# COURSE DATA

## DATA SUBJECT

**Code:** 44696
**Name:** Bioinformatic techniques
**Cycle:** Master's Degree
**ECTS Credits:** 2.5
**Academic year:** 2026-27

## STUDY (S)

| Degree | Center | Acad. year | Period |
|---|---|---|---|
| 2224 - Master's Degree in Research and Development in Biotechnology and Biomedicin | Facultat de Ciències Biològiques | 1 | First quarter |

## SUBJECT-MATTER

| Degree | Subject-matter | Character |
|---|---|---|
| 2224 - Master's Degree in Research and Development in Biotechnology and Biomedicin | New technology | COMPULSORY |

## COORDINATION

BARRIO ESPARDUCER ELADIO

# SUMMARY

Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data: nucleotide and protein sequences, genome organization, structure and variability, as well as the prediction of biomolecule structure and function.

With the advent of the era of genomics, bioinformatics has extended its field of study to the massive analysis of biological data, including those derived from humans, and therefore, has become as a very important support for biomedical and biotechnology research.

The contents of this course have been designed as an advanced extension of the basic knowledge and training acquired by the students during their degrees. The extended knowledge and skills acquired in this course must be defined as bioinformatic techniques, since the aim is not to train bioinformaticians, but biomedicine and biotechnology researchers able to understand and use bioinformatic analysis to the advanced user level.

# PREVIOUS KNOWLEDGE

## RELATIONSHIP TO OTHER SUBJECTS OF THE SAME DEGREE

There are no specified enrollment restrictions with other subjects of the curriculum.

## OTHER REQUIREMENTS

There are no enrollment restrictions with other subjects of the curriculum of the master or any prerequisites

# COMPETENCES / LEARNING OUTCOMES

## 2224 - Master's Degree in Research and Development in Biotechnology and Biomedicin

Access the necessary information within the specific subject area (databases, scientific articles, etc.) and have sufficient judgement to interpret and apply it.

Adquirir destrezas en el manejo de las metodologías avanzadas empleadas en las biociencias moleculares y en el registro anotado de actividades.

Apply critical reasoning and argumentation based on rational criteria.

Apply research experience acquired both in private companies and public organisations.

Aprendizaje en la redacción de artículos científicos en los campos de la Biomedicina y la Biotecnología.

Be able to integrate new technologies in their professional and/or research work.

Critically analyse one's own work and that of colleagues.

Design multidisciplinary experimental strategies in the field of molecular biosciences to solve complex biological problems, especially those related to human health.

Handle scientific information sources appropriately and assess them critically, integrating the information to contribute knowledge to multidisciplinary research teams.

Make proper use of IT tools, statistical and data simulation methods, applying IT tools and statistics to biomedical and biotechnological problems.

Make rapid and effective decisions in complex situations within one's professional or research activity by developing new and innovative working methodologies adapted to the scientific/research, technological or professional field in which the activity takes place.

Master the scientific method, the design of experimental protocols and the interpretation of results in the biomedical and biotechnological fields.

Prepare, write and present reports and projects in public in a clear and coherent manner, defend them with rigour and tolerance and respond satisfactorily to any criticism that may arise from the presentation.

Select and manage available resources (instrumental and human) to optimise research outcomes.

Students should apply acquired knowledge to solve problems in unfamiliar contexts within their field of study, including multidisciplinary scenarios.

Students should be able to integrate knowledge and address the complexity of making informed judgments based on incomplete or limited information, including reflections on the social and ethical responsibilities associated with the application of their knowledge and judgments.

Students should communicate conclusions and underlying knowledge clearly and unambiguously to both specialized and non-specialized audiences.

Students should demonstrate self-directed learning skills for continued academic growth.

Students should possess and understand foundational knowledge that enables original thinking and research in the field.

Work in a team, without discriminating between men and women, carry out professional or research work efficiently and acquire the ability to participate in research projects and scientific or technological collaborations.

## DESCRIPTION OF CONTENTS

### 1. Introduction to Bioinformatics

What is Bioinformatics? Multidisciplinary area. Bioinformatics applications. Basic knowledge on file formats and omics data analysis

### 2. Genome assembling methods

Types of data obtained by ultrasequencing (NGS), sequence formats and files. Sequence quality assessment and data pre-treatment. Genome assembly algorithms: 'de novo' and 'mapping'. Genome assembly quality assessment. Sequence data bases management: SRA. Re-sequencing data analysis: 'variant calling'

### 3. Genome annotation procedures

Genome annotation files and formats: GFF, EMBL, GenBank. Current annotation strategies: advance methods for gene prediction and annotation transfer. Gene identification algorithms. Genome annotation quality assessment

### 4. RNAseq gene expression analysis: data management.

Quality assessment of the RNAseq reads. Alignment of reads with a reference genome. Sequence structure and normalization. Gene transcription counting

## 5. Statistical concepts applied to bioinformatics

- Analysis design and data management
- Inferential Statistics. Confidence intervals and hypothesis testing
- Introduction to R and Bioconductor

## 6. Classical Statistics for omics data

- t test, ANOVA and non-parametric tests for omics data
- The problem of multiple comparisons. False Discovery Rate (FDR)
- Statistic analysis with R: performance and interpretation

## 7. Statistical modeling of omics data (I)

- Multivariable analysis concept
- Batch effect correction
- Clinical/Epidemiological data integration
- Lineal models

## 8. Statistical modeling of omics data (II)

- Negative binomial regression. RNAseq analysis.
- Statistic analysis with R: performance and interpretation

## 9. Predictive models with omics biomarkers

- Problem description
- Penalized regression methods
- Random Forest
- Penalized regression with R

## 10. Genome variation analysis

Phylogenomics: genome phylogenetic analysis. Population genomics: population structure analysis from genome variation. Selection detection at the genome level

# WORKLOAD

## PRESENCIAL ACTIVITIES

| Activity | Hours |
|---|---:|
| Computer classroom practice | 25,00 |
| **Total hours** | **25,00** |

## NON PRESENCIAL ACTIVITIES

| Activity | Hours |
|---|---:|
| Attendance at other activities | 1,00 |
| Individual or group project | 6,00 |
| Independent study and work | 10,00 |
| Preparation of lessons | 8,00 |
| Preparation for assessment activities | 5,00 |
| Resolution of case studies | 7,50 |
| **Total hours** | **37,50** |

# TEACHING METHODOLOGY

Classes will be theoretical-practical, so they will be hold in a computer room.

For each topic, there will be:

1) An introduction of theoretical concepts that require prior preparation tasks by the students (by reading texts provided by the teacher).

2) Active learning through problem solving and practical case studies, through which the students will acquire skills on different aspects of the subject.

3) Self learning to solve individual problems and new theoretical and practical exercises.

4) We can consider attending a conference, seminar, panel discussion, on topics related to the subject of the course and organized at the Burjassot-Paterna campus by the Master CCA or departments, centers, research institutes from this campus.

# EVALUATION

The evaluation will be based on a continuous assessment of each student, according to the different activities described in the Methodology section: attendance to all class activities, resolution, and presentation of all the works and activities, the degree of participation and involvement of the student in the process of learning. The specific aspects to evaluate are:

1. Objective exams on the contents of the course, either in the classroom or 'online' (based on a series of short questions or multiple-choice tests, regarding both theoretical and practical issues), and exercises and assignments along the course. These tests and exercises will be evaluated by each professor, of the different parts of the course. The total score will be a weighted mean of the scores of the different parts, according to the number of hours of each one. The score of this section will represent 80% of the final

course scoring.

2. Assessment of participation in classroom activities, tutorials, and other group activities. Among other things, this section will assess the ability to ask questions, propose answers and lead the group discussion, as well as attendance. The score of this section will represent 20% of the final course scoring.

To pass the course, it is mandatory to get a final score of 5 out of 10.

## REFERENCES

- Hedderich, J. Applied Statistics: Methods using R. Springer. 2024
- Korpelainen E, Tuimala J, Somervuo P, Huss M, Wong G. RNA-seq data analysis: a practical approach. CRC Press; 2014.
- Krijnen, WP. Applied Statistics for Bioinformatics using R 2nd edition, 2022
- Lee, JK (Ed.) Statistical Bioinformatics A Guide for Life and Biomedical Science Researchers WileyBlackwell, 2010
- Lemey P, Salemi M, Vandamme AM (Ed.) The phylogenetic handbook. A practical approach to phylogenetic analysis and hypothesis testing. 2nd ed. Cambridge: Cambridge University Press; 2009
- Masoudi-Nejad A, Narimani Z, Hosseinkhan N. Next Generation Sequencing and Sequence Assembly: Methodologies and Algorithms. Springer Briefs in Systems Biology, 2013
- Soh J, Gordon PMK, Sensen CW. Genome Annotation. Chapman & Hall/CRC, 2012
- Toll, AP (Ed.) Structural Bioinformatics. Ceed Publishing, 2012
- Yang, Z. Computational Molecular Evolution. Oxford: Oxford Univ. Press; 2006
- Baker, M. (2012). De novo genome assembly: what every biologist should know. Nature Methods, 9(4), 333337. http://doi.org/10.1038/nmeth.1935
- Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. Evolutionary Applications, 7(9), 10261042. http://doi.org/10.1111/eva.12178
- Gauthier J, Vincent AT, Charette SJ, Derome N. (2019). A brief history of bioinformatics. Briefings in Bioinformatics. 20(6):1981-1996. doi:10.1093/bib/bby063
- Li H, Durbin R. (2024) Genome assembly in the telomere-to-telomere era. Nat Rev Genet. doi:

10.1038/ s41576-024-00718-w

- Valiente-Mullor C, Beamud B, Ansari I, et al. (2021) One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. PLoS Comput Biol. 17(1): e1008678. doi:10.1371/journal.pcbi.1008678