



## Universidad Nacional de Ingeniería (UNI)

Escuela Profesional de  
Ciencia de la Computación  
Sílabo 2024-II

### 1. CURSO

CB309. Bioinformática (Obligatorio)

### 2. INFORMACIÓN GENERAL

2.1 Curso	:	CB309. Bioinformática
2.2 Semestre	:	9 <sup>no</sup> Semestre.
2.3 Créditos	:	2
2.4 horas	:	1 HT; 2 HP;
2.5 Duración del periodo	:	16 semanas
2.6 Condición	:	Obligatorio
2.7 Modalidad de aprendizaje	:	Presencial
2.8 Prerrequisitos	:	<ul style="list-style-type: none"><li>• CS212. Análisis y Diseño de Algoritmos. (5<sup>to</sup> Sem)</li><li>• MA307. Matemática aplicada a la computación. (6<sup>to</sup> Sem)</li></ul>

### 3. PROFESORES

Atención previa coordinación con el profesor

### 4. INTRODUCCIÓN AL CURSO

El uso de métodos computacionales en las ciencias biológicas se ha convertido en una de las herramientas claves para el campo de la biología molecular, siendo parte fundamental en las investigaciones de esta área.

En Biología Molecular, existen diversas aplicaciones que involucran tanto al ADN, al análisis de proteínas o al secuenciamiento del genoma humano, que dependen de métodos computacionales. Muchos de estos problemas son realmente complejos y tratan con grandes conjuntos de datos.

Este curso puede ser aprovechado para ver casos de uso concretos de varias áreas de conocimiento de Ciencia de la Computación como: Lenguajes de Programación (PL), Algoritmos y Complejidad (AL), Probabilidades y Estadística, Manejo de Información (IM), Inteligencia Artificial (IA).

### 5. OBJETIVOS

- Que el alumno tenga un conocimiento sólido de los problemas biológicos moleculares que desafían a la computación.
- Que el alumno sea capaz de abstraer la esencia de los diversos problemas biológicos para plantear soluciones usando sus conocimientos de Ciencia de la Computación

### 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

<b>Unidad 1: Introducción a la Biología Molecular (4 horas)</b>	
<b>Resultados esperados:</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje (<i>Learning Outcomes</i>)</b>
<ul style="list-style-type: none"> <li>• Revisión de la química orgánica: moléculas y macromoléculas, azúcares, ácidos nucleicos, nucleótidos, ARN, ADN, proteínas, aminoácidos y niveles de estructura en las proteínas.</li> <li>• El dogma de la vida: del ADN a las proteínas, transcripción, traducción, síntesis de proteínas</li> <li>• Estudio del genoma: Mapas y secuencias, técnicas específicas</li> </ul>	<ul style="list-style-type: none"> <li>• Lograr un conocimiento general de los tópicos más importantes en Biología Molecular. [Familiarizarse]</li> <li>• Entender que los problemas biológicos son un desafío al mundo computacional. [Evaluar]</li> </ul>
<b>Lecturas :</b> [CB00], [SM97]	

<b>Unidad 2: Comparación de Secuencias (4 horas)</b>	
<b>Resultados esperados:</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje (<i>Learning Outcomes</i>)</b>
<ul style="list-style-type: none"> <li>• Secuencias de nucleótidos y secuencias de aminoácidos.</li> <li>• Alineamiento de secuencias, el problema de alineamiento por pares, búsqueda exhaustiva, Programación dinámica, alineamiento global, alineamiento local, penalización por gaps</li> <li>• Comparación de múltiples secuencias: suma de pares, análisis de complejidad por programación dinámica, heurísticas de alineamiento, algoritmo estrella, algoritmos de alineamiento progresivo.</li> </ul>	<ul style="list-style-type: none"> <li>• Entender y solucionar el problema de alineamiento de un par de secuencias. [Usar]</li> <li>• Comprender y solucionar el problema de alineamiento de múltiples secuencias. [Usar]</li> <li>• Conocer los diversos algoritmos de alineamiento de secuencias existentes en la literatura. [Familiarizarse]</li> </ul>
<b>Lecturas :</b> [CB00], [SM97], [Pev00]	

<b>Unidad 3: Árboles Filogenéticos (4 horas)</b>	
<b>Resultados esperados:</b>	
<b>Temas</b>	<b>Objetivos de Aprendizaje (<i>Learning Outcomes</i>)</b>
<ul style="list-style-type: none"> <li>• Filogenia: Introducción y relaciones filogenéticas.</li> <li>• Árboles Filogenéticos: definición, tipo de árboles, problema de búsqueda y reconstrucción de árboles</li> <li>• Métodos de Reconstrucción: métodos por parsimonia, métodos por distancia, métodos por máxima verosimilitud, confianza de los árboles reconstruidos</li> </ul>	<ul style="list-style-type: none"> <li>• Comprender el concepto de filogenia, árboles filogenéticos y la diferencia metodológica entre biología y biología molecular. [Familiarizarse]</li> <li>• Comprender el problema de reconstrucción de árboles filogenéticos, conocer y aplicar los principales algoritmos para reconstrucción de árboles filogenéticos. [Evaluar]</li> </ul>
<b>Lecturas :</b> [CB00], [SM97], [Pev00]	

Unidad 4: Ensamblaje de Secuencias de ADN (4 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje ( <i>Learning Outcomes</i> )
<ul style="list-style-type: none"> <li>• Fundamento biológico: caso ideal, dificultades, métodos alternativos para secuenciación de ADN</li> <li>• Modelos formales de ensamblaje: <i>Shortest Common Superstring, Reconstruction, Multicontig</i></li> <li>• Algoritmos para ensamblaje de secuencias: representación de overlaps, caminos para crear <i>superstrings</i>, algoritmo voraz, grafos acíclicos.</li> <li>• Heurísticas para ensamblaje: búsqueda de sobreposiciones, ordenación de fragmentos, alineamientos y consenso.</li> </ul>	<ul style="list-style-type: none"> <li>• Comprender el desafío computacional que ofrece el problema de Ensamblaje de Secuencias. [Familiarizarse]</li> <li>• Entender el principio de modelo formal para ensamblaje. [Evaluar]</li> <li>• Conocer las principales heurísticas para el problema de ensamblaje de secuencias ADN [Usar]</li> </ul>
Lecturas : [SM97], [Alu06]	

Unidad 5: Estructuras secundarias y terciarias (4 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje ( <i>Learning Outcomes</i> )
<ul style="list-style-type: none"> <li>• Estructuras moleculares: primaria, secundaria, terciaria, cuaternaria.</li> <li>• Predicción de estructuras secundarias de ARN: modelo formal, energía de pares, estructuras con bases independientes, solución con Programación Dinámica, estructuras con bucles.</li> <li>• <i>Protein folding</i>: Estructuras en proteínas, problema de <i>protein folding</i>.</li> <li>• <i>Protein Threading</i>: Definiciones, Algoritmo <i>Branch &amp; Bound, Branch &amp; Bound</i> para <i>protein threading</i>.</li> <li>• <i>Structural Alignment</i>: definiciones, algoritmo DALI</li> </ul>	<ul style="list-style-type: none"> <li>• Conocer las estructuras protéicas y la necesidad de métodos computacionales para la predicción de la geometría. [Familiarizarse]</li> <li>• Conocer los algoritmos de solución de problemas de predicción de estructuras secundarias ARN, y de estructuras en proteínas. [Evaluar]</li> </ul>
Lecturas : [SM97], [CB00], [Alu06]	

Unidad 6: Modelos Probabilísticos en Biología Molecular (4 horas)	
Resultados esperados:	
Temas	Objetivos de Aprendizaje ( <i>Learning Outcomes</i> )
<ul style="list-style-type: none"> <li>• Probabilidad: Variables aleatorias, Cadenas de Markov, Algoritmo de Metropoli-Hasting, Campos Aleatorios de Markov y Muestreador de Gibbs, Máxima Verosimilitud.</li> <li>• Modelos Ocultos de Markov (HMM), estimación de parámetros, algoritmo de Viterbi y método Baul-Welch, Aplicación en alineamientos de pares y múltiples, en detección de Motifs en proteínas, en ADN eucariótico, en familias de secuencias.</li> <li>• Filogenia Probabilística: Modelos probabilísticos de evolución, verosimilitud de alineamientos, verosimilitud para inferencia, comparación de métodos probabilísticos y no probabilísticos</li> </ul>	<ul style="list-style-type: none"> <li>• Revisar conceptos de Modelos Probabilísticos y comprender su importancia en Biología Molecular Computacional. [Evaluar]</li> <li>• Conocer y aplicar Modelos Ocultos de Markov para varios análisis en Biología Molecular. [Usar]</li> <li>• Conocer la aplicación de modelos probabilísticos en Filogenia y compararlos con modelos no probabilísticos [Evaluar]</li> </ul>
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 Molecular Biology* 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.