

“Curso de Filogenia Molecular y Modelado de Proteínas”

Lugar: Departamento de Ciencias Biológicas, CENUR Litoral Norte, Salto

Fecha estimada del curso: 06 a 10 de octubre de 2024

Responsables y docentes del Curso: German Matias Traglia (CENUR), Samuel Miño (INTA Cerro Azul - Argentina), Nelida Rodríguez Osorio (CENUR), Pablo Dans (CENUR)

Público Objetivo:

Estudiantes de grado avanzados de las carreras de bioquímica, biología, biotecnología, medicina, agronomía, veterinaria y afines. Estudiantes de posgrado, Maestría o Doctorado.

Resumen (100 palabras)

El curso de Filogenia Molecular y Modelado de Proteínas combina bioinformática de vanguardia con prácticas para capacitar a estudiantes de grado avanzados y posgrado. A lo largo de cinco días, los participantes aprenderán a depurar secuencias, construir alineamientos múltiples rigurosos y seleccionar modelos evolutivos para inferir árboles filogenéticos robustos mediante métodos de distancia, máxima verosimilitud y bayesianos. Además, explorarán la predicción y refinamiento de estructuras proteicas 3D, visualización y acoplamiento molecular, orientando sus análisis a problemas reales en salud, biotecnología y conservación. El curso enfatiza la planificación de flujos de trabajo reproducibles y la evaluación estadística de resultados exitosos.

Introducción:

La filogenia molecular es una herramienta indispensable para reconstruir las relaciones evolutivas entre organismos a nivel de genes y proteínas, proporcionando un marco sólido para interpretar datos moleculares en contextos históricos y funcionales. En los últimos años se han desarrollado herramientas computacionales que permiten obtener secuencias de ADN o de proteínas fácilmente. Para entender las relaciones entre este tipo de datos (secuencias de ADN o proteínas), es fundamental el dominio de estas herramientas es necesario y fundamental para los profesionales y los estudiantes de las ciencias biológicas y de la salud.

La Filogenia Molecular permite el estudio del origen, evolución y dispersión de patógenos, posee un rol central en estudios de epidemiología, posibilitando el seguimiento de patógenos de interés para desarrollar una prevención mediante la selección de vacunas o terapias adecuadas.

Por otro lado, el Modelado Molecular permite la identificación de regiones conservadas o variables en biomoléculas, así como el estudio de proteínas, sus funciones biológicas, la

predicción de roles basados en homología y el desarrollo de hipótesis rigurosas sobre adaptaciones moleculares, procesos de coevolución y selección natural.

Estas disciplinas se desarrollan en un entorno informático (Bioinformática), desarrollando además análisis estadísticos como medidas de confiabilidad, reforzando las competencias analíticas y técnicas de los estudiantes con capacidades de vanguardia del conocimiento científico. Este curso brinda las herramientas necesarias para afrontar preguntas complejas y aportar de manera significativa al progreso de la ciencia, con aplicaciones de gran relevancia en salud, biotecnología, conservación de la biodiversidad, desarrollo de fármacos, evolución, etc.

Objetivo del aprendizaje:

La realización de este curso brindará al estudiante la capacidad de analizar secuencias de ADN o proteínas, basado en la filogenia molecular y el análisis de la estructura tridimensional de proteínas.

Contenidos Temáticos:

1. Introducción a la Bioinformática.

Teoría 1 (5 horas)

Introducción a la Bioinformática y Biología Computacional. Concepto de trash-in / trash-out. Conceptos de Evolución Molecular. Importancia de las mutaciones en los análisis. Importancia de realizar un buen armado de la matriz de datos. Flujo de trabajo (pipeline) en bioinformática y/o biología computacional. Cómo evaluar la calidad de los datos. Conceptos: “Los programas solo responden las preguntas que les hacemos”. Criterios para planificar un trabajo ordenado basado en análisis informáticos.

2. Construcción de la matriz.

Teoría 2 (5 horas)

Programas para visualizar un cromatograma: Bioedit, MEGA. Criterios para manipular una secuencia de ADN. Curado: desde el cromatograma hasta el Genbank. Programas para manipular secuencias: Bioedit, Aliview. BLASTn: Cómo realizar una búsqueda en una base de datos especializada, cómo descargar las secuencias de forma personalizada y cómo armar una matriz de datos diversa. Grupo externo (outgroup): cómo seleccionarlo. Conceptos: “un outgroup debe ser diferente a las cepas a analizar, pero no tanto”. Métodos de alineamientos de secuencias de ADN, fundamentos, ventajas y desventajas. Parámetros de un alineamiento, importancia de estos.

Práctica 1 (10 horas)

Instalación de los programas: NotePad++, MEGA, Bioedit, Aliview. Manejo y edición de secuencias de ADN: Trabajar un cromatograma para obtener una secuencia lista para subirla al Genbank. Cargado de la secuencia al Genbank. Búsquedas en BLASTn. Selección y descarga de secuencias de la base de datos. Armado de las matrices de análisis. Alineamiento múltiple inicial y ajuste de la matriz. Búsquedas de secuencias para armar el grupo externo.

3. Alineamiento de secuencias de ADN.

Teoría 3 (10 horas)

Fundamentos teóricos de alineamiento de secuencias. Distancia “p” y distancia genética. Alineamiento de pares y alineamientos múltiples. Alineamientos como hipótesis de homología, cómo seleccionar el alineamiento de trabajo. Programas para realizar alineamientos múltiples de secuencias, Bioedit, Aliview, MEGA, MAFFT. Control de calidad y edición de los alineamientos. Controles de calidad de la matriz de datos: frecuencias de bases y de sustituciones, efecto de la recombinación, contenido de información filogenética y contenido de información temporal.

Práctica 2 (10 horas)

Instalación de programas: Split Tree, IQ-Tree, RDP4. Alineamiento múltiple de secuencias on-line y off-line. Criterios de selección de programas, ventajas y desventajas. Evaluación y edición manual de los alineamientos. Evaluación de la presencia de recombinación “Phi-test”. Introducción al uso de RDP4.

4. Métodos de inferencia filogenética.

Teoría 4 (10 horas)

Conceptos de evolución molecular. Modelos evolutivos para secuencias de ADN. Introducción a los métodos de inferencia filogenética: Distancias, Parsimonia, Verosimilitud y Análisis Bayesianos. Búsquedas heurísticas, fundamentos. Fundamentos, ventajas y desventajas de los métodos UPGMA, Neighbor-Joining, Máxima Parsimonia, Maximum Likelihood. Programas para realizar cada tipo de análisis, aplicaciones y criterios para llevarlos adelante. Criterios de soporte estadístico: Bootstrap, Ultrafast bootstrap y Jackknife, fundamentos teóricos.

Práctica 3 (10 horas)

Determinación de modelo de evolución (MEGA, IQ-Tree). Construcción de matrices de distancia y árboles de Distancias con MEGA. Construcción de árboles filogenéticos de Máxima Verosimilitud: IQ-Tree. Visualización y edición de árboles con FigTree.

5. Nociones de modelado de proteínas.

Teoría 5 (10 horas)

Estructura de proteínas, primaria, secundaria, terciaria y cuaternaria. Fundamentos teóricos del modelado de proteínas. Bases de datos de estructuras. Modelado de novo y modelado basado en un templado. Modelado online a través del servidor I-Tasser. Visualización de estructuras atómicas, gráficos de cintas y modelos 3D. Alineamiento estructural. Análisis a posteriori (Coulombic Surface Coloring). Mapeo de sitios dentro de una estructura. Refinamiento de estructuras. Introducción al acoplamiento molecular. Fundamentos del acoplamiento molecular (docking). Cribado virtual a través de acoplamiento molecular. Mapas de calor y selección de compuestos.

Práctica 4 (10 horas)

Instalación y usos de los programas básicos para realizar el modelado de proteínas: Modeller, ITasser, UCSF-Chimera, AutoDock Tools y AutoDock VINA. Manipulación de modelos 3D, construcción de figuras, marcado de sitios de interés. Refinado de estructuras. Docking contra la base de datos ZINC. Construcción de mapas de calor.

Cronograma

Lunes	Martes	Miércoles	Jueves	Viernes
08:00 – 08:45. Acreditación.	07:00 – 07:45. Desayuno	07:00 – 07:45. Desayuno	07:00 – 07:45. Desayuno	07:00 – 07:45. Desayuno
08:45 – 09:00. Presentación del curso.	08:00 – 10:00. Práctico 1.	08:00 – 09:30. Teoría 3.	08:00 – 11:45. Práctico 3.	08:00 – 11:30. Práctico 5.
09:05 – 11:00. Introducción a la bioinformática. Teoría 1	10:15 – 11:45. Teoría 3.	9:45 – 12:00 Teoría 3		
12:30 – 13:30. Almuerzo	12:00 – 13:00. Almuerzo	12:00 – 13:00. Almuerzo	12:00 – 13:00. Almuerzo	12:00 – 13:00. Almuerzo
14:00 – 18:00. Teoría 2.	14:00 – 17:00. Práctico 2.	Integración / Paseo por instalaciones del CENUR *	13:30 – 15:00. Teoría 4.	Libre / Consulta
			15:10 – 18:00. Práctico 4.	Libre / Consulta

*=Integración/Paseo por instalaciones del CENUR: Tiene el objetivo de brindar un entorno de discusión sobre la aplicabilidad de las herramientas del curso en el área laboral/de estudio de cada uno de los participantes. Además, el paseo por las instalaciones del

CENUR tiene el objetivo de divulgar las capacidades regionales y temas de investigación en CENUR.

Bibliografía

1. Philippe Lemey, Marco Salemi and Anne-Mieke Vandamme. The Phylogenetic Handbook A Practical Approach to Phylogenetic Analysis and Hypothesis Testing. Cambridge University Press 2009, ISBN-13 978-0-511-71963-9. Cambridge University Press, The Edinburgh Building, Cambridge CB2 8RU, UK.
2. Xohua Xia. Bioinformatics and the Cell: Modern Computational Approaches in Genomics, Proteomics and Transcriptomics. Springer Cham, published on 05 July 2018 eBook ISBN978-3-319-90684-3. <https://doi.org/10.1007/978-3-319-90684-3>. Springer Int Publishing Ag, Gewerbestrasse 11, Cham, Ch 6330, Switzerland.
3. Gonzalez, F.N., Raticelli, F., Ferrufino, C. et al. Detection and characterization of Deformed Wing Virus (DWV) in apiaries with stationary and migratory management in the province of Entre Ríos, Argentina. *Sci Rep* 14, 16747 (2024). <https://doi.org/10.1038/s41598-024-67264-7>
4. Miño, S.O.; Badaracco, A.; Uriarte, E.L.; Ciarlet, M.; Parreño, V. Evolution of Animal South American RVA Told by the NSP4 Gene E12 Genotype. *Viruses* 2022, 14, 2506. <https://doi.org/10.3390/v14112506>
5. Miño S, Mojsiejczuk L, Guo W, Zhang H, Qi T, Du C, Zhang X, Wang J, Campos R, Wang X. 2019. Equine influenza virus in Asia: phylogeographic pattern and molecular features reveal circulation of an autochthonous lineage. *J Virol* 93:e00116-19. <https://doi.org/10.1128/JVI.00116-19>.
6. Traglia GM, Pasteran F, Moheb S, Akhtar U, Gonzalez S, Maldonado C, Furtado N, Mohamed A, Escalante J, Tuttobene MR, Quillen A, Fontan C, Albornoz E, Corso A, Bonomo RA, Rao GG, Tolmasky ME, Ramirez MS. Insights into *Acinetobacter baumannii* AMA205's Unprecedented Antibiotic Resistance. *Int J Mol Sci.* 2024 Oct 24;25(21):11424. doi: 10.3390/ijms252111424. PMID: 39518977; PMCID: PMC11546953.
7. Traglia GM, Poudel M, Miño S, Canteros BI, Minsavage GV, Sharma A, Goss EM, Jones JB, Gochez A. *Xanthomonas citri* pv. *eucalyptorum* strain 4866-2_S43 (formerly *X. axonopodis* pv. *eucalyptorum*): the causal agent of bacterial leaf blight on eucalypts recovered in Argentina. *Access Microbiol.* 2024 Oct 7;6(10):000827.v3. doi: 10.1099/acmi.0.000827.v3. PMID: 39411562; PMCID: PMC11475742.

8. Traglia GM, Betancor L, Yim L, Iriarte A, Chabalgoity JA. Genotypic and phenotypic analysis of *Salmonella enterica* serovar Derby, looking for clues explaining the impairment of egg isolates to cause human disease. *Front Microbiol.* 2024 Jun 6;15:1357881. doi: 10.3389/fmicb.2024.1357881. PMID: 38903793; PMCID: PMC11186997.