

UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE CIENCIAS

CC471 - BIOLOGÍA COMPUTACIONAL

ePydemic

Herramienta gráfica para el análisis genético de especies
endémicas

- Propuesta de proyecto -

Autores:

Ingrid Ipanaqué Casquina
Luis Vasquez Espinoza

Profesor:

Ciro Nuñez Iturri

September 25, 2018

1 Resumen ejecutivo

El presente proyecto busca integrar las herramientas disponibles en la librería biopython tales como **alineamiento de genes homólogos**, **alineamiento de proteínas** y **generación de árboles filogenéticos**. Se podrá acceder a estas herramientas mediante módulos incluidos en la interfaz gráfica presentada en nuestro producto final. **ePydemic** busca evitar al usuario la ardua tarea de redactar código para revisar los resultados de estas operaciones elementales del análisis biológico.

2 Descripción del proyecto

2.1 Alcances

Durante el trabajo de este proyecto se espera lograr cubrir las siguientes necesidades:

- Crear una interfaz intuitiva, responsiva y práctica para aplicarla al análisis biológico sin conocimientos de programación

Para el producto final de este trabajo se busca que el proyecto cumpla con los siguientes aspectos:

- Capacidad de cargar, leer e interpretar ficheros de elementos biológicos (ácidos nucleicos, aminoácidos, etc.) con contenido secuenciado.
- Capacidad de alinear dos secuencias previamente importadas, ya sean genes o proteínas.
- Capacidad de generar y visualizar árboles filogenéticos a partir de dos secuencias.

2.2 Cronograma

El cronograma de trabajo que especifica la naturaleza de los entregables durante el semestre se detalla en la siguiente tabla:

Fecha de entrega	Evaluación equivalente	Entregable
25/Set/18	-	- Propuesta del proyecto
29/Oct/18	Prac. Calificada 3	- Version final de la propuesta del proyecto - Definición de especies, genes, y proteínas - Demostración de distribución de paneles y herramientas a disponer en la versión final - Demostración de manipulación básica de datos (Importe de archivos formateados, lectura y escritura)
26/Nov/18	Prac. Calificada 4	- Demostración de operaciones básicas sobre genes y proteínas de las especies escogidas (carga, análisis y alineamientos)
03/Dic/18	Lab. Calificado 4	- Demostración pulida de las operaciones anteriores - Demostración de generación de árboles filogenéticos

3 Algoritmos e implementación computacional

3.1 Modelo de trabajo

Se utilizará el lenguaje **python** en su versión *3.6*. Las variables y los métodos serán escritos en *snake case*, las clases en *camel case* y las constantes en *mayúsculas*. Se utilizará la programación orientada a objetos con una perspectiva modular (debido a la necesidad de implementar la GUI descrita) y se conectarán los métodos a esta mediante *conectores* basados en programación orientada a eventos (Event-driven Programming).

3.2 Manejo de interfaz gráfica

Para la construcción de la interfaz de interacción se empleará la librería *Qt* en su implementación en python: **PyQt4**. Se emplearán paradigmas modulares y de encapsulamiento para distribuir el trabajo de manera colaborativa y que sea mantenible en el transcurso del desarrollo.

3.3 Herramientas de análisis biológico

Se emplearán la librería BioPython para la manipulación de formatos biológicos, ocupando módulos como **SeqIO** para asegurar el correcto formateo de los archivos de entrada hacia la aplicación. Además se utilizará la aplicación **clustalw** para operar alineamientos múltiples y generación de árboles filogenéticos como una alternativa a los métodos utilizados en Biopython. **Clustalw** requerirá la integración de manejo del sistema operativo via python para poder operar comandos sobre *bash*

4 Resultados esperados

Con ePydemic se podrá realizar el análisis bioinformático - filogenético de secuencias genéticas específicas y proteínas pertenecientes a especies endémicas del **Perú**. Aún se está evaluando exactamente qué especies se utilizarán para probar la plataforma, pero se espera emplear datos biológicos de aves pertenecientes a familias como *Scytalopus*, *Synalaxis*, *Geositta*; aplicando secuencias proteicas proporcionadas por la **NCBI** y mostrando como respuestas visual gráficas resultantes de comparaciones y alineamientos.

5 Conclusiones

El estudio de genes supone un reto para todo aquél sin conocimientos en programación, por lo que el proyecto tiene el reto de apoyar y suplir esta necesidad. El estudiar especies endémicas también nos favorece a nosotros, como estudiantes, dado que adquirir conocimientos por nuestra fauna local le da un valor agregado a los proyectos que tengamos en un futuro como científicos concientes en el impacto local.

6 Apéndice

- Especie Endémica: Son seres vivos, que incluyen tanto la flora como la fauna, cuya distribución se restringe a una determinada zona geográfica. [2]

- BioPython: Biopython permite representar secuencias biológicas y anotaciones de genomas y es capaz de comunicar con las bases de datos biológicos en línea del NCBI para hacer cálculos. [1]
- PyQt4: Es un conjunto de herramientas para crear aplicaciones con interfaz gráfica. Es una mezcla entre Python y la biblioteca Qt. [4]
- Clustalw: Es un programa de propósito general para el alineamiento de secuencias múltiples de ADN o proteínas. [3]

7 Bibliografía

References

- [1] BIOPYTHON: PYTHON TOOLS FOR COMPUTATIONAL MOLECULAR BIOLOGY
. Retrieved from <https://biopython.org/>.
- [2] CATEGORÍA:AVES ENDÉMICAS DE PERÚ
. Retrieved from <https://bit.ly/2pwwZLt>.
- [3] CLUSTALW: MULTIPLE SEQUENCE ALIGNMENT
. Retrieved from <http://www.clustal.org/>.
- [4] PYQT4: PYTHON BINDINGS FOR THE QT CROSS PLATFORM GUI TOOLKIT
. Retrieved from <https://pypi.org/project/PyQt4/>.