

## Bioinformatyka – Laboratorium nr 4.

### Biologiczne bazy danych

- EMBL – European Molecular Biology Laboratory - <https://www.embl.org/>
- NCBI – National Center for Biotechnology Information - <https://www.ncbi.nlm.nih.gov/>
- DDBJ – DNA Data Bank of Japan - <https://www.ddbj.nig.ac.jp/index-e.html>

### Genomy eukariotyczne

Genom jądrowy jest podzielony na dwie lub więcej liniowych cząsteczek DNA – chromosomów ( u człowieka 22 autosomy i dwa chromosomy płci X i Y)

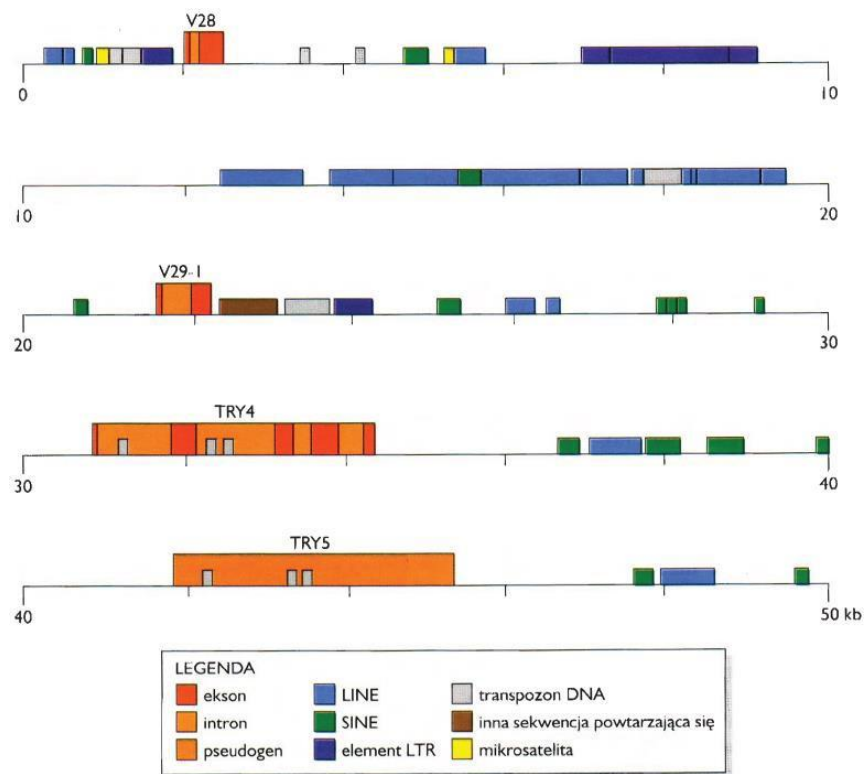
Genom mitochondrialny

Genom chloroplastowy tylko w organizmach fotosyntezujących

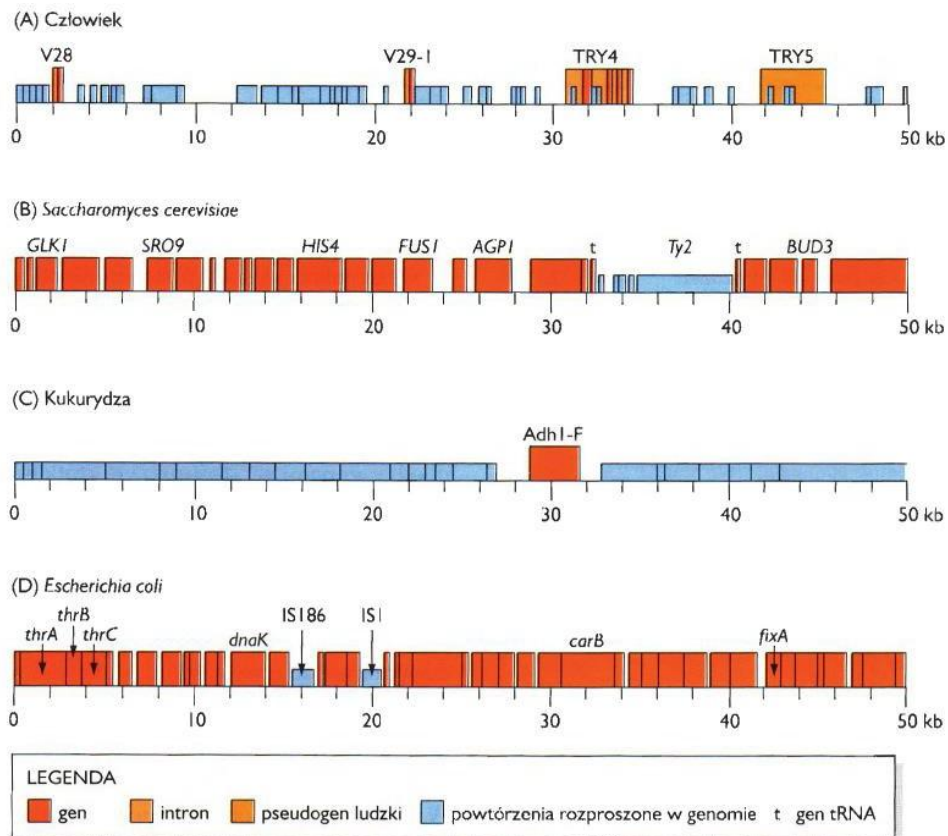
Organizm	Wielkość genomu (Mb)
<b>Prokariota</b>	
<i>Mycoplasma genitalium</i>	0,58
<i>Escherichia coli</i>	4,64
<i>Bacillus megaterium</i>	30
<b>Eukariota</b>	
Grzyby	
<i>Saccharomyces cerevisiae</i> (drożdże)	12,1
<i>Aspergillus nidulans</i>	25,4
Pierwotniaki	
<i>Tetrahymena pyriformis</i>	190
Bezkręgowce	
<i>Caenorhabditis elegans</i> (nicień)	100
<i>Drosophila melanogaster</i> (muszka owocowa)	140
<i>Bombyx mori</i> (jedwabnik)	490
<i>Strongylocentrus purpuratus</i> (jeżowiec)	845
<i>Locusta migratoria</i> (szarańcza)	5000
Kręgowce	
<i>Fugu rubripes</i> (ryba najeżka)	400
<i>Homo sapiens</i> (człowiek)	3000
<i>Mus musculus</i> (mysz)	3300
Rośliny	
<i>Arabidopsis thaliana</i> (rzodkiewnik)	100
<i>Oryza sativa</i> (ryż)	565
<i>Pisum sativum</i> (groch)	4800
<i>Zea mays</i> (kukurydza)	5000
<i>Triticum aestivum</i> (pszenica)	17000
<i>Fritillaria assyriaca</i> (szachownica)	120 000

Dane za: Brown (1998).

## Zawartość genetyczna genomu człowieka



## Porównanie genomów



## Ćwiczenia

1. Zapisać na dysku referencyjną sekwencję chromosomu 1, Kropidlaka popielatego (*Aspergillus fumigatus*)
2. Odszukać w bazie danych sekwencję genu FMR1 człowieka i zapisać ją w formacie FASTA.
3. Zapisać sekwencje aminokwasów kodowanych przez drugi ekson ludzkiego genu CREB5

**BLAST** (ang. *Basic Local Alignment Search Tool*) – narzędzie bioinformatyczne (algorytm) służące do lokalnego przyrównywania sekwencji aminokwasów (białek) lub nukleotydów DNA. BLAST umożliwia użytkownikom porównywanie zadanej sekwencji z sekwencjami zawartymi w biologicznych bazach danych i statystyczną ocenę ich podobieństwa.

Różne typy BLAST-a służą do porównywania różnych rodzajów sekwencji. Dla przykładu: po odkryciu nieznanego genu u myszy, przeszukują bazę z ludzkim genomem pod kątem obecności podobnych genów. BLAST znajdzie sekwencje podobne w bazie o ustalonych z góry parametrach (takich jak stopień podobieństwa). Dzięki prostocie obsługi jest bardzo użytecznym narzędziem.

Projektantami algorytmu byli: Stephen Altschul, Warren Gish, Webb Miller, Eugene Myers, i David J. Lipman z National Institutes of Health.

## Ćwiczenia

4. Dopasować sekwencję:  
AAAAAACA AAAAACAAGATTGATGTTCCGGATGATGGCTGCTAGCTATCCTTTTATGAAGG  
TGGGATATGTGGTCTCACTGCAGGGTGGCCTAGGAAGGGGAATCTGCTCATAGACATTAC  
CCAGAGATCTGTTCTCCTGCATTCTGACTGGAGGGGAGCTAAGACCCTAGAGGAGTCGGC  
CTGTCACCATAGACCAATCAATCACCTGTCATCCATAGACCAGTCACCTGTCACCTAAA  
CCCATGGCTGCTCTTTTTTCATCTTCATCTTCCAAGACCTAGTGACCCAGGTTTCATGC
5. Dopasować sekwencję:  
CCTCCTCTCTTTTTCTTCTTTCTTCTTTTTCTTTCTCTTCTTCTTCTCCTTCTTTCTCTTT  
CACTCCTCTTTCTTTCTTTTTCTTCTTTTTTTTTCTTCCCTCCTTCTTTCTCCCCCTTTTC  
TATTTTCCCTCCCTCCTCTTTCCCTCTTCTCCCCCCCCCTTCCCCCCCCCTCCTCTCTCCC  
TTTTCTCCTTCTCCTCCCCCTCCCCCTTTCTCCCCCTCTCCCTTCTCCTCTCCCCCCCCCCCC  
CCCCCTCCCCCCCCCTTCCCTCTCCCTCT
6. Dopasować sekwencję:  
caggaggctt ataagttttg ccaagagtat aggtatctga attaactgta atcgacttaa  
tggttttcac taaatcctcc cgtaataacc attttaccat aataccatta ccatcattaa  
aaaaaggtaa caaacagacc taacataaag aaaatacctc cgcaccacaa aattagatat  
atgaatccaa tcaagactct aacacaaaa tatttactgc tcatgacaat aaaaa

7. Dopasować sekwencję:

```
MVGGGTKRRRRRLQLSKLYTLTCAQACFKQDHSQIGGGPFSRVVYCNEPDSPEADSRNYSNDY  
VRTTKYTLATFLPKSLFEQFRRVANFYFLVTGVLAF
```

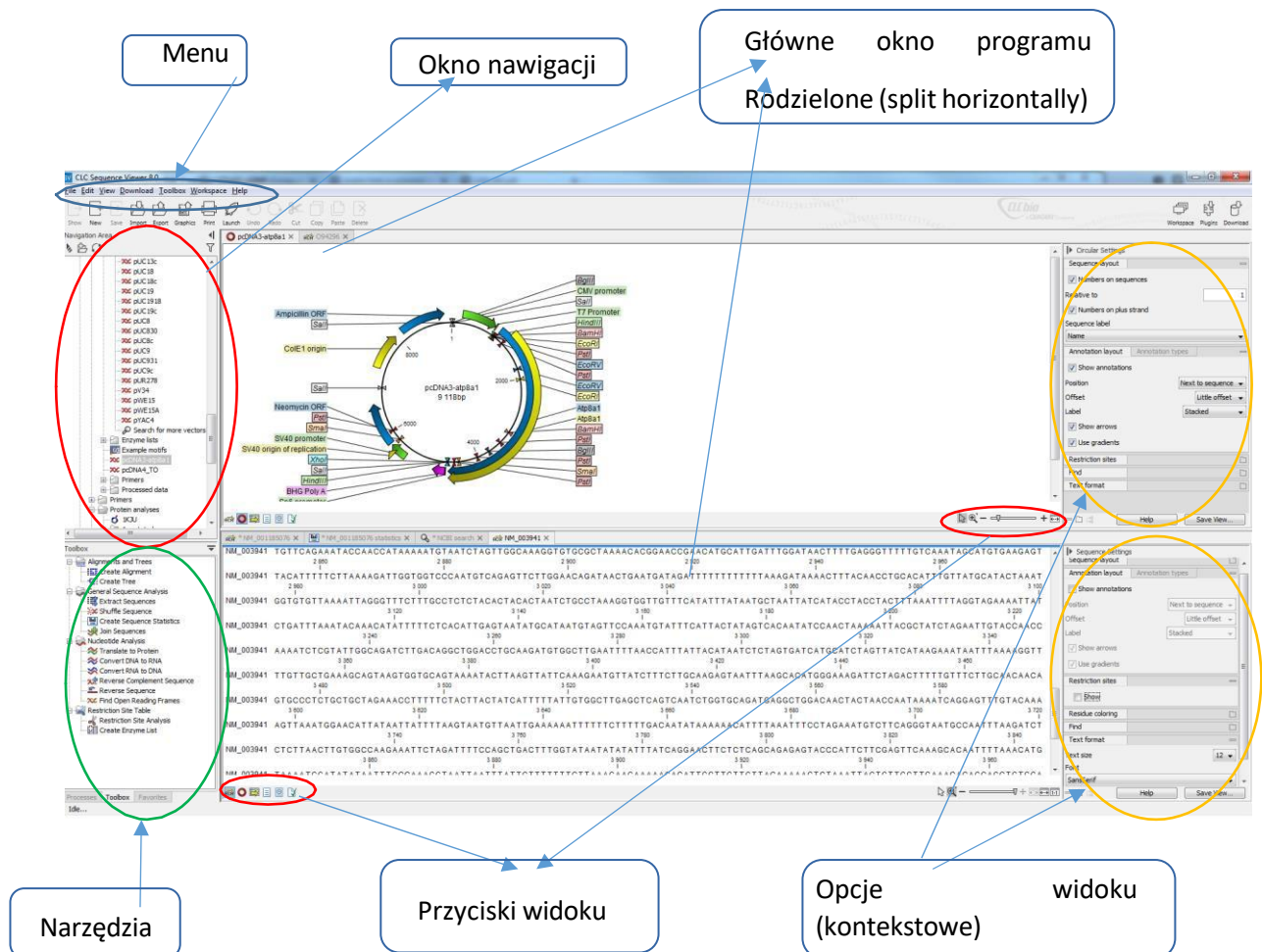
8. Dopasować sekwencję:

```
CTGCATGGCCGTCCTTTGAAGGTTCAAACGAACCTTTTTACAATTTGCACTCACTGTTG
```

9. Wykonać BLAST dwóch sekwencji z pliku sample.fa.

## Programy do pracy z sekwencjami – CLC Sequence Viewer


Okno programu



## Menu Narzędzia (Toolbox)

- Alignments and Trees - Dopasowania i drzewa (filogenetyczne);
- General sequence analysis - Ogólna analiza sekwencji;
- Extract – wyodrębnienie sekwencji;
- Shuffle – wymieszanie;
- Statistics – statystyka ;
- Join – łączenie;
- Nucleotide analyses – analiza nukleotydów;
- Convert – przekształcenie;
- Reverse complements of sequence – tworzenie odwrotnej sekwencji komplementarnej;
- Reverse sequence – odwrócenie sekwencji;
- Find Open Reading Frames – znajdowanie otwartych ramek odczytu;
- Translation of DNA or RNA to protein – przetłumaczenia DNA lub RNA na białko (aminokwasy);
- Restriction site analyses – analiza miejsc restrykcyjnych.

W programie CLC Sequence Viewer (wersja do instalacji <https://clc-sequence-viewer.software.informer.com>) wykonać:

10. Pobieranie danych z zewnętrznej bazy danych. Pobrać sekwencję ludzkiego genu WASL w formacie Genbank i wykonać Import do programu CLC.
  - Polecenie Download
  - Wybrać bazę danych – nukleotydy (ang. Nucleotide)
  - Organism – human
  - Gene Name – WASL
  - Odpowiedni wiersz z wynikiem pobieramy i otwieramy (prawy przycisk myszy Download and Open)
11. Prezentacja sekwencji (gen WASL)
12. Sprawdzić działanie przycisków widoku 
13. Eksport grafiki oraz eksport do wybranych formatów plików.
14. Dopasowanie sekwencji białkowych (pliki O94296, P39524, P57792, Q9NTI2, Q9SX33, Q29449)
15. Dopasowywanie sekwencji DNA (pliki Fwd1, Fwd2, Fwd3, Fwd4, Fwd5)

