

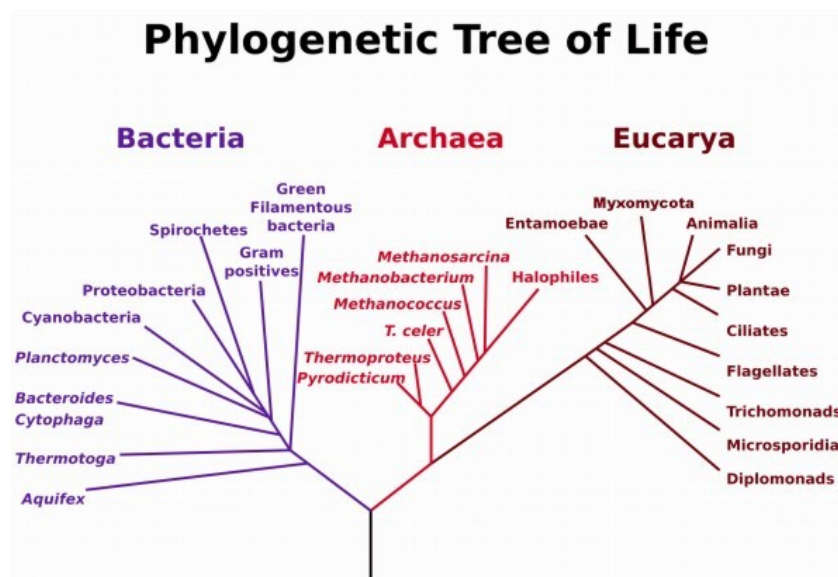
*Instytut Informatyki i Matematyki Komputerowej UJ,  
opracowanie: mgr Ewa Matczyńska, dr Jacek Śmietański*

## Analizy filogenetyczne

### 1. Wprowadzenie

**filogeneza** – droga rozwoju rodowego, pochodzenie i zmiany ewolucyjne grupy organizmów, zwykle gatunków

**filogenetyka** – dział biologii zajmujący się badaniem drogi rozwojowej (filogenezy) organizmów

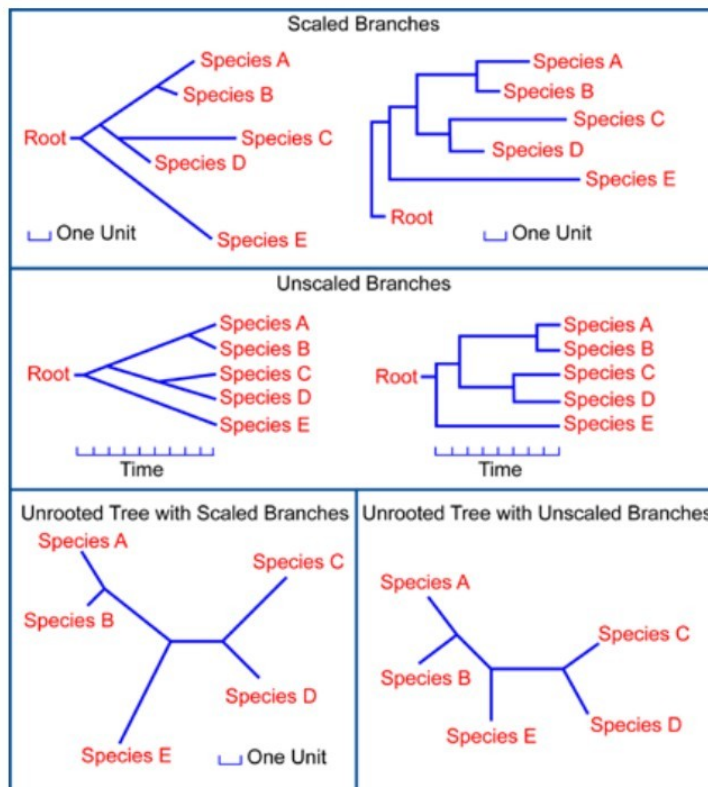


Rysunek 1: Filogenetyczne drzewo życia.

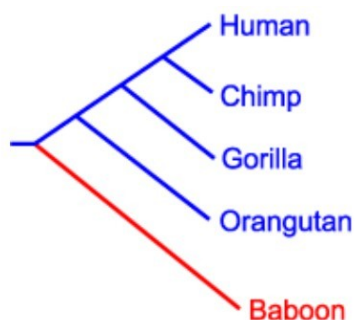
Zadania filogenetyki :

- zrekonstruowanie ewolucyjnej historii wszystkich organizmów
- odkrycie przodka wszystkich organizmów żyjących na Ziemi
- segregacja i klasyfikacja organizmów
- poznanie mechanizmów ewolucji

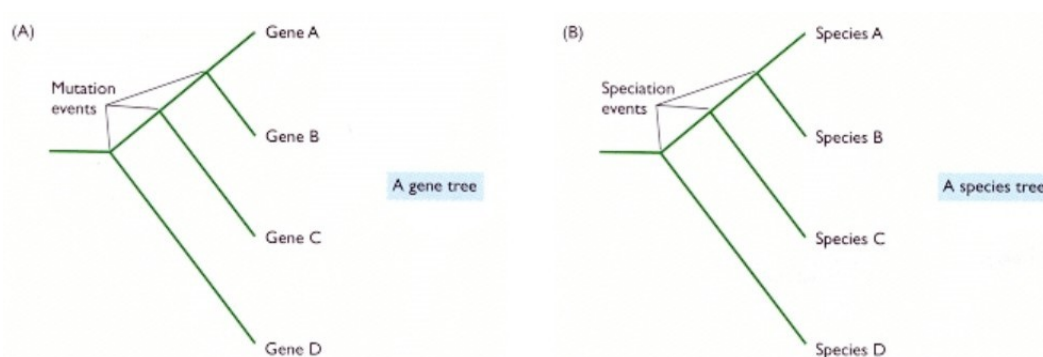
Od lat posługiwano się filogenetyką bazując na cechach morfologicznych organizmów, w tej chwili możemy mówić o filogenetyce molekularnej, gdyż możemy porównywać ze sobą informacje zawarte w DNA organizmów. Zakładamy, że sekwencje te są homologiczne i zmieniały się podlegając tym samym prawom ewolucji.



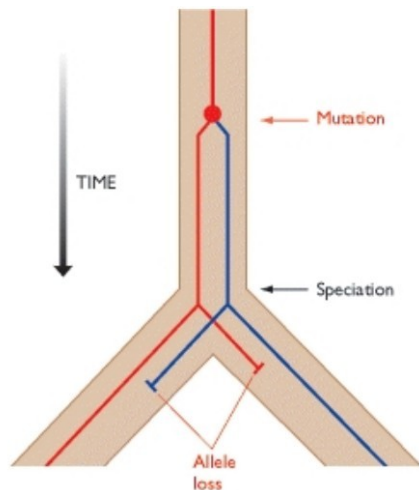
Rysunek 2: Różne rodzaje drzew filogenetycznych.



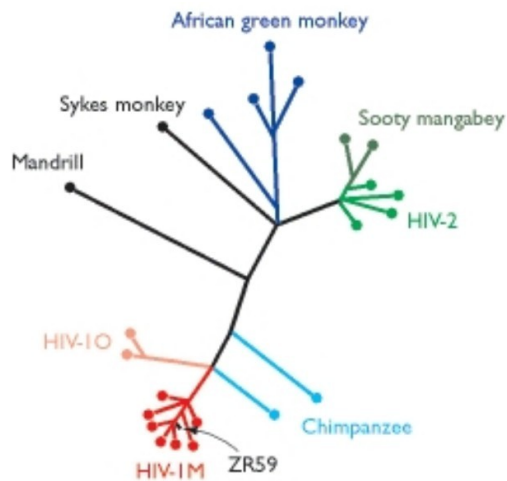
Rysunek 3: Wprowadzenie grupy zewnętrznej umożliwia ulokowanie korzenia drzewa.



Rysunek 4: Drzewo filogenetyczne dla genów nie zawsze jest zgodne z drzewem filogenetycznym gatunków. Rozgałęzienie drzewa dla genów oznacza mutację, ale dwa rodzaje genów mogą nadal występować w ramach jednego gatunku. Rozgałęzienie w drzewie gatunków oznacza specjację.



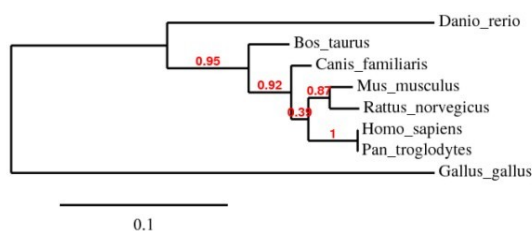
Rysunek 5: Mutacja, specjacja i utrata alleli danych genów w gatunkach potomnych.



Rysunek 6: Drzewo filogenetyczne dla wirusa HIV. Odmiana ZR59 uzyskana z krwi Arfykańczyka w 1959 jest jedną z najwcześniejszych wersji wirusa HIV-1.

## 2. Ocena wiarygodności drzewa filogenetycznego

Aby sprawdzić czy drzewo, które obliczyliśmy jest sensowne często stosuje się podejście bootstrap, które pozwala na ocenienie istotności informacji biologicznej zawartej w drzewie. Podejście to polega na tworzeniu nowych dopasowań (tej samej długości) na podstawie tego, które badamy poprzez wielokrotne losowanie ze zwracaniem kolumn dopasowania. Dlatego pewne kolumny mogą pojawić się więcej niż raz, inne wcale.



Rysunek 7: Bootstrap dla rhodopsyny uzyskany na serwerze - [www.phylogeny.fr](http://www.phylogeny.fr)

Jeśli pomimo tak zmienionych dopasowań, drzewo, które uzyskujemy jest podobne wtedy możemy mówić, że wynik jest biologicznie wiarygodny. Dla każdego poddrzewa możemy wyznaczyć współczynniki – odsetek drzew wyprodukowanych metodą bootstrap w którym się pojawiły. W ten sposób możemy się dowiedzieć, które fragmenty drzewa są istotne biologicznie.

### 3. Bio.Align i Bio.Phylo

W Biopythonie mamy moduł Bio.Align, dzięki któremu możemy analizować dopasowania wielosekwencyjne, oraz Bio.Phylo, który pozwala na wizualizację i analizę drzew filogenetycznych. Na wcześniejszych zajęciach korzystaliśmy z programu Clustal aby wyprodukować dopasowanie wielosekwencyjne dla białka rodopsyny - światłoczułego barwnika siatkówki oka. Plik rho.clustalw zawiera output programu z dopasowaniem.

#### Zadanie 1.

Wejdź na stronę [http://www.ebi.ac.uk/Tools/phylogeny/simple\\_phylogeny](http://www.ebi.ac.uk/Tools/phylogeny/simple_phylogeny). Wklej output z dopasowaniem wielosekwencyjnym oraz zaznacz w ustawieniach *Clustering Method* UMPGA i uruchom Clustal. Zapisz plik ze strukturą drzewa, który jest dostępny w *Result Summary* jako **rhoTree.ph**.

#### Zadanie 2.

Zadanie w Biopythonie – wykonaj wszystkie poniższe polecenia:

##### 1. Przeczytaj i wypisz dopasowanie wykonane przez Clustal

```
from Bio import AlignIO
alignment = AlignIO.read("rho.clustalw", "clustal")
print(alignment)
```

Możemy wypisać dopasowanie również w formacie fasta:

```
print(alignment.format("fasta"))
```

jak i dostać się do wszystkich sekwencji dopasowania :

```
for record in alignment:
    print(record.seq, record.id)
```

Sprawdź, co można otrzymać wywołując następujące komendy:

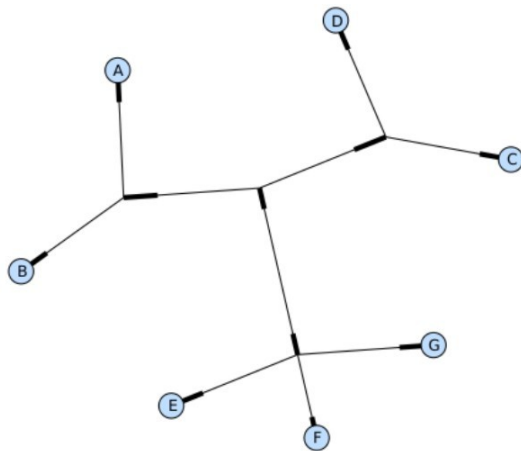
```
print(alignment[4:7])
print(alignment[3].seq[6:20])
print(alignment[:,1:6])
```

##### 2. Drzewo filogenetyczne możemy wczytać używając *Bio.Phylo* z pliku, w którym mamy zapisane drzewo filogenetyczne w formacie newick (zob. na wiki format newick):

```
from Bio import Phylo
tree = Phylo.read("rhoTree.ph", "newick")
print(tree)
Phylo.draw_ascii(tree)
```

Jeżeli posiadasz zainstalowane biblioteki matplotlib, NetworkX, PyGraphviz i Graphviz możesz wyrysować drzewo w ładniejszej graficznej postaci (ten punkt nie jest obowiązkowy):

```
Phylo.draw_graphviz(tree)
```



3. Na obiekcie drzewa można wykonywać różne operacje np.:

```
terminals = tree.get_terminals()
for terminal in terminals :
    print(terminal)

path = tree.get_path("Homo")
print(path)

common_ancestor = tree.common_ancestor("Homo", "Danio")
print common_ancestor

depths = tree.depths()
for d in depths.keys():
    print(d ,":", depths.get(d))
```

Zajrzyj też do dokumentacji:

<http://biopython.org/DIST/docs/api/Bio.Phylo.BaseTree.TreeMixin-class.html>

gdzie znajdziesz więcej funkcji dla drzew.

### Zadanie 3. (4pkt)

Rozwiązanie zadania prześlij mailem do niedzieli, **27.11.2018** włącznie, na adres:

**jacek.smietanski@ii.uj.edu.pl**

Temat wiadomości proszę opatrzyć przedrostkiem **[Bio] Lab 08**. Rozwiązaniem ma być **tylko jeden plik** – dokument PDF. Proszę o nazwanie pliku wg schematu: **Imie.Nazwisko.08.pdf**.

1. Wybierz sobie jakiś dowolny gen i wyszukaj czy jest on w bazie **HomoloGene** na NCBI.
2. Za pomocą *download* po prawej na stronie rekordu bazy zapisz sekwencje homologiczne tego genu u różnych organizmów (minimum 4-5, może być więcej – wtedy będziemy mogli lepiej ocenić drzewo). Wykonaj dopasowanie wielosekwencyjne za pomocą Clustala a następnie wygeneruj drzewo filogenetyczne.
3. Wykorzystując funkcję *distance* oblicz i wypisz dystans pomiędzy człowiekiem a innymi organizmami w twoim drzewie.