UNIVERSIDADE DA CORUÑA

Teaching Guide				
Identifying Data				**2023/24**
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	Face-to-face			
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.			

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 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 AJ2 AJ3 AJ8 AJ9	BJ1 BJ2 BJ8	CJ6 CJ7

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**	**Ordinary class hours**	**Student?s personal work hours**	**Total hours**
ICT practicals	A2 A3 B1 B2 B8 C7 C6	14	70	84
Mixed objective/subjective test	A1 A2 A3 A8 A9 B2	3	0	3
Guest lecture / keynote speech	A1 A2 A3 A8 A9	28	32	60
Personalized attention		3	0	3
(*)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.
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. Active participation will be monitored on a continuous and objective basis.

Personalized attention	
Methodologies	**Description**
Guest lecture / keynote speech 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	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 4). If that minimum grade is not achieved, the maximum grade cannot exceed 4.9 (and therefore the course is failed)	40
Guest lecture / keynote speech	A1 A2 A3 A8 A9	There will be a continuous and objectifiable follow-up of the active participation during the master classes, by means of exercises delivered during these classes. This part of the evaluation will not be able to be recovered in the second opportunity.	10
ICT practicals	A2 A3 B1 B2 B8 C7 C6	The work done by the students during practice sessions will be assessed. Students must submit and defend their work in front of the teaching staff. Submissions out of time may result in a grade penalty.	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 circumpstances 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.Written test (40%): in the same conditions as in the first opportunity.The grade obtained from the continuous and objective monitoring of the active participation during the lectures cannot be recovered.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 4).Students that do not retake any part will obtain a grade of "Non presentado" (Absent).ADVANCED OPPORTUNITY:

The assessment for the advanced opportunity will be: 50% ICT practicals, 50% 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.

ACADEMIC FRAUD:

The fraudulent performance of tests or evaluation activities, once verified, will directly imply the qualification of failure in the call in which it is committed: the student will be graded with "suspenso" (numerical grade 0) in the corresponding call of the academic year, whether the commission of the fault occurs in the first opportunity or in the second. For this, the student's grade will be modified in the first opportunity report, if necessary.

Sources of information	
Basic	- 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, Djamal Belazzougui, Fabio Cunial, Alexandru I. Tomescu (2015). Genome-Scale Algorithm Design. Cambridge University Press
Complementary	- 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. Clearly 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

<div align="center">**Subjects that are recommended to be taken simultaneously**</div>

<div align="center">**Subjects that continue the syllabus**</div>

Advanced processing of biological sequences/614522020
New trends and applications in bioinformatics and biomedical engineering/614522021

<div align="center">**Other comments**</div>

Gender perspective: According to the different regulations applicable to university teaching, the gender perspective must be incorporated in this subject (use of non-sexist language, etc.). Work will be done to identify and modify sexist prejudices and attitudes and influence the environment to modify them and promote values of respect and equality. The aim will be to detect situations of gender discrimination and to propose actions and measures to correct them.