

Teaching guide

IDENTIFICATION DETAILS

Degree:	Diploma in Biotechnological Research Methodology (UFV-Awarded Title associated with Biotechnology)
---------	--

Field of Knowledge:	Science
---------------------	---------

Faculty/School:	Experimental Science
-----------------	----------------------

Course:	ADVANCED BIOINFORMATICS
---------	-------------------------

Type:	Compulsory Internal
-------	---------------------

ECTS credits:	3
---------------	---

Year:	4
-------	---

Code:	20117
-------	-------

Teaching period:	Seventh semester
------------------	------------------

Teaching type:	Classroom-based
----------------	-----------------

Language:	English
-----------	---------

Total number of student study hours:	75
--------------------------------------	----

Teaching staff	E-mail
Rocío Rama Ballesteros	rocio.rama@ufv.es

SUBJECT DESCRIPTION

Bioinformatics is an interdisciplinary scientific field that brings together methods from various domains to manage, organize, and understand all types of data from biological systems. It combines computational, statistical, mathematical, and biotechnological tools.

With the advent of the genomic era, traditional methods and tools, which were useful for storing and studying biological molecules on a small scale on a personal computer, are no longer valid. Omics and post-genomic technologies, especially next-generation sequencing, that have emerged in the last decade have led to a paradigm shift in both the bioinformatics tools used and the design of experiments, organization of teams, and scientific and technical objectives to be achieved. It is important for students to realize that these technologies are revolutionizing many aspects of biomedical research, the pharmaceutical industry, diagnostic and clinical management practices, and associated biotechnological applications. Furthermore, the quantitative analysis methodologies provided by Bioinformatics are now systematically applied.

This course focuses on the synergies between biotechnology and new bioinformatics methods, their joint evolution,

and their main applications. The classes are predominantly practical, and after a theoretical introduction, students will directly experiment on the computer with the main bioinformatics applications in next-generation sequencing technology: generation and processing of massive sequences, short read alignment, genomic variant analysis, transcriptome analysis, genome assembly, and visualization and analysis of data at the genomic level.

GOAL

The main objective of the course is to introduce students to the bioinformatics concepts and techniques that are commonly used today in research and analysis laboratories through practice lessons. These techniques have changed significantly in recent years with the implementation of high-throughput sequencing technology and, therefore, are not part of the usual knowledge covered in classical bioinformatics courses. Classes are held in the computer lab and have a practical orientation so that students understand the types of tools (hardware and software), the format of the data, and the methods underlying data analyses for the most commonly used applications.

The specific aims of the subject are:

Know and understand the potential of multidisciplinary bioinformatics techniques in the post-genomic era, especially in the context of massive sequencing.

Apply bioinformatics to obtain and analyze information from NGS experiments, in particular: quality control, format conversion, massive alignment of short reads, characterization of genomic variants, transcriptomics by RNA-seq, ChIP-seq data analysis, genome assembly, and access to genomic database information.

Contextualize these post-genomic techniques and describe their applications in various fields of medicine, biology, pharmacy, and agriculture.

Apply the theoretical knowledge acquired to solve problems and practical cases related to different subjects.

Know the history and evolution of post-genomic bioinformatics.

Know the main methods, tools, and information resources to support large sequencing projects (annotation, assembly, etc.) and experiments in molecular biology (transcriptomics, proteomics, etc.) using next-generation sequencing technology.

Know the most common tools and operating systems (software) in bioinformatics.

Apply programming to analyze and visualize genomic data.

PRIOR KNOWLEDGE

- Basic fundamentals of bioinformatics techniques: sequence analysis, biological databases.
- Basic fundamentals of biochemistry, molecular biology, and omics disciplines.
- Knowledge of nucleic acid sequencing techniques.
- Familiarity with computers and their components.
- Familiarity with programming in Python.
- Theoretical and/or practical familiarity with experimental techniques for generating large-scale data, especially next-generation sequencing.

COURSE SYLLABUS

Lesson 1. Introduction and Next-Generation Sequencing: This topic reviews the history of molecular biology in the context of numerical and analytical methods that allow its quantification at the molecular level. Special emphasis is

placed on introducing computationally based analytical techniques, setting students in the context of massive sequencing, its technological foundation, challenges, and applications.

Lesson 2. Practical Introduction to the Linux Operating System: This topic covers basic commands for managing files and directories as well as executing programs from the command line.

Lesson 3. Programming in Python for Bioinformatics: This topic reinforces programming skills in Python and their application to biological data.

Lesson 4. Tools and Libraries in Bioinformatics: This topic explores tools for data visualization. Introduction to Biopython. Introduction to NumPy and Pandas.

Lesson 5. Variant Analysis: Students receive basic concepts of bioinformatics methods used for the detection of genomic point variants (mutations and SNPs) in massive sequencing experiments. The methods, information standards, and technologies used by modern sequencers to generate sequence information and their qualities on a large scale are reviewed. Additionally, notions on annotation methods and functional impact prediction of these variants are acquired. All class contents are covered in a theoretical-practical manner.

Lesson 6. RNA-seq: This topic reviews the history and concepts associated with the systematic analysis of gene expression, and a practical protocol for expression analysis based on data from massive sequencing is carried out.

Lesson 7. Introduction to Machine Learning and Its Application to Solving Biological Problems: This topic introduces machine learning and how to apply it to help solve biological problems.

EDUCATION ACTIVITIES

The proposed methodology for this course will consist of:

- Conducting classes in the computer lab, with three main methodologies:
 - a) Biotechnological, biomedical, or molecular context of the practice: Theoretical introduction, prior explanation of the practices by the instructor.
 - b) Open discussion of these aspects: All participate, with open questions posed by the instructor or by other students.
 - c) Solving problems on the computers, supported and guided by the instructor, reviewing the results together.
- Utilizing the University's Virtual Classroom for monitoring student activities.
- Tutoring and problem-solving related to each unit by the students with supervision from the instructor.

Student learning activities will consist of:

- Practical exercises in which the problems posed will be carried out on the computers under the supervision of the instructor.
- Raising questions and elements for open discussion in class.

DISTRIBUTION OF WORK TIME

CLASSROOM-BASED ACTIVITY	INDEPENDENT STUDY/OUT-OF-CLASSROOM ACTIVITY
30 hours	45 hours
<ul style="list-style-type: none"> • Lecture classes • Practical classes (completion of projects and programming problems) 	<ul style="list-style-type: none"> • Theoretical study • Preparation of exercises and practical cases

SKILLS

Apply bioinformatics to obtain and analyze information from NGS experiments, in particular: quality control, format conversion, massive alignment of short reads, characterization of genomic variants, transcriptomics by RNA-seq,

genome assembly, and access to information from genomic databases.

Contextualize post-genomic techniques and describe their applications in various fields of medicine, biology, pharmacy, and agriculture.

Know the history and evolution of post-genomic bioinformatics and the most common bioinformatics tools and operating systems (software).

Acquire the necessary skills to search, compare, annotate, and download genomic information.

Apply theoretical knowledge to solve problems and practical cases related to different subjects.

Know the main methods, tools, and information resources to support large sequencing projects (annotation, assembly,...) and experiments in molecular biology (transcriptomics, proteomics...) using next-generation sequencing technology.

LEARNING RESULTS

Introduction to using the Linux command terminal, by executing a wide range of commands necessary to carry out the exercises throughout the course.

Introduction to the analysis of large-scale data from DNA-seq, RNA-seq, and de novo Genome Assembly.

Introduction to the most common file formats and data in Next Generation Sequencing (NGS) technologies.

LEARNING APPRAISAL SYSTEM

The percentages for the learning evaluation system will be as follows:

(1) Regular evaluation system: The evaluation system, based on continuous assessment, distributes the students' grades as follows:

- Class activities and submissions (40%)
- Class attendance, participation, and completion of practical cases (10%)
- Continuous assessment tests (50%) A minimum of 5 out of 10 must be achieved in each part to average the grades.

(2) Alternative evaluation system: Students in their second or subsequent enrollments must contact the instructor to request to follow this system.

*** Plagiarism and the use of illegitimate means in evaluation tests will be sanctioned in accordance with the University's Evaluation Regulations and Code of Conduct.

BIBLIOGRAPHY AND OTHER RESOURCES

Basic

ean-Michel Claverie and Cedric Notredame. Bioinformatics for dummies 2nd ed., Indianapolis :Wiley,2007.

Adriana López, Alex L. Rojas Introducción a Python para estudiantes de ciencias Tunja :Editorial UPTC,2021.