

1. Podstawowe pojęcia dotyczące sekwencji biologicznych:

- RNA i DNA
- transkrypcja, odwrotna transkrypcja i translacja
- budowa genu

2. Biologiczne bazy danych:

- I- i II-rzędowe bazy danych
- Rodzaje baz danych
- Przykłady baz danych i narzędzia do ich przeszukiwania:
 - a) GenBank i ENA
 - b) UniProt
 - c) ENSEMBL (BioMart)
- formaty rekordów danych:
 - a) format zapisu sekwencji FASTA
 - b) format zapisu sekwencji FASTQ i jakość sekwencji
 - b) formaty zapisu GenBank i EMBL-ENA/UniProt i identyfikatory sekwencji;
 - c) formaty uniwersalne (XML);

3. Porównywanie sekwencji biologicznych:

- relacje ewolucyjne między sekwencjami
- porównanie dwóch sekwencji ('pairwise'):
 - a) metoda graficzna – 'dot plot' / 'dot matrix'
 - b) punktowanie dopasowania: matryce podstawień i kary za przerwy
- wielu sekwencji (msa):
 - a) dopasowanie globalne i lokalne
 - b) metoda progresywna
 - c) analiza profili i bloków
 - d) PSI-BLAST
- statystyka dopasowania sekwencji: e-value
- przeszukiwanie zasobów przy pomocy sekwencji: BLAST;

4. Analiza sekwencji biologicznych:

- analiza i adnotacja sekwencji genomowych:
 - a) metody adnotacji sekwencji genomowych:
 - Maskowanie powtórzeń
 - Programy do przewidywania genów *ab initio*
 - Analiza sekwencji EST
 - Adnotacja funkcjonalna poprzez homologię (BLASTp)
 - b) narzędzia służące do wizualizacji genomów i ich adnotacji (Genome Browsers)
 - c) narzędzia do analizy porównawczej adnotacji genomów (Galaxy)
- elementy analizy filogenetycznej:
 - a) relacje ewolucyjne między sekwencjami (podobieństwo i homologia)
 - b) drzewa filogenetyczne i formaty zapisu
 - c) etapy i metody rekonstrukcji filogenetycznej
- struktura RNA i białek:
 - a) poziomy analizy strukturalnej
 - b) motywy funkcjonalne w strukturze białek
 - c) sekwencje konsensusowe, wzorce, motywy, profile
 - d) struktura drugorzędowa RNA