Use of languages

Principal working language: english (eng)

Contact

Name: Sònia Casillas Viladerrams
Email: Sonia.Casillas@uab.cat

Teachers

Antoni Barbadilla Prados
Leonardo Pardo Carrasco
Pere Puig Casado
Alfredo Ruiz Panadero
Miquel Àngel Senar Rosell
Jean Didier Pie Marechal
Daniel Yero Corona
Raquel Egea Sánchez
Xavier Daura Ribera

External teachers

Cedric Notredame
Emanuel Raineri
Josep Abril
Sebastián Ramos

Prerequisites

Level B2 of English or equivalent is recommended.

Objectives and Contextualisation

This module focuses on the development of diverse bioinformatic tools and resources commonly used in Omics research. Our intention is that it covers several aspects of bioinformatics in a series of brief topics, in the form of “tastings”. Therefore, it is not an accumulative module, but a transversal one, which should provide with a wide range of ideas and approaches that bioinformatics offers, through the hands of experts. The main objective is to provide students with the necessary foundation to apply bioinformatics to different areas of scientific research. Over time, each student will be able to gain all the depth they propose on any of these topics, the one which finally represents their research framework.
Skills

- Analyse and interpret data deriving from omic technology using biocomputing methods.
- Design and apply scientific methodology in resolving problems.
- Possess and understand knowledge that provides a basis or opportunity for originality in the development and/or application of ideas, often in a research context.
- Propose biocomputing solutions for problems deriving from omic research.
- Propose innovative and creative solutions in the field of study
- Student should possess the learning skills that enable them to continue studying in a way that is largely student led or independent.
- Understand the molecular bases and most common standard experimental techniques in omic research (genomics, transcriptomics, proteomics, metabolomics, interactomics, etc.)
- Use and manage bibliographical information and computer resources in the area of study
- Use operating systems, programs and tools in common use in biocomputing and be able to manage high performance computing platforms, programming languages and biocomputing analysis.

Learning outcomes

1. Create and promote algorithms, calculation and statistical techniques and theories to resolve formal and practical problems deriving from the handling and analysis of biological data.
2. Design and apply scientific methodology in resolving problems.
3. Identify and apply algorithms in which the programs are based bioinformatic analysis.
4. Identify and classify the principle types of biomolecular data obtained from omic technology.
5. Possess and understand knowledge that provides a basis or opportunity for originality in the development and/or application of ideas, often in a research context.
6. Propose innovative and creative solutions in the field of study
7. Search for specific bioinformatics tools and bioinformatics resources in the network.
8. Student should possess the learning skills that enable them to continue studying in a way that is largely student led or independent.
9. Synthesise and interpret in a logical and reasoned manner the information from the molecular data bases and analyse it using biocomputing tools.
10. Understand the theoretical, statistical and biological bases, in which the programs are based bioinformatic analysis: sequence alignment, similarity search and multiple alignment, structure prediction, genome annotation, phylogenetic and evolutionary analysis.
11. Use and manage bibliographical information and computer resources in the area of study
12. Use the main molecular databases, the main standard formats of molecular data and integrate data from different data sources

Content

Statistical Inference
Professor Antonio Barbadilla

- Statistics: bridge between data and models
- Data Types
- Population and sample
- Experimental design
- Data Quality
- Exploration of Data
- Sample distribution and law of large numbers
- Statistical inference
- Central Limit Theorem
- Point estimation
- Estimation of confidence interval
- Hypothesis
- Elements of a test: H0, H1, statistical test, p value, significance level, type I and II errors, power
- Z test, t test, chi-square test, correlation test, regression, analysis of variance
- Interpretation of statistical significance
- Parametric versus nonparametric tests
- Selecting the appropriate statistical test (decision tree)
- Multivariate Testing
- Resampling

Statistics and Stochastic Processes for Sequence Analysis
Professor Pere Puig

a. Probability basics

b. The multinomial model
Simulating a multinomial sequence. Estimating probabilities.

c. The seqinr package

d. Markov chain models

e. Higher order Markov chain models
Concept and examples. Estimating the probabilities of transition. Comparison of higher order Markov chains.

f. Hidden Markov chain models
Concept and examples. Parameter estimation. Hidden states estimation.

g. An introduction to Generalized Linear Models
GLM basics. The Logistic model. The Poisson model.

Bayesian Inference
Professor Emmanuele Raineri

1. Curve fitting.
- Estimation of parameters of probability distributions: binomial, poisson and gaussian.
- Example: fitting a noisy dataset.
- Cross validation, overfitting and regularization.

2. Dimensional reduction.
- Principal component analysis, multidimensional scaling.
- Example: distinguishing cell types using methylation profiles.

3. Lasso regression.
- Variable selection in linear models.
- Penalized regression: Lasso and Elastic Net.
- Example: lasso regression in R.

Bioinformatics Formats and Databases
Professor Daniel Yero

a. Sequence formats
Nomenclature. Text editors. FASTA format and its variants. Raw/Plain format. Genbank sequence format. EMBL sequence format. GCG, NBRF/PIR, MSA, PHYLIP, NEXUS. Format conversion.

b. Databases
Workflows with Galaxy
Professor Raquel Egea

a. Introduction to workflow managers
Concept, origin and design of workflow managers. Workflow patterns. Existing workflow managers. APIs and
Web Servers.

b. Galaxy: basics, interface and practical uses.

Software Engineering
Professor Miquel Àngel Senar

a. Parallelization strategies and HPC
b. Cloud computing with Amazon Web Services
c. Version control system with Git and GitHub

Sequence Alignment
Professor Cedric Notredame

a. Evolution and comparison Models
Molecular clock. Protein structure and evolution. Substitution Matrices.

b. Dynamic Programming based sequence comparisons
Needleman and Wunsch algorithm. Smith and Waterman algorithm. Afine gap penalties computation. Linear
space computation of pairwise algorithms.

c. Blast and Database searches
The Blast algorithm. E-values and estimates of statistical significance. Database search strategies. PSI-Blast
and other evolutionary approaches.

d. Multiple Sequence Alignments: algorithms and strategies
Main applications of multiple sequence alignments. Most common algorithms. Multiple sequence alignment
strategies.

Gene and Control Region Finding
Professor Josep Abril

a. Gene prediction
Annotation: concept, databases, problems. Gene finding: search by signal, search by content, approaches
(ab-initio, homology search, comparative genomics, NGS), evaluation of software accuracy.

b. Finding regulatory motifs
DNA motifs: exact matching, regular expressions, position weight matrices, search trees, profiles, randomized
algorithms, logos and pictograms, software for motif finding. Regulatory domains. Histones. CRMs architecture
shadowing. NGS.

Population Genomics
Professor Alfredo Ruiz

a. Population genomics under neutrality in a finite population

b. Population genomics under selection
Natural selection. Probability of fixation of selected mutants. Fitness distribution of new mutants. Rate of
evolution.

c. Adaptive evolution and population size
Phylogeny and Molecular Evolution  
Professor Sebastián Ramos

a. Models of sequence evolution  

b. Phylogeny  

Introduction to Genomes  
Professor Alfredo Ruiz

a. Introduction to genomes  

b. The human genome: where are we now?  
Current assembly of the human genome. The ENCODE project: functional elements in the human genome. Repetitive content of the human genome.

Structural Bioinformatics  
Professors Xavier Daura, Leonardo Pardo and Jean-Didier Maréchal

a. Structural biology and interactions  

b. Structure databases  
PDB. PyMOL.

c. Molecular modeling  

d. Docking  

Systems Biology  
Professor TBD

a. Classical and Genomic age Systems Biology  
The systems biology paradigm in light of technological developments over the last 100 years. Data integration bottlenecks.

b. Mathematical modeling of molecular circuits.  
Conceptual models. From conceptual models to mathematical models. Mathematical formalisms. Data driven models.

c. Design and organization principles in molecular circuits.  

Methodology

The methodology will combine master classes, solving practical problems and real cases, working in the computing lab, performing individual and team work, reading articles related to the thematic blocks, and independent self-study. The virtual platform will be used.
Activities

<table>
<thead>
<tr>
<th>Title</th>
<th>Hours</th>
<th>ECTS</th>
<th>Learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Type: Directed</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solving problems in class and work in the biocomputing lab</td>
<td>39</td>
<td>1.56</td>
<td>1, 2, 4, 5, 6, 8, 9, 10, 11, 12</td>
</tr>
<tr>
<td>Theoretical classes</td>
<td>39</td>
<td>1.56</td>
<td>1, 2, 3, 4, 5, 6, 8, 9, 10, 11</td>
</tr>
<tr>
<td><strong>Type: Supervised</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performing individual and team works</td>
<td>40</td>
<td>1.6</td>
<td>2, 4, 5, 6, 7, 8, 9, 11</td>
</tr>
<tr>
<td><strong>Type: Autonomous</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regular study</td>
<td>178</td>
<td>7.12</td>
<td>2, 4, 5, 6, 7, 8, 9, 11</td>
</tr>
</tbody>
</table>

Evaluation

- Student's portfolio (55%): work done and presented by the student.
- Individual theoretical and practical test (35%): a final exam will take place at the end of this module. It will consist of one or two multiple-choice or short questions by each professor teaching in this module.
- Soft skills (10%): assistance, arrival on time and active participation in class.

Evaluation activities

<table>
<thead>
<tr>
<th>Title</th>
<th>Weighting</th>
<th>Hours</th>
<th>ECTS</th>
<th>Learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual theoretical and practical test</td>
<td>35%</td>
<td>4</td>
<td>0.16</td>
<td>2, 3, 4, 5, 6, 8, 9, 10, 11, 12</td>
</tr>
<tr>
<td>Soft skills</td>
<td>10%</td>
<td>0</td>
<td>0</td>
<td>2, 6, 11</td>
</tr>
<tr>
<td>Student's portfolio</td>
<td>55%</td>
<td>0</td>
<td>0</td>
<td>1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12</td>
</tr>
</tbody>
</table>

Bibliography

Updated bibliography will be recommended in each session of this module by the professor, and links will be made available on the Student's Area of the MSc Bioinformatics official website (http://mscbioinformatics.uab.cat/base/base3.asp?sitio=bioinformaticsintranet&anar=module_2 &item=&subitem=).

6