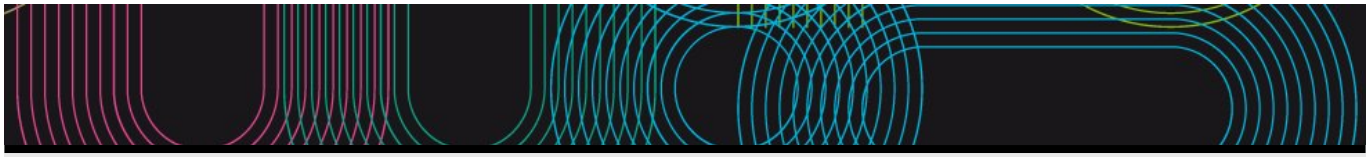

IDENTIFYING DATA				
Bioinformatics				
Subject	Bioinformatics			
Code	V02G031V01403			
Study programme	Grado en Biología			
Descriptors	ECTS Credits	Choose	Year	Quadmester
	6	Optional	4th	2nd
Teaching language	#EnglishFriendly Spanish			
Department				
Coordinator	Posada González, David			
Lecturers	Arenas Busto, Miguel Posada González, David			
E-mail	dposada@uvigo.es			
Web	http://www.uvigo.gal/es/universidad/administracion-personal/pdi/david-posada-gonzalez			
General description	This subject is intended to provide students with a first outlook into current bioinformatics. The approach will not consist of offering an overview of the various topics that bioinformatics contemplates today, which, by necessity, would imply a shallow perspective. On the contrary, we will focus on the analysis of high-throughput sequencing data for identifying genomic variants, deciphering gene expression, genome assembly and metagenomic characterization of the microbiome. On the one hand, the analysis of high-throughput sequencing data is nowadays extremely popular and transversal in multiple areas of biology. On the other hand, in order to perform this type of analysis reliably, the student must first acquire a series of concepts and transversal skills that will greatly facilitate subsequent learning of other aspects of bioinformatics. IMPORTANT: Using a laptop with the ability to connect to the internet via Wi-Fi is essential in all sessions. Those students who have problems meeting this requirement can go to the dean's office to borrow a laptop. The teaching methodology will consist mainly of lectures, practical sessions dedicated to problem solving, problem solving outside the classroom, and the use of the forum on the Moovi platform. Apart from asynchronous communication with the teacher through the Moovi platform, students can arrange virtual or face-to-face tutoring with the teacher via email or in person at any time. To be qualified, the student must upload a photo to the platform Moovi. (*) English Friendly subject. International students may request from the teachers: a) resources and bibliographic references in English, b) tutoring sessions in English, c) exams and assessments in English.			

Training and Learning Results	
Code	
A2	Students should know how to apply their knowledge to their work or vocation in a professional way. They also should have the competences that are usually proved through the elaboration and defence of arguments and the resolution of problems within their study field.
A5	Students should develop the necessary learning skills to undertake further studies with a high degree of autonomy.
B1	Developing autonomous learning by identifying their own training need and organizing and planning tasks and time.
B3	Apply the knowledge acquired in the degree and use the scientific-technical instrumentation and CIT in contexts of Biology and/or related to the professional practice.
B6	Develop analysis and synthesis, critical reasoning and argumentation skills, applying them in Biology and other scientific-technical disciplines.
C1	Solve problems by applying the scientific method, the concepts and terminology specific to biology, mathematical models and statistical and computer tools.
C2	Identify levels of organisation of living beings through the study of current specimens and fossils. Carry out phylogenetic analyses and study the mechanisms of heredity, evolution and biodiversity.
C5	Manipulate and analyse genetic material and determine its alterations and pathological implications. Knowing the applications of genetic engineering.

D4	Collaborate and work in teams or multidisciplinary groups, promote negotiation skills and the ability to reach agreements.
D5	Communicate effectively and appropriately, including the use of computer tools and English.

Expected results from this subject

Expected results from this subject	Training and Learning Results			
To recognize the role of Bioinformatics in the analysis and generation of hypotheses in Biology.	A2	B3 B6	C2	D5
To describe and understand key computational concepts, such as algorithms and relational databases, and their applications in biology.	A2 A5	B1 B3 B6	C1	D5
To understand and apply statistical methods commonly used in bioinformatics.	A5	B1 B3 B6	C1 C5	D4
To find, retrieve and organize different types of biological data.	A2 A5	B1 B3	C1 C2 C5	D4
To design simple bioinformatics applications.	A2 A5	B1 B3 B6	C1 C5	D4 D5
To practice reproducibility in bioinformatics.	A5	B1 B3	C1	D4 D5

Contents

Topic	
Lesson 1. Unix for Bioinformatics	Unix environment and command line. Remote servers. File access and manipulation. Regular expressions. Bash utilities and scripts.
Lesson 2. High-throughput DNA sequencing.	Sequencing platforms. Sequencing libraries. Sequencing coverage. FASTQ format. Read quality control.
Lesson 3. Sequence alignment	Alignment. Scoring. Alignment algorithms. Sequencing read mapping. SAM/BAM formats Post-processing.
Lesson 4. Variant calling	Types of variants. Identification. Calling strategies. VCF format. Structural variation. Filtering. Annotation
Lesson 5. Quantification of gene expression	RNA-seq. Experimental design. RNA-seq alignment. Quantification. Differential expression analysis.
Lesson 6. Genome assembly and annotation	Assembly. Algorithms. Evaluation. K-mer analysis. Genome annotation.
Lesson 7. Metagenomic analysis	Microbiome. Metagenomics. 16S analysis. Shotgun analysis. Alpha and beta diversity. Metagenomic annotation

Planning

	Class hours	Hours outside the classroom	Total hours
Lecturing	14	14	28
Problem solving	30	82	112
Discussion Forum	0	4	4
Objective questions exam	0.5	0	0.5
Objective questions exam	0.5	0	0.5
Objective questions exam	0.5	0	0.5
Objective questions exam	0.5	0	0.5
Problem and/or exercise solving	1	0	1
Problem and/or exercise solving	1	0	1
Problem and/or exercise solving	1	0	1
Problem and/or exercise solving	1	0	1

*The information in the planning table is for guidance only and does not take into account the heterogeneity of the students.

Methodologies

	Description
Lecturing	Basic concepts and methodologies of the subject will be explained and developed.

Problem solving	Practical sessions of a computational nature will be carried out in which students will apply the concepts and methodologies developed theoretically. Students are required to use their own laptop. These practical sessions will include the manipulation and transfer of files in Unix, use of remote servers, bioinformatics programming, data quality control, read mapping, identification and annotation of genomic variants, quantification of gene expression, representation of data in R, genome assembly and metagenomic analysis. Furthermore, students will solve problems outside the classroom on the Moovi platform to strengthen the concepts and methodologies of the subject.
Discussion Forum	All students are expected to actively participate in the subject's discussion forum on the moovi platform, raising their doubts and responding to questions and comments from the teacher and other students.

Personalized assistance

Methodologies	Description
Lecturing	Students will be able to interact directly with the teacher in relation to the lectures through individualized tutorials to resolve doubts. This personalized attention can be given in the classroom, agreeing on the date and time for a face-to-face or virtual meeting, via email, or using the Moovi platform messaging.
Problem solving	Students will be able to interact directly with the teacher in relation to problem solving through individualized tutorials to clarify doubts. This personalized attention can be given in the classroom, agreeing on the date and time for a face-to-face or virtual meeting, via email, or using the Moovi platform messaging.

Assessment

	Description	Qualification	Training and Learning Results			
Objective questions exam	Partial 1: Lesson 1	5	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5
Objective questions exam	Partial 2: Lessons 2-3	5	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5
Objective questions exam	Partial 3: Lesson 4	5	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5
Objective questions exam	Partial 4: Lessons 5-7	5	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5
Problem and/or exercise solving	Partial 1: Problem-solving sessions 1-3	20	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5
Problem and/or exercise solving	Partial 2: Problem-solving sessions 4-5	20	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5
Problem and/or exercise solving	Partial 3: Problem-solving sessions 6-7	20	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5
Problem and/or exercise solving	Partial 4: Problem-solving sessions 8-10	20	A2 A5	B1 B3 B6	C1 C2 C5	D4 D5

Other comments on the Evaluation

Throughout the course there will be four eliminatory partial tests, each one with a weight of 25% of the final grade. Each midterm will contain objective questions (20% of the grade) and problems (80% of the grade).

In June and July, students may retake any of these midterm exams.

Students who choose the global evaluation modality within the deadline established by the center may opt in June and/or July for 100% of the grade by taking the four midterm exams.

In all cases, in order to pass the subject it will be necessary to obtain 5 points out of 10 in the final grade.

Students who take a test will be considered as having presented themselves.

Dishonest behavior (e.g., plagiarism, cheating during exams, falsification of documents) may result in a failure of the subject.

The exam schedule is available at http://bioloxia.uvigo.es/es/docencia/examenes.

Sources of information

Basic Bibliography

Kappelmann-Fenzl M (editor), **Next Generation Sequencing and Data Analysis**, 1, Springer, 2021

Kappelmann-Fenzl M (editor), **Next Generation Sequencing and Data Analysis**, 1, Springer, 2021

Lloyd L, Tammi M (editors), **Bioinformatics: A Practical Handbook of Next Generation Sequencing and Its Applications.**, 1, World Scientific, 2017

Lesk A, **Introduction to Bioinformatics**, 5, Oxford University Press,, 2019

Complementary Bibliography

Pevsner J, **Bioinformatics and Functional Genomics.**, 3, Wiley, 2015

Buffalo V, **Bioinformatics Data Skills**, 1, O'Reilly, 2015

Allesina S., Wilmes M., **Computing Skills for Biologists.**, 1, Princeton University Press, 2019

Recommendations

Subjects that it is recommended to have taken before

Biology: Informatic tools in biology/V02G031V01110
Statistics: Biostatistics/V02G031V01107
Mathematics: Mathematics applied to Biology/V02G031V01104
Genetics I/V02G031V01209
Genetics II/V02G031V01304

Other comments

Considerations before enrolling in Bioinformatics: https://darwin.uvigo.es/docencia/binf2425/matricula.html

This subject is almost entirely problem-solving based. It will imply a continuous effort of several hours throughout the weeks of the course. Learning is sequential and each new step depends on the previous ones, much like mathematics in that sense. You will have to work a lot on your own, repeating tasks and checking solutions. We have 100 non face-to-face hours available, and you will have to use them.

No prior knowledge of any programming language is required to take this course, but basic knowledge of the use of computer tools (e.g., operating a laptop; opening and closing programs; accessing the internet) is required.