



| Teaching Guide | | | | |
|--------------------------|---|--------|--|---------|
| Identifying Data | | | | 2017/18 |
| Subject (*) | Data structures and algorithms for biological sequences | Code | 614522013 | |
| Study programme | Mestrado Universitario en Bioinformática para Ciencias da Saúde | | | |
| Descriptors | | | | |
| Cycle | Period | Year | Type | Credits |
| Official Master's Degree | 2nd four-month period | First | Obligatoria | 6 |
| Language | SpanishGalicianEnglish | | | |
| Teaching method | Face-to-face | | | |
| Prerequisites | | | | |
| Department | Computación | | | |
| Coordinador | Ladra González, Susana | E-mail | susana.ladra@udc.es | |
| Lecturers | Bernardo Roca, Guillermo de Ladra González, Susana | E-mail | guillermo.debernardo@udc.es susana.ladra@udc.es | |
| Web | | | | |
| General description | This course introduces the basics of the algorithms and data structures that are commonly used in the field of computational biology. | | | |

| Study programme competences / results | |
|---------------------------------------|--|
| Code | Study programme competences / results |
| A1 | CE1 - Ability to know the scope of Bioinformatics and its most important aspects |
| A2 | CE2 ? To define, evaluate and select the architecture and the most suitable software for solving a problem in the field of bioinformatics |
| A3 | CE3 ? To analyze, design, develop, implement, verify and document efficient software solutions based on an adequate knowledge of the theories, models and techniques in the field of Bioinformatics |
| A8 | CE8 - Understanding the basis of the information of the hereditary material, its transmission, analysis and evolution |
| A9 | CE9 ? To understand the benefits and the problems associated with the sequencing and the use of biological sequences, as well as knowing the structures and techniques for their processing |
| B1 | CB6 - Own and understand knowledge that can provide a base or opportunity to be original in the development and/or application of ideas, often in a context of research |
| B2 | CB7 - Students should know how to apply the acquired knowledge and ability to problem solving in new environments or little known within broad (or multidisciplinary) contexts related to their field of study |
| B8 | CG3 - Be able to work in a team, especially of interdisciplinary nature |
| C6 | CT6 - To assess critically the knowledge, technology and information available to solve the problems they face to. |
| C7 | CT7 ? To maintain and establish strategies for scientific updating as a criterion for professional improvement. |

| Learning outcomes | | | |
|--|---------------------------------------|-------------------|------------|
| Learning outcomes | Study programme competences / results | | |
| To know the data structures and the algorithms used for compactly storing and processing of biological sequences. | AJ1 AJ2 AJ9 | | |
| To analyze and compare the data structures and the complexity of the algorithms used. | AJ2 AJ3 | BJ1 | CJ6 CJ7 |
| To understand, analyze, design and implement solutions for different fundamental problems of sequence alignment, read error correction, contig assembly, gap filling, etc. | AJ1 AJ2 AJ3 AJ8 AJ9 | BJ1 BJ2 BJ8 | CJ6 CJ7 |



| | | | |
|--|-----|-----|-----|
| To explain, analyze, design and implement solutions to the problems related with evolution, such as haplotype assembly, motif finding, permutation patterns, genomic rearrangement, etc. | AJ1 | BJ1 | CJ6 |
| | AJ2 | BJ2 | CJ7 |
| | AJ3 | BJ8 | |
| | AJ8 | | |
| | AJ9 | | |

| Contents | |
|---|--|
| Topic | Sub-topic |
| Introduction to algorithms complexity analysis | Algorithms analysis Complexity |
| Sequence pattern search | Exact string matching methods Approximate string matching methods Suffix trees and suffix arrays |
| Introduction to sequence compression and indexing | Compression techniques Indexes and self-indexes |
| Applications to biological sequences | Sequence comparison Motif finding Genomic rearrangements Sequence alignment Sequence assembly Phylogenetic analysis |

| Planning | | | | |
|---------------------------------|-------------------------------|--------------------------------------|-------------------------------|-------------|
| Methodologies / tests | Competencies / Results | Teaching hours (in-person & virtual) | Student?s personal work hours | Total hours |
| ICT practicals | A2 A3 B1 B2 B8 C6 C7 | 14 | 42 | 56 |
| Supervised projects | A1 A2 A3 A8 A9 B1 B2 B8 C6 C7 | 3 | 30 | 33 |
| Mixed objective/subjective test | A1 A2 A3 A8 A9 B2 | 4 | 0 | 4 |
| Guest lecture / keynote speech | A1 A2 A3 A8 A9 | 28 | 28 | 56 |
| Personalized attention | | 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 | |
|---------------------------------|---|
| Methodologies | Description |
| ICT practicals | Students will complete practical exercises to develop all the knowledge acquired during lectures. |
| Supervised projects | Students will develop a work, individually or in small group, under the supervision of the teachers. |
| Mixed objective/subjective test | It consists of a written test to show that the student has acquired the knowledge and skills during lectures and practice sessions. |
| Guest lecture / keynote speech | Lectures where the course contents are exposed. |

| Personalized attention | |
|---------------------------------------|---|
| Methodologies | Description |
| Supervised projects ICT practicals | There may exist differences among the students regarding their background on algorithms and data structures. Thus, teachers will provide personalized attention for practice sessions and for the supervised project, both individual or in small groups. |



Assessment

| Methodologies | Competencies / Results | Description | Qualification |
|---------------------------------|-------------------------------|---|---------------|
| Mixed objective/subjective test | A1 A2 A3 A8 A9 B2 | It will consist of a written test where the students must prove the knowledge and competences acquired during lectures and practice sessions. To pass the course globally it is necessary to obtain in the mixed test a minimum grade of 1.5 (over 3). If that minimum grade is not achieved, the maximum grade cannot exceed 4.9 (and therefore the course is failed) | 30 |
| Supervised projects | A1 A2 A3 A8 A9 B1 B2 B8 C6 C7 | Students must complete a project, individually or in small groups, related with a scientific article. It must be presented orally. For the second opportunity the defenses will be done with a written test. | 20 |
| ICT practicals | A2 A3 B1 B2 B8 C6 C7 | The work done by the students during practice sessions will be assessed. Students must submit bulletins with their solutions to proposed problems and defend them orally. For the second opportunity the defenses will be done with a written test. | 50 |

Assessment comments

FIRST OPPORTUNITY:

Students that do not take the written exam will obtain a grade of "Non presentado" (Absent).

SECOND OPPORTUNITY:

Only those students that have not passed the course in the first opportunity can be evaluated in the second opportunity. Students can recover any of the parts as follows:

ICT practicals (50%): students must submit the exercises bulletins and defended them in a written test. Supervised project (20%): the defense of the project will be done using a written test. Written test (30%): in the same conditions as in the first opportunity. In case of not retaking one of the parts, the grade obtained in the first opportunity for that part will be kept. To pass the course globally it is necessary to obtain in the mixed test a minimum grade of 1.5 (over 3). Students that do not retake any part will obtain a grade of "Non presentado" (Absent).

ADVANCED OPPORTUNITY:
The assessment for the advanced opportunity will consist of a written exam that will compute for the 100% of the grade, and will include all the knowledge and skills acquired during lectures, practice sessions and supervised project.

ACADEMIC DISPENSATION:
Students officially enrolled part-time who have been granted an official dispensation from attending classes, as stipulated in the regulations of this University, must contact with the responsible of the course within the first two weeks to establish the conditions for submitting and defending the practical exercises and the supervised project.

Sources of information

| | |
|----------------------|--|
| Basic | <ul style="list-style-type: none"> - Dan Gusfield (1997). Algorithms on Strings, Trees and Sequences. Cambridge University Press - Neil C. Jones, Pavel A. Pevzner (2004). An Introduction to Bioinformatics Algorithms. MIT Press - Veli Mäkinen, Djamel Belazzougui, Fabio Cunial, Alexandru I. Tomescu (2015). Genome-Scale Algorithm Design. Cambridge University Press |
| Complementary | <ul style="list-style-type: none"> - Enno Ohlebusch (2013). Bioinformatics Algorithms: Sequence Analysis, Genome Rearrangements, and Phylogenetic Reconstruction. Oldenbusch Verlag - G. Navarro y M Raffinot (2002). Flexible Pattern Matching in Strings. Cambridge University Press - A. Moffat y A. Turpin (2002). Compression and Coding Algorithms. Kluwer Academic Publishers - T. C. Bell, J. G. Cleary y I. H. Witten (1990). Text Compression. Prentice Hall |

Recommendations

Subjects that it is recommended to have taken before



Introduction to molecular biology/614522004

Genetics and molecular evolution/614522005

Genomics/614522006

Fundamentals of bioinformatics/614522008

Introduction to programming/614522001

Subjects that are recommended to be taken simultaneously

Subjects that continue the syllabus

Advanced processing of biological sequences/614522020

New trends and applications in bioinformatics and biomedical engineering/614522021

Other comments

(*)The teaching guide is the document in which the URV publishes the information about all its courses. It is a public document and cannot be modified. Only in exceptional cases can it be revised by the competent agent or duly revised so that it is in line with current legislation.