



Licenciatura en **CIENCIAS GENÓMICAS**

Formato para proponer cursos Semestre 2024-1

El curso ya ha sido impartido: Sí No

1. Indicar modalidad: (Optativo, Seminario, curso regular (con profesor invitado)
Curso regular

2. Título: Se sugiere que sea conciso y refleje el contenido general
Python II (Bioinformática y Estadística 1)

3. Tutor responsable:

Nombre completo

LI Heladia Salgado Osorio

6. Descripción del curso

En Python se han desarrollado muchas bibliotecas para la extracción, manipulación y presentación de datos. El objeto central en bioinformática es la secuencia, por lo tanto, el objetivo principal de BioPython es desarrollar bibliotecas y aplicaciones de Python que aborden la necesidad del trabajo actual y futuro en bioinformática.

Python también cuenta con diferentes módulos para analizar estos tipos de datos, como lo son NumPy, Pandas y SciPy. Además de sus complementos para hacer gráficos y hacer más fácil su interpretación, como lo son Matplotlib y Seaborn.

El curso tiene como objetivo, ayudar a los estudiantes a conocer algunas de las bibliotecas que se han desarrollado para el análisis de secuencias, el acceso a base de datos de secuencias, la homología entre secuencias, por mencionar algunas, y seguir aplicando las buenas prácticas para el desarrollo de software.

Además durante el curso se hará énfasis al control del código a través de versiones usando GitHub, y se impulsará al desarrollo de software colaborativo.

Pre-requisitos

1. Introducción al bio-cómputo con sistema Linux (welcome)
2. Programación práctica con Python

Los alumnos deberán cumplir con TODOS los requisitos que se listan a continuación :

- Se requieren conocimientos básicos de programación en Python (estructuras de control (condicionales y ciclos); estructuras de datos (arreglos, hash, listas); funciones y paso de parámetros. El uso de lenguaje en línea de comando también sería útil.

- Los participantes necesitan tener conocimientos de nivel intermedio en la línea de comandos de UNIX/Linux lo que implica manejo de archivos y directorios (rm, mkdir, cd, less, cut, sort, uniq, etc).
- Comprensión del Inglés.

Forma y criterio de evaluación

Para la evaluación final se considera lo siguiente:

1. Todas las entregas electrónicas deberán entregarse a través de la plataforma de cursos en formato PDF antes de la fecha límite. TAREAS QUE NO CUMPLAN CON ESTAS ESPECIFICACIONES NO SERÁN CALIFICADAS. Sólo se harán excepciones por motivos de salud presentando el correspondiente certificado médico.
2. Todos los programas que se pidan como tareas o ejercicios, deben
 - ❖ Ser colocados en el repositorio de GitHub
 - ❖ Cumplir con las mejores prácticas de desarrollo de software que se vayan explicando.
3. La calificación final comprende:
 - a. 40% tareas y ejercicios que se soliciten durante el curso y que cumplan con los dos puntos anteriores.
 - b. 10% participación en clase
 - c. 50% presentación de proyecto de software colaborativo, que incluye una presentación de 15 min del proyecto, la revisión del repositorio en github, la aplicación de las mejores prácticas en el proyecto.

Contenido

GitHub Intermedio [3 sesiones]

- Especificar versiones en Git con tag
- Trabajar con ramas en Git: git branch
- Manejo de proyectos
 - Issues
 - proyectos

BioPython

- Introducción a librerías
 - ¿Qué son SciPy, NumPy, Pandas, Matplotlib, Seaborn y BioPython?
 - Instalar SciPy, NumPy, Pandas, Matplotlib, Seaborn y BioPython
 - Secuencias y formatos
 - Leer y escribir archivos de secuencias
 - Trabajar con secuencias biológicas
 - Parseo de archivos con formato biológico
 - Acceso a Bases de datos
 - Aprendiendo como acceder a datos desde NCBI
 - Accediendo y recuperando datos de bases de datos como
 - ExpASy -- <http://www.expasy.ch/>

- Entrez from NCBI -- <http://www.ncbi.nlm.nih.gov/Entrez/>
- PubMed from NCBI -- <http://www.ncbi.nlm.nih.gov/PubMed/>
- SCOP -- <http://scop.mrc-lmb.cam.ac.uk/scop/>
- Vectores y Matrices con NumPy (ver <https://numpy.org/> DATA SCIENCE section)
 - Tipos de datos
 - Generar Arrays
 - Entrada / Salida
 - Métodos y operaciones
 - Indexar y asignar
- DataFrames y Series con Pandas
 - Tipos de datos
 - Entrada / Salida
 - Series y DataFrames
 - Métodos y operaciones
- Jupyter Notebooks
 - Cómo trabajar en Jupyter
 - Presentación de datos
- Creación de gráficos con Matplotlib
 - Tipos de gráficas y cómo crearlas
 - Gráficas con Pandas
 - Formatos de salida
- Seaborn
 - Tipos de gráficas y diferencias con Matplotlib
 - Gráficas con Pandas
 - Formato de salida
- Algoritmos matemáticos con SciPy
 - Entrada / Salida
 - Funciones
 - Álgebra lineal
 - Optimización y ajuste

7. Características para la impartición del curso :

Lugar donde se realizará	LCG, CCG
Duración en horas por sesión y número de sesiones	2 horas
Disponibilidad de impartirlo por videoconferencia	Sí <u>X</u> No <u> </u> Curso presencial, pero también se mantiene la modalidad virtual para alumnos enfermos o con otra circunstancia.

8. Método de evaluación:

Por favor incluya en este apartado el % de la contribución relativa de:

Participación en clase	10%
------------------------	-----

Presentación en clase	0%
Proyecto de investigación	50%
Trabajos	40%
Otros	

10. Bibliografía

Referencias:

- Scipy:
<https://www.oreilly.com/library/view/elegant-scipy/9781491922927/ch01.html>
- BioPython: <http://biopython.org/DIST/docs/tutorial/Tutorial.pdf>
- Python Data Analytics: With Pandas, NumPy, and Matplotlib
<https://www.amazon.com.mx/Python-Data-Analytics-Matplotlib-English-ebook/dp/B07FT6FB6Y>