

Herramientas computacionales para abordar problemas biológicos (HC4APB)

Objetivos:

El objetivo general del curso es introducir a los estudiantes de posgrado en una selección de herramientas fundamentales de la Biología Computacional, entendida como la aplicación de métodos computacionales, matemáticos y estadísticos para modelar, analizar e interpretar datos biológicos.

Se trata de un curso de carácter introductorio, orientado a brindar una visión general de enfoques computacionales que permiten abordar problemáticas biológicas actuales. El curso combinará clases teóricas con sesiones prácticas de laboratorio. Las clases teóricas estarán enfocadas en la comprensión conceptual de los distintos métodos, mientras que los laboratorios permitirán aplicar estas herramientas en escenarios representativos, favoreciendo la integración entre teoría y práctica.

Programa:

Semana 1: Presentación curso y lineamientos actividad integradora – *todos los docentes*

En esta primera clase se presentará la estructura general del curso, los objetivos de aprendizaje, la modalidad de evaluación y los criterios de participación. Además, se introducirá la actividad integradora que se desarrollará a lo largo del curso, detallando su propósito, formato y entregables. Participan todos los docentes responsables.

Semana 2: “Control de versiones con Git” – *Ignacio Alcántara*

Este módulo cubre el uso de Git, una herramienta de control de versiones esencial para gestionar cambios en archivos binarios (formato típico de “scripts”). Se explorarán conceptos básicos como commits y seguimiento de cambios, instalación de Git, plataformas de hosting (GitHub, GitLab), y distintas interfaces de usuario. La segunda parte incluirá explorar distintas aplicaciones prácticas como la escritura y revisión de tesis en Git, gestión de repositorios de laboratorio, y manejo de archivos grandes con Git LFS basado en casos reales presentados por los docentes. Se espera que al finalizar el módulo, el estudiante pueda manejar a nivel de usuario el sistema de versión de control de sus propios archivos así como colaborar en proyectos compartidos.

Semana 3: “Manejo de secuencias en línea de comando” – *Martín Rivara, Paola Sosa*

En este módulo se introducirá el manejo de datos de secuenciación de ácidos nucleicos mediante herramientas bioinformáticas en línea de comando. Se abordarán los conceptos fundamentales sobre archivos FASTA, FASTQ, entre otros y se trabajará con herramientas como grep, awk, cut, y sed para explorar, filtrar y transformar secuencias. Asimismo, se presentarán programas específicos para tareas frecuentes en bioinformática y se discutirá cómo integrarlos en flujos de trabajo reproducibles. Al finalizar, los estudiantes estarán capacitados para inspeccionar y manipular archivos de secuencias desde la línea de comando.

Semana 4: “Análisis de secuencias en SeqinR” – *Diego Simón*

Este módulo introduce a los participantes en el uso del paquete SeqinR del lenguaje R, una herramienta versátil para la manipulación y para análisis composicionales de secuencias biológicas en el entorno R. Se abordarán funciones básicas para importar, explorar y procesar secuencias de ADN (o ARN), con énfasis en tareas como el cálculo del contenido GC, uso de dinucleótidos y uso de codones. El objetivo es familiarizarse con la generación de estadísticas descriptivas muy utilizadas en genómica y evolución molecular.

Semana 5: “Minería de datos genómicos con Ensembl” – *Carla Filippi*

En este módulo práctico, los participantes explorarán las funcionalidades del navegador Ensembl y sus

herramientas asociadas (biomart, variant effect predictor, entre otras) para el análisis de datos genómicos. Se abordará el uso de Ensembl para consultar anotaciones funcionales, descargar secuencias génicas y promotoras, identificar estructuras génicas, explorar variantes y sus posibles efectos, y realizar conversiones entre distintos identificadores, todo mediante una interfaz accesible.

Semana 6: “Redes de co-expresión génica con Python” – *Sofía Zeballos, Lucas Inchausti*

En este módulo se introducirá los conceptos de grafos y redes de co-expresión para la predicción de función de genes y exploración de datos. Para esto utilizaremos el lenguaje de programación Python en notebooks de Google Colab y paquetes de funciones ya diseñadas para construir redes. Veremos también cómo interpretar los resultados y analizar los clusters de genes co-expresados.

Semana 7: “Modelado de redes ecológicas en R” – *Esteban Ortiz Grandal, Felipe Maresca*

En este módulo se presentarán herramientas de simulación de redes ecológicas basadas en ecuaciones diferenciales en el lenguaje de programación R. Para ello, se trabajará con modelos Lotka-Volterra para simular dinámicas de competencia y consumo. Durante el módulo se abordarán conceptos de la teoría ecológica asociados a interacciones de competencia y redes tróficas, y se demostrará cómo estas interacciones se representan en lenguaje matemático sencillo. Por último, se ejemplificará cómo programar, simular y visualizar este tipo de dinámicas ecológicas en R. Hacia el final del módulo, los estudiantes podrán explorar diferentes escenarios analizando cómo distintas condiciones iniciales afectan tanto la dinámica como la estructura final de las redes simuladas.

Semana 8: “Simulaciones biomoleculares con el campo de fuerzas SIRAH” – *Martín Soñora*

En este módulo se introducirá a los estudiantes en el uso de simulaciones de dinámica molecular a nivel de resolución coarse-grained como herramienta para estudiar sistemas biomoleculares complejos. Se trabajará específicamente con el campo de fuerzas SIRAH, un modelo desarrollado para representar proteínas, lípidos, ácidos nucleicos y solventes de forma eficiente, preservando aspectos estructurales, fisicoquímicos y funcionales relevantes. A lo largo del taller se abordarán conceptos clave sobre la preparación de sistemas, ejecución de simulaciones y análisis de trayectorias mediante herramientas como GROMACS y VMD. Los estudiantes tendrán la oportunidad de explorar casos prácticos relacionados con la dinámica de proteínas y la organización de membranas, así como de discutir las ventajas y limitaciones del enfoque coarse-grained frente a modelos de resolución atómica.

Semana 9: “Machine learning con SciKitLearn” – *Flavio Pazos*

Se abordará, de forma introductoria los fundamentos del aprendizaje automático y su aplicación en el análisis de datos biológicos. La parte teórica se centrará en los conceptos clave del aprendizaje supervisado y no supervisado, incluyendo tipos de algoritmos, nociones de entrenamiento y validación de modelos, y ejemplos de problemas biológicos abordables con estas herramientas. En la parte práctica, trabajaremos con un conjunto de datos reales utilizando Python y bibliotecas como scikit-learn, para aplicar un flujo básico de análisis: preprocesamiento, entrenamiento de modelos supervisados y visualización de resultados. Se fomentará la interpretación biológica de las salidas del modelo y se introducirá brevemente cómo evaluar su desempeño. La clase está orientada a estudiantes sin experiencia previa en machine learning, con énfasis en la comprensión conceptual y la aplicación práctica guiada.

Semana 10: “Introducción a Nextflow y flujos de trabajo reproducibles” – *Martin Beracochea, Joaquín Pereira*

Este módulo tiene como objetivo introducir a las y los estudiantes en el uso de Nextflow, una herramienta moderna para el diseño y ejecución de flujos de trabajo computacionales de manera eficiente y reproducible. Se abordarán los conceptos fundamentales del paradigma de flujos de trabajo, la escritura de scripts en Nextflow, el manejo de canales y procesos, así como la integración con contenedores (Docker, Singularity) y gestores de paquetes (Conda). También se discutirán buenas prácticas para la reproducibilidad computacional y la trazabilidad de resultados en investigaciones bioinformáticas.

Semana 11: Instancia globalizadora – todos los docentes

En esta clase se ofrecerá un espacio de consulta y retroalimentación para resolver dudas sobre los contenidos abordados durante el curso. Será también un espacio donde los estudiantes podrán presentar una versión inicial de su proyecto final, con el fin de obtener *feedback*. Participan todos los docentes.

Bibliografía:

- Introduction to Machine Learning with Python, Müller A C, Guido S. O'Reilly Media, Inc. ISBN: 978-1-449-36941-5.
- Braun, E., et al. 2019. Best Practices for Foundations in Molecular Simulations [Article v1.0]. Living journal of computational molecular science, 1(1), 5957. <https://doi.org/10.33011/livecoms.1.1.5957>
- Molecular Modelling: Principles and Applications, 2nd Edition. Dr Andrew Leach
- Gotelli, N. J. (2008). A primer of ecology (4th ed.). Sunderland, MA: Sinauer Associates
- Stevens, M. H. H. (2009). A Primer of Ecology with R. Springer, New York
- Bryan J., Happy Git with R. <https://happygitwithr.com/>

OTRA INFORMACIÓN RELEVANTE

Fecha de inicio y finalización:

Inicio: Jueves 31 de julio

Finalización: Jueves 16 de octubre

Institución, Salón y Horario:

Clases virtuales, sincrónicas y con control de asistencia.

Teóricos: Martes de 18 a 20:00 hs

Prácticos: Jueves de 18:00 a 20:30 hs

Dedicación horaria:

Horas presenciales: 54 (22 horas teóricas, 25 horas prácticas, 7 horas de discusión)

Horas de trabajo domiciliaria: 36 (estudio, realización de proyecto final)

Conocimientos previos requeridos:

Conocimientos básicos de estadística y programación.

Metodología de enseñanza:

El curso se dictará en modalidad virtual, sincrónica, con control de asistencia. Se desarrollará en formato de talleres semanales que combinarán exposiciones teóricas con actividades prácticas aplicadas. Cada encuentro estará centrado en una herramienta o conjunto de herramientas específicas, introducidas a partir de casos de uso reales o simulados que reflejan problemáticas biológicas actuales.

Las actividades prácticas se realizarán en computadoras personales, ya sea utilizando software previamente instalado o plataformas en la nube (como GitHub, Google Colab, entre otras), promoviendo el trabajo autónomo guiado, la resolución de problemas concretos y la participación activa durante los encuentros. Como cierre del curso, se propondrá un proyecto final integrador que permitirá a los estudiantes aplicar de manera articulada los conocimientos y habilidades adquiridos a lo largo de los módulos.

Asistencia:

La asistencia será obligatoria al 80% de las clases teóricas y prácticas. Dado que el curso se dictará en modalidad virtual sincrónica, se requerirá la participación con la cámara encendida durante los encuentros, como forma de garantizar la presencia activa y el involucramiento con las actividades propuestas.

Forma de evaluación:

Para obtener la ganancia del curso, los estudiantes deberán cumplir con los requisitos de asistencia (mínimo 80%) y aprobar la evaluación final integradora. La ganancia tendrá una vigencia de un año.

El curso contará con una evaluación continua, mediante entregas prácticas individuales o grupales al finalizar cada módulo. Estas actividades permitirán aplicar los contenidos trabajados y se deberá completar al menos el 70% de ellas para habilitar la evaluación final.

La evaluación final consistirá en la resolución de un problema biológico utilizando una o más de las herramientas presentadas en el curso. Los estudiantes elaborarán un informe escrito y lo defenderán en una instancia oral ante el equipo docente. El problema y los datos a utilizar serán definidos en conjunto entre cada estudiante y el equipo docente.

La calificación final será determinada a partir de la calidad del informe escrito y de su presentación oral, cuya fecha se acordará con cada estudiante de manera flexible.
