

PROGRAMA de Introducción a la Bioinformática

Carreras: Tecnicatura Universitaria en Programación Informática - Licenciatura en Informática

Asignatura: Introducción a la Bioinformática

Núcleo al que pertenece: Orientación

Profesora: Ana Julia Velez Rueda

Asignaturas Correlativas: No posee

Objetivos:

- Proporcionar una visión general de la Bioinformática, así como sus campos de aplicación.
- Presentar algunos de los problemas más generales en la temática
- Proporcionar el mínimo conocimiento básico, para resolver problemas típicos.
- Presentar el análisis y desarrollo de nuevos algoritmos que permitan abordar problemas genéticos.
- Proporcionar una familiarización con los tipos de datos manejados por la Bioinformática.
- Presentar aplicaciones reales de la Bioinformática
- Mostrar técnicas matemáticas que permitan obtener indicios de posibles procesos celulares

Contenidos mínimos

- Conceptos básicos de la genética molecular: leyes de la herencia, genética de poblaciones, genética evolutiva, replicación del ADN, mutación y reparación.
- Acceso remoto a bancos de datos, bancos genéticos.
- Análisis de secuencias biológicas, algoritmos asociados.
- Homologías secuenciales y estructurales.

Carga horaria semanal: 4 hs

Programa analítico:

Unidad 1: Introducción a la Bioinformática.

Introducción de algunos conceptos biológicos. Descripción de algunos problemas biológicos. Datos de expresión génica. Aplicaciones. DNA y Proteínas. El código genético. Control de la expresión de genes

Unidad 2: Secuencias de ADN

Concepto de secuencias. Análisis de secuencias. Formato FASTA. Secuencias reguladoras. BLAST y NCBI Entrez. Algoritmos genéticos. Algoritmos de clustering. Simulaciones. Modelos computacionales de poblaciones.

Unidad 3: Genes

Características del código genético. Splicing. Marco abierto de lectura. Modelos estocásticos. Modelo oculto de Markov. Predicción de genes.

Unidad 4: Secuencias y proteínas

Alineamiento de secuencias. Alineamiento múltiple de secuencias. Árbol filogenético. Clustal W. Proteínas y proteómica. Análisis comparativo.

Unidad 5: Bases de datos en bioinformática

Revisión de conceptos de bases de datos. Modelo jerárquico. Modelo en red. Modelo orientado a objetos. Modelo declarativo. Descomposición y normalización. Biología de visualización de datos. Bases de datos biológicas

Unidad 6: Estructuras biológicas

Estructura de ADN (primaria, secundaria y terciaria). Estructura del ARN (primaria, secundaria y terciaria). Estructura de proteínas. Protein Data Bank. Folding de proteínas.
Alineamiento de estructuras.

Unidad 6: Búsqueda Exhaustiva

Mapas de restricción. Algoritmos para la construcción de mapas de restricción. Motivos reguladores en secuencias de ADN. Búsqueda de motivos

Unidad 7: Algoritmos Voraces.

Re-arreglo de genomas. Re-arreglo mediante permutaciones. Enfoque voraz para la búsqueda de motivos Algoritmos en grafos. Secuenciación de ADN por hibridación. Ensamblado por Shotgun. Secuenciación de proteínas

Unidad 8: Programación Dinámica

Alineamiento de pares de secuencias. Alineamiento múltiple de secuencias. Alineamiento clásico y probabilístico

Bibliografía obligatoria:

1. Neil C. Jones and Pavel A. Pevzner. An Introduction to Bioinformatics Algorithms. The MIT Press; 1 edition (August 1, 2004)
2. Michael S. Waterman. Introduction to Computational Biology: Maps, sequences and genomes. Chapman & Hall/CRC c2000
3. Dan Gusfield. Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press. c1999
4. Pavel Pevzner. Computational Molecular Biology: An Algorithmic Approach. The MIT Press; 1st edition (August 21, 2000)
5. I. Eidhammer, I. Jonassen, and W. R. Taylor. Protein Bioinformatics: An algorithmic approach to sequence and structure analysis. John Wiley & Sons, 2004.

Bibliografía de consulta:

- Jonathan Pevsner Bioinformatics and Functional Genomics. 2nd edition, Wiley -Blackwell, 2009).
- Klug-Spencer-Cummings, Conceptos de genética. Pearson. 2006.
- David Mount. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, 2001.
- Richard C. Deonier, Simon Tavaré, Michael S. Waterman. Computational Genome Analysis: An Introduction. Springer Verlag. 2005

Organización de las clases:

La materia se divide en clases teóricas (2hs. semanales), en las que se introducen los conceptos fundamentales, y clases de practica/laboratorio (2hs semanales), en las que se aplican los conceptos fundamentales para la resolución de distintos ejercicios prácticas introductorios en la temática en las primeras prácticas, y en las subsiguientes problemas de implementación y aplicación de los algoritmos en casos de estudios provistos por profesores/estudiantes investigadores de la carrera de Licenciatura en Bioinformática.

Trabajos Prácticos

Trabajo Práctico N.º 1: Biomoléculas. Los objetivos del trabajo práctico son: 1) familiarizarse con la terminología y conceptos propios de la Biología, necesarios para la comprensión de la materia; 2) estudiar las características y niveles estructurales de proteínas y ácidos nucleicos; 3) manipular la información secuencial. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta, el desarrollo de programas básicos para la manipulación de información secuencial y estructural.

Trabajo Práctico N.º 2: La vida en tres dimensiones. Los objetivos del trabajo práctico son: 1) familiarizarse con el uso de bases de datos de estructuras proteicas, en particular PDB; 2) comprender el formato PDB de estructuras proteicas y los alcances de los modelos de estructuras de biomoléculas; 3) manipular la información estructurada de los archivos PDB. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta, la obtención de estructuras cristalográficas de distintas biomoléculas, el análisis de la calidad y parseo de archivos PDB.

Trabajo Práctico N.º 3: Expresión génica. Los objetivos del trabajo práctico son: 1) familiarizarse con el código genético y la manipulación e interconversión de secuencias de proteínas y ácidos nucleicos mediante el uso del mismo; 2) manipular secuencias en formato FASTA y otros formatos de uso corriente en bioinformática. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta a esta temática, desarrollo de programas sencillos para la traducción y transcripción, y reconocimiento de motivos secuenciales (incluidos codones de stop y de inicio) y regiones promotoras.

Trabajo Práctico N.º 4: NCBI e introducción a alineamiento de secuencias. Los objetivos del trabajo práctico son: 1) familiarizarse con el uso de la base de datos NCBI; 2) familiarizarse con la búsqueda de secuencias proteicas y regiones codificantes y la anotación de secuencias, 3) introducir la complejidad y uso de

los alineamientos secuencias mediante el uso de CLUSTAL. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta a esta temática, búsqueda de secuencias proteicas y de ácidos nucleicos en la base de datos de NCBI y construcción y análisis de alineamiento orientativo de secuencias proteicas.

Trabajo Práctico N.º 5: Programación dinámica y algoritmos de alineamiento de secuencias. Los objetivos del trabajo práctico son: 1) familiarizarse con el uso de matrices de sustitución para la construcción de alineamientos de secuencias; 2) comprender las limitaciones y complejidades del alineamiento de secuencias biológicas, 3) comprender la necesidad de parametrización y penalización de apertura y cierre de gaps, 4) construcción de matrices de score y uso para la construcción de alineamientos de secuencias. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta a esta temática, comparación de algoritmos para alineamientos de secuencias (Needleman Wunsch y Smith Waterman) y comparación de resultados según parametrización diferencial.

Trabajo Práctico N.º 6: Búsqueda de similitud secuencial. Los objetivos del trabajo práctico son: 1) familiarizarse con el concepto de homología; 2) identificación de secuencias problema mediante uso de BLAST, 3) familiarizarse con los parámetros de corrida de BLAST para mejorar los resultados. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta a esta temática, identificación de una secuencia problema, comparación y análisis de resultados obtenidos con distintas parametrizaciones y bases de datos.

Trabajo Práctico N.º 7: Inferencias evolutivas. Los objetivos del trabajo práctico son: 1) familiarizarse con el uso de modelos de evolución para la construcción de árboles filogenéticos; 2) comprender las limitaciones y complejidades de las inferencias evolutivas e indagar en los distintos algoritmos que las calculan (Bayesianos, de Máxima Verosimilitud, NJ, etc.), 3) analizar el soporte de las inferencias evolutivas mediante métodos de bootstrapping. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta a esta temática, búsqueda de secuencias homólogas y outgroups para la construcción de árboles filogenéticos, análisis, lectura y visualización de los árboles filogenéticos.

Trabajo Práctico N.º 8: Análisis de estructuras proteicas. Los objetivos del trabajo práctico son: 1) familiarizarse con los métodos de predicción de subestructuras proteicas (túneles y cavidades); 2) comprender las relaciones entre secuencia y estructura sobre ejemplos de aplicación de diseño de fármacos in silico, 3) la relación estructura función en proteínas problema, 4) inferir plegamiento proteico a partir de una proteína homóloga con estructura conocida. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta a esta temática, búsqueda de estructuras de proteínas, predicción de túneles y cavidades, implementación de programas para la medida de distancia entre residuos.

Trabajo Práctico N.º 9: Ciencia de datos. Los objetivos del trabajo práctico son: 1) familiarizarse con el manejo de datos a gran escala; 2) estimaciones

estadísticas básicas, 3) estimación de correlaciones entre parámetros y clustering sobre set de datos biológicos. Las actividades incluyen la resolución de interrogantes teóricos de aplicación concreta a esta temática, análisis de un set de datos reales mediante el uso de PANDAS, graficación exploratoria con Seaborn, Clustering mediante uso de la librería scikit learn.

Modalidad de evaluación:

Los mecanismos de evaluación en modalidades libre y presencial de esta asignatura están reglamentados según los siguientes artículos del Régimen de estudios de la UNQ (Res. CS 201/18).

En la modalidad de libre, se evaluarán los contenidos de la asignatura con un examen escrito, un examen oral e instancias de evaluación similares a las realizadas en la modalidad presencial.

CRONOGRAMA TENTATIVO

Semana	Tema/unidad	Actividad*				Evaluación
		Teórico	Práctico			
			Res Prob.	Lab.	Otros Especificar	
1	Introducción a la Bioinformática.	x			Presentación de 2 TPs	
2	Biomoléculas – Práctica 1	x	x			
3	La vida en 3 dimensiones - Práctica 2	x	x			
4	Expresión Génica - Práctica 3	x	x			
5	NCBI e introducción a alineamiento de secuencias - Práctica 4		x	x		x
6	Entrega del TP 1	x	x			
7	Programación dinámica y algoritmos de alineamiento de secuencias - Práctica 5		x	x		
8	Búsqueda de similitud secuencial - Práctica 6	x	x			
9	Inferencias evolutivas - Práctica 7		x	x		
10	Análisis de estructuras proteicas - Práctica 8	x	x			
11	Ciencia de datos – Práctica 9		x	x		
12	Entrega TP2	x	x			
13	Parcial Teórico		x	x		x
14	Recuperatorio		x			
15	Consultas					x
16	Integrador					x

