

1. Biologiczne bazy danych:

- rodzaje i przykłady:
 - a) I i II rządowe bazy danych;
 - b) specjalistyczne i uniwersalne bazy danych;
 - c) zasoby NCBI
- formaty rekordów danych:
 - a) format zapisu sekwencji FASTA;
 - b) formaty zapisu GenBank i EMBL-ENA/UniProt;
 - c) formaty specjalne (TPA) i uniwersalne (ASN.1, XML);
- strategie i narzędzia przeszukiwania bdd:
 - a) przeszukiwanie przy pomocy słów kluczowych;
 - b) NCBI Entrez;
 - c) PubMed i Automatyczne Mapowanie Haseł.

2. Porównywanie i dopasowanie sekwencji:

- dwóch sekwencji ('pairwise'):
 - a) metoda graficzna - dotPlot
 - b) metoda optymalna - programowanie dynamiczne
 - c) punktowanie dopasowania: matryce podstawień i kary za przerwy
- wielu sekwencji (msa):
 - a) dopasowanie globalne i lokalne;
 - b) metoda progresywna i metody statystyczne;
- statystyka dopasowania sekwencji: p-value, e-value i z-score;
- przeszukiwanie zasobów przy pomocy sekwencji: BLAST i FASTA;
- analiza danych NGS:
 - a) format FASTQ;
 - b) metody dopasowania: drzewa i tablice sufiksowe, algorytm haszujący i metoda Burrows'a- Wheeler'a;
 - c) zastosowania technologii NGS.

3. Analiza sekwencji biologicznych:

- analiza i adnotacja sekwencji genomowych:
 - a) dekodowanie informacji zawartych w genomach;
 - b) metody *ab initio* i oparte o homologie;
 - c) adnotacja funkcjonalna w oparciu o pokrewieństwo lub identyfikację funkcjonalnych domen;
- elementy analizy filogenetycznej:
 - a) drzewa filogenetyczne
 - b) modele zmian ewolucyjnych
 - c) metody rekonstrukcji filogenetycznej (minimalna ewolucja, metody oparte o dystans, metoda maksymalnej wiarygodności)
- bioinformatyka strukturalna: modelowanie RNA i białek:
 - a) poziomy analizy strukturalnej;
 - b) elementy strukturalne białek i RNA;
 - c) metody i strategie przewidywania struktury białeczek.