



Pontificia Universidad
Católica del Ecuador

FACULTAD DE INGENIERIA

MAESTRIA EN BIOLOGIA COMPUTACIONAL

**ANÁLISIS METAGENÓMICO DE VIRUS DE PLANTAS PRESENTES EN
MUESTRAS DE HECES DE PRIMATES DEL ECUADOR**

**TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE MÁSTER EN
BIOLOGÍA COMPUTACIONAL**

AUTORA: ING. JOSELIN CAROLINA CAIZA GUAMBA

DIRECTOR: ING. FRANCISCO FLORES PH.D.

QUITO, 2024

CONTENIDO

| | |
|---|----|
| INDICE DE TABLAS | 3 |
| INDICE DE FIGURAS | 4 |
| DEDICATORIA | 5 |
| AGRADECIMIENTOS | 6 |
| RESUMEN | 7 |
| 1. INTRODUCCION | 8 |
| 2. PLANTEAMIENTO DEL PROBLEMA | 10 |
| 3. OBJETIVOS | 11 |
| 3.1 Objetivo general | 11 |
| 3.2 Objetivos específicos | 11 |
| 4. METODOLOGIA | 11 |
| 4.1 Datos metagenómicos | 11 |
| 4.2 Instrumentos y herramientas para el procesamiento de datos metagenómicos | 12 |
| 4.2.1 Selección del pipeline bioinformático | 12 |
| 4.2.2 Procesamiento de los datos metagenómicos | 14 |
| 4.2.2.1 Control de calidad | 14 |
| 4.2.2.2 Limpieza de secuencias | 15 |
| 4.2.2.3 Concatenación | 16 |
| 4.2.2.4 Asignación taxonómica | 16 |
| 4.2.2.5 Análisis taxonómico | 18 |
| 4.2.2.6 Ensamblaje | 18 |
| 5. RESULTADOS | 20 |
| 5.1 Selección del pipeline bioinformático | 20 |
| 5.2 Análisis Taxonómico | 21 |
| 5.3 Abundancia y distribución de los virus de plantas en todas las muestras | 32 |
| 5.4 Confirmación de virus de plantas presentes en cada muestra | 33 |
| 6. DISCUSION | 35 |
| 7. CONCLUSIONES | 39 |
| 8. RECOMENDACIONES | 40 |
| 9. REFERENCIAS BIBLIOGRAFICAS | 40 |

INDICE DE TABLAS

| | |
|---|----|
| Tabla 1. Datos metagenómicos | 11 |
| Tabla 2. Tiempo de ejecución del pipeline bioinformático sin hospedero y con hospedero | 20 |
| Tabla 3. Número de lecturas obtenidas con el pipeline bioinformático sin hospedero y con hospedero | 21 |
| Tabla 4. Resultados confirmatorios obtenidos del ensamblaje de novo | 34 |

INDICE DE FIGURAS

| | |
|---|----|
| Figura 1. Pipeline bioinformático sin el hospedero | 13 |
| Figura 2. Pipeline bioinformático con el hospedero | 14 |
| Figura 3. Asignación taxonómica muestra 449 | 22 |
| Figura 4. Asignación taxonómica muestra 452 | 23 |
| Figura 5. Asignación taxonómica muestra 455 | 24 |
| Figura 6. Asignación taxonómica muestra 456 | 25 |
| Figura 7. Asignación taxonómica muestra 459 | 26 |
| Figura 8. Asignación taxonómica muestra 460 | 27 |
| Figura 9. Asignación taxonómica muestra 461 | 28 |
| Figura 10. Asignación taxonómica muestra 462 | 29 |
| Figura 11. Asignación taxonómica muestra 463 | 30 |
| Figura 12. Asignación taxonómica muestra 464 | 31 |
| Figura 13. Asignación taxonómica muestra 469 | 32 |
| Figura 14. Comparación de abundancia y distribución de los virus de plantas..... | 33 |

DEDICATORIA

A mis queridos padres: José y Rocío

Mi gran hermano: Henry

Mis pequeños: Willy, Keyla, Leona y Osita

AGRADECIMIENTOS

Deseo expresar mi más sincero agradecimiento a todas las personas e instituciones que han contribuido de alguna manera con las herramientas para la realización de esta tesis.

En primer lugar, quiero agradecer a mi director de tesis, Ing. Francisco Flores PhD, por su apoyo incondicional y asesoría a lo largo de este proyecto. Su experiencia y orientación han sido elementales para la culminación de este trabajo.

Agradezco a Sara Martin PhD por facilitar los recursos de secuenciación y datos necesarios para llevar a cabo esta investigación.

A mi familia y Paúl, por su amor y paciencia. A mis amigos, por su constante ánimo y apoyo, Sebastián, con quien construí una excelente amistad durante la maestría.

A todos ustedes, mi más profundo agradecimiento.

RESUMEN

La secuenciación de nueva generación ha revolucionado la detección y el descubrimiento de virus, permitiendo la caracterización no dirigida de viromas completos. Los estudios de metagenómica viral han demostrado la pequeña fracción de la biodiversidad viral descrita hasta la fecha. La mayoría de los estudios que utilizan secuenciación para caracterizar virus de plantas se han centrado en cultivos económicamente importantes, sólo un pequeño número de estudios han considerado diferentes fuentes de muestras como heces de primates. Caracterizar los virus de plantas ingeridas por los primates es muy relevante, ya que estas plantas pueden afectar la dinámica en los cultivos, actuando como reservorios virales. En el presente estudio, se trabajó con las secuencias de 11 muestras de heces de primates obtenidas por secuenciación masiva shotgun. Se seleccionó un pipeline bioinformático considerando la secuencia del hospedador y se procesó los datos a través del clúster de CEDIA en Linux. La asignación taxonómica se realizó en MEGAN6 v 6.25.9 y el ensamblaje en Geneious prime v 2024.0.5. Se detectaron algunos géneros virales en donde predominaron *Fabavirus* y *Tobamovirus* con sus especies representativas *Broad bean wilt virus 1* y *Tobacco mosaic virus* respectivamente. En el ensamblaje, se confirmó la presencia de *Tobacco mosaic virus* para algunas muestras. Estos datos sugieren la presencia de virus en plantas asociados con leguminosas y frutales. Además, se demostró las tendencias inesperadas en la distribución de virus, resaltando el potencial de la metagenómica como enfoque de conservación de la biodiversidad y seguridad alimentaria.

1. INTRODUCCION

Las plantas representan cerca del 80% de la biomasa de la Tierra y desempeñan un papel importante en la seguridad alimentaria y la producción agrícola a nivel mundial (Dombrovsky & Smith, 2017). Los virus que infectan las plantas amenazan continuamente, la sostenibilidad de los alimentos al afectar el rendimiento y la calidad de los cultivos, generando pérdidas económicas sustanciales en todo el mundo (Nicaise, 2014).

Los virus se han asociado con enfermedades de las plantas desde que fueron descritos por primera vez en 1898 (Beijerinck, 1898). Actualmente, también se han descrito virus con impactos positivos en las plantas hospedantes que implican la protección ambiental de la planta huésped, la protección contra otros patógenos o el control de las respuestas de la planta a las necesidades nutricionales. Por otro lado, las interacciones negativas se estudian principalmente como síntomas de enfermedades, como el retraso del crecimiento o la necrosis, por lo tanto, la gran mayoría de las investigaciones sobre virus se han centrado en el aspecto patológico de estas interacciones (Roossinck, 2011).

El Comité Internacional para la Taxonomía de Virus (ICTV), es el encargado de aprobar los nombres científicos de los virus y en su noveno informe mencionan que existen alrededor de 900 especies de virus de plantas (King et al., 2012). Aunque, algunos estudios preliminares de plantas silvestres indican que aún existen miles de virus nuevos por descubrir en plantas (Roossinck, 2012; Roossinck et al., 2010).

Históricamente, el estudio de las comunidades virales se ha llevado a cabo mediante co-cultivo de virus y sus huéspedes celulares (Tennant et al., 2018). Hoy en día, el estudio de virus se puede realizar utilizando estrategias de base metagenómica que no dependen de enfoques de cultivo celular (Nooij et al., 2018). La metagenómica representa una oportunidad única para describir la

composición de comunidades virales en diferentes ambientes, así como para analizar los reservorios genéticos virales para caracterizar nuevas proteínas y compuestos bioactivos de potencial utilidad biotecnológica (Dávila et al., 2019).

Varios estudios metagenómicos han identificado virus de plantas similares en una amplia variedad de muestras como: heces de humanos y otros mamífero, fuentes de agua dulce y salada, suelo de cultivos de arroz e insectos que se alimentan de plantas (Roossinck, 2012). La mayoría de los virus de estos entornos son virus vegetales conocidos, y su ingreso al medio ambiente fue a través de la ingestión animal de plantas y el posterior paso de los virus a través del intestino. Los virus encontrados en estos lugares son de géneros con estructuras de cápside muy estables, como Tobamovirus (Roossinck, 2013).

Según Susi y sus colaboradores (2019), mencionan que existen varias aplicaciones para los enfoques metagenómicos en el diagnóstico de virus de plantas: (i) identificar las causas de enfermedades virales en cultivos económicamente importantes; (ii) detectar virus específicos cuando se sospecha su presencia, (iii) detectar virus asintomáticos o crípticos, y (iv) descubrir virus nuevos cuya presencia no se haya informado antes.

En este contexto, la metagenómica permitirá la vigilancia a largo plazo de los virus dentro de plantas silvestres y cultivadas (Roossinck et al., 2015). Si bien los virus de plantas no afectan a la salud de los hospedadores, pone en riesgo los cultivos que son parte de la cadena alimenticia de los primates y otros individuos, alterando el equilibrio y la dinámica ecológica del ecosistema.

La presente investigación forma parte de un proyecto cuyo objetivo es identificar y reportar los principales virus que afectan la salud de los primates. Esta investigación, al estar relacionada con virus de plantas facilitará el desarrollo de un pipeline bioinformático optimizado en términos de tiempo y resultados. Este pipeline será implementado en la siguiente fase del proyecto.

2. PLANTEAMIENTO DEL PROBLEMA

La aparición de virus en los cultivos es un proceso complejo que implica interacciones entre huéspedes silvestres, domesticados, insectos, entre otros vectores de virus, y los cambios en los ecosistemas. Aunque los datos son bastante limitados, se cree que la reducción de la biodiversidad en los agroecosistemas, tanto en términos de riqueza de especies de plantas como de diversidad genética dentro de las especies, conducirá a aumentos en la incidencia de enfermedades virales (Roossinck & García, 2015).

La metagenómica consiste en la secuenciación masiva de todos los ácidos nucleicos y tiene la capacidad de detectar en un solo ensayo todos los potenciales virus de plantas de una muestra (incluyendo los viables no cultivables), generando una ventaja en relación con los métodos de diagnóstico tradicionales (Comas et al., 2020). Esta herramienta permite obtener resultados completos que sirven para crear estrategias de monitoreo de virus de plantas y futuros planes de acción, en el caso de detectar infecciones de plantas domesticadas que formen parte de la dieta de mamíferos, como los primates.

Sin embargo, pocos estudios han utilizado la metagenómica para catalogar los virus vegetales a partir de muestras como: plantas, fuentes acuáticas, heces de animales e insectos. Otra dificultad en este tipo de investigaciones va relacionado con que las lecturas de secuencia obtenidas de virus no tienen homólogos detectables en las bases de datos de secuencias como GenBank (Roossinck et al., 2015).

Por lo tanto, los estudios metagenómicos de virus de plantas están en constante desarrollo y actualización de información. Es por esto, que es indispensable un estudio metagenómico en varias muestras que permita encontrar coincidencias de virus de plantas descritas y el reporte de nuevas especies, relacionadas con la dieta vegetariana de los primates. Siendo así, uno de los primeros

reportes en el país que busca sentar las bases de futuras investigaciones que empleen este tipo de análisis.

3. OBJETIVOS

3.1 Objetivo general

Analizar mediante un estudio metagenómico los virus de plantas presentes en muestras de heces de primates del Ecuador.

3.2 Objetivos específicos

- Desarrollar una estrategia de análisis bioinformático que permita pre-procesar los datos metagenómicos.
- Comparar las secuencias obtenidas con bases de datos de referencia para identificar los virus de plantas presentes.
- Estimar la abundancia y distribución de los virus de plantas de todas las muestras.

4. METODOLOGIA

4.1 Datos metagenómicos

Los datos metagenómicos proporcionados fueron 11 muestras secuenciadas por secuenciación masiva shotgun con la Tecnología Illumina. En la siguiente tabla se visualiza el código de la muestra, el estado de conservación del hospedero y la provincia referente.

Tabla 1. Datos metagenómicos

| No | Código de la muestra | Estado de conservación del hospedero | Provincia referente |
|-----------|-----------------------------|---|----------------------------|
| 1 | M449 | Libre | Napo |
| 2 | M452 | Cautiverio | Pastaza |

| | | | |
|-----------|------|------------|---------|
| 3 | M455 | Cautiverio | Pastaza |
| 4 | M456 | Cautiverio | Pastaza |
| 5 | M459 | Cautiverio | Pastaza |
| 6 | M460 | Cautiverio | Pastaza |
| 7 | M461 | Cautiverio | Pastaza |
| 8 | M462 | Cautiverio | Pastaza |
| 9 | M463 | Cautiverio | Pastaza |
| 10 | M464 | Cautiverio | Pastaza |
| 11 | M469 | Cautiverio | Pastaza |

Fuente: Elaboración propia

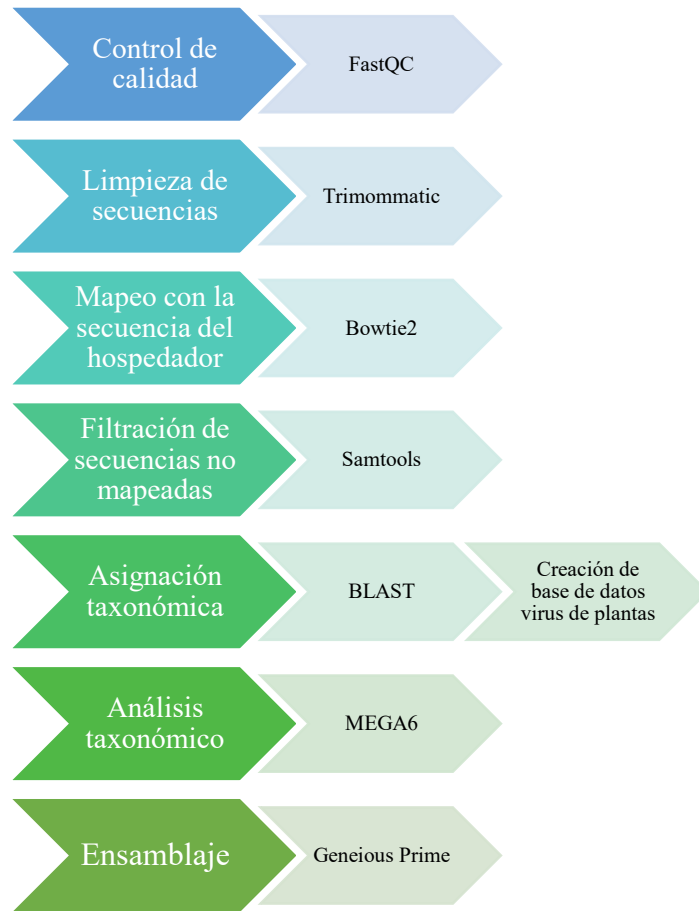
4.2 Instrumentos y herramientas para el procesamiento de datos metagenómicos

El procesamiento de los datos se realizó en Linux por medio del clúster de CEDIA a través de la herramienta PuTTY y WinSCP.

4.2.1 Selección del pipeline bioinformático

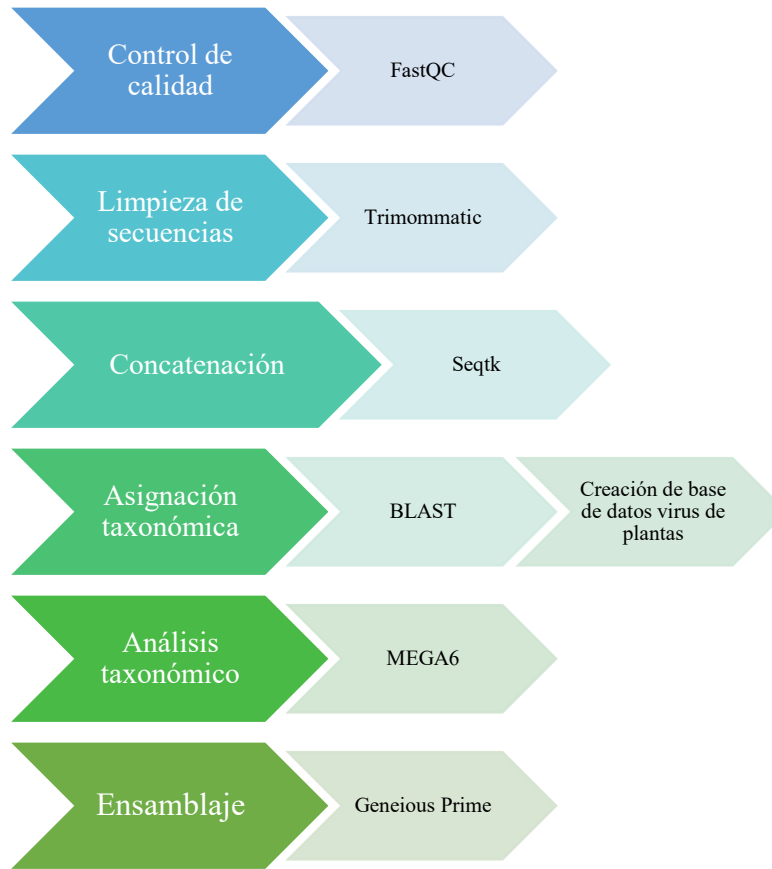
Inicialmente, se trabajó con una muestra para seleccionar el mejor pipeline bioinformático. El primer pipeline bioinformático se enfocó en eliminar la secuencia del hospedero y constó de la siguiente manera (Figura 1):

Figura 1. Pipeline bioinformático sin el hospedero



El segundo pipeline bioinformático se enfocó en mantener la secuencia del hospedero y constó de la siguiente manera (Figura 2):

Figura 2. Pipeline bioinformático con el hospedero



4.2.2 Procesamiento de los datos metagenómicos

Considerando el tiempo de ejecución y el número de lecturas obtenidas, se procedió a procesar los datos metagenómicos con el pipeline bioinformático con el hospedero.

4.2.2.1 Control de calidad

El control de calidad de los datos de secuenciación se realizó con la herramienta **FastQC**. En esta herramienta no se especificó ningún parámetro, pues el único requisito para su ejecución fue seleccionar el objeto de entrada, que, en este caso, fueron los archivos tanto de la secuencia forward como de la secuencia reverse de cada muestra. FastQC generó un informe, con una serie de tablas y gráficas por muestra. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ /home/francisco.flores/PROGRAMS/FastQC/fastqc /home/francisco.flores/DATA/Primates/M452_1.fastq.gz
```

```
(base) francisco.flores@dgx-node-0-0:~$ /home/francisco.flores/PROGRAMS/FastQC/fastqc /home/francisco.flores/DATA/Primates/M452_2.fastq.gz
```

4.2.2.2 Limpieza de secuencias

La limpieza de secuencias se realizó con el comando **Trimmomatic**, que es una herramienta para recortar y limpiar secuencias de datos de lectura masiva. Los parámetros de ejecución del comando fueron:

- Codificación de calidad Phred +33
- Recorte de los adaptadores: 2 (El número máximo de desajustes permitidos en la alineación de adaptadores):30 (La puntuación de corte en la cual se considera que un adaptador ha sido encontrado):10 (La longitud mínima de la superposición entre la secuencia y el adaptador).
- SLIDINGWINDOW: 5:20 para realizar un recorte con una ventana deslizante de tamaño 5, eliminando bases cuando la calidad promedio dentro de la ventana cae por debajo de 20.
- LEADING:5 para eliminar bases del extremo 5' si tienen una calidad menor que 5.
- TRAILING:5 para eliminar bases del extremo 3' si tienen una calidad menor que 5.
- MINLEN:50 para descartar las lecturas que tienen una longitud menor que 50 nucleótidos después del trimming.

Trimmomatic generó cuatro archivos, dos archivos forward y reverse pareados y dos archivos forward y reverse no pareados. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ java -jar /home/francisco.flores/PROGRAMS/Trimmomatic-0.39/trimmomatic-0.39.jar PE -phred33 -threads 16 /home/francisco.flores/DATA/Primates/M452_1.fastq.gz /home/francisco.flores/DATA/Primates/M452_2.fastq.gz /home/francisco.flores/DATA/Primates/M452_leftP.fq /home/francisco.flores/DATA/Primates/M452_leftU.fq /home/francisco.flores/DATA/Primates/M452_rightP.fq /home/francisco.flores/DATA/Primates/M452_rightU.fq ILLUMINACLIP:/home/francisco.flores/PROGRAMS/Trimmomatic-0.39/adapters/Illumina_adapt.fa:2:30:10 SLIDINGWINDOW:5:20 LEADING:5 TRAILING:5 MINLEN:50
```

4.2.2.3 Concatenación

La concatenación de secuencias se realizó con el comando **Seqtk mergepe**, que es una herramienta para manipular archivos FASTQ y permitió intercalar lecturas pareadas en un solo archivo. En esta herramienta no se especificó ningún parámetro, pues el único requisito para su ejecución fue seleccionar el objeto de entrada, que, en este caso, fueron los archivos FASTQ con las lecturas pareadas (post-trimming). Seqtk generó un archivo con las lecturas de ambos extremos intercaladas, de modo que cada par de lecturas secuenciales pertenece a la misma pareja de fragmentos de ADN. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ seqtk mergepe M452_rightP.fq M452_leftP.fq > M452_interleaved.fq
```

4.2.2.4 Asignación taxonómica

- *Cambio de formato FASTQ a FASTA*

Antes de la asignación taxonómica se realizó el cambio de formato de FASTQ a FASTA del archivo concatenado. El comando **sed** permitió realizar el cambio de formato. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ sed -n '1~4s/^@/>/p;2~4p' /home/francisco.flores/DATA/Primates/M452_interleaved.fq > /home/francisco.flores/DATA/Primates/M452_interleaved.fasta
```

- *Creación de la base de datos de virus de plantas*

En la página web del National Center for Biotechnology Information (NCBI), se descargó todos los virus de plantas en un solo archivo en formato fasta.

Luego, se creó una base de datos BLAST en Linux mediante el comando **makeblastdb**. Los parámetros de ejecución del comando fueron:

- `dbtype nucl` para indicar que las secuencias en el archivo de entrada son de tipo nucleotídico (ADN o ARN).
- `parse_seqids` para permitir que `makeblastdb` procese los identificadores de las secuencias en el archivo FASTA.

El comando `makeblastdb` generó una base de datos BLAST nucleotídica a partir del archivo FASTA descargado del NCBI con el nombre `PlantVirus2024`. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ makeblastdb -in /home/francisco.flores/DATABAcdfasta -dbtype nucl -parse_seqids -out PlantVirus2024 -title PlantVirus2024
```

- *BLAST*

Para realizar una búsqueda BLAST se utilizó el comando **`blastn`**. Los parámetros de ejecución del comando fueron:

- `db PlantVirus2024` para especificar la base de datos BLAST que se utilizó para la búsqueda. En este caso, la base de datos fue `PlantVirus2024`.
- `query` para especificar el archivo de secuencias en formato FASTA que se utilizaron como consulta en la búsqueda.
- `outfmt 6` para definir el formato de salida como tabular, donde cada línea representó una alineación y los campos están separados por tabulaciones.
- `perc_identity 90` para filtrar los resultados con las alineaciones con un porcentaje de identidad de al menos el 90%.
- `qcov_hsp_perc 25` para filtrar los resultados que muestren las alineaciones en las que al menos el 25% de la secuencia de consulta fue cubierta por el HSP (High-scoring Pair).

- `max_target_seqs 1` para limitar el número de alineaciones que se mostraron para cada secuencia de consulta a una sola, seleccionando la mejor alineación según la puntuación.

El comando `blastn` generó un archivo con los resultados de la búsqueda. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ blastn -db PlantVirus2024 -query /home/francisco.flores/DATA/Primates/M452_interleaved.fasta -out /home/francisco.flores/DATA/Primates/M452_interleavedBlastPlantVir -outfmt 6 -perc_identity 90 -qcov_hsp_perc 25 -num_threads 64 -max_target_seqs 1
```

4.2.2.5 Análisis taxonómico

Una vez procesada la muestra en un archivo con los resultados de la búsqueda en BLAST, se realizó un análisis taxonómico utilizando la taxonomía NCBI en el programa **MEGAN6**, que proporciona una amplia gama de métodos de análisis y visualización para el análisis de datos metagenómicos de lectura corta y larga con enfoques de análisis de abundancia taxonómica y funcional (Bağcı et al., 2019).

En el programa se seleccionó la pestaña File y se escogió la opción Import From BLAST. En el cuadro de diálogo se importó el archivo obtenido de BLAST y se cargó la base de datos *megan-nucl-Feb2022.db*, para asignar secuencias a taxones específicos basándose en comparaciones con secuencias de referencia conocidas.

El programa MEGAN6 generó un archivo en formato RMA6 que contiene la composición taxonómica visual de cada muestra analizada.

En el mismo programa, se realizó la comparación de todos los archivos RMA6 de cada muestra para obtener un gráfico que permitió identificar la abundancia de virus de todas las 11 muestras.

4.2.2.6 Ensamblaje

- *Extracción de códigos para obtener secuencias a partir del archivo BLAST*

Antes de ensamblar se realizó la extracción de códigos del archivo BLAST con el comando **sed** y **awk** para procesar un archivo y extraer la primera columna de datos, realizando una sustitución de caracteres en el proceso. Los comandos utilizados generaron un archivo de texto con la primera columna del archivo original que correspondían a los identificadores de las secuencias. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ sed 's|/| /' /home/francisco.flores/DATA/Primates/M452_interleavedBlastPlantVir | awk -v 'OFS='\t' '{print $1}'>/home/francisco.flores/DATA/Primates/M452_interleavedBlastPlantVirnames.txt
```

- *Extracción de secuencias*

La extracción de secuencias se realizó con el comando **seqtk subseq**, para extraer secuencias específicas de un archivo FASTA basado en una lista de identificadores de secuencias proporcionada en un archivo de texto. El comando **seqtk subseq** generó un archivo con las secuencias extraídas en formato FASTA. A continuación, se indica el comando procesado en Linux:

```
(base) francisco.flores@dgx-node-0-0:~$ seqtk subseq /home/francisco.flores/DATA/Primates/M452_interleaved.fasta /home/francisco.flores/DATA/Primates/M452_interleavedBlastPlantVirnames.txt > /home/francisco.flores/DATA/Primates/M452_interleavedhits.fna
```

- *Ensamblaje de novo y BLAST en Geneious Prime*

En el programa se seleccionó la pestaña File y se escogió la opción Import. En el cuadro de diálogo se importó el archivo con las secuencias extraídas en formato FASTA. Luego, se seleccionó la opción Align/Assemble y se procedió con la opción De Novo Assemble.

Se analizó el reporte del ensamblaje y se seleccionó el contig con mayor número de pares de bases. Después, se eligió la opción de BLAST para el contig mencionado y se comprobó que el resultado sea un virus de planta.

5. RESULTADOS

5.1 Selección del pipeline bioinformático

La selección del pipeline bioinformático fue en base al tiempo de ejecución y el número de lecturas obtenidas. En la tabla 2 se observa el tiempo de ejecución de cada proceso con el pipeline bioinformático sin el hospedero y con el hospedero.

Tabla 2. Tiempo de ejecución del pipeline bioinformático sin hospedero y con hospedero

| Proceso | Tiempo sin hospedero | Tiempo con hospedero |
|--------------------------------------|-----------------------------|-----------------------------|
| Control de calidad | 2 min | 2 min |
| Limpieza de secuencias | 10 min | 10 min |
| Concatenación | No aplica | 5 min |
| Mapeo con el hospedero | 1h40min | No aplica |
| Filtración de secuencias no mapeadas | 6 min | No aplica |
| Asignación taxonómica | 30 min | 2 h |
| Análisis taxonómico | 10 min | 10 min |
| Ensamblaje | 10 min | 10 min |
| TOTAL | 2h48min | 2h37min |

Fuente: Elaboración propia

En la tabla 3 se observa el número de lecturas obtenidas de cada proceso con el pipeline bioinformático sin el hospedero y con el hospedero de las 3 especies de virus abundantes de la muestra.

Tabla 3. Número de lecturas obtenidas con el pipeline bioinformático sin hospedero y con hospedero

| Especie | Número de lecturas | |
|--------------------------------|------------------------|------------------------|
| | Pipeline sin hospedero | Pipeline con hospedero |
| <i>Alfalfa mosaic virus</i> | 339 | 1060 |
| <i>Citrus exocortis viroid</i> | 280 | 546 |
| <i>Papaya Mosaic virus</i> | 146 | 372 |

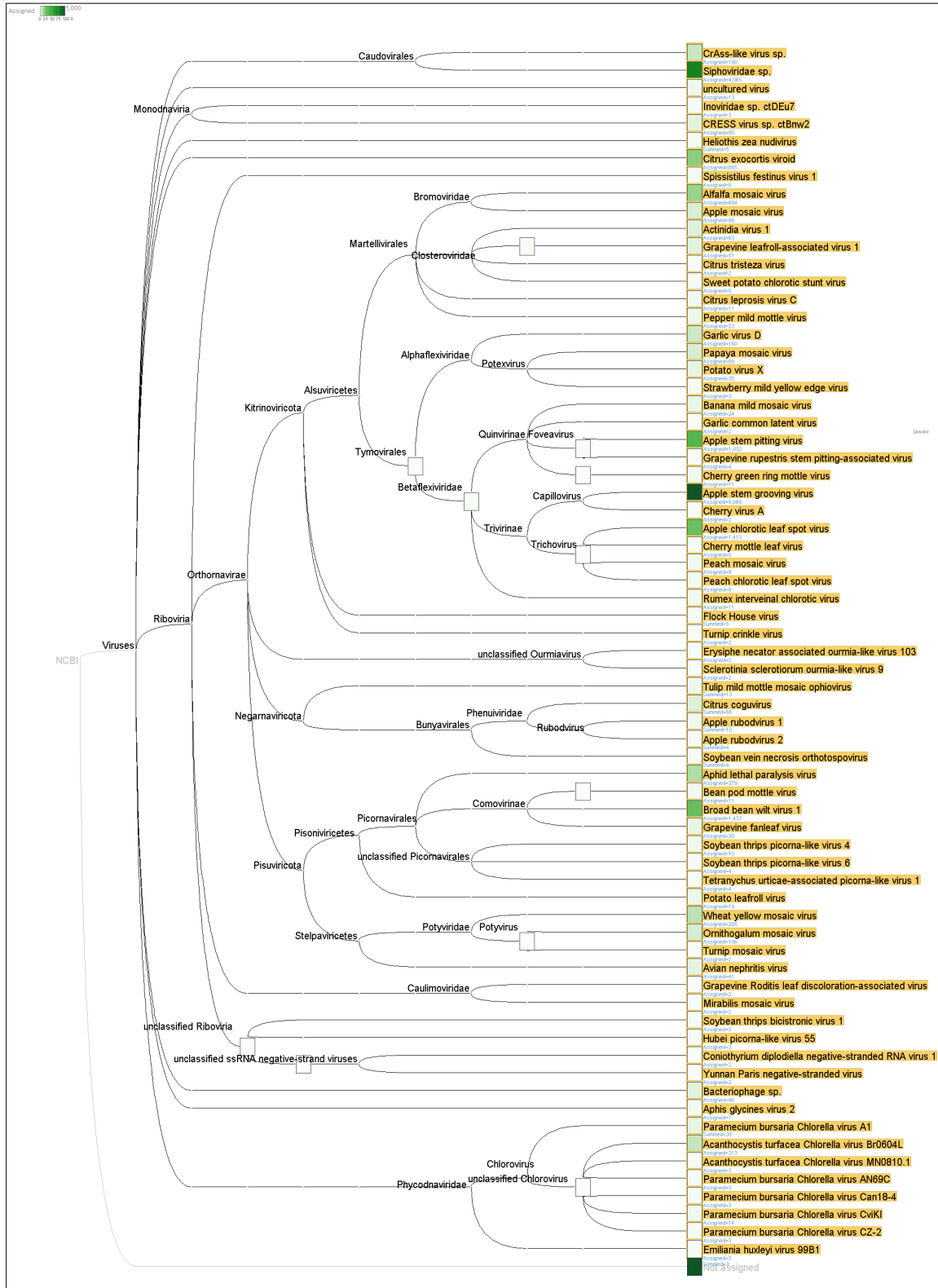
Fuente: Elaboración propia

El pipeline bioinformático seleccionado fue con el hospedero porque el tiempo de ejecución fue de 2 horas con 37 minutos y se obtuvo un mayor número de lecturas de cada virus.

5.2 Análisis Taxonómico

La muestra 449 recolectada en Napo, presentó una asignación taxonómica dominada por virus del género *Capillovirus*, siendo *Apple stem grooving virus* (5668 reads) su especie representante más abundante (Figura 3).

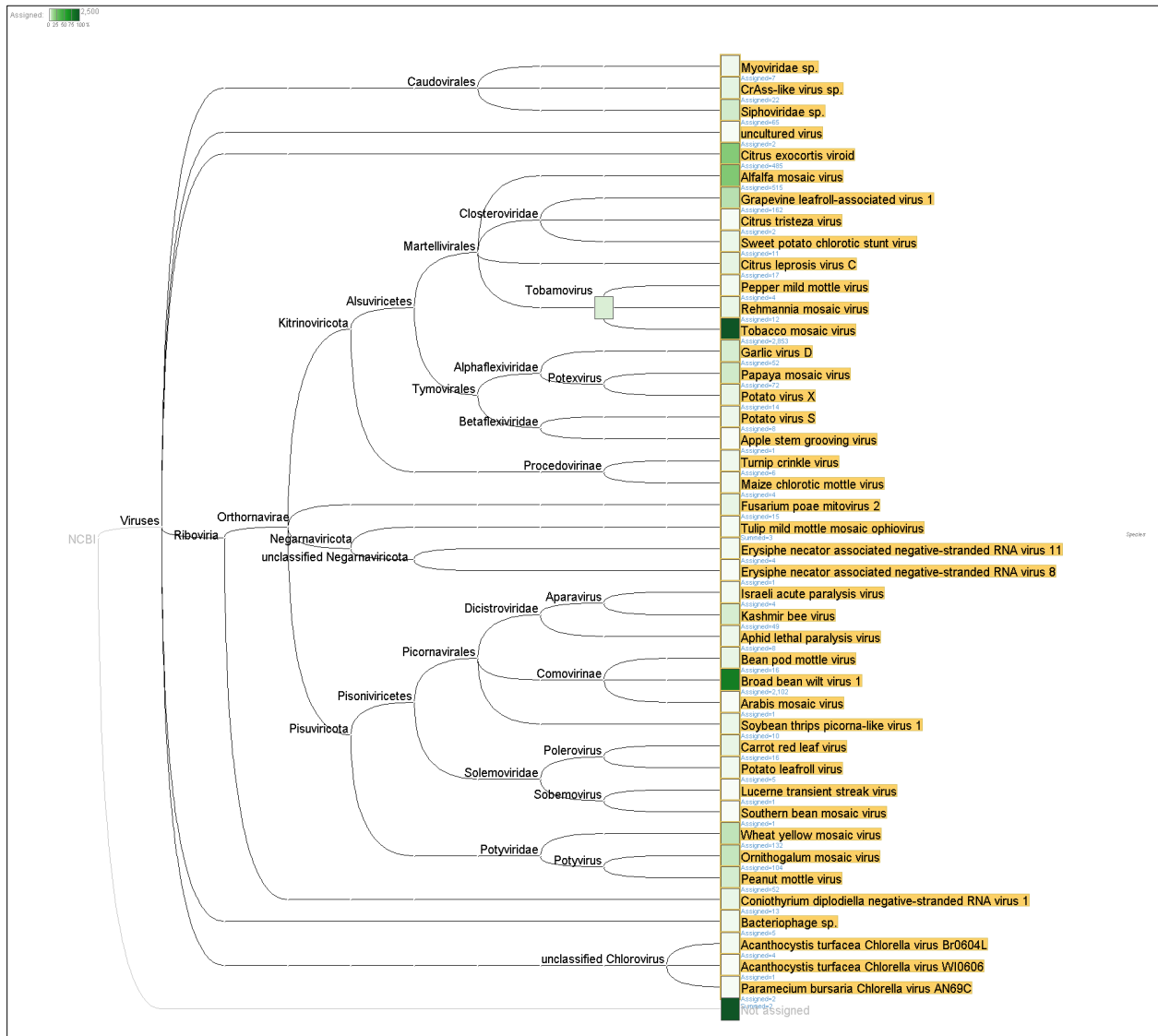
Figura 3. Asignación taxonómica muestra 449



Fuente: Elaboración propia

La muestra 452 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Tobamovirus* y *Fabavirus*, siendo *Tobacco mosaic virus* (2853 reads) y *Broad bean wilt virus 1* (2102 reads) respectivamente, su especie representante más abundante (Figura 4).

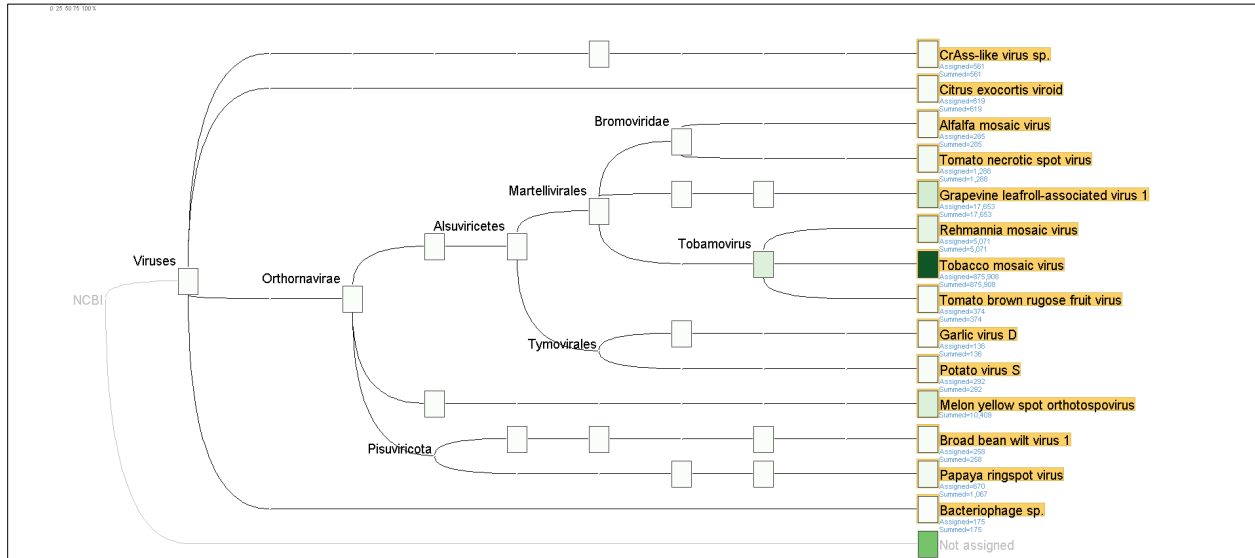
Figura 4. Asignación taxonómica muestra 452



Fuente: Elaboración propia

La muestra 455 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Tobamovirus*, siendo *Tobacco mosaic virus* (875908 reads) su especie representante más abundante (Figura 5).

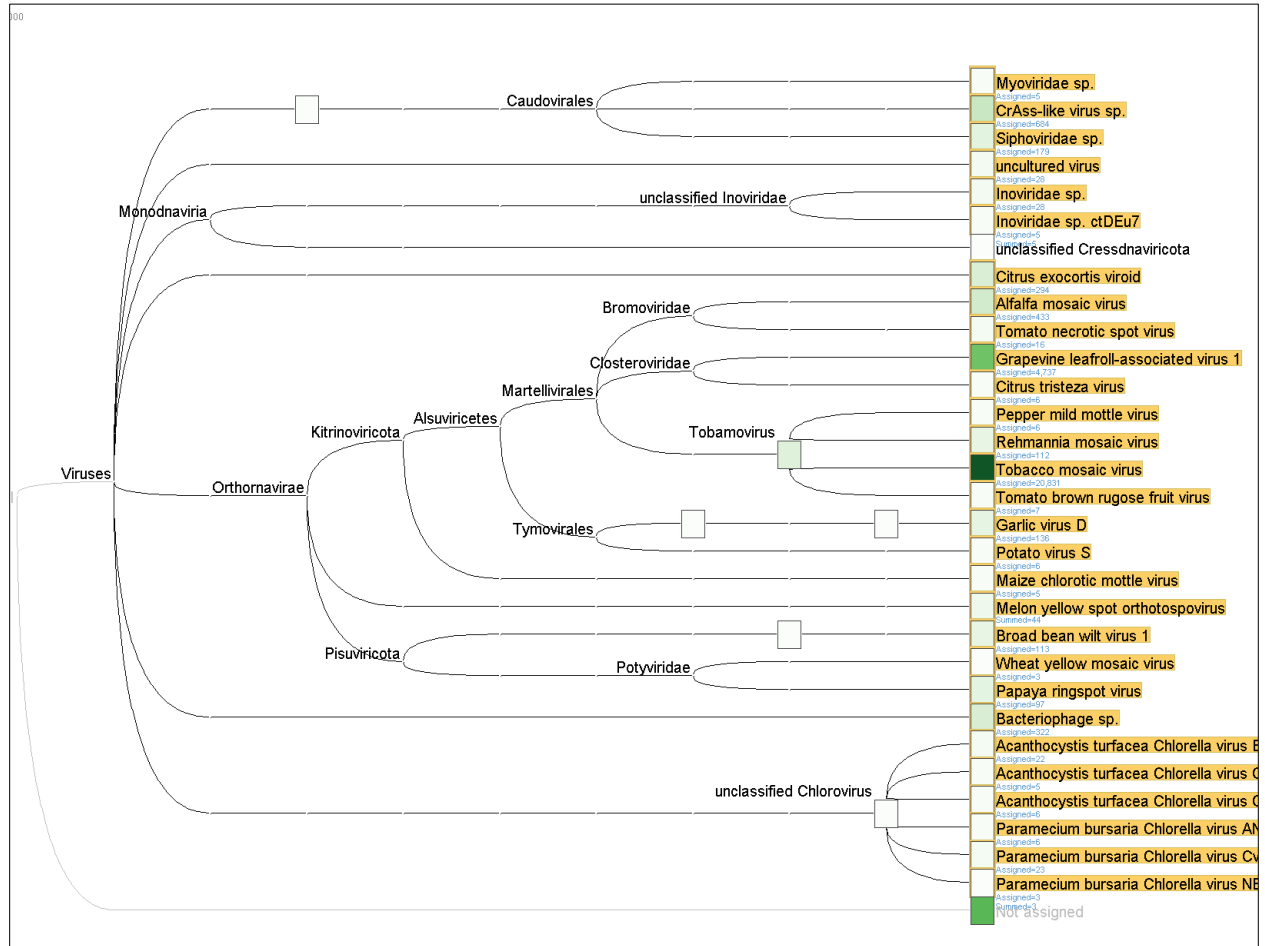
Figura 5. Asignación taxonómica muestra 455



Fuente: Elaboración propia

La muestra 456 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Tobamovirus*, siendo *Tobacco mosaic virus* (20831 reads) su especie representante más abundante (Figura 6).

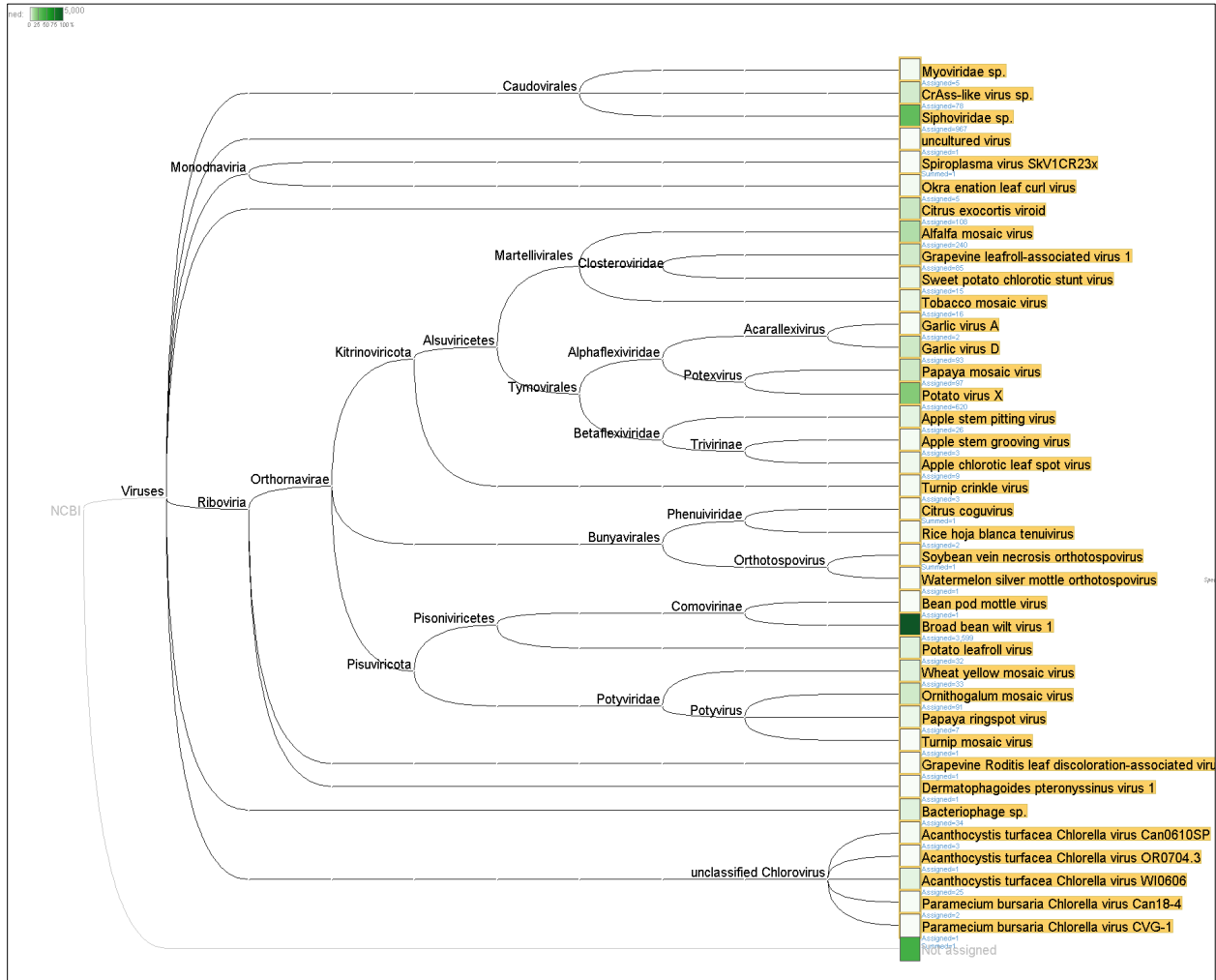
Figura 6. Asignación taxonómica muestra 456



Fuente: Elaboración propia

La muestra 459 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Fabavirus*, siendo *Broad bean wilt virus 1* (3599 reads) su especie representativa más abundante (Figura 7).

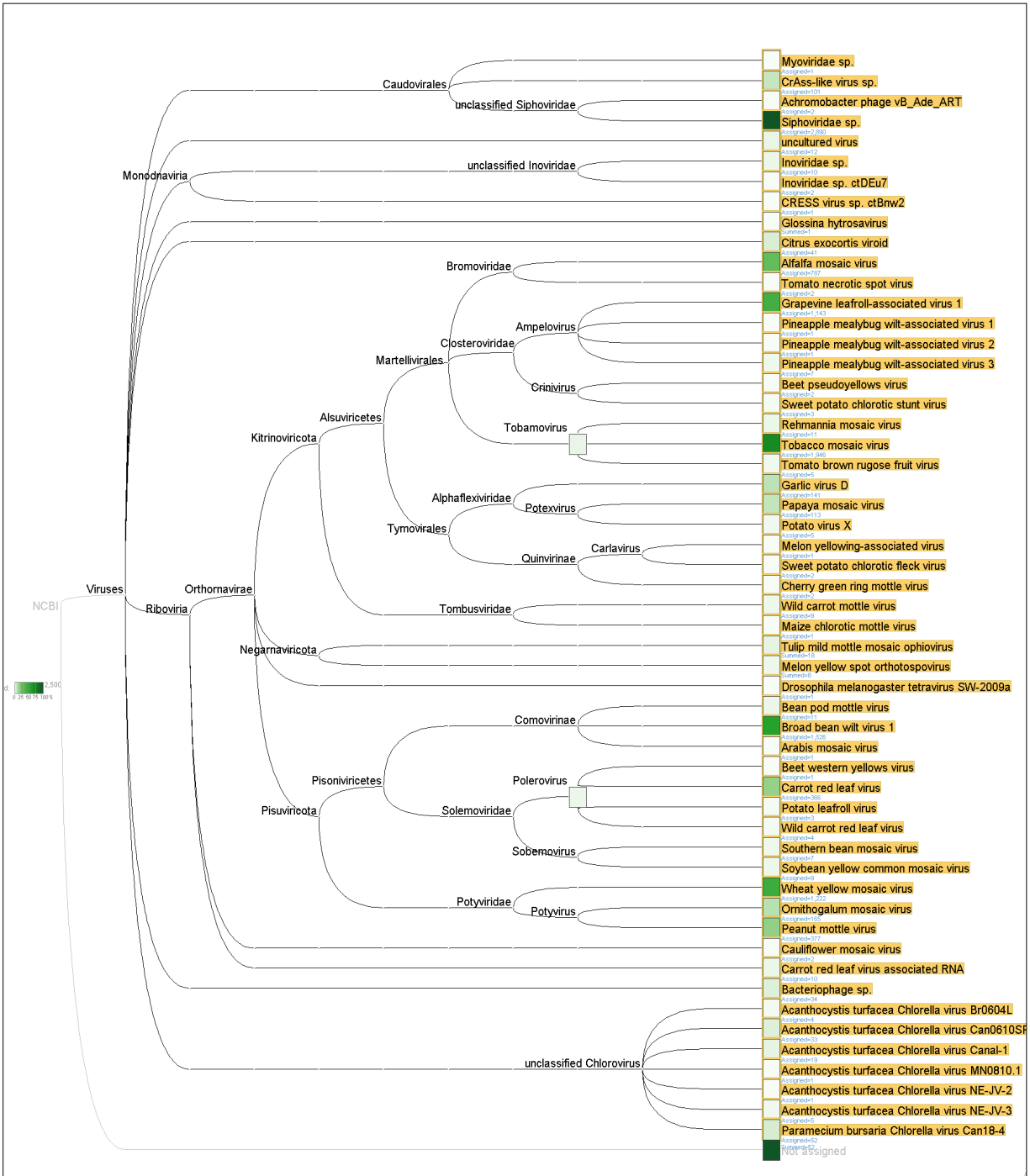
Figura 7. Asignación taxonómica muestra 459



Fuente: Elaboración propia

La muestra 460 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Tobamovirus* y *Fabavirus*, siendo *Tobacco mosaic virus* (1946 reads) y *Broad bean wilt virus 1* (1526 reads) respectivamente, su especie representante más abundante (Figura 8).

Figura 8. Asignación taxonómica muestra 460

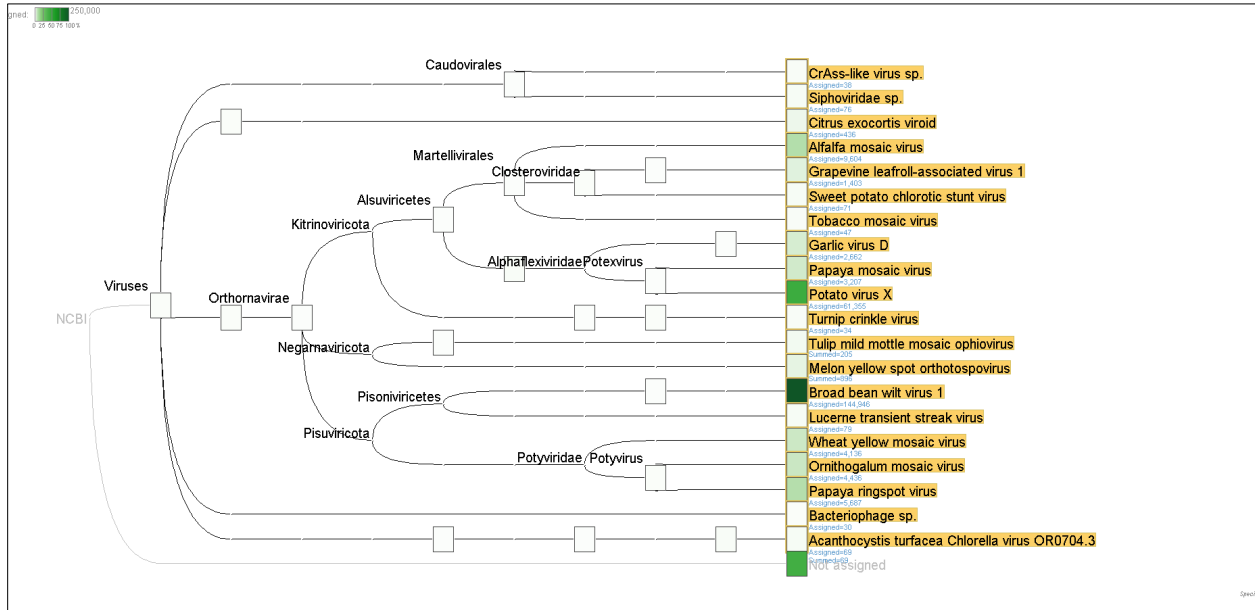


Fuente: Elaboración propia

La muestra 461 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Fabavirus*, siendo *Broad bean wilt virus 1* (8840 reads) respectivamente, su especie representante más abundante (Figura 9).

La muestra 462 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Fabavirus*, siendo *Broad bean wilt virus 1* (144946 reads) respectivamente, su especie representante más abundante (Figura 10).

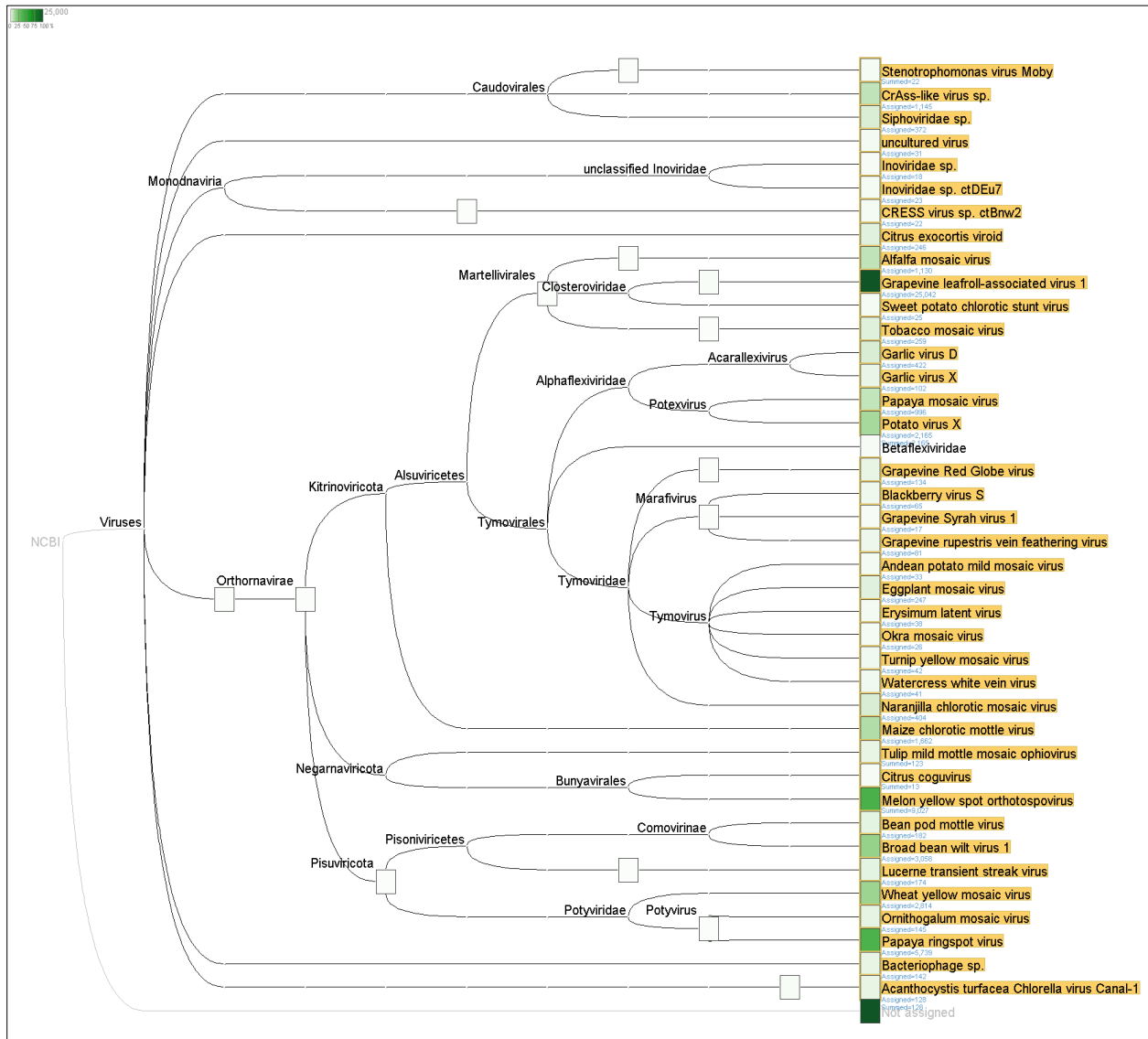
Figura 10. Asignación taxonómica muestra 462



Fuente: Elaboración propia

La muestra 463 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Ampelovirus*, siendo *Grapevine leafroll-associated virus 1* (25042 reads) respectivamente, su especie representante más abundante (Figura 11).

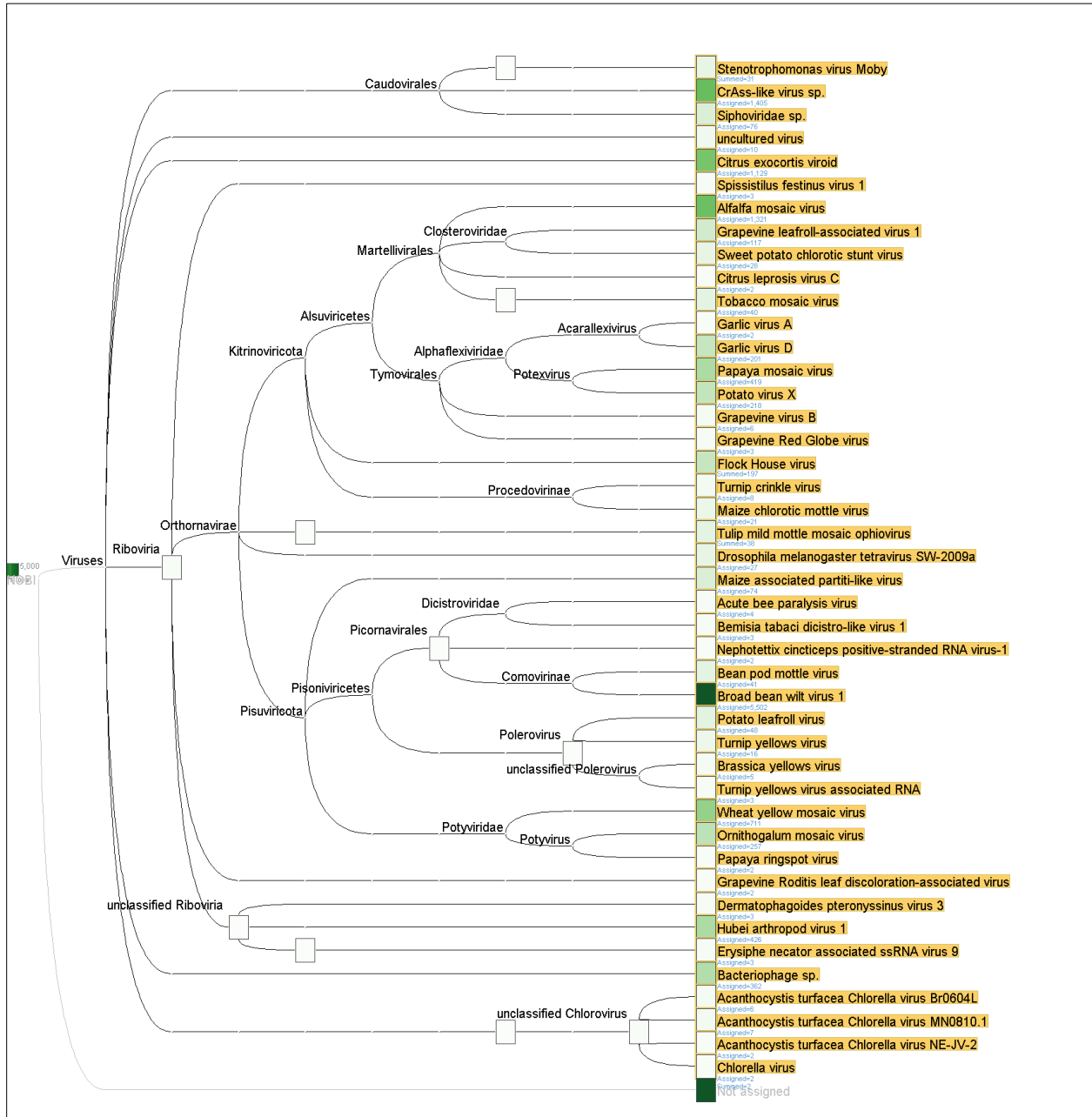
Figura 11. Asignación taxonómica muestra 463



Fuente: Elaboración propia

La muestra 464 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Fabavirus*, siendo *Broad bean wilt virus 1* (5502 reads) respectivamente, su especie representante más abundante (Figura 12).

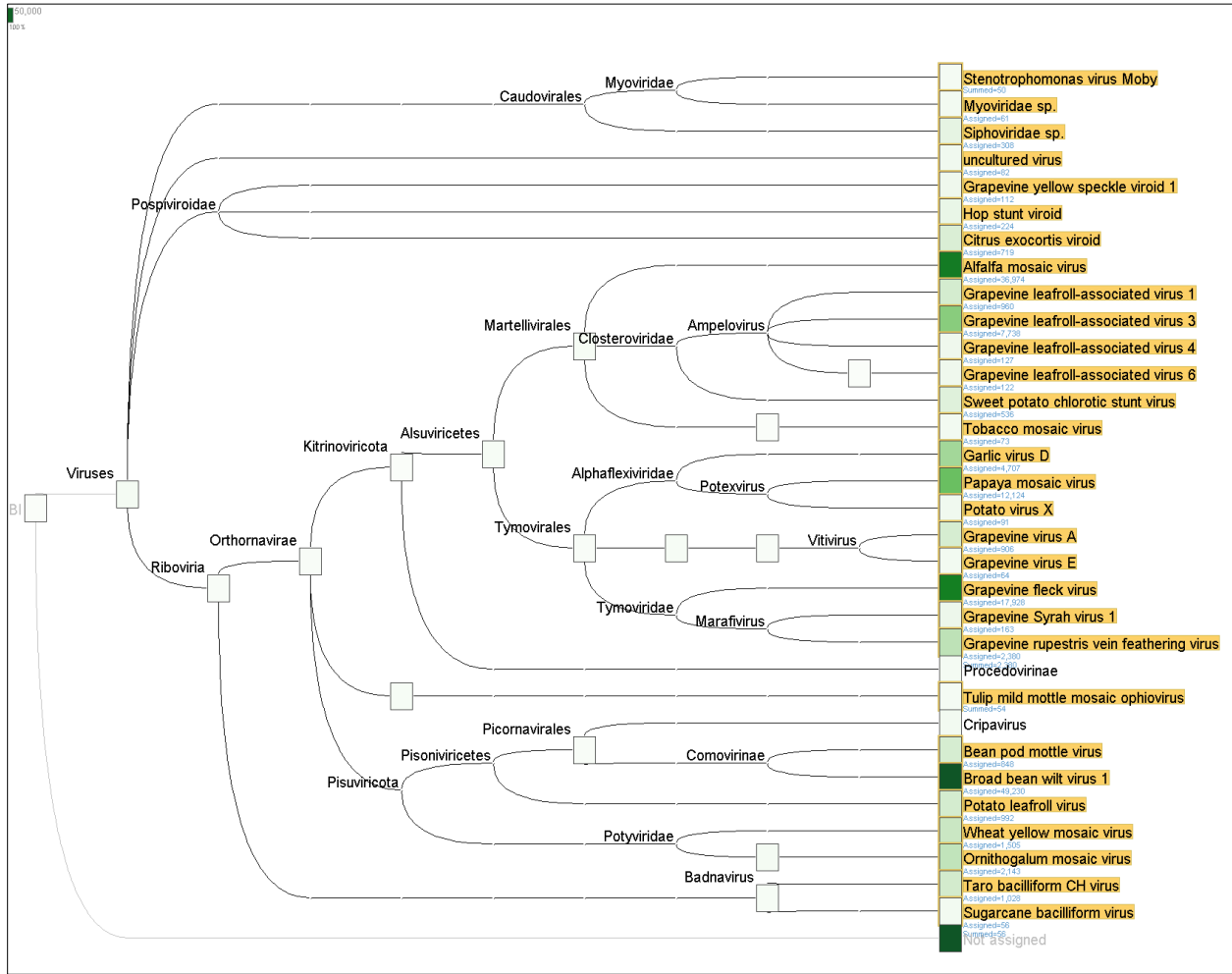
Figura 12. Asignación taxonómica muestra 464



Fuente: Elaboración propia

La muestra 469 recolectada en Pastaza, presentó una asignación taxonómica dominada por virus del género *Fabavirus* y *Alfamovirus*, siendo *Broad bean wilt virus 1* (49230 reads) y *Alfalfa mosaic virus* (36974 reads) respectivamente, su especie representante más abundante (Figura 13).

Figura 13. Asignación taxonómica muestra 469

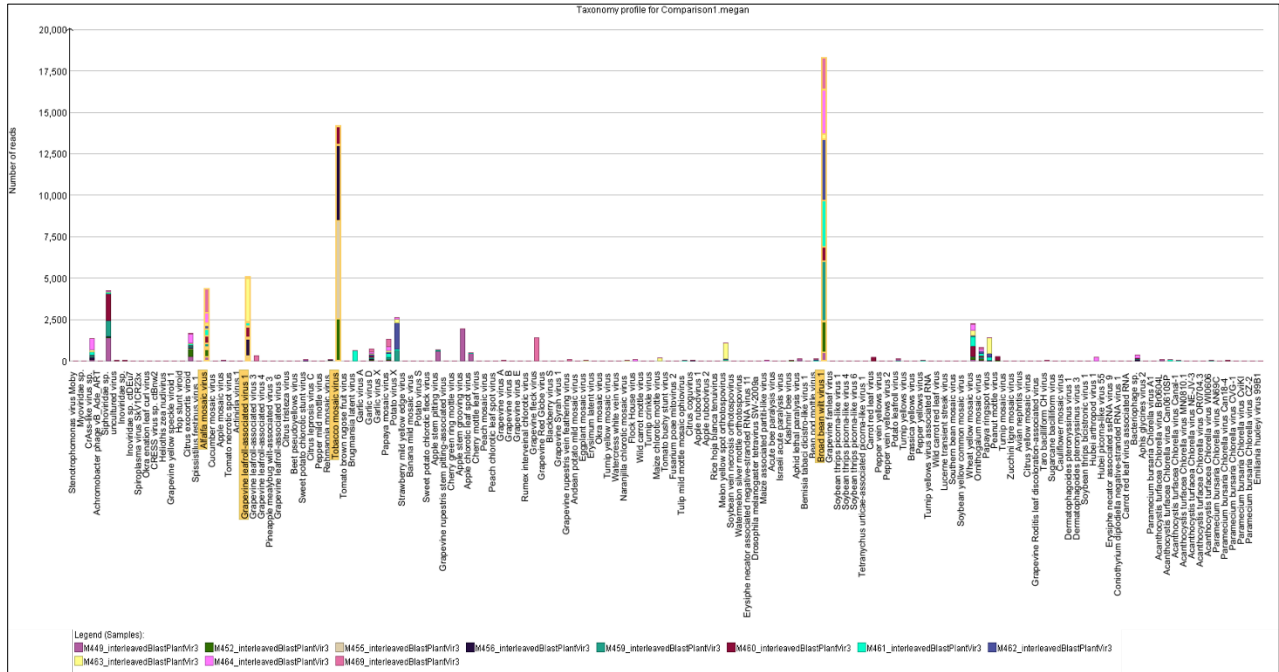


Fuente: Elaboración propia

5.3 Abundancia y distribución de los virus de plantas en todas las muestras

En la figura 14, se visualizó un gráfico de barras apiladas para mostrar la abundancia y distribución de virus presentes en plantas, cada barra dividida en segmentos representó la abundancia dentro de cada muestra del virus.

Figura 14. Comparación de abundancia y distribución de los virus de plantas



Fuente: Elaboración propia

Estos resultados sugieren una distribución no uniforme de la diversidad viral en las muestras. Algunos virus presentaron una alta prevalencia como: *Broad bean wilt virus*, *Tobacco mosaic virus*, *Grapevine leafroll-associated virus 1* y *Alfalfa mosaic virus*, mientras que otros permanecieron presentes en menor cantidad.

En la única muestra recolectada en Napo, se detectaron únicamente *Broad bean wilt virus* y *Alfalfa mosaic virus*, en contraste con las muestras recolectadas en Pastaza. Esta muestra presentaba en su mayoría virus relacionados con leguminosas y frutales.

5.4 Confirmación de virus de plantas presentes en cada muestra

En la tabla 3 se presentaron los contigs de mayor longitud obtenidos del ensamblaje de cada muestra. El tamaño de los contigs se encontró entre 1163bp a 10314bp. La mayoría de los contigs en las muestras se asociaron con fragmentos del genoma de virus de plantas.

Tabla 4. Resultados confirmatorios obtenidos del ensamblaje de *novo*

| Muestra | Tamaño contig | Similitud | Porcentaje de Identidad |
|----------------|----------------------|--|--------------------------------|
| 449 | 4401 bp | <i>Capillovirus mali</i> isolate BR-Brae1, complete genome | 98.4% |
| 452 | 2414 bp | <i>Tobacco mosaic virus</i> isolate BPP12, complete genome | 91.8% |
| 455 | 2873 bp | <i>Melon yellow spot virus</i> isolate MYSV-Ec segment L RNA-dependent RNA polymerase gene, partial cds. | 99.3% |
| 456 | 2144 bp | <i>Tobacco mosaic virus</i> isolate BPP12, complete genome | 93.1% |
| 459 | 1163 bp | <i>Segatella copri</i> strain HDE04 chromosome, complete genome | 86.2% |
| 460 | 4123 bp | <i>Carrot red leaf virus</i> HK genomic RNA, complete genome | 99.3% |
| 461 | 4226 bp | <i>Brugmansia latent virus</i> isolate 1996- 4584, complete genome | 97.4% |
| 462 | 10314 bp | <i>Papaya ringspot virus</i> isolate Bab-Ec polyprotein gene, complete | 92.3% |
| 463 | 8005 bp | Melon yellow spot virus isolate IN:Kar:Cucumber:21 segment L, complete sequence | 97.3% |

| | | | |
|------------|---------|---|-------|
| 464 | 2310 bp | MAG TPA_asm: Crassvirales sp. isolate ctM8r20, partial genome | 100% |
| 469 | 1461 bp | Grapevine leafroll-associated virus 3 isolate 14G466, complete genome. | 99.2% |

Fuente: Elaboración propia

6. DISCUSION

Los pipelines bioinformáticos constituyen flujos de trabajo automatizados que permiten procesar y analizar datos metagenómicos de manera eficiente. Sin embargo, la selección adecuada de las herramientas y algoritmos para cada etapa del proceso representa un desafío significativo, dando lugar a un posible cuello de botella bioinformático (Lorenzana López, 2020).

Los principios de buenas prácticas computacionales sugieren que el algoritmo usado sea un código claro, comprensible y obtenga los resultados esperados por el investigador (Wilson et al., 2017). En este estudio se seleccionó el pipeline bioinformático que incluía la secuencia del hospedador por el menor tiempo de ejecución y mayor obtención de lecturas.

Al tener un mayor número de lecturas, mayor será la posibilidad de conseguir un ensamblaje completo de los genomas de los virus de plantas (Pop, 2009). Además, la identificación taxonómica será más precisa y existe una mejor representación de la abundancia relativa de los virus asociados con las plantas.

Actualmente, existen dos enfoques principales que se utilizan para las asignaciones taxonómicas de lecturas metagenómicas. El primer enfoque denominado clasificación supervisada se basa en algoritmos de aprendizaje automático entrenados con bases de datos de secuencias de referencia para asignar cada lectura a un taxón conocido (McHardy et al., 2007). Algunas

herramientas como PhyloPythia y MetaPhyler utilizan este método (McHardy, 2015). El segundo enfoque evalúa la identidad taxonómica de una lectura a partir de los resultados de una búsqueda basada en homología contra la base de datos de secuencias de referencia conocidas (generalmente la base de datos no redundante del NCBI) (Sharma et al., 2012).

En este caso se aplicó el segundo enfoque con la herramienta MEGAN6, la cual, realiza la clasificación taxonómica basada en la puntuación de bits del NCBI BLAST, basado en el Ancestro Común Más Bajo (LCA) (Huson et al., 2007). Además, se obtuvo dendogramas para mostrar los datos metagenómicos de virus de plantas para cada muestra. Esta opción facilita la visualización, interpretación y el análisis de grupos taxonómicos particulares (Thomas et al., 2012).

Los grupos taxonómicos particulares en este estudio fueron especies del género *Fabavirus* y *Tobamovirus*. El género *Fabavirus* se caracteriza por infectar una amplia gama de plantas hospedantes monocotiledóneas, dicotiledóneas, particularmente en las leguminosas y son transmitidos de forma no persistente por pulgones (Sanfaçon et al., 2012).

La especie representativa de este género reportado en el estudio fue *Broad bean wilt virus 1*. BBWV-1 se distribuye a nivel mundial e infecta importantes cultivos agrícolas como la haba (*Vicia faba*), el pimiento (*Capsicum annuum*), el tomate (*Solanum lycopersicum*), la espinaca (*Spinacia oleracea*), y cultivos ornamentales como la petunia (*Petunia sp.*) y el narciso (*Narcissus sp.*) (King et al., 2012).

El BBWV-1 tiene un genoma de ARN monocatenario bipartito, que se divide en dos segmentos: ARN1 y ARN2 (Carpino et al., 2020). El ARN1 (5.8 kb) es responsable de codificar las proteínas necesarias para la replicación viral, mientras que el ARN2 (3.4 kb) codifica las

proteínas de la cubierta y las proteínas de movimiento esenciales para el ciclo de vida del virus (Sanfaçon, 2015).

El siguiente grupo taxonómico en este estudio fueron especies del género *Tobamovirus*. El género *Tobamovirus* se encuentra entre los virus de plantas más estudiados e infectan a diferentes plantas de la familia Solanaceae, Cucurbitaceae, Apocynaceae, Brassicaceae, Cactaceae y Malvaceae. Estos virus se transmiten principalmente por contacto directo entre plantas (Ilyas et al., 2022).

La especie particular de este género reportado en la investigación fue *Tobacco mosaic virus*. El virus del mosaico del tabaco (TMV) es el primer virus descubierto y es un patógeno vegetal que infecta a las plantas de la familia Solanaceae, en particular al tabaco (*Nicotiana tabacum*), tomate (*Solanum lycopersicum*), la papa (*Solanum tuberosum*) y el pimiento (*Capsicum annuum*) (Balique et al., 2012; Yamaya et al., 1988).

El TMV tiene un genoma de ARN monocatenario de 6.3 a 6.6 kb (International Committee on Taxonomy of Viruses, 2018). El ARN es monopartito y está compuesta por una hélice de 2130 subunidades de proteína de cobertura denominada CP que cubre al ARN monocatenario de 6400 nucleótidos. El ARN codifica cuatro genes: dos proteínas asociadas a la replicación, la proteína de movimiento MP y la proteína de cobertura que fueron traducidas a partir de ARNs subgenómicos (Hull, 2002).

Una limitación del estudio fue no disponer equitativamente de muestras de heces de primates en estado de cautiverio y libre, por lo que la comparación no fue significativa. Sin embargo, se observó que en la única muestra del primate libre perteneciente a Napo se detectaron algunos virus relacionados con plantas de frutales, mientras que el resto de las muestras de Pastaza coincidían con los virus mencionados anteriormente.

Al ser muestras provenientes de primates en estado de cautiverio de una misma provincia es importante relacionar con la dieta del individuo. Generalmente, la mayor parte de la dieta de los primates en estado de cautiverio se basa en el consumo de hojas de leguminosas (folívora) y frutas (frugívora). Mientras que la dieta de primates en estado libre es diversa y se relaciona con el consumo de frutas de temporada (frugívora) y de insectos (insectívora) (Urbina, 2010).

En este contexto, los resultados obtenidos son consistentes con la diversidad de virus de plantas previamente mencionada. Los virus asociados a plantas frutales pertenecieron a la muestra del primate en estado libre y los virus de plantas leguminosas se relacionaron con los primates en estado de cautiverio. Los primates al ser frugívoros pueden ser dispersores de semillas (Andresen, 2005). Por lo tanto, los virus de plantas pueden ser transmitido por semillas, representando un riesgo significativo para la propagación de la enfermedad en nuevos cultivos (Hunter et al., 1969).

Por otro lado, el proceso de ensamblaje confirmó la presencia de ciertos virus asociados a plantas como *Tobacco mosaic virus*. Sin embargo, hubo muestras en las que el reporte del virus no coincidió entre los resultados de abundancia obtenidos con MEGAN6 y el ensamblaje. Existen varios factores que respaldan estas diferencias. Un factor se relaciona con que MEGAN6 puede destacar un virus abundante basado en la similitud de secuencias con referencias conocidas, debido a que la consideración de los hits se basa únicamente en puntajes de bits, esto puede conducir a un mayor número de asignaciones taxonómicas no específicas debido a la consideración de los hits esperados (correctos) e inesperados (nivel taxonómico más alto) (Sharma et al., 2012).

Otro factor, está relacionado con el ensamblaje de las secuencias. El ensamblaje de *nov* depende de referencias y puede identificar virus nuevos o divergentes (Aguilar-Bultet & Falquet, 2015). Un virus puede parecer más abundante en este proceso si tiene un genoma que se ensambla mejor debido a una mayor cobertura, menor complejidad, o si su secuencia es más completa en la

muestra (López-Pérez & Fernández, 2013). En algunos casos, el resultado de los ensamblajes pueden ser falsos positivos al considerar pequeños fragmentos del genoma del virus (Gama, 2017). Un virus identificado con pocas lecturas o de menor tamaño del contig en comparación con otros virus en la muestra puede ser un candidato a falso positivo.

Los estudios metagenómicos abarcan una amplia gama de campos de investigación, entre los que se incluyen la investigación medioambiental, la biotecnología vegetal, agrícola y la virología (Mokili et al., 2012). Este estudio hace referencia a que la mayoría de los virus reportados están asociados con legumbres y frutales. Las legumbres son una de las fuentes más importantes de alimentación humana y animal, siendo susceptibles a una gran cantidad de virus vegetales (Chatzivassiliou, 2021). Muchos de estos virus causan enfermedades que afectan al manejo agronómico de los cultivos y su valor agregado en el país.

Finalmente, este estudio contribuye con los primeros reportes de virus de plantas a partir de muestras de heces de primates en el Ecuador, como base para futuros proyectos de investigación relacionados con metagenómica. Favoreciendo así, a la implementación de programas de manejo de virus de plantas con tecnologías de secuenciación de nueva generación para reducir el impacto en la producción agrícola y conservar la biodiversidad.

7. CONCLUSIONES

Se analizó mediante un estudio metagenómico los virus de plantas presentes en muestras de heces de primates del Ecuador, reportando datos de virus conocidos para futuras investigaciones.

Se desarrolló una estrategia de análisis bioinformático que permitió preprocesar los datos metagenómicos, considerando un pipeline con el hospedador debido al menor tiempo de

ejecución y mayor cantidad de lecturas. Este pipeline bioinformático optimizado será implementado en la siguiente fase del proyecto.

Se comparó las secuencias obtenidas con bases de datos de referencia para identificar los virus de plantas presentes, mediante la herramienta BLAST con el 90% de identidad.

Se estimó la abundancia y distribución de los virus de plantas de todas las muestras, en donde especies del género *Fabavirus*, *Tobamovirus*, *Alfamovirus*, *Ampelovirus* y *Capillovirus* presentaron mayor abundancia.

8. RECOMENDACIONES

Se recomienda ampliar el número de muestras en estado libre de los primates para realizar análisis de diversidad alfa y beta, con el objetivo de comprender las dinámicas de infección, la evolución viral, el impacto en la agricultura y la biodiversidad.

Se recomienda analizar la posibilidad de falsos positivos o la identificación de virus que no están realmente presentes en la muestra mediante análisis filogenético de las secuencias ensambladas para identificar secuencias que no son congruentes con las especies esperadas.

Se recomienda continuar con estudios relacionados con vigilancia metagenómica viral en plantas, para monitorear y caracterizar la diversidad de virus en diferentes entornos, con el objetivo de detectar y estudiar virus emergentes, así como su impacto en la agricultura.

9. REFERENCIAS BIBLIOGRAFICAS

Aguilar-Bultet, L., & Falquet, L. (2015). Secuenciación y ensamblaje de novo de genomas bacterianos: Una alternativa para el estudio de nuevos patógenos. *Revista de Salud Animal*, 37(2), 125-132.

- Andresen, E. (2005). Interacción entre primates, semillas y escarabajos coprófagos en bosques húmedos tropicales: Un caso de diplocoria. *Universidad y Ciencia, Especial(II)*, 73-84.
- Bağcı, C., Beier, S., Górska, A., & Huson, D. H. (2019). Introduction to the Analysis of Environmental Sequences: Metagenomics with MEGAN. En M. Anisimova (Ed.), *Evolutionary Genomics: Statistical and Computational Methods* (pp. 591-604). Springer.
https://doi.org/10.1007/978-1-4939-9074-0_19
- Balique, F., Colson, P., & Raoult, D. (2012). Tobacco mosaic virus in cigarettes and saliva of smokers. *Journal of Clinical Virology*, 55(4), 374-376.
<https://doi.org/10.1016/j.jcv.2012.08.012>
- Beijerinck, M. (1898). Concerning a contagium vivum fluidum as cause of the spot disease of tobacco leaves. En J. Johnson, *Phytopathological Classics: Vol. VII* (pp. 33-52). American Phytopathological Society.
- Carpino, C., Ferriol Safont, I., Elvira-González, L., Medina, V., Rubio, L., Peri, E., Davino, S., & Galipienso Torregrosa, L. (2020). RNA2-encoded VP37 protein of Broad bean wilt virus 1 is a determinant of pathogenicity, host susceptibility, and a suppressor of post-transcriptional gene silencing. *Molecular Plant Pathology*, 21(11), 1421-1435.
<https://doi.org/10.1111/mpp.12979>
- Chatzivassiliou, E. K. (2021). An Annotated List of Legume-Infecting Viruses in the Light of Metagenomics. *Plants*, 10(7), Article 7. <https://doi.org/10.3390/plants10071413>
- Comas, I., Cancino-Muñoz, I., Mariner-Llicer, C., Goig, G. A., Ruiz-Hueso, P., Francés-Cuesta, C., García-González, N., & González-Candelas, F. (2020). Uso de las tecnologías de secuenciación masiva para el diagnóstico y epidemiología de enfermedades infecciosas.

Enfermedades Infecciosas y Microbiología Clínica, 38, 32-38.

<https://doi.org/10.1016/j.eimc.2020.02.006>

Dávila, S., Castelán, H., Martínez, L., Sánchez, M., Peralta, R., Hernández, A., Dobson, A., Gonzalez, R., Pastor, N., & Batista, R. (2019). A Review on Viral Metagenomics in Extreme Environments. *Front. Microbiol*, 10. <https://doi.org/10.3389/fmicb.2019.02403>

Dombrovsky, A., & Smith, E. (2017). Seed transmission of tobamoviruses: Aspects of global disease distribution. En J. Jimenez, *Advances in Seed Biology* (pp. 233-240). BoD – Books on Demand.

Gama, A. (2017). *Ensamble y exploración de genomas de peces mediante métodos computacionales* [Tesina para obtener el grado de Ingeniero en Informática, Universidad Politécnica de Sinaloa].

<http://repositorio.upsin.edu.mx/formatos/TesinaAndreaGama7300.pdf>

Hull, R. (2002). *Matthews' Plant Virology* (Cuarta, pp. 13-45). Academic Press.

<https://doi.org/10.1016/B978-012361160-4/50053-0>

Hunter, D. E., Darling, H. M., & Beale, W. L. (1969). Seed transmission of potato spindle tuber virus. *American Potato Journal*, 46(7), 247-250. <https://doi.org/10.1007/BF02863021>

Huson, D. H., Auch, A., Qi, J., & Schuster, S. (2007). MEGAN analysis of metagenomic data.

Genome research, 17(3), 377-386. <https://doi.org/doi/10.1101/gr.5969107>

Ilyas, R., Rohde, M. J., Richert-Pöggeler, K. R., & Ziebell, H. (2022). To Be Seen or Not to Be Seen: Latent Infection by Tobamoviruses. *Plants*, 11(16), Article 16.

<https://doi.org/10.3390/plants11162166>

International Committee on Taxonomy of Viruses. (2018). *Genus: Tobamovirus*.

https://ictv.global/ictv-reports/ictv_online_report/positive-sense-rna-viruses/w/virgaviridae/672/genus-tobamovirus

King, A. M. Q., Adams, M. J., Carstens, E. B., & Lefkowitz, E. J. (2012). *Virus taxonomy:*

Classification and nomenclature of viruses. Ninth report of the International Committee on Taxonomy of Viruses (A. M. Q. King, M. J. Adams, E. B. Carstens, & E. J. Lefkowitz, Eds.; p. 1327pp). Elsevier, Amsterdam.

<https://www.sciencedirect.com/book/9780123846846/virus-taxonomy>

López-Pérez, M., & Fernández, F. J. (2013). Metavirómica en masas de agua: Pasado, presente y perspectivas futuras. *Hidrobiológica*, 23(3), 287-302.

Lorenzana López, E. (2020). *Flujo de trabajo bioinformático para el análisis de secuencias de microbioma derivadas de un enfoque metagenómico: Metabiome*.

<https://bibliotecadigital.udea.edu.co/handle/10495/18105>

McHardy, A. C. (2015). PhyloPythia(S). En K. E. Nelson (Ed.), *Encyclopedia of Metagenomics:*

Genes, Genomes and Metagenomes: Basics, Methods, Databases and Tools (pp. 582-585). Springer US. https://doi.org/10.1007/978-1-4899-7478-5_736

McHardy, A. C., Martín, H. G., Tsirigos, A., Hugenholtz, P., & Rigoutsos, I. (2007). Accurate phylogenetic classification of variable-length DNA fragments. *Nature Methods*, 4(1), 63-72. <https://doi.org/10.1038/nmeth976>

Mokili, J., Rohwer, F., & Dutilh, B. (2012). Metagenomics and future perspectives in virus discovery. *Current Opinion in Virology*, 2(1), 63-77.

<https://doi.org/10.1016/j.coviro.2011.12.004>

- Nicaise, V. (2014). Crop immunity against viruses: Outcomes and future challenges. *Frontiers in Plant Science*, 5. <https://doi.org/10.3389/fpls.2014.00660>
- Nooij, S., Schmitz, D., Vennema, H., Kroneman, A., & Koopmans, M. P. G. (2018). Overview of Virus Metagenomic Classification Methods and Their Biological Applications. *Frontiers in Microbiology*, 9. <https://doi.org/10.3389/fmicb.2018.00749>
- Pop, M. (2009). Genome assembly reborn: Recent computational challenges. *Briefings in Bioinformatics*, 10(4), 354-366. <https://doi.org/10.1093/bib/bbp026>
- Roossinck, M. J. (2011). The good viruses: Viral mutualistic symbioses. *Nature Reviews Microbiology*, 9(2), 99-108. <https://doi.org/10.1038/nrmicro2491>
- Roossinck, M. J. (2012). Plant Virus Metagenomics: Biodiversity and Ecology | Annual Reviews. *Annual Reviews of Genetics*, 46, 359-369.
- Roossinck, M. J. (2013). Plant Virus Ecology. *PLOS Pathogens*, 9(5), e1003304. <https://doi.org/10.1371/journal.ppat.1003304>
- Roossinck, M. J., & García, F. (2015). Ecosystem simplification, biodiversity loss and plant virus emergence. *Current Opinion in Virology*, 10, 56-62. <https://doi.org/10.1016/j.coviro.2015.01.005>
- Roossinck, M. J., Martin, D. P., & Roumagnac, P. (2015). Plant Virus Metagenomics: Advances in Virus Discovery. *Phytopathology*, 105(6), 716-727. <https://doi.org/10.1094/PHYTO-12-14-0356-RVW>
- Roossinck, M. J., Saha, P., Wiley, G. B., Quan, J., White, J. D., Lai, H., Chavarría, F., Shen, G., & Roe, B. A. (2010). Ecogenomics: Using massively parallel pyrosequencing to understand virus ecology. *Molecular Ecology*, 19(s1), 81-88. <https://doi.org/10.1111/j.1365-294X.2009.04470.x>

- Sanfaçon, H. (2015). Secoviridae: A Family of Plant Picorna-Like Viruses with Monopartite or Bipartite Genomes. En *eLS* (pp. 1-14). John Wiley & Sons, Ltd.
<https://doi.org/10.1002/9780470015902.a0000764.pub3>
- Sanfaçon, H., Iwanami, T., Karasev, A., Van der Vlugt, T., Wellink, J., Wetzel, J., & Yoshikawa, N. (2012). Family Secoviridae. En A. M. Q. King, M. J. Adams, E. B. Carstens, & E. J. Lefkowitz, *Virus taxonomy. Ninth report of the International Committee on Taxonomy of Viruses*. (pp. 881-899). Elsevier Academic.
- Sharma, V. K., Kumar, N., Prakash, T., & Taylor, T. D. (2012). Fast and Accurate Taxonomic Assignments of Metagenomic Sequences Using MetaBin. *PLOS ONE*, 7(4), e34030.
<https://doi.org/10.1371/journal.pone.0034030>
- Susi, H., Filloux, D., Frilander, M. J., Roumagnac, P., & Laine, A.-L. (2019). Diverse and variable virus communities in wild plant populations revealed by metagenomic tools. *PeerJ*, 7, e6140. <https://doi.org/10.7717/peerj.6140>
- Tennant, P., Fermin, G., & Foster, J. E. (2018). *Viruses: Molecular Biology, Host Interactions, and Applications to Biotechnology*. Academic Press.
- Thomas, T., Gilbert, J., & Meyer, F. (2012). Metagenomics—A guide from sampling to data analysis. *Microbial Informatics and Experimentation*, 2(1), 3.
<https://doi.org/10.1186/2042-5783-2-3>
- Urbina, N. S. U. (2010). *Patrón de actividad, dieta y dispersión de semillas por tres especies de primates en un fragmento de bosque en San Juan de Carare, Santander (Colombia)*. [Trabajo de grado para optar al título de Biólogo, Pontificia Universidad Javeriana].
<http://hdl.handle.net/10554/8669>

Wilson, G., Bryan, J., Cranston, K., Kitzes, J., Nederbragt, L., & Teal, T. K. (2017). Good enough practices in scientific computing. *PLOS Computational Biology*, *13*(6), e1005510. <https://doi.org/10.1371/journal.pcbi.1005510>

Yamaya, J., Yoshioka, M., Meshi, T., Okada, Y., & Ohno, T. (1988). Expression of tobacco mosaic virus RNA in transgenic plants. *Molecular and General Genetics MGG*, *211*(3), 520-525. <https://doi.org/10.1007/BF00425710>