

**INTER AMERICAN UNIVERSITY OF PUERTO RICO  
METROPOLITAN CAMPUS  
SCHOOL OF SCIENCE AND TECHNOLOGY  
NATURAL SCIENCE DEPARTMENT  
MASTER'S IN SCIENCE IN MOLECULAR MICROBIOLOGY**

**SYLLABUS**

**I. GENERAL INFORMATION**

Course Title	:	Programming and Data Mining
Code and number	:	MOMI 5320
Credits	:	3
Academic term	:	
Instructor	:	
Office hours and location	:	
Office telephone	:	
E-mail	:	

**II. DESCRIPTION**

Application of open-source platforms (pipeline) for the analysis and mining of data in biological sequences. Emphasis on the analysis of DNA, RNA and protein sequences using computational tools such as Python, UNIX and R, among others. Requires 45 hours of conference-laboratory.

**III. OBJETIVES**

It is expected that at the end of the course, the student will be able to:

1. Apply programming languages and high-performance parallel processing computational platforms for biological data analysis.
2. Formulate an ethical approach to the management and transformation of scientific data and intellectual property.

**Competencies of the graduate profile addressed in this course:**

1. Evaluate the most useful platforms for analyzing molecular microbiology data using emerging tools.
2. Value the importance of ethical standards related to scientific conduct in research, respect for confidentiality and the defense of intellectual property.

#### IV. CONTENT

- A. Introduction to programming: the UNIX environment
- B. Unix Tutorial
  - 1. Tabular files
  - 2. Combining commands
  - 3. Modifying text files in the terminal (Command line interfaces)
  - 4. Creating a remote connection
  - 5. Copy files between machines
  - 6. Change file permissions
  - 7. Emacs text editor
- C. Introducción a *Python* y *IPython* (funciones interactivas)
- D. Inferential statistics and figures
- E. Design of Primers in Primer Prospector and using web tools
- F. Perl applications in bioinformatics
  - 1. Microarray data management
  - 2. Gene detection
- G. Creation of working pipelines
- H. Redistribution (parsing) of sequences
- I. Iterations in Python
- J. DNA sequence assembly
- K. Hierarchical clustering and clustering algorithms
- L. Tools in Galaxy for DNA and protein sequencing workflows

#### V. LEARNING ACTIVITIES

- 1. Illustrated lectures in power point format
- 2. Invited speakers
- 3. Computer lab
- 4. Exercises and computer simulations

#### VI. EVALUATION

The evaluation of the course will be based on the execution of 1 project, 2 mid-term exams and a final exam. The final grade will be calculated based on 100% as follows:

	Score	% of Final Grade
Project	100	25
2 Finals	200	50
Final Exams	100	25
<b>Total</b>	<b>400</b>	<b>100</b>

#### Class Attendance and Exams

Class attendance is mandatory. A student who needs to be absent from class should contact the professor before the class by phone or email. There will be no make-up exams, except for reasons of illness. In such case, make-ups will be offered with a doctor's excuse during the final exam period during the professor's office hours.

## **VII. SPECIAL NOTES**

### **A. Auxiliary services or special needs**

All students who require auxiliary services or special assistance must request these at the beginning of the course or as soon as they know that they need them, through the proper registry, in the Office of Orientation with Sr. José Rodríguez.

### **B. Honesty, fraud, and plagiarism**

Dishonesty, fraud, plagiarism and any other inappropriate behavior in relation to academic work constitutes major infractions sanctioned by the General Student Regulations. The major infractions, as stated in the General Student Regulations, may have as a consequence, suspension from the University for a definite period greater than one year or the permanent expulsion from the University, among others sanctions.

### **C. Use of electronic devices**

Cellular telephones and any other electronic device that could interrupt the teaching and learning processes or alter the environment leading to academic excellence will be deactivated. Any urgent situation will be dealt with, as appropriate. The handling of electronic devices that allow students to access, store or send data during evaluations or examinations is prohibited.

### **D. Compliance with the Provisions of Title IX**

The Federal Higher Education Act, as amended, prohibits discrimination because of sex in any academic, educational, extracurricular, and athletic activity or in any other program or function, sponsored or controlled by a higher education institution, whether or not it is conducted within or outside the property of the institution, if the institution receives federal funds.

In harmony with the current federal regulation, in our academic unit an Assistant Coordinator of Title IX has been designated to offer assistance and orientation in relation to any alleged incident constituting discrimination because of sex or gender, sexual harassment or sexual aggression. The Assistant Coordinator, Sr. George Rivera, can be reached by phone at 787-250-1912, extension 2262 or 2147, or by e-mail [griverar@metro.inter.edu](mailto:griverar@metro.inter.edu).

The Normative Document titled Norms and Procedures to Deal with Alleged Violations of the Provisions of Title IX is the document that contains the institutional rules to direct any complaint that appears to be this type of allegation. This document is available in the Web site of Inter American University of Puerto Rico ([www.inter.edu](http://www.inter.edu)).

## **VIII. EDUCATIONAL RESOURCES**

### **Supplementary readings**

Adler P, Kolde R, Kull M, Tkachenko A, Peterson H, Reimand J, Vilo J. Mining for coexpression across hundreds of datasets using novel rank aggregation and visualization methods. *Genome Biology* 2009;

Barret T, Troup DB, Wilhite SE, Ledoux P, Rudnev D, Evangelista C, Kim IF, Soboleva A, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Muerlter RN, Edgar R. NCBI GEO: archive for high-throughput functional genomic data. *Nucleic Acids Research* 2009; 37:D885-D890.

## **IX. BIBLIOGRAPHY**

### **Text**

Brian Kernighan, et al (1999), *The Practice of Programming*, W. Addison-Wesley, **ISBN-13: 978-0201615869**

Mark Lutz (2006) *Programming python 3<sup>rd</sup> edition*, O'Reilly Media, Inc. ISBN13: **978-0596009250**

**REv: 2022**