

Universidad Católica San Pablo (UCSP)
Escuela Profesional de
Ciencia de la Computación
SILABO



CB309. Computación Molecular Biológica (Electivo)

1. Información general

1.1 Escuela	:	Ciencia de la Computación
1.2 Curso	:	CB309. Computación Molecular Biológica
1.3 Semestre	:	9 ^{no} Semestre.
1.4 Prerrequisitos	:	<ul style="list-style-type: none">• CS212. Análisis y Diseño de Algoritmos. (5^{to} Sem)• MA307. Matemática aplicada a la computación. (6^{to} Sem)
1.5 Condición	:	Electivo
1.6 Modalidad de aprendizaje	:	Virtual
1.7 horas	:	2 HT; 2 HP; 2 HL;
1.8 Créditos	:	4

2. Profesores

3. Fundamentación del 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), Sistemas Inteligentes (IS).

4. Resumen

1. Introducción a la Biología Molecular 2. Comparación de Secuencias 3. Árboles Filogenéticos 4. Ensamblaje de Secuencias de ADN 5. Estructuras secundarias y terciarias 6. Modelos Probabilísticos en Biología Molecular

5. Objetivos Generales

- 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. Contribución a los resultados (*Outcomes*)

Esta disciplina contribuye al logro de los siguientes resultados de la carrera:

- 1) Analizar un problema computacional complejo y aplicar los principios computacionales y otras disciplinas relevantes para identificar soluciones. (**Evaluar**)
- 2) Diseñar, implementar y evaluar una solución basada en computación para cumplir con un conjunto determinado de requisitos computacionales en el contexto de las disciplinas del programa. (**Usar**)
- 6) Aplicar fundamentos de teoría de ciencias de la computación y desarrollo de software para producir soluciones basados en computación. (**Usar**)
- 7) Desarrollar tecnología computacional buscando el bien común, aportando con formación humana, capacidades científicas, tecnológicas y profesionales para solucionar problemas sociales de nuestro entorno. (**Usar**)

7. Contenido

UNIDAD 1: Introducción a la Biología Molecular (4)

Competencias:

Contenido

- Revisión de la química orgánica: moléculas y macromoléculas, azúcares, ácidos nucleicos, nucleótidos, ARN, ADN, proteínas, aminoácidos y niveles de estructura en las proteínas.
- El dogma de la vida: del ADN a las proteínas, transcripción, traducción, síntesis de proteínas
- Estudio del genoma: Mapas y secuencias, técnicas específicas

Objetivos Generales

- Lograr un conocimiento general de los tópicos más importantes en Biología Molecular. [Familiarizarse]
- Entender que los problemas biológicos son un desafío al mundo computacional. [Evaluar]

Lecturas: Clote and Backofen (2000), Setubal and Meidanis (1997)

UNIDAD 2: Comparación de Secuencias (4)

Competencias:

Contenido

- Secuencias de nucleótidos y secuencias de aminoácidos.
- Alineamiento de secuencias, el problema de alineamiento por pares, búsqueda exhaustiva, Programación dinámica, alineamiento global, alineamiento local, penalización por gaps
- Comparación de múltiples secuencias: suma de pares, análisis de complejidad por programación dinámica, heurísticas de alineamiento, algoritmo estrella, algoritmos de alineamiento progresivo.

Objetivos Generales

- Entender y solucionar el problema de alineamiento de un par de secuencias. [Usar]
- Comprender y solucionar el problema de alineamiento de múltiples secuencias. [Usar]
- Conocer los diversos algoritmos de alineamiento de secuencias existentes en la literatura. [Familiarizarse]

Lecturas: Clote and Backofen (2000), Setubal and Meidanis (1997), Pevzner (2000)

UNIDAD 3: Árboles Filogenéticos (4)	
Competencias:	
Contenido	Objetivos Generales
<ul style="list-style-type: none"> • Filogenia: Introducción y relaciones filogenéticas. • Árboles Filogenéticos: definición, tipo de árboles, problema de búsqueda y reconstrucción de árboles • Métodos de Reconstrucción: métodos por parsimonia, métodos por distancia, métodos por máxima verosimilitud, confianza de los árboles reconstruidos 	<ul style="list-style-type: none"> • Comprender el concepto de filogenia, árboles filogenéticos y la diferencia metodológica entre biología y biología molecular. [Familiarizarse] • Comprender el problema de reconstrucción de árboles filogenéticos, conocer y aplicar los principales algoritmos para reconstrucción de árboles filogenéticos. [Evaluar]
Lecturas: Clote and Backofen (2000), Setubal and Meidanis (1997), Pevzner (2000)	

UNIDAD 4: Ensamblaje de Secuencias de ADN (4)	
Competencias:	
Contenido	Objetivos Generales
<ul style="list-style-type: none"> • Fundamento biológico: caso ideal, dificultades, métodos alternativos para secuenciación de ADN • Modelos formales de ensamblaje: <i>Shortest Common Superstring</i>, <i>Reconstruction</i>, <i>Multicontig</i> • Algoritmos para ensamblaje de secuencias: representación de overlaps, caminos para crear <i>superstrings</i>, algoritmo voraz, grafos acíclicos. • Heurísticas para ensamblaje: búsqueda de sobreposiciones, ordenación de fragmentos, alineamientos y consenso. 	<ul style="list-style-type: none"> • Comprender el desafío computacional que ofrece el problema de Ensamblaje de Secuencias. [Familiarizarse] • Entender el principio de modelo formal para ensamblaje. [Evaluar] • Conocer las principales heurísticas para el problema de ensamblaje de secuencias ADN [Usar]
Lecturas: Setubal and Meidanis (1997), Aluru (2006)	

UNIDAD 5: Estructuras secundarias y terciarias (4)	
Competencias:	
Contenido	Objetivos Generales
<ul style="list-style-type: none"> • Estructuras moleculares: primaria, secundaria, terciaria, cuaternaria. • Predicción de estructuras secundarias de ARN: modelo formal, energía de pares, estructuras con bases independientes, solución con Programación Dinámica, estructuras con bucles. • <i>Protein folding</i>: Estructuras en proteínas, problema de <i>protein folding</i>. • <i>Protein Threading</i>: Definiciones, Algoritmo <i>Branch & Bound</i>, <i>Branch & Bound</i> para <i>protein threading</i>. • <i>Structural Alignment</i>: definiciones, algoritmo DALI 	<ul style="list-style-type: none"> • Conocer las estructuras protéicas y la necesidad de métodos computacionales para la predicción de la geometría. [Familiarizarse] • Conocer los algoritmos de solución de problemas de predicción de estructuras secundarias ARN, y de estructuras en proteínas. [Evaluar]
Lecturas: Setubal and Meidanis (1997), Clote and Backofen (2000), Aluru (2006)	

UNIDAD 6: Modelos Probabilísticos en Biología Molecular (4)	
Competencias:	
Contenido	Objetivos Generales
<ul style="list-style-type: none"> • Probabilidad: Variables aleatorias, Cadenas de Markov, Algoritmo de Metropoli-Hasting, Campos Aleatorios de Markov y Muestreador de Gibbs, Máxima Verosimilitud. • 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. • Filogenia Probabilística: Modelos probabilísticos de evolución, verosimilitud de alineamientos, verosimilitud para inferencia, comparación de métodos probailísticos y no probabilísticos 	<ul style="list-style-type: none"> • Revisar conceptos de Modelos Probabilísticos y comprender su importancia en Biología Molecular Computacional. [Evaluar] • Conocer y aplicar Modelos Ocultos de Markov para varios análisis en Biología Molecular. [Usar] • Conocer la aplicación de modelos probabilísticos en Filogenia y comparalos con modelos no probailísticos [Evaluar]
Lecturas: Durbin et al. (1998), Clote and Backofen (2000), Aluru (2006), Krogh et al. (1994)	

8. Metodología

El profesor del curso presentará clases teóricas de los temas señalados en el programa propiciando la intervención de los alumnos.

El profesor del curso presentará demostraciones para fundamentar clases teóricas.

El profesor y los alumnos realizarán prácticas

Los alumnos deberán asistir a clase habiendo leído lo que el profesor va a presentar. De esta manera se facilitará la comprensión y los estudiantes estarán en mejores condiciones de hacer consultas en clase.

9. Evaluar

Evaluación Continua 1 : 20 %

Examen parcial : 30 %

Evaluación Continua 2 : 20 %

Examen final : 30 %

References

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