



Identificación de virus fitopatógenos en Estátice (*Limonium sinuatum* (L.) Mill.) mediante la aplicación de secuenciación de alto rendimiento (HTS-High Throughput Sequencing)

Taípe Bolaños, Marco Vinicio

Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Maestría de Investigación en Biotecnología Vegetal

Trabajo de titulación, previo a la obtención del título de Magister en Biotecnología Vegetal

Flores Flor, Francisco Javier, Ph.D

21 de diciembre de 2022



Tesis_maestria_BIOVEG_MarcoTaipe fjf_02_SinBiblio.docx

Scanned on: 14:40 December 11, 2022 UTC



Overall Similarity Score



Results Found



Total Words in Text

Identical Words	172
Words with Minor Changes	6
Paraphrased Words	523
Omitted Words	0



Firmado electrónicamente por:

FRANCISCO
JAVIER FLORES
FLOR



Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Certificación

Certifico que el trabajo de titulación: **“Identificación de virus fitopatógenos en Estátice (*Limonium sinuatum* (L.) Mill.) mediante la aplicación de secuenciación de alto rendimiento (HTS-High Throughput Sequencing)”** fue realizado por el señor **Taípe Bolaños, Marco Vinicio** el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizado en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 12 de diciembre de 2022

Firma:



firmas asociadas por:
FRANCISCO
JAVIER FLORES
FLOR

Flores Flor Francisco Javier, PhD.

Director

C.C.: 1713443479



Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Responsabilidad de Autoría

Yo **Taipe Bolaños, Marco Vinicio** con cédula de ciudadanía n° 1712485091, declaro que el contenido, ideas y criterios del trabajo de titulación: "**Identificación de virus fitopatógenos en Estátice (*Limonium sinuatum* (L.) Mill.) mediante la aplicación de secuenciación de alto rendimiento (HTS-High Throughput Sequencing)**", es de mi autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 21 de diciembre de 2022

Taipe Bolaños, Marco Vinicio

C.C.: 1712485091



Vicerrectorado de Investigación, Innovación y Transferencia de Tecnología

Centro de Posgrados

Autorización de Publicación

Yo **Taipe Bolaños, Marco Vinicio** con cédula de ciudadanía n° 1712485091, autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: "Identificación de virus fitopatógenos en Estátice (*Limonium sinuatum* (L.) Mill.) mediante la aplicación de secuenciación de alto rendimiento (HTS-High Throughput Sequencing)" en el Repositorio Institucional, cuyo contenido, ideas y criterios son de mi responsabilidad.

Sangolquí, 21 de diciembre de 2022

Taipe Bolaños, Marco Vinicio

C.C.: 1712485091

Dedicatoria

Dedico el presente trabajo a mi familia, mi esposa Enith Jacqueline, mis tres amados hijos Gabriela, José y Daniel, quienes han sido el soporte fundamental en la obtención y culminación del presente trabajo.

A mis padres Marco y Beatriz, mis hermanos Amparo, Lucía, Roberto y Patricio.

A todos quienes con su apoyo y palabras de aliento aportaron significativamente para lograr conseguir este objetivo.

Agradecimiento

A Dios, por darme vida, salud y fuerza durante este tiempo de estudio para terminar con éxito el proyecto de investigación.

A mi familia, Enith, Gabriela, José y Daniel, quienes me apoyaron incondicionalmente en este proceso, y están siempre a mi lado.

A todos mis familiares y amigos, que han apoyado desinteresadamente el proceso de ejecución del trabajo, brindándome sus palabras de aliento.

A Francisco Flores, compañero y amigo, director de mi tesis de maestría, por su constante apoyo y preocupación en el desarrollo del proyecto.

A Carlos Noceda y Ligia Ayala, por su sincera amistad y soporte.

A todos los colegas que de una u otra forma, colaboraron en mi investigación y apoyaron la iniciativa del presente trabajo: Karina Proaño, Pablo Landázuri, Valeria Ochoa, Mónica Jadán, Blanquita Naranjo y Gabriela Miño.

A todos mis profesores de la maestría, por guiarme con sabiduría y brindarme su conocimiento.

A la Universidad de las Fuerzas Armadas ESPE y Centro de Posgrados, por brindarme la oportunidad de realizar el estudio investigativo.

A la Universidad Tecnológica Equinoccial, por el apoyo en la publicación del paper en la revista indexada.

A la Corporación Ecuatoriana para el Desarrollo de la Investigación y la Academia CEDIA, por facilitarme el uso del *cluster* para ejecutar el estudio bioinformático.

Índice de contenidos

Resumen.....	13
Abstract.....	14
Capítulo 1.....	15
Introducción.....	15
Tema.....	15
Formulación del problema.....	15
Objetivos de la investigación.....	16
<i>Objetivo general</i>	16
<i>Objetivos específicos</i>	16
Marco teórico.....	17
Floricultura en Ecuador.....	17
Estátice: <i>Limonium sinuatum</i> (L.) Mill.....	17
Métodos de detección de virus fitopatógenos.....	18
Detección por Sintomatología.....	19
Detección por Serología.....	19
Detección mediante métodos moleculares.....	20
<i>Hibridación de ácidos nucleicos</i>	20
<i>Reacción en cadena de la polimerasa (PCR)</i>	20
<i>Secuenciación de alto rendimiento (HTS o NGS)</i>	22
Metagenómica de virus en plantas.....	23

Análisis bioinformático metagenómico	23
Ensamblaje de genomas metagenómicos.....	24
Virus fitopatógenos	25
Género tospovirus (Orthotospovirus).....	25
<i>Alstroemeria necrotic streak virus</i>	26
Análisis filogenéticos (Filogenia)	27
Métodos de construcción de árboles filogenéticos.....	27
Anotación genómica.....	27
Hipótesis	28
Capítulo 2.....	29
Materiales y métodos.....	29
Zona de estudio.....	29
Obtención de muestras vegetales.	29
Extracción de ARN total.	29
Secuenciación y preparación de librerías.....	30
Análisis bioinformático.	30
Control de calidad de las muestras secuenciadas.	31
Remoción de adaptadores y lecturas de baja calidad	32
Eliminación de lecturas duplicadas.....	32
Ensamblaje de secuencias genómicas de virus fitopatógenos.....	33
Alineamiento con bases de datos de secuencias de referencia para virus, (Blasteo o Mapeo)	33

	10
Visualización de resultados en KronaTools.....	34
Ensamblaje final del virus encontrado.....	34
Anotación de virus.....	35
Análisis filogenético: reconstrucción histórica del virus identificado.....	35
Capítulo 3.....	37
Resultados.....	37
Características de la muestra secuenciada.....	37
Control de la calidad inicial de la muestra secuenciada.....	38
Reporte final de los datos depurados y limpios.....	41
Visualización de secuencias virales en KronaTools.....	44
Características de los segmentos virales de <i>Alstroemeria necrotic streak virus</i> ANSV ensamblado.....	45
Anotación del genoma viral identificado.....	46
Reconstrucción histórica evolutiva de <i>Alstroemeria necrotic streak virus</i> ANSV.....	48
Capítulo 4.....	52
Discusión.....	52
Capítulo 5.....	57
Conclusiones.....	57
Recomendaciones.....	59
Bibliografía.....	60

Índice de tablas

Tabla 1 <i>Accesiones de virus del genero Orthospovirus seleccionados del GenBank para el análisis filogenético.....</i>	36
Tabla 2 <i>Características de la muestra secuenciada (M69) de hojas de estática Limonium sinuatum, enviado por la Compañía ©Macrogen Inc. Corea del Sur.</i>	38
Tabla 3 <i>Resultados de la identificación de organismos presentes en muestras de Limonium sinuatum (datos de secuenciación y de referencia para virus del GenBank)</i>	45
Tabla 4 <i>Características de los segmentos del virus de Alstroemeria necrotic streak virus ANSV ensamblado y contrastado con las secuencias de los segmentos de accesiones de referencia del virus de ANSV del GenBank.....</i>	46
Tabla 5 <i>Anotación para los segmentos del virus de Alstroemeria necrotic streak virus (ANSV) identificado en la muestra foliar de Limonium sinuatum.....</i>	47
Tabla 6 <i>Resultado del mejor modelo evolutivo, obtenido en MEGA-X V10.2.6 para las 30 secuencias genómicas del gen N de la nucleocapside</i>	49

Índice de figuras

Figura 1 <i>Flujo de trabajo (pipeline) para el análisis bioinformático.</i>	31
Figura 2 <i>Lectura con Bioanalyzer (Macrogen Inc.)</i>	37
Figura 3 <i>Control de la calidad inicial y final de archivos secuenciados</i>	39
Figura 4 <i>Calidad del parámetro contenido de secuencias por base</i>	40
Figura 5 <i>Calidad del parámetro contenido GC por secuencia</i>	41
Figura 6 <i>Reporte final de la calidad de los archivos depurados</i>	42
Figura 7 <i>Reporte final para la calidad de datos, archivos M69_1.Rded.fq y M69_2.Rded.fq ...</i>	43
Figura 8 <i>Diagrama con la distribución en porcentajes de los virus encontrados en la muestra de Limonium sinuatum (M69) con el programa KronaTools</i>	44
Figura 9 <i>Gráfico para el segmento L del virus de Alstroemeria necrotic streak virus ANSV-L.</i>	47
Figura 10 <i>Gráfico para el segmento M del virus de Alstroemeria necrotic streak virus ANSV-M.</i>	48
Figura 11 <i>Gráfico para el segmento S del virus de Alstroemeria necrotic streak virus ANSV-S.</i>	48
Figura 12 <i>Árbol filogenético con el valor más alto de probabilidad logarítmica para el virus ANSV El Quinche ON149271</i>	50
Figura 13 <i>Árbol filogenético de la inferencia bayesiana para la probabilidad posterior de la muestra de ANSV El Quinche ON149271</i>	51

Resumen

La detección de virus fitopatógenos por medio del análisis de datos de secuenciación de alto rendimiento (*HTS-High Throughput Sequencing*) y el análisis con herramientas bioinformáticas se ha incrementado en los últimos años. La presente investigación se centró en la aplicación de este método de diagnóstico para el reconocimiento de virus fitopatógenos a partir de muestras foliares de plantas de flores de corte de *Limonium sinuatum* (L.) Mill, una flor de relleno para florero cultivada en plantaciones florícolas del Ecuador. Con el flujo de trabajo utilizado, fue posible ensamblar el genoma completo de *Alstroemeria Necrotic Streak Virus* (ANSV), un virus del género *Orthospovirus* conformado por tres segmentos (*L*, *M* y *S*) y cuyas longitudes fueron de 8755 pb, 4869 pb y 3135 pb respectivamente. La anotación de este virus fue enviada al repositorio de GenBank y registrada con números de acceso *ON149273*, *ON149272* y *ON149271* para cada uno de los segmentos del virus. La reconstrucción histórica evolutiva realizada con los métodos de máxima verosimilitud e inferencia bayesiana, utilizando los aminoácidos del gen *N* de la nucleocapside de los *orthospovirus* registrados en GenBank, ubica al virus ensamblado de ANSV de la muestra recolectada en Ecuador dentro del mismo clado que las accesiones reportadas de estudios sobre el virus de ANSV en Colombia. Se presume que el agente causal de la sintomatología de virus presentada en campo para las hojas de *Limonium sinuatum* sea provocada por el virus de ANSV detectado en el análisis HTS, lo que pone en alerta la necesidad de implementar planes de controles fitosanitarios integrales a nivel de pequeños, medianos y grandes agricultores.

Palabras clave: Secuenciación de alto rendimiento HTS, *Limonium sinuatum*, *Alstroemeria necrotic streak virus*, Virus fitopatógeno, Ensamblaje

Abstract

The detection of phytopathogenic viruses through the analysis of High Throughput Sequencing data (HTS) and the analysis with bioinformatics tools has increased in recent years. The present investigation was focused on the application of this diagnostic method for the recognition of phytopathogenic viruses from foliar samples of cut flower plants of *Limonium sinuatum* (L.) Mill, a vase filler flower grown in flower plantations in Ecuador. With the workflow used, it was possible to assemble the complete genome of Alstroemeria Necrotic Streak Virus (ANSV), a virus of the *Orthospovirus* genus made up of three segments (*L*, *M* and *S*) and whose lengths were 8755 bp, 4869 bp and 3135 bp respectively. The annotation of this virus was sent to the GenBank repository and registered with accession numbers ON149273, ON149272 and ON149271 for each of the virus segments. The historical evolutionary reconstruction carried out with the methods of Maximum Likelihood and Bayesian Inference, using the amino acids of the N gene of the orthospoviruses's nucleocapsid registered in GenBank, that places the assembled ANSV virus from the sample collected in Ecuador within the same clade as the accessions reported from studies on the ANSV virus in Colombia. It is presumed that the causative agent of the virus symptoms presented in the field for the leaves of *Limonium sinuatum* is caused by the ANSV virus detected in the HTS analysis, which alerts the need to apply comprehensive phytosanitary control plans for small, medium and large farmers.

Keywords: High Throughput Sequencing HTS, *Limonium sinuatum*, *Alstroemeria necrotic streak virus* ANSV, Phytopathogens viruses, Assembly

Capítulo 1

Introducción

Tema

Identificación de virus fitopatógenos en Estátice (*Limonium sinuatum* (L.) Mill.) mediante la aplicación de secuenciación de alto rendimiento (HTS-High Throughput Sequencing)

Formulación del problema

Las tecnologías de secuenciación de alto rendimiento (*HTS High throughput sequencing*) han revolucionado la investigación de enfermedades de plantas y cada vez son más utilizadas para el diagnóstico de virus de plantas (Olmos et al., 2018). La secuenciación de nueva generación (HTS), también conocida como secuenciación masiva paralela o secuenciación de alto rendimiento, incluye métodos que ayudan a detectar e identificar en plantas, infecciones virales en etapas tempranas de desarrollo, utilizando herramientas bioinformáticas para obtener genomas completos de virus (Jaramillo Mesa, 2017).

La floricultura es considerada como uno de los sectores agrícolas más importantes de la sierra ecuatoriana y de la economía del país, es la tercera actividad agrícola que más divisas genera después del banano y el camarón, representando aproximadamente el 1 % del total del Producto Interno Bruto (PIB) (Paredes Barros, 2019).

El cultivo de flores de corte o de verano, es una actividad realizada por varias empresas florícolas además de medianos y pequeños agricultores, quienes cultivan estas especies consideradas de complemento para el florero, una de estas especies es el estátice (*Limonium sinuatum* (L.) Mill.). Si bien el diagnóstico de enfermedades es llevado por técnicos especializados, los métodos utilizados son los tradicionales de laboratorio y basados en la evaluación de la sintomatología que presenta la planta hospedera, lo que puede ser impreciso y

muchas veces no permite establecer la presencia del agente patógeno como bacterias, hongos o virus.

La presente investigación se centró en la aplicación de métodos de diagnóstico para el reconocimiento y detección de virus fitopatógenos mediante el análisis de datos de secuenciación de alto rendimiento (*HTS*). Los virus son difíciles de estudiar mediante los métodos tradicionales, por lo que el uso de estas nuevas tecnologías generan mayor conocimiento de la etiología sintomática asociada a infecciones virales de plantas, lo que sin duda trasciende en la toma de decisiones e implementación eventual de controles fitosanitarios. Además, se pretende reconstruir la historia evolutiva de los posibles virus encontrados en las muestras de hojas de plantas de estátice *Limonium sinuatum* (L.) Mill, lo que permitirá entender mejor su epidemiología.

Objetivos de la investigación

Objetivo general

Identificar virus fitopatógenos en Estátice (*Limonium sinuatum* (L.) Mill.) mediante la aplicación de secuenciación de alto rendimiento (*HTS-High Throughput Sequencing*)

Objetivos específicos

- Ensamblar secuencias genómicas de virus fitopatógenos presentes en muestras foliares de plantas de *Limonium sinuatum* (L.) Mill, aplicando herramientas bioinformáticas.
- Anotar los genomas de los virus fitopatógenos presentes en muestras foliares de plantas de *Limonium sinuatum* (L.) Mill. aplicando métodos basados en homología.
- Reconstruir la historia evolutiva de los virus encontrados utilizando las secuencias de sus genomas.

Marco teórico

Floricultura en Ecuador

En Ecuador, uno de los principales productos no tradicionales de exportación son las flores, entre enero y noviembre de 2018 las exportaciones de estas suman USD 792.7 millones, lo que representa el 4% del total de las exportaciones ecuatorianas (Castro et al., 2020).

El nombre de flores de verano o también flores de corte, de sebe a que en ciertos países del globo, estas flores se cultivan solo en la estación de verano. En nuestro país, las condiciones climáticas favorecen el florecimiento de estas flores durante todo el año cultivándose en altitudes de entre 2200 y 2700 m.s.n.m. produciendo flores atractivas y de calidad (PRO-ECUADOR, 2015).

Algunos tipos de flores que se cultivan en Ecuador son: Alstroemeria, Anastasia, Aster, Calas, Campanula, Cartuchos, Chrysanthemums, Claveles, Craspedia, Delphinium, Dianthus, Ginger, Girasol, Gypsophila, Heliconias, Hypericum, Lilium, Limonium, Statice entre otras más (PRO-ECUADOR, 2015).

Las exportaciones de flores de corte ecuatorianas, de acuerdo a lo registrado durante 2014, tienen destinos como: Estados Unidos (54%); Unión Europea (20%); Rusia (7%) y por último los demás países con 19% (PRO-ECUADOR, 2015).

Estátice: *Limonium sinuatum* (L.) Mill.

El estátice es originario del Mediterráneo (Larson, 2012), el *Limonium sinuatum* pertenece a la familia de las Plumbaginaceae de la que podemos encontrar gran número de variedades en toda la cuenca del mediterráneo (Pedros, 1994). El género *Limonium* comprende más de 300 especies ornamentales, muchas de porte bajo y perennes conocidas como “siempreviva azul” (de Blas, 1992).

Su clasificación taxonómica es: Reino: *Plantae*; Subreino: *Viridiplantae*; División: Tracheophyta; Subdivisión: Spermatophytina; Clase: Magnoliopsida; Superorden: Caryophyllanae; Orden: Caryophyllales; Familia: *Plumbaginaceae*; Género: *Limonium*; Especie: *sinuatum*. * **Fuente:** Integrated Taxonomic Information System (ITIS), www.itis.gov.

Las flores tanto permanentes como transitorias se cultivan en invernaderos y a campo abierto, según estudio realizado en 2011 por el Instituto Nacional de Estadística y Censos (INEC) ESPAC la producción de flores permanentes en invernadero es de 4393 ha y en campo abierto de 1396 ha (Riera Yépez & Zuleta Colimba, 2013).

A nivel nacional, Ecuador produce 21% de flores de verano, la provincia con mayor superficie cultivada es Pichincha (661.70 ha); Cotopaxi, Azuay e Imbabura en conjunto poseen un área cultivada de 188.90 ha (PRO-ECUADOR, 2015).

Algunas de las variedades de flores de corte estática que se cultivan en Ecuador son: Pedro White, Cristal Dark Blue, Cristal Lilac, Ara Purple, Ara Blue, Ara Lavander, Ara Violet entre otras.

Métodos de detección de virus fitopatógenos

La detección e identificación de virus en plantas, debe ser efectiva, conveniente y rápida (Jeong et al., 2014). El control de las enfermedades virales se dirige al control indirecto del patógeno como es el combate al insecto vector, la eliminación de plantas enfermas o evitar la siembra de semillas infectadas (González-Garza, 2017).

Entre los métodos desarrollados en la detección de virus en plantas están la observación microscópica, técnicas serológicas (ELISA), métodos moleculares (PCR, HTS-secuenciación de alto rendimiento), entre otros (Jeong et al., 2014).

Detección por Sintomatología

La observación de la sintomatología que los patógenos virales provocan en las plantas es la primera forma de detección e identificación de los mismos. Entre 1900 a 1935, las inclusiones virales eran identificadas y descritas en el interior de las células con la ayuda de un microscopio. (González-Garza, 2017).

Es necesario tener mucha experiencia en la observación de la sintomatología viral para identificar tipos de enfermedades que provocan los virus en las plantas, además de realizar otras pruebas adicionales para tener un correcto diagnóstico (González-Garza, 2017; Naidu, 2001).

Detección por Serología

La detección por serología de virus en vegetales, tiene sus inicios en el método de precipitación, que consiste en la reacción del antisuero (*antisuero específico de virus conocido*) y el antígeno (*gota de sabia infectiva de una planta*) en un medio líquido, la reacción es observable a simple vista porque forma precipitados o aglutinados (González-Garza, 2017; Kleczkowski, 1965; Matthews, 2012).

Los métodos serológicos de precipitación usados por varios años, han sido sustituidos por los métodos de ensayo de inmunoabsorción ligado a enzimas (enzyme-linked immunosorbent assay) o ELISA por sus siglas en inglés (González-Garza, 2017).

Entre los métodos de ELISA más utilizados se encuentran: ELISA en placa, ELISA en membrana (immunoblotting), inmunoabsorción en microscopía electrónica (ISEM por sus siglas en inglés) e inmunocromatografía o tiras reactivas (Immunochromatographic or Strip Tests) (Hull, 2009; Matthews, 2012).

En patología vegetal se utilizan comúnmente los métodos de ELISA en placa, siendo los más conocidos: ELISA de sandwich de doble anticuerpo (DAS double antibody sandwich), ELISA de sandwich de triple anticuerpo (TAS Triple antibody sandwich) y ELISA de placa sensibilizada con antígeno (ACP antigen coated plate) (González-Garza, 2017). En la actualidad los diversos tipos de diagnóstico ELISA, siguen siendo los más utilizados en la detección de virus tanto en humanos como animales y vegetales (González-Garza, 2017).

Detección mediante métodos moleculares

Los métodos moleculares son alternativos a los métodos serológicos, se utilizan en el laboratorio por su alta precisión y sensibilidad en la detección de enfermedades virales siendo necesario contar con información genética de los virus (Jeong et al., 2014).

Hibridación de ácidos nucleicos

Este tipo de detección es común de utilizarla en virus de plantas, los ácidos nucleicos a prueba son inmovilizados en membranas de nylon o nitrocelulosa, técnica conocida como hibridación por puntos o manchas (Owens & Diener, 1984).

El ácido nucleico viral de la muestra se inmovilizado sobre una membrana de nitrocelulosa con carga positiva, los sitios libres de unión de la membrana se bloquean con un DNA no homólogo o con proteína, la hibridación ocurre entre la unión del ácido nucleico viral y la sonda, la secuencia final es detectada por la molécula reportera (González-Garza, 2017).

Reacción en cadena de la polimerasa (PCR)

La reacción en cadena de la polimerasa (PCR por sus siglas en inglés), es una reacción enzimática in vitro que amplifica millones de veces una secuencia específica de ADN (secuencia blanco) o de ADNc (ADN complementario) durante varios ciclos repetidos (Tamay de Dios et al., 2013). Se fundamenta en la actividad enzimática de la *ADN polimerasa* para

sintetizar naturalmente el ADN en las células, la técnica reúne tres etapas principales: desnaturalización, hibridación y extensión. (Naidu & Hughes, 2003; Tamay de Dios et al., 2013).

La técnica PCR es usada en el diagnóstico e identificación de enfermedades virales de plantas (Rodríguez Sánchez & Barrera Saldaña, 2004). En el caso de los virus con genoma de ARN, previo a una PCR, es necesario realizar un paso denominado retrotranscripción, transcripción reversa o también transcripción inversa, en donde una enzima, la transcriptasa inversa o retrotranscriptasa, transcribe el ARN a un ADN complementario (ADNc) (González-Garza, 2017).

Cuando en la reacción utilizamos como sustrato ADN genómico, hablamos de una PCR convencional, pero si en la reacción utilizamos ADN complementario (ADNc) se conoce a la técnica como RT-PCR (Reverse Transcription-PCR) (González-Garza, 2017; Naidu & Hughes, 2003).

La técnica de PCR ha evolucionado en los últimos tiempos, al punto de evitar el uso del gel de agarosa. La técnica se denomina PCR Tiempo Real o PCR cuantitativa (qPCR por sus siglas en inglés), el resultado de la detección del producto amplificado se visualiza en un computador (González-Garza, 2017).

La técnica PCR Anidada (Nested PCR por sus siglas en inglés), es una técnica más sensible, en esta se utilizan dos rondas de amplificación con distintos pares de *primers* en cada una, la primera amplificación se realiza con *primers* externos y el producto obtenido se envía a una segunda amplificación con *primers* internos (Rodríguez Sánchez & Barrera Saldaña, 2004).

Otra variante de la PCR convencional es la PCR multiplex, con esta técnica se pueden detectar dos o más virus en la misma reacción, la técnica utiliza dos o más pares de *primers* y acorta el tiempo de detección (Rodríguez Sánchez & Barrera Saldaña, 2004).

Secuenciación de alto rendimiento (HTS o NGS)

El reciente desarrollo de técnicas de secuenciación de alto rendimiento (HTS por sus siglas en inglés), también llamada secuenciación de nueva generación (NGS por sus siglas en inglés), secuenciación masiva o secuenciación profunda, ha revolucionado la investigación sobre organismos asociados a plantas, estas técnicas se están empezando a utilizar para el diagnóstico de plagas, en particular de virus (Olmos et al., 2018).

Las tecnologías HTS son de amplio espectro y genéricas, pueden reemplazar métodos convencionales de detección de patógenos, HTS permite la secuenciación simultánea, y por lo tanto, la detección de cualquier organismo presente en una muestra. Las tecnologías HTS tienen la capacidad de revelar en cualquier muestra de planta la presencia de organismos inesperados y desconocidos que podrían ser amenazas potenciales (Olmos et al., 2018).

La tecnología HTS puede detectar, identificar y descubrir nuevos virus en plantas sin necesidad de anticuerpos. La técnica permite secuenciar genomas virales completos o parciales, a partir de ácidos nucleicos totales que se obtienen de plantas sintomáticas o asintomáticas a una infección viral (Prabha et al., 2013).

La virología vegetal, se ha beneficiado del avance tecnológico computacional y bioinformático que utiliza las HTS, principalmente en las áreas de secuenciación de genomas, descubrimiento, epidemiología, ecología, transcriptomas, replicación, detección e identificación de virus y viroides. La tecnología de HTS secuenciando genomas de forma directa o ensamblando pequeños fragmentos de ARN (ADNc, ADN, ARNsi) han identificado nuevos virus y viroides (Barba et al., 2014).

Existen cuatro plataformas principales de secuenciación masiva, Illumina (Solexa) sequencing, Roche 454 sequencing, Ion torrent: Proton / PGM sequencing y SOLiD

sequencing, que generan miles de millones de secuencias cortas de entre 25 a 400 pb que se analizan a través de estudios metagenómicos (González-Garza, 2017).

Metagenómica de virus en plantas

En la metagenómica viral, tres procesos son los principales: el muestreo y recolección de virus en tejidos o muestras ambientales, la secuenciación de ácidos nucleicos (empleando básicamente técnicas HTS), y finalmente el análisis bioinformático, este último interpreta de manera adecuada los datos obtenidos y proporciona información sobre la diversidad viral asociada con la muestra. (Pantaleo & Chiumenti, 2018).

La metagenómica viral recupera y analiza los genomas completos o parciales de todos los potenciales virus presentes en una muestra. La metagenómica de virus, aplica procedimientos para remover células bacterianas y del hospedero, tratamientos para remover ácidos nucleicos libres, amplificación con cebadores aleatorios utilizando transcripción inversa y RT-PCR, para finalmente realizar la secuenciación masiva paralela (Nieuwenhuijse & Koopmans, 2017).

La tasa de descubrimiento de virus se ha visto incrementada gracias al uso de la metagenómica viral, las tecnologías HTS y el desarrollo de herramientas bioinformáticas de ensamblaje genómico, generando información sobre nuevos virus (Alavandi & Poornima, 2012).

Análisis bioinformático metagenómico

Los análisis bioinformáticos de los metagenomas virales, intentan resolver tres preguntas: cuántos virus hay (diversidad), ¿Qué son (taxonomía) y qué están haciendo (función)? (Willner, 2010). Los datos de secuenciación generados deben ensamblarse primero.

Las tecnologías de secuenciación metagenómica difieren en los métodos de preparación de las bibliotecas “*shotgun library*” y la longitud de las lecturas producidas. Un

enfoque importante es construir bibliotecas básicas virales para la secuenciación por métodos como Sanger (Breitbart & Rohwer, 2005).

El método de secuenciación de *Sanger* produce datos de secuencia de alta calidad con longitudes de lectura de alrededor de 400 pares de bases (Alavandi & Poornima, 2012). El sistema *454 Life Sciences GSFLX* proporciona longitudes de lectura que van desde 250 a 400 pares de bases. Tecnologías de mayor rendimiento como *Solexa/Illumina* y los sistemas *SOLiD* ya están disponibles y pueden proporcionar datos de hasta 3 a 6 Gb por ejecución, la longitud media de lectura obtenida con estos sistemas es de aproximadamente de 50 a 100 pb (Pareek et al., 2011).

El *screen* bioinformático identifica y filtra hospederos y otras secuencias no virales que podrían estar incluidas en el metagenoma. Software como *MEGAN*, *PathSeq* y *CAMERA*, entre otros, están disponibles para el análisis de secuencias metagenómicas (Alavandi & Poornima, 2012).

Varios paquetes de software pueden analizar y visualizar rápidamente los resultados de *BLAST*, y la asignación taxonómica. Herramientas bioinformáticas comparan las secuencias virales generadas con secuencias conocidas mantenidas en un base de datos anotada como *NCBI* utilizando programas como *Blastn* y *Blastx* para asignación taxonómica (Alavandi & Poornima, 2012).

Ensamblaje de genomas metagenómicos

Previa clasificación o asignación taxonómica, es necesario ensamblar las lecturas cortas en secuencias continuas (*contigs*) de una longitud de nucleótidos mayor y generar secuencias consenso al alinear dichos *contigs*. El ensamblaje de genomas se puede realizar de dos formas: 1) en base al mapeo de lecturas frente a un genoma de referencia o 2) el llamado

ensamblaje de novo, que consiste en unir las lecturas en base a solapamientos, frecuencia o lecturas de dos sentidos (Paired-end, PE) (Van der Walt et al., 2017).

Dado que para muchos virus no existen secuencias de referencia completas y curadas, el ensamblaje de novo es la mejor opción en análisis de virus. El ensamblaje de novo de metagenomas facilita el descubrimiento y reconstrucción de genomas nuevos gracias a una predicción de genes más confiable (Van der Walt et al., 2017).

Virus fitopatógenos

Los virus de plantas, son considerados de gran relevancia después de los hongos y se difunden a cultivos cercanos rápidamente, son patógenos de importancia económica estimándose pérdidas en miles de millones de dólares por año alrededor del mundo (Hull, 2009; Jeong et al., 2014).

Género tospovirus (Orthotospovirus)

El género de virus *Tospovirus* que infectan a plantas, toma su nombre del virus del marchitamiento del tomate (*Tomato spotted wilt virus*), primera especie de virus descrita en 1915 (Health, 2012). El nombre del género aceptado actualmente por el comité internacional de taxonomía de virus (ICTV por sus siglas en inglés) es Orthotospovirus (King et al., 2011).

Los *orthotospovirus* son capaces de infectar a un gran número de especies vegetales, tanto cultivos alimenticios como ornamentales, se consideran patógenos agrícolas y representan un grave riesgo para la seguridad alimentaria en todo el mundo (Health, 2012; Oliver & Whitfield, 2016).

Estos virus generalmente se transmiten de manera persistente y propagativa por insectos del orden *Thysanoptera*, pero también pueden iniciar la infección por medios mecánicos (Oliver & Whitfield, 2016). La infección produce necrosis de tejidos en hojas y frutos, marchitamiento, reducción del crecimiento vegetativo y muerte de la planta hospedera (Health, 2012).

Orthospovirus es un género de la familia *Tospoviridae*, comprende virus envueltos con forma esférica o viriones pleomórficos de alrededor de 80 a 120 nm de diámetro y un genoma tripartito de ssRNAs negativos o ambisentidos de aproximadamente 16,6 kb (King et al., 2011). Tienen un genoma de ARN monocatenario tripartito, clasificados según el tamaño de la molécula en pequeño *S (Small)*, mediano *M (Medium)* y grande *L (Large)* (de Oliveira et al., 2011; Nichol et al., 2005).

Alstroemeria necrotic streak virus

El virus *Alstroemeria necrotic streak virus (ANSV)* pertenece al género *Orthospovirus*, en Colombia fue reportado sobre plantas ornamentales de *Alstroemeria*. El virus está relacionado con miembros de *orthospovirus* americano y se ha demostrado que infecta naturalmente a lulo (*Solanum quitoense*), tomate (*Solanum lycopersicum*) y pimiento (*Capsicum annuum*) (Gallo et al., 2019).

La clasificación taxonómica de *Alstroemeria necrotic streak virus (ANSV)* de acuerdo al ICTV es la siguiente: Dominio: *Riboviria*; Reino: *Orthornavirae*; Filo: *Negarnaviricota*; Subfilo: *Polyploviricotina*, Clase: *Ellioviricetes*, Orden: *Bunyavirales*, Familia: *Tospoviridae*, Género: *Orthospovirus*, Especie: *Alstroemeria necrotic streak orthospovirus*. * **Fuente:** "International Committee on Taxonomy of Viruses (ICTV): <https://ictv.global/taxonomy/>"

Los síntomas asociados con la infección por ANSV incluyen rayas necróticas en hojas y tallos, anillos necróticos y necrosis venal (Gallo et al., 2019).

Análisis filogenéticos (Filogenia)

Los análisis filogenéticos, filogenia, reconstrucción evolutiva o histórica, hace referencia al estudio las relaciones evolutivas de grupos taxonómicos, secuencias o especies que permite la construcción de un árbol filogenético que evidencia la historia evolutiva entre ellos (Peña, 2011).

Métodos de construcción de árboles filogenéticos

Tres métodos comúnmente son utilizados en estos estudios de construcción de árboles filogenéticos: Máxima Parsimonia (MP); Maximum Likelihood (ML) o Máxima Verosimilitud e Inferencia Bayesiana (IB). De estos tres métodos válidos para inferencia filogenética, MP, ML, e IB, algunos sistemáticos utilizan solo uno de ellos para sus manuscritos, o combinaciones de los tres. (Peña, 2011).

Maximum Likelihood y la Inferencia Bayesiana son métodos estadísticos basados en modelos de evolución molecular, donde se toma en cuenta conocimiento a priori acerca de los caracteres, especialmente cuando son caracteres moleculares. El método de máxima verosimilitud (ML) estima la probabilidad de una matriz de caracteres explicada en los árboles filogenéticos (Felsenstein, 2004), mientras que la inferencia bayesiana (IB) estima la probabilidad de que los árboles filogenéticos sean explicados por los datos (Brooks et al., 2007; Huelsenbeck et al., 2001; Peña, 2011).

Anotación genómica

La anotación del ADN de un genoma consiste en la identificación estructural de genes, regiones codificantes y motivos, así como la identificación funcional de esas regiones; consiste en el proceso de identificación y etiquetado de todas las características relevantes en una secuencia del genoma (Richardson & Watson, 2013).

Una vez que se han identificado las regiones de interés, se les debe asignar una función, para ello se utilizan herramientas de búsqueda de secuencias homólogas. Básicamente estos buscadores (Blast, HHMER, Diamond) comparan las secuencias del genoma contra bases de datos como NCBI (National Center for Biotechnology Information) y COG (Cluster of Orthologous Groups). Una vez que se han alineado e identificado estas secuencias se les asigna una función si es posible. Estas herramientas son útiles para hacer comparaciones metabólicas e identificar genes de interés (Brenes-Guillén, 2019).

Existen programas computacionales como *GENEIOUS*, *GENEMARK* y *GLIMMER* que permiten anotar los genomas de manera fácil y rápida, comparándolos con genomas de referencia (RefSeq) descritos en bases de datos como GenBank (Lukashin & Borodovsky, 1998).

Hipótesis

Se pueden identificar virus fitopatógenos, en muestras de hojas de plantas de estátice *Limonium sinuatum* (L.) Mill, aplicando secuenciación de alto rendimiento (HTS-High Throughput Sequencing) y programas bioinformáticos.

Capítulo 2

Materiales y métodos

Zona de estudio

La zona de estudio hace relación a una finca florícola ubicada en las cercanías de Quito, Parroquia del Quinche, valle de Tumbaco, provincia de Pichincha. Por temas de confidencialidad el nombre de la empresa y localización geográfica no se menciona puesto que esto puede afectar acuerdos comerciales de la empresa.

Los procesos previos a la secuenciación realizada por la empresa ©Macrogen Inc. se realizaron en el laboratorio de diagnóstico molecular IDgen, que se encuentra ubicado en la Avenida de los Granados, cantón Quito 170513, provincia de Pichincha, Quito-Ecuador.

Obtención de muestras vegetales.

Las muestras vegetales provinieron de plantas de corte estática (*Limonium sinuatum* (L.) Mill.) recolectadas en campo de la plantación florícola cercana a Quito, Parroquia El Quinche, las mismas que fueron transportadas en cadena de frío hasta el laboratorio IDgen para su análisis.

Extracción de ARN total.

Para la extracción y obtención del ARN total, se consideró cuatro hojas de la muestra vegetal con sintomatología para virus, malformación de hojas, enanismo y rayado de hojas “streaking”, se tomó aproximadamente 100 mg de cada hoja para obtener la muestra final. Se utilizó el kit comercial de extracción *SV Total RNA Isolation System* de Promega. El ARN total extraído, se precipitó con acetato de sodio y etanol para mejorar las condiciones de conservación y envió a la compañía ©Macrogen Inc. La calidad del ARN obtenido se midió

mediante lectura espectrofotométrica en el nanodrop y una lectura con *Bioanalyzer* que la realizó luego la compañía ©Macrogen Inc.

Secuenciación y preparación de librerías.

La empresa ©Macrogen Inc. realizó la secuenciación en el sistema *NovaSeq 6000* de la plataforma *Illumina*, y se encargó de preparar las librerías a partir del ARN total enviado con el kit *TruSeq Stranded Total RNA with Ribo-Zero Plant*, realizando lecturas paired-end de 100 pb de tamaño, en cada extremo (forward y reverse) con un tamaño de inserto de 350 pb.

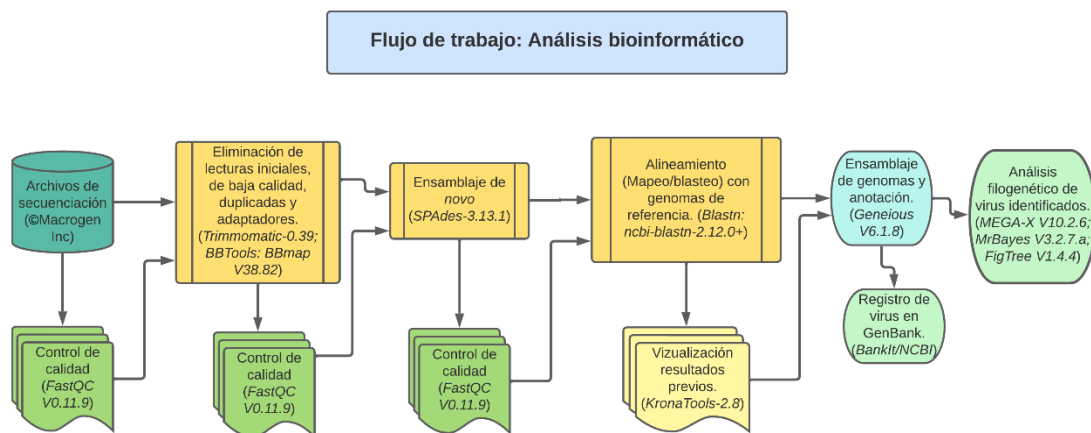
Análisis bioinformático.

Los datos obtenidos de secuenciación, se analizaron mediante software bioinformático de programas de código abierto como Trimomatic-0.39, BBTools:BBMap V38.82, SPAdes-3.13.1, KronaTools-2.7, MEGA-X V10.2.6, MrBayes V3.2.7a, entre otros y software licenciado como Geneious V6.1.8. Se utilizó el cluster de la Corporación Ecuatoriana para el Desarrollo de la Investigación y Academia (CEDIA) así como de un computador de sobremesa.

El flujo de trabajo -pipeline- (Figura 1) reúne los siguientes pasos: 1) Control de calidad de las secuencias (uso de programas que permiten remover los adaptadores, eliminar lecturas muy cortas con índice de calidad bajo). 2) Eliminación de secuencias contaminantes (de origen no viral) y secuencias duplicadas. 3) Ensamblaje *de novo* de las lecturas obtenidas en los pasos previos. 4) Mapeo o búsqueda de homólogos de los *contigs* ensamblados en bases de datos de genomas de referencia. 5) Ensamblaje de genomas completos y 6) Visualización de resultados.

Figura 1

Flujo de trabajo (pipeline) para el análisis bioinformático.



Nota. El flujo de trabajo inicia a partir de los archivos de secuenciados de la muestra foliar de *Limonium sinuatum*, en la figura de izquierda a derecha.

Control de calidad de las muestras secuenciadas.

Los archivos secuenciados (*reads iniciales*) y los archivos finales depurados (*sin lecturas duplicadas y limpios*) se analizaron con el programa FastQC V0.11.9, el software genera resultados en formato *.html.

RUTINA (reads iniciales): fastqc M69_1.fastq.gz M69_2.fastq.gz Archivos de entrada en formato *.fastq.gz: M69_1.fastq.gz y M69_2.fastq.gz Archivos de salida en formato *.fastqc.html: M69_1_fastqc.html y M69_2_fastqc.html

RUTINA (reads finales): fastqc M69_1.Rded.fq M69_2.Rded.fq Archivos de entrada en formato *.fq: M69_1.Rded.fq y M69_2.Rded.fq Archivos de salida en formato *.fastqc.html: M69_1.Rded_fastqc.html y M69_2.Rded_fastqc.html

Remoción de adaptadores y lecturas de baja calidad

Para la remoción de adaptadores y eliminación de lecturas de baja calidad, se utilizó el programa Trimmomatic-0.39. De los archivos secuenciados, se eliminan las secuencias de los adaptadores utilizados en la creación de las librerías (TruSeq3-PE-2.fa:2:30:10), las lecturas de baja calidad (Phred <30) así como las lecturas iniciales de secuenciación (HEADCROP:10).

```
RUTINA: java -Xmx1G -jar /home/marco.taipe/Programas/Trimmomatic-0.39/trimmomatic-0.39.jar PE -phred33 -threads 32 /home/marco.taipe/Datos/M69_1.fastq.gz /home/marco.taipe/Datos/M69_2.fastq.gz /home/marco.taipe/Datos/M69_1leftP.fq /home/marco.taipe/Datos/M69_1leftU.fq /home/marco.taipe/Datos/M69_1rightP.fq /home/marco.taipe/Datos/M69_1rightU.fq
```

```
ILLUMINACLIP:/home/marco.taipe/Programas/Trimmomatic-0.39/adapters/TruSeq3-PE-2.fa:2:30:10 SLIDINGWINDOW:5:20 LEADING:5 TRAILING:5 MINLEN:50 HEADCROP:10
```

* La rutina (script) genera cuatro archivos de salida (leftP.fq; leftU.fq; rightU.fq y rightP.fq) de los cuales solo se utilizan dos (rightP.fq y leftP.fq) para el siguiente paso.

Eliminación de lecturas duplicadas

Para la eliminación de lecturas duplicadas, se utilizó el paquete BBTools:BBMap V38.82, que incluye las herramientas `-dedupe-` y `-reformat-`. Los archivos que ingresaron fueron (rightP.fq y leftP.fq), `-dedupe.sh-` genera dos archivos `M69_Rded.fq` y `M69_Rdup.fq` y `-reformat.sh-` toma solo el archivo `M69_Rded.fq` para generar dos archivos finales (`M69_1.Rded.fq` y `M69_2.Rded.fq`) para el siguiente paso.

```
RUTINA: export PATH=$PATH:/home/marco.taipe/Programs/bbmap dedupe.sh in1=/home/marco.taipe/Datos/M69_leftP.fq in2=/home/marco.taipe/Datos/M69_rightP.fq out=/home/marco.taipe/Datos/M69_Rded.fq outd=/home/marco.taipe/Datos/M69_Rdup.fq ac=f
```



```
reformat.sh in=/home/marco.taipe/Datos/M69_Rded.fq
```

```
out1=/home/marco.taipe/Datos/M69_1.Rded.fq qfake=35
```

```
out2=/home/marco.taipe/Datos/M69_2.Rded.fq qfake=35
```

Ensamblaje de secuencias genómicas de virus fitopatógenos

Para el ensamblaje de *novo*, se utilizó el programa SPAdes-3.13.1, los datos de entrada fueron los archivos depurados *M69_1.Rded.fq* y *M69_2.Rded.fq*, los archivos y datos de salida del ensamblaje se crean en un directorio del archivo raíz, aquí se encuentran varias carpetas y archivos creados como resultado del algoritmo del programa SPAdes, de aquí se seleccionaron los archivos *scaffolds.fasta* y *contigs.fasta* que contienen las probables secuencias ensambladas de virus fitopatógenos.

```
RUTINA: export PATH=$PATH:/home/marco.taipe/Programas/SPAdes-3.13.1-Linux/bin
```

```
metaspades.py --pe1-1 /home/marco.taipe/Datos/M69_1.Rded.fq --pe1-2
```

```
/home/marco.taipe/Datos /M69_2.Rded.fq -o /home/marco.taipe/Datos/metaspadesoutput
```

Alineamiento con bases de datos de secuencias de referencia para virus, (Blasteo o Mapeo)

Para el alineamiento de las secuencias (*contigs.fasta*), se utilizó el programa ncbi-blast-2.12.0+ de (BLAST/BLASTN del NCBI). La base de datos de referencia utilizada (*subject*) corresponde a una base de datos creada con las secuencias de virus para plantas disponibles en la base de datos de referencia de virus del NCBI (*ref_viruses_rep_genomes*). El archivo de consulta (*query*) corresponde a *contigs.fasta* y el resultado de la alineación es un archivo de texto (*VIRUS_M69.txt*) que contiene los nombres de los organismos presentes en la muestra.

```

RUTINA: export PATH=$PATH:/home/marco.taipe/Programas/ncbi-blast-2.12.0+/bin
export BLASTDB=/database/blast_db
blastn -query /home/marco.taipe/Datos/metaspadesoutput /contigs.fasta -db
/home/marco.taipe/database/blast_db/ref_viruses_rep_genomes -out
/home/marco.taipe/database/VIRUS_M69.txt -outfmt 6 -perc_identity 65 -qcov_hsp_perc 30 -
num_threads 8 -max_target_seqs 5
*Archivo query (consulta): contigs.fasta (También se puede usar el archivo scaffolds.fasta).

```

Visualización de resultados en KronaTools

La alineación de las secuencias de virus potenciales, se observó con el programa KronaTools-2.8, este software genera un archivo gráfico en formato *.html (*VIRUS_M69.html*) con la proporción de virus que coincidieron con las secuencias de referencia de los organismos alineados del archivo de formato *.txt (*VIRUS_M69.txt*).

```

RUTINA: export PATH=$PATH:/home/marco.taipe/database
export PATH=$PATH:/home/marco.taipe/Programas/KronaTools-2.8/bin
ktImportBLAST -o /home/marco.taipe/database/VIRUS_M69.html
/home/marco.taipe/database/VIRUS_M69.txt

```

Ensamblaje final del virus encontrado

El organismo identificado con mayor proporción en las muestras de *Limonium*, fue nuevamente ensamblado utilizando el programa Geneious V6.1.8, para esto se obtuvieron las secuencias de referencia RefSeq NC_055298 del segmento L; NC_055297 del segmento M y NC_055299 del segmento S que corresponden a los segmentos del virus de *Alstroemeria necrotic streak virus* (ANSV) presentes en el GenBank y se contrastó con el archivo de datos resultado del ensamblaje (*contigs.fasta*).

Anotación de virus

Las secuencias del genoma viral ensamblado de *Alstroemeria necrotic streak virus* (ANSV), se anotaron con el programa Geneious V6.1.8, que ejecuta un pipeline de anotación propio del software, en este proceso se identificó y etiquetó todas las características relevantes en la secuencia del genoma, utilizando genomas de referencia virales de las bases de datos en línea del GenBank del NCBI.

Análisis filogenético: reconstrucción histórica del virus identificado

Para la reconstrucción histórica de la filogenia del virus identificado de *Alstroemeria necrotic streak virus* (ANSV), se procedió a realizar dos análisis: máxima verosimilitud con el programa MEGA-X V10.2.6 e inferencia bayesiana con el programa MrBayes V3.2.7a.

En MEGAX V10.2.6, se determinó el mejor modelo evolutivo, para lo que se creó una base de datos con la secuencia de aminoácidos del *gen N de la nucleocapside* de 29 accesiones de secuencias disponibles para *Orthospovirus* del GenBank y la secuencia reconstruida del virus ANSV (Tabla 1). Estos datos se alinearon con la herramienta *ClustalW* del mismo programa.

En el análisis de máxima verosimilitud, se utilizó el modelo General Time Reversible con distribución gamma (GTR+G) para el cálculo de su algoritmo, con un bootstrap de 1000 repeticiones; para el análisis bayesiano se utilizó los parámetros del mejor modelo evolutivo encontrado GTR+G y el algoritmo Markov Chain Monte Carlo (MCMC) en el cálculo de la probabilidad posterior, el gráfico resultante de este cálculo se observó con el software FigTree V1.4.4.

```
RUTINA: Mrbayes > execute C:\Marco_Trabajo\Tesis_maestria\arbol\archivo.nex lset nst=6
rates=gamma mcmc ngen=20000 samplefreq=100 sump burnin=2000 sumt burnin=2000
```

Tabla 1

Accesiones de virus del genero *Orthotospovirus* seleccionados del GenBank para el análisis filogenético

Orthotopovirus	Codigo virus	No. Accesoión en GenBank	Ubicación	Hospedero
<i>Capsicum chlorosis virus</i>	CaCV	AY036058	Queensland, Australia	<i>Capsicum, chilli and tomato</i>
<i>Calla lily chlorotic spot virus</i>	CCSV	AY867502	China: Zhejiang, Quzhou, Changshu	<i>Lactuca sativa</i>
<i>Chrysanthemum stem necrosis virus</i>	CSNV	AF067068	Brasil	<i>Chrysanthemum x morifolium</i>
<i>Groundnut bud necrosis virus</i>	GBNV	U27809	India	<i>Capsicum annum</i>
<i>Groundnut ringspot virus</i>	GRSV	S54327	South Africa	<i>groundnut</i>
<i>Impatiens necrotic spot virus</i>	INSV	X66972	Netherlands	<i>Impatiens</i>
<i>Iris yellow spot virus</i>	IYSV	AF001387	Netherlands	<i>Allium sp. /onion</i>
<i>Melon severe mosaic virus</i>	MeSMV	EU275149	Mexico	<i>Nicotiana benthamiana</i>
<i>Melon yellow spot virus</i>	MYSV	AF067151	Netherlands	<i>Melon</i>
<i>Peanut chlorotic fan-spot virus</i>	PCFV	AF080526	Taiwan	<i>Peanut</i>
<i>Peanut yellow spot virus</i>	PYSV	AF013994	USA: Florida	<i>Peanut</i>
<i>Polygonum ringspot virus</i>	PoIRSV	KX468765	Italy	<i>Nicotiana benthamiana</i>
<i>Tomato chlorotic spot virus</i>	TCSV	S54325	Brasil	<i>Tomato</i>
<i>Tomato spotted wilt virus</i>	TSWV	D00645	Netherlands	<i>Tomato</i>
<i>Tomato yellow ring virus</i>	TYRV	AY686718	Netherlands	<i>Tomato</i>
<i>Tomato zonate spot virus</i>	TZSV	EF552433	China: Yunan	<i>Lycopersicon esculentum</i>
<i>Watermelon bud necrosis virus</i>	WBNV	EU373749	India	<i>watermelon</i>
<i>Watermelon silver mottle virus</i>	WSMoV	Z46419	Germany	<i>Tomato</i>
<i>Zucchini lethal chlorosis virus</i>	ZLCV	AF067069	Brasil	<i>Zucchini</i>
<i>Alstroemeria necrotic streak virus</i>	ANSV-El Quinche	ON149271	Ecuador	<i>Limonium sinuatum</i>
<i>Alstroemeria necrotic streak virus</i>	ANSV-Hoya SD	MK801881	USA: San Diego Country, CA	<i>Hoya carnosa</i>
<i>Alstroemeria necrotic streak virus</i>	ANSV-San Vicente	MK275263	Colombia	<i>Solanum quitoense</i>
<i>Alstroemeria necrotic streak virus</i>	ANSV-Penol	MK275262	Colombia	<i>Capsicum annum</i>
<i>Alstroemeria necrotic streak virus</i>	ANSV-Marinilla	MK275264	Colombia	<i>Capsicum annum</i>
<i>Bean necrotic mosaic virus</i>	BNMV	JN587268	Brasil	<i>Phaseolus vulgaris</i>
<i>Hippeastrum chlorotic spot virus</i>	HCSV	JX833564	China	<i>Hymenocallis littoralis</i>
<i>Groundnut yellow spot virus</i>	GYSV	NC_043210	China	<i>Chili</i>
<i>Soybean vein necrosis virus</i>	SVNV	NC_055178	USA	<i>Soybean</i>
<i>Alstroemeria yellow spot virus</i>	AYSV	NC_040741	Netherlands	<i>Alstroemeria</i>
<i>Pepper chlorotic spot virus</i>	PCSV	NC_033772	China	<i>Pepper</i>

Capítulo 3

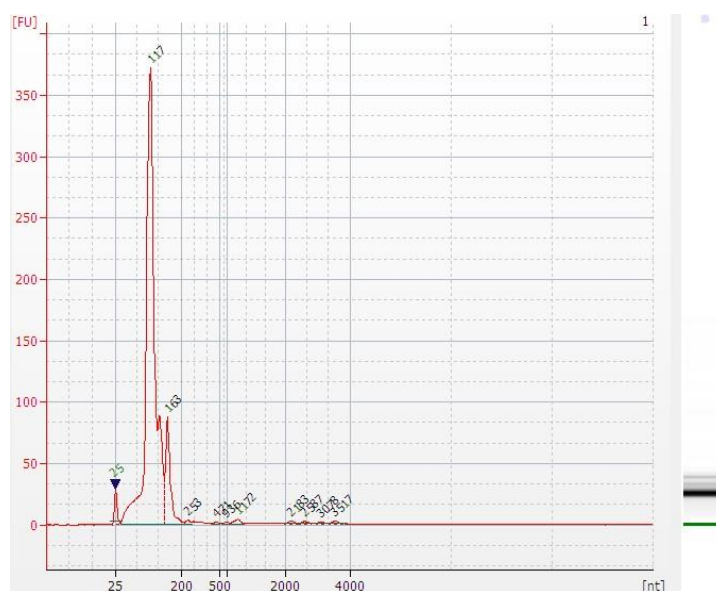
Resultados

Características de la muestra secuenciada.

El ARN total obtenido de la muestra foliar de estáctice *Limonium sinuatum* (codificado en IDgen como M69) tuvo una lectura espectrofotométrica de concentración 36.7 ng/uL en el nanodrop y una lectura de número de integridad de RNA de 2.5 con Bioanalyser (Figura 2).

Figura 2

Lectura con Bioanalyzer (Macrogen Inc.).



Nota. El pico a 25 nt corresponde al marcador, los picos más grandes corresponden al ARN total.

La cantidad de lecturas de bases en la muestra de hojas de estáctice *Limonium sinuatum* (M69) son de 5.3 G, con un contenido GC de 41.45%, contenido AT de 58.55% y con índices de calidad Q20 98.14% y Q30 94.91%, Tabla 2.

Tabla 2

*Características de la muestra secuenciada (M69) de hojas de estátice *Limonium sinuatum*, enviado por la Compañía ©Macrogen Inc. Corea del Sur.*

Sample ID	Total read bases (bp)	Total reads	GC(%)	AT(%)	Q20(%)	Q30(%)
M69-EI Quinche	5.278.306.258	52.260.458	41.45	58.55	98.14	94.91

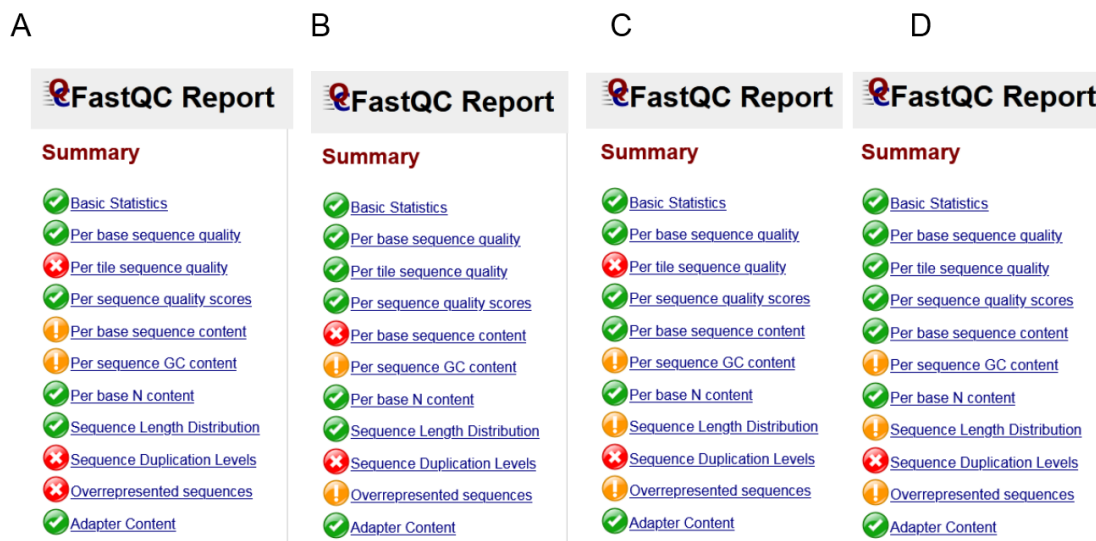
Nota. **Sample ID:** Nombre de la muestra; **Total read bases (bp):** Número total de bases secuenciadas; **Total reads:** Número total de lecturas; **GC(%):** Contenido GC; **AT(%):** Contenido AT; **Q20(%):** rango de bases con score de calidad superior a 20; **Q30(%):** rango de bases con score de calidad superior a 30.

Control de la calidad inicial de la muestra secuenciada.

La calidad inicial de los archivos de lecturas secuenciadas (*reads*) muestran un resumen “*FastQC report Summary*” con once parámetros que la herramienta considera para indicar la calidad de la información de cada archivo. Los archivos iniciales *M69_1.fastq.gz* y *M69_2.fastq.gz* presentan juntos, una gran cantidad de parámetros adecuados en color verde, cuatro parámetros aceptables y cinco parámetros no tan aceptables (Figura 3, A y B).

Figura 3

Control de la calidad inicial y final de archivos secuenciados



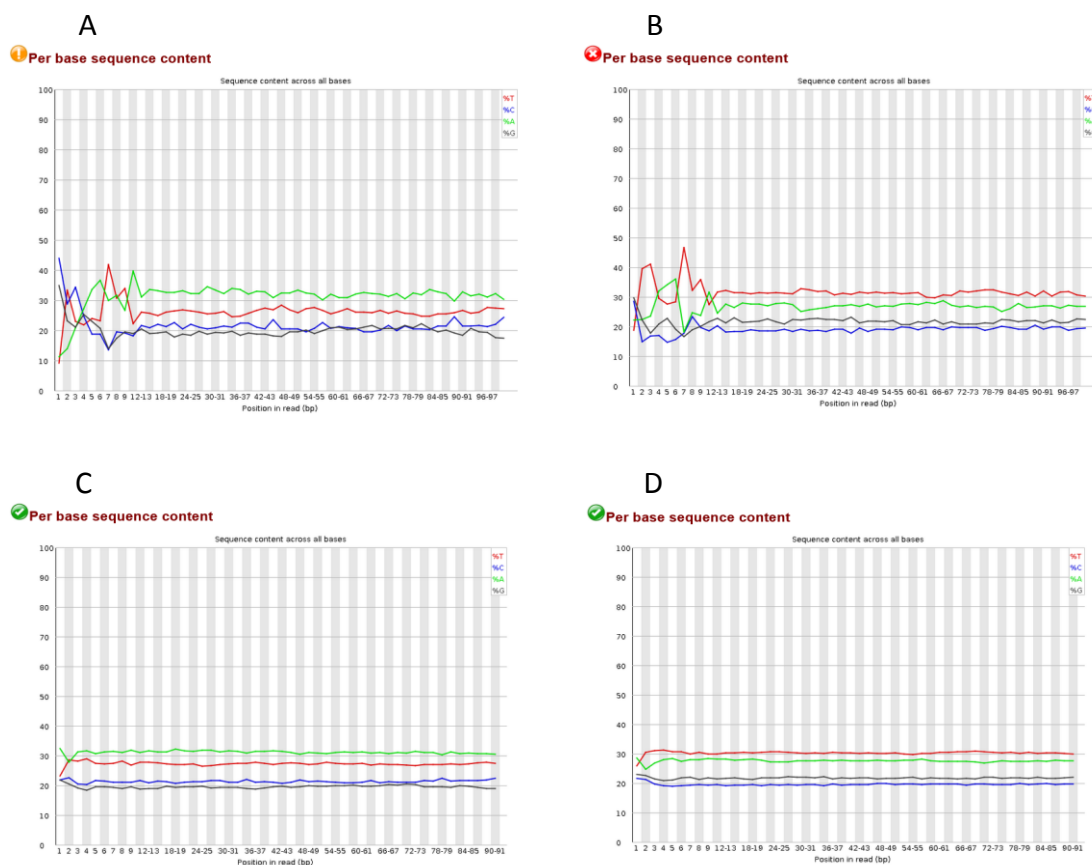
Nota. A) Reporte inicial del archivo M69_1.fastq.gz B) reporte inicial del archivo M69_2.fastq.gz C) reporte eliminado las lecturas iniciales (*head*) del archivo M69_1.fastq.gz B) reporte eliminado las lecturas iniciales (*head*) del archivo M69_2.fastq.gz; *El color verde corresponde a un parámetro adecuado, el naranja aceptable y el color rojo no tan aceptable.

La calidad de las lecturas secuenciadas *M69_1.fastq.gz* y *M69_2.fastq.gz*, luego de eliminar las lecturas iniciales (*head*), se puede observar en el resumen “*FastQC report*” (Figura 3, C y D), las lecturas presentan una mayor cantidad de parámetros adecuados en color verde, siete parámetros aceptables y tres parámetros no tan aceptables.

Para el parámetro contenido de secuencias por base, “*per base sequence content*”, se observa las 10 secuencias (*head*) iniciales a eliminar (Figura 4, A y B) y en la misma figura X, C y D las secuencias iniciales (*head*) ya eliminadas.

Figura 4

Calidad del parámetro contenido de secuencias por base

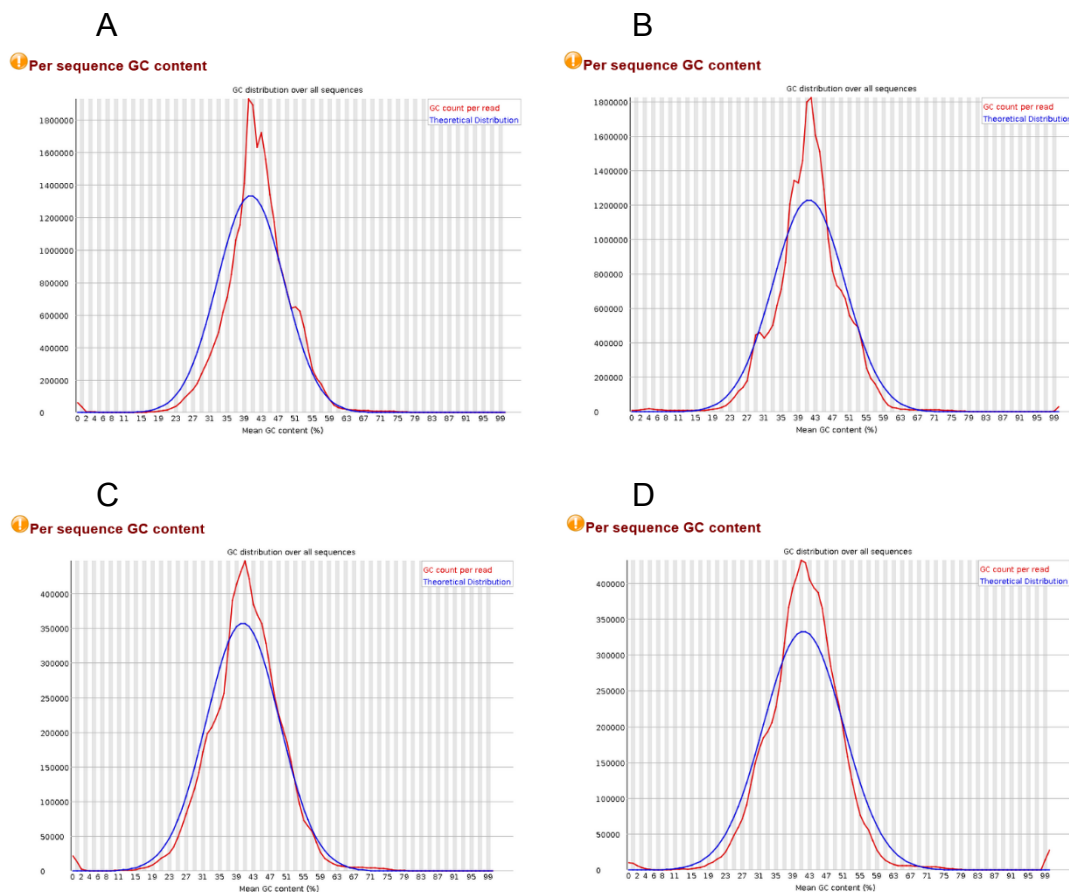


Nota. A) Secuencia de head a eliminar de M69_1.fastq.gz B) Secuencia de head a eliminar de M69_2.fastq.gz C) Secuencias de head eliminadas de M69_1.fastq.gz D) Secuencias de head eliminadas de M69_2.fastq.gz

Para el parámetro contenido GC por secuencia, “*per sequence GC content*”, (Figura 5, A y B) se observa la distribución inicial del conteo de los *reads* GC (en rojo) frente a la distribución teórica (en azul) y en la misma figura 5 en C y D, la distribución del conteo de *reads* más ajustada a la distribución teórica, luego de la limpieza de los datos.

Figura 5

Calidad del parámetro contenido GC por secuencia



Nota. A) Contenido GC del archivo M69_1.fastq.gz B) Contenido GC del archivo M69_2.fastq.gz
 C) Contenido GC del archivo M69_1.fastq.gz sin lecturas iniciales “head” D) Contenido GC del
 archivo M69_2.fastq.gz sin lecturas iniciales “head”.

Reporte final de los datos depurados y limpios.

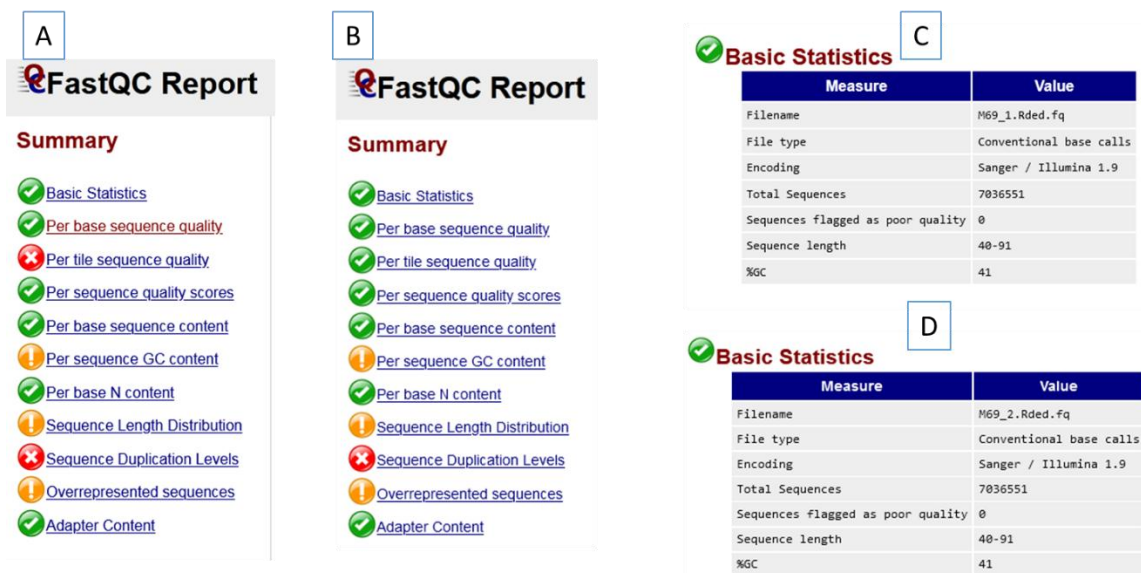
El reporte final de los archivos depurados *M69_1.Rded.fq* y *M69_2.Rded.fq*, sin lecturas duplicadas y remoción de adaptadores, se puede observar en la figura 6.




Estos archivos corresponden a los utilizados para el ensamblaje y procesamiento de la información generada de todos los resultados que se presentan. El resumen de calidad “*FastQC report*”, muestra una gran cantidad de parámetros adecuados en color verde, seis parámetros aceptables y tres no tan aceptables (Figura 6: A y B). El parámetro estadísticas

básicas, muestra un total de secuencias de 7036551 bp con un largo de secuencias entre 40-91 bp y contenido GC de 41% (Figura 6, C y D).

Figura 6

Reporte final de la calidad de los archivos depurados



Nota. Reporte final para la calidad de datos de los archivos *M69_1.Rded.fq* y *M69_2.Rded.fq*: A y B) Resumen de calidad “FastQC report”, C y D Estadísticas básicas; *El color verde  corresponde a un parámetro adecuado, el naranja  aceptable y el color rojo  no tan aceptable.

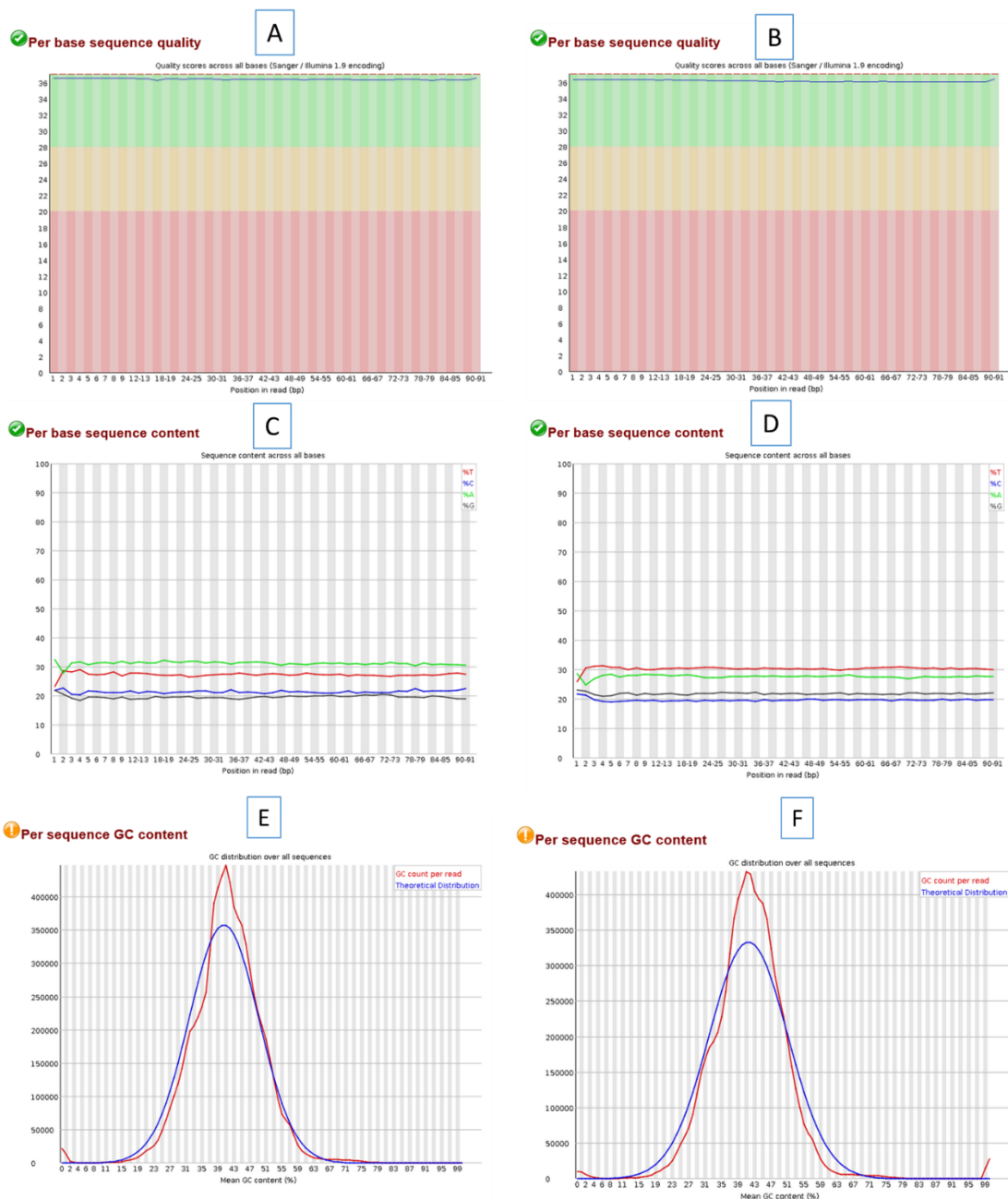
Para el parámetro “Per base sequence quality” todas las lecturas (*reads*) se encuentran dentro del área de color verde (28-36 del eje de ordenadas) del gráfico resultado, indicando una adecuada calidad de los datos (Figura 7, A y B).

Para el parámetro contenido de secuencias por base, “Per base sequence content”, se observa los porcentajes para Timina (%T: en rojo), porcentaje para Citocinina (%C: en azul), porcentaje para Adenina (%A: en verde) y porcentaje para Guanina (%G: en negro) configurando todos los porcentajes líneas horizontales en el grafico (Figura 7, C y D).

Para el parámetro contenido GC por secuencia, “*Per sequence GC content*”, se observa la distribución del conteo de los *reads* GC (en rojo) frente a la distribución teórica (en azul) bastante ajustada (Figura 7, E y F).

Figura 7

Reporte final para la calidad de datos, archivos *M69_1.Rded.fq* y *M69_2.Rded.fq*



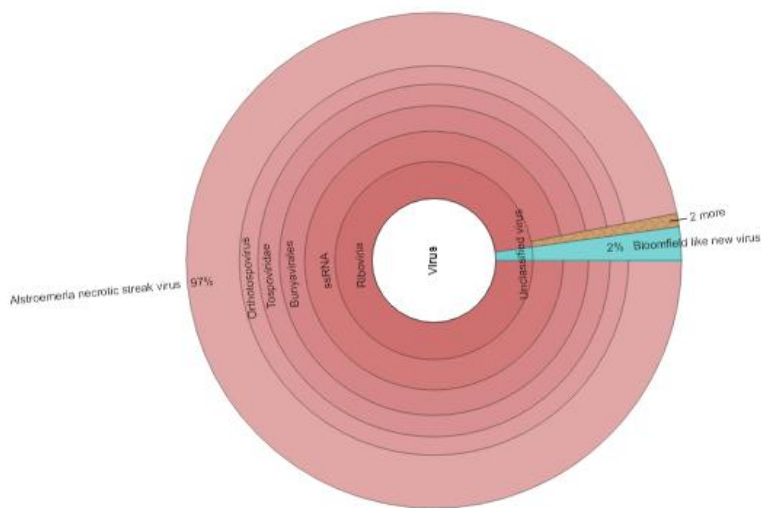
Nota. A y B) Calidad de las secuencias por base, C y D) contenido de secuencias por base, E y F) Contenido GC por secuencia.

Visualización de secuencias virales en KronaTools

En la identificación de secuencias genómicas de virus fitopatógenos, se puede observar que 97% de secuencias coinciden con el virus de *Alstroemeria necrotic streak virus*, 2% de secuencias coinciden con *Bloomfield like new virus*, 0.8% *New chrysovirus* y 0.2% para *Satellite tobacco necrosis virus 2*, (Figura 8 y Tabla 3).

Figura 8

Diagrama con la distribución en porcentajes de los virus encontrados en la muestra de *Limonium sinuatum* (M69) con el programa KronaTools.



Nota. Resultado del análisis realizado con el software *KronaTools*

Tabla 3

Resultados de la identificación de organismos presentes en muestras de Limonium sinuatum (datos de secuenciación y de referencia para virus del GenBank)

Muestra	Total de lecturas (bp)	Organismos identificados	% de lecturas virales	% identidad con secuencias reportadas
Hojas de Limonium sinuatum	5.3G	<i>Alstroemeria necrotic streak virus</i>	97	96
		<i>Satellite tobacco necrosis virus 2</i>	0.2	95
		<i>Bloomfield like new virus</i>	2	40
		<i>New chrysovirus</i>	0.8	50

Características de los segmentos virales de *Alstroemeria necrotic streak virus* ANSV ensamblado.

El organismo identificado con mayor proporción en las muestras de *Limonium*, *Alstroemeria necrotic streak virus* (ANSV) (Figura 8 y Tabla 3), fue nuevamente reensamblado en Geneious V6.1.8, el procedimiento generó cuatro contigs que coincidieron con las secuencias de referencia de los segmentos de ANSV del *GenBank*, dos de los contigs ensamblaron el segmento S, mientras que los otros dos *contigs* representan los segmentos L y M completos.

Las secuencias ensambladas para los segmentos del virus de *Alstroemeria necrotic streak virus* (ANSV), tuvieron las siguientes características: Segmento L, una longitud de 8755 pb, 98.9% de identidad con el segmento L de referencia, 34.2% de contenido GC y una profundidad de cobertura de 890X; Segmento M, una longitud de 4869 pb, 98.7% de identidad con el segmento M de referencia, 35.6% de contenido GC y profundidad de cobertura de 2870X; Segmento S, una longitud de 3135 pb, 99.3% de identidad con el segmento S de referencia, 35.1% contenido GC y una profundidad de cobertura de 3500X (Tabla 4).

Tabla 4

Características de los segmentos del virus de *Alstroemeria necrotic streak virus* ANSV ensamblado y contrastado con las secuencias de los segmentos de accesiones de referencia del virus de ANSV del GenBank

Segmento ensamblado del virus ANSV	# de Contigs	Refseq (Accesión de referencia)	Largo (pb)	% GC	% Identidad	Profundidad de cobertura
L	1	NC_055298	8755	34.2	98.9	890 X
M	1	NC_055297	4869	35.6	98.7	2870 X
S	2	NC_055299	3135	35.1	99.3	3500 X

Anotación del genoma viral identificado

En la predicción de marcos de lectura abiertos para los segmentos del virus de *Alstroemeria necrotic streak virus* (ANSV) se puede observar: Al segmento L con un ORF que codifica la *proteína L* con 8625 pb; al segmento M que tiene dos ORFs, el ORF1 que codifica la *proteína no estructural* con 912 pb y un ORF2 que codifica la *glicoproteína precursora* con 3381 pb; y finalmente al segmento S que tiene también dos ORFs, ORF1 codifica una *proteína no estructural* de 1404 pb y ORF2 que codifica el *gen N de la nucleocápside* con 777 pb. (Tabla 5)

Tabla 5

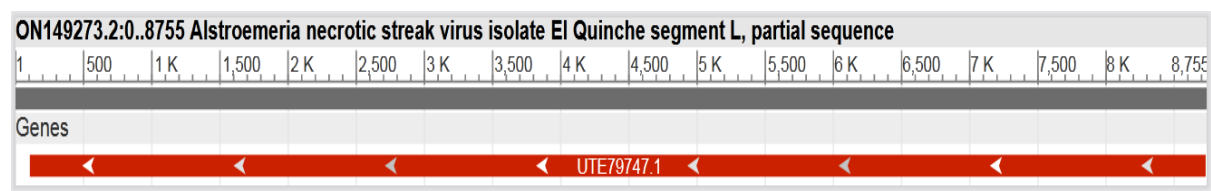
Anotación para los segmentos del virus de Alstroemeria necrotic streak virus (ANSV) identificado en la muestra foliar de Limonium sinuatum

Segmento del virus de ANSV	# de Accesoión	Largo de la secuencia (pb)	ORFs	Largo secuencia de ORFs (pb)	Proteína	ID_proteína	% Similaridad en la anotación
L	ON149273	8755	ORF	8625	Proteína L	AYJ76753.1	98.92
M	ON149272	4869	ORF1	912	Non-estructural	AYJ76755.1	99.12
			ORF2	3381	Glicoproteína precursora	AYJ76754.1	99.05
S	ON149271	3135	ORF1	1404	Non-estructural	YP_010086065.1	99.55
			ORF2	777	Proteína de la nucleocapside	YP_010086066.1	99.87

Las secuencias anotadas de los tres segmentos que conforman el virus de *Alstroemeria necrotic streak virus (ANSV)*, fueron registradas en el GenBank con la herramienta BankIt (<https://submit.ncbi.nlm.nih.gov/about/bankit/>) con los números de accesoión designados por GenBank, ON149273 para el *segmento L* (Figura 11), accesoión ON149272 para el *segmento M* (Figura 10) y accesoión ON149271 para el *segmento S* (Figura 9).

Figura 9

Gráfico para el segmento L del virus de Alstroemeria necrotic streak virus ANSV-L



Nota. Reporte gráfico, registro GenBank (ON149273) obtenido de:

<https://www.ncbi.nlm.nih.gov/nuccore/ON149273.2?report=graph>

Figura 10

Gráfico para el segmento M del virus de *Alstroemeria necrotic streak virus* ANSV-M.

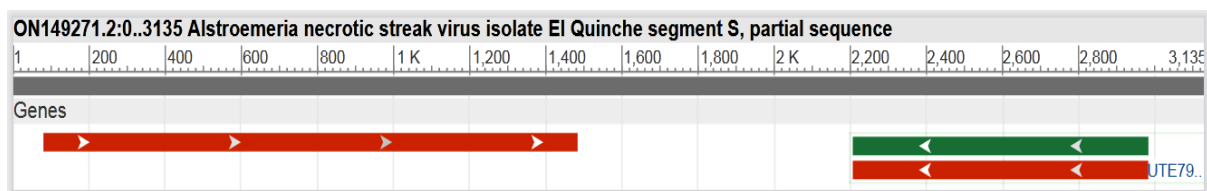


Nota. Reporte gráfico, registro GenBank (ON149272) obtenido de:

<https://www.ncbi.nlm.nih.gov/nuccore/ON149272.2?report=graph>

Figura 11

Gráfico para el segmento S del virus de *Alstroemeria necrotic streak virus* ANSV-S.



Nota. Reporte gráfico, registro GenBank (ON149271) obtenido de:

<https://www.ncbi.nlm.nih.gov/nuccore/ON149271.2?report=graph>

Reconstrucción histórica evolutiva de *Alstroemeria necrotic streak virus* ANSV

La reconstrucción histórica evolutiva, se realizó para el virus de *Alstroemeria necrotic streak virus* (ANSV) cuya proporción de *contigs* fue del 97% frente a la poca proporción de *contigs* de los otros virus identificados, *Satellite tobacco necrosis virus 2* (0.2%), *Bloomfield like new virus* (2%), *New chrysovirus* (0.8%) (Tabla 3), por lo que para estos virus no se realiza el análisis filogenético.

Con las secuencias de los aminoácidos del gen *N* de la *nucleocapside* para 30 secuencias genómicas de *Orthotospovirus* disponibles en GenBank (Tabla 1), se determinó el mejor

modelo evolutivo, que corresponde al de General Time Reversible con distribución gamma (GTR+G), cuyo puntaje de Criterio de Información Bayesiano (BIC), es más bajo que el resto de los otros modelos (Tabla 6).

Tabla 6

Resultado del mejor modelo evolutivo, obtenido en MEGA-X V10.2.6 para las 30 secuencias genómicas del gen N de la nucleocapside

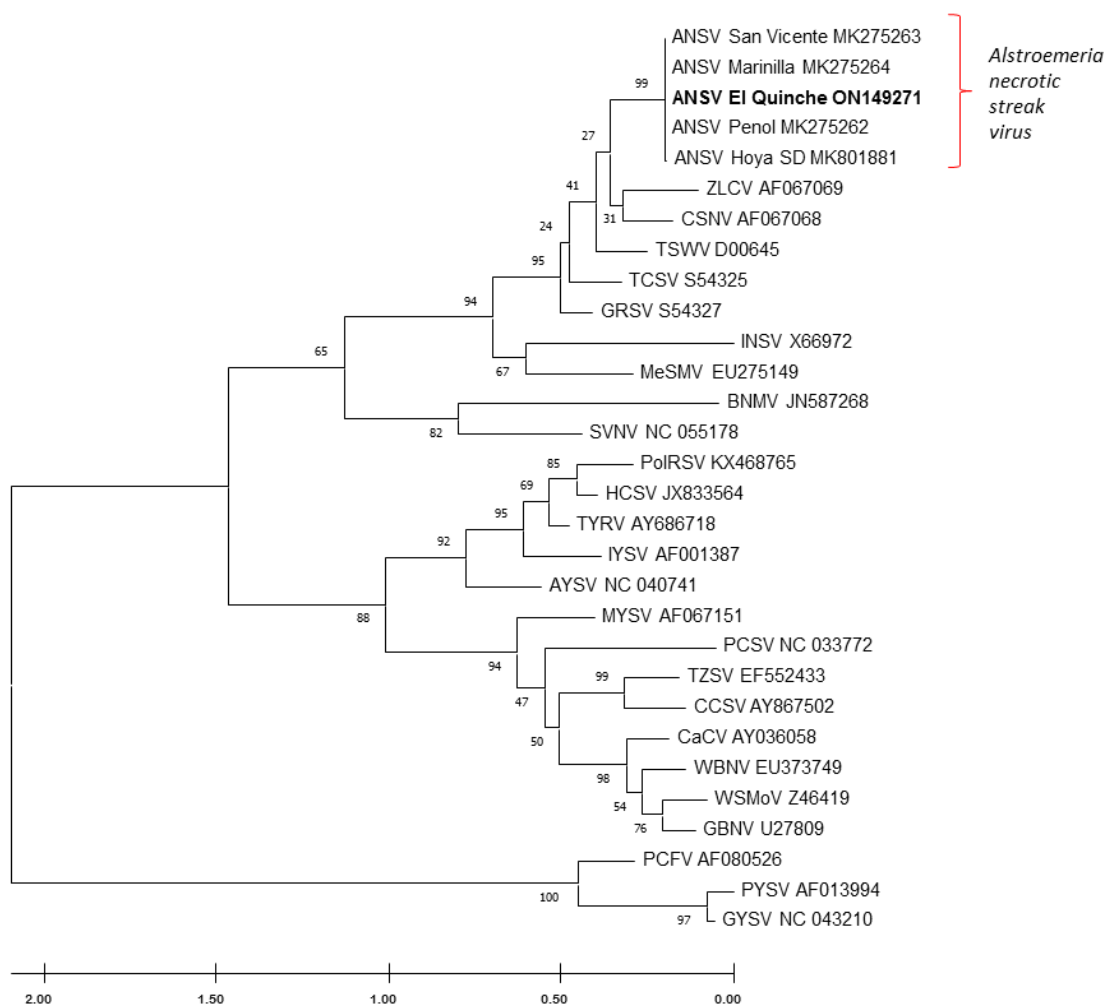
Model	#Param	BIC	AICc	lnL	Invariant	Gamma
GTR+G	66	16049,86726	15565,94179	-7716,581737	n/a	0,934852441
GTR+G+I	67	16058,56467	15567,31886	-7716,258445	0,016825113	1,004492644
T92+G	60	16069,03404	15629,03809	-7754,197118	n/a	0,966120468
T92+G+I	61	16076,71034	15629,39192	-7753,36327	0,03047218	1,111122075
TN93+G	63	16095,50387	15633,54156	-7753,416039	n/a	0,956804921
HKY+G	62	16099,4815	15644,84096	-7760,076851	n/a	0,962115559
TN93+G+I	64	16102,10429	15632,82057	-7752,044251	0,037856583	1,141357585
HKY+G+I	63	16106,82767	15644,86536	-7759,077937	0,032894549	1,119313468
K2+G	59	16155,03243	15722,35931	-7801,868309	n/a	0,986822086
K2+G+I	60	16161,55184	15721,55589	-7800,456018	0,038305238	1,177932239
T92+I	60	16495,72113	16055,72518	-7967,540663	0,08531854	n/a
GTR+I	66	16498,04311	16014,11764	-7940,669661	0,085298736	n/a
TN93+I	63	16509,87935	16047,91704	-7960,603775	0,085483245	n/a
JC+G	58	16523,06559	16097,71566	-7990,556888	n/a	1,072247998
HKY+I	62	16526,79956	16072,15901	-7973,735879	0,085373425	n/a
JC+G+I	59	16528,95517	16096,28205	-7988,82968	0,040721611	1,305130829
K2+I	59	16564,5555	16131,88238	-8006,629846	0,085831448	n/a
GTR	65	16686,72254	16210,11776	-8039,681374	n/a	n/a
T92	59	16693,72038	16261,04726	-8071,212285	n/a	n/a
TN93	62	16720,5001	16265,85955	-8070,586151	n/a	n/a
HKY	61	16728,05777	16280,73935	-8079,036985	n/a	n/a
K2	58	16786,82963	16361,4797	-8122,438911	n/a	n/a
JC+I	58	16892,72053	16467,3706	-8175,384361	0,085717773	n/a
JC	57	17105,33737	16687,31098	-8286,364778	n/a	n/a

Nota. Abreviaciones TR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor. Discrete Gamma distribution (+G), evolutionarily invariable (+I), BIC (Criterio de Información Bayesiano), AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (lnL).

El árbol filogenético para el virus de ANSV, con el valor más alto de probabilidad logarítmica (-7697.19), aplicando el modelo General Time Reversible de distribución gamma (GTR+G) y bootstrap de 1000, se observa en la figura 12. La muestra de ANSV El Quinche ON149271 (en negrillas) se encuentra en el mismo clado (llave en rojo) que las accesiones de los virus de *Alstroemeria necrotic streak virus* (ANSV) disponibles en el GenBank (Figura 12).

Figura 12

Árbol filogenético con el valor más alto de probabilidad logarítmica para el virus ANSV El Quinche ON149271

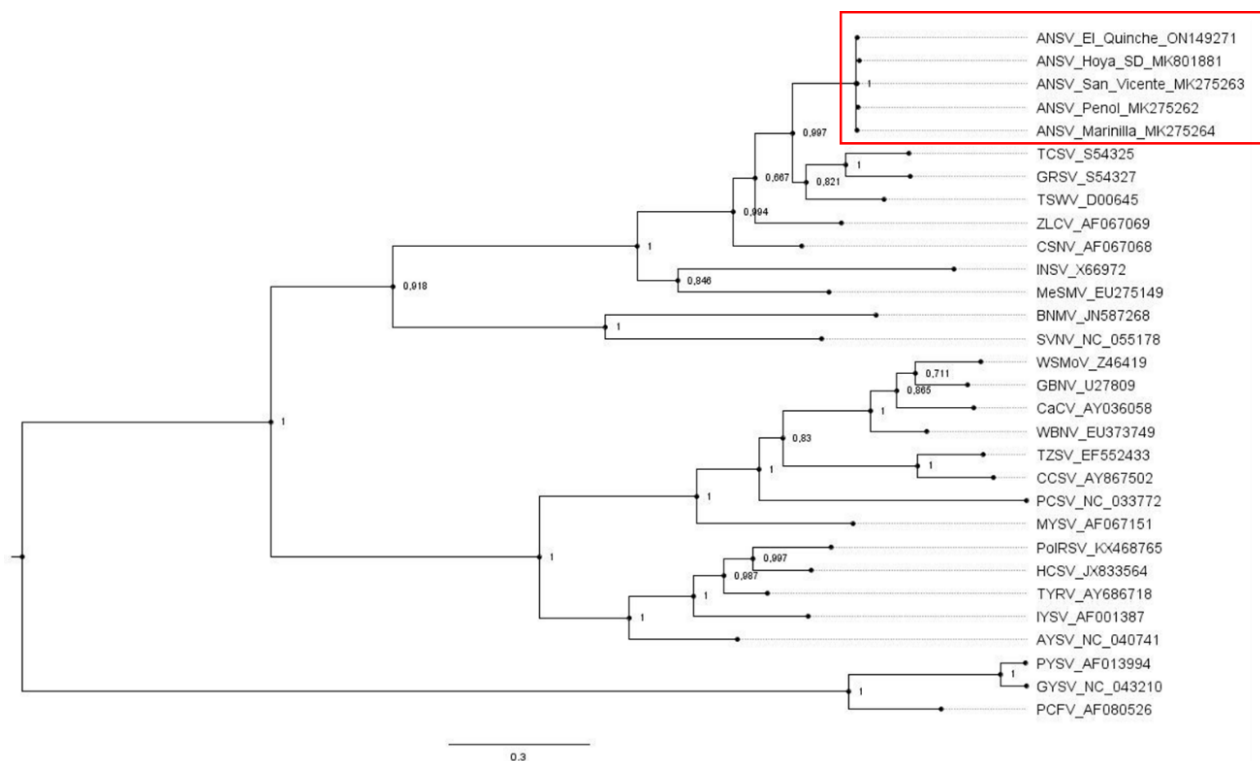


Nota. Análisis filogenético: Máxima Verosimilitud con MEGA-X V10.2.6

En el análisis de inferencia bayesiana y el cálculo de las probabilidades posteriores, aplicando el modelo evolutivo General Time Reversible con distribución gamma GTR+G y el algoritmo MCMC, se puede observar en árbol filogenético resultante (Figura 13) una topología semejante con el árbol filogenético de máxima verosimilitud (Figura 12), la probabilidad posterior para el clado de los virus de ANSV es de uno (rectángulo rojo), y donde está ubicada la muestra de ANSV El Quinche ON149271, que contiene el *gen N* de la nucleocapside.

Figura 13

Árbol filogenético de la inferencia bayesiana para la probabilidad posterior de la muestra de ANSV El Quinche ON149271



Nota. Análisis filogenético: Inferencia Bayesiana con MrBayes V3.2.7a, observación con FigTree V1.4.4.

Capítulo 4

Discusión

En el presente trabajo, la extracción de ARN-total de la muestra foliar a partir de la planta de corte estática *Limonium sinuatum* con sintomatología para virus, fue realizada exitosamente, el valor de concentración de 36.7 ng/uL en el nanodrop resulta ser relativamente bajo, aunque fue suficiente material para permitir realizar la secuenciación de la muestra por la empresa ©Macrogen Inc. Corea del Sur, quien confirmó una lectura de número de integridad de RNA de 2.5 con Bioanalyser. Las características del cDNA de la muestra secuenciada presentaron contenidos GC de 41.45% y AT de 58.55%, así como índices de calidad Q20 98.14% y Q30 94.91% es decir proporciones adecuadas para las lecturas de secuenciación.

Los procesos bioinformáticos que se llevan a cabo en flujos de trabajo “pipeline”, pueden ser desarrollados bajo diferentes criterios que dependen de cada investigador, pero siempre con la finalidad de garantizar y asegurar que los datos secuenciados metagenómicos sean de calidad. Previo al procesamiento bioinformático es de gran importancia, la limpieza de lecturas, remoción de adaptadores, eliminación de lecturas duplicadas, observación de los parámetros de calidad inicial de las lecturas, lo que permite incrementar la confiabilidad y calidad final de los resultados metagenómicos. En el presente estudio, la calidad de las lecturas secuenciadas (*reads*) fue mejorada, al final del pre-procesamiento, los parámetros fueron más aceptables al eliminar lecturas iniciales, adaptadores y duplicados, lo que permitió obtener una distribución del conteo de *reads* más ajustado a la distribución teórica inicial del “*FastQC report*”. El total de secuencias finales fue de 7036551 bp con un largo de las secuencias entre 40-91 bp y contenido GC de 41%.

En el análisis bioinformático se debe considerar la capacidad de memoria computacional requerida para varios procesos informáticos, en muchas ocasiones se requiere de gran

equipamiento, en nuestro trabajo se requirió del uso del cluster de la Corporación Ecuatoriana para el Desarrollo de la Investigación y Academia (CEDIA) para cumplir los objetivos planteados. También es necesario considerar el uso de programas bajo licencia y de uso libre pues esto determina la facilidad de elegir un mejor flujo de trabajo (Van der Walt et al., 2017).

El ensamblaje de *novο* de las secuencias generadas con metaSPAdes, y su comparación con genomas de referencia para virus de plantas del GenBank, identificó secuencias genómicas de virus fitopatógenos en un 97% coincidentes con el virus de *Alstroemeria necrotic streak virus*, 2% de secuencias con *Bloomfield like new virus*, 0.8% con *New chrysovirus* y 0.2% para *Satellite tobacco necrosis virus 2*, valores visualizados con ayuda del programa KronaTools.

En las muestras foliares de las plantas de corte de *Limonium sinuatum*, el organismo identificado con mayor proporción fue el *Alstroemeria necrotic streak virus* (ANSV), un virus fitopatógeno del género *Orthospovirus* cuya característica principal es la de poseer 3 segmentos virales como lo mencionan (Gallo et al., 2019; Hassani-Mehraban et al., 2010), estos tres segmentos fueron ensamblados a partir de cuatro *contigs* que coincidieron con las secuencias de referencia de los segmentos de ANSV del GenBank, dos de los *contigs* ensamblaron el segmento S, mientras que los otros dos *contigs* representan los segmentos L y M completos, permitiendo de esta manera realizar un informe indexado en la revista Microbiology Resource Announcements donde se expone la presencia del virus ANSV en el Ecuador.

En 2008, un estudio en plantas ornamentales de *Alstroemeria sp.* de Colombia, reveló la existencia de un nuevo *orthospovirus* que causa síntomas de estrías necróticas en hojas y tallos, por lo que se propuso el nombre de virus de estrías necróticas de alstroemeria (ANSV) (Hassani-Mehraban et al., 2010). Para el 2017, (Olaya et al., 2017) reportan la presencia de ANSV en plantas de tomate (*Solanum lycopersicum* L.) y pimiento (*Capsicum annuum*).

Nuevos estudios para el 2018 se centran en el cultivo de lulo “naranjilla” (*Solanum quitoense*) cultivado en áreas de departamentos colombianos donde también se reporta la presencia de ANSV (Gallo et al., 2018). Un estudio realizado en San Diego, Estados Unidos, en plántulas de flor de nacar o flor de cera bajo invernadero, indicó la presencia de ANSV, este es el primer reporte de este virus fuera de Colombia (Tian et al., 2020). El presente estudio realizado en muestras foliares de las plantas ornamentales de corte de *Limonium sinuatum*, identificó la presencia de *Alstroemeria necrotic streak virus* (ANSV), lo que convierte a este trabajo en una evidencia de la presencia del virus fitopatógeno del género *Orthotospovirus* en Ecuador.

Las secuencias ensambladas para los segmentos del virus de *Alstroemeria necrotic streak virus* (ANSV), tuvieron una longitud de 8755 pb para el *segmento L*, 4869 pb para el *segmento M* y 3135 pb para el *segmento S*. En comparación con las longitudes encontradas por Gallo Yuliana 2019, de 8967 pb para el *segmento L*, 4859 pb para el *segmento M* y 3132 pb para el *segmento S* y también en 2018 por la misma autora, las longitudes de 3113 pb para el *segmento S*, 4839 pb para el *segmento M* y 8756 pb para el *segmento L* (Gallo et al., 2019; Gallo et al., 2018).

Para el virus ensamblado de *Alstroemeria necrotic streak virus* (ANSV), se determinó en el *segmento L* un ORF (marco abierto de lectura) que codifica la *proteína L* con 8625 pb; en el *segmento M* se determinó dos ORFs, el ORF1 que codifica la *proteína no estructural* con 912 pb y un ORF2 que codifica la *glicoproteína precursora* con 3381 pb; finalmente el *segmento S* también presentó dos ORFs, el ORF1 que codifica una *proteína no estructural* de 1404 pb y el ORF2 que codifica el *gen N de la nucleocápside* con 777 pb; estos marcos de lectura son similares a los determinados por Gallo Y. en su estudio de 2018 quien para el *segmento L* presenta un ORF, para el *segmento M* dos ORFs ambisentidos y dos ORFs para el *segmento S* uno de los cuales codifica la *proteína N de la nucleocápside* en el sentido complementarios al

virion (vc); así como también en su siguiente estudio en pimiento (*Capsicum annum*) 2019 donde vuelve a ensamblar el virus con las mismas características de los marcos de lecturas.

En la reconstrucción histórica evolutiva del virus de *Alstroemeria necrotic streak virus* (ANSV), se utilizó las secuencias de los aminoácidos del *gen N de la nucleocapside* para 30 secuencias genómicas de *Orthospovirus* disponibles en GenBank incluida la muestra de El Quinche. El análisis filogenético de máxima verosimilitud mostró la ubicación de *Alstroemeria Necrotic Steak Virus* muestra El Quinche, en el mismo clado donde se encuentran los estudios colombianos de *Alstroemeria Necrotic Steak Virus* (ANSV) con un valor alto de soporte bootstrap, al igual que el análisis a través de inferencia bayesiana donde el clado con la accesión de Ecuador y las accesiones de Colombia, vuelven a ser observadas con valor alto de probabilidad, lo que sugiere una cercanía evolutiva del virus de ANSV encontrada en Ecuador y las accesiones colombianas. El estudio filogenético de máxima verosimilitud realizada por Gallo Yuliana (Gallo et al., 2019) ubica en el mismo clado americano a virus del género *Orthospovirus* con las accesiones que presenta en sus estudios de Colombia que mostraron infección natural en naranjilla (*Solanum quitoense*), tomate (*S. lycopersicum*) y pimiento (*Capsicum annum*).

La European Food Safety Authority (EFSA), en un reporte de sanidad vegetal para el año 2012, categoriza en diferentes grupos de riesgo a 24 *Orthospovirus*, considerando parámetros como la presencia del virus, la existencia de plantas hospederas, la existencia de especies de trips vectores y el potencial daño a los cultivos, siete de estos 24 virus, dentro los cuales está *Alstroemeria necrotic streak virus* (ANSV), representan el mayor riesgo como plaga en la Unión Europea (Health, 2012).

La identificación de ANSV en el presente trabajo, debe encender las alertas sobre los riesgos fitosanitarios que podrían desarrollarse en Ecuador. Olaya en 2017, ya menciona los posibles riesgos sobre cultivos de importancia económica en Colombia y otros países andinos,

así como menciona la necesidad de evaluar la distribución geográfica del patógeno (Olaya et al., 2017), al igual que la preocupación que muestra (Tian et al., 2020) en San Diego California sobre la presencia del virus en plantas ornamentales. Los *orthospovirus* dentro de los cuales se ubica ANSV, se consideran patógenos agrícolas que representan un grave riesgo para la seguridad alimentaria en todo el mundo (Oliver & Whitfield, 2016).

En un estudio sobre la determinación de especies de insectos del orden *Thysanoptera* realizada en fincas florícolas de Ecuador para el año 2014, se determinó la presencia de insectos *thrips* en un 80 % de *Frankliniella occidentalis*, 7.5% de *Frankliniella Tuberosi* y *Thrips tabaci* y un 5% de *Thrips palmi*, principales vectores de virus fitopatógenos (Godoy et al., 2014). Como también lo menciona (Jones, 2005; Oliver & Whitfield, 2016) donde los virus generalmente se transmiten de manera persistente y propagativa por insectos del orden *Thysanoptera*, es necesario entonces tomar consideraciones sobre los cuidados agronómicos que deberían implementarse en programas nacionales para el cuidado de la propagación de patógenos virales a plantas de interés económico y alimenticio.

Capítulo 5

Conclusiones

- Aplicando herramientas de secuenciación de alto rendimiento (HTS-High Throughput Sequencing) y programas bioinformáticos, fue posible identificar virus fitopatógenos presentes en la muestra de hojas de flores de corte de *Limonium sinuatum*.
- La presencia de *Alstroemeria Necrotic Streak Virus* (ANSV) fue de un 97%, el 3% restante de los virus fitopatógenos encontrados en la muestra foliar, corresponden a la secuencia de otros tres virus: *Satellite tobacco necrosis virus 2* (0.2%), *Bloomfield like new virus* (2%) y *New chrysovirus* (0.8%), de los cuales aún no existen reportes.
- Fue posible ensamblar el genoma completo del virus *Alstroemeria Necrotic Streak Virus* (ANSV) con sus tres segmentos, *segmento L* con 8755 pb, *segmento M* con 4869 pb y *segmento S* con 3135 pb.
- Se realizó la anotación del virus de ANSV y su reporte se encuentra registrado en GenBank con los números de accesión *ON149273* para el *segmento L*, accesión *ON149272* para el *segmento M* y accesión *ON149271* para el *segmento S*.
- Se realizó una publicación indexada en la revista *Microbiology Resource Annoucenment* con la información de la presencia del virus fitopatógeno *Alstroemeria Necrotic Streak Virus* (ANSV) en plantas de corte de *Limonium sinuatum* en Ecuador.
<https://journals.asm.org/eprint/MXKPEHT43FUXI9HXXAGQ/full>.

- Se realizó la reconstrucción evolutiva para *Alstroemeria Necrotic Streak Virus* (ANSV) en base a las secuencias de los aminoácidos del *gen N de la nucleocapside* y de las secuencias genómicas de los *Orthospovirus* disponibles en GenBank, determinando su cercanía evolutiva del virus de ANSV- muestra El Quinche, al clado donde se encuentran los reportes colombianos.
- Existe alta posibilidad que el agente causal de la sintomatología observada en campo para malformación de hojas, enanismo y rayado necrótico de hojas “streaking” en las hojas de flores de corte de *Limonium sinuatum*, sean provocadas por *Alstroemeria Necrotic Streak Virus* (ANSV).

Recomendaciones

- Se recomienda la aplicación de la secuenciación de alto rendimiento (HTS-High Throughput Sequencing) propuestas en el pipeline del presente trabajo, como una herramienta en la determinación de secuencias de virus fitopatogénos en muestras de flores de corte de *Limonium sinuatum* (L.) Mill con posible extrapolación a otros cultivos de interés.
- Realizar más muestreos foliares a nivel de campo e invernaderos, en fincas pequeños y grandes agricultores de plantas de corte, para determinar si el virus encontrado es el único presente en el Ecuador.
- Se recomienda desarrollar un método rápido de diagnóstico mediante el uso de PCR para la detección de ANSV en *Limonium sinuatum* (L.) Mill.
- Desarrollar un plan estratégico de control fitosanitario integral a nivel de fincas florícolas y cultivos de interés económico, dado que el vector de propagación del virus ANSV pertenece al orden de insectos *Thysanoptera* cuya presencia en Ecuador ha sido ya reportada como problema sanitario.

Bibliografía

- Alavandi, S., & Poornima, M. (2012). Viral metagenomics: a tool for virus discovery and diversity in aquaculture. *Indian Journal of Virology*, 23(2), 88-98.
- Barba, M., Czosnek, H., & Hadidi, A. (2014). Historical perspective, development and applications of next-generation sequencing in plant virology. *Viruses*, 6(1), 106-136.
- Breitbart, M., & Rohwer, F. (2005). Method for discovering novel DNA viruses in blood using viral particle selection and shotgun sequencing. *Biotechniques*, 39(5), 729-736.
- Brenes-Guillén, L. (2019). Ensamblaje y anotación de genomas microbianos. *Revista de biología tropical*, Blog-Blog.
- Brooks, D. R., Bilewitch, J., Condy, C., Evans, D. C., Folinsbee, K. E., Fröbisch, J., . . . Mattern, M. (2007). Quantitative phylogenetic analysis in the 21st century. *Revista mexicana de biodiversidad*, 78(2), 225-252.
- Castro, C., Castañeda, Z., Ruiz, K., González, G., & Poveda, G. (2020). El sector florícola ecuatoriano y su afectación en el mercado internacional producto de la pandemia causada por el Covid-19. Congreso INternacional Virtual sobre COVID-19. Consecuencias psicológicas, sociales, políticas y económicas,
- de Blas, I. O. (1992). Morfología de *Limonium sinuatum* (L.) Miller (Plumbaginaceae). *Revista de biología tropical*, 40, 11-17.
- de Oliveira, A. S., Bertran, A. G. M., Inoue-Nagata, A. K., Nagata, T., Kitajima, E. W., & Oliveira Resende, R. (2011). An RNA-dependent RNA polymerase gene of a distinct Brazilian tospovirus. *Virus Genes*, 43(3), 385-389.
- Felsenstein, J. (2004). *Inferring phylogenies* (Vol. 2).

- Gallo, Y., Sierra, A., Muñoz, L., Marín, M., & Gutiérrez, P. A. (2019). Genome characterization of three Alstroemeria necrotic streak orthotospovirus (ANSV) isolates naturally infecting bell pepper (*Capsicum annuum*) in Antioquia (Colombia). *Tropical Plant Pathology*, *44*(4), 326-334.
- Gallo, Y., Toro, L. F., Jaramillo, H., Gutiérrez, P. A., & Marín, M. (2018). Identificación y caracterización molecular del genoma completo de tres virus en cultivos de lulo (*Solanum quitoense*) de Antioquia (Colombia). *Revista Colombiana de Ciencias Hortícolas*, *12*(2), 281-292.
- Godoy, S., Tigrero, J., & Taípe, M. (2014). *Determinación de especies de insectos de la familia Thysanoptera: thripidae que afectan al cultivo de rosas en dos zonas florícolas de Pichincha–Ecuador* [Pregrado, Universidad de las Fuerzas Armadas ESPE]. Agrocalidad.
- González-Garza, R. (2017). Evolución de técnicas de diagnóstico de virus fitopatógenos. *Revista mexicana de fitopatología*, *35*(3), 591-610.
- Hassani-Mehraban, A., Botermans, M., Verhoeven, J. T. J., Meekes, E., Saaijer, J., Peters, D., . . . Kormelink, R. (2010). A distinct tospovirus causing necrotic streak on *Alstroemeria* sp. in Colombia. *Archives of Virology*, *155*(3), 423-428.
- Health, E. P. o. P. (2012). Scientific Opinion on the pest categorisation of the tospoviruses. *EFSA Journal*, *10*(7), 2772.
- Huelsenbeck, J. P., Ronquist, F., Nielsen, R., & Bollback, J. P. (2001). Bayesian inference of phylogeny and its impact on evolutionary biology. *science*, *294*(5550), 2310-2314.
- Hull, R. (2009). *Comparative plant virology*. Academic press.
- Jaramillo Mesa, H. (2017). Analisis del transcriptoma y viroma de *Passiflora edulis* f. *edulis* en cultivos de Antioquia utilizando métodos de secuenciación de nueva generación. *Escuela de Biociencias*.
- Jeong, J.-j., Ju, H.-j., & Noh, J. (2014). A review of detection methods for the plant viruses. *Research in Plant Disease*, *20*(3), 173-181.

- Jones, D. R. (2005). Plant viruses transmitted by thrips. *European journal of plant pathology*, 113(2), 119-157.
- King, A. M., Lefkowitz, E., Adams, M. J., & Carstens, E. B. (2011). *Virus taxonomy: ninth report of the International Committee on Taxonomy of Viruses* (Vol. 9). Elsevier.
- Kleczkowski, A. (1965). A study of the effects of salt and of pH on precipitation of antigen—antibody compounds. *Immunology*, 8(2), 170.
- Larson, R. A. (2012). *Introduction to floriculture*. Elsevier.
- Lukashin, A. V., & Borodovsky, M. (1998). GeneMark. hmm: new solutions for gene finding. *Nucleic acids research*, 26(4), 1107-1115.
- Matthews, R. E. F. (2012). *Fundamentals of plant virology*. Academic Press.
- Naidu, R. (2001). Hughes, JdA Methods for the detection of plant virus diseases. Proceedings of the Conference on Plant Virology in Sub Saharan Africa organized by IITA, Ibadan, Nigeria,
- Naidu, R., & Hughes, J. d. A. (2003). Methods for the detection of plant virus diseases. Plant Virology in Sub-Saharan Africa: Proceedings of a Conference Organized by IITA: 4-8 June 2001, International Institute of Tropical Agriculture, Ibadan, Nigeria,
- Nichol, S., Beaty, B., Elliot, R., Goldbach, R., Pljusnin, A., Schmaljohn, C., & Tesh, R. (2005). VIIIth report of the International Committee on Taxonomy of Viruses. In *Virus taxonomy: VIIIth report of the International Committee on Taxonomy of Viruses* (pp. 695-716). Elsevier Academic Press.
- Nieuwenhuijse, D. F., & Koopmans, M. P. (2017). Metagenomic sequencing for surveillance of food-and waterborne viral diseases. *Frontiers in microbiology*, 8, 230.
- Olaya, C., Velásquez, N., Betancourt, M., Cuellar, W., & Pappu, H. (2017). First report of natural infection of Alstroemeria necrotic streak virus on tomato (*Solanum lycopersicum*) and bell pepper (*Capsicum annuum*) in Colombia. *Plant disease*, 101(6), 1065-1065.

- Oliver, J., & Whitfield, A. (2016). The genus Tospovirus: emerging bunyaviruses that threaten food security. *Annual Review of Virology*, 3, 101-124.
- Olmos, A., Boonham, N., Candresse, T., Gentit, P., Giovani, B., Kutnjak, D., . . . Moreira, A. (2018). High-throughput sequencing technologies for plant pest diagnosis: challenges and opportunities. *EPPO Bulletin*, 48(2), 219-224.
- Owens, R., & Diener, T. (1984). Spot hybridization for detection of viroids and viruses. In *Methods in virology* (Vol. 7, pp. 173-187). Elsevier.
- Pantaleo, V., & Chiumenti, M. (2018). *Viral Metagenomics*. Springer New York.
- Paredes Barros, M. X. (2019). *Posicionamiento de los floricultores ecuatorianos, frente a los cambios y tendencias del mercado ruso, americano, europeo y chino en la exportación de rosas de corte, como fuente para la toma de decisiones* [Maestría, QUITO/UIDE/2019].
<https://repositorio.uide.edu.ec/handle/37000/3834>
- Pareek, C. S., Smoczynski, R., & Tretyn, A. (2011). Sequencing technologies and genome sequencing. *Journal of applied genetics*, 52(4), 413-435.
- Pedros, J. R. B. (1994). *Limonium sunuatum* " Statice" en alza. *Horticultura: Revista de industria, distribución y socioeconomía hortícola: frutas, hortalizas, flores, plantas, árboles ornamentales y viveros*(94), 13-19.
- Peña, C. (2011). Métodos de inferencia filogenética. *Revista peruana de Biología*, 18(2), 265-267.
- Prabha, K., Baranwal, V., & Jain, R. (2013). Applications of next generation high throughput sequencing technologies in characterization, discovery and molecular interaction of plant viruses. *Indian Journal of Virology*, 24(2), 157-165.
- PRO-ECUADOR. (2015). Análisis Sectorial Flores de Verano 2015. *Instituto de Promoción de Exportaciones e Inversiones*, 14. WWW.proeducador.gob.ec

- Richardson, E. J., & Watson, M. (2013). The automatic annotation of bacterial genomes. *Briefings in bioinformatics*, 14(1), 1-12.
- Riera Yépez, S. J., & Zuleta Colimba, V. N. (2013). *Análisis y evaluación de las exportaciones de flores desde Ecuador a la federación de Rusia para el período 2006-2011*
<http://dspace.ups.edu.ec/handle/123456789/5867>
- Rodríguez Sánchez, I. P., & Barrera Saldaña, H. A. (2004). La reacción en cadena de la polimerasa a dos décadas de su invención. *Ciencia UANL*, 7(3).
- Tamay de Dios, L., Ibarra, C., & Velasquillo, C. (2013). Fundamentos de la reacción en cadena de la polimerasa (PCR) y de la PCR en tiempo real. *Investigación en discapacidad*, 2(2), 70-78.
- Tian, T., Guo, Y., Posis, K., Belanger, C. A., & Roy, A. (2020). First Report of Alstroemeria Necrotic Streak Virus on *Hoya carnosa* and *H. kerrii* in the United States. *Plant disease*, 104(1), 298-298.
- Van der Walt, A. J., Van Goethem, M. W., Ramond, J.-B., Makhalaryane, T. P., Reva, O., & Cowan, D. A. (2017). Assembling metagenomes, one community at a time. *BMC genomics*, 18(1), 1-13.
- Willner, D. L. (2010). *Viral metagenomics in host-associated systems*. University of California, San Diego and San Diego State University.