



UNIVERSITA' DEGLI STUDI DI MESSINA
*Dipartimento di Scienze Biomediche, Odontoiatriche e delle Immagini
Morfologiche e Funzionali*

Corso di Laurea in Biotecnologie Mediche (Classe LM-9)
BIOINFORMATICS AND MOLECULAR NETWORKS
Docente: Prof. Orazio Romeo

COURSE PROGRAM

PART I:

Introduction to bioinformatics and biological databases

The objectives of bioinformatics; Contribution of bioinformatics toward modern genetics and genomics; Introduction to biological databases: types of databases, classification and data retrieval; Nucleotide sequence databases (EMBL/DDBJ/GenBank; RefSeq); The Ensembl database; Protein sequence databases (TrEMBL; Uniprot; Entrez Protein).

Introduction to nucleic acid sequencing

Overview of first-, second-, and third-generation sequencing; Whole genome/transcriptome sequencing using Next Generation Sequencing (NGS) and Single-Molecule Real-Time (SMRT) Sequencing; Ins and outs of different NGS platforms; Metagenomics; Targeted and *De-novo* sequencing; Raw data generated by NGS and SMRT sequencing technologies.

Comparison of DNA and amino acid sequences and molecular phylogeny

Biological sequences and their analysis; Introduction to pairwise alignment; Global and local alignment; Sequence similarity and scoring methods; Dot-Matrix method; Scoring matrices (PAM and BLOSUM matrices); The BLAST algorithm and database search; Multiple sequence alignment and phylogenetic analysis; Distance-based methods (Neighbor-Joining and UPGMA Method); Statistical validation methods: Bootstrap analysis; Tools and software used for tree construction.

PART II:

Bioinformatics tools and methods in NGS data analysis

FASTQ file format, and base quality score; NGS data quality control and preprocessing; Read mapping; SAM/BAM as the standard mapping file format; *De novo* genome assembly; Contig

assembly algorithms; Scaffolding; Assembly quality evaluation; Gap closure; RNA-Seq data analysis; RNA-Seq data normalization; Identification of differentially expressed genes; Visualization of RNA-Seq data; Functional analysis of identified genes; Identification of long non-coding RNAs; Metagenome Analysis by NGS data; Taxonomic characterization of a microbial community; Identification of differentially abundant species or operational taxonomic units (OTUs).

Biological Networks: Tools, Methods, and Analysis

Types of biological networks: protein-protein interaction networks, disease-gene interaction networks; metabolic networks, and gene regulatory networks; Gene Ontology (GO); GO annotation; The database for annotation, visualization, and integrated discovery (DAVID); Bioinformatics utilities for GO annotation/enrichment analysis.

Textbook and course materials

- **Wang Xinkun**, Next-Generation Sequencing Data Analysis, CRC Press 2016.
- **Noor Ahmad Shaik** et al. Essentials of Bioinformatics, Vol. I. Springer, Cham 2019.
- Material (links, slides, files) and papers provided by the professor

Contacts

e-mail: oromeo@unime.it