przez mechanizmy korekcyjne polimerazy DNA (**mutacje spontaniczne**);
- bezpośredniego lub pośredniego działania czynników mutagennych (**mutacje indukowane**). Mutagenami mogą być czynniki fizyczne (np. promieniowanie jonizujące lub UV), chemiczne (np. reaktywne formy tlenu, benzopiren, bromek etyldyny) bądź biologiczne (np. transpozony, wirusy).

Bez mechanizmów przeciwdziałania mutacjom genom nie byłby w stanie utrzymać swojej roli dłużej niż kilka godzin. Dlatego też w komórkach nieustannie działają **enzymy naprawy DNA**, których rolą jest minimalizowanie częstości pojawiania się mutacji. Część z nich działa bezpośrednio po replikacji, naprawiając błędy w nowo zsintetyzowanym DNA, natomiast inne wymieniają nieprawidłowe nukleotydy przed replikacją, zapobiegając tym samym ich powieleniu.



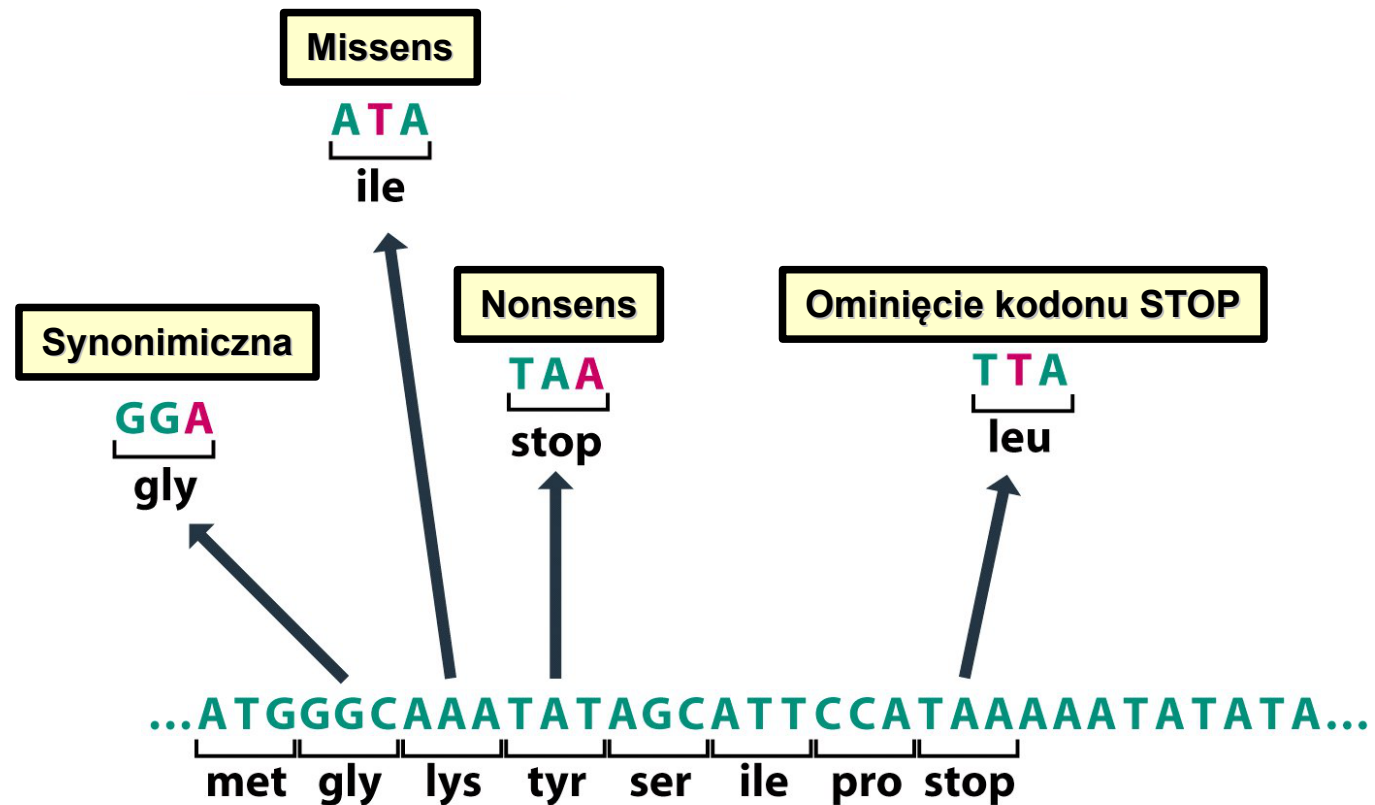
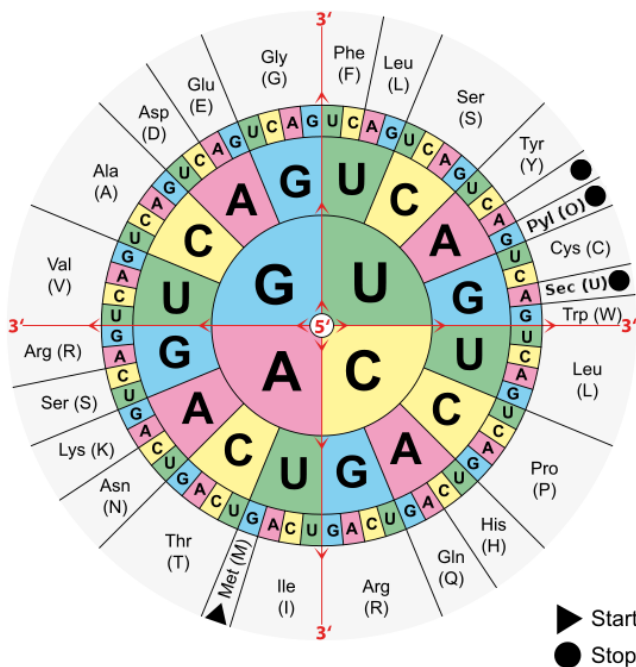
Ze względu na zasięg i charakter zmian sekwencji mutacje można podzielić na:

- **punktowe (substytucje)**, czyli zmiany pojedynczych nukleotydów, w tym:
  - **tranzycja** – zastąpienie jednej puryny przez inną purynę lub pirymidyny przez drugą pirymidynę (A↔G, C↔T);
  - **transwersja** – zastąpienie zasady purynowej przez pirymidynową lub odwrotnie (A↔C, A↔T, G↔C, G↔T);
  
- **insercje** – wstawienie przynajmniej jednego nukleotydu do sekwencji;
  
- **delecje** – usunięcie przynajmniej jednego nukleotydu z sekwencji;
  
- **chromosomowe**, dotyczące dużych fragmentów chromosomów:
  - **deficjencja** – utrata fragmentu chromosomu;
  - **duplikacja** – podwojenie fragmentu chromosomu;
  - **inwersja** – obrócenie fragmentu chromosomu o 180 stopni;
  - **translokacja** – przeniesienie fragmentu chromosomu na inny (nie homologiczny).

# Mutacje

Mutacje punktowe w obszarze genu, można podzielić na:

- **synonimiczne**, które zmieniają kodon w inny kodon tego samego aminokwasu;
- **niesynonimiczne**, zmieniające znaczenie kodonu:
  - **missens** – zmiana kodonu aminokwasu w kodon innego aminokwasu;
  - **nonsens** – zmiana kodonu aminokwasu w kodon terminujący translację (STOP);
  - **ominięcie kodonu terminującego** – zmiana kodonu STOP w kodon aminokwasu.

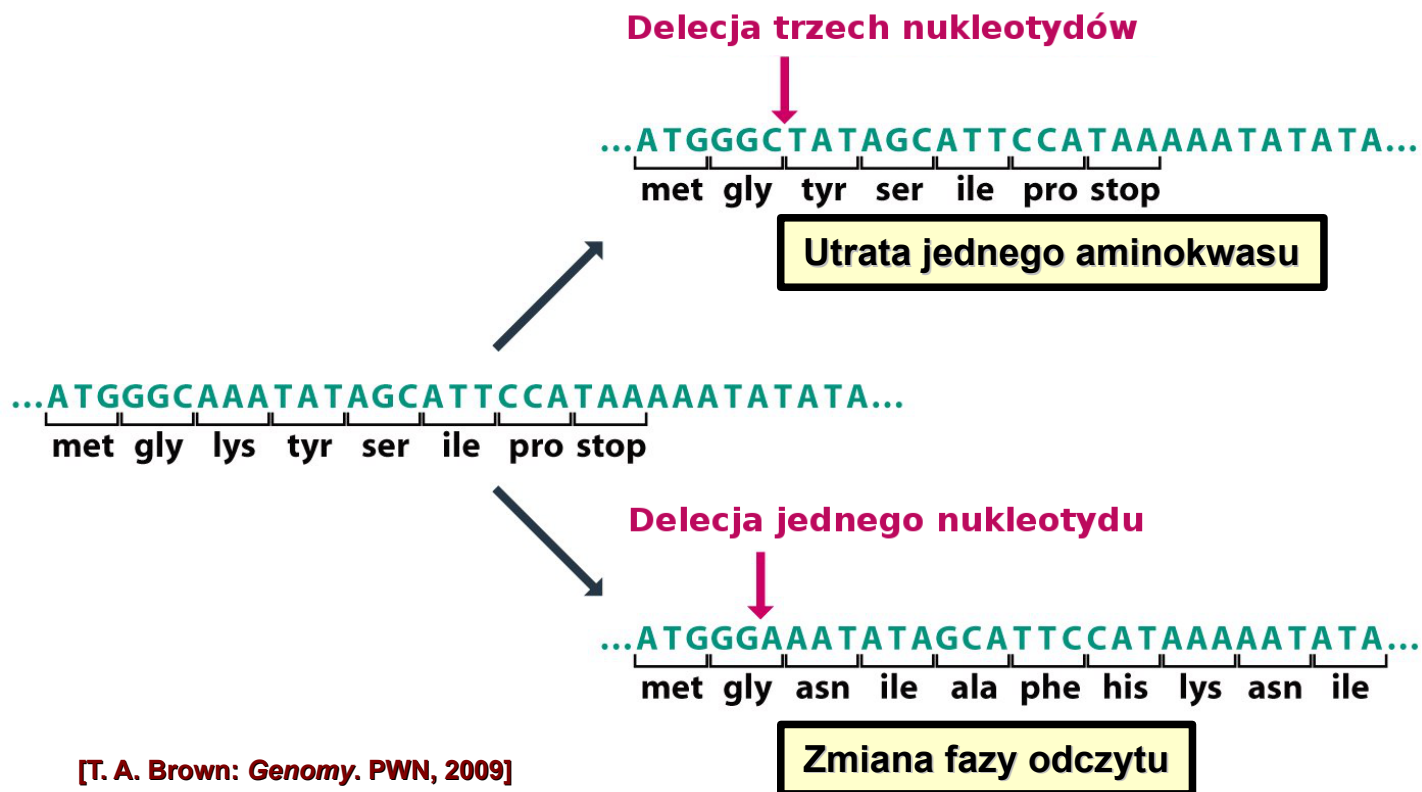
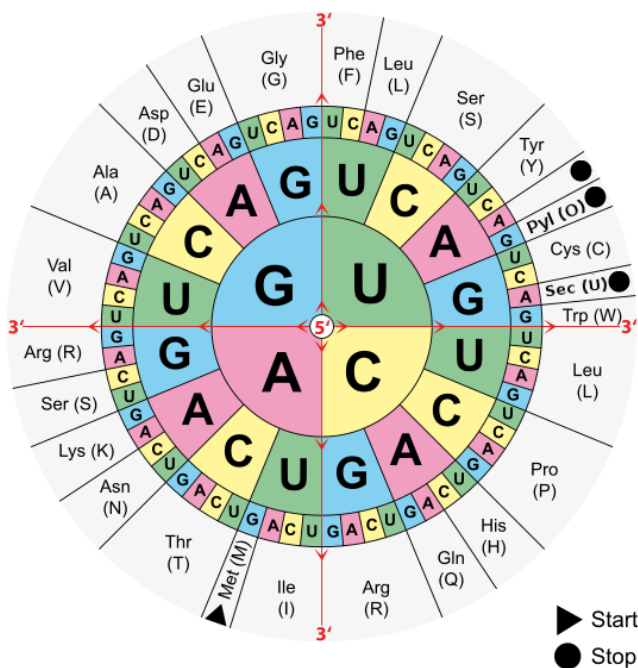


[T. A. Brown: *Genomy*. PWN, 2009]

# Mutacje

Insercje i delecje mogą w różny sposób wpływać na zdolność kodowania genu:

- **gdy dotyczą całych kodonów** powodują zmianę liczby aminokwasów w sekwencji kodowanego białka. Tego typu mutacje mogą nie wpłynąć znacząco na funkcje białka (pod warunkiem, że nie powodują zmian struktur drugorzędowych);
- **gdy nie są zgodne z podziałem na kodony** (np. gdy dotyczą pojedynczych nukleotydów) powodują przesunięcie fazy odczytu i całkowitą zmianę sekwencji aminokwasów kodowanych przez nukleotydy znajdujące się poniżej miejsca mutacji. Takie mutacje zwykle mają poważny wpływ na funkcję białka.



# Mutacje

Pośrednimi skutkami mutacji mogą być natychmiastowe lub opóźnione zmiany fenotypu organizmu. Ze względu na efekt fenotypowy wyróżnia się mutacje:

- **obojętne**, nie wpływające na zdolność organizmu do przeżycia;
- **niekorzystne**, upośledzające funkcje organizmu lub prowadzące do jego śmierci;
- **korzystne**, powodujące nabycie funkcji zwiększającej przystosowanie organizmu.

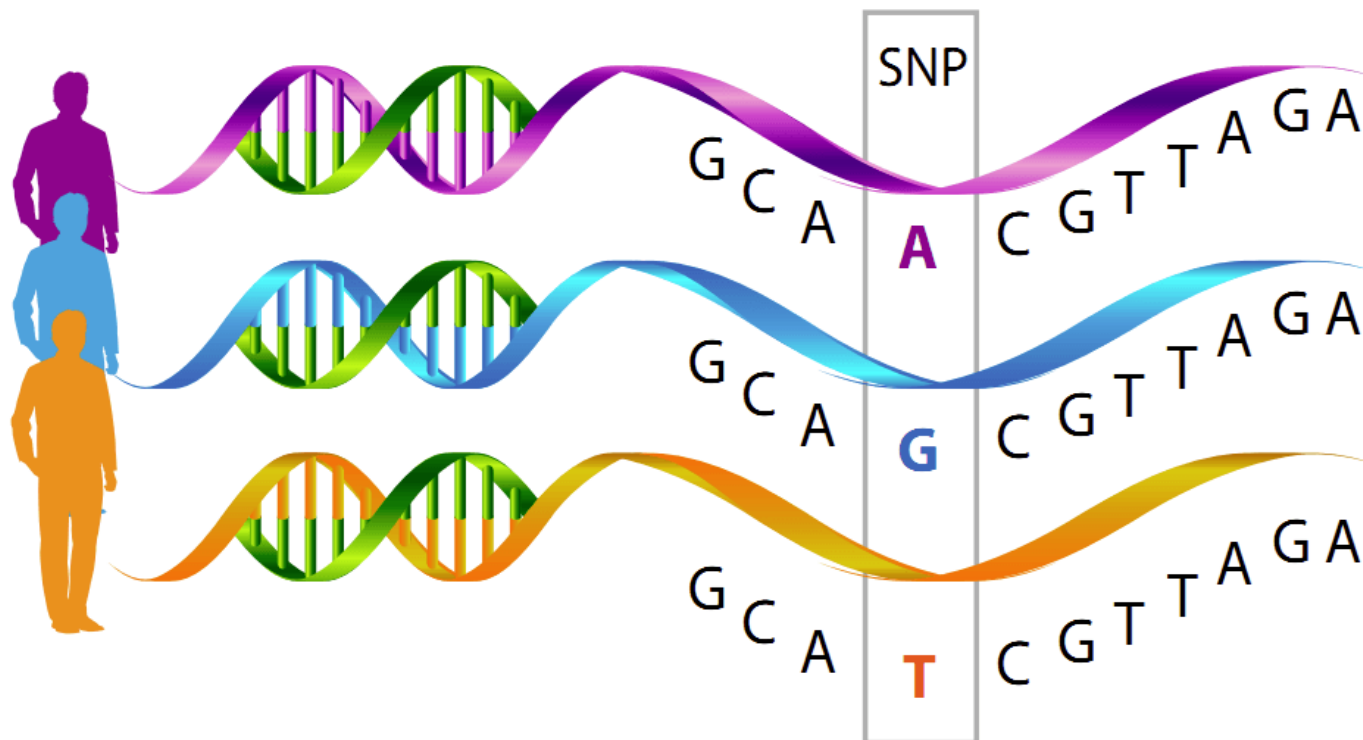
Mutacje nieprowadzące do śmierci organizmu mogą przyczynić się do **ewolucji genomu**. Aby tak się stało muszą zostać odziedziczone przez organizmy potomne:

- u organizmów jednokomórkowych wszystkie zmiany genomu (nie prowadzące do śmierci) są dziedziczone przez organizmy potomne;
- u organizmów wielokomórkowych jedynie mutacje zachodzące w komórkach linii płciowej są istotne z punktu widzenia ewolucji genomu. Zmiany w genomach komórek somatycznych mają jedynie znaczenie dla fenotypu danego organizmu.

# Mutacje

Utrwalenie mutacji na drodze dziedziczenia prowadzi do występowania różnic w sekwencjach DNA w ramach populacji. Zmianę występującą wystarczająco często, aby nie można było już mówić o mutacji osobniczej określa się **polimorfizmem**. Typowo jako próg przyjmuje się występowanie zmiany u nie mniej niż 1% populacji.

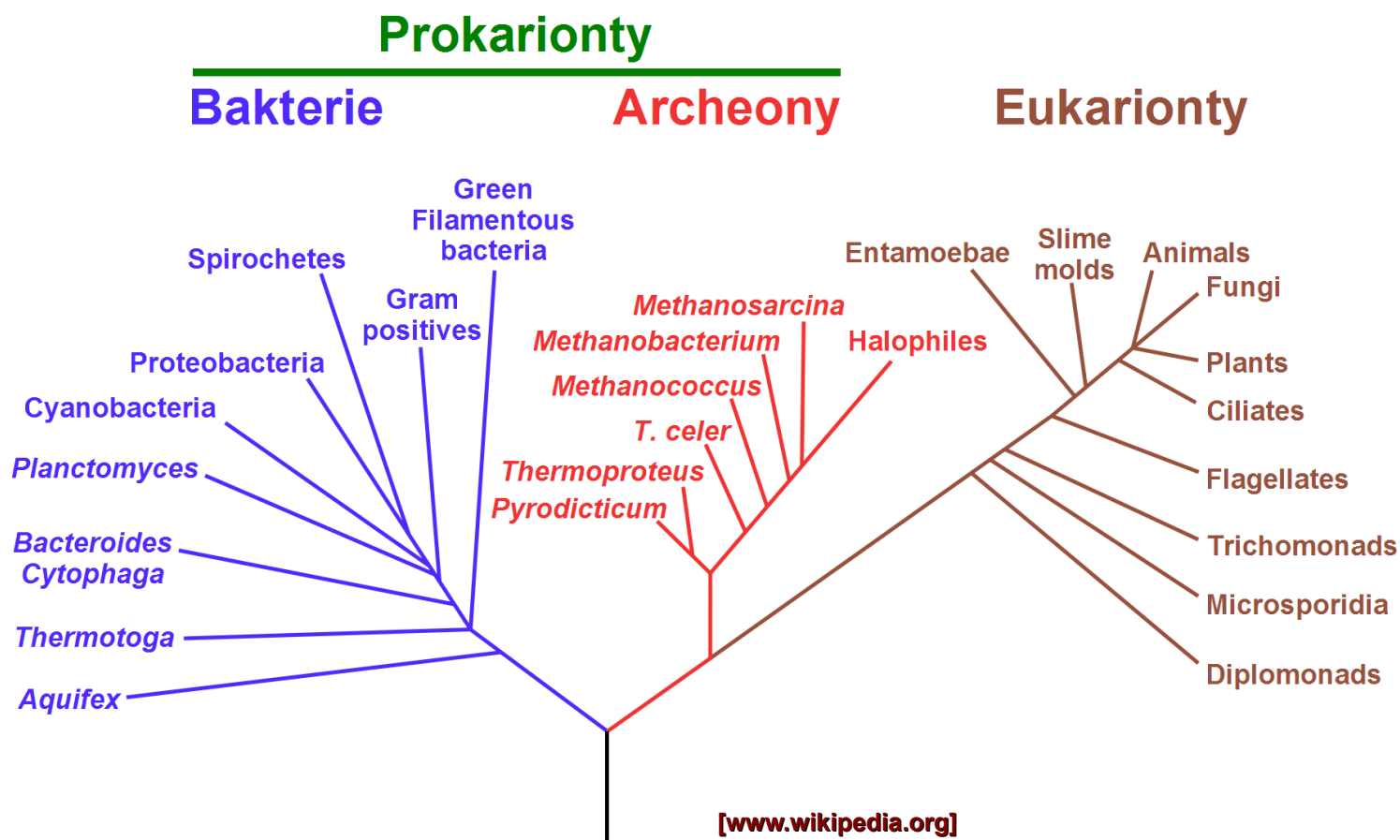
Przykładem tego zjawiska mogą być **polimorfizmy pojedynczych nukleotydów (Single Nucleotide Polymorphisms – SNP)**, których liczbę w ludzkim genomie szacuje się na wiele milionów.



# Ewolucja genomów

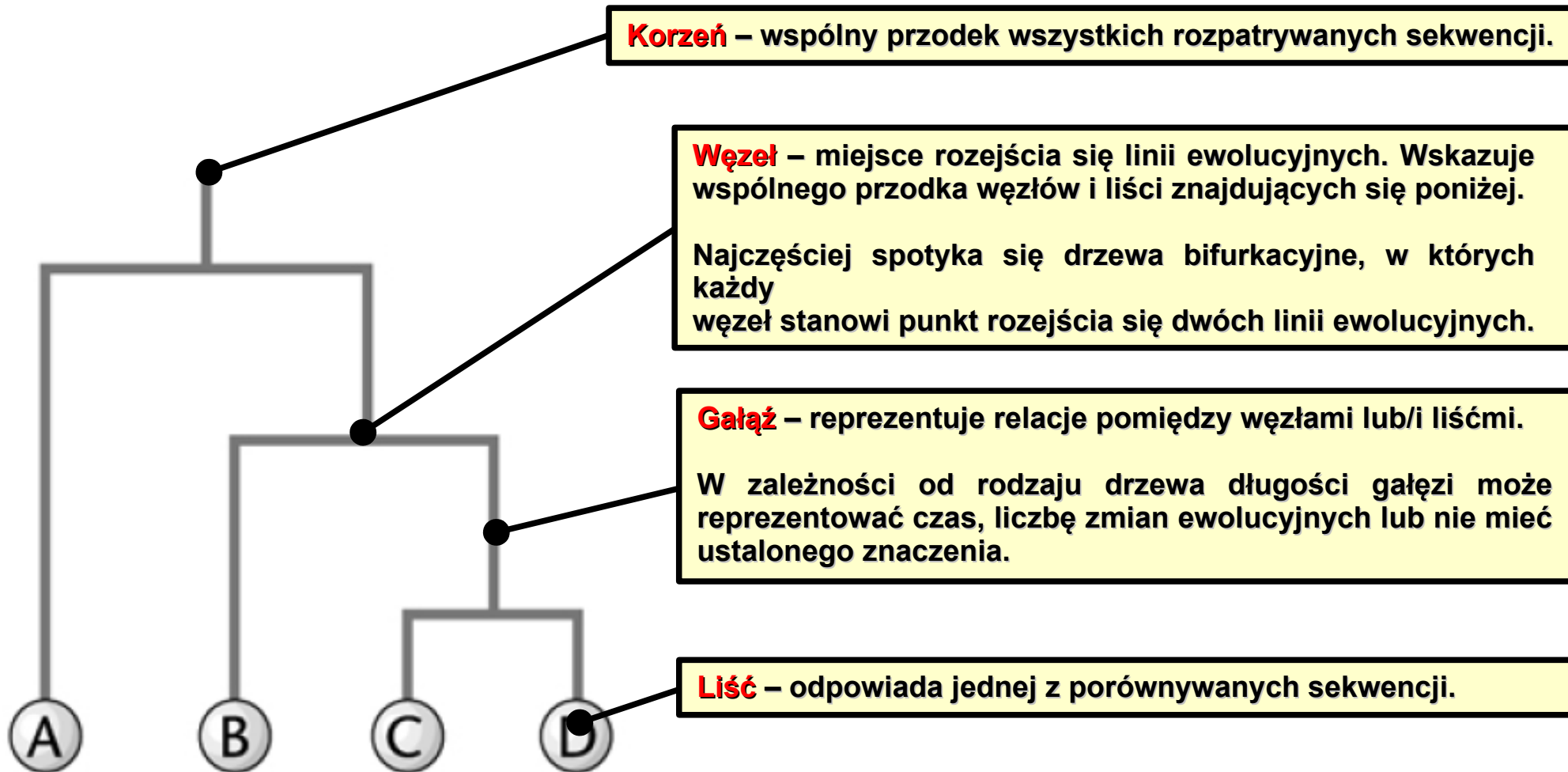
Wraz z ewolucją genomu następuje stopniowa akumulacja mutacji. Można więc oczekiwać, że genomy, których linie ewolucyjne rozdzieliły się w niedalekiej przeszłości będą różnić się od siebie w mniejszym stopniu od genomów, których wspólny przodek jest bardziej odległy w czasie.

Założenie to leży u podstaw **filogenetyki molekularnej**, dziedziny wiedzy zajmującej się określaniem związków ewolucyjnych pomiędzy organizmami poprzez badanie podobieństwa sekwencji DNA lub białek.



# Drzewa filogenetyczne

**Drzewo filogenetyczne** jest grafem acyklicznym przedstawiającym ewolucyjne zależności pomiędzy porównywanymi gatunkami lub sekwencjami.

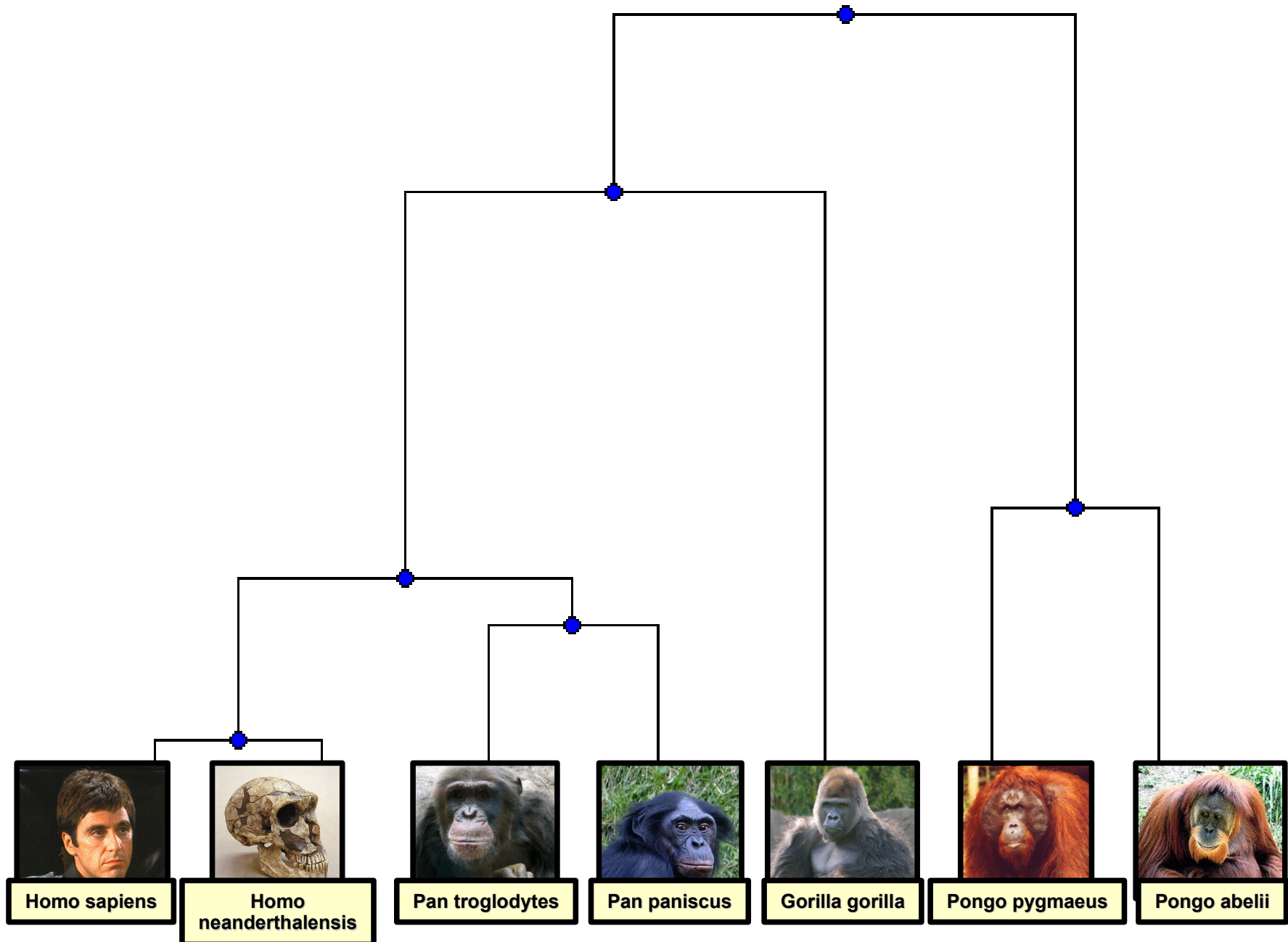


Może wyznaczyć różne drzewa o tej samej **topologii** (sposobie rozgałęziania się i długości gałęzi). Każda forma drzewa zachowująca topologię jest właściwa.



# Drzewa filogenetyczne

## Drzewo filogenetyczne rodziny człowiekowatych (*Hominidae*)

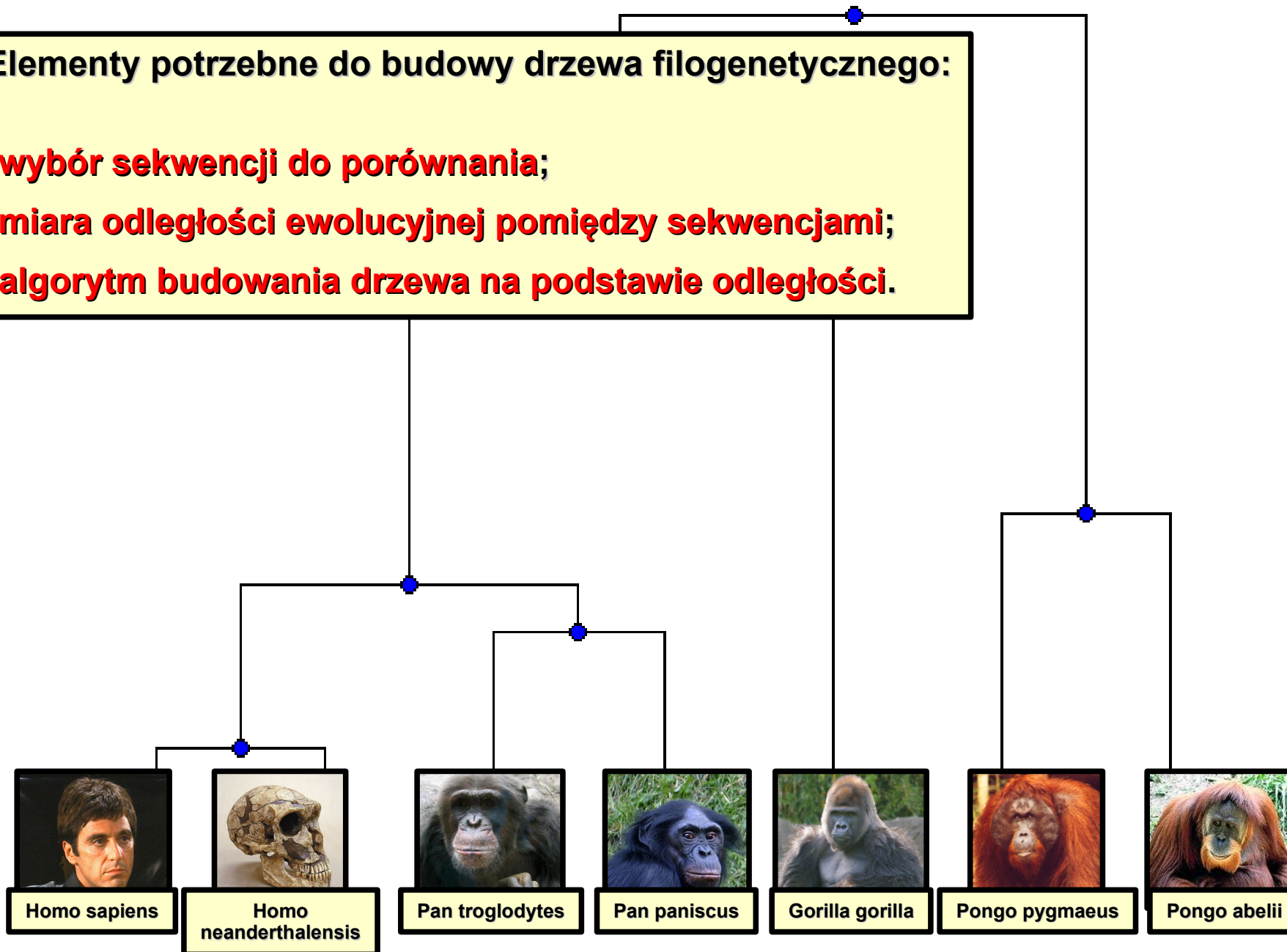


# Drzewa filogenetyczne

## Drzewo filogenetyczne rodziny człowiekowatych (*Hominidae*)

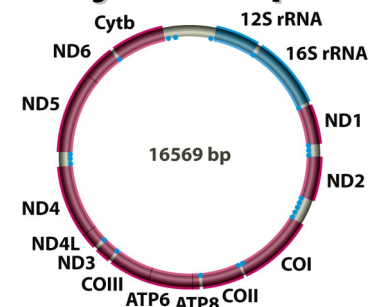
Elementy potrzebne do budowy drzewa filogenetycznego:

- **wybór sekwencji do porównania;**
- **miara odległości ewolucyjnej pomiędzy sekwencjami;**
- **algorytm budowania drzewa na podstawie odległości.**



# Drzewa filogenetyczne: wybór sekwencji

Fragment sekwencji mitochondrialnego DNA charakteryzujący się dużym tempem akumulacji mutacji (Hypervariable Region 2 – HVR2).



## Homo sapiens

```
TTTTCGTCTGGGGGGTGTGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCCTATGTCGCAGTA  
TCTGTCTTTGATTCCTGCCTCATCCTATTATTTATCGCACCTACGTTCAATATTACAGGCGAACATACTT  
ACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATAACAATTGAATGTCTGCACAGCCGCTT  
TCCACACAGACATCATAACAAAAAATTTCCACCAAACCCCCCTCCCCCGCTTCTGGCCACAGCACTTA  
AACACATCTCTGCCAAACCCCAAAAACAAAGAACCCTAACACCAGCCTAACCCAGATT
```

## Homo neanderthalensis

```
TTTTCGTCTGGGGGGTGTGCACGCGATAGCATTGCGAGACGCTGGAGCCGGAGCACCCCTATGTCGCAGTA  
TCTGTCTTTGATTCCTGCCCCATTCCATTATTTATCGCACCTACGTTCAATATTACAGGCGAGCATACTT  
ACTAAAGTGTGTTAATTAATTAATGCTTGTAGGACATAATAATAACGACTAAATGTCTGCACAGCTGCTT  
TCCACACAGACATCATAACAAAAAATTTCCACCAAACCCCCCTTTCCTCCCCCGCTTCTGGCCACAGCACT  
TAAACACATCTCTGCCAAACCCCAAAAACAAAGAACCCTAACACCAGCCTAACCCAGACTTCAAAT
```

## Pan troglodytes

```
TTTTCGTCTGGGGGGTGTGCACGCGATAGCATTGCGAAACGCTGGCCCCGGAGCACCCCTATGTCGCAGTA  
TCTGTCTTTGATTCCTGCCCCATTGTATTATTTATCGCACCTACGTTCAATATTACGACCTAGCATAACCT  
ACTAAAGTGTGTTGATTAATTAATGCTTGCAGGACATAACAACAGCAGCAAATGCTCACATAACTGCTT  
TCCACACCAACATCATAACAAAAAATTTCCACCAAACCCCCCTTCCCCCGGCCACAGCACTCAAACAAA  
TCTCTGCCAAACCCCAAAAACAAAGAACCAGACGCCAGCCTAGCCAGACTTCAAAT
```

# Drzewa filogenetyczne: odległość ewolucyjna

**Metryką (funkcją odległości)** w zbiorze  $X$  może być funkcja  $d: X \rightarrow [0, \infty)$ , która dla dowolnych elementów  $x, y, z$  tego zbioru spełnia warunki:

$$d(x, y) = 0 \iff x = y$$

$$d(x, y) > 0 \iff x \neq y$$

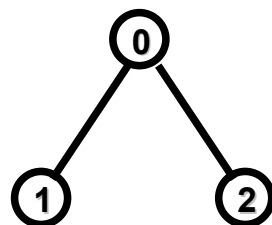
$$d(x, y) = d(y, x)$$

$$d(x, z) \leq d(x, y) + d(y, z)$$

W przypadku odległości ewolucyjnej dodatkowymi, pożądanymi cechami są:

- liniowa zależność od czasu: np. dwukrotnie dłuższemu okresowi czasu od rozejścia się linii ewolucyjnych odpowiada dwukrotny wzrost odległości;

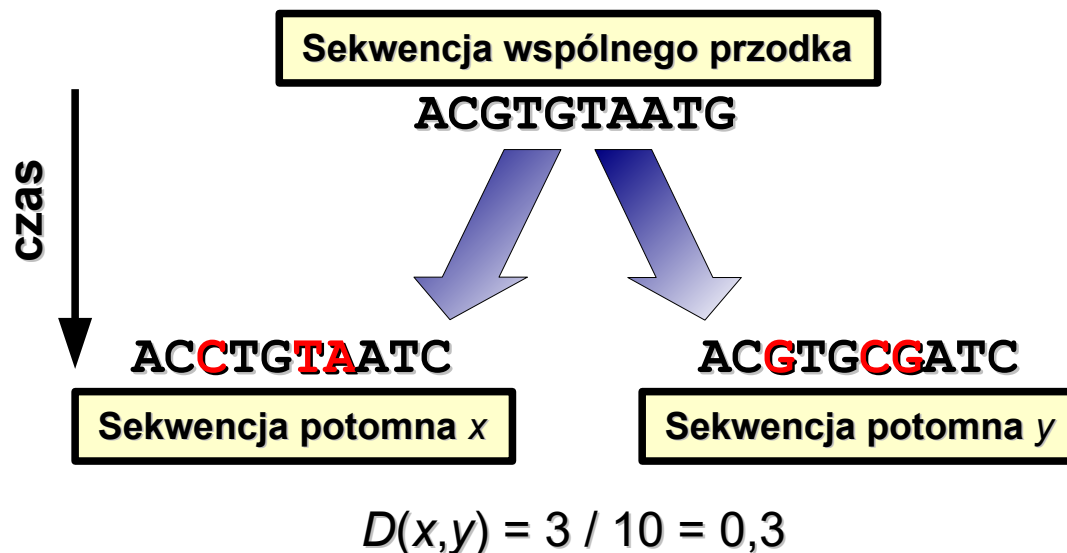
- addytywność:



$$d(1, 2) = d(1, 0) + d(0, 2)$$

# Drzewa filogenetyczne: odległość ewolucyjna

Najprostszą miarą odległości pomiędzy sekwencjami  $x$  i  $y$  jest  $D(x,y)$ , liczone jako odsetek pozycji, na których w dopasowanych globalnie sekwencjach występują różne nukleotydy.



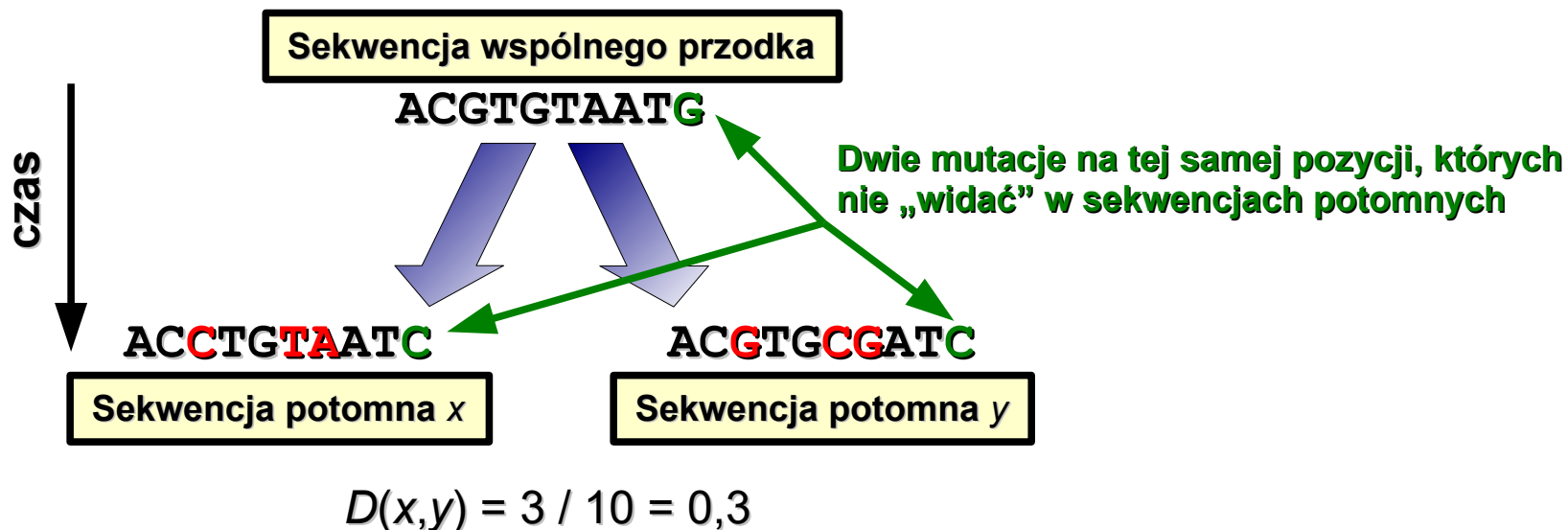
Tak zdefiniowana odległość ma jednak pewną wadę: nie uwzględnia możliwości wystąpienia wielokrotnych mutacji na tej samej pozycji.

Ze względu na wielokrotne zmiany tych samych pozycji, wraz ze wzrostem czasu od rozejścia się linii ewolucyjnych liczba widocznych różnic pomiędzy porównywanymi sekwencjami coraz wyraźniej niedoszacowuje rzeczywistej liczby mutacji.

Po odpowiednio długim okresie czasu obie sekwencje potomne staną się losowe wobec sekwencji przodka. W efekcie, przy założeniu równych częstości występowania nukleotydów,  $D$  dążyć będzie do wartości  $\frac{3}{4}$ .

# Drzewa filogenetyczne: odległość ewolucyjna

Najprostszą miarą odległości pomiędzy sekwencjami  $x$  i  $y$  jest  $D(x,y)$ , liczone jako odsetek pozycji, na których w dopasowanych globalnie sekwencjach występują różne nukleotydy.



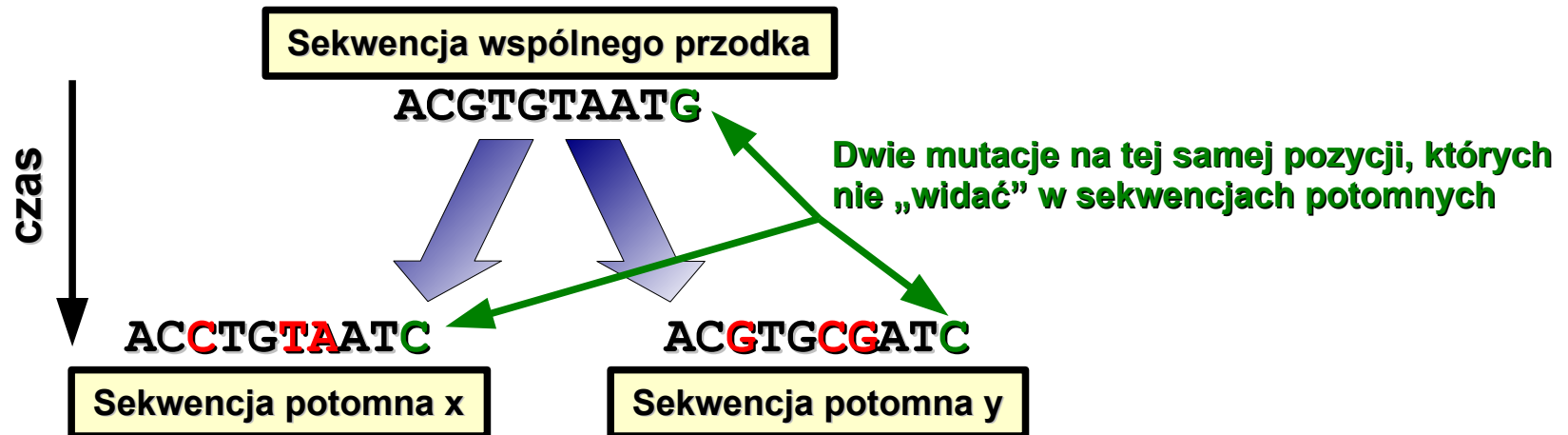
Tak zdefiniowana odległość ma jednak pewną wadę: nie uwzględnia możliwości wystąpienia wielokrotnych mutacji na tej samej pozycji.

Ze względu na wielokrotne zmiany tych samych pozycji, wraz ze wzrostem czasu od rozejścia się linii ewolucyjnych liczba widocznych różnic pomiędzy porównywanymi sekwencjami coraz wyraźniej niedoszacowuje rzeczywistej liczby mutacji.

Po odpowiednio długim okresie czasu obie sekwencje potomne staną się losowe wobec sekwencji przodka. W efekcie, przy założeniu równych częstości występowania nukleotydów,  $D$  dążyć będzie do wartości  $\frac{3}{4}$ .

# Drzewa filogenetyczne: odległość ewolucyjna

Najprostszą miarą odległości pomiędzy sekwencjami  $x$  i  $y$  jest  $D(x,y)$ , liczone jako odsetek pozycji, na których w dopasowanych globalnie sekwencjach występują różne nukleotydy.

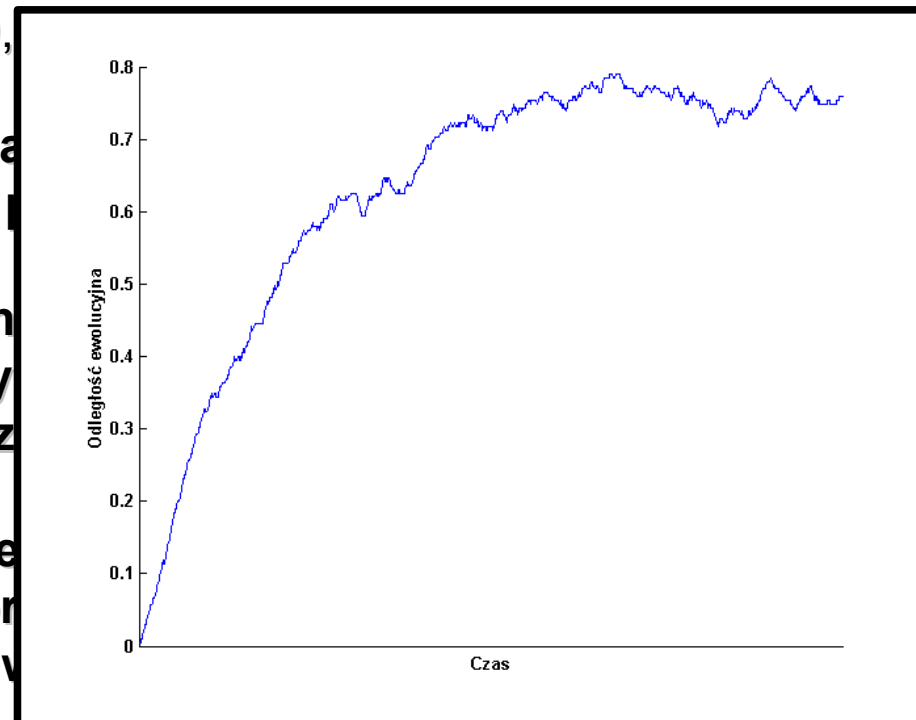


$$D(x,y) = 3 / 10 = 0,3$$

Tak zdefiniowana odległość ma jednak pewną wada. Ze względu na wielokrotne wystąpienia wielokrotnych mutacji na tej samej pozycji, liczba widocznych różnic w sekwencjach coraz wyraźniej niedoszacowuje rzeczywistą odległość ewolucyjną.

Po odpowiednio długim okresie czasu obie sekwencje różnią się od sekwencji przodka. W efekcie, przy długim czasie występowania nukleotydów,  $D$  dążyć będzie do wartości 1,0.

Po odpowiednio długim okresie czasu obie sekwencje różnią się od sekwencji przodka. W efekcie, przy długim czasie występowania nukleotydów,  $D$  dążyć będzie do wartości 1,0.



# Drzewa filogenetyczne: odległość ewolucyjna

**Odległość ewolucyjna**  $d(x,y)$ , definiowana jest zwykle jako średnia liczba mutacji przypadająca na pojedynczą pozycję pary sekwencji.

Jeżeli mutacje zachodzą losowo i w stałym tempie, to tak zdefiniowana odległość jest wprost proporcjonalna do czasu. Jednocześnie jest ona addytywna.

Odległość ewolucyjna jest wyznaczana w oparciu o **model ewolucji sekwencji** uwzględniający częstości wystąpień nukleotydów ( $\pi_A, \pi_G, \pi_C, \pi_T$ ) oraz tempo mutacji ( $\alpha_{AG}, \alpha_{AC}, \alpha_{AT}, \alpha_{GC}, \alpha_{GT}, \alpha_{TC}$ ).

Parametry modelu mogą zostać zapisane w postaci **macierzy współczynników tempa mutacji**:

	$A$	$G$	$C$	$T$
$A$	*	$\alpha_{AG} \pi_G$	$\alpha_{AC} \pi_C$	$\alpha_{AT} \pi_T$
$G$	$\alpha_{AG} \pi_A$	*	$\alpha_{GC} \pi_C$	$\alpha_{GT} \pi_T$
$C$	$\alpha_{AC} \pi_A$	$\alpha_{GC} \pi_G$	*	$\alpha_{CT} \pi_T$
$T$	$\alpha_{AT} \pi_A$	$\alpha_{GT} \pi_G$	$\alpha_{CT} \pi_C$	*

$$* = r_{ii} = - \sum_{j \neq i} r_{ij}$$



## MODEL JUKESA-CANTORA

Założenia:

$$\pi_A = \pi_G = \pi_C = \pi_T = \frac{1}{4}$$

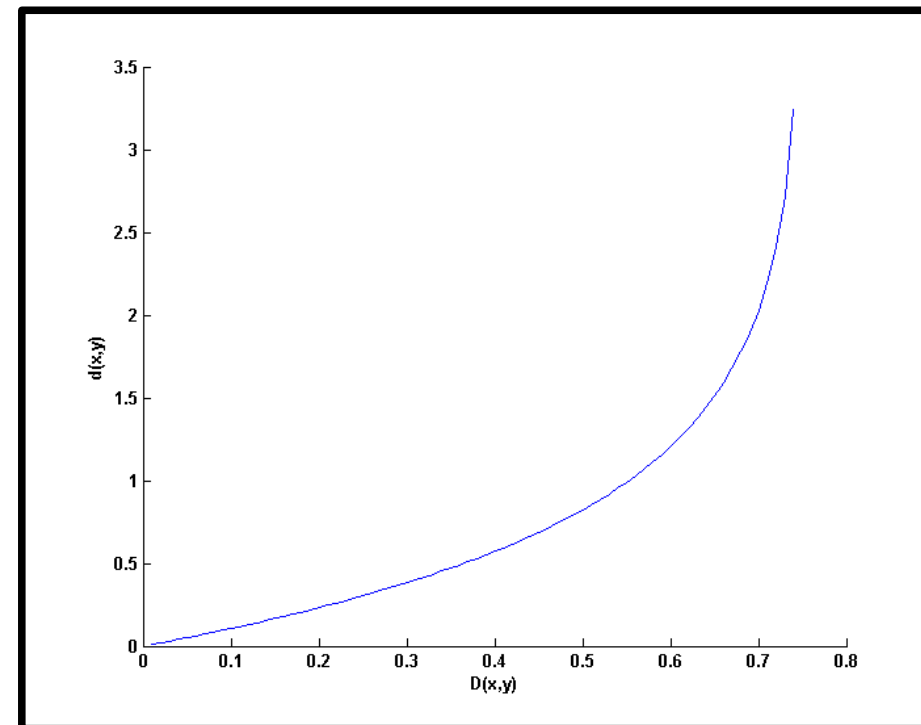
$$\alpha_{AG} = \alpha_{AC} = \alpha_{AT} = \alpha_{GC} = \alpha_{GT} = \alpha_{CT} = \alpha$$

Macierz tempa mutacji:

	<i>A</i>	<i>G</i>	<i>C</i>	<i>T</i>
<i>A</i>	$-3 \cdot \frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$
<i>G</i>	$\frac{\alpha}{4}$	$-3 \cdot \frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$
<i>C</i>	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$-3 \cdot \frac{\alpha}{4}$	$\frac{\alpha}{4}$
<i>T</i>	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$-3 \cdot \frac{\alpha}{4}$

Odległość ewolucyjna:

$$d(x, y) = -\frac{3}{4} \ln\left(1 - \frac{4}{3} D(x, y)\right)$$



## MODEL JUKESA-CANTORA

Założenia:

$$\pi_A = \pi_G = \pi_C = \pi_T = \frac{1}{4}$$

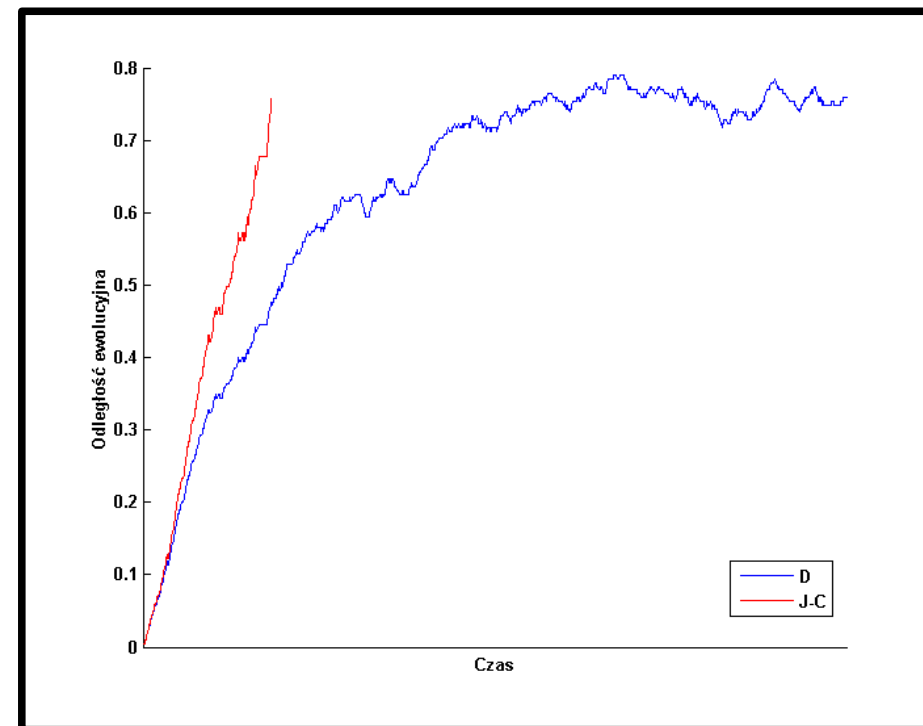
$$\alpha_{AG} = \alpha_{AC} = \alpha_{AT} = \alpha_{GC} = \alpha_{GT} = \alpha_{CT} = \alpha$$

Macierz tempa mutacji:

	$A$	$G$	$C$	$T$
$A$	$-3 \cdot \frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$
$G$	$\frac{\alpha}{4}$	$-3 \cdot \frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$
$C$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$-3 \cdot \frac{\alpha}{4}$	$\frac{\alpha}{4}$
$T$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$\frac{\alpha}{4}$	$-3 \cdot \frac{\alpha}{4}$

Odległość ewolucyjna:

$$d(x, y) = -\frac{3}{4} \ln\left(1 - \frac{4}{3} D(x, y)\right)$$



## MODEL KIMURY

Założenia:

$$\pi_A = \pi_G = \pi_C = \pi_T = \frac{1}{4}$$

$$\alpha_{AG} = \alpha_{CT} = \alpha \quad \alpha_{AC} = \alpha_{AT} = \alpha_{GC} = \alpha_{GT} = \beta$$

Macierz tempa mutacji:

	A	G	C	T
A	$-\frac{1}{4} \cdot (\alpha + 2\beta)$	$\frac{\alpha}{4}$	$\frac{\beta}{4}$	$\frac{\beta}{4}$
G	$\frac{\alpha}{4}$	$-\frac{1}{4} \cdot (\alpha + 2\beta)$	$\frac{\beta}{4}$	$\frac{\beta}{4}$
C	$\frac{\beta}{4}$	$\frac{\beta}{4}$	$-\frac{1}{4} \cdot (\alpha + 2\beta)$	$\frac{\alpha}{4}$
T	$\frac{\beta}{4}$	$\frac{\beta}{4}$	$\frac{\alpha}{4}$	$-\frac{1}{4} \cdot (\alpha + 2\beta)$

Odległość ewolucyjna:

$$d(x, y) = -\frac{1}{2} \ln(1 - 2 \cdot S(x, y) - V(x, y)) - \frac{1}{4} \ln(1 - 2 \cdot V(x, y))$$

gdzie:  $S(x, y)$  to liczba tranzycji ( $A \leftrightarrow G$ ,  $C \leftrightarrow T$ ) odniesiona do długości sekwencji  
 $V(x, y)$  to liczba transwersji ( $A \leftrightarrow C$ ,  $A \leftrightarrow T$ ,  $G \leftrightarrow C$ ,  $G \leftrightarrow T$ ) odniesiona do długości sekwencji

$$S(x, y) + V(x, y) = D(x, y)$$

# Drzewa filogenetyczne: tworzenie drzewa

Jednym z często spotykanych sposobów tworzenia drzewa filogenetycznego jest użycie **algorytmów aglomeracyjnej (wstępującej) klasteryzacji hierarchicznej**. Działają one w oparciu o symetryczną **macierz odległości**, której element  $d_{ij}$  reprezentuje odległość ewolucyjną  $d(i,j)$  pomiędzy  $i$ -tą a  $j$ -tą sekwencją:

$$D = \begin{bmatrix} 0 & d_{12} & d_{13} & \cdots & d_{1N} \\ d_{21} & 0 & d_{23} & \cdots & d_{2N} \\ d_{31} & d_{32} & 0 & \cdots & d_{3N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ d_{N1} & d_{N2} & d_{N3} & \cdots & 0 \end{bmatrix} ; \quad d_{ij} = d_{ji} ; \quad d_{ii} = 0$$

Ogólny schemat działania algorytmów aglomeracyjnej klasteryzacji hierarchicznej:

- Inicjalizacja: utworzenie  $N$  klastrów odpowiadających liściom drzewa (sekwencjom)
- Iteracyjne powtarzanie dwóch kroków aż do momentu gdy pozostanie jeden klastery:
  1. połączenie klastrów  $A$  i  $B$ , pomiędzy którymi jest najmniejsza odległość w jeden klastery  $A \cup B$  (zawierający elementy obu łączonych klastrów). Towarzyszy temu utworzenie nowego węzła w drzewie (długość gałęzi jest zależna od algorytmu)
  2. usunięcie z macierzy odległości wierszy i kolumn klastrów  $A$  i  $B$  oraz dodanie kolumny i wiersza odpowiadających połączonemu klastrowi  $A \cup B$  (sposób wyznaczania odległości pomiędzy nowym klastrem a pozostałymi zależy od konkretnego algorytmu)

# Drzewa filogenetyczne: tworzenie drzewa algorytmem UPGMA

W algorytmie **średnich połączeń (UPGMA – Unweighted Pair Group Method with Arithmetic Mean)** odległość  $d(A,B)$  pomiędzy klastrami  $A$  i  $B$  definiowana jest jako średnia arytmetyczna odległości  $d(a,b)$  między wszystkimi parami elementów należących do różnych klastrów:

$$d(A, B) = \frac{1}{|A| + |B|} \sum_{a \in A} \sum_{b \in B} d(a, b)$$

Schemat działania algorytmu UPGMA jest taki sam jak ten przedstawiony wcześniej, z dwoma uwagami szczegółowymi:

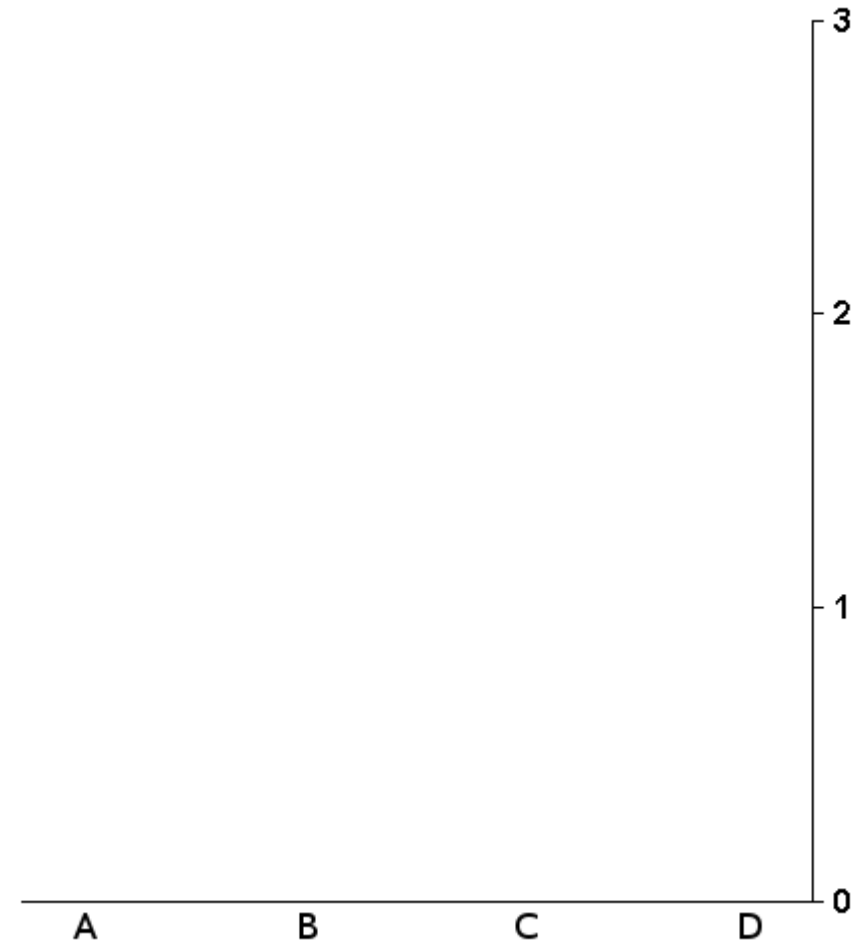
- połączeniu klastrów towarzyszy utworzenie węzła na wysokości równej połowie odległości pomiędzy łączonymi klastrami
- aktualizacja odległości  $d(C, (A \cup B))$  pomiędzy nowo utworzonym klastrem  $A \cup B$  a klastrem  $C$  następuje zgodnie z zależnością:

$$d(C, (A \cup B)) = \frac{|A|}{|A| + |B|} d(C, A) + \frac{|B|}{|A| + |B|} d(C, B)$$

# Drzewa filogenetyczne: tworzenie drzewa algorytmem UPGMA

## INICJALIZACJA

	A	B	C	D
A	0			
B	2	0		
C	6	6	0	
D	6	6	4	0

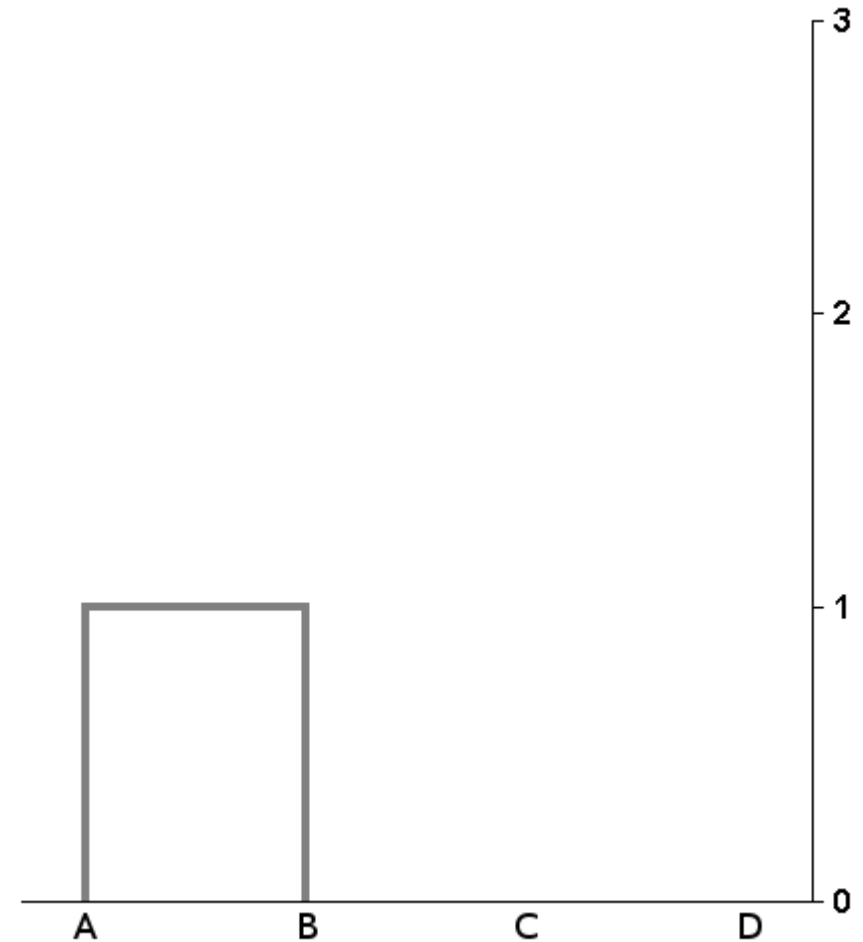


# Drzewa filogenetyczne: tworzenie drzewa algorytmem UPGMA

## ITERACJA 1

	A	B	C	D
A	0			
B	2	0		
C	6	6	0	
D	6	6	4	0

	AB	C	D
AB	0		
C		0	
D		4	0



# Drzewa filogenetyczne: tworzenie drzewa algorytmem UPGMA

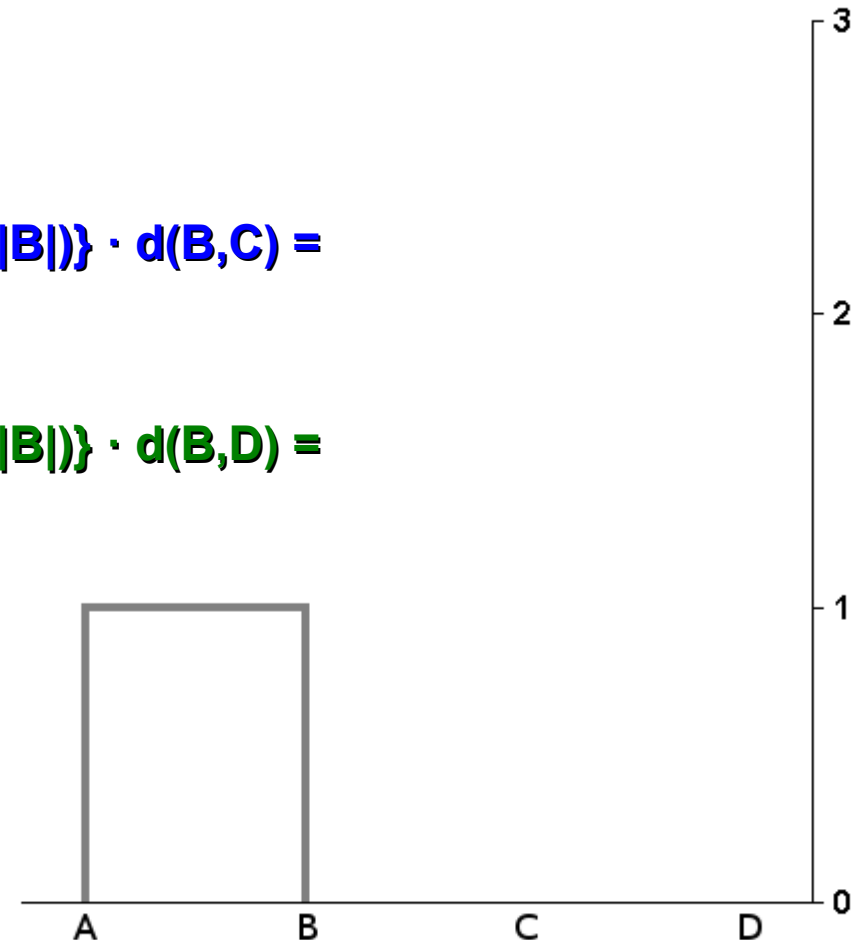
## ITERACJA 1

	A	B	C	D
A	0			
B	2	0		
C	6	6	0	
D	6	6	4	0

	AB	C	D
AB	0		
C	6	0	
D	6	4	0

$$\begin{aligned}d(A \cup B, C) &= \{|A| / (|A| + |B|)\} \cdot d(A, C) + \{|B| / (|A| + |B|)\} \cdot d(B, C) = \\ &= \{1 / (1+1)\} \cdot 6 + \{1 / (1+1)\} \cdot 6 = 6\end{aligned}$$

$$\begin{aligned}d(A \cup B, D) &= \{|A| / (|A| + |B|)\} \cdot d(A, D) + \{|B| / (|A| + |B|)\} \cdot d(B, D) = \\ &= \{1 / (1+1)\} \cdot 6 + \{1 / (1+1)\} \cdot 6 = 6\end{aligned}$$



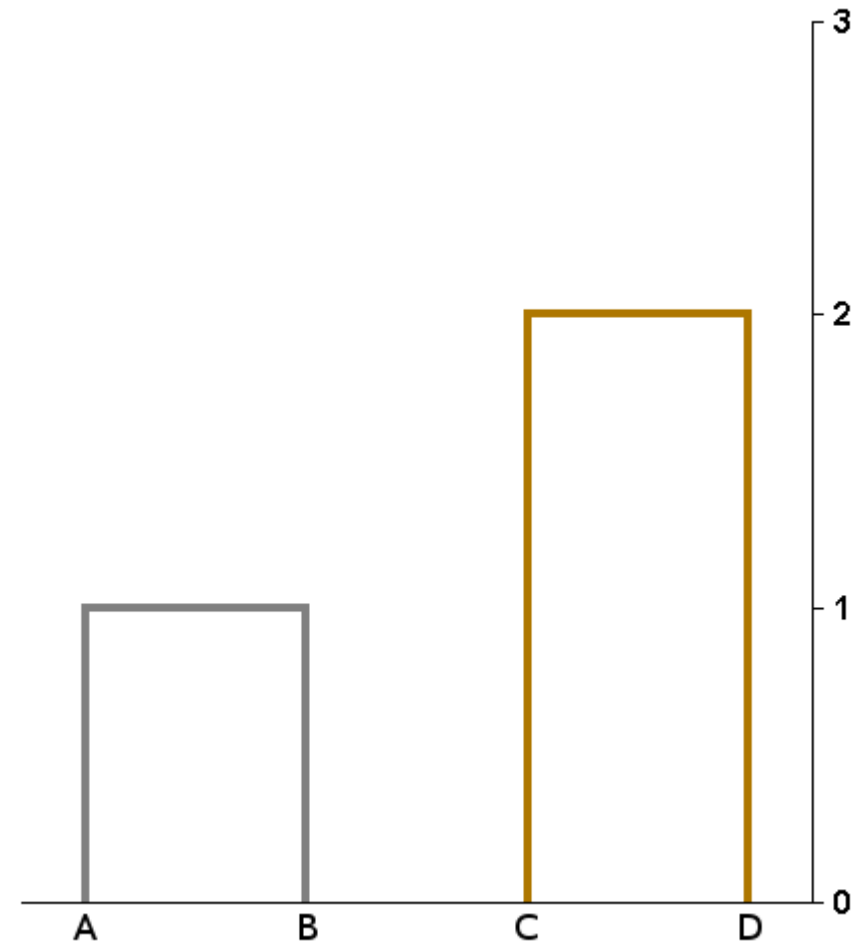


# Drzewa filogenetyczne: tworzenie drzewa algorytmem UPGMA

## ITERACJA 2

	AB	C	D
AB	0		
C	6	0	
D	6	4	0

	AB	CD
AB	0	
CD		0

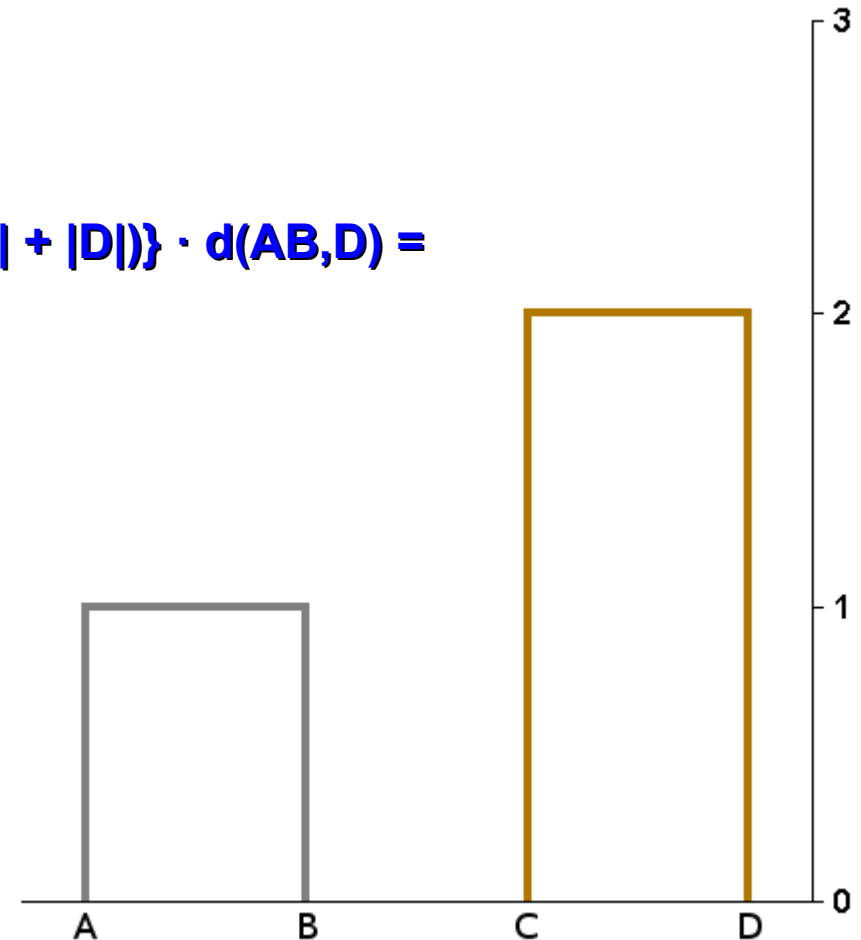


## ITERACJA 2

	AB	C	D
AB	0		
C	6	0	
D	6	4	0

	AB	CD
AB	0	
CD	6	0

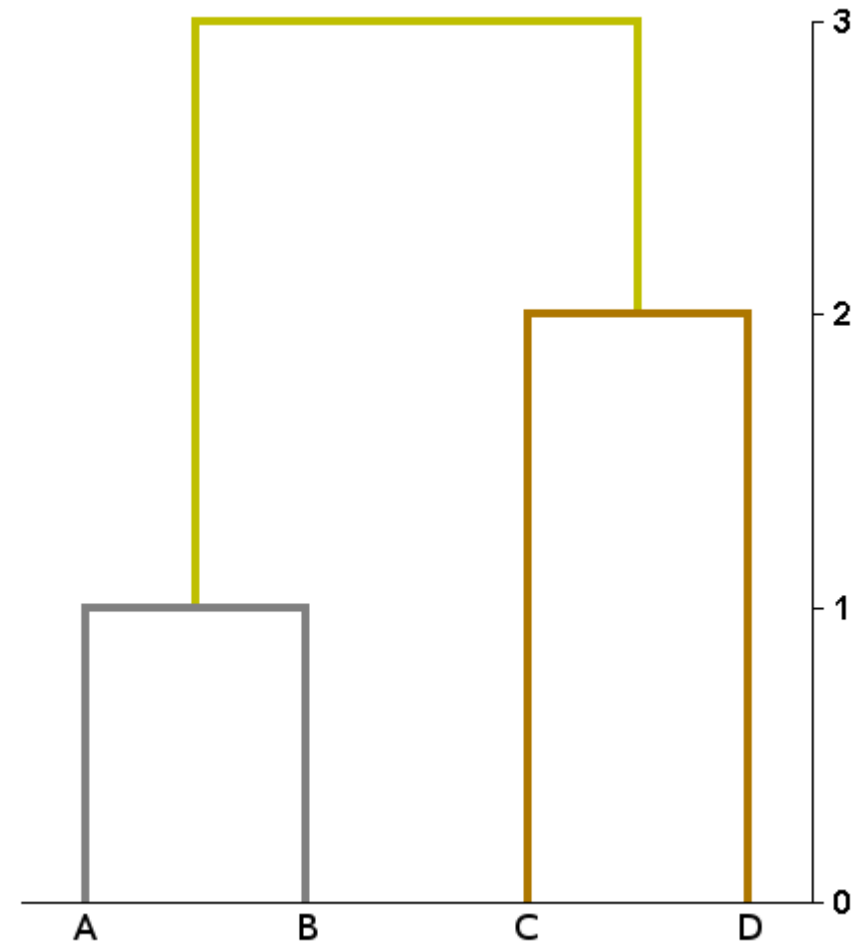
$$\begin{aligned}d(AB, C \cup D) &= \{|C| / (|C| + |D|)\} \cdot d(AB, C) + \{|D| / (|C| + |D|)\} \cdot d(AB, D) = \\ &= \{1 / (1+1)\} \cdot 6 + \{1 / (1+1)\} \cdot 6 = 6\end{aligned}$$



# Drzewa filogenetyczne: tworzenie drzewa algorytmem UPGMA

## ITERACJA 3

	AB	CD
AB	0	
CD	6	0



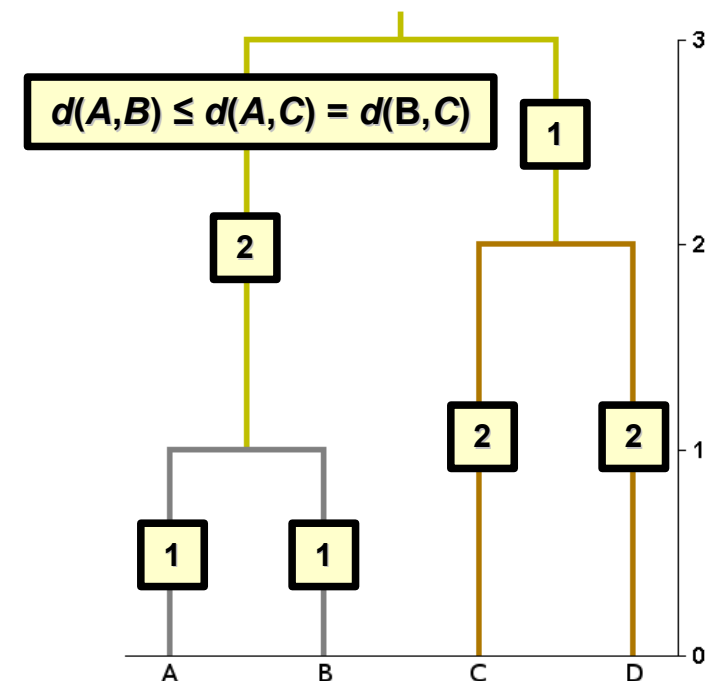
# Porównywanie sekwencji: tworzenie drzewa algorytmem UPGMA

## Cechy drzewa wyznaczonego algorytmem UPGMA:

- drzewo jest zawsze ukorzenione;
- całkowita długość gałęzi prowadząca od korzenia do dowolnego liścia jest zawsze taka sama;
- całkowita długość gałęzi prowadząca od dowolnego węzła do znajdujących się liści jest zawsze taka sama;
- drzewo jest **ultrametryczne**, czyli dla dowolnych trzech gatunków dwie najdłuższe z trzech dzielących je odległości są sobie równe (dla dowolnych trzech gatunków dwa zawsze są bliżej spokrewnione ze sobą niż z trzecim).

Właściwości wynikają z leżącego u podstaw metody UPGMA założenia, mówiącego że zegar molekularny „tyka” w jednakowym tempie dla wszystkich gatunków.

Jeżeli rzeczywiste tempo mutacji różni się pomiędzy gatunkami, wówczas **topologia otrzymanego drzewa może być nieprawidłowa**.



# Drzewa filogenetyczne: tworzenie drzewa algorytmem UPGMA

