

Planificación de la Asignatura: Análisis y Alineam. de Secuen.

Fecha: 23/10/2024 13:02

Código: L1335

Carrera: Licenciatura en Bioinformática

Departamento Académico: Biología

Docente a cargo:

Correo del docente a cargo: ileana.tossolini@uner.edu.ar

Régimen de Dictado: Cuatrimestral 2º Cuatrimestre

Carga Horaria Semanal: 6 horas semanales

Carga Horaria Total: 84 horas

Contenidos Mínimos:

Análisis de secuencias de ADN, técnicas de alineamiento de pares de secuencias, técnicas de alineamiento de secuencias múltiples

Correlativas Regulares para cursar:

Genética

Probabilidad y Estadística

Correlativas Aprobadas para cursar:

Algoritmos y Estructuras de Datos

Correlativas Aprobadas para promocionar o rendir el examen final:

Primer año

Algoritmos y Estructuras de Datos

Objetivo General:

La presente asignatura pretende que los alumnos conozcan los formatos de secuencias, las bases de datos de secuencias de ácidos nucleicos y proteínas mayormente utilizadas, los algoritmos principales para el análisis de dichas secuencias, que los estudiantes puedan implementar soluciones simples a problemas biológicos y que se familiaricen con los recursos disponibles. Para ello, las clases teóricas englobarán los conocimientos esenciales y las clases prácticas cubrirán la utilización de los recursos disponibles y la implementación de dichas soluciones. Se busca producir un espacio de aprendizaje que permita al alumno adquirir competencias que favorezcan su futuro desempeño como profesional, asumiendo una actitud ética, responsable, crítica, proactiva, creativa y comunicativa para la identificación y resolución de problemas contextualizados en la disciplina.

Objetivos Particulares:

1. Objetivos relacionados con las competencias específicas (conocimiento disciplinar) y competencias genéricas tecnológicas. Se espera que los/as alumnos/as:

- Conozcan los distintos formatos de anotación de secuencias.
- Identifiquen y utilicen las principales bases de datos de secuencias, y utilicen herramientas computacionales de búsqueda para recuperar secuencias particulares de estas bases de datos.
- Comprendan los métodos y herramientas de análisis de características de una secuencia.
- Comprendan lo que es un alineamiento de secuencias y cómo se determinan sus puntajes.
- Entiendan la diferencia de un alineamiento global y local de pares de secuencias y los usos de cada uno.
- Comprendan cómo utilizar los métodos de matrices de puntos para analizar secuencias.
- Entiendan los pasos realizados por los algoritmos de Needleman-Wunsch y Smith-Waterman para producir un alineamiento de pares de secuencias.
- Comprendan el significado del Score y el “E-value” en un alineamiento de pares de secuencias.
- Entiendan lo que es un alineamiento de secuencias múltiple y conozcan los métodos computacionales para realizarlo.
- Efectúen búsquedas de secuencias similares en bases de datos de secuencias.
- Utilicen las principales herramientas de análisis y alineamiento de secuencias: Blast suite, ClustalW, EMBOSS, Glimmer, Prokka, entre otras.
- Realicen el análisis de datos reales, y obtengan conclusiones acerca de las mejores prácticas para los análisis así como conclusiones biológicas a partir de los mismos.
- Puedan establecer criterios de selección de las distintas herramientas bioinformáticas para su implementación, según el problema y/o contexto analizado.

- Realicen scripts simples en Perl para la manipulación de datos de secuencia y los diversos resultados obtenidos por otros programas, y utilizarlos para automatizar diferentes tareas bioinformáticas.
- Comparen el lenguaje Perl con los otros lenguajes de programación empleados en bioinformática para utilizar en problemas biológicos concretos.
- Realicen y comprendan los pasos para el ensamblado de secuencias de ADN, a partir de datos de secuenciación de primera y segunda generación, y se familiaricen con el vocabulario específico de análisis de secuencias provenientes de técnicas de secuenciación masiva.
- Analicen trabajos científicos, tanto de software desarrollado para el análisis de secuencias, como trabajos que utilizan las técnicas y/o procedimientos vistos en la asignatura para abordar problemáticas biológicas actuales.

2. Objetivos del desarrollo de competencias genéricas sociales, políticas y actitudinales. Se espera que los/as alumnos/as:

- Adquieran y/o potencien un modo de pensar crítico, reflexivo y comprometido.
- Potencien su aprendizaje de manera continua y autónoma, y puedan comunicar contenidos técnicos vinculados a la profesión de manera efectiva.
- Exploten sus aptitudes para el desempeño en equipos de trabajo.
- Puedan resolver situaciones problemáticas, integrando conocimientos y alentando a la búsqueda creativa de soluciones.
- Apliquen la autoevaluación y la evaluación entre pares, para identificar fortalezas, debilidades y potencialidades de las diversas producciones de los demás estudiantes (informes escritos, exposiciones orales, trabajo final).
- Comprendan el rol de un/a Licenciado/a en Bioinformática en la interpretación de los resultados a partir de la información que se genera principalmente a partir de datos biológicos.

Programa Analítico:

En esta planificación, siempre que se hable de “secuencias” se hace referencia indistintamente a secuencias de ADN o de proteínas. Dado que los principios para el alineamiento tanto de a pares como múltiple de secuencias son los mismos, solo en aquellos casos que la metodología lo requiere se hace la diferenciación. La asignatura parte de los conocimientos adquiridos en las materias Biología Celular y Molecular, y Genética, reforzando los conceptos de información contenida en el genoma. Asimismo, se organizan los contenidos rescatando los conocimientos previos adquiridos en asignaturas como las vinculadas a la Programación, Base de Datos, Probabilidad y Estadística, de manera de establecer una integración entre las asignaturas del plan de estudio.

En la primera Unidad se cubren los aspectos informáticos de la organización de la información de forma sistemática en bases de datos, la forma de navegar y obtener información, así como el análisis de las características diferenciales de las secuencias. A partir de la segunda Unidad, se aborda la problemática del alineamiento entre secuencias, partiendo de la problemática más sencilla, aumentando en cada tema en nivel de complejidad. Se llega así a la tercera Unidad, la cual aborda las metodologías para comparar un número muy grande de secuencias alojado en bases de datos y los algoritmos desarrollados para resolver de forma rápida este complejo problema. La siguiente Unidad aborda la problemática de alinear un gran número de secuencias al mismo tiempo, y también se hace una diferenciación entre alineamiento local y global. Las dos últimas Unidades corresponden particularmente al análisis de secuencias, desde un punto de vista puramente informático en la Unidad de Programación, la cual presenta herramientas para desarrollar programas que solucionen tareas complejas. Por último, la Unidad de Secuenciación de Genomas integra gran parte de los contenidos desarrollados anteriormente, y permite la integración de muchos de los conceptos desarrollados en la cursada volcados en aplicaciones reales y complejas.

UNIDAD 1. GENOMAS Y SECUENCIAS COMO INFORMACIÓN BIOLÓGICA: El ADN como información genética. Formatos de datos de secuencias. Bases de datos de secuencias. Formatos de intercambio. Análisis de características de secuencias. Contenido de GC, regiones codificantes, repeticiones, tamaño de genomas y complejidad.

UNIDAD 2. ALINEAMIENTO DE SECUENCIAS: Alineamiento de secuencias. Alineamiento de pares de secuencias global y local. Métodos de matrices de puntos. Algoritmos de Needleman-Wunsch para alineamiento global y Smith-Waterman para alineamiento local. Matrices de puntuación y penalidades Blosum y PAM.

UNIDAD 3. ALINEAMIENTO CONTRA BASES DE DATOS: Búsqueda de secuencias en bases de datos. Algoritmos FASTA y BLAST. Conceptos básicos de probabilidad y estadística asociados al alineamiento de secuencias. Significancia estadística de los resultados obtenidos en un alineamiento. Puntaje (score), estadísticas, significado de los parámetros utilizados y resultados.

UNIDAD 4. ALINEAMIENTO MÚLTIPLE: Alineamiento múltiple de secuencias. Métodos computacionales de alineamiento múltiple. Matriz de puntuación de posición-específica, patrones, dominios. Búsquedas de secuencias similares utilizando modelos ocultos de Markov (HMMs). Conceptos de evolución. Inferencias filogenéticas.

UNIDAD 5. PROGRAMACIÓN PARA ANÁLISIS DE SECUENCIAS: Bioinformática bajo Linux, Bash, Perl, BioPerl. Conceptos básicos de programación en Perl, variables, control de flujo, manejo de archivos. Expresiones regulares. BioPerl como repositorio de módulos para análisis bioinformáticos. Comparación de Perl/BioPerl con Python/Biopython.

UNIDAD 6. SECUENCIACIÓN DE GENOMAS: Metodologías clásicas de secuenciación de genomas. Proyecto Genoma Humano. Ensamblado de secuencias de ADN. Algoritmos de overlap-layout-consensus y grafos. Conceptos básicos de cobertura, calidad, lecturas pareadas. Historia de las tecnologías de secuenciación, introducción al funcionamiento de las nuevas tecnologías de secuenciación (NGS). Predicción y anotación de genes codificantes en secuencias. Concepto de ORF, concepto de gen. Métodos intrínsecos y extrínsecos para la predicción de genes (técnicas ab initio, comparación por homología y métodos experimentales). Compilación de múltiples fuentes de información. Ontología génica (términos GO).

Listado de Actividades de Formación Práctica:

Son las actividades prácticas planificadas para el cuatrimestre (detalladas por semana en el cronograma extendido).

Utilización de las siguientes bases de datos y software específico para la realización de las prácticas de laboratorio propuestas:

- Readseq, SnapGene Viewer, Staden Package.
- International Nucleotide Sequence Databases (GenBank/EMBL/DDBJ). Entrez (NCBI), SRS (EBI). Uniprot. InterProScan. Prosite.
- EMBOSS, SMS, Codon Usage DB.

- Dotplot, dotmatcher, dotter, dotup.
- Needle, Water, Lalign.
- Blast web, Blast+.
- ClustalW, T-Coffee, MUSCLE.
- PHYLIP, treeview, figtree.
- HMMer, MEME/MAST, PSI-Blast.
- Glimmer, EasyGene, FgenesB, Prodigal, Prokka, Artemis.

Las prácticas van creciendo en nivel de complejidad e integración de conocimientos conforme se avanza en el cursado de la asignatura. Las mismas se basan principalmente en la resolución de casos y problemas fundamentados en situaciones reales a las cuales los bioinformáticos se enfrentan en el ejercicio profesional, y tienen la intención de desarrollar la fundamentación, pensamiento crítico y el criterio profesional.

Programación de scripts en lenguaje Perl: uso de Bioperl. Desarrollo de scripts para automatización de análisis de resultados, reformato de archivos, manejo de expresiones regulares. Comparación con Python y Biopython.

Trabajo Práctico Integrador:

Incluye ejercicios de programación de análisis de secuencias donde se buscará solución a un problema biológico real. Se trabajará bajo Linux utilizando programas de ensamblado de secuencias, alineamiento múltiple, comparación con bases de datos, programación básica de análisis de secuencias utilizando Perl / BioPerl, favoreciendo así al desarrollo de habilidades bioinformáticas avanzadas.

Metodología de Evaluación Durante el cursado:

Se realizó la modificación de la metodología de evaluación a partir del 2022, teniendo en cuenta el resultado de las encuestas internas de la cátedra y autoevaluación respondidas por los alumnos durante años anteriores, en las cuales la mayoría de los estudiantes manifiesta que un único parcial al final del cursado no les resulta adecuado para el seguimiento y aprobación de la materia. Durante las cursadas posteriores esta metodología de evaluación (2 parciales) arrojó buenos resultados y fue bien recibida por parte de los alumnos, por lo cual se decide continuar de esta manera.

Los alumnos serán evaluados cuantitativamente en dos Parciales individuales. Asimismo, se realizarán actividades semanalmente durante el coloquio, que serán evaluadas de manera conceptual, y que ayudarán a los alumnos a estudiar los contenidos de la materia durante el cursado.

El Parcial N° 1 se realizará en la Semana 7 y el N° 2 en la Semana 15. En estos parciales se evaluarán los diferentes conceptos teóricos vistos hasta la semana anterior al parcial, así como la utilización de las distintas herramientas bioinformáticas empleadas durante las clases prácticas de la materia. Los problemas tienen una estructura similar a los desarrollados en las prácticas de la asignatura.

Se plantea la realización de un Trabajo Práctico Integrador que no tendrá una nota numérica y ayudará a los alumnos a integrar los diversos contenidos prácticos y herramientas usadas durante el cursado, sirviendo de preparación además para el Parcial N° 2.

Respecto al Trabajo Práctico Integrador, del informe escrito presentado se evaluarán conceptualmente (cualitativamente) los resultados obtenidos y los criterios con los que se llevó a cabo. Se tendrá en cuenta que los resultados obtenidos sean coherentes con lo que se ha trabajado en clase, las respuestas acorde a la temática y completitud de las mismas, autonomía en el desarrollo de scripts de Perl, justificación pertinente de los programas elegidos, funcionalidad de los archivos generados, etc. Los resultados se expondrán de manera oral ante el resto de la clase.

Además se les pide a los estudiantes que al final del cursado realicen una Autoevaluación, la cual contribuye a un mayor entendimiento del propio proceso de aprendizaje. La misma consta de una rúbrica que deben completar respecto al desarrollo del Trabajo Integrador y su desempeño en los Parciales, y diversas preguntas relacionadas con los conceptos previos que se retomaron y/o ampliaron, con las nuevas habilidades adquiridas y respecto a la proyección de dichas actividades en el futuro profesional.

Metodología de Evaluación en Exámenes Finales:

Aquellos estudiantes regulares que no cumplan con la promoción directa (Nota final <80%), deberán rendir un examen final escrito en las fechas dispuestas según calendario académico sobre los contenidos desarrollados en las clases teórico-prácticas, siendo el resultado de esta instancia la nota final de la materia.

Los alumnos serán evaluados en examen final en dos etapas. En primer lugar se realizará una evaluación teórica, la cual consistirá en responder preguntas y resolver problemas que no requieran de computadoras. Luego de esta evaluación de conceptos teóricos, se les dará una serie de problemas prácticos que deben resolverse en la computadora y para los cuales los alumnos pueden hacer uso de toda la bibliografía disponible, apuntes, etc.

Para los alumnos libres, la evaluación cuenta con una pregunta teórica adicional, de mayor grado de desarrollo e interpretación, así como un problema práctico extra.

Condiciones de Regularidad :

Los alumnos deberán aprobar ambos parciales con un mínimo de 60% (promedio entre la parte teórica y la práctica) para lograr la regularización de la materia. La nota final de la asignatura será un promedio de ambos exámenes, considerando además la presentación del Trabajo Final Integrador. Cada parcial contará con un recuperatorio para aquellos que no logren el 60%.

La promoción de la materia se logrará si la nota promedio final supera el 80%, los alumnos que la logren no realizarán examen final. Los alumnos que regularicen pero no promocionen la materia deberán aprobar un examen final.

Bibliografía Principal:

Básica:

- BAXEVANIS, A. D. Y OUELLETTE, B. F. F. 2001. Bioinformatics – A practical guide to the analysis of genes and proteins. Wiley-Interscience (Second Edition)
- BEDELL, J.; KORF, I.; YANDELL, M. 2003. BLAST. O Reilly & Associates.
- DURBIN, R., S. R. EDDY, A. KROGH Y G. MITCHISON. 1999. Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids. Cambridge University Press.
- GIBAS, C. Y P. JAMBECK. 2001. Developing Bioinformatics Computer Skills. O Reilly & Associates.
- MOUNT, D. W. 2004. Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory. (Second edition)
- TISDALL, J. 2001. Beginning Perl for Bioinformatics. O Reilly & Associates.

Bibliografía Complementaria:

Se le brinda a los alumnos una carpeta compartida con trabajos científicos/manuales de los programas y metodologías utilizadas durante el dictado; así como también con artículos científicos y notas actuales sobre técnicas y avances actuales en el área.