



Universidad Nacional Mayor de San Marcos
School of Computer Science
Syllabus of Course
Academic Period 2018-II

1. **Code and Name:** BI0021. Bioinformatics and Biostatistics (Mandatory)
2. **Credits:** 4
3. **Hours of theory and Lab:** 2 HT; 4 HL; (15 weeks)
4. **Professor(s)**

Meetings after coordination with the professor

5. Bibliography

- [Alu06] Srinivas Aluru, ed. *Handbook of Computational Molecular Biology*. Computer and Information Science Series. Boca Raton, FL: Chapman & Hall, CRC, 2006.
- [CB00] P. Clote and R. Backofen. *Computational Molecular Biology: An Introduction*. 279 pages. John Wiley & Sons Ltd., 2000.
- [Dur+98] R. Durbin et al. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998, p. 357. ISBN: 9780521629713.
- [Kro+94] Anders Krogh et al. "Hidden Markov Models in Computational Biology, Applications to Protein Modeling". In: *J Molecular Biology* 235 (1994), pp. 1501–1531.
- [Pev00] Pavel A. Pevzner. *Computational Molecular Biology: an Algorithmic Approach*. Cambridge, Massachusetts: The MIT Press, 2000.
- [SM97] João Carlos Setubal and João Meidanis. *Introduction to computational molecular biology*. Boston: PWS Publishing Company, 1997, pp. I–XIII, 1–296. ISBN: 978-0-534-95262-4.

6. Information about the course

- (a) **Brief description about the course** The use of computational methods in the biological sciences has become one of the key tools for the field of molecular biology, being a fundamental part of research in this area. In Molecular Biology, there are several applications that involve both DNA, protein analysis or sequencing of the human genome, which depend on computational methods. Many of these problems are really complex and deal with large data sets. This course can be used to see concrete use cases of several areas of knowledge of Computer Science such as Programming Languages (PL), Algorithms and Complexity (AL), Probabilities and Statistics, Information Management (IM), Intelligent Systems (IS).
- (b) **Prerequisites:** CS2102. Analysis and Design of Algorithms. (5th Sem)
- (c) **Type of Course:** Mandatory
- (d) **Modality:** Face to face

7. Specific goals of the Course

- That the student has a solid knowledge of molecular biological problems that challenge computing.
- That the student is able to abstract the essence of the various biological problems to pose solutions using their knowledge of Computer Science

8. Contribution to Outcomes

- a) An ability to apply knowledge of mathematics, science. (**Usage**)

b) An ability to design and conduct experiments, as well as to analyze and interpret data. (**Assessment**)

l) Develop principles research in the area of computing with levels of international competitiveness. (**Assessment**)

9. Competences (IEEE)

C1. An intellectual understanding and the ability to apply mathematical foundations and computer science theory.⇒
Outcome a,b

C3. An intellectual understanding of, and an appreciation for, the central role of algorithms and data structures.⇒
Outcome b,l

C5. Ability to implement algorithms and data structures in software.⇒ **Outcome a,b**

10. List of topics

1. Introduction to Molecular Biology
2. Sequence Comparison
3. Phylogenetic Trees
4. DNA Sequence Assembling
5. Secondary and tertiary structures
6. Probabilistic Models in Molecular Biology

11. Methodology and Evaluation

Methodology:

Theory Sessions:

The theory sessions are held in master classes with activities including active learning and roleplay to allow students to internalize the concepts.

Lab Sessions:

In order to verify their competences, several activities including active learning and roleplay will be developed during lab sessions.

Oral Presentations:

Individual and team participation is encouraged to present their ideas, motivating them with additional points in the different stages of the course evaluation.

Reading:

Throughout the course different readings are provided, which are evaluated. The average of the notes in the readings is considered as the mark of a qualified practice. The use of the UTEC Online virtual campus allows each student to access the course information, and interact outside the classroom with the teacher and with the other students.

Evaluation System:

12. Content

Unit 1: Introduction to Molecular Biology (4)	
Competences Expected: CS1	
Learning Outcomes	Topics
<ul style="list-style-type: none">• Achieve a general knowledge of the most important topics in Molecular Biology. [Familiarity]• Understand that biological problems are a challenge to the computational world. [Assessment]	<ul style="list-style-type: none">• Review of organic chemistry: molecules and macromolecules, sugars, nucleic acids, nucleotides, RNA, DNA, proteins, amino acids and levels of structure in proteins.• The Dogma of Life: From DNA to Proteins, Transcription, Translation, Protein Synthesis.• Genome study: Maps and sequences, specific techniques
Readings : [CB00], [SM97]	

Unit 2: Sequence Comparison (4)	
Competences Expected: CS2	
Learning Outcomes	Topics
<ul style="list-style-type: none"> • Understand and solve the problem of aligning a pair of sequences. [Usage] • Understand and solve the problem of multiple sequence alignment. [Usage] • Know the various algorithms for aligning existing sequences in the literature . [Familiarity] 	<ul style="list-style-type: none"> • Sequences of nucleotides and amino acid sequences. • Sequence alignment, paired alignment problem, exhaustive search, Dynamic programming, global alignment, local alignment, gaps penalty • Comparison of multiple sequences: sum of pairs, complexity analysis by dynamic programming, alignment heuristics, star algorithm, progressive alignment algorithms.
Readings : [CB00], [SM97], [Pev00]	

Unit 3: Phylogenetic Trees (4)	
Competences Expected: CS2	
Learning Outcomes	Topics
<ul style="list-style-type: none"> • Understand the concept of phylogeny, phylogenetic trees and the methodological difference between biology and molecular biology. [Familiarity] • Understand the problem of the reconstruction of phylogenetic trees, to know and apply the main algorithms for the reconstruction of phylogenetic trees. [Assessment] 	<ul style="list-style-type: none"> • Phylogeny: Introduction and phylogenetic relations • Phylogenetic trees: definition, type of trees, problem of search and reconstruction of trees • Reconstruction methods: parsimony methods, distance methods, maximum likelihood methods, confidence of reconstructed trees
Readings : [CB00], [SM97], [Pev00]	

Unit 4: DNA Sequence Assembling (4)	
Competences Expected: CS2	
Learning Outcomes	Topics
<ul style="list-style-type: none"> • Understand the computational challenge of the Sequence Assembly problem. [Familiarity] • Understand the principle of formal model for assembly. [Assessment] • Know the main heuristics for the problem of assembly of DNA sequences[Usage] 	<ul style="list-style-type: none"> • Biological basis: ideal case, difficulties, alternative methods for DNA sequencing • Formal Assembly Models: Shortest Common Superstring, Reconstruction, Multicontig • Algorithms for sequence assembly: representation of overlaps, paths to create superstrings, voracious algorithm, acyclic graphs. • Assembly heuristics: search for overlaps, ordering fragments, alignments and consensus.
Readings : [SM97], [Alu06]	

Unit 5: Secondary and tertiary structures (4)	
Competences Expected: CS2	
Learning Outcomes	Topics
<ul style="list-style-type: none"> • Know the protein structures and the necessity of computational methods for the prediction of the geometry. [Familiarity] • Know the algorithms for solving prediction problems of secondary structures RNA, and structures in proteins. [Assessment] 	<ul style="list-style-type: none"> • Molecular structures: primary, secondary, tertiary, quaternary. • Prediction of secondary structures of RNA: formal model, pair energy, structures with independent bases, solution with Dynamic Programming, structures with loops. • <i>Protein folding</i>: Estructuras en proteínas, problema de protein folding. • <i>Protein Threading</i>: Definitions, Branch Bound Algorithm, Branch Bound for protein threading. • <i>Structural Alignment</i>: Definitions, DALI algorithm
Readings : [SM97], [CB00], [Alu06]	

Unit 6: Probabilistic Models in Molecular Biology (4)	
Competences Expected: CS2	
Learning Outcomes	Topics
<ul style="list-style-type: none"> • Review concepts of Probabilistic Models and understand their importance in Computational Molecular Biology. [Assessment] • Know and apply Hidden Markov Models for various analyzes in Molecular Biology.. [Usage] • Know the application of probabilistic models in Phylogeny and to compare them with non-probabilistic models[Assessment] 	<ul style="list-style-type: none"> • Probability: Random Variables, Markov Chains, Metropoli-Hasting Algorithm, Markov Random Fields, and Gibbs Sampler, Maximum Likelihood. • Hidden Markov Models (HMM), parameter estimation, Viterbi algorithm and Baul-Welch method, Application in paired and multiple alignments, Motifs detection in proteins, in eukaryotic DNA, in sequences families. • Probabilistic phylogeny: probabilistic models of evolution, likelihood of alignments, likelihood for inference, comparison of probailistic and non-probabilistic methods
Readings : [Dur+98], [CB00], [Alu06], [Kro+94]	