Identificación de genes envueltos en la interacción entre *Fusarium oxysporum* f. sp. *cubense* RT4 y *Musa acuminata* subgrupo Cavendish utilizando minería de datos.

Montesinos Ludeña, Stefanía Soledad

Departamento de Ciencias de la Vida y de la Agricultura

Carrera de Ingeniería en Biotecnología

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

Flores Flor, Francisco Javier PhD.

5 de marzo del 2021

URKUND

Document Information

Analyzed document	Montesinos_Stefania_tesis_borrador .docx (D96768460)
Submitted	2/28/2021 3:41:00 PM
Submitted by	
Submitter email	pjludena@utpl.edu.ec
Similarity	1%
Analysis address	pjludena.utpl@analysis.urkund.com

Sources included in the report

W	URL: https://prod.senasica.gob.mx/SIRVEF/ContenidoPublico/Fichas%20tecnicas/Ficha%20T%C ... Fetched: 7/26/2020 3:49:52 AM	1
SA	**Monografia Fusarium Oxysporum raza 4 en el cultivo de Banano.docx** Document Monografia Fusarium Oxysporum raza 4 en el cultivo de Banano.docx (D66810483)	1
W	URL: https://www.redalyc.org/jatsRepo/437/43761812020/html/index.html Fetched: 12/19/2020 9:27:57 AM	1
SA	**Tesis Figueroa URKUND.doc** Document Tesis Figueroa URKUND.doc (D77228082)	1
W	URL: https://theses.liacs.nl/pdf/2018-2019-ParsaniaMina.pdf Fetched: 2/28/2021 3:42:00 PM	1
W	URL: https://es.qaz.wiki/wiki/List_of_RNA-Seq_bioinformatics_tools Fetched: 1/15/2021 11:31:56 AM	2
SA	**M0.181_20192_PEC4 -Cierre de la memoria_12683828.txt** Document M0.181_20192_PEC4 -Cierre de la memoria_12683828.txt (D75574025)	1
SA	**M0.181_20192_PEC4 -Cierre de la memoria_12683976.txt** Document M0.181_20192_PEC4 -Cierre de la memoria_12683976.txt (D75575174)	1

DEPARTAMENTO DE CIENCIAS DE LA VIDA Y DE LA AGRICULTURA

CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

CERTIFICACIÓN

Certifico que el trabajo de titulación, **"Identificación de genes envueltos en la interacción entre *Fusarium oxysporum* f. sp. *cubense* RT4 y *Musa acuminata* subgrupo Cavendish utilizando minería de datos"** fue realizado por la señorita ***Montesinos Ludeña, Stefania Soledad*** el mismo que ha sido revisado en su totalidad, analizado por la herramienta de verificación de similitud de contenido; por lo tanto cumple con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, razón por la cual me permito acreditar y autorizar para que lo sustente públicamente.

Sangolquí, 29 de marzo de 2021

Firma:

Firmado electrónicamente por:
FRANCISCO JAVIER FLORES FLOR

Francisco Javier Flores Flor , Ph.D.

C.C. 1713443479

DEPARTAMENTO DE CIENCIAS DE LA VIDA Y DE LA AGRICULTURA

CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

RESPONSABILIDAD DE AUTORÍA

Yo, **Montesinos Ludeña Stefanía Soledad**, con C.C. 1105054363, declaro que el contenido, ideas y criterios del trabajo de titulación: **"Identificación de genes envueltos en la interacción entre *Fusarium oxysporum* f. sp. *cubense* RT4 y *Musa acuminata* subgrupo *Cavendish* utilizando minería de datos"** es de mi autoría y responsabilidad, cumpliendo con los requisitos teóricos, científicos, técnicos, metodológicos y legales establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 23 de marzo 2021

Montesinos Ludeña, Stefanía Soledad

C.C. 1105054363

ESPE
UNIVERSIDAD DE LAS FUERZAS ARMADAS
INNOVACIÓN PARA LA EXCELENCIA

DEPARTAMENTO DE CIENCIAS DE LA VIDA Y DE LA AGRICULTURA

CARRERA DE INGENIERÍA EN BIOTECNOLOGÍA

AUTORIZACIÓN DE PUBLICACIÓN

Yo, **Montesinos Ludeña Stefanía Soledad, con C.C. 1105054363** autorizo a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de titulación: **Identificación de genes envueltos en la interacción entre *Fusarium oxysporum* f. sp. *cubense RT4 y Musa acuminata* subgrupo *Cavendish* utilizando minería de datos** en el Repositorio Institucional, cuyo contenido, ideas y criterios es de mi responsabilidad

Sangolquí, 23 de marzo 2021

Montesinos Ludeña, Stefanía Soledad

C.C. 1105054363

DEDICATORIA

A mis padres Soledad y Manuel

por su incondicional apoyo

en cada momento de mi vida

y por mantener siempre constante

su confianza en mi.

A mi hermana Romina,

por brindarme todo su apoyo.

A mis abuelitos, tíos y primos,

que han representado mi apoyo

en todo momento.

Stefania Soledad Montesinos Ludeña

AGRADECIMIENTOS

A mis padres, por ser mi constante apoyo e impulso como modelos a seguir durante esta etapa de mi vida. Cada uno contribuyendo con su tiempo y enseñanzas para mi formación.

A mi hermana, que ha sido mi mejor amiga y me ha ayudado a aprender en cada una de las aventuras que hemos vivido juntas.

A mi mentor Dr. Francisco Flores, por depositar su confianza en mí, permitiéndome ser parte de su equipo de trabajo y por encaminarme en la rama de Bioinformática. A Alma Koch, por brindarme todo su apoyo tanto en mi vida personal como en la tesis.

A todos mis profesores, que han contribuido en esta fase de mi vida y que siempre llevaré en mi memoria sus enseñanzas tanto de vida como académicas para plasmarlas en mi vida profesional. A todos mis compañeros de aula, que directa o indirectamente han contribuido en mi desarrollo académico.

A mis amigos, Pamela, Camila, Gabriela, Jennifer y Paul por haber compartido conmigo toda esta experiencia dentro y fuera de la universidad.

Stefania Soledad Montesinos Ludeña

ÍNDICE DE CONTENIDO

LISTADO DE TABLAS

LISTADO DE FIGURAS

LISTADO DE ABREVIATURAS

NRPS	Sintetasa peptídica no ribosomal
krs	Kurstakina
ADNc	Ácido desoxirribonucleico complementario
ARNm	Ácido ribonucleico mensajero
RNA-seq	Secuencias de ADNc útiles para análisis transcriptómicos
BLAST	Herramienta de búsqueda de secuencias de ADN o proteicas
NCBI	Centro Nacional de Información Biotecnológica de Estados Unidos
PCR	Reacción en cadena de polimerasa
SinR	Factor de represión de transcripción génica
SpoA	Factor de activación de transcripción en una porción de células
AbrB	Factor de represión de transcripción génica
ComK	Factor de activación de transcripción génica
mL	Mililitros
g	Gramos
µg	Microgramos
µL	Microlitros
h	Horas
ºC	Grados centígrados

RESUMEN

En la actualidad, la economía mundial se encuentra afectada debido a las grandes pérdidas de plantaciones de banano (*Musa acuminata* sub. Cavendish) por infección del mal de Panamá causado por *Fusarium oxysporum* f.sp. *cubense* raza tropical 4 (FOCRT4). Es imperativo dar alternativas para la sostenibilidad del cultivo, por ello se requieren medidas para evitar la diseminación del patógeno. Una vez que el hongo está presente en las plantaciones no puede ser controlado con fungicidas, siendo como alternativa rápida y eficiente la incineración de los sembríos para su eliminación. Por ende, la alternativa de recopilación de genes esenciales para el hongo al momento de la interacción con la planta es clave para atacar eficientemente al hongo, mediante el silenciamiento de genes que codifiquen a proteínas esenciales en la infectibilidad o patogenicidad, y así evitar la infección y colonización de la planta. En la presente investigación, mediante minería de datos, se detectaron 12.810 genes del hongo que fueron expresados diferencialmente (DEGs) a los cero, dos y cuatro días después de la inoculación, durante la interacción FOCRT4-banano. Se filtraron por su p-value y se encontraron aproximadamente 1249 DEGs y de ellos 310 representaban proteínas caracterizadas como DNA polimerasa, proteína SIX, betagalactosidasa entre otras. La mayoría de los 939 genes restantes fueron encontrados como proteínas hipotéticas sin caracterizar.

PALABRAS CLAVES:

- ***FUSARIUM OXYSPORUM CUBENSE RAZA TROPICAL 4***
- ***BANANO***
- ***DEGs.***

ABSTRACT

At present, the world economy is affected due to the great losses of banana plantations (Musa acuminata sub. Cavendish) due to infection of the Panama disease caused by Fusarium oxysporum f.sp. cubense tropical race 4 (FOCRT4). It is imperative to provide alternatives for the sustainability of the crop, therefore measures are required to prevent the spread of the pathogen. Once the fungus is present in the plantations, it cannot be controlled with fungicides, being the incineration of the crops for their elimination as a quick and efficient alternative. Therefore, the alternative of collecting essential genes for the fungus at the time of interaction with the plant is key to efficiently attack the fungus, by silencing genes that encode essential proteins in infectibility or pathogenicity, and thus avoid infection. and colonization of the plant. In the present investigation, through data mining, 12,810 genes of the fungus that were differentially expressed (DEGs) were detected at zero, two and four days after inoculation, during the FOCRT4-banana interaction. They were filtered by their p-value and approximately 1249 DEGs were found, of which 310 represented proteins characterized as DNA polymerase, SIX protein, beta-galactosidase among others. Most of the remaining 939 genes were found as uncharacterized hypothetical proteins.

KEY WORDS:

- ***FUSARIUM OXYSPORUM CUBENSE RAZA TROPICAL 4***
- ***BANANO***
- ***DEGS***

Capítulo I: INTRODUCCIÓN

1.1 Formulación del problema

Ecuador es uno de los principales exportadores de banano en el mundo, atribuyéndosele un 33% del mercado internacional (FAO, 2020). Desde la perspectiva económica, en 2017, la exportación bananera representó el 2% del producto interno bruto con un rubro de 3000 millones de dólares (Ministerio de Comercio Exterior, 2017). El Ministerio de Agricultura, Ganadería, Acuacultura y Pesca (MAGAP) informó que existen aproximadamente 162,039 ha de cultivos de banano, de las cuales el 12% pertenece a banano orgánico y el 88% a banano convencional, siendo estos últimos susceptibles al hongo.

La fusariosis causada por FOCRT4, es una de las enfermedades más destructivas del plátano y banano que existen. Se ha reportado pérdidas de alrededor de 500 millones de dólares en más de 16 países. Actualmente, el patógeno se encuentra en Colombia, perdiéndose, hasta el momento, aproximadamente 175 ha de plantaciones de banano tipo Cavendish (AUGURA, Asociación de bananeros de Colombia, 2019). Al estar presente en el país vecino, Ecuador se encuentra en gran amenaza, pudiéndose ver afectadas aproximadamente 300 000 ha de plantaciones.

Según estudios epidemiológicos realizados por Scheerer y colaboradores (2018), se estima que FOCRT4 llegué a Ecuador en los próximos 10 años. Por ende, el Ministerio de la Agricultura, la Ganadería y la Pesca ha empleado protocolos de prevención, tales como: desinfección de vehículos de transporte, control en fronteras y aduanas, control fitosanitario en explantes, etc. (MAGAP, 2019). Esto último, ya que el hongo tiene alta tasa de diseminación a través del suelo, aire y agua. Se ha reportado la presencia de FOCRT4 en riachuelos, charcos

cercanos a plantaciones, rizomas, herramientas de trabajo e, incluso, en el aire como clamidosporas (Ordonez, et al., 2015).

Esfuerzos por la prevención de la entrada del patógeno a países productores han sido en vano, ya que se ha confirmado su presencia en cuatro continentes: Asía, África, Oceanía y América. Por lo que, se debe enfocar el trabajo en el desarrollo de técnicas de diagnóstico temprano y combate del hongo, con el fin de ser un apoyo extra para las estrategias de acordonamiento y cuarentena. Además, existen pocos estudios de posibles tratamientos para el hongo al momento que se encuentra colonizando las raíces. El sistema más usado es la incineración de los cultivares, perdiéndose la mayoría de la producción.

De tal manera, el hongo al ser un patógeno agresivo para las plantaciones de banano tipo Cavendish y estar cada más cerca de las fronteras ecuatorianas, con este estudio se prevé identificar genes que estén involucrados en la infección y colonización de las raíces por el hongo, con el fin de servir para futuros estudios de posibles ataques directos a genes esenciales del fitopatógeno

1.2 Justificación del problema

En Ecuador el proceso de diagnóstico del FOC RT4, debe realizarse a través de la Agencia de Regulación y Control Fito y Zoosanitario (Agrocalidad). Esta entidad posee una serie de parámetros y procesos que dependen de la ubicación de las haciendas, del tiempo hasta que se observaron síntomas y del número de muestras por analizar. El procesamiento óptimo puede demorar desde 3 a 12 días desde el momento de la toma de la muestra hasta la determinación de la presencia del hongo (Agrocalidad, 2018). Debido a la alta tasa de propagación del hongo,

este tiempo es perjudicial para las plantaciones, teniéndose que eliminar más del 30% por cada hectárea de cultivo si existe infección en una sola planta y con el peligro que el resto de plantas ya hayan sido contaminadas.

Con respecto a la observación de la presencia de síntomas, se ha reportado un lapso considerable entre la intrusión del fitopatógeno a un país y su detección, ocasionando mayores pérdidas tanto de sembríos como económicas. La razón es que FOC RT4 causa sintomatología similar a otras enfermedades como, virosis, daños por insectos, deshidratación, presencia de otros hongos, etc., confinando el diagnóstico temprano y las medidas de contención que se pudieran tomar (FAO, 2020). Este caso se presenció en Jordania, el hongo pasó desapercibido por más de 8 años, causando la pérdida del 80% de hectáreas productivas de banano (Magdama F. , 2019).

FOCRT4 no puede ser erradicado mediante métodos convencionales como el uso de fungicidas, sino la planta infectada debe ser eliminada junto a las plantas vecinas. Por esta razón, se busca identificar los genes envueltos en la colonización, infección e interacción del hongo con el hospedero, con el fin de reportarlos y que sean útiles para nuevas investigaciones de edición génica para controlar la plaga o evitar la colonización de la planta.

1.3 Objetivos de la investigación

1.3.1 Objetivo general

Identificar genes envueltos en la interacción entre *Fusarium oxysporum* f. sp*. cubense* RT4 y *Musa acuminata* subgrupo Cavendish utilizando minería de datos.

1.3.2 Objetivos específicos

- Analizar datos de RNA-seq existentes para estudios de la interacción entre *Fusarium oxysporum* f. sp*. cubense* RT4 y *Musa acuminata* subgrupo Cavendish utilizando referencias actualizadas.

- Identificar genes con expresión diferencial durante la colonización de *Fusarium oxysporum* f. sp*. cubense* RT4 y *Musa acuminata* subgrupo Cavendish.

- Anotar genes efectores del hongo *Fusarium oxysporum* f. sp*. cubense* RT4 que no han sido identificados previamente.

1.4 Marco teórico

1.4.1 *Musa acuminata* subgrupo Cavendish

Musa acuminata subgrupo Cavendish es un cultivar de tipo triploide, se distinguen por ser una planta alta y de cosecha perenne de frutos característicos durante todo el año (Gamez, et al., 2019). La parte visible de la planta, el pseudotema y las hojas, mueren después que da fruto para permitir que los brotes jóvenes suplanten a la madre y así se rejuvenezcan los rizomas. Estos ciclos suelen ser perpetuados por ilimitadas generaciones.

El consumo de frutos frescos de musáceas (plátano o banano) es alto en comparación a las demás frutas (piña, papaya, uva, manzana, etc.), por su alto contenido nutricional y sabor característico; además representan un alimento básico para más de 400 millones de personas a nivel mundial (Martínez-Solórzano, Rey-Brina, Pargas-Pichardo, & Manzanilla, 2020).

Únicamente ha sido superado por el consumo de cítricos, sin embargo, la demanda de banano va en aumento.

El subgrupo Cavendish, es la variedad de mayor producción a nivel mundial, acreedor del 80%, principalmente de los clones “Gran Enano” y “Valery”. El porcentaje restante se divide entre más cultivares como Lady’s Finger (bananos de postre), Switer (consumo cocido), Prata (sabor agridulce), entre otros (Soto, 2011).

En comparación con los genotipos diploides silvestres, la mayoría de cultivares diploides y triploides comerciales son infértiles, es decir se reproducen asexualmente. La micropropagación de tejidos vegetales es la clave para la creación, expansión y mantenimiento de los monocultivos. Esto último puede ser contraproducente, ya que al contener una base genética estrecha los cultivares van a carecer de resistencia a plagas (Castañeda, et al., 2017).

1.4.1.1 Resistencia de cultivares de banano

La mayoría de la producción de banano tipo Gros Michel, durante la década de 1960, fue perdida a causa de la susceptibilidad que presentaba frente a *Fusarium oxysporum* f.sp. *cubense* raza tropical 1. Se desecharon cuantiosas hectáreas de sembríos y la economía de cada uno de los países productores decayó considerablemente. Por ello, bananeros eludieron la infección de fusarium raza 1, reemplazando la especie Gros Michel por el subgrupo Cavendish, que era un cultivar resistente al mencionado patógeno y se convirtió en la variedad predominante en plantaciones de África, Caribe y América del Sur (INIBAP, 1994).

La variedad Cavendish es susceptible a Sigatoka negra (*Mycosphaerella fijiensis*) y Moko (*Ralstonia solanacearum*), por ello, se han realizado estudios para la prevención y el tratamiento

de esta enfermedad, con el fin de aplacar la muerte de las plantaciones y además la reducción del uso de fungicidas (Fullerton & Olsen, 1995). Sin embargo, la sintomatología de problemas abióticos como deficiencia de nitrógeno, deficiencia de potasio, sequía o frío, es similar a FOC RT4 como el amarillamiento de las hojas, pudrición de las raíces y ennegrecimiento del tallo, consiguientemente se dificulta el diagnóstico temprano ante la presencia de fusarium, siendo una barrera para la aplicación de sistemas de contención (Molina, 2009). Según Scheerer y colaboradores (2018), alrededor el 80% de las plantaciones de banano serán devastados en los próximos 20 años, debido al tardío reconocimiento y la alta agresividad de FOC RT4. La fusariosis en banano se considera una de las diez enfermedades vegetales más importantes, por ser letal e infecciosa, al no existir control y perdurar en el suelo durante largos períodos en estado latente (Li, et al., 2017).

1.4.1.2 Importancia económica de *Musa acuminata* sub. Cavendish

La producción de banano y plátano se da principalmente en América Latina, el Caribe, Asia (Filipinas), África y Europa (Islas Canarias) (Robinson & Saúco, 2010). Ecuador se encuentra como líder en la exportación de las musáceas comestibles con un total de 32.52% en 2020, seguido por Filipinas con 13.2% y Colombia con el 11% (AEBE, 2020).

En 2020, se exportaron 352.317.647 cajas de banano y plátano desde Ecuador, con destino a África, Oceanía, Asia, Medio Oriente, Estados Unidos y la Unión Europea. Argelia, aumentó la importación de banano en un 25% este año. Además, hubo crecimiento del 9,92% de exportaciones de banano en el Ecuador, comparado con el año 2019, lo que indica un crecimiento del consumo y del PIB (producto interno bruto) (AEBE, 2020).

La cosecha de frutos de plátano y banano genera una entrada de $2254.832.941. El MAGAP (2020) estableció que el costo por caja en 2020 sería de $6,40. Sin embargo, el fruto no es la única parte útil de la planta, también el pseudotallo puede ser utilizado como materia prima para la elaboración de papel, como alternativa sustentable. Muchas familias, reutilizan las hojas de banano o plátano para la elaboración de platos típicos o para decoración (Alarcón & Marzocchi, 2015).

Ecuador, en el 2020, llegó a exportar casi el 25% de la producción de banano en todo el mundo. El comercio de banano contribuye enormemente a la economía nacional, en Ecuador, es la segunda fuente de ingreso después de la exportación de petróleo (Soto, 2011).

1.4.2 *Fusarium oxysporum* f. sp. *cubense* raza tropical 4

Fusarium oxysporum f.sp. *cubense* raza tropical 4 es el causante de la fusariosis, o también conocida como el Mal de Panamá. Es un hongo asexual que produce tres tipos de esporas, microconidos, macroconidios y clamidosporas durante todo el ciclo de vida, y así garantiza su supervivencia. Las clamidosporas son estructuras latentes, pueden permanecer hasta 30 años en ausencia de huéspedes adecuados o en malezas hospedantes de manera no patógena (Guo, et al., 2014). El suelo infestado con FOC RT4 se vuelve inadecuado para la plantación de nuevos explantes, ya que a diferencia de las otras enfermedades de banano, no se puede controlar mediante fungicidas (Stover, 1962).

El hongo FOC RT4 es patógeno de todos los hospederos de razas 1 y 2 (“Gros Michel”, “Bluggoe”) y del subgrupo Cavendish. A la vez existen dos razas de *fusarium oxysporum* f.sp. *cubense*: raza tropical (RT4) y la raza subtropical (RST4), causando enfermedad para las zonas

tropicales y subtropicales y exclusivamente para los subtrópicos, respectivamente (Nasdir, 2003).

1.4.2.1 Interacción con el huésped

FOC RT4 coloniza a la musácea a través de sus raíces, atraviesa el rizoma y asciende por el pseudotallo por el xilema, bloqueando el paso de agua y nutrientes hacia las partes aéreas de la planta (Guo, et al., 2014).

Los primeros síntomas externos del mal de Panamá comienzan con el amarillamiento y marchitamiento de las hojas más viejas y continúa hacia las más jóvenes, terminando en la muerte de la planta. Los síntomas internos, muestran necrosis y coloración marrón en el xilema tanto de pseudotallos como en los rizomas (FAO, 2020).

FOC RT4 ha limitado drásticamente la producción de banano desde su detección en 1990. Afecta a las plantaciones de Cavendish, principal subgrupo productor. Si bien está claro que la incidencia varía dependiendo de piso climático, calidad del cultivo, ésta enfermedad puede extenderse al 100% de los cultivos si no se detecta oportunamente.

La propagación de material de plantación contaminado y la falta de esfuerzos por diversificar los sembríos de banano, han ocasionado un aumento en la vulnerabilidad hacia patógenos incontrolables. Durante todos estos años, se han realizado estudios basados en la prevención del hongo (Molina, 2009). En Filipinas, se llevó a cabo un plan de diagnóstico en campo mediante un biosensor, lo que proporcionó mayor cobertura de detección, identificación temprana y decisiones rápidas y eficientes. En función de la alta tasa de diseminación del hongo y su rápida infectividad del huésped, urge tomar medidas de bioseguridad a nivel internacional,

regional y local para evitar la diseminación de FOC-RT4. Sin embargo, la investigación se ve limitada a la prevención y no al tratamiento de la enfermedad una vez colonizada la planta (FAO, 2020).

1.4.2.2 Impacto y sistemas de regulación

En el Ecuador, la Agencia de Regulación y Control Fitosanitario (Agrocalidad) es la encargada de tanto realizar y aplicar el plan de contingencia para la prevención de la entrada de FOC RT4 al país como del diagnóstico de fusariosis en banano (MAGAP, 2019).

Para el mecanismo de exclusión, Agrocalidad ha implementado un sistema monitorizado en aduanas, haciendas, aeropuertos y vehículos (Agrocalidad, 2018). Estas medidas son extremadamente dependientes de preparación del personal, herramientas de diagnóstico, asesoramiento legal. Estas incluyen: medidas fitosanitarias para la importación de plantas, productos vegetales y del material vegetal para propagación; desinfección de herramientas y maquinaria agrícola, control en puntos de entrada (aduanas y fronteras), control de equipaje, tratamiento de calzado, control de medios de transporte, tratamientos fitosanitarios en haciendas y vigilancia fitosanitaria constante (Martínez-Solórzano, et al., 2020).

El diagnóstico de FOC RT4 será mediante PCR en tiempo real de extractos de ADN de tejidos vegetales y aislados fúngicos. Desafortunadamente, se carece de herramientas eficientes y rápidas de detección de la enfermedad. Una vez que el hongo ha incursionado en un área, las medidas de exclusión serán inútiles y en tal caso, las medidas de contención deberán ser preparadas y analizadas. Estas comprenden, el acordonamiento y cuarentena de la zona, semaforización dependiendo del porcentaje de la zona diagnosticada. El control mediante la incineración de plantas infectadas y de plantaciones aledañas. Además, el ajuste de medidas de

bioseguridad para agricultores así como de herramientas y maquinarias de trabajo (Agrocalidad, 2018).

El patógeno tiene naturaleza compleja, al colonizar los rizomas y los vasos del xilema, no puede ser atacado por fungicidas. Además los aspectos epidemiológicos, no pueden ser determinados con eficiencia debido a la capacidad de supervivencia del hongo en el suelo (más de 30 años), diseminación rápida mediante vectores o fómites y se le agrega que el monocultivo de banano Cavendish es más susceptible por poseer variabilidad genética baja (Li C, et al., 2015). FOC RT4 recoloniza rápidamente el área ya que no existe la eliminación total del hongo.

1.4.3 Secuenciación de RNA

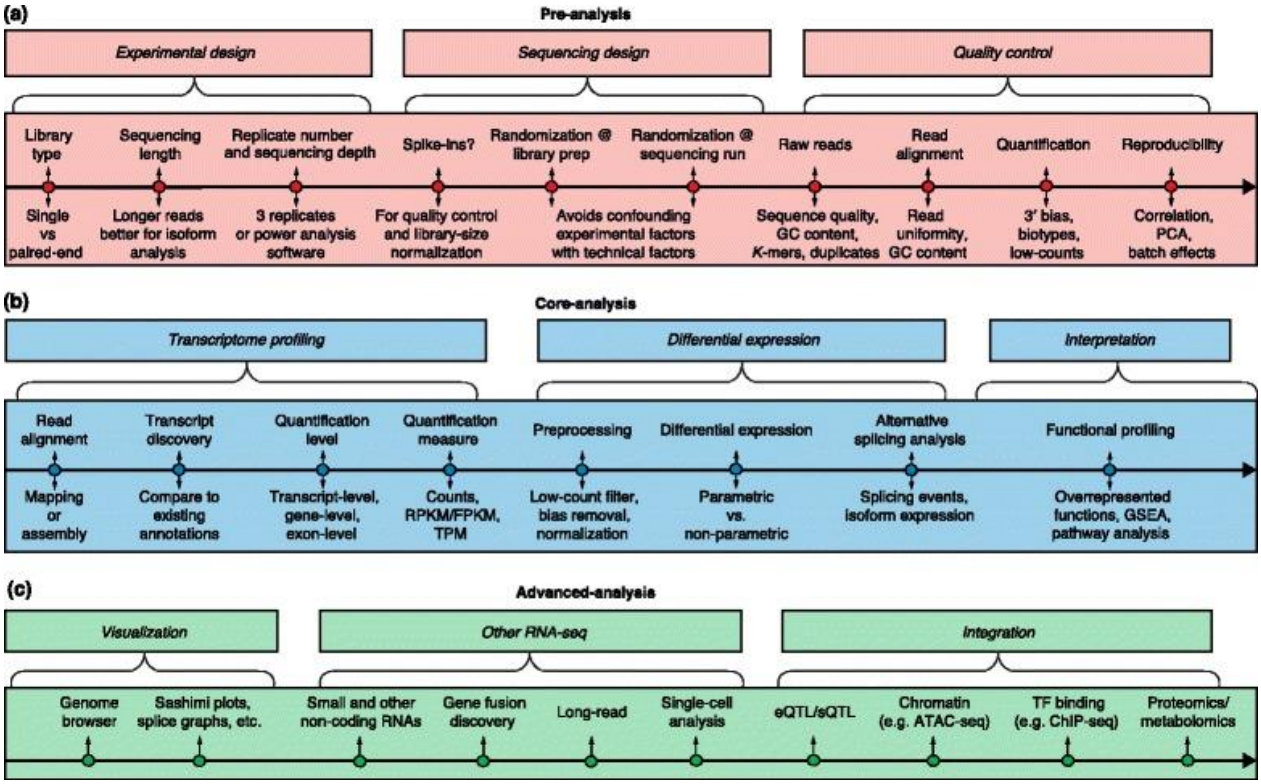
La secuenciación del RNA permite la identificación de transcritos y la cuantificación de la expresión génica, desde que se conoce que el RNA juega un papel importante como intermediario entre el genoma y el proteasoma. La clave de RNA-seq es la secuenciación de alto rendimiento para el descubrimiento de perfiles de expresión génica basado en variaciones de condiciones de tratamiento y así obtener la expresión diferencial de genes (DEG) (Conesa, et al., 2016).

RNA-seq no tiene limitación de uso para diferentes aplicaciones y escenarios de análisis, los científicos planifican estrategias dependiendo del organismo en estudio y los objetivos planteados. Es decir, si el organismo en cuestión ya posee una secuencia de genoma disponible en bases de datos, el estudio consistirá en la comparación y la identificación de las transcripciones mapeando las lecturas obtenidas de la secuenciación al genoma antes mencionado. Si el organismo, no posee genoma de referencia, el flujo de trabajo de RNA-seq

iniciará con el ensamblaje de novo de lecturas en contigs, para luego mapear las transcripciones contra el genoma ensamblado (Tong, et al.,2020).

Figura 1

Flujo de trabajo genérico para los análisis de RNA-seq.



Nota. El gráfico representa los diferentes pipelines que se pueden seguir dependiendo del fin del análisis. *Recuperado de* Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., Szcześniak, M. W., Gaffney, D. J., Elo, L. L., Zhang, X., & Mortazavi, A. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology,* 17, 13. https://doi.org/10.1186/s13059-016-0881-8

1.4.3.1 Minería de datos

El avance de la tecnología de información y comunicación ha permitido el desarrollo de grandes bancos inteligentes de datos que van desde posiciones geográficas hasta genoma de la mayoría de organismos conocidos (Wang, Miwa, & Morikawa , 2020). La minería de datos también conocida como el descubrimiento de conocimiento, es un proceso que recupera información útil de altos volúmenes de datos. Además, es muy utilizado en técnicas fundamentales para el preprocesamiento, agrupación, clasificación, identificación de nuevos valores o de valores atípicos, etc. (Bellinger, et al., 2017).

En FOC RT4 los estudios para comprender los genes envueltos en la infección y colonización del hospedero son limitados, sin embargo el genoma de referencia y archivos de anotaciones de genes se encuentran disponibles. Esto último, sirve de gran apoyo para la investigación del transcriptoma y así conocer los factores claves que permiten al hongo colonizar la planta.

1.5 Hipótesis

Existen genes involucrados en la interacción entre *Fusarium oxysporum* f. sp*. cubense* RT4 y *Musa acuminata* subgrupo Cavendish que no han sido reportados.

Capítulo II: MATERIALES Y MÉTODOS

2.1 Participantes

La presente investigación ha sido elaborada por Stefanía Soledad Montesinos Ludeña egresada de la Carrera Ingeniería en Biotecnología, bajo la tutoría del Ing. Francisco Flores Ph.D. docente investigador del departamento de Ciencias de la Vida y la Agricultura en la Universidad de las Fuerzas Armadas ESPE. El financiamiento de la investigación estuvo a cargo del estudiante.

2.2 Zona de estudio

El presente proyecto de investigación se realizó en la ciudad de Loja, mediante el uso de programas bioinformáticos disponibles en línea.

2.3 Duración de la investigación

El tiempo de duración de esta investigación fue de aproximadamente 5 meses. Se inició en el mes de septiembre de 2020 y culminó en enero del 2021 .

2.4 Metodología

El presente estudio se realizó en tres fases: La primera fase consistió en la obtención de la información de secuenciaciones de RNA previas, mediante la minería de datos. En la segunda fase se estandarizó el pipeline del workflow a seguir para el análisis diferencial y en la tercera fase se anotaron los genes expresados diferencialmente haciendo comparaciones con genes previamente reportados.

2.4.1 Elección de la plataforma de estudio

Se escogió la plataforma Galaxy (https://usegalaxy.org) que posee su servidor en alianza entre la Universidad de Texas y Cyverse®. Su interfaz amigable con el usuario y la simplicidad al omitir la codificación de programas, permitieron la estandarización del pipeline adecuado para el análisis. Los datos de secuenciación se cargaron en la plataforma web Galaxy y se usó el servidor público en usegalaxy.org para analizar los datos (Afgan, et al., 2016).

2.4.2 Obtención de información mediante minería de datos

Se utilizó minería de datos para la recopilación de datos actualizados del análisis transcriptómico de la interacción entre banano Cavendish y FOC RT4. La información fue recolectada de bases de datos como Ensembl Fungi, NCBI, Mycocosms, KEGG, Google scholar y Elsevier. Se tomaron artículos publicados actuales con un rango de fecha de 2014 en adelante. Para la obtención del genoma de referencia de FOCRT4, se compararon tres bases de datos Ensembl Fingi, NCBI y Mycocosm, se eligió el genoma que tenía mayor cantidad de pares de bases, menos scaffolds y más similitud al momento de realizar un blasteo con genoma de *Fusarium oxysporum* f.sp. *cubense raza tropical 1*. Se realizó el análisis con datos de tres estudios diferentes de raíces de banano infectadas con FOCRT4, para así comparar la fiabilidad de los nuevos genes anotados.

Para el primer análisis, se obtuvieron muestras del tratamiento para tres distintos tiempos de inoculación, a 0DPI, 2DPI y 4DPI (Guo, et al., 2014). Los controles fueron secuencias de raíces de banano sin inocular, para así eliminar las secuencias de RNA-seq del hospedero, como recomienda Wang, Z y colaboradores (2012).

Para el segundo análisis, los datos de tratamiento de las raíces de banano fueron a las 3, 27 y 51 horas después de la inoculación. Los datos de control se tomaron de la misma referencia, con el mismo tiempo de inoculación (Li C, et al.,2015).

Para el tercer análisis, los datos utilizados correspondían a las evaluaciones a las 0, 24 y 48 HPI. Tanto los datos de tratamiento como los de control fueron obtenidos de la Universidad de Agricultura de China Sur, (2020).

2.4.3 Estandarización del pipeline para RNA-seq

La estandarización del pipeline se estableció mediante consulta bibliográfica, en la que se menciona los pasos a seguir para tener un análisis de expresión diferencial preciso (Qi, et al.,2017).

2.4.3.1 Ajuste del workflow

El workflow a seguir va a contener los siguientes puntos:

1. Análisis de calidad (FastQC)
2. Recorte de los datos (adaptadores) (Trimmomatic)
3. Análisis de calidad luego del recorte (FastQC)
4. Alineación de los transcritos con el genoma de referencia del hongo (RNA STAR)
5. Ensamblaje y cuantificación de los transcritos (StringTie)

6. Unión de transcritos ensamblados (StringTie merge)

7. Comparación de transcripciones ensambladas con una anotación de referencia (FeatureCounts)

8. Análisis de expresión diferencial (DESeq2)

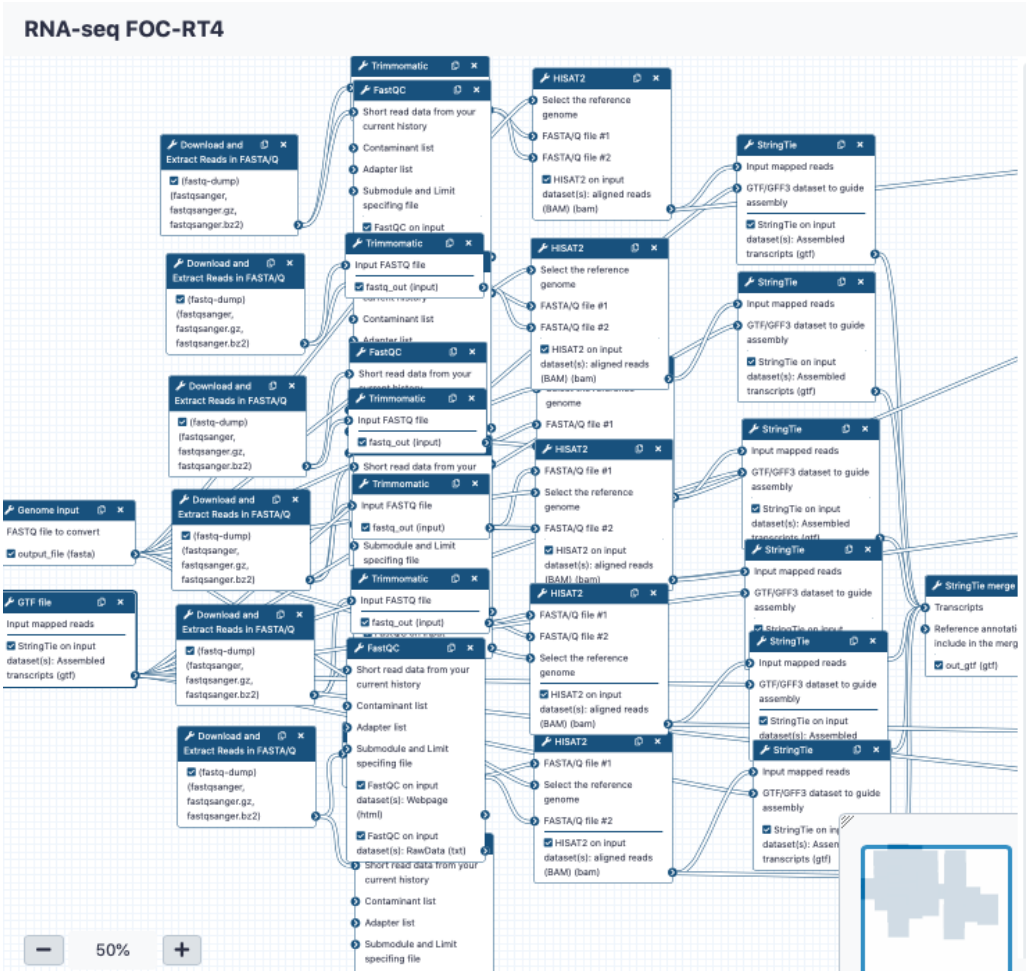
9. Filtración de genes con expresión diferencial significativa (Annotate)

Desarrollo del workflow

Se desarrollo el workflow en la plataforma de Galaxy, se construyó el pipeline con base en los pasos para un buen análisis de expresión diferencial. Se tomó en cuenta que cada output esté conectado correctamente con el siguiente paso. Se tomó en cuenta un cuadro de herramienta por cada dato obtenido; se diseñó el workflow de manera que el genoma de referencia y el archivo gft sean elegidos desde la historia.

Figura 2

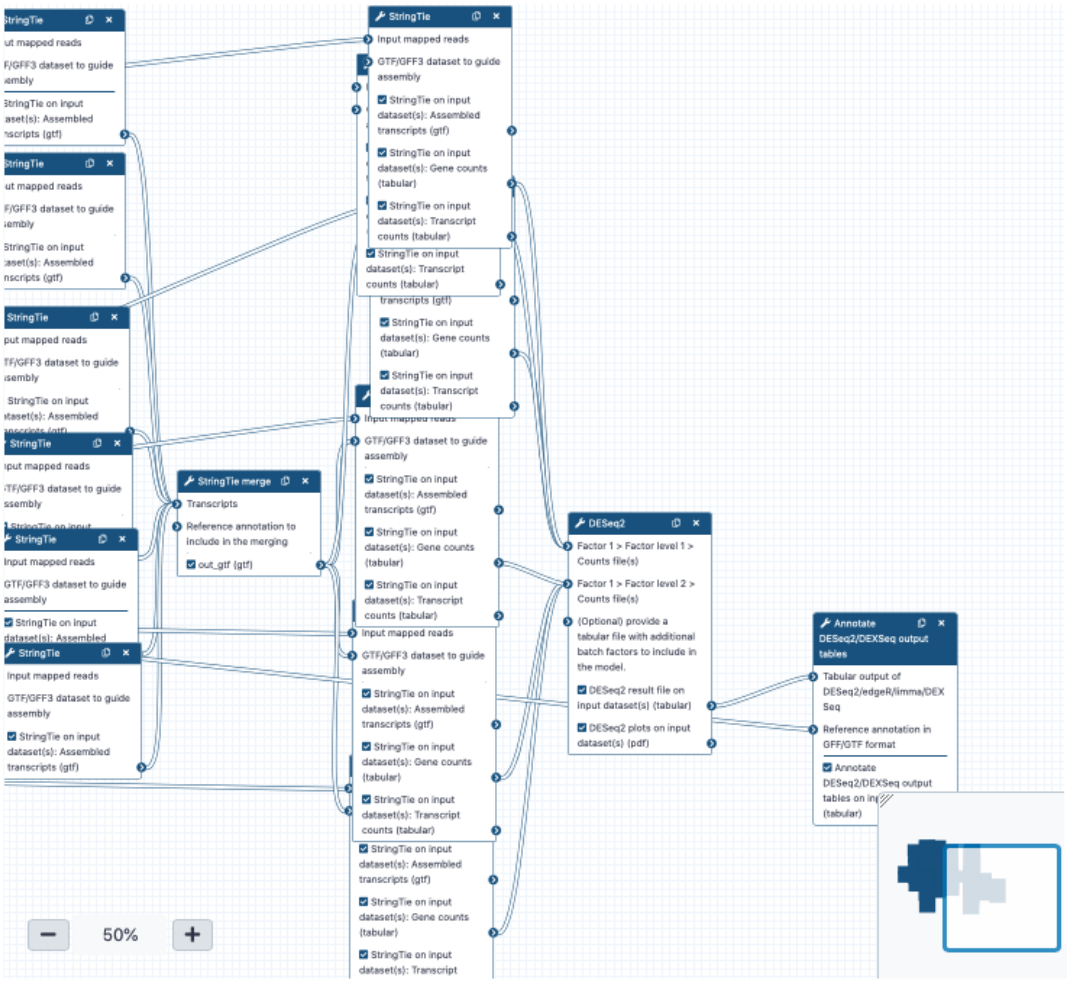
Diagrama del workflow de Galaxy server. Parte de análisis de datos y evaluación de datos.



Nota. Pipeline realizado en el servidor Galaxy, que va desde los pasos de subida de datos, análisis de calidad, recorte de secuencias y alineamiento de secuencias.

Figura 3

Diagrama de workflow de Galaxy server. Parte de análisis diferencial y conteo de genes reportados.



Nota. Pipeline para análisis de RNA-seq utilizado en el ensayo, comprende el análisis diferencial y el conteo de genes para su reporte posterior.

2.4.4.2 Análisis de los datos obtenidos

Para la evaluación de la información obtenida se utilizó FastQC, tiene como objetivo comprobar de forma sencilla la calidad de los datos de secuencia que provengan de pipelines de secuenciación de alto rendimiento, en este caso de RNA-seq. Proporciona un conjunto de análisis fáciles de comprender, con el fin de determinar si los datos tienen algún problema que debe ser arreglado antes de continuar (Andrews, 2010). Se suben todas las muestras en formato fasta y se mantiene un k-mer de búsqueda estándar de 7.

2.4.4.3 Selección de los parámetros a evaluar

Los parámetros se seleccionaron con base en los resultados de FastQC, además al momento de descargar los datos de las fuentes se observó con que tipo de secuenciación habían sido obtenidos y si los datos eran de un solo extremo o emparejado. Los parámetros de cada uno de los datos fueron:

Parámetros de las fuentes:

- Longitud de read
- Secuenciación
- Layout (Simple o Pareado)
- Contiene adaptadores

Parámetros de resultados de FastQC

- Calidad por base de secuencia
- Contenido GC
- Contenido de N

2.4.4.4 Recorte de secuencias

El recorte de datos fue importante para obtener datos crudos limpios, es decir sin adaptadores, contaminantes o secuencias específicas. Al poseer secuencias generadas con Illumina Sanger, se procedió a utilizar Trimmomatic, ya que es específico para este tipo de secuenciación. Esta herramienta también permite eliminar lecturas si están por debajo de la longitud especificada. Además, corta bases al final de una lectura o secuencia si se encuentra por debajo del umbral de calidad (Bolger, et al.,2014). Se discernieron los datos que eran de uno solo extremo y los emparejados, posteriormente se escogieron las muestras de tratamiento y control. Para el caso de muestras de un solo extremo, se utilizaron los datos una sola vez como inputs para Trimmomatic y se mantuvo todo lo demás como predeterminado. Para las muestras emparejadas, se seleccionaron los datos y ser marcó como paired-end. Se mantuvo todo lo demás como predeterminado. Una vez terminado el recorte de los datos, se realizó una nueva ronda de FastQC para comparar las secuencias antes y después del tratamiento con Trimmomatic.

2.4.4.5 Alineamiento con genoma de referencia

Se utilizó RNA STAR, el cual es un programa de alineamiento empalmado rápido y sensible. Además, el programa posee varias estrategias de alineación que están diseñadas para el mapeo de diferentes tipos de lecturas de RNA-seq (Dobin, et al., 2012). El genoma de referencia para FOC RT4 no se encuentra disponible en las bases de datos de Galaxy, por ende se usó el genoma de Ensembl Fungi. Nuevamente, se discriminaron los datos con base en si son de un solo extremo o emparejado. En este programa se añadió un nuevo parámetro de análisis, el sentido de la secuencia; unstranded (sin sentido), forward (adelante), reverse (atrás). Los datos obtenidos fueron todos unstranded.

2.4.4.6 Ensamblaje de transcritos con anotaciones de genes

Para el ensamblaje de transcritos se utilizó StringTie, el cual es un ensamblador rápido y altamente eficiente de alineaciones de RNA-seq. Asimismo, permite realizar ensamblajes de novo y cuantificaciones de transcripciones de longitud completa que representan múltiples variantes de empalme para cada locus génico (Kovaka, et al., 2019). Se utilizaron los outputs de RNA STAR para el experimento. Una vez que se obtuvieron los outputs de StringTie junto con el archivo gtf de referencia obtenido de Ensembl Fungi, se procedió a realizar una ronda de StringTie merge, para ensamblar todas las transcripciones resultantes en un solo conjunto de transcripciones no redundantes. De este modo, se genera un conjunto unificado global de las muestras de RNA-seq del estudio (Pertea, et al., 2015).

2.4.5 Análisis de expresión diferencial

2.4.5.1 Determinación de los genes con expresión diferencial a partir de tablas de anotaciones

Se realizó la comparación de transcripciones ensambladas con una anotación de referencia mediante FeatureCounts. Esta herramienta compara y evalúa la precisión de los ensambladores de transcripciones de RNa-Seq, como es el caso de StringTie. Para ello se utilizaron los outputs tipo gtf resultantes de StringTie y el archivo gtf de referencia (Trapnell, et al., 2010). Para la medición de la expresión génica, de igual manera se utilizó FeatureCounts, la cual sirve conteo de lecturas de RNA-seq con características genómicas y así medir la expresión génica (Liao , et al. , 2013).

2.4.5.2 Cuantificación de la expresión génica

El siguiente paso del pipeline fue realizar el análisis con DESeq2. Este es un programa que estima la dependencia media de la varianza en los datos de ensayos de secuenciación de alto rendimiento. Asimismo, prueba la expresión diferencial basada en un modelo de distribución binomial negativa (Love, et al., 2014). Los outputs de featureCounts (outputs tipo DESeq2) se utilizaron para este programa. Se clasificó en dos grupos, tratamiento y control; y se seleccionó la tabla normalizada de conteo y la tabla normalizada de rLog como outputs adicionales.

2.4.5.3 Anotación de genes sin reportar

La visualización de los genes diferencialmente expresados se utilizó la herramienta Annotate, en donde se observó los genes mapeados con el genoma de referencia del hongo.

2.5 Análisis estadístico

Los resultados de los ensayos se analizaron con DESeq2, se obtuvo la media, error estándar y el log2 de las anotaciones de los genes.

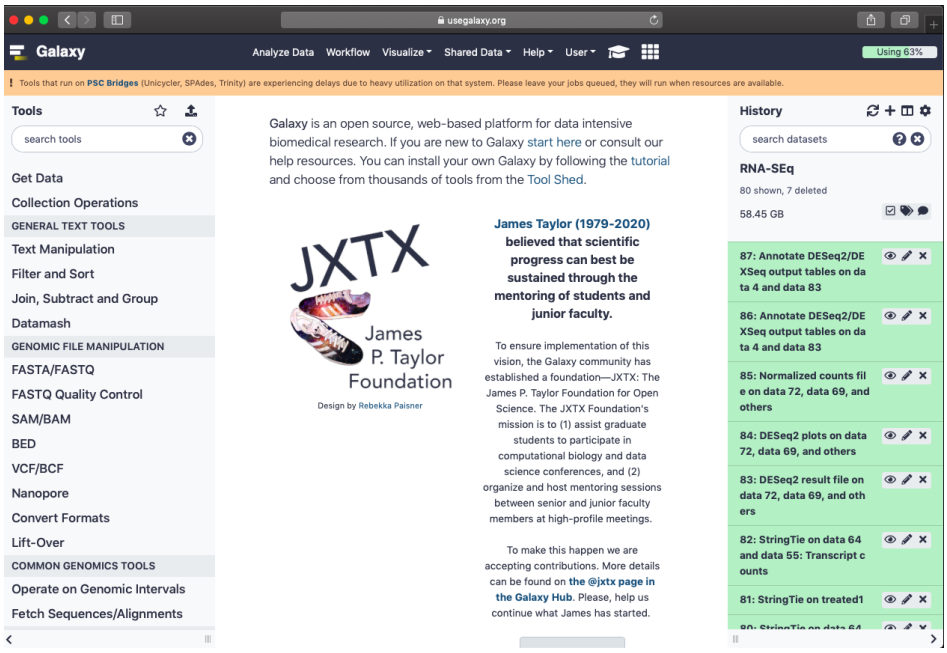
Capítulo III: RESULTADOS

3.1 Plataforma utilizada

La plataforma Galaxy server permitió la ejecución del proyecto. Para el análisis se crearon tres bases de datos o historias. El primer ensayo se denominó RNA-seq1, el segundo RNA-seq2 y el tercer ensayo RNA-seq3. En la parte inferior derecha, se pueden visualizar los archivos y outputs creados por cada uno de los trabajos en el pipeline.

Figura 4

Interfaz de Galaxy server.



Nota. Interfaz de la plataforma utilizada, en donde se observa cada uno de los procesos realizados. Tomado de https://usegalaxy.org

3.2 Extracción de datos mediante minería de datos

Tabla 1

Primer ensayo RNA-seq1 se utilizaron datos con los siguientes números de accesión de NCBI-SRA.

Ensayo RNA-seq1: Infección de las raíces de banano a los 0, 2 y 4 DPI	
Control	**Tratamiento**
• **Control 1 SRR8872344**	• Tratamiento 1 SRR554395
• **Control 2 SRR8872934**	• Tratamiento 2 SRR554396
• **Control 3 SRR10137402**	• Tratamiento 3 SRR495216

Tabla 2

Ensayo RNA-seq2 se utilizaron datos con los siguientes números de accesión de NCBI-SRA.

Ensayo RNA-seq2: Infección de las raíces de banano a los 0, 24 y 48 HPI	
Control	**Tratamiento**
• **Control 1 SRR12400582**	• Tratamiento 1 SRR372769
• **Control 2 SRR12400581**	• Tratamiento 2 SRR3727687
• **Control 3 SRR12400580**	• Tratamiento 3 SRR3727684

Tabla 3

Ensayo RNA-seq3 se utilizaron datos con los siguientes números de accesión de NCBI-SRA.

Ensayo RNA-seq3: Infección de las raíces de banano a los 0, 24 y 48 HPI	
Control	**Tratamiento**
• **Control1 SRR12400582**	• Tratamiento 1 SRR10424537
• **Control2 SRR12400581**	• Tratamiento 2 SRR10424536
• **Control3 SRR12400580**	• Tratamiento 3 SRR10424535

3.1.2 Análisis de calidad

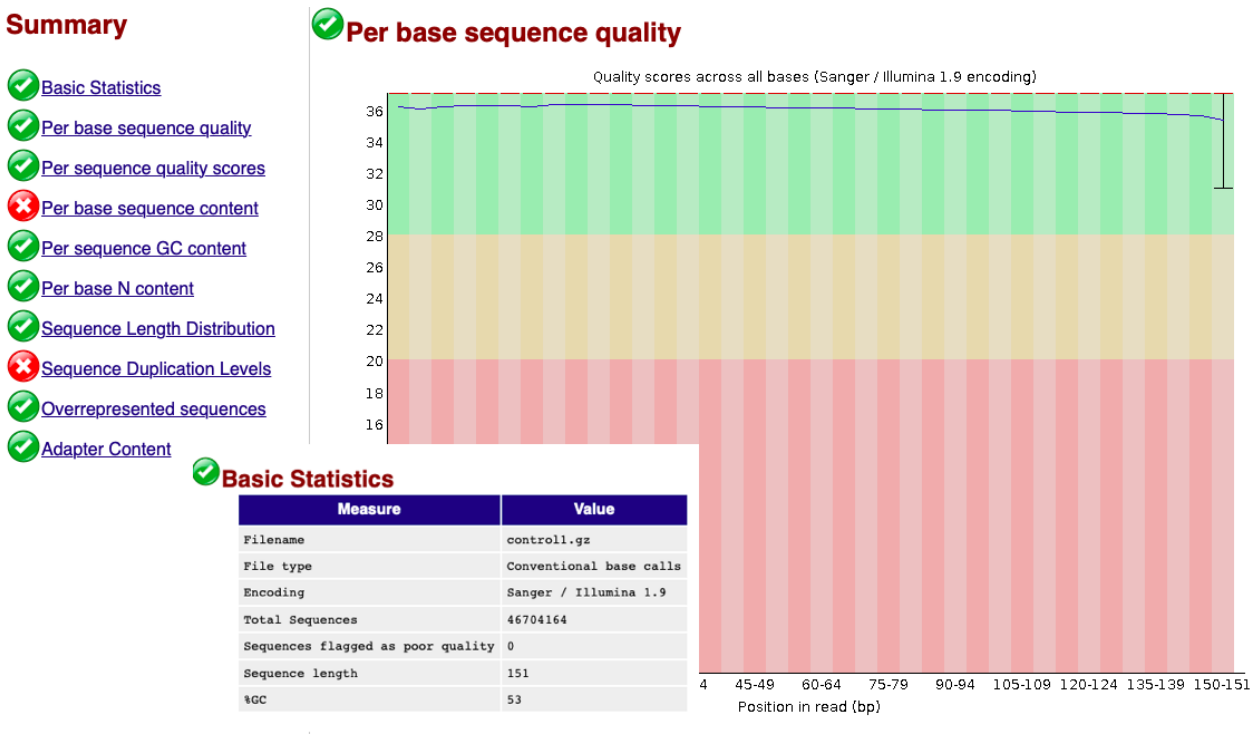
Al momento de realizar el análisis de calidad con FastQC, se obtuvo los resultados de estadísticas básicas, calidad de secuencia por base, calidad por score de secuencia, contenido de secuencia por base y contenido GC.

Se evaluaron los 6 datos del primer ensayo RNA-seq 1 (controles y tratamiento) y se obtuvieron los siguientes resultados:

Los datos de control 1, según los resultados arrojados por FastQC, fueron secuenciados mediante Sanger/Illumina con un contenido de GC de 53%, óptimo para el análisis de expresión diferencial. Además la calidad de secuencia por base es óptima según se muestra en la Figura 5.

Figura 5

Representación gráfica de los resultados de FastQC de muestra control1.

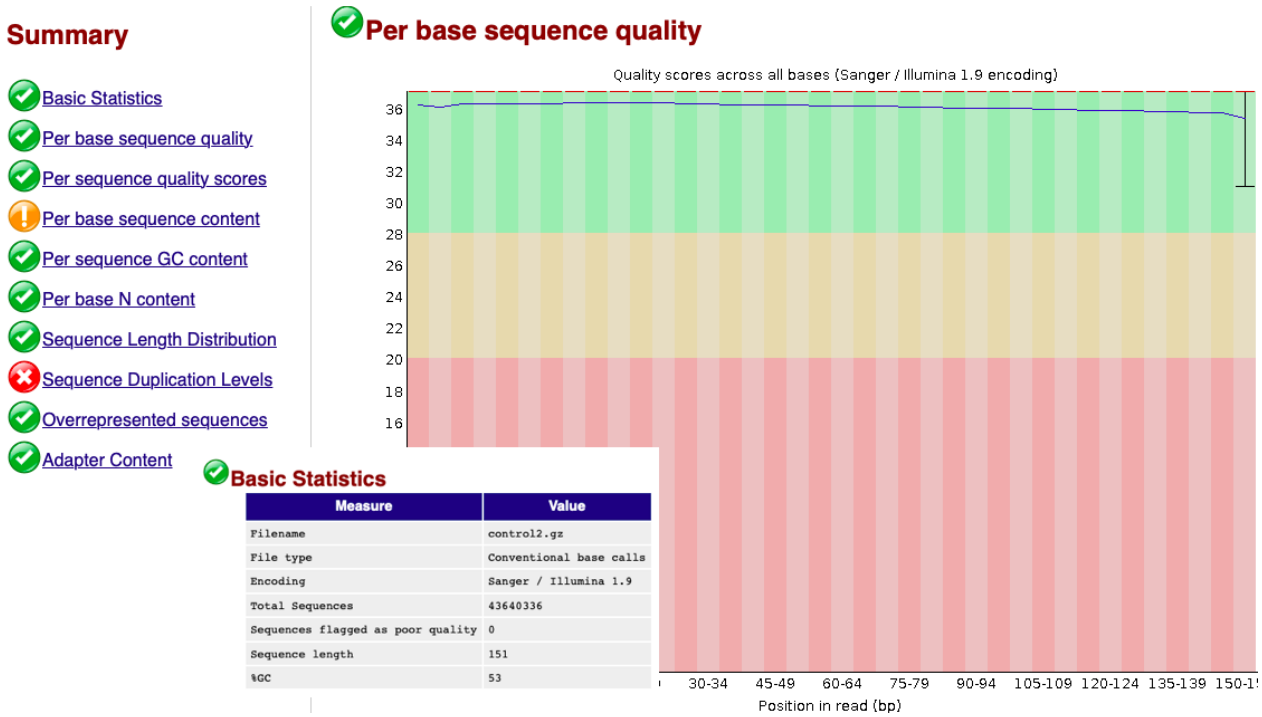


Nota. *Se puede observar que los reads tienen buena calidad por secuencia, además de contenido de GC aceptable y alta calidad por score de secuencia.*

Los resultados del control 2 muestran que fue generado con Sanger/Illumina y que posee un porcentaje de GC de 53%. Los datos se encuentran óptimos para realizar el análisis, como se muestra en la Figura 6.

Figura 6

Representación gráfica de los resultados de calidad de FastQC de datos de control2.

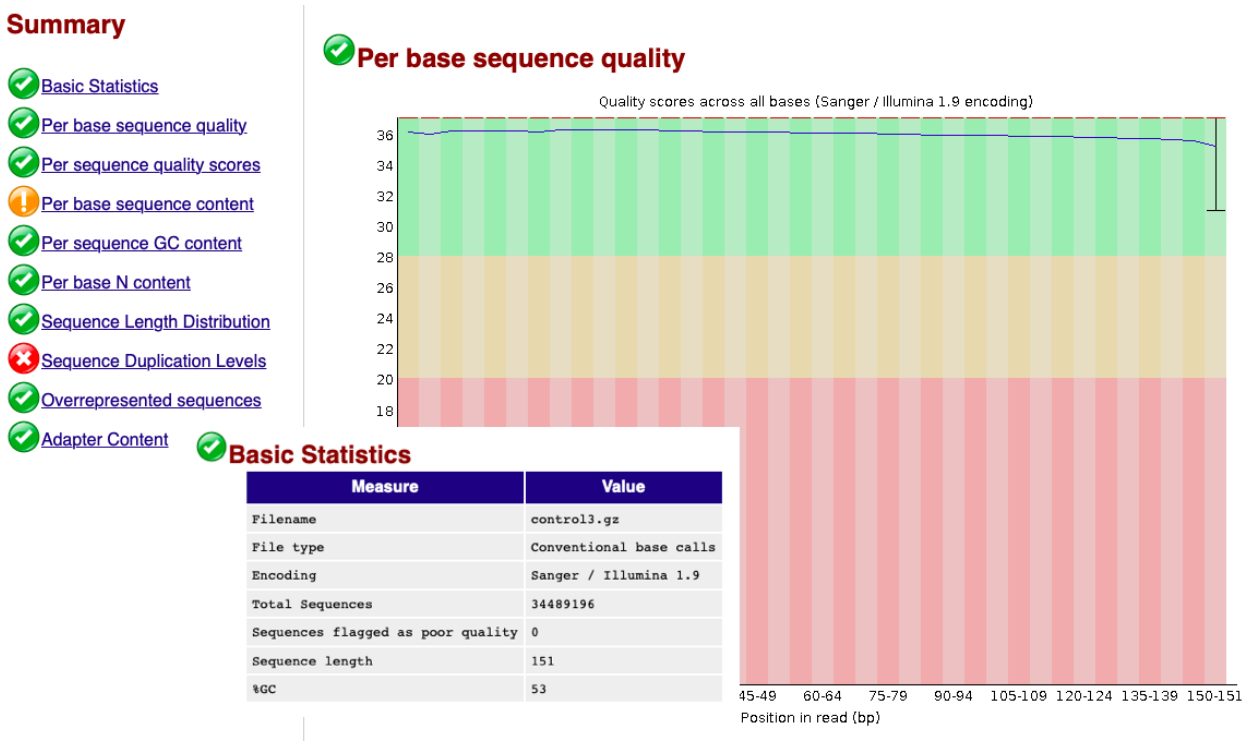


Nota. *Se puede observar que existe buena calidad por base de secuencia, buen contenido de GC y de N, y no existe contaminación por adaptadores.*

El control 3 muestra similar naturaleza que los otros dos controles, con un contenido de GC de 53% óptimo, buen contenido de Nitrógeno y calidad de secuencia por base, como se muestra en la Figura 7.

Figura 7

Resultados del análisis de la calidad de la secuencia del control 3.

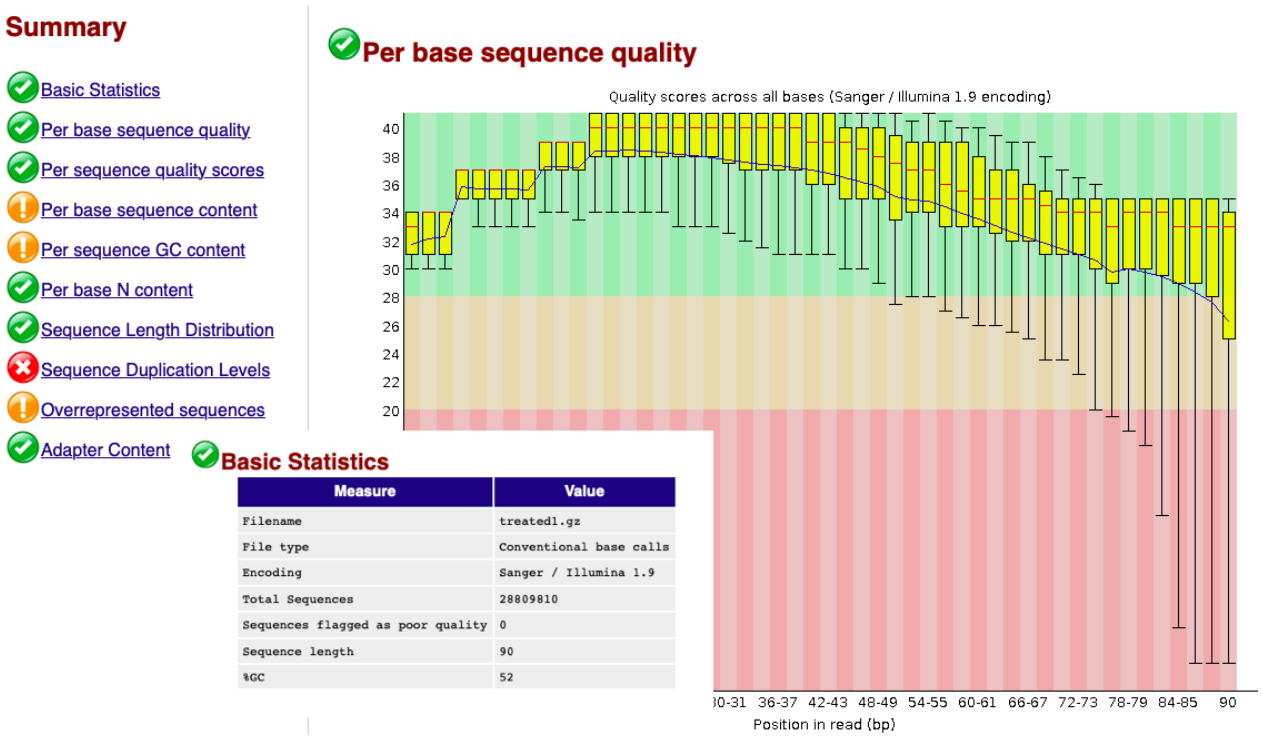


Nota. *Se observa que los datos de control3 poseen buena calidad de secuencia por base y contenido de Nitrógeno por base.*

El tratamiento 1 muestra mayor inconsistencia en su calidad, tiene un contenido de GC óptimo de 52%, sin embargo esto se puede ver afectado por la baja calidad de secuencia por base, como se muestra en la Figura 8. Debido a esto se procedió a realizar el paso de recorte de datos para estabilizar la secuencia.

Figura 8

Resultados del análisis de calidad de los datos de tratamiento 1.

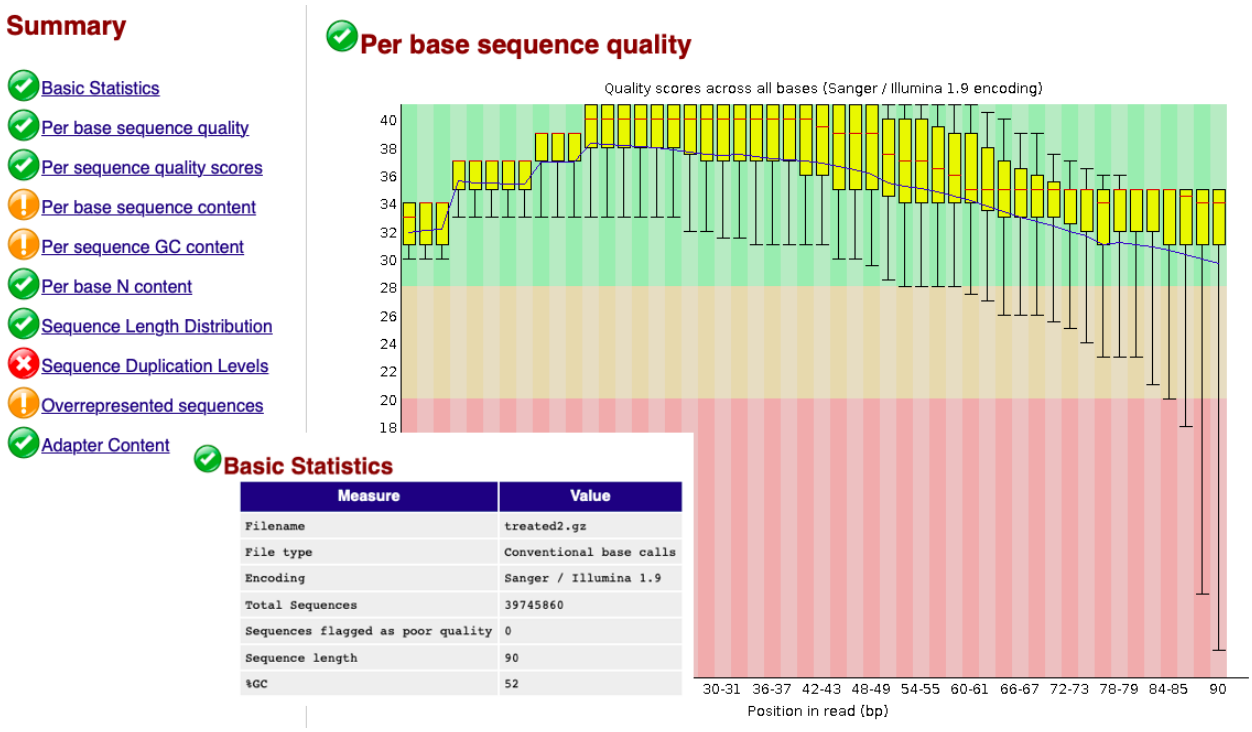


Nota. *Se puede observar que existe lecturas en mal estado, con reads de 90. No obstante, posee buen contenido de Nitrógeno por base y bajo contenido de adaptadores.*

El tratamiento 2 tuvo un contenido de GC de 52%, pero al igual que el tratamiento 1, este se ve influenciado por la baja calidad de secuencia como se puede observar en la Figura 9.

Figura 9

Resultados del análisis de calidad en la muestra de tratamiento 2.

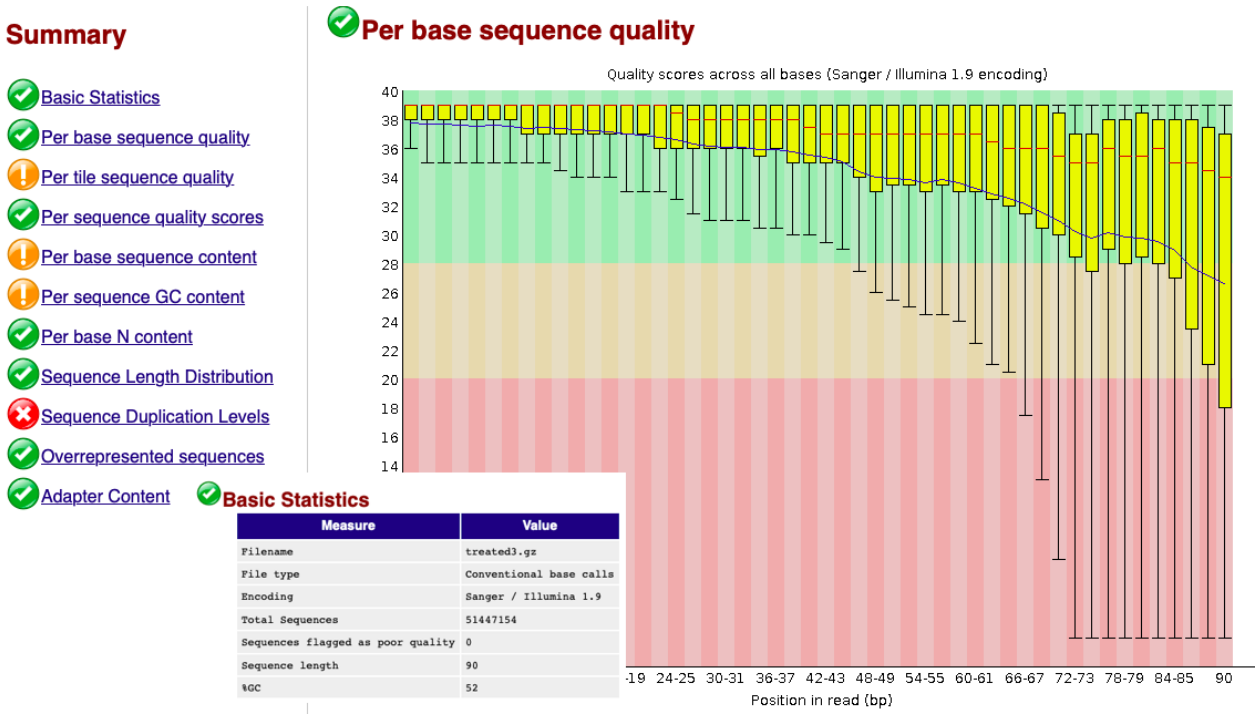


Nota. *Se puede observar que existen secuencias de baja calidad, lo que afectaría en el análisis de expresión diferencial.*

En el caso del tratamiento 3, los resultados arrojaron que tiene un contenido de GC del 52%, con secuencias de baja calidad. El contenido de Nitrógeno es bueno al igual que contenido de adaptadores como se evidencia en la Figura 10.

Figura 10

Resultados del análisis de calidad de secuencias del tratamiento 3.



Nota. *Se puede observar que existe contenido de GC óptimo. Además, se observa que existen niveles altos de duplicación de secuencia.*

Los resultados del análisis del ensayo RNA-seq 1 demuestran que todos los datos de muestra se pueden utilizar para el análisis de expresión diferencial. Sin embargo, se debe proceder a realizar el siguiente paso, para mayor eficiencia en el mapeo de secuencias.

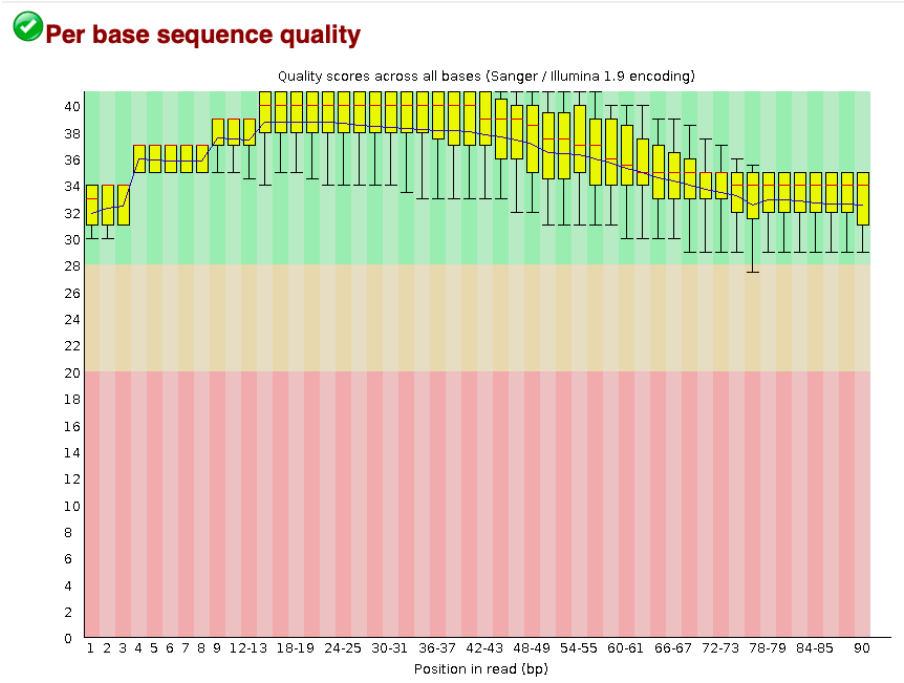
Los datos de los ensayos RNA-seq 2 y RNA-seq 3 mostraron comportamiento parecido a los resultados obtenidos de FastQC. Por ello, se procedió a realizar el recorte de secuencias.

3.1.3 Recorte de datos

El recorte de datos se realizó para cada uno de los datos tanto de control como de tratamiento para los tres análisis. Utilizando Trimmomatic se logró obtener secuencias sin adaptadores y con longitudes de reads de 90. Como se puede ver en la Figura 11, se puede ver que como las secuencias de baja calidad fueron cortadas y únicamente quedan de buena calidad, comparado con la Figura 8.

Figura 11

Representación gráfica de la calidad de secuencias resultantes del recorte de datos con Trimmomatic en el tratamiento 1 forward del análisis RNA-seq 1.



Nota. *Se observó que existen secuencias de longitud de 90, con buena calidad. Se hace la distinción ya que se usó secuencias emparejadas.*

Los resultados del análisis de calidad de secuencias después de aplicación de recorte de datos mediante Trimmomatic para las muestras de los ensayos RNA-seq 2 y RNA-seq 3, se encuentran en los anexos 1 y 2 respectivamente.

3.1.4 Alineamiento de los transcritos con el genoma de referencia

El alineamiento se realizó mediante el uso de RNA STAR, y se obtuvo una tabla por cada una de las muestras. En ellas se encuentra representado el número de alineamientos por secuencia de contigs, además si existe la presencia de scaffolds y se puede ver representado en la Figura 12.

Figura 12

Representación de tabla de datos con la agrupación de los mapeos con el genoma de FOC RT4 con datos de tratamiento 1 del análisis RNA-seq1.

@SQ SN:KK036454 LN:1014

QNAME	FLAG	RNAME	POS	MAPQ			MRNM	MPOS	SEQ
5796354	419	JH658272	717720	1	66M	=	928329	210675	ATAACGTCCGTCACCCAAAGACTGGCCACCCCCACAGACCGGCCAC
5797409	99	JH658272	717720	1	76M	=	928319	210675	ATAACGTCCGTCACCCAAAGACTGGCCACCCCCACAGACCGGCCAC
5797409	419	JH658272	717720	1	76M	=	928319	210675	ATAACGTCCGTCACCCAAAGACTGGCCACCCCCACAGACCGGCCAC
6166637	99	JH658272	717720	1	95M	=	928300	210675	ATAACGTCCGTCACCCAAAGACTGGCCACCCCCACAGACCGGCCAC
6166637	419	JH658272	717720	1	95M	=	928300	210675	ATAACGTCCGTCACCCAAAGACTGGCCACCCCCACAGACCGGCCAC
10596385	99	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
10596385	419	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
10595829	99	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
10595829	419	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
20206672	99	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
20206672	419	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
20207844	99	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
20207844	419	JH658272	717720	1	16S134M	=	928261	210675	CTCTTCGTAACTTCTTATAACGTCCGTCACCCAAAGACTGGCCACCC
1205365	99	JH658272	717721	1	15S135M	=	928259	210673	GCACTAAACATGCATTAACGTCCGTCACCCAAAGACTGGCCACCCCC
1205365	419	JH658272	717721	1	15S135M	=	928259	210673	GCACTAAACATGCATTAACGTCCGTCACCCAAAGACTGGCCACCCCC
2326259	99	JH658272	717721	1	15S135M	=	928259	210673	GCACTAAACATGCATTAACGTCCGTCACCCAAAGACTGGCCACCCCC
2326259	419	JH658272	717721	1	15S135M	=	928259	210673	GCACTAAACATGCATTAACGTCCGTCACCCAAAGACTGGCCACCCCC
2636213	99	JH658272	717721	1	15S135M	=	928259	210673	GCACTAAACATGCATTAACGTCCGTCACCCAAAGACTGGCCACCCCC
2636213	419	JH658272	717721	1	15S135M	=	928259	210673	GCACTAAACATGCATTAACGTCCGTCACCCAAAGACTGGCCACCCCC
5956571	99	JH658272	717721	1	13S137M	=	928257	210673	GTAGTAGTAGTAGTAACGTCCGTCACCCAAAGACTGGCCACCCCCAC
5956571	419	JH658272	717721	1	13S137M	=	928257	210673	GTAGTAGTAGTAGTAACGTCCGTCACCCAAAGACTGGCCACCCCCAC
16884965	163	JH658272	717721	1	13S137M	=	928257	210673	GGAGTAGTAGTAGTAACGTCCGTCACCCAAAGACTGGCCACCCCCAC
16884965	355	JH658272	717721	1	13S137M	=	928257	210673	GGAGTAGTAGTAGTAACGTCCGTCACCCAAAGACTGGCCACCCCCAC
17561060	99	JH658272	717721	0	15S135M	=	928259	210673	GCACTAAACATGCATTAACGTCCGTCACCCAAAGACTGGTCACCCCC

Nota. *Las filas representan cada uno de los parámetros que se consideraron: 1. Nombre contig, 2. Marcaje, 3. Nombre del read, 4. Posición, 5. Mapeo de secuencia, 6. Media, 7. Media posición, 8. Secuencia.*

Las columnas, de la tabla obtenida por RNA-STAR, más importantes son el nombre del read, posición, mapeo de secuencia y la secuencia, ya que esos servirán de base para el reconocimiento de la expresión diferencial en el siguiente paso.

3.1.5 Ensamblaje y cuantificación de transcritos

Se aplicó FeatureCounts en los outputs (mapped.bam) obtenidos de RNA STAR y de igual manera se obtuvo una tabla para cada una de las muestras. En la tabla se detalla el número e

identificación del gen (geneID) y el número de transcripts que fueron mapeados y ensamblados, se muestra en la Figura 13.

Figura 13

Ensamblaje y mapeo de transcritos mediante RNA STAR.

Geneid	RNA STAR on data 2	data 1	and others: mapped.bam
Geneid	RNA STAR on data 2, data 1, and others: mapped.bam		
FOIG_00001	2		
FOIG_00002	2		
FOIG_00003	0		
FOIG_00004	0		
FOIG_00005	8		
FOIG_00006	2		
FOIG_00007	0		
FOIG_00008	0		
FOIG_00009	1		
FOIG_00010	30		
FOIG_00011	296		
FOIG_00012	4		
FOIG_00013	12		
FOIG_00014	3		
FOIG_00015	2		
FOIG_00016	24		
FOIG_00017	18		
FOIG_00018	64		
FOIG_00019	0		
FOIG_00020	4		
FOIG_00021	1056		
FOIG_00022	1892		
FOIG_00023	1473		
FOIG_00024	486		
FOIG_00025	108		
FOIG_00026	0		
FOIG_00027	0		
FOIG_00028	24		
FOIG_00029	164		
FOIG_00030	8		

Nota. *Se observa la cantidad de transcritos que fueron mapeados al genoma y a cada uno de los genes que se encuentran anotados dentro de los archivos gft.*

Los resultados del análisis de secuencias mediante featureCounts para las muestras de los ensayos RNA-seq 2 y RNA-seq 3, se encuentran en los anexos 3 y 4 respectivamente.

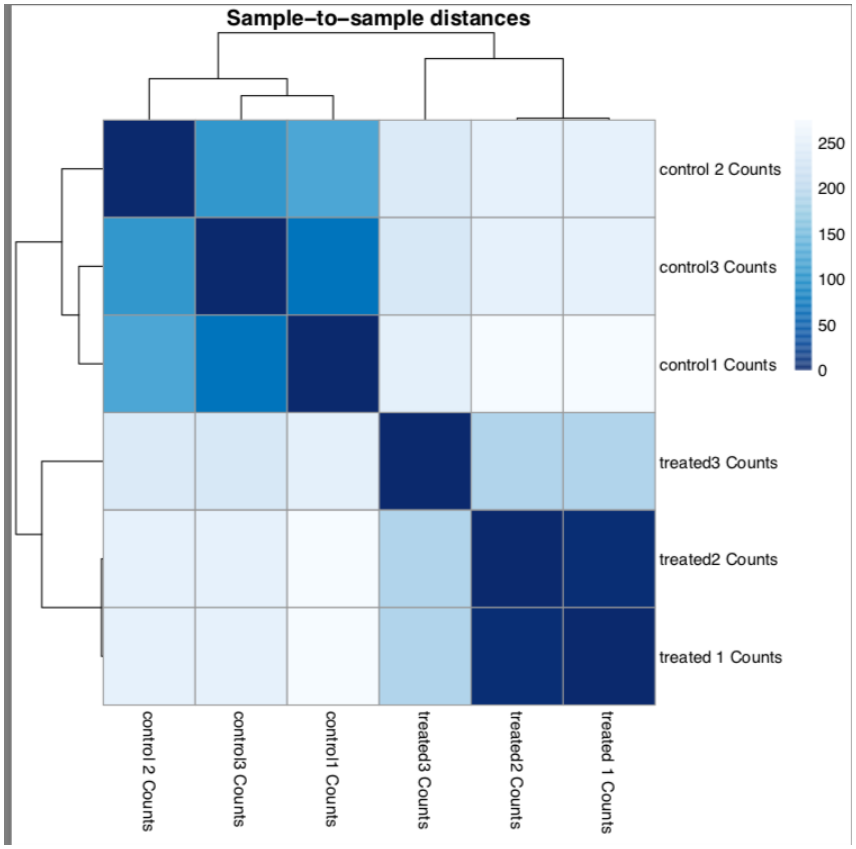
3.2 Análisis de expresión diferencial

3.2.1 Análisis de expresión diferencial a partir de tablas de anotaciones

Se realizó el análisis de expresión diferencial para los tres ensayos, mediante la utilización de DESeq2 y se generaron las siguientes gráficas. Las Figuras 14, 15 y 16 representan los resultados del ensayo RNA-seq1.

Figura 14

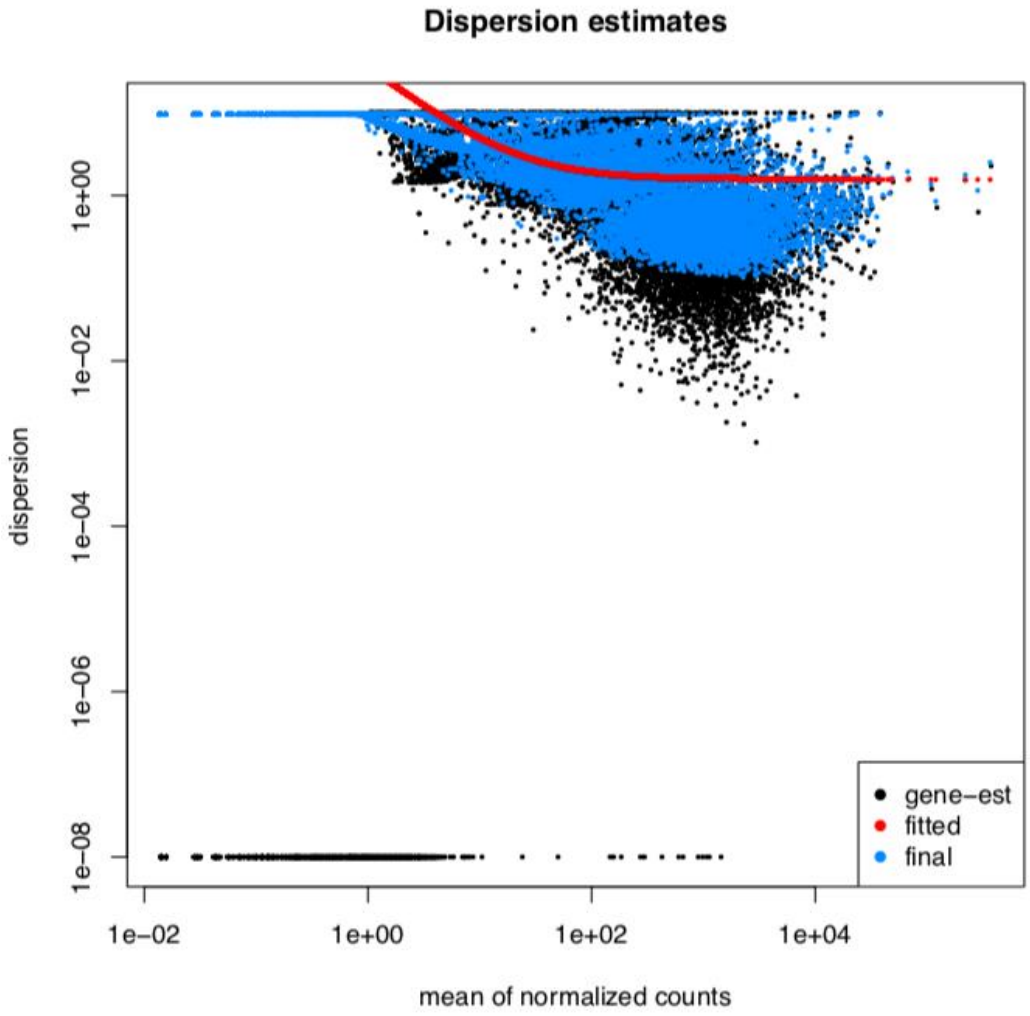
Heatmap de distancias entre muestras del ensayo 1.



Nota. *Se agruparon por tratamiento y control, se puede observar la distancia que existe entre los datos inoculados en raíces de banano e inoculados en medio de cultivo.*

Figura 15

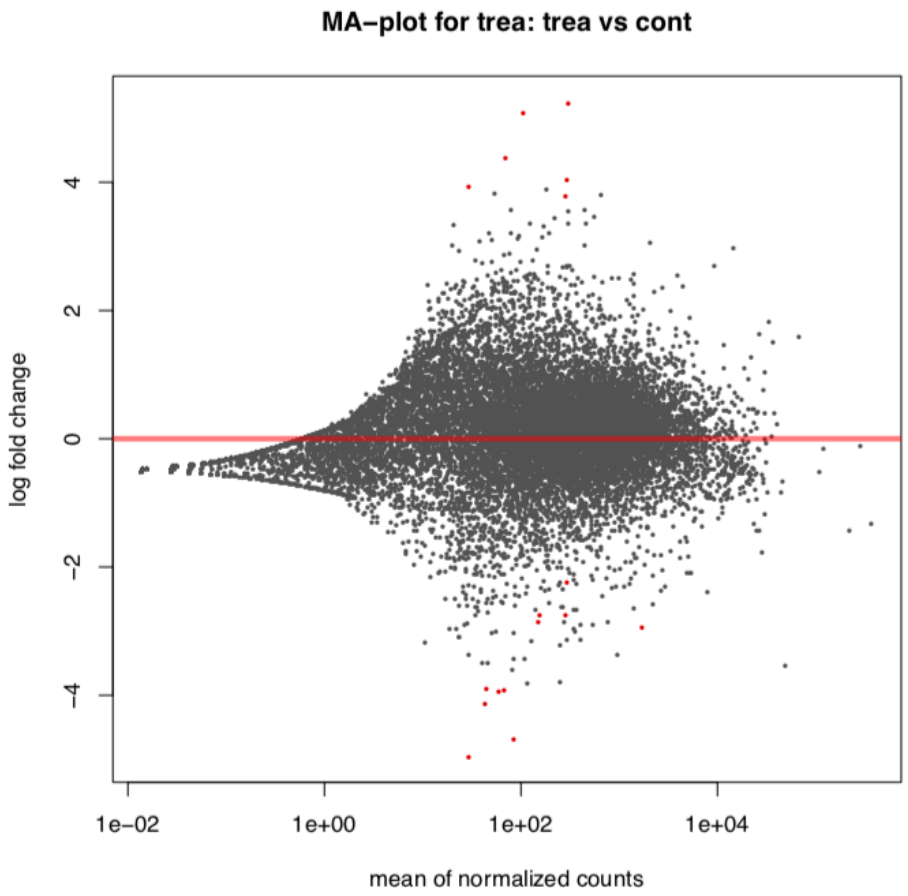
Diagrama de dispersión de las muestras de los tres ensayos de RNA-seq.



Nota. *Representa la estimación ajustada de la cantidad de genes que fueron expresados diferencialmente (rojo) en comparación con los genes que fueron expresados (azul). Los puntos en color negro representan la cantidad de genes que fueron encontrados pero no necesariamente diferencialmente expresados.*

Figura 16

Diagrama MA que representa tratamiento frente a control.



Nota. *Los puntos de las muestras con valores extremos a lo largo del eje, representan los genes que tienen niveles de expresión altamente diferenciales. Además al normalizar los datos se observa que existe linearidad en los datos.*

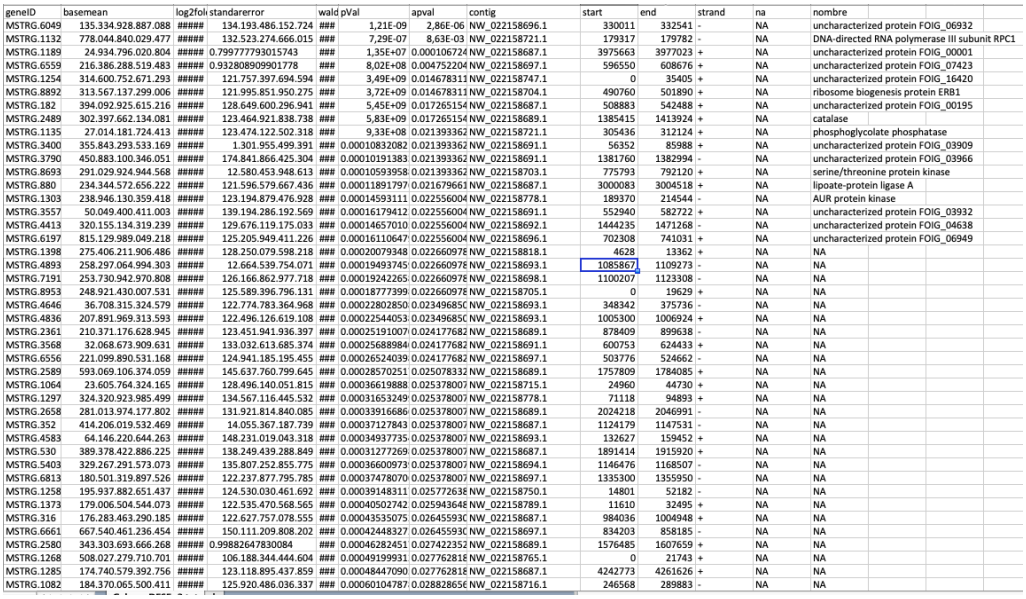
3.2.2 Anotación de todos los genes encontrados del patógeno

Se obtuvieron los siguientes genes que fueron expresados diferencialmente, se tomaron en cuenta aquellos con p-value menor a 0.05. Los genes que se obtienen con nombre N/A son

aquellos que no han sido reportados previamente y que se debe comparar con bases de datos existentes y así observar si hay similitudes.

Figura 17

Reporte de genes expresados diferencialmente encontrados en los tres ensayos de expresión diferencial.



geneID	basemean	log2fol	standarerror	wald	pVal	apval	contig	start	end	strand	na	nombre
MSTRG.6049	135.334.928.887.088	#####	134.193.486.152.724	###	1,21E-09	2,86E-06	NW_022158696.1	330011	332541	-	NA	uncharacterized protein FOIG_06932
MSTRG.1132	778.044.840.029.477	#####	132.523.274.666.015	###	7,29E-07	8,63E-03	NW_022158721.1	179317	179782	-	NA	DNA-directed RNA polymerase III subunit RPC1
MSTRG.1189	24.934.796.020.804	#####	0.799777793015743	###	1,35E+07	0.000106724	NW_022158687.1	3975663	3977023	+	NA	uncharacterized protein FOIG_00001
MSTRG.6559	216.386.288.519.483	#####	0.932808909901778	###	8,02E+08	0.004752204	NW_022158697.1	596550	608676	+	NA	uncharacterized protein FOIG_07423
MSTRG.1254	314.600.752.671.293	#####	121.757.397.694.594	###	3,49E+09	0.014678311	NW_022158747.1	0	35405	+	NA	uncharacterized protein FOIG_16420
MSTRG.8892	313.567.137.299.006	#####	121.995.851.950.275	###	3,72E+09	0.014678311	NW_022158704.1	490760	501890	+	NA	ribosome biogenesis protein ERB1
MSTRG.182	394.092.925.615.216	#####	128.649.600.296.941	###	5,45E+09	0.017265154	NW_022158687.1	508883	542488	+	NA	uncharacterized protein FOIG_00195
MSTRG.2489	302.397.662.134.081	#####	123.464.921.838.738	###	5,83E+09	0.017265154	NW_022158689.1	1385415	1413924	+	NA	catalase
MSTRG.1135	27.014.181.724.413	#####	123.474.122.502.318	###	9,33E+08	0.021393362	NW_022158721.1	305436	312124	+	NA	phosphoglycolate phosphatase
MSTRG.3400	355.843.293.533.169	#####	1.301.955.499.391	###	0.00010832082	0.021393362	NW_022158691.1	56352	85988	+	NA	uncharacterized protein FOIG_03909
MSTRG.3790	450.883.100.346.051	#####	174.841.866.425.304	###	0.00010191383	0.021393362	NW_022158691.1	1381760	1382994	-	NA	uncharacterized protein FOIG_03966
MSTRG.8693	291.029.924.944.568	#####	12.580.453.948.613	###	0.00010593958	0.021393362	NW_022158703.1	775793	792120	+	NA	serine/threonine protein kinase
MSTRG.880	234.344.572.656.222	#####	121.596.579.667.436	###	0.00011891797	0.021679661	NW_022158687.1	3000083	3004518	+	NA	lipoate-protein ligase A
MSTRG.1303	238.946.130.359.418	#####	123.194.879.476.928	###	0.00014593111	0.022556004	NW_022158778.1	189370	214544	-	NA	AUR protein kinase
MSTRG.3557	50.049.400.411.003	#####	139.194.286.192.569	###	0.00016179412	0.022556004	NW_022158691.1	552940	582722	+	NA	uncharacterized protein FOIG_03932
MSTRG.4413	320.155.134.319.239	#####	129.676.119.175.033	###	0.00014657010	0.022556004	NW_022158692.1	1444235	1471268	-	NA	uncharacterized protein FOIG_04638
MSTRG.6197	815.129.989.049.218	#####	125.205.949.411.226	###	0.00016110647	0.022556004	NW_022158696.1	702308	741031	+	NA	uncharacterized protein FOIG_06949
MSTRG.1398	275.406.211.906.486	#####	128.250.079.598.218	###	0.00020079348	0.022660978	NW_022158818.1	4628	13362	+	NA	NA
MSTRG.4893	258.297.064.994.303	#####	12.664.539.754.071	###	0.00019493745	0.022660978	NW_022158693.1	1085867	1109273	-	NA	NA
MSTRG.7191	253.730.942.970.808	#####	126.166.862.977.718	###	0.00019242265	0.022660978	NW_022158698.1	1100207	1123308	-	NA	NA
MSTRG.8953	248.921.430.007.531	#####	125.589.396.796.131	###	0.00018777399	0.022660978	NW_022158705.1	0	19629	+	NA	NA
MSTRG.4646	36.708.315.324.579	#####	122.774.783.364.968	###	0.00022802850	0.023496850	NW_022158693.1	348342	375736	-	NA	NA
MSTRG.4836	207.891.969.313.593	#####	122.496.126.619.108	###	0.00022544053	0.023496850	NW_022158693.1	1005300	1006924	+	NA	NA
MSTRG.2361	210.371.176.628.945	#####	123.451.941.936.397	###	0.00025191007	0.024177682	NW_022158689.1	878409	899638	-	NA	NA
MSTRG.3568	32.068.673.909.631	#####	133.032.613.685.374	###	0.00025688984	0.024177682	NW_022158691.1	600753	624433	+	NA	NA
MSTRG.6556	221.099.890.531.168	#####	124.941.185.195.455	###	0.00026524039	0.024177682	NW_022158697.1	503776	524662	-	NA	NA
MSTRG.2589	593.069.106.374.059	#####	145.637.760.799.645	###	0.00028570251	0.025078332	NW_022158689.1	1757809	1784085	+	NA	NA
MSTRG.1064	23.605.764.324.165	#####	128.496.140.051.815	###	0.00036619888	0.025378007	NW_022158715.1	24960	44730	+	NA	NA
MSTRG.1297	324.320.923.985.499	#####	134.567.116.445.532	###	0.00031653249	0.025378007	NW_022158778.1	71118	94893	+	NA	NA
MSTRG.2658	281.013.974.177.802	#####	131.921.814.840.085	###	0.00033916686	0.025378007	NW_022158689.1	2024218	2046991	-	NA	NA
MSTRG.352	414.206.019.532.469	#####	14.055.367.187.739	###	0.00037127843	0.025378007	NW_022158687.1	1124179	1147531	-	NA	NA
MSTRG.4583	64.146.220.644.263	#####	148.231.019.043.318	###	0.00034937735	0.025378007	NW_022158693.1	132627	159452	+	NA	NA
MSTRG.530	389.378.422.886.225	#####	138.249.439.288.849	###	0.00031277269	0.025378007	NW_022158687.1	1891414	1915920	+	NA	NA
MSTRG.5403	329.267.291.573.073	#####	135.807.252.855.775	###	0.00036600973	0.025378007	NW_022158694.1	1146476	1168507	-	NA	NA
MSTRG.6813	180.501.319.897.526	#####	122.237.877.795.785	###	0.00037478070	0.025378007	NW_022158697.1	1335300	1355950	-	NA	NA
MSTRG.1258	195.937.882.651.437	#####	124.530.030.461.692	###	0.00039148311	0.025772638	NW_022158750.1	14801	52182	-	NA	NA
MSTRG.1373	179.006.504.544.073	#####	122.535.470.568.565	###	0.00040502742	0.025943648	NW_022158789.1	11610	32495	+	NA	NA
MSTRG.316	176.283.463.290.185	#####	122.627.757.078.555	###	0.00043535075	0.026455936	NW_022158687.1	984036	1004948	+	NA	NA
MSTRG.6661	667.540.461.236.454	#####	150.111.209.808.202	###	0.00042448327	0.026455936	NW_022158697.1	834203	858185	-	NA	NA
MSTRG.2580	343.303.693.666.268	#####	0.99882647830084	###	0.00046282451	0.027422352	NW_022158689.1	1576485	1607659	+	NA	NA
MSTRG.1268	508.027.279.710.701	#####	106.188.344.444.604	###	0.00049199931	0.027762818	NW_022158765.1	0	21743	+	NA	NA
MSTRG.1285	174.740.579.392.756	#####	123.118.895.437.859	###	0.00048447090	0.027762818	NW_022158687.1	4242773	4261626	+	NA	NA
MSTRG.1082	184.370.065.500.411	#####	125.920.486.036.337	###	0.00060104787	0.028828656	NW_022158716.1	246568	289883	-	NA	NA

Nota. *Se realizó la comparación y la unión de las tablas de las muestras de los tres análisis, ya que se observó que existen los mismos genes y se agregaron aquellos que eran distintos en cada uno de los análisis.*

3.2.3 Genes sin reportar

Los genes sin reportar fueron comparados, mediante alineación, con bases de datos como Kegg, MycoCosm, Gene Ontology (GO) y NCBI. La mayoría de genes se encuentran anotados como proteínas sin caracterizar. En la Tabla 1, se puede observar la cantidad de DEGs y las características que presentan.

Tabla 4

Genes expresados diferencialmente, comparados con bases de datos KEGG.

Características	Cantidad de DEGs
Involucrados en ataque de pared celular	
Up-regulated	
Ribosome	69
Biosynthesis of amino acids	55
Carbon metabolism	39
Ribosome biogenesis in eukaryotes	35
RNA transport	28
2-Oxocarboxylic acid metabolism	23
Citrate cycle (TCA cycle)	20
Carbon fixation pathways in prokaryotes	9
Nitrogen metabolism	9
Down-regulated	
Tyrosine metabolism	17
Styrene degradation	6

La comparación se realizó mediante las plataformas KEGG y GO, utilizando como referencia estudios realizados por Zuo y colaboradores (2017). Hubo un total de 18.065 genes codificadores que existen en el genoma completo, de ellos, como resultado de DESeq2, 12.810 fueron expresados diferencialmente. Se hizo la filtración de los datos mediante el p-value. Aquellos que tenían p-value menor o igual 0.5 se escogieron para el estudio. Se obtuvieron 1249 DEGs, de los cuales 310 representaban proteínas caracterizadas como DNA polimerasa, proteína SIX y beta-galactosidasa. Los 939 genes restantes se encontraban descritos como proteínas hipotéticas sin caracterizar. La comparación se realizó mediante las plataformas KEGG y GO, utilizando como referencia estudios realizados por Zuo y colaboradores (2017). Los resultados del análisis con DESeq2 de los tres ensayos mostraron que los DEGs que se repetían fueron:

- Argininosuccinate lyase
- Chitin synthase
- Protein kinase
- G-protein
- Mitochondrial protein
- F-box protein
- Glucosidase 1
- Transcription factor
- Secreted protein SIX1
- GMC-oxidoreductase
- D-amino-acid oxidase
- Amidohydrolase ytcJ-like
- Triacylglycerol lipase
- pH-response regulator protein palA/rim-20

- flotillin-2
- beta-glucosidase
- 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate synthase
- striatin Pro11
- taurine dioxygenase
- DNAJ like subfamily B member 12
- Ubiquitin conjugation factor E4 B
- Xanthine dehydrogenase

Los DEGs que codifican para las proteínas caracterizadas fueron Argininosuccinate lyase, Chitin synthase, Protein kinase, G-protein, Mitochondrial protein, F-box protein, Glucosidase 1, Transcription factor y Secreted protein SIX1, siendo estas proteínas previamente descritas. Las 13 proteínas anotadas restantes, son proteínas caracterizadas, sin embargo no son DEGs anotados en el genoma de referencia del hongo o en los archivos gtf de FOC RT4. Por ende, se conoce como DEGs nuevos anotados, después del ensayo de análisis de expresión diferencial delos datos de secuenciación de RNA-seq.

Capítulo IV: DISCUSIÓN

Genes expresados diferencialmente en FOC RT4 inoculado en raíces de banano

El objetivo de la presente investigación es la identificación y anotación de genes sin reportar de FOC RT4 al ser inoculado en raíces de *Musa acuminata* sub. Cavendish.

En primer lugar, es necesario obtener data que sea confiable y que haya sido generada con buenas técnicas de secuenciación, ya que al momento de realizar los análisis para datos resultantes de secuenciación avanzada de RNA, se pueden observar secuencias de mala calidad que afectarían el ensayo de expresión diferencial, considerándose como contaminantes (Andrews, 2010). Desde la Figura 5 hasta la Figura 10, se observa que existen secuencias de baja calidad sin embargo, con la aplicación de software de recorte de datos, la calidad de secuencias fue óptima para el desarrollo del análisis. Además, que todas las secuencias presentaron contenidos de GC (Guanina-Citosina) aceptables para el estudio. El recorte de datos de secuenciación es fundamental para obtener muestras de alta calidad y alta confianza para ser utilizadas en datos posteriores. Además, los datos pueden sufrir contaminación con adaptadores, sesgos de contenido de bases y secuencias con representaciones inexactas de las secuencias originales (Chen, Zhou, Chen, & Gu, 2018). En la Figura 11, se aprecia el recorte de datos que fue realizado en una de las muestras del ensayo RNA-seq1; este, comparado con los resultados previos del análisis de calidad, se puede observar que existe disminución de secuencias de baja calidad y contaminantes.

En ausencia de anotaciones recientes tanto de genes como de genoma, el desarrollo del workflow con base en el ensamblaje de novo de datos de RNA-Seq fue la opción más viable para

el estudio del transcriptoma de FOC RT4. Para obtener la descripción general del perfil de expresión de FOC RT4 frente a su hospedero, banano Cavendish, se tomaron muestras de estudios que realizaron la inoculación del hongo en raíces de plantas desarrolladas. Los datos fueron comparados con datos de secuenciación del hongo sembrado en medio de cultivo artificial, para así adquirir genes, cuya expresión haya sido alterada específicamente cuando el hongo se encuentre en contacto con las raíces (Wang , et al., 2012).

El alineamiento de los transcritos a la secuencia de genoma, fue un punto clave para un correcto mapeo de los genes que pudieron tener expresiones altamente diferenciales. Según Parsania (2019), el alineamiento es importante para lograr que lecturas no contiguas se asignen directamente al genoma y no existan errores en el mapeo. En los resultados de RNA STAR, se observa que existen varias secuencias que fueron alineadas al genoma, en donde se agruparon por el nombre del gen y por el tamaño de los reads. Además, al momento de hacer el alineamiento de los genes se comprueba que los datos están correctos ya que presentan similitudes entre el genoma de referencia y las anotaciones.

El ensamblaje y la cuantificación de los genes agrupa de mejor manera, los reads que fueron alineados al genoma previamente. featureCounts permite la asignación de reads precisos mediante la correlación que existe entre la ubicación del mapeo de cada base con la región genómica para cada una de las características. Además, permite la asignación de reads en cualquier espacio que pueden ser inserciones, deleciones, unión exón-exón o fusiones (Yang , Gordon , & Wei , 2014). Por eso, como se muestra en la Figura 13, para cada uno de los genes (geneID), se asignaron reads dependiendo de sus características que más tarde pueden servir para comparar la expresión en cada uno de los tratamientos.

El análisis de expresión diferencial se realizó mediante DESeq2, en la Figura 14 se observa un diagrama de calor de distancias entre muestras. Existe correlación entre cada una de las muestras que fueron ensambladas. El diagrama fue generado utilizando la relación log2 de los datos de transcripción de FOC RT4 en medio de cultivo frente a los datos de transcripción de FOC RT4 en la etapa vegetativa dentro de las raíces de banano (0DPI), a los 2DPI y a los 4DPI para el ensayo RNA-seq1.

Luego de realizar el análisis con DESeq2, se procedió a realizar la anotación de los DEGs mediante Annotate, en donde se encontraron que de los 18.065 genes codificadores que existen en el genoma completo, 12.810 fueron expresados diferencialmente. Para la mayoría de ellos no existió expresión diferencial significativa, por ende, se escogieron aquellos genes cuyo p-value fuese menor o igual a 0.05. Después del filtrado, se encontraron aproximadamente 1249 DEGs y de ellos 310 representaban proteínas conocidas como DNA polimerasa, proteína SIX, betagalactosidasa entre otras. La mayoría de los 939 genes restantes están descritos como proteínas hipotéticas sin caracterizar, siendo estos genes que no han sido reportados y que se expresan diferencialmente cuando el hongo está colonizando las raíces (Guo, et al., 2014).

La anotación de los genes de expresión diferencial encontrados, será la base para nuevos estudios sobre el uso de genes involucrados en la infección y la colonización de las raíces de banano, por parte de FOC RT4. Esta investigación se centró en la anotación de los genes importantes para el hongo, al momento de encontrarse en contacto con el huésped. Esto debido a la importancia económica que conlleva la producción de banano y las pérdidas causadas por el marchitamiento por Fusarium. Las siguientes proteínas que son codificadas por DEGs y que fueron encontradas en los tres ensayos, fueron previamente reportadas por Guo y colaboradores (2014): Argininosuccinate lyase, Chitin synthase, Protein kinase, G-protein,

Mitochondrial protein, F-box protein, Glucosidase 1, Transcription factor y Secreted protein SIX1. Hay otras proteínas para las que se encontró expresión diferencial significativa en el análisis, pero no han sido reportadas previamente y que cumplen un papel importante en la virulencia y patogenicidad del hongo. Por ejemplo, la proteína resultante GMS-oxidoreductase es sobreexpresada por *Fusarium oxysporum* f. sp. *lycopersici,* estudios de Kawabe y colaboradores (2011) muestran que es una proteína utilizada como biocontrol de hongos no patógenos de plantas. La enzima D-amino-acid oxidase, tiene como actividad la oxidación enantioselectiva de aminoácidos (Gabler & & Fischer, 1999). La proteína Triacylglycerol lipase tiene actividad de metabolismo lípidico, lo que sugiera que puede ser una proteína que permita el ingreso del hongo a través de las células, conjuntamente con flotillin-2 que según Haney y Long (2010) son moléculas que se encuentran activas durante la patogenicidad, endocitosis y la formación de membranas en hongos. La proteína de regulación pH-response regulator protein pala/rim-20 es propio de *Fusarium oxysporum*, el cual le permite la escisión proteolítica de factores de transcripción en presencia de medio alcalino. Los demás genes están envueltos en la patogenicida del hongo, ya que permiten la colonización del hongo a las células de las raíces de la planta. Los genes que codifican para las proteínas previamente descritas, fueron expresados diferencialmente en los tres ensayos, indicando que estos genes se encuentran al momento de la infección del hongo a las raíces de la planta de banano tipo Cavendish.

Capítulo V: CONCLUSIONES

- Los datos de secuenciación de alto rendimiento de RNA-seq reportados en múltiples estudios, sirvieron para la detección de genes que son expresados diferencialmente al momento de la interacción entre el hongo y el huésped Cavendish.

- Se identificaron alrededor de 12.810 genes que fueron expresados diferencialmente y mediante filtración de p-value se encontraron 1.249 DEGs.

- Se encontraron 310 DEGs reportados bajo proteínas efectoras conocidas como SIX, DNA polimerasa, etc., los 939 restantes se anotaron como proteínas hipotéticas sin caracterizar. Se anotaron 13 genes sin reportar que fueron expresados diferencialmente en los tres ensayos, y cuya función tiene relación con la patogenicidad del hongo.

Capítulo VI: RECOMENDACIONES

- Se recomienda realizar el ensayo de expresión diferencial en hojas y frutos, para así conocer si estos sirven como fómites. Además, para obtener una idea clara de cómo se encuentra el hongo atacando a estas partes del huésped.

- Se recomienda realizar la selección de hifas y conidios como posibles objetivos para hacer el análisis de expresión diferencial de genes y así desarrollar el tratamiento de eliminación de las esporas latentes del hongo.

Capítulo VII: BIBLIOGRAFÍA

MAGAP. (2019 de septiembre de 2019). *II Simulacro Nacional de Actuación para Minimizar el Riesgo de Introducción de Foc R4T en Ecuador*. Recuperado el 6 de enero de 2021, de Ministeria de Agricultura y Ganadería : https://www.agricultura.gob.ec/ii-simulacro-nacional-de-actuacion-para-minimizar-el-riesgo-de-introduccion-de-foc-r4t-en-ecuador/

Ordonez, N., Seidl, M., Waalwijk, C., Drenth, A., Kilian, A., Thomma, B., y otros. (2015). Worse comes to worst: bananas and Panama disease—when plant and pathogen clones meet. *PLoS pathogens , 11* (11).

Agrocalidad. (9 de 09 de 2018). *Agrocalidad .* Recuperado el 7 de 08 de 2020, de PREOCUPACIÓN DE LA DISPERSIÓN DE FOC R4T EN AMÉRICA LATINA Y EL CARIBE (ALC): https://www.ippc.int/static/media/files/publication/es/2018/09/9.2._Preocupacion_ingreso_FOC._AGROCALIDAD.pdf

FAO. (2020). «Banana Market Review: Preliminary Results 2019,.

Magdama, F. (2019). FUSARIUM OXYSPORUM-EL HONGO MÁS TEMIDO EN LA INDUSTRIA DEL BANANO. *ECUADOR ES CALIDAD , 6* (1).

Gamez, R., Rodríguez, F., Vidal, N., Ramirez, S., Vera Alvarez, R., Landsman, D., y otros. (2019). Banana (Musa acuminata) transcriptome profiling in response to rhizobacteria: *Bacillus amyloliquefaciens* Bs006 and Pseudomonas fluorescens Ps006. *BMC genomics , 20* (1), 378.

Martínez-Solórzano, G., Rey-Brina, J., Pargas-Pichardo, R., & Manzanilla, E. (2020). Marchitez por Fusarium raza tropical 4: Estado actual y presencia en el continente americano. *Agronomía Mesoamericana , 31* (1), 259-276.

Soto, M. (2011). Situación y avances tecnologicos en la producción bananera mundial. . *Revista Brasileira de Fruticultura , 33* (spe1), 13-28.

Castañeda, N., Alves, G., Almeida, R., Togawa, R., Fortes Ferreira, C., Costa, M., y otros. (2017). Gene expression analysis in Musa acuminata during compatible interactions with Meloidogyne incognita. *Annals of botany , 119* (5), 915-930.

INIBAP. (1994). *The Improvement and Testing of Musa: a Global Partnership.* (D. Jones, Ed.) France: Parc Scientifique Agropolis.

Fullerton, R., & Olsen, T. (1995). Pathogenic variability in Mycosphaerella fijiensis Morelet, cause of black Sigatoka in banana and plantain. *New Zealand Journal of Crop and Horticultural Science , 23* (1), 39-48.

Molina, A. B. (2009). RECENT OCCURRENCE OF FUSARIUM OXYSPORUM F. SP. CUBENSE TROPICAL RACE 4 IN ASIA. . *Acta Horticulturae , 828*, 109–116.

Li, C., Yang, J., Li, W., Sun, J., & Peng, M. (2017). Direct root penetration and rhizome vascular colonization by Fusarium oxysporum f. sp. cubense are the key steps in the successful infection of Brazil Cavendish. *Plant disease , 101* (12), 2073-2078.

Robinson, J., & Saúco, V. (2010). *Bananas and plantains* (Vol. 19). Cabi.

AEBE. (2020). La exportación de banano mantiene su ritmo creciente. *Bananotas , XV* (142), 16-21.

Alarcón, L., & Marzocchi, V. (2015). Evaluation for Paper Ability to Pseudo Stem of Banana Tree. *Procedia Materials Science , 8*, 814-823.

Guo, L., Han, L., Yang, L., Zeng, H., Fan, D., Zhu, Y., y otros. (2014). (2014). Genome and transcriptome analysis of the fungal pathogen Fusarium oxysporum f. sp. cubense causing banana vascular wilt disease. P. *PloS one , 9* (4).

Stover, R. H. (1962). *Fusarial wilt (Panama Disease) of bananas and other Musa species. Fusarial wilt (Panama disease) of bananas and other Musa species.* United Kingdom : Commonwealth Mycological Institute.

Nasdir, N. (2003). Fusarium wilt race 4 in Indonesia. Research Institute for Fruits west. Sumatra, Indonesia. . *Abstracts of Papers 2nd. International Symposium on Fusarium wilt on banana.* .

Li C, S. J. (2015). Analysis of banana transcriptome and global gene expression profiles in banana roots in response to infection by race 1 and tropical race 4 of Fusarium oxysporum f. sp. cubense. *BMC Genomics. , 14* (1), 851.

Conesa, A., Madrigal, P., Tarazona, S., Gomez-Cabrero, D., Cervera, A., McPherson, A., y otros. (2016). A survey of best practices for RNA-seq data analysis. *Genome biology, , 17*, 13.

Tong, L., Wu, P., Phan, J., Hassazadeh, H., Tong, W., & Wang, M.D. (2020). Impact of RNA-seq data analysis algorithms on gene expression estimation and downstream prediction. *Scientific reports , 10* (1), 17925.

Wang, D., Miwa, T., & Morikawa , T. (2020). Big Trajectory Data Mining: A Survey of Methods, Applications, and Services. *Sensors (Basel, Switzerland) , 20* (16), 4571.

Bellinger, C., Mohomed, J., Zaïane, O., & Osornio-Vargas, A. (2017). A systematic review of data mining and machine learning for air pollution epidemiology. *BMC public health , 17* (1), 907.

Afgan, E., Baker, D., Van den Beek, M., Blankenberg, D., Bouvier, D., Cech, M., y otros. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic acids research , 44* (1), 3-10.

Qi, W. S. (2017). RNA-Seq Data Analysis: From Raw Data Quality Control to Differential Expression Analysis. . *Plant Germline Development , 1669*, 295–307.

Andrews, S. (24 de abril de 2010). *Babraham Bioinformatics* . Recuperado el 6 de enero de 2021, de FastQC A Quality Control tool for High Throughput Sequence Data. : http://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Bolger, A. M. (2014). *Bioinformatics*. Recuperado el 6 de enero de 2021, de Trimmomatic: a flexible trimmer for Illumina sequence data.: http://dx.doi.org/10.1093/bioinformatics/btu170

Dobin, A., Davis, C., Schlesinger, F., Drenkow, J., .Zaleski, C., Jha, S., y otros. (2012). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics , 29* (1), 15-21.

Kovaka, S., Zimin, A., Pertea, G., Razaghi, R., Salzberg, S., & Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biology , 20* (1).

Pertea, M., Pertea, G., Antonescu, C., Chang, T.-C., Mendell, J., & Salzberg, S. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology , 33* (3), 290-295.

Trapnell, C., Williams, B., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M., y otros. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnology , 28* (5), 511-515.

Liao , Y., Smyth, G., & Shi , W. (2013). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics , 30* (7), 923-930.

Love, M., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology , 15* (12).

Wang , Z., Zhang , J., Jia , C., Liu, J., Li , Y., Yin , X., y otros. (2012). De novo characterization of the banana root transcriptome and analysis of gene expression under Fusarium oxysporum f. sp. Cubense tropical race 4 infection. B. *MC Genomics. , 13*, 650.

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics , 34* (17), i884–i890.

Yang , L., Gordon , K., & Wei , S. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics , 30* (7), 923–930.

Global Info Research. (2018). *Bussiness Risk & Industry Analysis Reports*. Obtenido de https://www.globalinforesearch.com/

Woodard, S., Jocelyne, M., Bailey, M., Barker, D., Love, R., Lane, J., et al. (2003). Maize (Zea mays)-derived bovine trypsin: characterization of the first large-scale, commercial protein product from transgenic plants. *Biotechnol.Appl.Biochem* , 123-130.

MAGAP. (2018). *Manual para el registro de empresas y productos de uso veterinario.* Obtenido de http://www.agrocalidad.gob.ec/documentos/MANUAL-PARA-EL%20REGISTRO-DE-EMPRESAS-Y-PRODUCTOS-DE-USO-VETERINARIO.pdf

Ramírez, M. (2018). *Libro de Memorias del XV Foro Internacional del Banano .* Guayaquil : Centro de Investigaciones Biotecnológicas del Ecuador .

Ministerio de Comercio Exterior. (2017). «Informe sector bananero ecuatoriano,».

South China Agricultural University. (3 de Marzo de 2020). *Cavendish banana variety 'Brailian' root inoculated with Fusarium oxysporum f.sp. cubense race1 and race4 as well as sterile water*. Recuperado el 16 de Octubre de 2020, de NCBI: https://www.ncbi.nlm.nih.gov/sra/SRX7120632[accn]

Pegg , K., Coates , L., O'Neill Wayne T, & Turner , D. (2019). The Epidemiology of Fusarium Wilt of Banana . *Frontiers in Plant Science , 10*, 1395.

Magdama , F., Monserrate-Maggi, L., Serrano, L., García Onofre, J., & Jimenez-Gasco, M. (2020). Genetic Diversity of Fusarium oxysporum f. sp. cubense, the Fusarium Wilt Pathogen of Banana, in Ecuador. *Plants , 9*, 1123.

Scheerer, L., Pemsl, D., Dita, M., Pérez-Vicente, L., & Staver, C. (2018). A quantified approach to projecting losses caused by Fusarium wilt Tropical race 4. *Acta Horticulturae. , 11* (96), 211-218.

Zuo, C., Zhang, W., Chen, Z., Chen, B., & Huang, Y. (2017). RNA Sequencing Reveals that Endoplasmic Reticulum Stress and Disruption of Membrane Integrity Underlie Dimethyl Trisulfide Toxicity against Fusarium oxysporum f. sp. cubense Tropical Race 4. *Frontiers in microbiology , 8* (8), 1365.

Parsania, M. (01 de Enero de 2019). *Performance Evaluation of Transcript-level RNA-Seq Aligners 'HISAT vs STAR'.* Recuperado el 15 de Enero de 2021, de Leiden Institute of Advanced Computer Science: https://theses.liacs.nl/pdf/2018-2019-ParsaniaMina.pdf

AUGURA, Asociación de bananeros de Colombia. (2019). Augura intensificó medidas de control y prevención ante sospecha de presencia del hongo Fusarium R4T en Colombia alertado por el ICA.

Kawabe, M., Okabe Onokubo, A., Arimoto, Y., Yoshida, T., Azegami, K., Teraoka, T., y otros. (2011). GMC oxidoreductase, a highly expressed protein in a potent biocontrol agent

Fusarium oxysporum Cong:1-2, is dispensable for biocontrol activity. . *The Journal of General and Applied Microbiology , 57* (4), 207–217.

Gabler, M., & & Fischer, L. (1999). Production of a new D-amino acid oxidase from the fungus Fusarium oxysporum. . *Applied and environmental microbiology, , 65* (8), 3750–3753.

Haney, C. H., & & Long, S. R. (2010). Plant flotillins are required for infection by nitrogen-fixing bacteria. *Proceedings of the National Academy of Sciences , 107* (1), 478.