



Universidad Nacional de Luján
Departamento de
Ciencias Básicas



DISPOSICION CONSEJO DIRECTIVO DEPARTAMENTAL DE CIENCIAS BÁSICAS DISPCD-CB : 247 /
2025

LUJAN, 10 DE JULIO DE 2025

VISTO: El programa de la asignatura Métodos de análisis de secuencias (18930) para la carrera Especialización y Bioinformación; y

CONSIDERANDO:

Que tomó intervención la Comisión Académica de la Carrera.

Que se ha tratado y aprobado por el Consejo Directivo Departamental de Ciencias Básicas en su Sesión Ordinaria del día 3 de julio de 2025

Por ello,

EL CONSEJO DIRECTIVO DEPARTAMENTAL
DE CIENCIAS BÁSICAS

D I S P O N E :

ARTÍCULO 1°.- APROBAR el programa de la asignatura Métodos de análisis de secuencias (18930) para la carrera Especialización en Bioinformática que como anexo I forma parte de la presente Disposición.-

ARTICULO 2°.- ESTABLECER que el mismo tendrá vigencia para los años 2025-2027.-

ARTÍCULO 3°.- Regístrese, comuníquese, cumplido, archívese.-

Lic. Ariel H. REAL - Secretario Académico - Departamento de Ciencias Básicas

Lic. Emma L. FERRERO - Directora Decana - Departamento de Ciencias Básicas

PROGRAMA OFICIAL

DENOMINACIÓN DE LA ACTIVIDAD: 18930 Métodos de análisis de secuencias

TIPO DE ACTIVIDAD ACADÉMICA: Curso.

CARRERA/S: Especialización en Bioinformática

PLAN DE ESTUDIOS: Plan de Estudios: 76.01 (Resolución HCS N° 711/22)

DOCENTE/S RESPONSABLE/S: Agustin Arce

EQUIPO DOCENTE: Alberto Penas Steinhardt, Agustin Arce

ACTIVIDADES CORRELATIVAS PRECEDENTES:

PARA CURSAR: TRAMO NIVELATORIO

PARA APROBAR: TRAMO NIVELATORIO

MODALIDAD DE DICTADO:

PRESENCIAL: X

CARGA HORARIA TOTAL: 36 horas

DISTRIBUCIÓN INTERNA DE LA CARGA HORARIA:

TEÓRICO 18 horas
PRÁCTICO 18 horas

PERÍODO DE VIGENCIA DEL PRESENTE PROGRAMA: 2025-2027.

CONTENIDOS MÍNIMOS O DESCRIPTORES

El curso tiene como objetivo conocer las principales herramientas para el análisis de secuencias biológicas. Para ello se abordarán los siguientes tópicos:

- Uso de las principales bases de datos para la obtención de secuencias
- Alineamiento de pares de secuencias
- Búsqueda en Bases de Datos de secuencias con Blast
- Alineamiento de múltiples secuencias
- Identificación de motivos de secuencia y detección de variaciones
- Procesamiento de información genómica (genomas, transcriptomas y metagenomas)

PROGRAMA OFICIAL

- Genómica funcional
- Estudio y comparación de estructuras proteicas

FUNDAMENTACIÓN, OBJETIVOS, COMPETENCIAS

OBJETIVO GENERAL:

El objetivo principal de este curso es que los inscriptos desarrollen las habilidades y conocimientos para el empleo de herramientas informáticas de análisis de secuencias biológicas; comprendiendo los conceptos y modelos matemáticos en los que se basan, las características de los algoritmos y los programas informáticos que los implementan.

OBJETIVOS ESPECÍFICOS:

Utilizar las principales bases de datos de secuencias y estructuras biológicas, y extraer información de ellas.

Profundizar los conocimientos conceptuales y algorítmicos de las herramientas de alineación de pares y múltiples secuencias.

Aplicar herramientas bioinformáticas de análisis de secuencias para la detección de variantes genómicas, y de motivos y dominios proteicos.

Integrar los resultados de análisis de secuencias para resolver problemas biológicos: comparación de genomas, metagenómica y transcriptómica.

Implementar métodos para el análisis de datos masivos en genómica funcional.

Visualizar y comparar estructuras de proteínas, así como evaluar la calidad de los resultados del modelado de estructuras.

CONTENIDOS

UNIDAD 1: Principales base de datos de secuencias

Tipos de Bases de Datos: de secuencias, estructurales, funcionales, especializadas en un organismo. Bases de Datos primarias y secundarias. Niveles de curado. Tipo de información almacenada. Bases de datos de nucleótidos (NCBI GenBank, NCBI RefSeq, NCBI Nucleotide, etc). Bases de Datos de secuencias de proteínas (Uniprot, NCBI RefSeq). Formatos de archivos de secuencias: FASTA y GenBank. Recuperación de datos de múltiples Bases de Datos interconectadas: el uso de Entrez de NCBI.

UNIDAD 2: Alineamiento de secuencias

PROGRAMA OFICIAL

Matrices de sustitución (PAM, Blosum). Método de alineamiento gráfico: dot-plot. Programación dinámica y métodos de alineamientos optimal de a pares: global (Needleman-Wunsch) y local (Smith-Waterman). Formatos de entrada y de salida de un alineamiento. Análisis estadístico de las puntuaciones de los alineamientos. Programas para edición de alineamientos de secuencias (Jalview o similares).

UNIDAD 3: Búsqueda en Bases de Datos de secuencias con Blast

El algoritmo de Blast. Parámetros básicos de la búsqueda. Interpretación de los resultados. El E-value: significado estadístico y criterios para establecer homología. Usos de diferentes tipos de blast: tblastn, blastx y tblastx. Parámetros avanzados de búsqueda. Bases de datos disponibles en NCBI.

UNIDAD 4: Alineamiento múltiple de secuencias

Alineamiento múltiple de secuencias de nucleótidos o de proteínas. Función de Scoring. Métodos ClustalW (y sus sucesores), T-Coffee, MUSCLE, MAFFT, etc. Formatos de entrada y salida. Identificación de Motivos y Dominios. Representación de alineamientos múltiples mediante Sequence Logos. Uso de cadenas ocultas de Markov (HMM). Bases de Datos de Perfiles. Búsquedas de alineamientos tipo: secuencia-perfil (PFAM) y perfil-perfil (HMM-HMM).

UNIDAD 5: Identificación de motivos de secuencia y detección de variaciones

Búsqueda y descubrimiento de motivos (MEME). Detección de SNPs y Variación Estructural (SVs, deleciones, duplicaciones, inversiones, translocaciones, SNP/indel).

UNIDAD 6: Genomas, transcriptomas y metagenomas

Tecnologías de secuenciación de próxima generación. Bases de datos de secuenciación (GEO). Formato de las lecturas (Fastq). Evaluación de calidad de lecturas (FastQC). Filtrado y recorte de lectura (Trimmomatic, fastp). Mapeo de lecturas (STAR, HISAT2). Servidor Galaxy. Aplicaciones: ensamblado de genomas y transcriptomas, transcriptómica y metagenómica.

UNIDAD 7: Genómica funcional

Herramientas para el análisis de datos masivos en genómica funcional y proteómica. Ontología génica (GO), dominios y estructura (DAG). Códigos de evidencia. KEGG. Enriquecimiento funcional. GSEA.

UNIDAD 8: Estructuras de proteínas

PROGRAMA OFICIAL

Análisis y clasificación estructural de proteínas: Introducción a las principales metodologías de predicción de estructuras proteicas. Alphafold y la evaluación de los modelos (pLDDT, PAE, etc). Alineamiento estructural.

METODOLOGÍA

Se realizarán prácticas en laboratorios de computación de acuerdo a los requerimientos de cada una, de manera de ejercitar e integrar los contenidos abordados durante las clases teóricas. Cada unidad tiene asociada actividades prácticas de aplicación de las herramientas, que serán realizadas durante los Trabajo Práctico. Estos contendrán problemas a resolver con el fin de desarrollar habilidades y criterios para aplicar las herramientas e interpretar los resultados. En las horas destinadas a las actividades prácticas se resolverán algunos ejercicios testigo y el resto deberán ser resueltos por el estudiante en forma grupal.

La resolución de los trabajos se realizará a través del uso servidores/clusters de libre acceso o de aquellos instalados en la universidad, en caso de ser requerido. Estas herramientas permitirán el trabajo colaborativo remoto entre todos los estudiantes del curso. Tiene un doble objetivo: comprender y aplicar los contenidos de cada unidad, al mismo tiempo que se aprende a trabajar en forma remota con una máquina virtual, la cual contendrá todo el software necesario para desarrollar cada unidad del programa de estudios.

TRABAJOS PRÁCTICOS

TP 1: Base de Datos de ácidos nucleicos y proteínas. Uso del portal de NCBI.

TP 2: Alineamiento de a pares de secuencias de nucleótidos y de proteínas.

TP 3: Búsqueda en Bases de Datos con BLAST

TP 4: Alineamiento múltiple de secuencias de nucleótidos y de proteínas. Perfiles HMM y búsqueda de dominios. Identificación de motivos de secuencia y detección de variaciones

TP 5: Análisis de datos genómicos: genomas, transcriptomas y metagenomas. Genómica funcional.

TP 6: Visualización de estructuras proteicas, evaluación de modelos y alineamiento estructural.

PROGRAMA OFICIAL

REQUISITOS DE APROBACIÓN Y CRITERIOS DE CALIFICACIÓN

A lo largo de todo el curso se deberán entregar y aprobar, en las fechas previstas, un trabajo práctico por cada unidad del programa del curso, los cuales se utilizarán para realizar una evaluación formativa del estudiante de carácter continuo.

A su vez, al final del curso deberán aprobar un examen integrador de carácter individual.

BIBLIOGRAFÍA

Introduction to Computational Genomics: A Case Studies Approach. Nello Cristianini and Matthew W. Hahn. Cambridge University Press (2006). (temas/units 1-4)

Essential Bioinformatics. Jin Xiong. Cambridge University Press (2006).

Bioinformatics and functional genomics. Jonathan Pevsner. Wiley-Blackwell. Second edition (2009).

Bioinformatics --From Genomes to Therapies. Edited by Thomas Lengauer. Wiley-VCH (2007).

Bioinformatics (Sequence and Genome analysis). David W. Mount. Cold Spring Harbor Laboratory Press (2001)

Biological sequence analysis. Richard Durbin, Sean Eddy, Anders Krogh and Graeme Mitchison. Cambridge University Press (2007, 10th reprint).

Perera J., Tormo A., García J.L. (2002) Ingeniería Genética, Vol I: Preparación, Análisis, Manipulación y Clonaje de DNA. Editorial Síntesis.

Reece R.J. (2004) Analysis of Genes and Genomes. John Wiley & Sons Ltd.

Hoja de firmas