Universidad Nacional San Cristobal de Huamanga (UNSCH)
Programa Profesional de
Ciencia de la Computación
Sílabo 2024-II

1. CURSO

CB309. Bioinformatics (Mandatory)

2. INFORMACIÓN GENERAL

2.1 Curso	:	CB309. Bioinformatics
2.2 Semestre	:	9th Semester.
2.3 Créditos	:	2
2.4 horas	:	1 HT; 2 HP;
2.5 Duración del periodo	:	16 semanas
2.6 Condición	:	Mandatory
2.7 Modalidad de aprendizaje	:	Face to face
2.8 Prerrequisitos	:	

- CS212. Analysis and Design of Algorithms. (5th Sem)
- MA307. Mathematics applied to computing . (6th Sem)

- CS212. Analysis and Design of Algorithms. (5th Sem)
- MA307. Mathematics applied to computing . (6th Sem)

3. PROFESORES

Atención previa coordinación con el profesor

4. INTRODUCCIÓN AL CURSO

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).

5. OBJETIVOS

- 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

6. RESULTADOS DEL ESTUDIANTE

1) Analyze a complex computing problem and apply principles of computing and other relevant disciplines to identify solutions. (Assessment)

2) Design, implement, and evaluate a computing-based solution to meet a given set of computing requirements in the context of the program's discipline. (Usage)

6) Apply computer science theory and software development fundamentals to produce computing-based solutions. (Usage)

7) Develop computational technology for the well-being of all, contributing with human formation, scientific, technological and professional skills to solve social problems of our community. (Usage)

7. TEMAS

Unidad 1: Introduction to Molecular Biology (4)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• 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	• Achive a general knowledge of the most important topics in Molecular Biology. [Familiarizarse] • Understand that biological problems are a challenge to the computational world. [Evaluar]
Lecturas : [CB00], [SM97]	

Unidad 2: Sequence Comparison (4)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• 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.	• Understand and solve the problem of aligning a pair of sequences. [Usar] • Understand and solve the problem of multiple sequence alignment. [Usar] • Know the various algorithms for aligning existing sequences in the literature . [Familiarizarse]
Lecturas : [CB00], [SM97], [Pev00]	

Unidad 3: Phylogenetic Trees (4)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• 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	• Understand the concept of phylogeny, phylogenetic trees and the methodological difference between biology and molecular biology. [Familiarizarse] • Understand the problem of the reconstruction of phylogenetic trees, to know and apply the main algorithms for the reconstruction of phylogenetic trees. [Evaluar]
Lecturas : [CB00], [SM97], [Pev00]	

Unidad 4: DNA Sequence Assembling (4)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• 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 overlays, ordering fragments, alignments and consensus.	• Understand the computational challenge of the Sequence Assembly problem. [Familiarizarse] • Understand the principle of formal model for assembly. [Evaluar] • Know the main heuristics for the problem of assembjale of DNA sequences[Usar]
Lecturas : [SM97], [Alu06]	

Unidad 5: Secondary and tertiary structures (4)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• 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. • *Protein folding*: Estructuras en proteinas, problema de protein folding. • *Protein Threading*: Definitions, Branch Bound Algorithm, Branch Bound for protein threading. • *Structural Alignment*: Definitions, DALI algorithm	• Know the protein structures and the necessity of computational methods for the prediction of the geometry. [Familiarizarse] • Know the algorithms for solving prediction problems of secondary structures RNA, and structures in proteins. [Evaluar]
Lecturas : [SM97], [CB00], [Alu06]	

Unidad 6: Probabilistic Models in Molecular Biology (4)	
Resultados esperados:	
Temas	**Objetivos de Aprendizaje (*Learning Outcomes*)**
• 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	• Review concepts of Probabilistic Models and understand their importance in Computational Molecular Biology. [Evaluar] • Know and apply Hidden Markov Models for various analyzes in Molecular Biology.. [Usar] • Know the application of probabilistic models in Phylogeny and to compare them with non-probabilistic models[Evaluar]
Lecturas : [Dur+98], [CB00], [Alu06], [Kro+94]	

8. PLAN DE TRABAJO

8.1 Metodología

Se fomenta la participación individual y en equipo para exponer sus ideas, motivándolos con puntos adicionales en las diferentes etapas de la evaluación del curso.

8.2 Sesiones Teóricas

Las sesiones de teoría se llevan a cabo en clases magistrales donde se realizarán actividades que propicien un aprendizaje activo, con dinámicas que permitan a los estudiantes interiorizar los conceptos.

8.3 Sesiones Prácticas

Las sesiones prácticas se llevan en clase donde se desarrollan una serie de ejercicios y/o conceptos prácticos mediante planteamiento de problemas, la resolución de problemas, ejercicios puntuales y/o en contextos aplicativos.

9. SISTEMA DE EVALUACIÓN

********* EVALUATION MISSING ********

10. BIBLIOGRAFÍA BÁSICA

[Kro+94] Anders Krogh et al. “Hidden Markov Models in Computational Biology, Applications to Protein Modeling”. In: *J Mol. Biol* 235 (1994), pp. 1501–1531.

[SM97] João Carlos Setubal and João Meidanis. *Introduction to computational molecular biology*. Boston: PWS Publishing Company, 1997, pp. I–XIII, 1–296.

[Dur+98] R. Durbin et al. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, 1998, p. 357.

[CB00] P. Clote and R. Backofen. *Computational Molecular Biology: An Introduction*. 279 pages. John Wiley & Sons Ltd., 2000.

[Pev00] Pavel A. Pevzner. *Computational Molecular Biology: an Algorithmic Approach*. Cambridge, Massachusetts: The MIT Press, 2000.

[Alu06] Srinivas Aluru, ed. *Handbook of Computational Molecular Biology*. Computer and Information Science Series. Boca Raton, FL: Chapman & Hall, CRC, 2006.