

BIOINFORMÁTICA

Carrera: Microbiología

Plan de estudios: 2023

Área de Formación: Aplicada

Año: Segundo

Régimen de Cursada: Cuatrimestral

Carácter: Obligatoria

Carga horaria total: 70 horas

Carga horaria teórica: 25 horas

Carga horaria práctica: 45 horas

OBJETIVO GENERAL DEL CURSO

Aportar contenidos sobre las principales metodologías de análisis bioinformático aplicado al ámbito de la microbiología desde un enfoque teórico hasta su implementación en la práctica. Sentar las bases para los cursos de Herramientas de Ingeniería Genética, Ingeniería Genética aplicada, Análisis Genéticos aplicados a la Microbiología y Bioprocesos.

OBJETIVOS DE APRENDIZAJE

Se espera que al final del curso el estudiante:

- Comprenda los algoritmos y metodologías de alineamiento de secuencias que se utilizan para el análisis funcional y estructural del ADN, ARN y proteínas mediante la aplicación y discusión de ejercicios prácticos.
- Adquiera y apropie el manejo de las principales bases de datos primarias y secundarias de ácidos nucleicos y proteínas a través de la explicación teórica, videos tutoriales y el desarrollo de ejercicios prácticos.
- Desarrolle habilidades metodológicas y lógicas sobre plataformas bioinformáticas de código abierto tales como R/Bioconductor que le posibiliten construir, apropiarse y aplicar los conocimientos adquiridos en el análisis e interpretación de datos biológicos mediante el uso de dichos programas.
- Utilice los principales programas bioinformáticos disponibles y aplicaciones webs en la exploración, obtención y análisis de datos de genomas, metagenomas y transcriptomas celulares y microbianos.
- Desarrolle un pensamiento crítico sobre los alcances y limitaciones de las metodologías de análisis bioinformático en el contexto de estudios de microbiología clínica mediante la discusión de casos de estudio.

CONTENIDOS MÍNIMOS

Introducción a la bioinformática y bases de datos. Secuenciación y análisis de secuencias biológicas. Recursos bioinformáticos. Genómica, metagenómica y microbiomas. Genómica funcional y análisis de transcriptomas. Diseño de cebadores para la detección de microorganismo mediante PCR.

PROGRAMA ANALÍTICO

UNIDAD N° I: INTRODUCCIÓN A LA BIOINFORMÁTICA Y BASES DE DATOS

Aplicaciones de la bioinformática y uso de bases de datos biológicas

Definición de bioinformática, objetivos, reseña histórica, áreas de estudio. El dogma central de la biología molecular: flujo de información genética, conceptos secuencia > estructura > función y diferencias entre dato e información. Aplicaciones de la bioinformática en el estudio de variantes genéticas asociadas a enfermedades, la medicina de precisión y la microbiología. Conceptos sobre bases de datos informáticas: tipos, estructura y organización. Sistemas de gestión de bases de datos e infraestructura de las bases de datos: servidores. Bases de datos biológicas ácidos nucleicos / proteínas primarias (GenBank, UniProt, Protein Data Bank), secundarias (RefSeq, InterPro, Pfam) y compuestas (NCBI, EMBL, DDBJ). Acceso, búsquedas, obtención e interpretación de datos disponibles mediante NCBI Entrez y UniProt en el contexto de la estructura génica de procariontes y eucariotes. Estructura de datos en GenBank y secuencias biológicas en formato FASTA. Bases de datos curadas de genes (Entrez Gene), genomas (NCBI Genome, UCSC Genome Browser) y factores de transcripción (JASPAR).

UNIDAD N° II: SECUENCIACIÓN Y ANÁLISIS DE SECUENCIAS BIOLÓGICAS

Análisis pareado y múltiple de secuencias de ADN y proteínas

Importancia del estudio de secuencias biológicas en la predicción de su estructura, función y relaciones evolutivas. Conceptos de similitud, homología, identidad de secuencias, secuencias ortólogas y parálogas. Metodologías de secuenciación de péptidos (Edman) y ácidos nucleicos (Sanger). Conceptos generales sobre el alineamiento pareados vs. múltiples y globales vs. locales. Metodología gráfica y cualitativa para la comparación de dos secuencias: matrices de puntos (Dot-plot) y su aplicación en el reconocimiento de patrones. Metodologías cuantitativas para el alineamiento de secuencias: enfoques computacionales exhaustivos vs. heurísticos. Elementos característicos de un alineamiento: match, mismatch y gaps. Matrices de identidad y de sustitución (PAM y BLOSUM) y sistema de puntuación de alineamientos. Fundamentos de los algoritmos de Needleman-Wunsch y Smith-Waterman para alineamientos pareados de secuencias. Ventajas y desventajas de los algoritmos de programación dinámica. Algoritmos de alineamiento para la búsqueda en bases de datos: Basic Local Alignment Search Tool (BLAST). Evaluación de la significancia del alineamiento (Valor E). Uso de NCBI-BLAST y sus diferentes variantes. Conceptos sobre alineamiento múltiple de secuencias (AMS). Algoritmos para el alineamiento múltiple de secuencias: métodos progresivos (ClustalW), no progresivos (MUSCLE) y combinados (T-Coffee).

Secuencias consenso, dominios protéicos y visualización de AMS. Conceptos sobre análisis filogenético: anatomía de un árbol filogenético y construcción de un árbol filogenético a partir de AMS. Aplicaciones del análisis filogenético molecular en microbiología.

UNIDAD N° III: RECURSOS BIOINFORMÁTICOS I

Introducción a Linux/Bash y R/Bioconductor en el análisis bioinformático

Conceptos de infraestructura computacional: software + hardware. Introducción al sistema operativo GNU/Linux: estructura de directorios, localización de archivos, terminal de comando y el intérprete de líneas de comandos Bash. Uso de comandos de utilidad para la manipulación de archivos en bioinformática. Entorno de implementación y desarrollo en bioinformática: R y RStudio. Conceptos básicos sobre el lenguaje de programación estadístico R: función + objeto y estructura de datos en R. Ejecución de códigos en R. Introducción al proyecto R/Bioconductor: acceso programático a datos curados y paquetes con diversas funcionalidades para el análisis de datos en biología molecular y genómica. Implementación de flujos de trabajo bioinformáticos para el alineamiento de secuencias biológicas utilizando paquetes de R/Bioconductor.

UNIDAD N° IV: GENÓMICA, METAGENÓMICA Y MICROBIOMAS

Análisis de genomas y metagenomas en la caracterización de microbiomas

Organización del genoma, tamaño, número de genes y complejidad biológica. El genoma de los procariontes, la diversidad de los genomas virales y bases de datos para la extracción de sus genomas de referencia. Organización funcional y estructural del genoma humano: acervo genético, variabilidad genética (polimorfismos, variantes raras, mutaciones) y sus consecuencias fenotípicas. Conceptos de genómica y metagenómica. Fundamentos y aplicaciones de las plataformas tecnológicas basadas en secuenciadores de próxima generación (NGS): secuenciación de novo vs. resecuenciación de genomas (genoma completo, exomas o paneles de genes) y metagenomas (secuenciación completa de metagenomas vs. Secuenciación del ADN de 16S/18S). Aplicaciones de la genómica en salud humana y en la microbiología para el diagnóstico, tratamiento y vigilancia de agentes infecciosos. Taxonomía microbiana genómica y conceptos sobre microbiota, microbioma, índices de diversidad, disbiosis y metaboloma. Aplicaciones del estudio de comunidades microbianas en salud humana (Proyecto Microbioma Humano), salud animal, metagenomas de suelos y en la biorremediación. Análisis de datos de secuenciadores NGS: formato de datos FASTQ y sistema de calidad de secuenciación Phred. Control de calidad de datos genómicos: uso de los programas FASTQC y Rfastp. Algoritmos de alineamiento de secuencias cortas y largas (ADN y ARN). Análisis e interpretación de datos genómicos: formatos SAM/BAM y VCF. Secuenciación y ensamblaje de un genoma microbiano y análisis de metagenomas para la caracterización estructural y funcional de microbiomas. Metagenoma de referencia.

UNIDAD N° V: GENÓMICA FUNCIONAL Y ANÁLISIS DE TRANSCRIPTOMAS

Análisis de bioprocesos celulares en enfermedades infecciosas

Principios y fundamentos de la bioinformática aplicada a la genómica funcional: transcriptomas, proteomas y metabolomas. Análisis de transcriptomas para la caracterización de bioprocesos celulares: vías de señalización y vías metabólicas. Aplicaciones de la genómica funcional en salud humana: biomarcadores con valor pronóstico y/o predictivo. Secuenciación de transcriptomas / meta-transcriptomas (RNA-seq). Flujos de trabajo bioinformático para el análisis de transcriptomas: métodos no supervisados y supervisados. Conceptos de significancia estadística y significancia biológica. Anotación funcional de genes (Gene Ontology y KEGG) y análisis de agrupamiento jerárquico: fundamentos, métricas de similitud. Identificación de genes diferencialmente expresados y análisis de enriquecimiento funcional para la identificación de bioprocesos desregulados en sistemas huésped-patógeno. Flujos de trabajo para el análisis de transcriptomas en R/Bioconductor. Visualización de datos: volcano plots y mapas de calor. Bases de datos de genómica funcional (Gene Expression Omnibus, GEO-NCBI): acceso, búsquedas, obtención e interpretación de resultados mediante el uso de GEO2R y ShinyGO.

UNIDAD N° VI: RECURSOS BIOINFORMÁTICOS II

Diseño de cebadores para la detección de microorganismos mediante PCR

Conceptos básicos sobre la reacción en cadena de la polimerasa (PCR): componentes y sus aplicaciones en microbiología. Características y criterios para el diseño de cebadores: temperatura de fusión (T_m), temperatura de templado (T_a), estructuras secundarias. Utilización de los programas Primer3Plus y Primer-Blast para el diseño de cebadores. Simulación de la PCR *in-silico* para la evaluación de los cebadores con los programas *in-silico* PCR amplification y UCSC *in-silico* PCR. Diseño de cebadores de una PCR multiplex para la detección de agentes infecciosos y un gen de referencia. Análisis y visualización de cromatogramas de secuencias Sanger (archivos .ab1) obtenidas a partir de productos de PCR con R/Bioconductor.

METODOLOGÍA DE ENSEÑANZA

La metodología adoptada para el dictado de las clases es teórico/práctica. Los programas necesarios para el desarrollo de las actividades son provistos en computadoras personales disponibles en el Aula de Informática de la FCV-UNLP. El docente a cargo del curso se ocupa del dictado de aquellos temas con contenidos teóricos que describen y profundizan conceptos relevantes en la especialidad. Se procede a describir conceptos, características y el uso de herramientas informáticas de aplicación en el ámbito de la microbiología. A cada clase teórica se corresponde una clase práctica que consiste en un conjunto de problemas y ejercicios computacionales que permiten afianzar los contenidos teóricos que fueron previamente desarrollados. Las actividades prácticas están completamente integradas en el marco conceptual de la teoría por lo que generará un ámbito de reflexión y discusión de los temas presentados facilitando la integración entre conceptos aprendidos.

Inicialmente, se buscará familiarizar al alumno con los conceptos básicos vinculados a la biología molecular e informática que le permitan hacer frente a una pregunta compleja o problema biológico. Luego, se introduce al estudiante en las aplicaciones informáticas de uso frecuente en el ámbito de la microbiología en particular, comenzando por el manejo de las principales bases de datos de secuencias biológicas y los algoritmos de alineamiento de secuencias. Seguidamente, se realiza una introducción al uso de lenguaje Linux/Bash y R/Bioconductor, para la manipulación de archivos y la ejecución de programas bioinformáticos. A continuación, se aborda el análisis de genomas, metagenomas y transcriptomas mediante técnicas de secuenciación, con especial énfasis en la caracterización de microbiomas y la identificación de bioprocesos celulares. Al finalizar la cursada se realiza la integración de recursos bioinformáticos en el diseño de cebadores para la detección de agentes infecciosos.

DESCRIPCIÓN DE LAS ACTIVIDADES TEÓRICAS Y PRÁCTICAS

■ **Actividades teóricas:** Se desarrollarán clases teóricas donde se introducirá cada aspecto disciplinar de la bioinformática y se impartirá el marco conceptual y los fundamentos de biología computacional necesario para el desarrollo de las actividades prácticas. Como complemento a las actividades teóricas se dispondrán contenidos audiovisuales en el sistema de Aula Virtual de Bioinformática para que el alumnado profundice en contenidos específicos.

■ **Actividades prácticas:** Se desarrollarán en el aula de informática, donde los alumnos tendrán acceso a una infraestructura computacional y el software preinstalado (RStudio -R/Bioconductor) para llevar adelante las actividades prácticas propuestas. Los alumnos utilizarán guías de actividades prácticas que permitirán implementar de manera ordenada y exitosa el desarrollo de las labores propuestas y coordinadas por los docentes a cargo.

METODOLOGÍA DE EVALUACIÓN

De acuerdo al Reglamento sobre el desarrollo de los cursos de la Facultad, para acceder a la evaluación parcial los estudiantes deberán asistir al 75% de las actividades planificadas. Los alumnos que hayan estado presentes en más del 60% pero no alcancen el 75%, deberán recuperar mediante la resolución de la actividad práctica correspondiente.

La evaluación consiste en un único parcial con sus 3 instancias. La modalidad de evaluación es escrita con soporte informático donde el alumno resuelve problemas prácticos en el área de la bioinformática aplicado a la microbiología.

La calificación de las evaluaciones se realizará por el sistema de puntaje de 0 a 10 puntos, aprobándose con 4 (cuatro) puntos (art. 9 del Reglamento). Para promocionar el curso la nota debe ser de 7 (siete) puntos como mínimo, obtenidos en la última instancia de aprobación (art. 10a).

Los estudiantes cuya nota sea entre 4 y 6 puntos deberán rendir una EFI (art. 10b).

La modalidad de evaluación de la EFI será una parte escrita similar al parcial y otra oral.

BIBLIOGRAFÍA

- **Bioinformatics in Tropical Disease Research: A Practical and Case-Study Approach** [Internet]. Gruber A, Durham AM, Huynh C, et al., editors. Bethesda (MD): National Center for Biotechnology Information (US); 2008. <https://www.ncbi.nlm.nih.gov/books/NBK6818/?term=Bioinformatics>
- **Bioinformatics**. Editor: Nakaya Helder I., PhD. Brisbane (AU): Exon Publications; 2021 Mar 20. ISBN-13: 978-0-6450017-1-6. <https://www.ncbi.nlm.nih.gov/books/NBK569562/>
- Canzoneri R, Lacunza E, Abba MC. **Genómica y bioinformática como pilares de la medicina de precisión en oncología**. Medicina (Buenos Aires). 2019 Oct 2;79(6). <https://www.medicinabuenosaires.com/indices-de-2010-a-2019/volumen-79-ano-2019-no-6-1-indice/genomica/>
- **Computational Biology** [Internet]. Husi H, editor. Brisbane (AU): Codon Publications; 2019 Nov 21. <https://www.ncbi.nlm.nih.gov/books/NBK550339/?term=Bioinformatics>
- Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: **NCBI gene expression and hybridization array data repository**. Nucleic acids research. 2002 Jan 1;30(1):207-10. <https://academic.oup.com/nar/article/30/1/207/1332640>
- Ge SX, Jung D, Yao R. **ShinyGO: a graphical gene-set enrichment tool for animals and plants**. Bioinformatics. 2020 Apr 15;36(8):2628-9. <https://academic.oup.com/bioinformatics/article/36/8/2628/5688742>
- Guía de actividades elaboradas por los docentes del curso. Disponibles en la plataforma Moodle.
- Hernández M, Quijada NM, Rodríguez-Lázaro D, Eiros JM. **Aplicación de la secuenciación masiva y la bioinformática al diagnóstico microbiológico clínico**. Revista argentina de microbiología. 2020 Apr 1;52(2):150-61. http://www.scielo.org.ar/scielo.php?script=sci_arttext&pid=S0325-75412020000200101
- Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R. **Orchestrating high-throughput genomic analysis with Bioconductor**. Nature Methods 2015. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4509590/>
- Méndez-Álvarez S, Pérez-Rotha E. **La PCR múltiple en microbiología clínica**. Enfermedades Infecciosas y Microbiología Clínica. 2004 Jan 1;22(3):183-92. <https://www.elsevier.es/es-revista-enfermedades-infecciosas-microbiologia-clinica-28-articulo-la-pcr-multiple-microbiologia-clinica-13058027>
- Morelli, I.S., Coppotelli, B.M., Madueño, L. and Del Panno, M.T., 2015. **La biorremediación en la era post-genómica**. <https://ri.conicet.gov.ar/handle/11336/16076>
- Sepulveda JL. **Using R and bioconductor in clinical genomics and transcriptomics**. The Journal of Molecular Diagnostics. 2020 Jan 1;22(1):3-20. [https://www.jmdjournal.org/article/S1525-1578\(19\)30397-6/fulltext](https://www.jmdjournal.org/article/S1525-1578(19)30397-6/fulltext)