



| Teaching Guide           |   |        |                       |         |
|--------------------------|---|--------|-----------------------|---------|
| Identifying Data         |   |        | 2020/21               |         |
| Subject (*)              | Data structures and algorithmics 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  | Obligatory            | 6       |
| Language                 | SpanishEnglish  |        |                       |         |
| Teaching method          | Hybrid  |        |                       |         |
| Prerequisites            |   |        |                       |         |
| Department               | Ciencias da Computación e Tecnoloxías da InformaciónComputación   |        |                       |         |
| Coordinador              | Ladra González, Susana  | E-mail | susana.ladra@udc.es   |         |
| Lecturers                | Ladra González, Susana  | E-mail | susana.ladra@udc.es   |         |
|                          | Silva Coira, Fernando   |        | fernando.silva@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. |        |                       |         |

## Contingency plan

### 1. Modifications to the contents

- No changes will be made

### 2. Methodologies

\*Teaching methodologies that are maintained

- Guest lecture / keynote speech
- ICT practicals
- Supervised projects
- Personalized attention

\*Teaching methodologies that are modified

- Mixed test: in case it cannot be performed in person, it will be replaced by "Problem solving")

### 3. Mechanisms for personalized attention to students

- Email: Daily. Used to make queries, request virtual meetings to resolve doubts and monitor the work being supervised.
- Moodle: Daily. According to the needs of the students. They will have "thematic forums" associated to the modules of the subject, to formulate the necessary consultations.
- Teams: Telematic session in the time slot assigned to the subject in the faculty's classroom calendar for the advancement of the expository and practical subject. Individual or small group telematic sessions to resolve doubts about the subject of the presentation, practice or supervised work

### 4. Modifications in the evaluation

The 30% of the mixed test is replaced by "Problem solving", which will also have a 30% weight in the evaluation. It will consist of the resolution of four exercises in which the knowledge and the competencies acquired during the master classes. The students will have to hand in the solution of the proposed exercises, having to defend it orally.

\*Evaluation observations:

The minimum requirements for the different parts of the evaluation will not be applied. The final mark will be the sum of those obtained in each part.

#### FIRST OPPORTUNITY:

Any student who does not send a proposal for the proposed problem-solving tasks will be marked as "Non presentado" (Absent).

#### SECOND OPPORTUNITY:

Only those students who do not pass the subject in the first opportunity will be able to present themselves to the second opportunity. The recovery of each part will be done in the following way:

- ICT practicals (50%): students will be able to repeat the proposed ICT practicals during the course in the same conditions as in the first opportunity (the internships delivered late will obtain a maximum of 80%). Thus, in case of repeating all the practices, the maximum mark that can be obtained is 4 points.
- Supervised project (20% of the final mark): carried out under the same conditions as at the first opportunity, but on an individual basis.
- Problem solving (30%): carried out under the same conditions as the first time.
- In case of not recovering any of the parts, the grade obtained at the first opportunity in that part will be kept.
- Any student who does not opt for the recovery of any of the parties will have a grade of "Non presentado" (Absent).



5. Modifications to the bibliography or webgraphy

- No changes will be made.



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

| Contents  |  |
|---|--|
| Topic   | Sub-topic  |
| Introduction to algorithms complexity analysis    | Algorithms analysis<br>Complexity  |
| Sequence pattern search                           | Exact string matching methods<br>Approximate string matching methods<br>Suffix trees and suffix arrays |
| Introduction to sequence compression and indexing | Compression techniques<br>Indexes and self-indexes   |



|                                      |  |
|--------------------------------------|--|
| Applications to biological sequences | Sequence comparison<br>Motif finding<br>Genomic rearrangements<br>Sequence alignment<br>Sequence assembly<br>Phylogenetic analysis |
|--------------------------------------|--|

| Planning  |                               |                      |                               |             |
|---|-------------------------------|----------------------|-------------------------------|-------------|
| Methodologies / tests   | Competencies                  | Ordinary class hours | Student's personal work hours | Total hours |
| ICT practicals  | A2 A3 B1 B2 B8 C6 C7          | 14                   | 60                            | 74          |
| 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             | 0                    | 5                             | 5           |
| Guest lecture / keynote speech  | A1 A2 A3 A8 A9                | 28                   | 10                            | 38          |
| Personalized attention  |                               | 0                    | 0                             | 0           |
| (*)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<br>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                  | Description  | Qualification |
| Mixed objective/subjective test | A1 A2 A3 A8 A9 B2             | <p>It will consist of a written test where the students must prove the knowledge and competences acquired during lectures and practice sessions.</p> <p>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)</p> | 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 in front of the teaching staff.   | 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 in front of the teaching staff.  | 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%): the students can repeat the ICT practicals under the same circumstances than in the first opportunity (those submitted out of time can obtain a maximum of 80% of the grade). Thus, in case of repeating all the assignments, the maximum grade will be 4 points.

Supervised project (20%): in the same conditions as in the first opportunity. 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 is equivalent to that of the first opportunity (50% ICT practicals, 20% supervised project, which must be done individually, and 30% written text).

**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

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

## 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.