

Schedule for Lectures and Practicals in Bioinformatics - 2019

1. What does bioinformatics mean?

9 Sep

Review of the schedule and the requirements of the lectures. Introduction to the course: When do we need computers for biological research? The history, meaning and definition of bioinformatics. Bioinformatics applications. Trends in biological research. Bioinformatics on the web. The most used bioinformatic methods and software. Introduction to Linux operating system.

Dávid Fazekas

Practical: Introduction to the usage of the computer room

Review of the schedule and the requirements of the practicals. Introduction of the computer room. Introduction of Linux operating system: graphical interface, text editor, how to start a software. Usage of simple Linux commands to solve bioinformatic problems.

Dávid Fazekas, Balázs Bohár

2. Databases in molecular biology

16 Sep

The structure and content of databases. Relational databases. How to apply molecular biology databases in biological research? Nucleotide (EMBL, GenBank) and protein databases (UniProt). Entry formats of sequence databases. ID mapping. Other biological databases (Gene Ontology, Tree of life).

Eszter Ari

Practical: Application of molecular biology databases

Introduction to the usage of NCBI databases and the Uniprot protein database. How to search in databases? Searching in Uniprot database: Retrieve, ID mapping. What does an NCBI or a Uniprot entry contain?

Eszter Ari

3. Sequence comparison and alignment

23 Sep

Clarification of conceptions: sequence similarity vs. homology. Pairwise sequence alignments, pairwise comparisons: dot-plots. Scoring system: substitution matrices: PAM, BLOSUM. Optimal global and local alignments. Gap-penalties. Multiple sequence alignment methods.

Eszter Ari

Practical: The basics of sequence analysis and sequence alignment

Manipulation of sequences, different sequence file formats. Introduction to the usage of the UGENE software. Using UGENE for database searching and downloading sequences. Analysis of sequences using UGENE: Dot-plots, pairwise and multiple alignments.

Eszter Ari, Dániel Gerber

4. Sequence similarity searching

30 Sep

Heuristic algorithm for similarity searching: BLAST software. Statistical bases: evaluate the significance of hits. Filtering false positive hits: low complexity and repetitive regions. Different BLAST versions and other similarity searching software.

Eszter Ari

Practical: Searching sequences in databases based on their similarity

Similarity searches in online sequence databases. Using the online version of BLAST software on NCBI webpage. Evaluating the results. Downloading, aligning and editing the hit sequences.

Balázs Egyed

5. Structural bioinformatics I

7 Oct

Introduction to protein structures: Structure determination, Protein Data Bank, basic principles of protein structure

Zsuzsanna Dosztányi

Practical: Visualization of protein structures

Zsuzsanna Dosztányi

6. Structural bioinformatics II

Introduction to predicting various properties of proteins.

14 Oct

Zsuzsanna Dosztányi

Practical: Bioinformatic analysis of protein structures II

Sequence analysis, secondary structure predictions, prediction of transmembrane regions and intrinsically disordered protein regions.

Zsuzsanna Dosztányi

7. Molecular phylogenetics

21 Oct

What is phylogenetics? How can we apply it? Phylogenetic signal, phylogenetic trees. Prediction of phylogenetic changes and distances, nucleotide substitution models. Methods for phylogenetic reconstruction. Distance based methods: neighbor-joining. Character based methods: maximum parsimony. Consensus trees. Statistical tests: bootstrap analysis.

Dániel Gerber

Practical: Molecular phylogenetics in practice I

Introduction to the usage of PHYLIP software package. Calculating maximum parsimony tree using PHYLIP. Reconstructing distance based phylogenetic tree with UGENE. Evaluate the reliability of phylogenetic trees by bootstrapping. Tree visualization using the Figtree software.

Balázs Egyed

fall break

28 - 31 Oct

8. Genomics and transcriptomics I + Mid term test I (Lecture: 1-6)

4 Nov

Genome sequencing methods. Analysis of high-throughput sequencing data with bioinformatic methods. Genome assembly, mapping reads to reference genome, annotation, SNP analysis.

Eszter Ari

Practical: Genome databases

Introduction to the Ensembl genome database. What is Biomart and how to use it?

Eszter Ari, Dániel Gerber

9. Genomics and transcriptomics II

11 Nov

Introduction to the microarray technique. Analysis and possible applications of microarray data. Introduction to the RNA-seq method and data analysis. *De novo* assembly, mapping reads to reference genome, differential expression analysis. Functional enrichment analysis.

Eszter Ari

Practical: Differential expression analysis of microarray and RNA-seq data

Introduction to the NCBI GEO database. Differential expression analysis. Functional enrichment analysis using the Gene Ontology database.

Eszter Ari

10. Network and systemsbiology I

18 Nov

Introduction to network and systemsbiology. What kind of data can be used? Bioinformatic approaches in network and systemsbiology.

Balázs Bohár

Practical: Cytoscape I

Introduction to the usage of Cytoscape network analyzer and visualization software.

Balázs Bohár

11. Network and systemsbiology II

25 Nov

Introduction to the “world of networks”: topology, hierarchy and dynamics. Different types and sources of molecular networks. Creation of protein-protein interaction (PPI) networks, high-throughput methods, their advantages and disadvantages.

Balázs Bohár

Practical: Cytoscape II

Creating a tissue specific network using the Cytoscape software.

Balázs Bohár

12. Genetic background of haplotype reconstruction

2 Dec

Introduction to the relationship between chromosome, haplotype, genotype and crossing over by parent-child trio. Genetic linkage, linkage analysis, the discovery of the genes being responsible for monogenic diseases. Extension to population, linkage disequilibrium, D , D' and r^2 . Molecular haplotyping. Bioinformatic haplotype reconstruction (phasing): introduction on algorithms and programs. Impact on human genetics.

Márton Doleschall

Practicals: A taste of haplotype reconstruction

Introduction on PHASE v2.1 program: genotypic data, input format, command line, parameters and output format. Generation of family (trio) input file, related command line and the evaluation of output file. Command line and output of population input file.

Márton Doleschall

13. Mid term test 2 (Lecture 7-12)

9 Dec

Practical: Consultation about the project work

Eszter Ari, Zsuzsa Dosztányi, Balázs Bohár

Evaluation

Evaluation of the theoretical part (lecture grade):

The average grade of the 2 mid term tests (written, 45 min.).

You have to take an oral exam during the examination period if one (or both) mid term tests resulted with mark 1 or you missed the mid term exam. Oral exam will be about that part in which you failed or missed. You also can improve your mark (what you got as an average) if you wish in an oral exam – but be aware that you can also decrease the mark.

x.5 averages will be rounded upwards.

Evaluation of the practical part (term grade):

Criterion: Being active on the practicals. You can miss maximum 3 practicals.

Write a project work essay in groups of 3 people (using 3 different proteins) at the end of the semester using the results and the knowledge you have learned during the practicals.