

UNIVERSIDAD NACIONAL DE INGENIERIA

FACULTAD DE CIENCIAS

CIENCIAS DE LA COMPUTACIÓN



PROYECTO DE BIOLOGIA COMPUTACIONAL

Título del Trabajo

**Creación de un sistema de ayuda para el analisis de
información Genética de Especies endémicas de las regiones
del Perú**

Autores

Lázaro Camasca, Edson Nicks
Leon Rios, Marco Naro

Profesor

Nuñez Iturri, Ciro Javier

Lima - Peru
(2019)

Contents

1	Objetivos	2
1.1	Objetivos Generales	2
1.2	Objetivos Especificos	2
2	Resumen Ejecutivo	2
3	Descripción del Proyecto	2
3.1	Determinar las secuencias a utilizar	2
3.2	Elegir los marcadores moleculares	2
3.3	Realizar el lineamiento multiple de genes homologos	2
3.4	Elegir un modelo evolutivo	2
3.5	Aplicar un método para la construcción del árbol filogenético	2
3.6	Verificar la fiabilidad del árbol construido	3
3.7	Analizar el arbol filogenetico	3
3.7.1	Modelamiento de la Estrutura de Proteinas	3
3.8	Cronograma	3
4	Algoritmos e implementación computacional	3
5	Resultados	4
6	Conclusiones	4
7	Apéndice	4

1 Objetivos

1.1 Objetivos Generales

- Creación de un software grafico con una base de datos para el analisis de infomacion genética.

1.2 Objetivos Especificos

- Recolectar informacion genetica de especies endemicas.
- Desarrollar la aplicación para el analisis de secuencias,.
- Desarrollar algoritmos para obtener árboles filogenéticos
- Evaluar el árbol filogenético

2 Resumen Ejecutivo

Se pretender crear un software y una base de datos con la informacion genetica de las especies endemicas del peru, el software procesara las secuencias, crara el árbol filogenético, mostrara los resultados y analizara las relaciones evolutivas de las especies escogidas.

3 Descripción del Proyecto

El proyecto sera implementado netamente en el lenguaje Pyhton Las librerias utilizadas seran:

- BioPython para el procesamiento de secuencias
- Tkinter o Gt para el entorno grafico.

Dentro de la GUI, se pobra escoger /Especie/Genes/Proteina para el análisis posterior.

Los datos recolectados seran reales de la base de datos de NCBI.

Se implementara algoritmos para el alineamiento de Genes homólogos.

Se implementara algoritmos para el alineamiento de Proteinas.

Se implementara la algoritmos para la Generacion de Arboles Filogenticos de acuerdo a un modelo.

Para el desarrollo del proyecto se seguira la siguiente metodologia:

3.1 Determinar las secuencias a utilizar

El alineamiento de secuencias debe permitir la determinación de regiones homologas

3.2 Elegir los marcadores moleculares

3.3 Realizar el lineamiento multiple de genes homologos

3.4 Elegir un modelo evolutivo

Entre los modelos tenemos:

- Modelo Jukes-Cantor: asumiento que los nucleotidos son substituidos con igual probabilidad
- Modelo Kimura: realista ya que considera las diferentes tasas de mutación

3.5 Aplicar un método para la construcción del árbol filogenético

Entre los metodos tenemos

- Basados en distancia
- Basados en secuencia
- Basados en agrupamiento

- Basados en optimalidad
- Basados en caracteres

3.6 Verificar la fiabilidad del árbol construido

3.7 Analizar el árbol filogenético

Apartir del árbol filogenético se podrá descubrir/Mostrar/Analizar las relaciones evolutivas de las especies endémicas escogidas.

3.7.1 Modelamiento de la Estructura de Proteínas

En el análisis se encuentra el modelamiento de la Estructura de Proteínas

3.8 Cronograma

Para el desarrollo del proyecto se emplea un cronograma por semanas, las fechas del cronograma coinciden con las fechas propuestas para evaluaciones de práctica del proyecto. En dichas ocasiones se presentarán y evaluarán los avances del proyecto.

DESCRIPCION	SEMANA									
	7	8	9	10	11	12	13	14	15	
Propuesta inicial										
Determinar los datos										
Elegir las especies endemicas										
Recolectar las seuencia y proteinas										
Alineamiento										
Alineamiento en genes homologos y proteinas										
Analisis de los alineamientos										
Creacion del arbol filogenetico										
Elegir el modelo evolutivo y metodo										
Implementar el algoritmo y verificar funcionamiento										
Analizar el arbol filogenetico										
Modelamiento de la estrutura de proteinas										
Avance del proyecto	5%	10%	15%	30%	40%	50%	60%	70%	100%	

Fig: Cronograma por semanas

SEMANA	FECHA	ENTREGABLE	DESCRIPCION
7	29/04/19 al 06/05/19	Entregable 1	Propuesta inicial del proyecto
8	06/05/19 al 13/05/19		
9	13/05/19 al 20/05/19		
10	20/05/19 al 27/05/19	Entregable 2	Propuesta final y elección de especies, genes, proteínas
11	27/05/19 al 03/06/19		
12	03/06/19 al 10/06/19		
13	10/06/19 al 17/06/19		
14	17/06/19 al 24/06/19	Entregable 3	Realización y análisis de alineamiento de genes y proteínas
15	24/06/19 al 01/07/19	Entregable 4	Análisis y generación de árboles filogenéticos

Fig: Cronograma coincidente a los entregables

4 Algoritmos e implementación computacional

Una descripción de los algoritmos y herramientas que se [planean utilizar en caso de la propuesta] utilizados incluyendo pseudo código y código fuente

5 Resultados

Una descripción de los resultados [esperados en el caso de la propuesta]. Un reporte integrando los resultados proporcionados por la herramienta

6 Conclusiones

Incluye las ventajas y desventajas del enfoque utilizado, aspectos inesperados del proyecto, trabajo futuro, etc.

7 Apéndice