

#### Aspectos generales

Título:	Bioinformática, Linux y bases de programación para el análisis de datos biológicos
Programas de posgrado o planes de estudio en donde se ofertará adicionalmente:	
Ninguno	
Área del conocimiento:	Genética, genómica y bioinformática
Semestre:	2026-2
Modalidad:	Curso fundamental
Horario:	Viernes 9-13 hrs
No. sesiones:	16
Horas por sesión:	4.0
Total alumnos PDCB:	6
Total alumnos:	12
Videoconferencia:	No
Lugar donde se imparte:	Unidad de Bioinformática y Manejo de la Información del Instituto de Fisiología Celular.
Informes:	cperalta@ifc.unam.mx

#### Métodos de evaluación

MÉTODO	PORCENTAJE	NOTAS
Examen práctico	20%	El examen práctico, está planeado para que cada estudiante aborde una pregunta de investigación a través de métodos bioinformáticos y sets de datos de su elección, su evaluación constará de la evaluación del reporte escrito de los resultados así como la p
Examen teórico	20%	El examen teórico, será un instrumento de opción múltiple que busque evaluar la comprensión de los conceptos fundamentales del curso.
Proyecto de Python	20%	Los items identificados como proyectos son ejercicios prácticos que buscan poner en práctica los conocimientos adquiridos en los módulos correspondientes, se evaluarán a través del reporte de los resultados de los mismos ( scripts, notebooks de jupyter,
Proyecto de R	20%	Los items identificados como proyectos son ejercicios prácticos que buscan poner en práctica los conocimientos adquiridos en los módulos correspondientes, se evaluarán a través del reporte de los resultados de los mismos ( scripts, notebooks de jupyter,
Proyecto GNU/Linux	20%	Los items identificados como proyectos son ejercicios prácticos que buscan poner en práctica los conocimientos adquiridos en los módulos correspondientes, se evaluarán a través del reporte de los resultados de los mismos ( scripts, notebooks de jupyter,

#### Contribución de este curso/tópico en la formación del alumnado del PDCB:

La contribución de este curso reside en capacitar al estudiante para que utilice la computadora de manera autónoma y eficaz en el contexto de la investigación biomédica. Se prioriza el desarrollo de un entendimiento fundamental sobre el aprendizaje de aplicaciones particulares, sentando las bases del pensamiento algorítmico aplicado a problemas biológicos. Esta formación se complementa con una experiencia directa y guiada en el entorno de cómputo de alto rendimiento del Instituto, habilitando al alumnado para enfrentar los desafíos del análisis de datos masivos en sus propios proyectos de tesis.

#### Profesor (a) responsable

Nombre:	Peralta Alvarez Carlos Alberto
Teléfono:	
Email:	cperalta@ifc.unam.mx

#### Profesores (as) participantes

PARTICIPANTE	ENTIDAD O ADSCRIPCIÓN	SESIONES

<b>PERALTA ÁLVAREZ CARLOS ALBERTO</b> Responsable	Instituto de Fisiología Celular	Examen teórico-práctico final Instalación y gestión de un sistema operativo GNU/Linux -Parte 2 Instalación y gestión de un sistema operativo GNU/LINUX Parte I. Introducción a R para bioinformática (Parte I) Introducción a R para bioinformática (Parte II) Introducción a R para bioinformática (Parte III) Introducción al cómputo científico para las ciencias biológicas Presentación de resultados por parte de los alumnos Proyecto de bioinformática
<b>POOT HERNÁNDEZ AUGUSTO CÉSAR</b> Integrante	Instituto de Fisiología Celular	Introducción a Python para bioinformática (parte II) Introducción a Python para bioinformática (parte III) Bibliotecas para manejo y visualización de datos Numpy, scipy, matplotlib Introducción a Python para bioinformática (parte IV) Bibliotecas especializadas ScikitLearn, ScikitImage, Biopython, Keras Presentación de resultados por parte de los alumnos (Augusto Poot 2h / Carlos Peralta 2h) Presentación de resultados por parte de los alumnos (Augusto Poot 2h / Carlos Peralta 2h)

## Introducción

La investigación en ciencias biológicas modernas genera datos a una escala sin precedentes. Para analizar genomas, estructuras de proteínas o cualquier gran conjunto de datos biológicos, es indispensable tener una sólida base computacional. Sin embargo, esta formación rara vez forma parte de los planes de estudio de pregrado.

Más que un recetario de protocolos específicos (como RNA-Seq o Docking), este curso se enfoca en construir los cimientos. Para ello, se prioriza el desarrollo del pensamiento algorítmico y el dominio de un ecosistema de herramientas transversales: el sistema operativo GNU/Linux, la lógica de la programación y el manejo de repositorios de software y bases de datos. Esta base fundamental permitirá al estudiante no solo operar software bioinformático, sino entender su lógica, adaptarlo a problemas novedosos y crear sus propias soluciones de análisis.

Este curso está pensado para que el alumno aprenda en su mayor parte, a través de la práctica, para ello cada estudiante establecerá y administrará un ambiente de trabajo personal para el desarrollo de las prácticas y evaluaciones. Adicionalmente tendrá acceso limitado a la infraestructura de cómputo científico del Instituto de Fisiología Celular durante la duración del mismo.

No es necesario ningún conocimiento previo de los temas a desarrollar durante el curso.

Si bien es muy recomendable que cada estudiante cuente con un equipo de cómputo adecuado para llevar a cabo las actividades del curso, el instituto puede proporcionar un número limitado de laptops para su uso dentro de las instalaciones del Instituto (favor de consultar la disponibilidad antes de solicitar su inscripción).

## Temario

1 - Introducción al cómputo científico para las ciencias biológicas (Carlos Peralta 2 h, César Poot 2h)

- Las computadoras y sus partes.
- Componentes fundamentales
- Sistemas operativos GNU/Linux y su uso en ciencias biológicas.
- ¿Cuál es la mejor distribución de GNU/Linux para bioinformática?
- Preparación para la instalación de un ambiente basado en Unix para bioinformática.
- Aprende a elegir/diseñar un equipo de cómputo en función de las necesidades de análisis.

2 - Instalación y gestión de un sistema operativo GNU/Linux - Parte I (Carlos Peralta 2 h, César Poot 2h)

- Usos y limitaciones de los diferentes tipos de instalación de sistemas GNU/Linux.
- Instalación física en equipo de alto rendimiento
- Instalación física en computadora personal (PC)
- Instalaciones virtualizadas:
  - Contenedores
  - Máquinas virtuales
  - Windows Subsystem for Linux.
  - Instalación frugal (live-USB)
- Práctica de conexión remota a un servidor de análisis para el aprendizaje de:
- Introducción al sistema de directorios
- Comandos fundamentales (parte I)

3 - Instalación y gestión de un sistema operativo GNU/Linux - Parte II (Carlos Peralta 2 h, César Poot 2h)

- Aspectos claves a tener en cuenta al instalar un sistema GNU/Linux
- La BIOS y sus opciones para instalación y/o virtualización

- Tabla de particiones
- Dual boot (convivencia con windows en un mismo equipo)
- La interfaz de texto y el Shell
- Comandos principales (parte II)
- Tipos de usuario de un sistema GNU/Linux; Root, superuser, usuario regular
- Permisos y privilegios
- El comando sudo
- Instalación de software bioinformático y de uso general
- Gestores de paquetes binarios (apt, dnf, pacman, etc.)
- Gestores de paquetes empaquetados (flatpak, snap, appimage)
- Aplicaciones distribuidas como contenedores (Docker, Podman, etc)
- Aplicaciones distribuidas dentro de ambientes virtuales (Anaconda)
- Editores de texto y el texto plano en bioinformática.
- Nano, Vim, Emacs, MS-edit y por qué Emacs es superior.
- El texto plano y su importancia para la preservación de la integridad de los datos en bioinformática.
- Tipos de archivos usados en bioinformática
- Práctica de instalación y uso de un entorno personalizado GNU/Linux para el trabajo bioinformático.
- Instalación de un sistema GNU/Linux.
- Instalación de paquetes binarios de bioinformática.
- Procesamiento de datos bioinformáticos.
- Creación de un script de automatización.
- Personalización de la experiencia del usuario, modificación de archivos de configuración
- Comandos útiles (parte III)

4 - Proyecto de bioinformática con GNU/Linux (Carlos Peralta 2 h, César Poot 2h)

- Implementación de un análisis de ChIP-Seq de principio a fin.
- Fundamento del ChIP-Seq
- Flujo de trabajo general para el análisis de datos genómicos
- Tipos de datos de entrada, intermedios y de salida
- Repositorios de información bioinformática
- Obtención de datos desde la terminal

5 - Brevísima introducción a la programación para bioinformática (Carlos Peralta 2 h, César Poot 2h)

- Introducción al pensamiento algorítmico
- Los lenguajes de programación usados frecuentemente en bioinformática, sus diferencias y casos de uso.
- Bash, Python y R
- Otros lenguajes emergentes y en desuso
- IDEs y cuadernos de programación
- Rstudio, Spyder, Jupyter, VScode
- Entornos colaborativos
- Cuándo y cómo usar los modelos de lenguaje (IA) para pedir ayuda.
- La documentación y su importancia
- Alternativas para no programar (Interfaces gráficas locales y en línea)
- Práctica de programación
- Apertura de cuenta en Google Colab
- Programa en R
- Programa en Python

6 - Introducción a R para bioinformática (parte I) (Carlos Peralta 4 h)

- Historia y filosofía de R
- Bioconductor
- Práctica en Google Colab para aprender:
- Sintaxis del lenguaje
- Estructuras de datos
- Ploteo (parte I)

7 - Introducción a R para bioinformática (parte II) (Carlos Peralta 4 h)

- Práctica en Google Colab para aprender:
- Iteración y automatización
- Llamadas a sistema
- Importación y exportación de archivos
- Ploteo (parte II)

8 - Introducción a R para bioinformática (parte III) (Carlos Peralta 4 h)

- Rstudio y otros IDEs

- El infierno de las dependencias (parte I)
- Condaforge al rescate
- Métodos de instalación de R.

9 - Proyecto de bioinformática (Carlos Peralta 4 h)  
- Análisis de expresión diferencial (RNA-Seq) con R

10 - Introducción a Python para bioinformática (parte I) (Augusto Poot 4h)  
- Principios de python como lenguaje de programación  
- Estructuras de datos  
- Sintaxis básica

11 - Introducción a Python para bioinformática (parte II) (Augusto Poot 4h)  
- Funciones  
- Elementos de control de un programa

12 - Introducción a Python para bioinformática (parte III) (Augusto Poot 4h)  
- Bibliotecas para manejo y visualización de datos  
- Numpy, scipy, matplotlib.

13 - Introducción a Python para bioinformática (parte IV) (Augusto Poot 4h)  
- Bibliotecas especializadas  
- ScikitLearn, ScikitImage, Biopython, Keras

14 - Examen teórico-práctico final. (Augusto Poot 2h / Carlos Peralta 2h)

15 - Presentación de resultados por parte de los alumnos. (Augusto Poot 2h / Carlos Peralta 2h)

16 - Presentación de resultados por parte de los alumnos. (Augusto Poot 2h / Carlos Peralta 2h)

## Bibliografía

Blum, R., & Bresnahan, C. (2021). Linux command line and shell scripting bible (4th ed.). Wiley.

Wickham, H., Çetinkaya-Rundel, M., & Grolmund, G. (2023). R for data science (2nd ed.). O'Reilly Media

How to Think Like a Computer Scientist: Learning with Python. 3rd Edition. <https://openbookproject.net/thinkcs/python/english3e/>. (2012)

## Observaciones

N/A