

Práctica 3

MSc. Vicente Machaca Arceda

20 de agosto de 2020

DOCENTE	CARRERA	CURSO
MSc. Vicente Machaca Arceda	Escuela Profesional de Ciencias de la Computación	Computación molecular Biológica

PRÁCTICA	TEMA	DURACIÓN
01	Thresholding	3 horas

1. Competencias del curso

- Aplica las bases matemáticas y la teoría de la informática en algoritmos de Bioinformática.
- Analiza, diseña y propone soluciones frente a problemas bioinformáticos.
- Sabe cómo utilizar y conoce las bases computacionales de herramientas modernas de secuenciamiento, alineamiento, árboles filogenéticos y mapeo de genomas.

2. Competencias de la práctica

- Utilizar herramientas de Dot plot.
- Implementar el algoritmo Dot plot para alineamiento de secuencias.

3. Equipos y materiales

- Python
- Matplotlib
- Numpy
- BioPython
- Cuenta en Github

4. Entregables

- Se debe elaborar un informe en Latex donde se responda a cada ejercicio de la Sección 5.
- En el informe se debe agregar un enlace al repositorio Github donde esta el código.
- En el informe se debe agregar el código fuente así como capturas de pantalla de la ejecución y resultados del mismo.

5. Ejercicios

1. Descargue dos secuencias de proteínas (las utilizadas en clases) y ejecute el siguiente código. Comente sus resultados.

```
# dot_matrix.py
from Bio import SeqIO

sequences = SeqIO.parse("P21333.fasta", "fasta")
for record in sequences:
    data1 = str(record.seq.upper()) # the fasta file just have one sequence

sequences = SeqIO.parse("Q8BTM8.fasta", "fasta")
for record in sequences:
    data2 = str(record.seq.upper()) # the fasta file just have one sequence

print(data1)
print(data2)
```

2. Ahora usted debe implementar un programa que genere un Dot matrix. Se recomienda utilizar Matplot para la gráfica, de igual manera no es necesario dibujar las líneas, basta con dibujar los puntos por cada coincidencia (no olvide incluir el *windows size* y un *threshold*).
3. Descargue otras secuencias y genere el *Dot matrix*, evalúe sus resultados y comente.