BioByte

-IEEE EMBS-

Online

Short Course on BIOINFORMATICS

Course Outline

IEEE Engineering in Medicine and Biology Society

Student Branch Chapter

University of Sri Jayewardenepura

IEEE EMBS

University of Sri Jayewardenepura

IEEE Student Branch

Day 01: Introduction to Bioinformatics and In Silico Research

Objectives:

- Understand what bioinformatics is and its significance in modern biology and medicine.
- Omics-driven approach in bioinformatics.

Detailed Content:

- Introduction to bioinformatics: definitions, history, applications, and current status in bioinformatics.
- Genomics, transcriptomics, proteomics, metabolomics, and evolutionary studies.

Day 02: Bioinformatic Databases

Objectives:

- Become familiar with major international DNA sequence databases: GenBank (NCBI), EMBL (EBI), and DDBJ (Japan), Protein Data Bank.
- Learn how to navigate these databases to retrieve sequence data.
- Understand the purpose and functions of BLAST (Basic Local Alignment Search Tool) and perform basic sequence similarity searches.
- Create and manage a personal NCBI account for sequence data access and management.
- Retrieve nucleotide sequences from GenBank using accession numbers and keyword searches.
- Learn to download and save sequences in FASTA format for further analysis.
- Understand GenBank record entries: organism, gene information, molecule type, lineage, annotations, and metadata.

Detailed Content:

- Overview of the International Nucleotide Sequence Database Collaboration (INSDC) and roles of GenBank, EMBL, and DDBJ.
- Step-by-step guide on using NCBI's GenBank interface and accessing related resources.
- Introduction to BLAST: types of BLAST programs (blastn, blastp, blastx, tblastn, tblastx), their purposes, and practical exercises performing BLAST searches to identify unknown sequences or homologs.
- Procedures for searching and downloading DNA sequence files using accession numbers.

- Explanation of GenBank record structure: header, features (CDS, genes), source organism, sequence length, and annotations.
- Practical exercise to locate information such as host, strain, country of origin, and molecular function.
- Hands-on with the extraction of sequence data in various formats and preparing FASTA files.

Day 03: Next Generation Sequencing: DNA Sequence Read Quality Analysis

Objectives:

- Learn about sequencing and sequencing technologies.
- Learn the typical workflow behind sequencing DNA using a popular Next Generation Sequencing (NGS) platform like Illumina and raw data generated from a sequencing run.
- Conduct a simple quality analysis on DNA sequencing reads generated from a sequencing run.
- Learn how to interpret results generated from a sequence read quality analysis

Detailed Content:

- Sequencing and its importance. The 3 generations of sequencing technologies.
- Library preparation using specific adapters and sequencing of DNA in Illumina sequencers.
- The general format and data in FASTQ files including Phred quality scores generated at the end of a sequencing run.
- Linux based NGS read quality control on a sample DNA read file using tools like FastQC, MultiQC and Trimmomatic in a Conda environment.
- Interpreting different plots like per base sequence content, per sequence GC content, adapter content, sequence duplication levels etc. and associated errors reflected in these output results.

Day 04: Sequence Alignment Techniques

Objectives:

- Understand the biological importance of sequence alignment for homology and evolutionary studies.
- Perform pairwise and multiple sequence alignments.
- Use alignment software tools such as ClustalW, Clustal Omega, MEGA, and BioEdit.
- Learn how to manually inspect and refine alignments.

Detailed Content:

- Concepts of gap penalties, scoring matrices, and algorithms behind alignment tools.
- Progressive alignment methods and guide trees in multiple sequence alignment (MSA).
- Step-by-step creation of MSA for gene families or homologous sequences.
- Interpretation of alignment results in terms of conserved regions and variable sites.
- Exporting alignments in formats compatible with phylogenetic analysis.

Day 05: Construction of Phylogenetic Trees from Molecular Data and Tree Interpretation

Objectives:

- Comprehend the concept of phylogenetics and its importance in understanding evolutionary relationships.
- Learn different phylogenetic tree-building methods: distance-based (UPGMA, Neighbor-Joining) and character-based methods (Maximum Parsimony, Maximum Likelihood, Bayesian Inference).
- Build and visualize single-gene phylogenetic trees using software and online resources such as MEGA and CIPRES Science Gateway.
- Perform multi-locus Maximum Likelihood phylogenetic analyses.
- Interpret the resulting trees in terms of evolutionary patterns, support values, and biological implications.

Detailed Content:

- The theory behind molecular phylogenetics: homology vs analogy, rooting trees, and evolutionary models.
- Detailed explanation of each tree-building algorithm with pros and cons.
- Use of CIPRES Gateway for computationally intensive analyses, uploading alignments and running RAxML or MrBayes.
- Bootstrapping and statistical support for tree nodes.
- Visualizing and interpreting trees including clades, sister groups, and outgroups.
- Comparison of tree topologies from single-locus vs multi-locus analyses.
- Definitions and distinctions among monophyletic, paraphyletic, and polyphyletic groups.
- Understanding polytomies, resolving ambiguous relationships, and the biological meaning of tree topology.

Day 06: Proteomics and Metabolomic Analysis.

Objectives:

- Understand the principles, experimental workflows, and biological questions addressed by proteomics and metabolomics.
- Learn common analytical platforms (LC–MS/MS, GC–MS, NMR) and how experimental choices affect downstream bioinformatic analysis.
- Learn to do a basic molecular docking analysis using the PyRx tool.
- Understand multi-locus phylogenetics and why combining multiple gene regions can improve tree resolution.

Detailed Content

- What proteomics and metabolomics measure (proteins as functional effectors; metabolites as final products of biochemical activity).
- Targeted vs untargeted approaches and when to use each (hypothesis-driven biomarker validation vs discovery).
- Typical study aims at biomarker discovery, pathway perturbation, mechanistic inference, and Enzyme inhibition visualization using molecular docking.
- Visualization and comparison of an enzyme inhibition activity of a readily available drug and a natural compound (Ex: Amylase and Acarbose, Amylase with Fucoxanthin)