

**Bioinformatics**

Code: 101951  
ECTS Credits: 6

Degree	Type	Year	Semester
2500890 Genetics	OB	3	1

**Contact**

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

**Use of languages**

Principal working language: catalan (cat)  
Some groups entirely in English: No  
Some groups entirely in Catalan: No  
Some groups entirely in Spanish: No

**Teachers**

Antoni Barbadilla Prados  
Raquel Egea Sánchez

**Prerequisites**

It is convenient for the student to review the basic concepts acquired in the subjects of the second grade:

The module of Databases and fundamentals of programming within the subject Instrumental Techniques  
Molecular Genetics of Eukaryotes  
Molecular Genetics of Prokaryotes  
Population Genetics

The content on structure and function of proteins of the subject of Biochemistry

Also, this subject is complemented by the Genomics, proteomics and interactomics course that is done during the same semester, so taking both together is recommended.

Knowledge of English is highly recommended for reading information.

**A basic level of any programming language (e.g. Perl or Python) will be essential to be able to follow the sessions, both theoretical and practical, and complete the proposed activities.**

**Objectives and Contextualisation**

Bioinformatics -the treatment, computer analysis and the interpretation and modeling of molecular and genomic data- has acquired a fundamental role in genetics nowadays. The subject taught during this course constitutes an introductory vision to bioinformatics. The main objectives are:

- To provide students with the basic bioinformatics knowledge that will allow them both to use a variety of tools to perform information searches in the databases and to approach the computational analysis of sequences of nucleic acids, proteins and genomes
- To apply scripting for the automation of repetitive bioinformatic tasks
- To give a perspective of the potential of this discipline both in the research and professional spheres

## Skills

- Be able to analyse and synthesise.
- Be able to communicate effectively, orally and in writing.
- Describe and identify the structural and functional characteristics of nucleic acids and proteins including their different organisational levels.
- Describe the organisation, evolution, inter-individual variation and expression of the human genome.
- Develop self-directed learning.
- Know and apply the omic tools of genomics, transcriptomics and proteomics.
- Perceive the strategic, industrial and economic importance of genetics and genomics to life sciences, health and society.
- Reason critically.
- Use and interpret data sources on the genomes and macromolecules of any species and understand the basics of bioinformatics analysis to establish the corresponding relations between structure, function and evolution.
- Use and manage bibliographic information or computer or Internet resources in the field of study, in ones own languages and in English.

## Learning outcomes

1. Be able to analyse and synthesise.
2. Be able to communicate effectively, orally and in writing.
3. Defend the relevance of progress in the generation and interpretation of data on a genomic scale for our understanding and technological manipulation of organisms.
4. Develop self-directed learning.
5. Explain and apply the methods for the analysis and annotation of genomes.
6. List and explain the content of bioinformatics databases and perform searches for information.
7. Reason critically.
8. Use and interpret the results of bioinformatics applications in the molecular analysis of sequences.
9. Use and manage bibliographic information or computer or Internet resources in the field of study, in ones own languages and in English.
10. Use bioinformatics techniques and tools to describe and analyse the human genome.
11. Use the techniques, tools and methodologies used to describe, analyse and interpret the enormous amounts of data produced by high performance technologies.

## Content

**Topic 1.** Introduction to bioinformatics. Bioinformatics in genomic research. The three corners of bioinformatic analysis. The bioinformatic challenges of data integration.

**Topic 2.** Databases of interest for biosciences. Bibliographic databases. Construction of search strategies. Molecular databases. Search engines.

**Topic 3.** Sequence formats. Extraction of sequences from the primary databases. Sequence Formats. Tools for conversion between formats.

**Topic 4.** Sequence statistics. Sequence visualization, editing and analysis tools. Sequence statistics.

**Topic 5.** Sequence alignment. Comparison of sequences. Local and global alignment for dynamic programming. Scoring matrices: identity and similarity. Penalties for gaps. Dot-plots.

**Topic 6.** Similarity searches. Heuristic algorithms for similarity searching against databases: BLAST. Applications of BLAST. Search for remote counterparts using PSI-Blast. Tools for the automation of similarity searches.

**Topic 7.** Multiple alignment. Multiple sequence alignment. Progressive alignments. Distance matrices. The CLUSTAL package. Applications of multiple alignment. Tools for the automation of multiple alignment. Matrices of position (weight) and LOGOS.

**Topic 8.** Molecular phylogenetic reconstruction. The molecular phylogeny. Methods of phylogenetic inference. Tools for phylogenetic reconstruction. Examples of phylogenetic reconstruction.

**Topic 9.** Genome assembly. Algorithms for mapping reads against reference genomes. Formats and tools.

**Topic 10.** Genome annotation. Annotation of genes. Annotation of regulatory regions. Visualization of annotations through genome browsers.

**Topic 11.** Comparative and functional genomics. Different levels of comparison. Synthetic blocks. Rearrangement hotspots. Comparative view using genome browsers.

## Methodology

The teaching methodology includes two types of differentiated activities: classes of theory and practical classes in the computer room. Learning will also have individual and/or collective tutoring of students that will support them to solve specific questions or that they require because of their complexity or difficulty.

### Theoretical classes

Classes to transmit the basic concepts and information necessary to develop self-learning. Promotion of active participation of students.

### Computer classroom practices

These practices will be organized based on problems posed by teachers that will need to be solved using the different tools and bioinformatics analyzes.

### Seminars and/or conferences

In the last sessions of the course there will be seminars and/or conferences of researchers in the field of bioinformatics.

### Tutorials

Individuals or small groups for the resolution of doubts related to the subject. This type of activity can be done at the request of the students.

## Activities

Title	Hours	ECTS	Learning outcomes
<b>Type: Directed</b>			
Computer room practices	27	1.08	3, 4, 5, 6, 7, 2, 1, 9, 8, 10, 11
Seminars	3	0.12	3, 5, 8, 10, 11
Theory	25	1	3, 4, 5, 6, 7, 1, 9, 8, 10, 11
<b>Type: Supervised</b>			
Tutoring	4	0.16	4, 7, 2, 1
<b>Type: Autonomous</b>			
Bibliography reading	2	0.08	3, 4, 5, 6, 7, 2, 1, 9, 8, 10, 11
Elaboration of works	15	0.6	3, 4, 5, 6, 7, 2, 1, 9, 8, 10, 11

Problem solving	25	1	3, 4, 5, 6, 7, 1, 9, 8, 10, 11
Study	39	1.56	3, 4, 6, 7, 2, 1, 9, 8, 10, 11

## Evaluation

The evaluation system is organized in **five main activities** and there will be, in addition, a recovery exam and an optional final improvement activity. The details of the activities are:

### Main evaluation activities

- **Partial evaluations (1 and 2). Overall Weight 60%**

Partial assessments are combined tests that can consist of test questions, written answers and problem solving. These tests will evaluate all the contents taught in the subject in the form of a theoretical-practical test.

The weight of each one of the assessments will be **30%**.

In order to pass these evaluations, a **minimum grade of 4.0 must be obtained in each of them**.

- **Continuous assessment. Overall Weight 25%**

Throughout the course, teachers will present computational problems or issues related to the subject matter that students will have to solve in the form of evaluations.

The continued nature of this assessment means that the subject can not be evaluated unless there is a **minimum participation of 50% of the proposed tests**.

- **Integrative work. Overall Weight 15%**

The work will consist of solving a problem that will contain questions related to the different thematic blocks presented during the classes of theory and practices.

These works will be carried out in groups of 3-5 students and are of an autonomous nature. The teaching staff will tutor the work and solve doubts and questions of approach.

The students will deliver the work to the professors following the content guidelines, presentation and deadlines.

This activity is **compulsory** and non-participation will entail **non-evaluation of the subject**.

- **Computerclassroom practices.** They can **subtract** up to **1 point** from the final grade

This activity is **compulsory** and the absence without justification or non-profit of the activity may **subtract up to 1 point from the final grade** of the subject.

The continuous nature of this assessment means that the subject can not be evaluated unless there is a **minimum participation in 50% of the proposed sessions**.

- **Optional problems or the student's own creations.** They can **add** up to **1 point** on the final grade

Throughout the course the professors will propose complex computational problems that will be **optional**. Their resolution may **add up to 1 point on the final grade** of the subject.

### Retake exam

To be eligible for the retake process, the student should have been previously evaluated in a set of activities equaling **at least two thirds of the final score** of the course.

The **evaluations 1 and/or 2** can be retaken individually and the grade, if  $\geq 4$ , will make an average with those passed in the partial exams. In the event that the grade of the retake exam does not reach a **minimum of 4.0**, it cannot make an average and the student does not pass the subject.

Continuous assessment, integrative work and attendance to the computer classroom practices, due to its continued nature, **cannot be retaken**.

### Improvement of the final qualification

Students who have passed the assessments 1 and 2 and want to **improve their final qualification**, may opt for a final test. This test will include **all of the subject matter**. It is not possible to improve the note through works or other types of activities.

The degree of **difficulty** of this test will correspond to the objective of the test and, therefore, **may be superior to that of the partial evaluations**.

The student who submits to this test **waives the previous qualifications** and therefore, the **grade** of this **improvement test** will be the one that will **prevail** in the final grade even though it is lower than that obtained by partial.

### Formula of weighting of the final note

**Final note** = [(Evaluation 1 x 0.3) + (Evaluation 2 x 0.3) + (Continuous assessment x 0.25) + (Integrative work x 0.15)]

### Passed

The subject is considered passed if the **final grade is  $\geq 5.0$**

### Not evaluable

The student will be graded as "No Avaluable" if the weighthin of all conducted evaluation activities is less than 67% of the final score.

## Evaluation activities

Title	Weighting	Hours	ECTS	Learning outcomes
Computer room practices	Can subtract up to 1 point from the final grade	0	0	3, 4, 5, 6, 7, 2, 1, 9, 8, 10, 11
Continuous assessment	25	2.5	0.1	3, 4, 5, 6, 7, 2, 1, 9, 8, 10, 11
Integrative work	15	1.5	0.06	3, 4, 5, 6, 7, 2, 1, 9, 8, 10, 11
Optional problems or the student's own creation	Can add up to 1 point on the final grade	0	0	3, 4, 5, 6, 7, 2, 1, 9, 8, 10, 11
Partial exam 1 (theoretical-practical contents)	30	3	0.12	4, 6, 1, 9, 8, 11
Partial exam 2 (theoretical-practical contents)	30	3	0.12	3, 4, 5, 6, 1, 9, 8, 10, 11

## Bibliography

### Books

- Compeau, P i Pevzner, P. 2015. Bioinformatics Algorithms. An active learning approach. Volumes I & II. Active Learning Publishers LLC.
- Attwood, TK. i Parry-Smith, J. 1999. Introduction to Biocomputing. Longman, UK. (Introducción a la Bioinformática. Attwood y Parry-Smith. 2002. Pearson Educación, S.A.)
- Cristianini, N. Y M. W. Hahn. 2007. Introduction to Computational Genomics. A case studies approach. Cambridge Univ. Press.
- Pevzer, P & R Shamir. 2011. Bioinformatics for biologists. Cambridge Univ. Press.
- Samuelsson, T. 2012. Genomics and Bioinformatics. Cambridge Univ. Press.
- Scherer, S. 2008. A short guide to the Human Genome. Cold Spring Harbor Laboratory Press.
- Xiong, J. 2006. Essential bioinformatics. Cambridge Univ. Press.

### MOOCS and online resources:

- Biology meets Programming: Bioinformatics for Beginners (course by Compeau&Pevzner):  
<https://www.coursera.org/learn/bioinformatics>
- Journey to the Frontier of Computational Biology (Specialization by Compeau&Pevzner):  
<https://www.coursera.org/specializations/computational-biology>
- Web associated to the book "Compeau, P i Pevzner, P. 2015. Bioinformatics Algorithms. An active learning approach. Volumes I & II. Active Learning Publishers LLC":  
<http://bioinformaticsalgorithms.com/index.htm>
- Computational problems in Rosalind: <http://rosalind.info/problems/locations/>