



*Facultad de Ciencias Exactas y Naturales es parte de la  
Universidad de Buenos Aires*

Trabajo Final de Materia Bioinformática Avanzada

*Biog5: Sistema para el análisis del virus del papiloma humano*

Integrantes:

- Ernesto Rafael Perez
- Sofia Erdozain

## Índice de contenido

Capítulo 1: Introducción y marco teórico .....	4
1.1 Introducción .....	4
1.2 Objetivo general .....	4
1.3 Objetivos específicos .....	4
1.4 Fundamentación .....	4
1.4.1. Base datos de proteínas y genomas .....	5
1.4.2. Análisis de a pares utilizando BLAST .....	5
1.4.3. Análisis múltiple utilizando Clustal Omega .....	5
1.4.4. Análisis comparativo usando modelos HMM utilizando HAMMER .....	5
1.4.5. Módulos para Análisis de Variantes en Proteínas y Genomas .....	6
Capítulo 2: Metodología .....	7
2.1 Back logs: Tareas a realizar .....	7
2.2 Análisis de requerimientos de software .....	8
2.2.1 Descripción general .....	8
2.3 Especificaciones .....	9
2.3.1 Módulo Main .....	9
2.3.2 Módulo CargaBD .....	9
2.3.3 Módulo Utils .....	10
2.3.4 Módulo alineamientos de a pares para proteínas .....	10
2.3.5 Módulo alineamientos de a múltiples para proteínas .....	11
2.3.6 Módulo modelos HMM .....	11
2.3.7 Módulo variantes en proteínas .....	12

2.3.8 Módulo variantes en genomas .....	12
2.4 Diagrama de caso de uso.....	13
2.5 Diagrama entidad relación .....	13
Capítulo 3: Herramientas y lenguajes de programación .....	15
3.1 Lenguajes de programación y herramientas .....	15
3.1.1 Requerimientos e Instalaciones: .....	15
3.2 Servidores y entornos de desarrollo .....	16
3.3 Estándares de codificación.....	17
Capítulo 4: Resultados .....	18
4.1 Módulo principal.....	18
4.2 Análisis de alineamientos de a pares.....	18
4.3 Análisis de alineamientos múltiples MSA .....	22
4.4 Análisis de cepas usando modelos HMM .....	24
4.5 Análisis de variantes en proteínas .....	26
4.6 Análisis de variantes en genomas.....	29
Capítulo 5: Conclusión .....	32
Referencias bibliográficas .....	33

# Capítulo 1: Introducción y marco teórico

## 1.1 Introducción

Existen más de 100 tipos de HPV de los cuales al menos 14 son causantes de cáncer. Los tipos HPV 16 y HPV18 son las principales variantes encontradas en cáncer de cérvix y en lesiones de cérvix precancerosas por lo que son consideradas de alto riesgo. Si bien cada vez más tipos de HPV son considerados de alto riesgo, aunque no se encuentren en tejido de cáncer de cérvix debido a que se los encuentra con frecuencia en lesiones escamosas intraepiteliales, las vacunas aprobadas protegen contra los HPV tipo 16 y 18 [1].

El virus de papiloma humano (HPV) es un virus no envuelto, icosaedro de aproximadamente unos 60nm. Su material genético se compone de una sola molécula de ADN circular de 8kb asociado a histonas. Solo una de las hebras se transcribe y da lugar a dos clases de proteínas expresadas por splicing alternativo: las proteínas tempranas que son proteínas no estructurales regulatorias (E1-E2-E7), y las tardías que son las proteínas estructurales L1 y L2 que conforman la cápside [2]. E1 es la Proteína de replicación, encargada de replicación del ADN en conjunto con la Proteína reguladora E2. L1 forma los pentámeros que conforman la cápside y L2 se localiza en la superficie interna del virión en los espacios entre los pentámeros de E1.

## 1.2 Objetivo general

Desarrollar “**Biog5**” un sistema que integre múltiples herramientas bioinformáticas que permitan estudiar las posibles variaciones de las proteínas E1, E2, E7, L1 y L2 entre los tipos de HPV considerados de alto riesgo respecto a los de riesgo bajo/no determinado. Incluyendo el estudio de características de grupos de riesgos y proteínas de interés.

Buscar variantes clínicas en dichas proteínas y determinar su frecuencia. ¿hay mayor frecuencia de variantes en alguna?

## 1.3 Objetivos específicos

1. Desarrollar un módulo principal, integrador con interfaz de usuario interactiva.
2. Desarrollar un módulo para análisis comparativo de a pares entre proteínas de interés.
3. Desarrollar un módulo para análisis comparativo múltiple entre proteínas de interés.
4. Desarrollar un módulo para análisis comparativo usando modelos HMM entre proteínas de interés.
6. Desarrollar un módulo para análisis de variantes en proteínas de interés.
7. Desarrollar un módulo para análisis de variantes en genomas HPV.
8. Publicar y Difundir.
9. Obtener colaboraciones científicas para optimización e implementación de algoritmos en base a nuevos estudios.

## 1.4 Fundamentación

Este trabajo tiene como objetivo encontrar características semejantes y divergentes de manera cuantitativa en las distintas cepas del virus del papiloma humano, agrupando y analizando relaciones

internas en genomas clasificados en grupos de riesgo y proteínas de interés y a su vez interacciones entre las mismas. Para realizar un sistema que permita estos análisis se requiere crear múltiples módulos cada uno con algoritmos bioinformáticos propios incluyendo herramientas y recursos ya desarrollados especializados para en análisis comparativo. En las siguientes secciones se explican herramientas y recursos aplicados a cada módulo del sistema.

#### 1.4.1. Base datos de proteínas y genomas

Se realizó primeramente una búsqueda de información de manera manual sobre las cepas HPV, en la plataforma “The National Center for Biotechnology Information (NCBI)”. Esta plataforma promueve la ciencia y la salud al brindar acceso a información biomédica y genómica [3], se obtuvieron los identificadores de los genomas y proteínas con los cuales se generaron dos tablas **proteínas y genomas**, luego una tercera **clasificación**, que contiene el nombre de genomas clasificados en riesgo a salud.

La clasificación se basa en 3 grupos de riesgos.

- Alto riesgo
- Bajo riesgo
- No especificado riesgo

Agrupadas las tablas se procede a realizar un módulo donde se utilizarán librerías de biopython, para descargar las secuencias en formato fasta de manera automática.

#### 1.4.2. Análisis de a pares utilizando BLAST

La herramienta de búsqueda de alineación local básica (BLAST) encuentra regiones de similitud local entre secuencias. El programa compara secuencias de nucleótidos o proteínas con bases de datos de secuencias y calcula la significación estadística de las coincidencias. BLAST se puede utilizar para inferir relaciones funcionales y evolutivas entre secuencias, así como para ayudar a identificar miembros de familias de genes [4].

El módulo Alineamiento de a pares, y Variantes en genomas, incluye esta herramienta (Debe instalarse se describe cómo hacerlo en próximas secciones) la misma es utilizada para comparar proteínas de distintas cepas y genomas pertenecientes a estas cepas.

#### 1.4.3. Análisis múltiple utilizando Clustal Omega

Clustal Omega es un nuevo programa de alineación de secuencias múltiples que utiliza árboles guía sembrados y técnicas de perfil de perfil HMM para generar alineaciones entre tres o más secuencias [5].

El trabajo realizado integro esta herramienta para hallar familias y árboles filogenéticos en distintos módulos; se incluye el aplicativo en el directorio raíz como un ejecutable portable para el Sistema Operativo Windows.

#### 1.4.4. Análisis comparativo usando modelos HMM utilizando HAMMER

HMMER se utiliza para buscar en las bases de datos de secuencias homólogos de secuencia y para realizar alineaciones de secuencias. Implementa métodos que utilizan modelos probabilísticos llamados modelos de perfil oculto de Markov o perfil HMM [6].

En este trabajo la herramienta se utiliza en un entorno Linux a través del desarrollo de un bash script que crea modelos HMM de proteínas del grupo de cepas de alto riesgo que se compraran con otros grupos de riesgos.

Los resultados comparativos del script, se analizaron en el módulo HMM, donde sobre cada grupo clasificado en tipo de riesgo y clase de proteína, se procede a filtrar valores de certeza informando solo las cepas con mejores puntajes que han sido asignadas a cada modelo.

#### **1.4.5. Módulos para Análisis de Variantes en Proteínas y Genomas**

Se define el Análisis de Variantes como un tipo de análisis de datos de secuenciación cuyo objetivo es ver las variaciones que presenta la secuencia genómica de la muestra respecto a la del genoma considerado de referencia.

Se crearon dos módulos para estudiar las variantes en distintas proteínas y genomas. En ambos se utilizó el programa Custal Omega para el alineamiento de secuencias.

Módulos:

- **Análisis de Variantes en Proteínas:** Analiza variantes en grupos de proteínas clasificadas en riesgos.
- **Análisis de Variantes en Genomas:** Analiza dos o más genomas HPV.

En los anteriores se utilizan de algoritmos que analizan las frecuencias de aparición y se informan a petición y formato a elección las diferencias, y aminoácidos más conservados.

## Capítulo 2: Metodología

Las metodologías son marcos de trabajo para planificar y gestionar el proceso de un desarrollo de software, las cuales combinan ciclos de vida con tecnologías y herramientas.

Para este trabajo final de la materia, se ha optado por utilizar el ciclo de vida iterativo incremental. El un modelo iterativo e incremental dividido en tiempos cortos y rápidos llamados sprints (primaveras). Cada sprint se centra en unos requisitos concretos, también llamados historias de usuarios, que tienen una duración definida, como así también cada sprint recoge las fases de desarrollo de software comunes, como ser el análisis, diseño, implementación y pruebas. Mediante esta metodología se van finalizando los requisitos sencillos (incremental) y se van perfeccionando aquellos y nos brinda gran flexibilidad, adaptabilidad e integración en diversos módulos de proyectos. [7]

### 2.1 Back logs: Tareas a realizar

El product backlog (tareas a realizar del producto) se puede ver desde la perspectiva de una iteración o sprint, de una release (lanzamiento) o de todo el producto. En cualquier caso, sigue siendo una lista priorizada de historias de usuario más o menos detalladas, aunque hablemos en cada caso de sprint backlog, release backlog o product backlog.

Primeramente, se establece una lista de tareas a desarrollar en todo el proyecto. Esto se hace para tener una perspectiva general de todo lo que se quiere hacer y tener claras las prioridades del proyecto.

A continuación, se muestra la lista de tareas generales iniciales, a las cuales se le incluyo algunas tareas extras que fueron agregadas, durante el proceso de desarrollo.

- **Back logs del producto:**

1. Configurar entorno de Python con librerías necesarias.
2. Crear y configurar un repositorio Git.
3. Realizar un módulo para la carga de datos de proteínas y genomas.
4. Realizar el módulo de alineamientos de a pares para proteína.
5. Realizar el módulo de alineamientos múltiples proteína.
6. Realizar el módulo de modelo HMM.
7. Realizar el módulo de variantes en proteínas.
8. Realizar el módulo de variantes en genomas.
9. Optimizar algoritmos generalizando funciones reutilizables.
10. Unificar en el un módulo Main.
11. Realizar los manuales de usuario.
12. Mejorar interacción con el usuario.

Las tareas anteriores se establecen como metas a cumplir, que se van seleccionando según su prioridad para ser desarrolladas en cada sprint, las mismas se suelen dividir en sub tareas a realizar. Al final

de cada sprint de una duración aproximada de 4 semanas a 5 semanas se realiza un release de la nueva versión del software.

Se planifica para el primer reléase de Biog5, se incluyan los siguientes componentes en el sistema:

- Tag 1: Interfaz de usuario integral que ejecute un módulo destinado a la gestión de datos.
- Tag 2: El sistema basado en un menú por consola, que permite realizar el módulo de alineamientos de a pares para proteína.
- Tag 3: Un módulo de alineamientos múltiples proteína.
- Tag 4: Un módulo de modelos HMM.
- Tag 5: Un módulo de variantes en proteínas.
- Tag 6: Un módulo incluyo el módulo de variantes en genomas.

Las pruebas se hicieron con datos obtenidos de referencias bibliográficas y búsqueda manual de información en plataformas especializadas.

## **2.2 Análisis de requerimientos de software**

En esta sección se proporcionará información sobre características generales que posee el sistema.

### **2.2.1 Descripción general**

En esta sección se presenta una vista general del sistema que incluye todos los módulos del mismo.

**El sistema consta de ocho módulos, los cuales se detallan en la siguiente sección.** En la Fig. 1 se muestra un diagrama general de la estructura del sistema con sus interacciones, obtenido a partir del estudio de todos los requerimientos del usuario y del sistema que se ha analizado a través de las reuniones con el grupo Biog5.



## Presentación del sistema

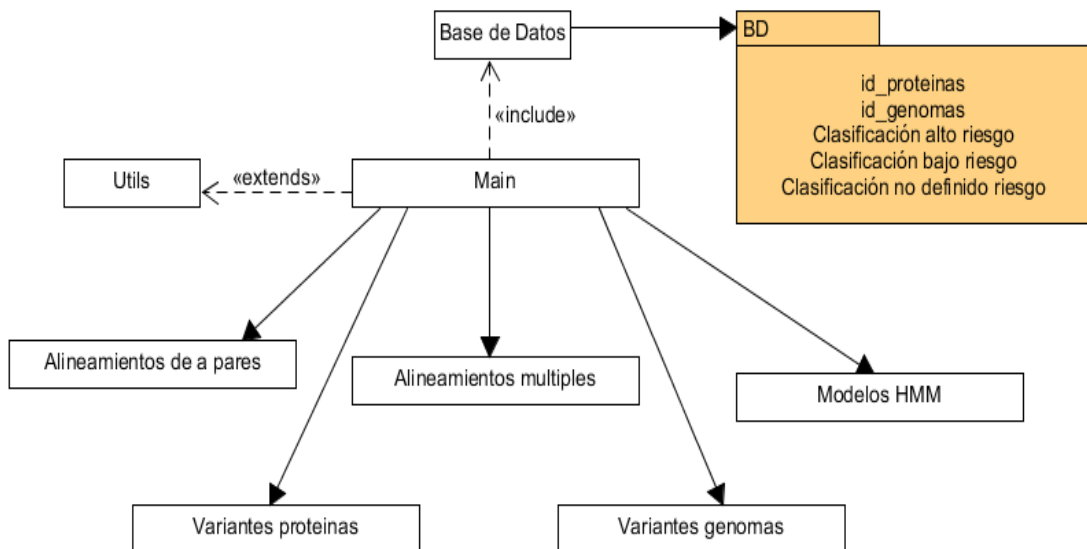


Fig. 1: Estructura general del sistema. Fuente: Elaboración propia.

## 2.3 Especificaciones

Se detallan módulos entradas, salidas parámetros, funciones principales.

### 2.3.1 Módulo Main

Este módulo tiene la interfaz principal para el acceso a los distintos Módulo del sistema.

Parámetros

- opción\_principal      type: String
- carga\_bd              type: CargaBD

Funciones y procedimientos

- Menu                      ()

Salidas

- Alineamientos          type: Módulo
- AlineamientosMSA      type: Módulo
- ModelosHMM            type: Módulo
- VariantesProteinas      type: Módulo
- VariantesGenomas        type: Módulo

### 2.3.2 Módulo CargaBD

Este módulo no tiene una interfaz interactiva, pero es la que se encarga de gestionar la base de datos y en caso de ser necesario conectarse al servidor online de NCBI y descargar proteínas y genomas en formato fasta.

Parámetros

- genomas                  type: Diccionario
- unspecified              type: Diccionario
- high\_risk                type: Diccionario

- low\_risk type: Diccionario
- proteinas type: Diccionario
- archivos\_agrupados type: Lista

#### Funciones y procedimientos

- DescargarProteinas ()
- DescargarGenomas ()
- LeerInicio ()
- LeerListaClasificacion ()
- LeerListaProteinas ()
- LeerListaGenomas ()

#### Salidas

- genomas type: Diccionario
- unspecified type: Diccionario
- high\_risk type: Diccionario
- low\_risk type: Diccionario
- proteinas type: Diccionario
- archivos\_agrupados type: Lista

### 2.3.3 Módulo Utils

Este módulo es similar al anterior, no tiene la interfaz interactiva, pero contiene múltiples métodos que son reutilizados por varios módulos.

#### Parámetros

- np type: Numpy

#### Funciones y procedimientos

- Existe ()
- GraficarMatriz ()
- GraficarBarras ()
- ListarProteinas ()
- ListarGenomas ()
- MenuRango ()

#### Salidas

- Lista de proteínas type: Impresión inline
- Lista de genomas type: Impresión inline
- Gráfica de barras type: Gráfica inline and imagen .jpg
- Lista de archivos existentes type: Lista

### 2.3.4 Módulo alineamientos de a pares para proteínas

Este módulo realiza el alineamiento de a pares de distintas proteínas, requiere tener instalado el programa "blastp" (Ver sección: Requerimientos).

#### Parámetros

- low\_risk, unspecified\_risk, high\_risk type: CargaBD
- proteinas type: CargaBD
- archivos\_agrupados type: Lista
- principal type: Main
- Utils type: Utils

#### Funciones y procedimientos

- Main ()
- Menu ()
- LeeryGrabar ()
- BuscarNombres ()
- Limpiar ()
- AgruparRiesgos ()
- Recorrer ()
- AlinearRiesgos ()
- Graficar ()

#### Salidas

- Archivo alineamientos type: Archivo Blast
- Gráfica de mejor score type: Gráfica inline and imagen .jpg

### 2.3.5 Módulo alineamientos de a múltiples para proteínas

Este módulo realiza alineamientos múltiples de distintas proteínas, requiere tener el ejecutable “ClustalOmega” (Ver sección: Requerimientos).

#### Parámetros

- low\_risk, unspecified\_risk, high\_risk type: CargaBD
- proteinas type: CargaBD
- archivos\_agrupados type: Lista
- archivos\_MSA type: Lista
- genes type: Lista
- principal type: Main
- Utils type: Utils

#### Funciones y procedimientos

- Main ()
- Menu ()
- CorrerClustalOmega ()
- LeerMSA ()
- ConstruirArboles ()

#### Salidas

- Archivo de alineamientos type: Archivo Clustal Omega
- Gráficas de árboles Filogenéticos type: Gráfica inline and imagen .jpg

### 2.3.6 Módulo modelos HMM

Este módulo realiza un análisis de proteínas con mayores similitudes a modelos HMM, el mismo requiere tener los resultados de los modelos en formato .out en la carpeta “BD/HMM/”, puede utilizar el pipeline biog5, para obtenerlos (Ver sección: Requerimientos)

#### Parámetros

- CargaBD type: CargaBD
- principal type: Main
- Utils type: Utils

- archivos\_resultados type: archivos HAMMER

#### Funciones y procedimientos

- Main ()
- Menu ()
- Leer ()
- Recorrer ()
- Comparar ()
- Analizar ()

#### Salidas

- Comparaciones de mayor score por familia type: Inline impresión

### 2.3.7 Módulo variantes en proteínas

Este módulo realiza la búsqueda de variantes y aminoácidos conservados en proteínas seleccionadas de distintas cepas, requiere tener el ejecutable “ClustalOmega”.

#### Parámetros

- low\_risk, unspecified\_risk, high\_risk type: CargaBD
- proteinas type: CargaBD
- archivos\_agrupados type: Lista
- archivos\_MSA type: Lista

#### Funciones y procedimientos

- Main ()
- Menu ()
- ClasificaciónRiesgos ()
- UnirArchivos ()
- QueRiesgoEs ()
- QueClaseEs ()
- CorrerClustalOmega ()
- LeerMSA ()
- CargaPorcentajesConservación ()
- GraficarBarras ()

#### Salidas

- Conservados type: Inline impresion
- Graficar conservados type: Gráfica inline and imagen .jpg

### 2.3.8 Módulo variantes en genomas

Este módulo realiza la búsqueda de variantes y aminoácidos conservados en cepas, requiere tener el ejecutable “ClustalOmega”.

Tiene dos funciones:

- Analizar 2 genomas a elección
- Analizar todos los genomas disponibles

#### Parámetros

- low\_risk, unspecified\_risk, high\_risk type: CargaBD
- genomas type: CargaBD
- archivos\_agrupados type: Lista
- archivos\_MSA type: Lista

#### Funciones y procedimientos

- Main ()
- Menu ()
- AlinearGenomasDeApareas ()
- AgruparArchivos ()
- LeerMSA ()
- CorrerClustalOmega ()
- CargaPorcentajes ()
- Graficar ()

Salidas

- Conservados type: Inline impresion
- Gráficas conservados type: Gráfica inline and imagen .jpg

## 2.4 Diagrama de caso de uso

Los casos de uso son los óvalos y las figuras con forma "humana" son los actores.

Casos de uso del análisis de requerimientos: Este diagrama especifica el comportamiento funcional, abstrayéndose solamente al sistema a desarrollar en sí y no a su entorno y correlaciones con otros que ya están implantados en el ente que utiliza el sistema.

En la Fig. 2 se muestra el caso de uso de grano grueso del sistema, donde se detallan cada módulo y su relación, se observa un menú principal la clase 'Main', donde se tiene cinco funcionalidades, un análisis por alineamiento de a pares de proteínas, un análisis por alineamiento múltiples de proteínas, un análisis de cepas usando modelos HMM, un análisis de variantes en proteínas y un análisis de variantes entre genomas.

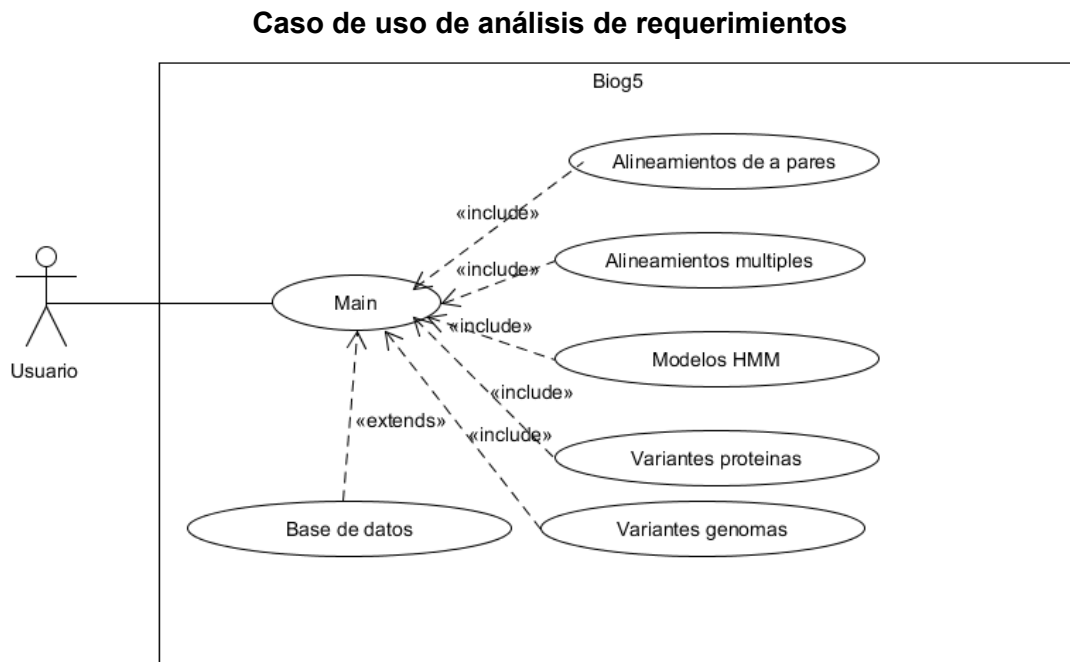


Fig. 2: Diagrama de casos de uso del análisis de requerimientos. Fuente: Elaboración propia.

## 2.5 Diagrama entidad relación

El modelo entidad-relación es una herramienta para generar el modelo de datos que describe la estructura y relaciones de una base de datos. Estos modelos se definen, a través de tablas que contienen un conjunto de columnas que suelen corresponderse con los atributos, así también sus relaciones con el entorno, utilizando atributos específicos, denominados claves que identifican en forma única a cada entidad [7, p 139-142]. En la Fig. 3 se muestra el diagrama de entidad relación del sistema.

### Diagrama entidad relación

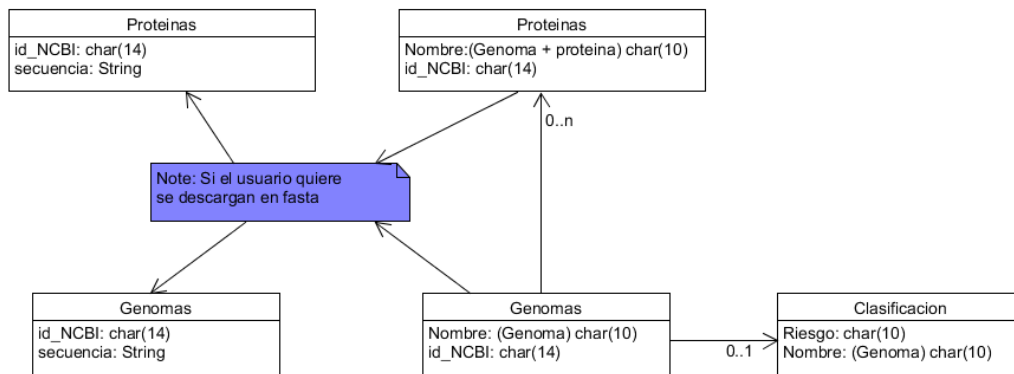


Fig. 3: Diagrama entidad relación (DER). Fuente: Elaboración propia.

## Capítulo 3: Herramientas y lenguajes de programación

En este capítulo se efectúa una breve descripción de las herramientas utilizadas para el análisis y desarrollo del presente proyecto tecnológico.

### 3.1 Lenguajes de programación y herramientas

- **Git:**

Es un software de control de versiones, se puede definir como, la gestión de los diversos cambios que se realizan sobre los elementos de algún producto o una configuración del mismo, es lo que se hace al momento de trabajar con un proyecto colaborativo de desarrollo.

Al trabajar en equipo resulta indispensable, tener una herramienta, que permita integrar, adecuada y controladamente las distintas funcionalidades desarrolladas por los integrantes del equipo. Por su facilidad y robustez se adoptó Git para gestionar y clonar el repositorio del proyecto.

- **Pip:**

Es un sistema de gestión de paquetes utilizado para instalar y administrar paquetes de software escritos en Python.

Se recomienda que use las versiones de 64 bits de estas herramientas.

- **Anaconda:**

Es un kit de herramientas de ciencia de datos. Con más de 25 millones de usuarios en todo el mundo, la edición individual de código abierto (distribución) es la forma más fácil de realizar ciencia de datos Python/R y aprendizaje automático en una sola máquina. Desarrollado para profesionales independientes, es el conjunto de herramientas que lo equipa para trabajar con miles de paquetes y bibliotecas de código abierto [9].

- **Python:**

Es un lenguaje de programación, interpretado o de script el cual se ejecuta utilizando un programa intermedio llamado intérprete. Es muy utilizado en el ambiente científico, por la gran cantidad de librerías especializadas para el cálculo científico, orientado a objetos y multiplataforma, código abierto y gratuito [10].

Se utilizó python 3.5 (Recomendamos usar el intérprete de Python que viene con Anaconda). Debido a que sus características las hacen adecuadas para el propósito planteado en el proyecto.

#### 3.1.1 Requerimientos e Instalaciones:

**Python:** Descargar e instalar el kit Anaconda para Python 3.5 para su correspondiente sistema operativo

Link: <https://www.anaconda.com/products/individual>

- **Paquetes:**

A continuación, se detallan los principales paquetes utilizado por python y su instalación.

conda install biopython

conda install scipy numpy pandas matplotlib

Or

pip2 install scipy numpy pandas matplotlib *Pip install matplotlib*

- **Blast:**

Descargar e instalar el programa Blast para Windows

Link: <https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/2.10.0/>

En caso de ser necesario configurar el variable de entorno del sistema operativo con el siguiente dato:

BLASTDB\_LMDB\_MAP\_SIZE=1000000

- **Pycharm:**

Opcional, le brinda un ambiente más dinámico de desarrollo.

Descargar e instalar, recomendada la versión comunitaria:

Link: <https://www.jetbrains.com/es-es/pycharm/>

### 3.2 Servidores y entornos de desarrollo

- **Spyder:**

Es un entorno científico gratuito y de código abierto escrito en Python, para Python y diseñado por y para científicos, ingenieros y analistas de datos. Cuenta con una combinación única de la funcionalidad avanzada de edición, análisis, depuración y creación de perfiles de una herramienta de desarrollo integral con la exploración de datos, la ejecución interactiva, la inspección profunda y las hermosas capacidades de visualización de un paquete científico. [11]

- **Pychan:**

Es un entorno de desarrollo diseñado por programadores y para programadores, con el fin de proporcionarle todas las herramientas que necesita para un desarrollo de Python productivo. Proporciona una finalización del código inteligente, inspecciones del código, indicación de errores sobre la marcha y arreglos rápidos, así como refactorización de código automática y completas funcionalidades de navegación. [12]

- **UMLet:**

Es una herramienta UML (lenguaje de modelado unificado) gratuita y de código abierto con una interfaz de usuario sencilla para dibujar diagramas UML rápidamente, generar diagramas de secuencia y actividad, exportar diagramas a eps, pdf, jpg, svg y portapapeles, compartir diagramas usando Eclipse y crear nuevos elementos UML personalizados UMLet se ejecuta de manera independiente o como un complemento de Eclipse en Windows, OS X y Linux. [13]

- **Repositorio GitHub:**

Es un lugar de almacenamiento de los archivos del proyecto, como ser los códigos fuentes, documentación, librerías necesarias y otros tipos de datos relacionados al desarrollo del producto.

Sobre el repositorio usualmente se usan programas para control de versiones los cuales son un grupo de aplicaciones originalmente ideadas para gestionar ágilmente los cambios en el código fuente de los programas y poder revertirlos.

Existen varias herramientas para gestionar los cambios en los repositorios, algunas de ellas corren en la nube y brindan conjuntamente otros servicios orientados a equipos de desarrolladores, que ayudan a agilizar, organizar y simplificar el proceso de creación de software.



En el transcurso del proyecto se utilizó un repositorio de software en la nube, provista por la plataforma libre **Github**, especializada en la gestión de proyectos de desarrollo de software:

Link del proyecto Biog5: <https://github.com/biog5/HPV-master>

### 3.3 Estándares de codificación

La siguiente sección establece los estándares de codificación a ser usados durante el desarrollo de User Stories y Bug Fixes.

#### Coding Style

Python: Se usó la guía de estilo PEP-8. Se utilizó la herramienta autopep8 [14], para validar el cumplimiento de la norma.

Las opciones usadas con autopep son:

```
pep8 src/python/ --statistics --max-line-length=190 --show-source --exclude=matplotlib, numpy
```

Para utilizar esta herramienta, se puede descargar e instalar desde su sitio oficial, luego aplicar los comandos anteriores por terminal o usar de manera más sencilla a través de la interfaz gráfica del editor Pycharm, con la opción “formateo de código”.

## Capítulo 4: Resultados

En acorde al back log propuesto para el primer reléase en esta sección se presentan los distintos módulos, incluyendo diagramas descriptivos de los procesos intervinientes en los mismos y las interfaces asociadas a cada uno.

### 4.1 Módulo principal

El módulo Main es la primera interfaz que visualiza el usuario, y le permite elegir los distintos análisis sobre genomas y proteínas de distintas cepas HPV.

Para correr el sistema debe abrir en archivo Main.py desde la interfaz del IDE Spyder u Pycharm luego en el menú de estas herramientas clicar la opción “Run” o “Correr” o usar el icono con el símbolo “▶”, también es válida desde una consola de comandos de Windows estando en la carpeta principal del proyecto puede correr el sistema ejecutando la instrucción: “*python Main.py*”.

En la Fig. 4 se muestra el menú principal del sistema Biog5.

**Menú Principal BioG5**

```
#####
##### Bienvenidos a BIOG5 #####
#####

Que desea realizar:
1- Alineamientos de a pares usando Blast
2- Alineamientos multiples usando Clustal Omega
3- Modelos HMM
4- Analisis de variantes en proteinas
5- Analisis de variantes en genomas
6- Salir
Ingrese una opción:
```

Fig. 4: Se muestran los distintos análisis posibles sobre los datos disponibles. Fuente: Elaboración propia.

### 4.2 Análisis de alineamientos de a pares

La primera etapa de este trabajo se enfocó en un análisis sobre semejanzas de la clase **proteína a proteína**. Se toma cada cepa de bajo riesgo y riesgo no determinado las secuencias de ADN correspondiente a las proteínas E1, E2, E7, L1, L2, para compararlas con la secuencia de la misma proteína en cepas de alto riesgo.

En primera instancia el Módulo cuenta con un menú de opciones que se despliegan al momento de su ejecución. En la Fig. 5 se muestra la lista de opciones disponibles para el módulo Alineamiento de a pares.

### Menú Alineamiento de a pares

```
### Modulo: Alineamiento de a pares ###

Que desea realizar:
1- Listar Proteínas almacenadas
2- Listar Genomas almacenadas
3- Realizar alineamiento de a pares de proteínas con BLAST y analizar
4- Volver al menu principal
5- Salir
Ingrese una opción:
```

Fig. 5: Menú de Módulo Alineamiento de a pares. Fuente: Elaboración propia.

En la figura anterior se observa el menú de opciones, las dos primeras permiten **listar proteínas y genomas, estas son reutilizables en todos los módulos**. En la Fig. 6 se observa la función de listado de proteínas.

### Listar las proteínas almacenadas

```
Ingrese una opción: 1
Proteína número 1 : HPV16E1 - Cepa: HPV16 - ID: NP_041327.2
Proteína número 2 : HPV16E2 - Cepa: HPV16 - ID: NP_041328.1
Proteína número 3 : HPV16E7 - Cepa: HPV16 - ID: NP_041326.1
Proteína número 4 : HPV18E1 - Cepa: HPV18 - ID: NP_040312.1
Proteína número 5 : HPV18E2 - Cepa: HPV18 - ID: NP_040313.1
Proteína número 6 : HPV18E7 - Cepa: HPV18 - ID: NP_040311.1
Proteína número 7 : HPV45E1 - Cepa: HPV45 - ID: CAA52575.1
Proteína número 8 : HPV45E2 - Cepa: HPV45 - ID: CAA52576.1
Proteína número 9 : HPV45E7 - Cepa: HPV45 - ID: CAA52574.1
Proteína número 10 : HPV31E1 - Cepa: HPV31 - ID: AAA46952
Proteína número 11 : HPV31E2 - Cepa: HPV31 - ID: AAA46953.1
Proteína número 12 : HPV31E7 - Cepa: HPV31 - ID: AAA46951.1
Proteína número 13 : HPV33E1 - Cepa: HPV33 - ID: AAA46960.1
Proteína número 14 : HPV33E2 - Cepa: HPV33 - ID: AAA46961.1
Proteína número 15 : HPV33E7 - Cepa: HPV33 - ID: AAA46959.1
Proteína número 16 : HPV35HE1 - Cepa: HPV35H - ID: CAA52563.1
Proteína número 17 : HPV35HE2 - Cepa: HPV35H - ID: CAA52564.1
Proteína número 18 : HPV35HE7 - Cepa: HPV35H - ID: CAA52562.1
```

Fig. 6: Lista de proteínas, con sus correspondientes cepas asociadas y el ID publica en la Base de Datos del NCBI. Fuente: Elaboración propia.

La segunda función reutilizable permite visualizar el listado de los genomas almacenados. En la Fig. 7 se muestra la salida de la función.

### Listar los genomas almacenados

```
Ingrese una opción: 2
Genoma número 1 : HPV16
Genoma número 2 : HPV18
Genoma número 3 : HPV45
Genoma número 4 : HPV31
Genoma número 5 : HPV33
Genoma número 6 : HPV35H
Genoma número 7 : HPV39
Genoma número 8 : HPV51
Genoma número 9 : HPV52
Genoma número 10 : HPV56
Genoma número 11 : HPV58
Genoma número 12 : HPV59
Genoma número 13 : HPV68A
Genoma número 14 : HPV68B
Genoma número 15 : HPV73
Genoma número 16 : HPV82
Genoma número 17 : HPV23
Genoma número 18 : HPV53
```

Fig. 7: Lista de genomas disponibles. Fuente: Elaboración propia.

La opción número tres ejecuta el algoritmo que contiene el programa de alineamiento Blast. La primera etapa del algoritmo implica el agrupamiento de las distintas proteínas en función de sus cinco clases, y clasificación en alguno de los tres riesgos dados. En la Fig. 8 se observa un diagrama de objetos con los grupos y sus relaciones correspondientes.

## Agrupamientos

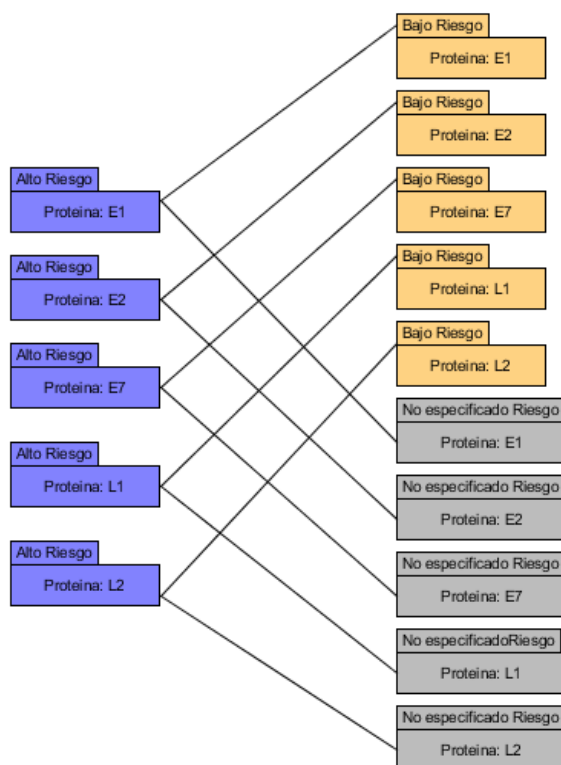


Fig. 8: Primera parte de la ejecución del algoritmo BlastP. Fuente: Elaboración propia.

Luego de tener la clasificación proteínas (5) – grupos (3) el siguiente proceso es comparar el grupo de alto riesgo contra el grupo de bajo riesgo y el grupo no especificado riesgo, y dentro de los mismos de proteína a proteína. En la Fig. 9 se muestra un ejemplo de cepas semejantes encontradas.

## Alineamientos Blast

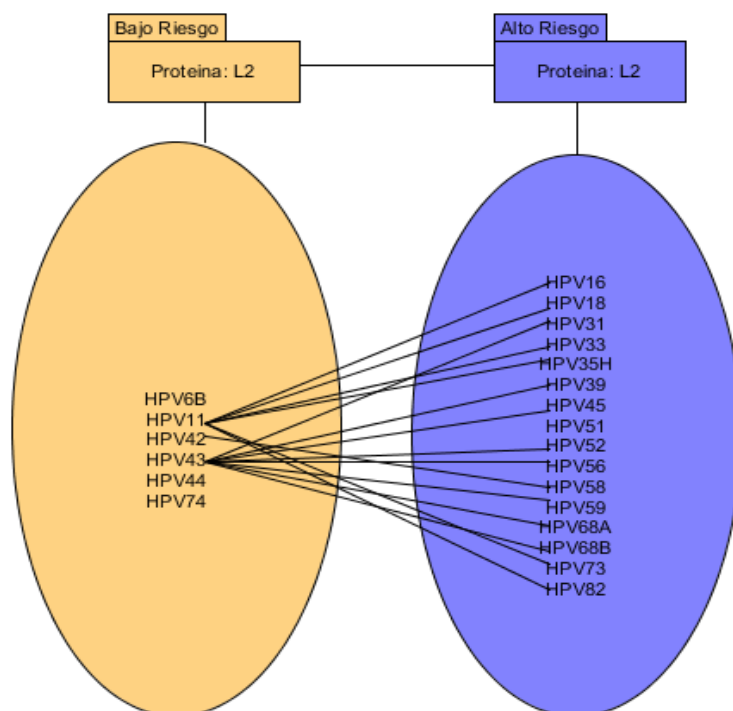


Fig. 9: Gráfica de resultados del alineamiento proteína L2 Fuente: Elaboración propia.

## Resultados Blasp – Bajo Riesgo

```
Relaciones evolutivas con mejores scores encontradas para la proteína L2
Cepa Bajo Riesgo  _____ Cepa Alto Riesgo
HPV11
| _____ HPV16
| _____ HPV18
| _____ HPV33
| _____ HPV35H
| _____ HPV73
| _____ HPV82
HPV43
| _____ HPV31
| _____ HPV39
| _____ HPV45
| _____ HPV52
| _____ HPV56
| _____ HPV59
| _____ HPV68A
| _____ HPV68B
HPV42
| _____ HPV58
```

Fig. 10: Comparación con la Cepa de bajo riesgo proteína L2. Fuente: Elaboración propia.

En la figura anterior se observa resultados alineamiento realizado sobre una cepa y proteína bajo riesgo contra cada cepa y proteína correspondiente de alto riesgo se tomas las mejores coincidencias.

### Pregunta Biológica

Se presenta el siguiente caso, donde se toma como ejemplo el resultado anterior. Teóricamente dada la proteína target L2 para que la vacuna de la Cepa 16 tenga un efecto similar en la Cepa 11 es necesaria una similitud en las proteínas L2. ¿Existe la similitud entre estas proteínas? Para responder el módulo utiliza programa Blast que brindar resultados que demuestran la similitud buscada.

Es necesaria esta comprobación, pero no es suficiente para probar igual efecto de la vacuna para ello se requerida un análisis más profundo.

De esta forma usando las coincidencias podemos obtener potenciales blancos de estudios las no semejanzas detectadas pueden servir para excluir blancos no relevantes.

### 4.3 Análisis de alineamientos múltiples MSA

En esta etapa de este trabajo se realizó un análisis sobre **semejanzas dentro de grupos de proteínas y riesgos**, primeramente, se agrupan las cepas de cada riesgo en función de sus proteínas E1, E2, E7, L1, L2, en segundo lugar, se realiza un alineamiento múltiple de cada una de la lista de secuencias agrupadas y se obtiene archivos de la clase, con el formato “riesgoProteina\_ MSA.phylip”. ejemplo: el archivo “high\_riskE1\_ MSA.phylip”, que contiene el alineamiento múltiple para las cepas de alto riesgo de la proteína E1. Con los archivos alineados se procede a generarse arboles filogenéticos para establecer las relaciones entre las distintas cepas.

En la Fig. 10 se ejemplifica los procesos llevados a cabo en cada grupo. Donde se observa un alineamiento múltiple o MSA sobre cada grupo de proteínas y luego la obtención de árboles filogenéticos de cada grupo.

## Procesos Módulo MSA

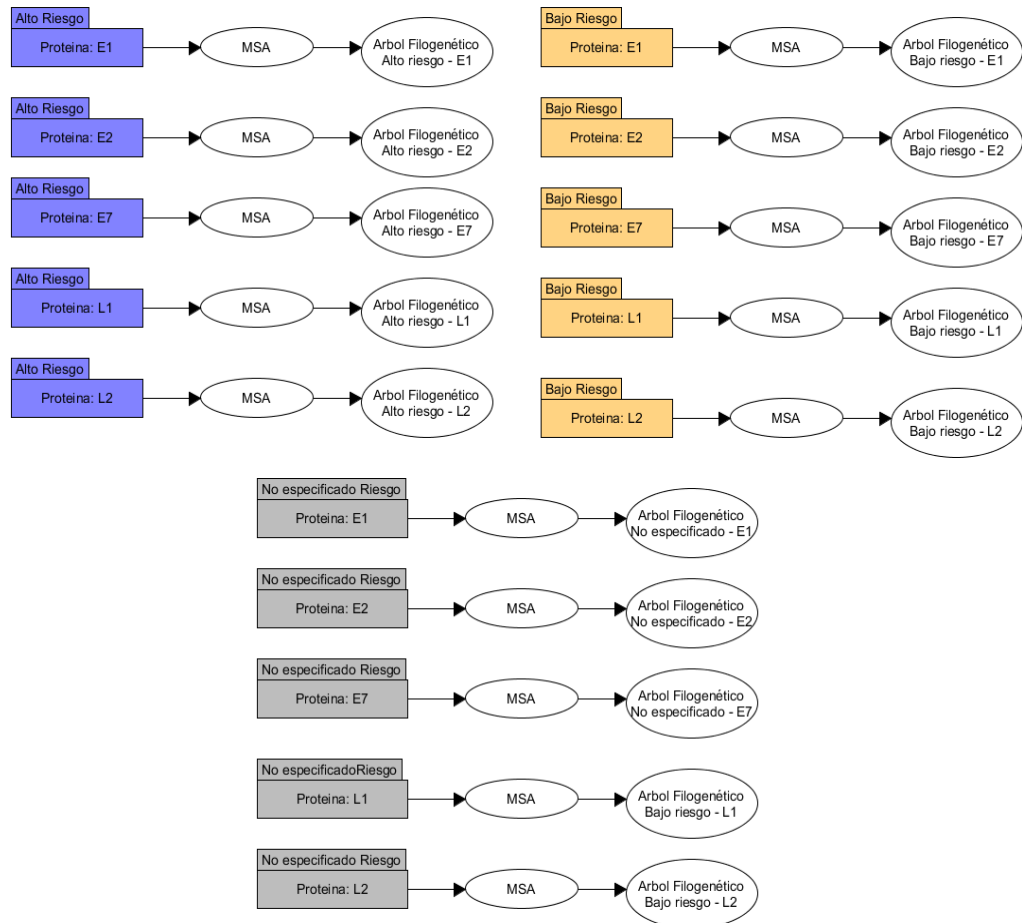


Fig. 10: Diagrama de procesos aplicados a grupos de proteínas. Fuente: Elaboración propia.

En la siguiente figura se observa a uno de los resultados obtenidos:

### Árbol filogenético proteína L2

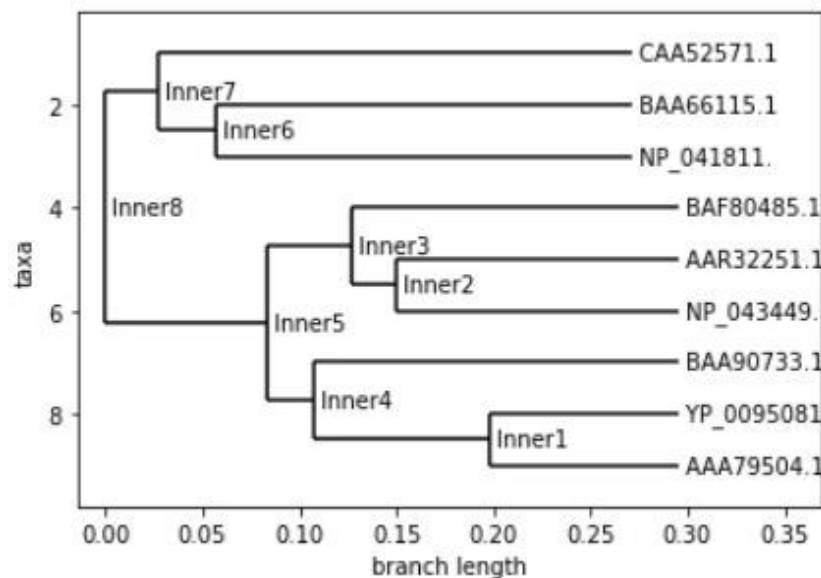


Fig. 11: Árbol filogenético proteína L2 cepa no especificado riesgo. Fuente: Elaboración propia.

## Pregunta Biológica

Encontrada una nueva cepa, clasificada con un determinado riesgo. ¿Es posible conocer las relaciones ancestrales con otras proteínas de las cuales se tienen registros?

Los resultados pueden posibilitar categorizar la proteína y priorizarla o no como posible blanco molecular para otro estudio.

## 4.4 Análisis de cepas usando modelos HMM

En esta sección se enfocará en los resultados del uso de modelos HMM.

Para utilizar esta herramienta se requiere correr pipeline biog5 en un sistema Linux el mismo hará uso de lo programa MAFFT y HMMER para generar semillas y modelos y correr los mismos generando un archivo de salida con resultados que serán analizados en esta herramienta. En la Fig. 12 se muestra el diagrama de caso de uso de pipeline biog5.

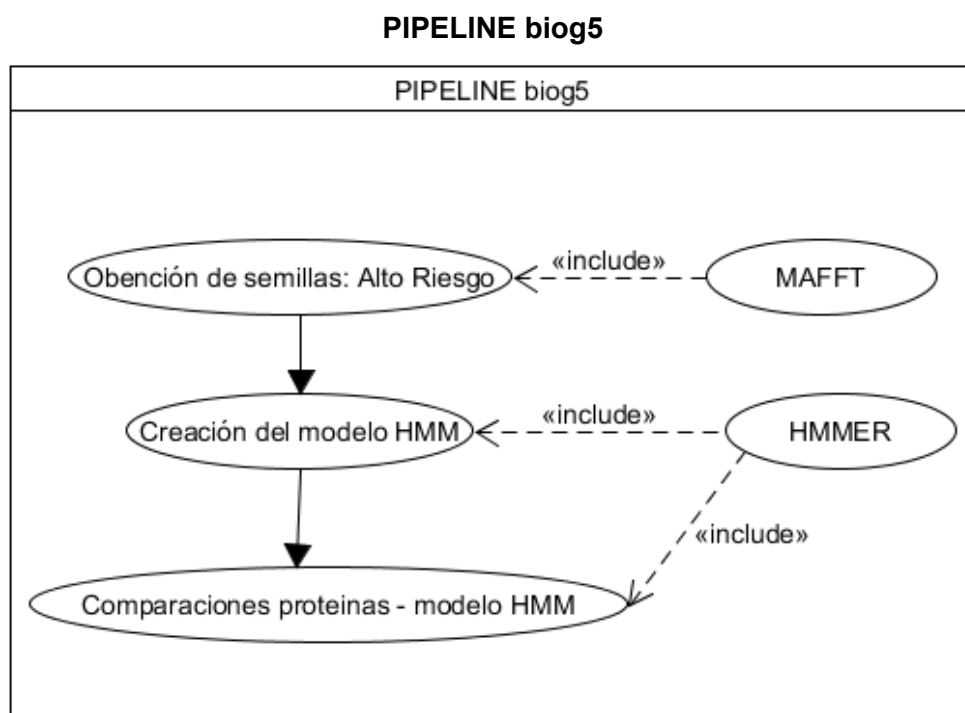


Fig. 12: Procesos aplicados por el pipeline biog5. Fuente: Elaboración propia.

En la siguiente figura se observa de forma detalladas todos los pasos del módulo HMM.



## Procesos Módulo HMM

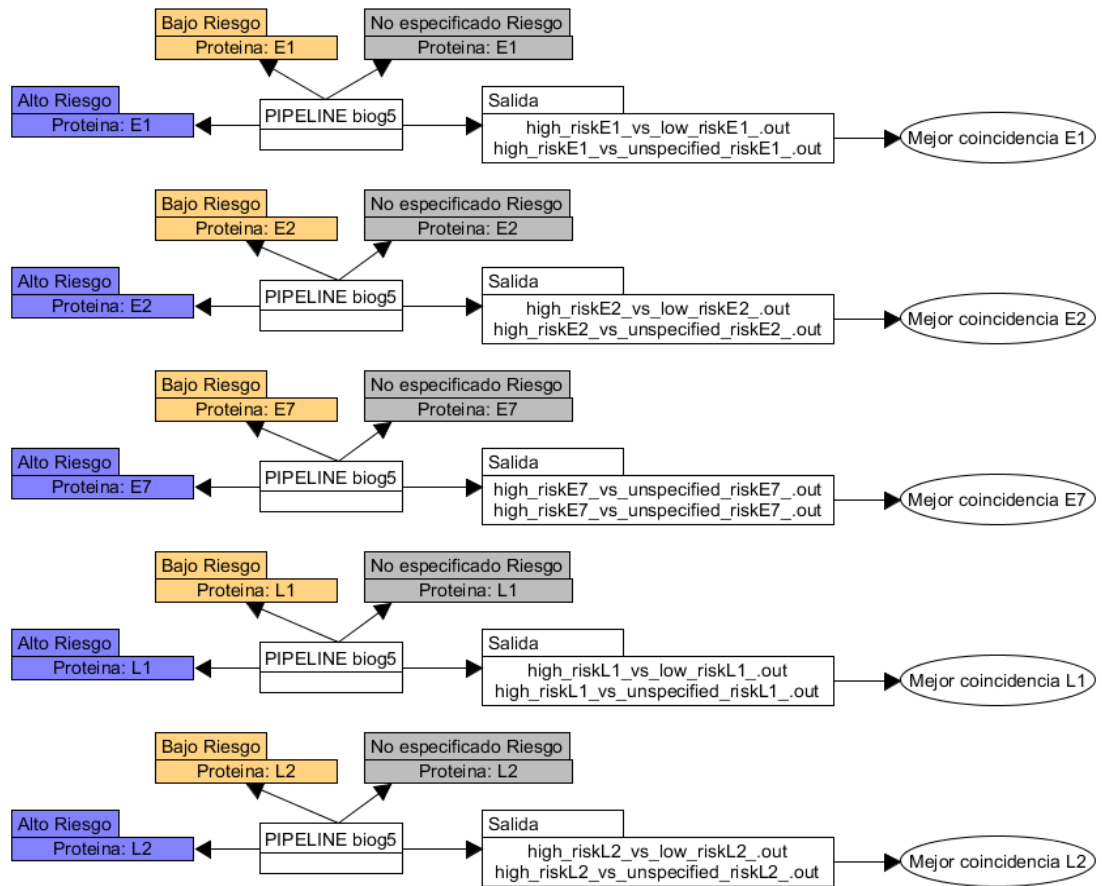


Fig. 13: Procesos aplicados a cada grupo de proteínas. Fuente: Elaboración propia.

En la Fig. 14 se presenta el análisis de resultados de modelos HMM para cepas de alto riesgo.

## Resultados modelo HMM

```
Ingrese una opción: 3
### Recuerde que debe tener los resultados del pipeline big5 en la carpeta /BD/HMM/ ###

1: Se compararan: Proteinas de cepas de alto riesgo vs proteinas de otros riesgos, espere por favor ...
Clasificacion: high Proteina: E1 vs Clasificacion: low Proteina: E1
Clasificacion: high Proteina: E1 vs Clasificacion: unspecified Proteina: E1
Clasificacion: high Proteina: E2 vs Clasificacion: low Proteina: E2
Clasificacion: high Proteina: E2 vs Clasificacion: unspecified Proteina: E2
Clasificacion: high Proteina: E7 vs Clasificacion: low Proteina: E7
Clasificacion: high Proteina: E7 vs Clasificacion: unspecified Proteina: E7
Clasificacion: high Proteina: L1 vs Clasificacion: low Proteina: L1
Clasificacion: high Proteina: L1 vs Clasificacion: unspecified Proteina: L1
Clasificacion: high Proteina: L2 vs Clasificacion: low Proteina: L2
Clasificacion: high Proteina: L2 vs Clasificacion: unspecified Proteina: L2

2: Relaciones evolutivas con mejores scores encontradas por cada modelo:
Modelo HMM de Alto Riesgo proteina E1
|_____ Mejor score - Cepa: HPV68a Riesgo: low Proteina: E1

Modelo HMM de Alto Riesgo proteina E1
|_____ Mejor score - Cepa: HPV68a Riesgo: unspecified Proteina: E1

Modelo HMM de Alto Riesgo proteina E2
|_____ Mejor score - Cepa: HPV68a Riesgo: low Proteina: E2
```

Fig. 14: Salidas de la ejecución del análisis de modelos HMM. Fuente: Elaboración propia.

Se logra observar, primeramente, la aclaración de la ubicación de los archivos resultantes de correr el pipeline biog5, luego se toman los datos y se clasifican, para obtener sus identificadores de cepas y proteínas, y finalmente se evalúa el score obtenido al comparar el grupo de cada proteína (de riesgo bajo y no especificado) con el modelo asociado a la misma, los resultados se ordenan y se toma solo uno, el de mayor puntuación de pertenencia.

### Pregunta Biológica

Encontrada una nueva cepa de interés, o una cepa ya conocida. ¿Es posible, dada una proteína, indistintamente de su riesgo inicial, compararla con un modelo HMM de alto riesgo? El resultado permite para verificar o descartar ciertas similitudes, deducir funciones y reacciones biológicas que son exclusivas de un modelo y así priorizar o clasificar en otros grupos en función a sus características.

## 4.5 Análisis de variantes en proteínas

En este módulo se agrupan las proteínas a través de alineamientos múltiples, para obtener un alineamiento de referencia y analizar sus variaciones y conservaciones. Para ello se agrupan las proteínas por riesgo y clase luego sobre cada una se aplica un algoritmo MSA, que devuelve un alineamiento de referencia sobre el cual se obtiene porcentajes de aparición de cada aminoácido en cada posición, y finalmente se filtran los mismos, basado en el criterio de conservación, esto implica que, dada las secuencias alineadas, una posición puede considerarse como “conservada” si contiene al menos el 90% de las secuencias con el mismo aminoácido en esa ubicación.

En la Fig. 15 se observa un diagrama con etapas del procesamiento de análisis de variantes en genomas.

## Procesos del módulo variantes en Proteínas

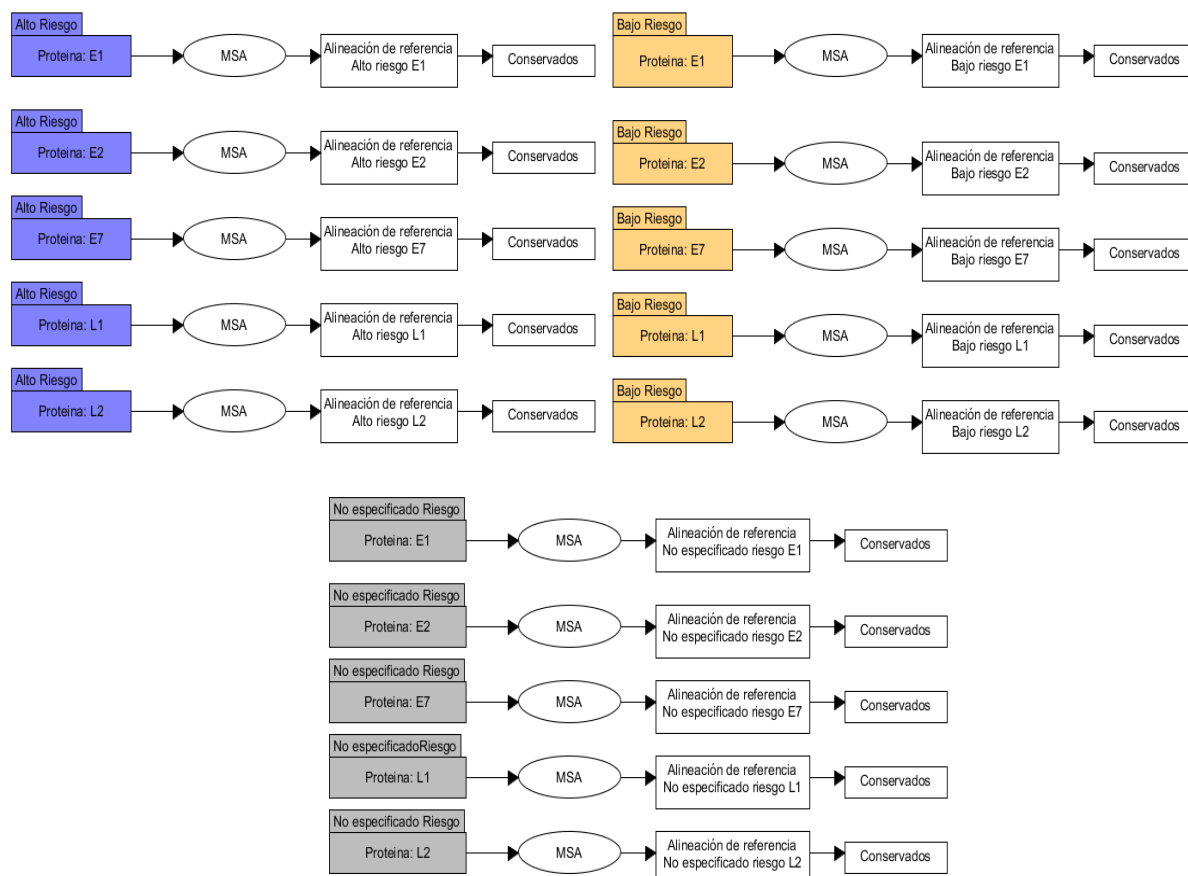


Fig. 15: Detalles los procesos aplicados a grupos de proteínas para el análisis de variantes. Fuente: Elaboración propia.

En la Fig. 16 se muestra la interfaz de usuario corriendo el proceso de agrupación, alineamiento y análisis de posiciones.

## Interfaz de usuario Análisis de variantes en proteínas

```

Ingrese una opción: 3
1: Agrupando datos por tipo de Riesgo y Proteína, espere por favor ...
2: Corriendo alineamiento multiple con Clustal Omega, espere por favor ...
3: Cargando alineamientos, espere por favor ...
4: Cargando porcentajes, espere por favor ...
5: Filtrando porcentajes con >90% de conservación, espere por favor ...
Calculos terminados analizar resultados:

### Modulo: Alineamiento multiples ###

Que desea realizar:
1- Imprimir lista de aminoacidos conservados
2- Graficar Riesgos y Aminoacidos
3- Volver al menu anterior
4- Volver al menu principal
5- Salir
Ingrese una opción:

```

Fig. 16: Detalles los procesos aplicados a cada grupo de proteínas para el análisis de variantes. Fuente: Elaboración propia.

Luego del análisis el módulo le brinda dos posibilidades para ver los resultados:

**1.Impresión en línea:** Se listarán los alineamientos de referencia por tipo de proteína y grupo de riesgo, luego se solicitará que elija uno para ver, y se imprimirá por consola el listado correspondiente a elección con la identificación del riesgo, proteína, posición y aminoácido. En la Fig. 17 se muestra la interfaz de la impresión en línea.

#### Impresión en línea conservados

```
Ingrese una opción: 1
Lista de proteínas agrupadas por riesgo con > 90% de conservación:
Número: 1 [ Alto Riesgo Proteína: E1 ]
Número: 2 [ Alto Riesgo Proteína: E2 ]
Número: 3 [ Alto Riesgo Proteína: E7 ]
Número: 4 [ Alto Riesgo Proteína: L1 ]
Número: 5 [ Alto Riesgo Proteína: L2 ]
Número: 6 [ Bajo Riesgo Proteína: E1 ]
Número: 7 [ Bajo Riesgo Proteína: E2 ]
Número: 8 [ Bajo Riesgo Proteína: E7 ]
Número: 9 [ Bajo Riesgo Proteína: L1 ]
Número: 10 [ Bajo Riesgo Proteína: L2 ]
Número: 11 [ Noespecificado Riesgo Proteína: E1 ]
Número: 12 [ Noespecificado Riesgo Proteína: E2 ]
Número: 13 [ Noespecificado Riesgo Proteína: E7 ]
Número: 14 [ Noespecificado Riesgo Proteína: L1 ]
Número: 15 [ Noespecificado Riesgo Proteína: L2 ]
Elija un número: 1
Alto Riesgo Proteína: E1 Posición: 0 Aminoácido: M
Alto Riesgo Proteína: E1 Posición: 5 Aminoácido: G
Alto Riesgo Proteína: E1 Posición: 6 Aminoácido: T
Alto Riesgo Proteína: E1 Posición: 13 Aminoácido: G
Alto Riesgo Proteína: E1 Posición: 14 Aminoácido: C
Alto Riesgo Proteína: E1 Posición: 16 Aminoácido: G
Alto Riesgo Proteína: E1 Posición: 17 Aminoácido: W
```

Fig. 17: Detalles de aminoácidos conservados para una proteína y riesgo determinado. Fuente: Elaboración propia.

**2. Impresión gráfica:** Se listarán los alineamientos de referencia por tipo de proteína y grupo de riesgo, luego se solicitará que elija uno para graficar, a su vez aparecerá la longitud de la secuencia, como puede ser muy extensa, aparecerá un nuevo menú con dos opciones, la primera graficar la totalidad de la secuencia o la segunda un intervalo dentro de la misma.

En la Fig. 18 se observa el menú con las anteriores opciones, y una salida gráfica de los resultados.

## Gráfica de variantes en proteínas

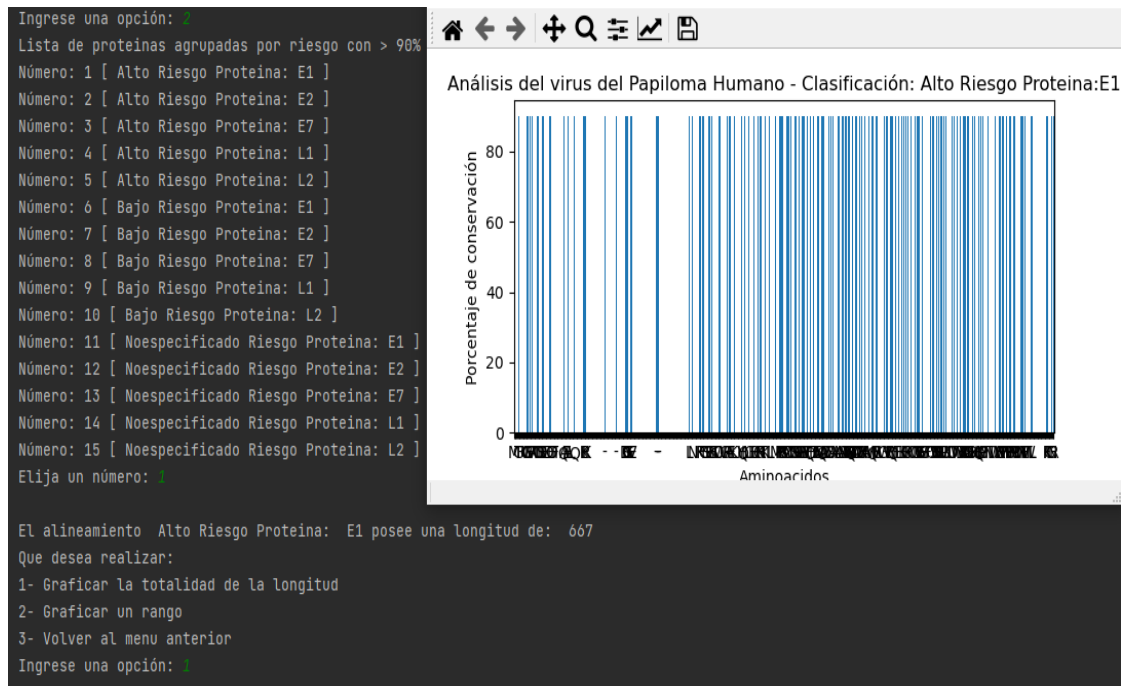


Fig. 18: Impresión con gráfica de barras del alineamiento de referencia. Fuente: Elaboración propia.

En la figura anterior se observa la gráfica de barras correspondiente al alineamiento de referencia de la proteína de alto riesgo E1, en este caso graficando la totalidad de la secuencia.

### Pregunta Biológica

¿Es posible detectar regiones de conservación comunes y asociarlas a un grupo de riesgo y proteína? El análisis permite a través del análisis de porcentajes de conservación definir posibles regiones de conservación como así también establecer como potenciales regiones donde podría llegar a existir patrones asociados a una proteína y riesgo.

## 4.6 Análisis de variantes en genomas

En este módulo se agrupan genomas a través de alineamientos múltiples, para obtener un alineamiento de referencia y analizar sus variaciones y conservaciones.

Se aplica un algoritmo MSA, solo dos o más genomas, luego sobre el alineamiento de referencia obtenido se calculan porcentajes de aparición de cada aminoácido en cada posición, y finalmente se filtran los mismos, basados en una conservación de la posición del 90%.

Este módulo tiene dos algoritmos:

**1. Análisis de 2 genomas:** Realiza un alineamiento a la par con dos genomas a elección y luego informa los resultados con gráficas de matrices.

**2. Análisis de múltiples genomas (Beta):** Realiza un alineamiento múltiple de genomas conocidos y obtienen los resultados de secciones que mejor alinean a los cuales se le puede asociar a un patrón. La principal desventaja de este algoritmo es que requiere un gran poder de cálculo dado el tamaño de los genomas y la cantidad de genomas, se ha probado con distintos números de genomas y dan resultados correctos, pero al incrementar o directamente tomar la totalidad de los genomas se requiere un gran poder de cómputo para completarlo.

En la Fig. 19 se observa el diagrama de procesos, para el análisis de variantes en genomas.

### Procesos análisis de variantes genomas

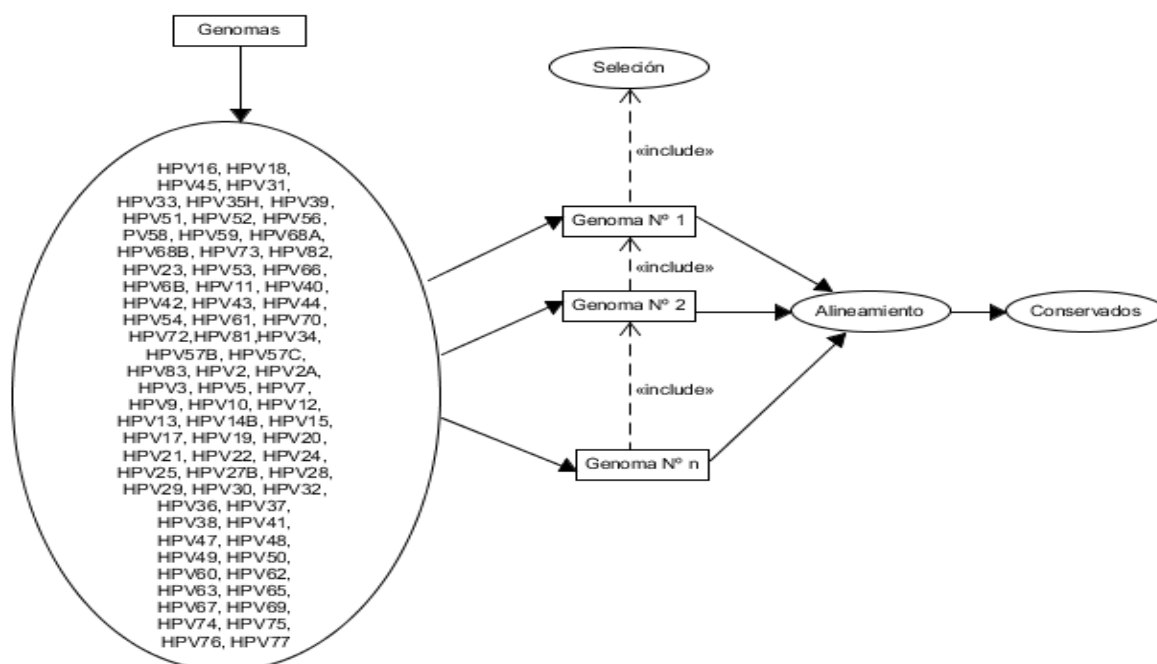


Fig. 18: Detalles de los procesos llevados a cabo para el análisis de variantes en genomas. Fuente: Elaboración propia.

En la Fig. 19 se observa la Interfaz de usuario Módulo variantes genomas análisis de 2 genomas.

### Análisis de dos genomas

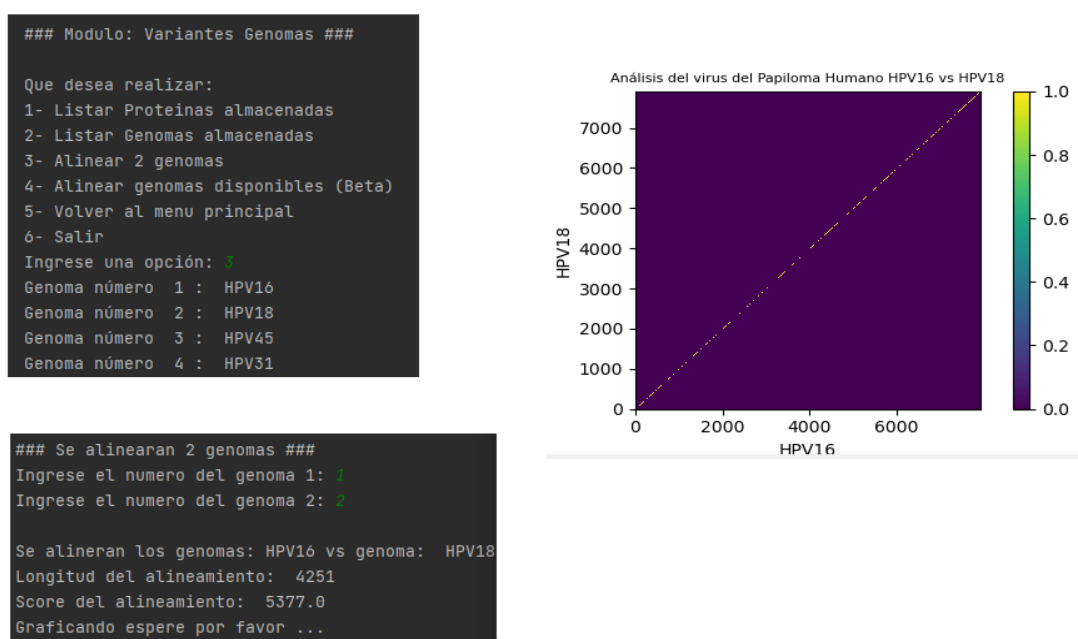


Fig. 19 Impresión con gráfica de barras del alineamiento de referencia. Fuente: Elaboración propia.

En la figura anterior se observa el menú para el análisis de dos genomas en el caso ejemplo se utiliza el N.º 1: HPV16 y el N.º 2: HPV18, a su lado la gráfica matricial resultante donde se destacan, las zonas de coincidencia de aminoácidos en todo el genoma.

### **Pregunta Biológica**

¿Es posible detectar regiones de conservación comunes a dos genomas distintos conocidos o nuevos? El análisis permite a través del cálculo de porcentajes de conservación, definir posibles regiones de conservación. Similar al realizado en proteínas, pero con una vista general a todo el genoma observándose de esta forma si es algo específico o en si ambos genomas o grupos de genomas poseen similitudes o diferencias destacables.

## Capítulo 5: Conclusión

El sistema Biog5 en su primera versión proporciona al usuario una herramienta eficaz, ágil y sencilla para distintos análisis de genomas y proteínas del virus del papiloma humano.

El sistema fue implementado en base a las buenas prácticas de análisis e implementación, con las funcionalidades documentadas para que las futuras modificaciones puedan ser hechas de la mejor manera posible, sin la demanda de mucho tiempo o esfuerzo para el entendimiento de lo que fue desarrollado, lo que permite que no solo su desarrollador inicial pueda hacer cambios.

Una fortaleza que posee el sistema es la aplicación de distintas herramientas destinadas a encontrar diversas correlaciones en blancos a estudio, con datos específicos de los mismos.

Una debilidad surge que esta versión, no tiene interfaces avanzadas para la configuración de parámetros o gestión de base de datos, o algoritmos predictivos.

Las siguientes versiones, aparte de mejoras estructurales se planea buscar e incluir recomendaciones de personas especializadas en el tema clínico y biológico que nos orienten a que resultados podrían ser más relevantes para estudios avanzados.

Se consideran como las líneas futuras de trabajo las siguientes:

- Crear módulos para la gestión de base de datos.
- Agregar funciones para parametrizar las distintas herramientas.
- Agregar mayor usabilidad y portabilidad a las interfaces.
- Agregar más opciones para graficar.
- Generar informes detallados.
- Desarrollar un modelo de machine learning para asociar a familias de proteínas y riesgos.
- Desarrollar un modelo de machine learning para identificar proteínas en genomas.
- Agregar parámetro en cada cepa para clasificar en función de taxonomías.



## Referencias bibliográficas

- [1] Who, Nota: Detalles Papilomavirus. fecha visitada (1, 12, 2021), [Online]. Disponible: [https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-\(hpv\)-and-cervical-cancer](https://www.who.int/news-room/fact-sheets/detail/human-papillomavirus-(hpv)-and-cervical-cancer),
- [2] NCBI, Artículo PMC. fecha visitada (1, 12, 2021), [Online]. Disponible: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC145302/>.
- [3] NCBI, fecha visitada (1, 12, 2021), [Online]. Disponible: <https://www.ncbi.nlm.nih.gov/>
- [4] NCBI, BLAST, fecha visitada (1, 12, 2021), [Online]. Disponible: <https://blast.ncbi.nlm.nih.gov>
- [5] NCBI, Clustal Omega, fecha visitada (1, 12, 2021), [Online]. Disponible: <https://www.ebi.ac.uk/Tools/msa/clustalo/>
- [6] HMMER, fecha visitada (1, 3, 2018), [Online]. Disponible: <http://hmmer.org/>
- [7] R. Pressman, "Ingeniería del Software. Un enfoque práctico", 7ma. Edición. ISBN: 978-607-15-0314-5. MC GRAW HILL 2008.
- [8] I. Sommerville, "Ingeniería del Software", 7ma. Edición, Barcelona, España Pearson Addison Wesley 2008.
- [9] Anaconda, fecha visitada (1, 12, 2021), [Online]. Disponible: <https://www.anaconda.com/products/individual>
- [10] Tutoriales Python, fecha visitada (8, 12, 2021), [Online]. Disponible: <https://docs.python.org/2/tutorial>
- [11] Spyder fecha visitada (12, 2, 2021), [Online]. Disponible: <https://www.spyder-ide.org/>
- [12] Pycharm, fecha visitada (12, 2, 2021), [Online]. Disponible: <https://www.jetbrains.com/es-es/pycharm/>
- [13] Umlet, fecha visitada (12, 2, 2021), [Online]. Disponible: <https://www.umlet.com/>
- [14] Automatically formats Python, fecha visitada (6, 11, 2021), [Online]. Disponible: <https://pypi.org/project/autopep8/>