

# **ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**

## **Facultad de Ciencias de la Vida**

Generación de una base de datos de genes involucrados en la resistencia y tolerancia del banano ante la infección por la Sigatoka Negra como instrumento para el control de enfermedades

### **PROYECTO INTEGRADOR**

Previo la obtención del Título de:

**Biólogo**

Presentado por:

Melissa Carolina Rivera Purizaga

GUAYAQUIL - ECUADOR

Año: 2022

# DEDICATORIA

A mi familia y a los que se han vuelto parte de ella.  
A Shumy por permitirme acompañarla por 13 años, desde el principio hasta el final.

## **AGRADECIMIENTOS**

Agradezco al laboratorio de biotecnología CIBE, a los doctores: Eduardo Sánchez Ph. D. por su dedicación y conocimiento brindado a lo largo de mi carrera; y Efrén Santos Ph. D por permitirme hacer uso de la información para realizar el proyecto. A Renesh Bedre Ph. D de la Universidad de Texas A&M por instruirme y guiarme en el campo de la bioinformática.

A mis compañeras de carrera, por siempre estar dispuestas a apoyarnos.

Mantener viva la memoria de los orígenes de mis antecesores me hace estar agradecida por todo el esfuerzo que hicieron, sin yo estar en la película, para darme el privilegio de poder ser la segunda generación de mi familia en tener una educación universitaria. Más que nada a mis abuelitos por darles esa oportunidad a mis padres y ellos a mí, como una de las mejores herencias. Hubiese querido que ustedes hayan tenido la misma oportunidad y espero que vivan sus sueños a través de los míos.

## DECLARACIÓN EXPRESA

“Los derechos de titularidad y explotación, me corresponde conforme al reglamento de propiedad intelectual de la institución; *Melissa Carolina Rivera Purizaga* y doy mi consentimiento para que la ESPOL realice la comunicación pública de la obra por cualquier medio con el fin de promover la consulta, difusión y uso público de la producción intelectual”



---

Melissa Carolina Rivera  
Purizaga

# EVALUADORES

---

**Diego Gallardo MSc**

PROFESOR DE LA MATERIA

---

**Luis Eduardo Sánchez Timm PhD**

PROFESOR TUTOR

## RESUMEN

La crisis climática vuelve más común el establecimiento de enfermedades de cultivos, como la Sigatoka Negra causada por el hongo *Pseudocercospora fijiensis*, que provoca alrededor de \$500 millones en pérdidas a nivel mundial anualmente en cultivos de banano. Las medidas tomadas actualmente no son viables a largo plazo por lo que se debe de optar de otras alternativas con nuevas tecnologías. El proyecto tiene como objetivo desarrollar una base de datos con genes de resistencia y/o tolerancia del banano (*M. acuminata*) ante la enfermedad de la Sigatoka Negra (*P. fijiensis*) mediante un análisis transcriptómico entre una variedad tolerante (Tuu Gia) y otra susceptible (Williams) al patógeno para el control de enfermedades.

El enfoque es netamente bioinformático, por lo que se utilizó ADNc secuenciado de un ensayo biológico previo como materia prima. Pasaron por un análisis pre-eliminar y luego un análisis de expresión diferencial para obtener una lista de genes que pasaron por un análisis comparativo con arroz para identificar homólogos. Finalmente, se escogieron genes responsables de resistencia mediante un análisis de enriquecimiento.

Se encontraron alrededor de 25 genes de tolerancia y/o resistencia. Los genes con esta característica se expresan incluso antes de la infección. Preponderan la ruta metabólica MAPK y las familias genéticas de, WRKY, AP2/EREBP, GST, entre otras.

Existe una fluctuación de la sobrerrepresentación de familias genéticas en cada tiempo post-infección muestreado. Se concluye que se pudo cumplir el desarrollo de la base de datos de genes de resistencia mediante el uso de la bioinformática.

**Palabras Clave:** Banano, Sigatoka Negra, Bioinformática, Transcriptómica, Genes resistentes.

## ***ABSTRACT***

The climate crisis makes the establishment of crop diseases more common, such as Black Leaf Streak caused by the fungus *Pseudocercospora fijiensis*, which causes around \$500 million in losses worldwide annually in banana crops. The measures currently taken are not viable in the long term, so other alternatives with new technologies should be chosen. The project aims to develop a database with genes for resistance and/or tolerance of banana (*M. acuminata*) to Black Sigatoka disease (*P. fijiensis*) through a transcriptomic analysis between a tolerant cultivar (Tuu Gia) and another susceptible (Williams) to the pathogen for disease control.

The approach is purely bioinformatic, so sequenced cDNA from a previous biological assay was used as raw material. They went through pre-eliminary analysis and then differential expression analysis to obtain a list of genes that went through comparative analysis with rice to identify homologues. Finally, genes responsible for resistance were chosen by enrichment analysis.

About 25 tolerance and/or resistance genes were found. Genes with this feature are expressed even before infection. The MAPK metabolic pathway and the genetic families of WRKY, AP2/EREBP, GST, among others, predominate.

There is a fluctuation of the overrepresentation of genetic families at each time post-infection sampled. It is concluded that the development of the resistance gene database could be accomplished using bioinformatics.

**Keywords:** Banana, Black Leaf Streak, Bioinformatics, Transcriptomics, Resistance genes.

# ÍNDICE GENERAL

<b>EVALUADORES .....</b>	<b>IV</b>
<b>RESUMEN.....</b>	<b>V</b>
<b>ABSTRACT .....</b>	<b>VI</b>
<b>ÍNDICE GENERAL.....</b>	<b>VII</b>
<b>ABREVIATURAS .....</b>	<b>IX</b>
<b>SIMBOLOGÍA .....</b>	<b>XI</b>
<b>ÍNDICE DE FIGURAS .....</b>	<b>XII</b>
<b>ÍNDICE DE TABLAS .....</b>	<b>XIV</b>
<b>CAPÍTULO 1 .....</b>	<b>1</b>
<b>1. Introducción.....</b>	<b>1</b>
<b>1.1 Descripción del problema.....</b>	<b>1</b>
<b>1.2 Justificación del problema.....</b>	<b>1</b>
<b>1.3 Objetivos .....</b>	<b>3</b>
1.3.1 Objetivo General .....	3
1.3.2 Objetivos Específicos.....	3
<b>1.4 Marco teórico .....</b>	<b>3</b>
1.4.1 El banano .....	3
1.4.2 Sigatoka Negra .....	5
1.4.3 Genes relacionados a la infección.....	6
1.4.4 RNA-seq y transcriptómica .....	7
<b>CAPÍTULO 2 .....</b>	<b>9</b>
<b>2. Metodología.....</b>	<b>9</b>
<b>2.1 Metodología preliminar .....</b>	<b>9</b>

2.2	Corte de secuencias.....	11
2.3	Limpieza de secuencias .....	11
2.4	Mapeo .....	12
2.5	Cuantificación de genes.....	12
2.6	Expresión diferencial.....	12
2.7	Análisis comparativo .....	15
2.8	Análisis funcional y de enriquecimiento .....	17
2.8.1	GenFam: Identificación de familias de genes enriquecidas .....	17
2.8.2	AgriGO: Identificación de términos GO enriquecidos .....	17
2.8.3	KOBAS: Identificación de rutas metabólicas enriquecidas.....	18
<b>CAPÍTULO 3.....</b>		<b>20</b>
<b>3.</b>	<b>Resultados y análisis .....</b>	<b>20</b>
3.1	Análisis de expresión diferencial .....	20
3.2	Análisis comparativo .....	22
3.3	Análisis funcional .....	23
<b>CAPÍTULO 4 .....</b>		<b>42</b>
<b>4.</b>	<b>Conclusiones Y Recomendaciones .....</b>	<b>42</b>
	Conclusiones .....	42
	Recomendaciones .....	43
<b>BIBLIOGRAFÍA .....</b>		<b>44</b>
<b>APÉNDICES .....</b>		<b>51</b>

## ABREVIATURAS

ESPOL Escuela Superior Politécnica del Litoral

ODS Objetivos de Desarrollo Sostenible

ADN Ácido Desoxirribonucleico

ARN Ácido Ribonucleico

BLSD Sigatoka Negra (Black Leaf Streak Disease)

Foc4 *Fusarium oxysporum* raza 4

SSH Supressive Subtractive Hybridization

CDS Secuencia codificante

ADNc ADN complementario

SSH Supressive Subtractive Hybridization

RNA-seq secuenciación de RNA

PAMP Patrones Moleculares Asociados a Patógenos

HTML Hypertext Markup Language

JSON JavaScript Object Notation

SAM Sequence Alignment Map format

Hpi Horas post infección

KEGG Kyoto Encyclopedia of Genes and Genome

Blastn Basic Local Alignment Search Tool nucleotide

GC guanina-citosina

GO Gene Ontology

DAG Direct Acyclic Graph

SEA Singular Enrichment Analysis

FC Fold of Change

CPM Counts Per Million

DGE Differential gene expression

# SIMBOLOGÍA

ha Hectárea

## ÍNDICE DE FIGURAS

Figura 2.1 Elementos que componen los diagramas DAG. Mientras más intenso es el color del recuadro, mayor es el nivel por lo que es una relación más significativa. Tomado de AgriGO. ....	18
Figura 3.1 Gráfica de exploración de datos de manera multi-dimensional entre los grupos de tratamiento. ....	20
Figura 3.2 Gráfico de expresión diferencial en la comparación entre los controles de Williams y Tuu Gia. ....	21
Figura 3.3 Las gráficas representan la expresión diferencial entre el control vs 3 días post-infección. A la derecha se puede observar la comparación en Tuu Gia y a la izquierda en Williams ....	21
Figura 3.4 Gráfico de expresión diferencial que representa la comparación entre ambos cultivares luego de 3 días post-infección. ....	22
Figura 3.5 Gráfico de enriquecimiento de rutas metabólicas expresadas positivamente generado en KOBAS de la comparación entre los controles de ambas variedades. ....	24
Figura 3.6 Familias genética expresadas positivamente sobrerrepresentadas en la comparación base entre los controles de ambas variedades. Obtenida de GenFam. ...	24
Figura 3.7 Familias genética expresadas negativamente sobrerrepresentadas en la comparación base entre los controles de ambas variedades. Obtenida de GenFam. ...	26
Figura 3.8 Ruta metabólica de señalización de MAPK obtenido de KEGG.....	27
Figura 3.9 Gráfico de enriquecimiento de rutas metabólicas expresadas negativamente generado en KOBAS de la comparación entre los controles de ambas variedades. ....	27
Figura 3.10 Mapa de ontología genética obtenido de AgriGO. Se muestran términos GO relevantes en el contraste entre los controles de Williams y Tuu Gia.....	28
Figura 3.11 Contraste entre Williams y Tuu Gia de genes expresados positivamente luego de 12 horas de infección. Obtenida de GenFam.....	30
Figura 3.12 Análisis de enriquecimiento de rutas metabólicas en KOBAS contrastando Williams y Tuu Gia luego de 12 horas de infección. ....	31
Figura 3.13 Análisis de enriquecimiento de familias genéticas luego de 24 hpi. Se muestran las familias sobrerrepresentadas positivamente. Obtenido de GenFam. ....	32

Figura 3.14 Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 24 hpi. ....	32
Figura 3.15 Familias de genes expresadas positivamente de manera diferencial luego de 72 hpi. Generado en GenFam. ....	33
Figura 3.16 Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 72 hpi .....	34
Figura 3.17 Familias de genes expresadas negativamente de manera diferencial 72 hpi. Generado en GenFam .....	35
Figura 3.18 Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 72 hpi .....	35
Figura 3.19 Familias de genes expresadas negativamente de manera diferencial luego de 144 hpi. Generado en GenFam .....	36
Figura 3.20 Familias de genes expresadas positivamente de manera diferencial luego de 168 hpi. Generado en GenFam .....	37
Figura 3.21 Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 72 hpi .....	38
Figura 3.22 Familias de genes expresadas positivamente de manera diferencial luego de 168 hpi. Generado en GenFam .....	39

## ÍNDICE DE TABLAS

Tabla 2.1 Esquema de ensayos y días cuando hubo recuperación de tejido vegetal luego de la infección. ....	9
Tabla 2.2 Información de los resultados de secuenciación en cada librería. Elaborado por: Efrén Santos Ph.D. ....	10
Tabla 2.3 Tratamientos usados en el análisis de expresión diferencial en R. Los tratamientos entre cultivares Williams vs Tuu Gia (12, 24, 72, 144, 168 hpi) tienen restado el control (6 hpi). ....	14
Tabla 2.4 Genes homólogos de <i>M. acuminata</i> encontrados por BLASTx con las especies <i>A. thaliana</i> , <i>O. sativa</i> y <i>Z. mays</i> . ....	16
Tabla 3.1 Genes seleccionados como candidatos de resistencia y/o tolerancia con sus respectivas familias genéticas o rutas metabólicas. ....	40

# CAPÍTULO 1

## 1. INTRODUCCIÓN

### 1.1 Descripción del problema

El banano (*Musa acuminata*) es una de las frutas que más se consume en el mundo por su alto valor nutritivo y por consecuencia, es uno de los más exportados a nivel global, con incrementos en su producción entre el 2019 y el 2020 en un 0,76%, lo que representa \$13,9 mil millones (Voora et al., 2020). Ecuador se posiciona como el primer país exportador de esta fruta, siendo una de las principales fuentes de ingresos con un promedio de \$28,4 millones al año en exportaciones a países de la Unión Europea (FAO, 2020).

La Sigatoka Negra es una enfermedad provocada por el hongo *Pseudocercospora fijensis*, el cual es uno de los más agresivos de su género que provoca pérdidas aproximadas de \$500 millones globalmente en cada ciclo de cultivo (Arango Isaza et al., 2016). Esta enfermedad es una epidemia creciente que podría afectar el cultivo de banano y a la economía de los países que dependen de ello, como lo ocurrido anteriormente con el Mal de Panamá, el cual extinguió una variedad completa a nivel comercial. Las medidas que se han tomado para frenar las enfermedades radican en el uso de pesticidas que a su vez incrementan el valor del fruto y puede provocar resistencia en las plagas. Otro enfoque novedoso es la de desarrollar plantas resistentes o tolerantes a la enfermedad, lo cual se puede lograr mediante la identificación de genes potenciales que participen en rutas metabólicas de resistencia durante la infección.

### 1.2 Justificación del problema

La globalización ha mantenido en pie el comercio del banano, viéndose Ecuador mayormente beneficiado. No obstante, al mismo tiempo este contacto constate entre diversas partes del mundo ha incurrido en la introducción de especies ajenas a áreas geográficas, provocando desequilibrios en los ecosistemas. Esto se puede trasladar a la

introducción de enfermedades, como la Sigatoka Negra que se detectó por primera vez en Honduras en 1969 para llegar a Ecuador el 1987. Actualmente la enfermedad trata de ser controlada por diversos métodos, sin embargo, el manejo y control inadecuado de la enfermedad puede provocar el contagio de cultivos sanos pudiendo causar una epidemia. Además de la aplicación indiscriminada de pesticidas lo cual genera cepas resistentes. Al tener ese riesgo inminente es imperativo tomar medidas preventivas, así como también planes alternos para cuando se convierta en una epidemia, ya que puede acabar con variedades que no poseen resistencia al patógeno. Dicha catástrofe no solo atenta contra el ODS 8 de trabajo decente y crecimiento económico, sino también el ODS de hambre cero porque afecta a población que tiene esta fruta como su principal alimento.

Además, a largo plazo el cambio climático contribuirá a un aumento de enfermedades de cultivos, ya que, en específico con la Sigatoka Negra, se crea un ambiente propicio para su crecimiento. Según Bebbber (2019) el riesgo de infección ha aumentado en un 44,2% desde 1960, especialmente en América del Sur.

El acercamiento que tiene este trabajo es el de identificar genes candidatos de resistencia y tolerancia que pueden ser utilizados para un programa de mejoramiento genético. La generación de variedades resistentes contribuiría en el aumento del rendimiento de la producción lo cual se necesita a largo plazo debido a que para el año 2050 la población crecerá en un 35%. Por otro lado, el proyecto brinda la metodología o la línea de trabajo para realizar este proceso con otras plantas y con otras enfermedades, siendo una plataforma para tener cultivos tolerantes o resistentes contra futuras enfermedades, asegurando los ODS de trabajo digno y hambre cero. Además, al implementar los resultados de la investigación, se facilita una base de datos lo cual es beneficioso ya que según Passos (2012) a pesar de que el banano sea una fruta ampliamente conocida, no se tiene mucha información de su transcriptoma. El análisis transcriptómico se realiza mediante RNA-seq (secuenciación de ARN), lo cual representa una ventaja sobre aplicaciones basadas en hibridación (Sánchez Timm et al., 2016; Swarupa et al., 2013) ya que brinda información más detallada de expresiones altas y

bajas de genes, sin limitarse a conocimiento previo del genoma (Marguerat & Bähler, 2009).

### **1.3 Objetivos**

#### **1.3.1 Objetivo General**

Desarrollar una base de datos con genes de resistencia y/o tolerancia del banano (*Musa acuminata*) ante la enfermedad de la Sigatoka Negra (*Pseudocercospora fijiensis*) mediante un análisis transcriptómico entre una variedad tolerante (Tuu Gia) y otra susceptible (Williams) al patógeno para el control de esta y otras enfermedades.

#### **1.3.2 Objetivos Específicos**

1. Comparar los genes relevantes expresados diferencialmente en ambas variedades con plantas modelos para la realización del análisis funcional.
2. Analizar las rutas metabólicas involucradas en las variedades susceptible y tolerante mediante el análisis funcional para la selección de genes candidatos tolerantes y resistentes involucrados en el proceso de infección.

### **1.4 Marco teórico**

#### **1.4.1 El banano**

##### **1.4.1.1 Generalidades**

*M. acuminata* es una herbácea gigante monocotiledónea monocárpica que posee un rizoma bajo tierra y un pseudo tallo, que son un grupo de hojas plegadas pudiendo alcanzar de 2 a 5 metros. Tienen la capacidad de producir nuevas plantas mediante su rizoma. Una vez terminada su cosecha muere. Tiene raíces adventicias por su naturaleza de propagación vegetativa. Posee diferentes tipos de hojas: escama, luego hojas espada y la hoja cigarro, la cual es la que nace en el centro; la emergencia completa de las hojas termina después de 10 o 12 semanas (Israeli & Lahav, 2016).

Además, es un buen recurso nutritivo por compuestos bioactivos como polifenoles y flavonoides. La fruta es fuente de hierro y potasio, así como de serotonina (Singh et al., 2016). Al banano se ubica como la tercera fruta más consumida en el mundo entero. Se estiman que las exportaciones del 2019 fueron de 21 millones de toneladas, aportando un 26% al PIB agrícola y generando más de 2,5 millones de puestos de trabajo (Betancourt et al., 2020). Además, según Apolo (2021), 190.381 ha de tierra se dedican a la agricultura del banano con un rendimiento de 35,91 cajas por ha.

#### **1.4.1.2 Variedades y mejoramiento genético**

El banano, con nombre científico *Musa acuminata* pertenece a la familia Musaceae y al orden Zingiberales. Se originó en parte de Oceanía y Asia; y su centro de diversidad está localizado en Malasia e Indonesia. Se distinguen 7 subespecies que se distribuyen en diferentes locaciones y que poseen distinta morfología. Su domesticación se dio con mayor facilidad mediante propagación vegetativa que por semillas, esto a largo plazo llevó a que se hibriden diferentes variedades para generar nuevas combinaciones de frutos comestibles partenocárpicos. La primera variedad conocida fue la de Gros Michel que fue luego reemplazada por Cavendish, siendo una de las más producidas actualmente junto con sus respectivos somaclones.

La variedad Williams fue el primer mutante en un Cavendish que surgió en Australia en 1910. Es utilizada por las características de su fruto y la eficiencia en su producción, así como también por la resistencia morfológica que posee ante fuertes vientos y condiciones ambientales desfavorables. No obstante, carece de resistencia a la Sigatoka Negra y a *Fusarium* raza 4 (ProMusa, 2020). Por otro lado, la variedad Tuu Gia es un diploide silvestre originaria de Vietnam resistente a la Sigatoka Negra y a *Fusarium* raza 4.

El mejoramiento genético del banano comenzó para poder generar una variedad parecida a Gros Michel, pero con resistencia al mal de Panamá, que fue la causa de su desaparición. Además, hay otros tipos de mejora mediante las mutaciones, pero no siempre se generan plantas con una buena calidad de fruta a pesar de que cumplan con tener resistencia (Israeli & Lahav, 2016).

## **1.4.2 Sigatoka Negra**

### **1.4.2.1 Generalidades**

La Sigatoka Negra es una enfermedad del banano provocada por el hongo *Pseudocercospora fijiensis* (previamente denominado *Mycosphaerella fijiensis*) que pertenece a la clase Dothideomycetes y fue descrito por primera vez en 1963 en las islas Fiji (Mulder & Stover, 1976). El hongo es hemibiotrófico, lo que significa que coloniza entre las células de los mesófilos sin causar estragos a la planta, este proceso suele durar de 3 a 4 semanas, dependiendo de la región climática. Posteriormente comienza su etapa necrótica en la cual obtiene nutrientes del tejido vegetal afectado por toxinas (Portal et al., 2011).

La expresión de los síntomas ocurre a una a dos semanas luego de la infección, que ocurren en hojas nuevas, comenzando desde la hoja cigarro y la primera hoja desplegada, lo que indica un gradiente de evolución de la enfermedad. En las plantas infectadas se pueden visualizar daños foliares tornando secciones de las hojas amarillas y luego negras que indican necrosis, lo cual afecta la actividad fotosintética y como consecuencia la producción de frutos. No obstante, los frutos que logra producir una planta infectada se maduran rápidamente (Castelan et al., 2012).

El hongo posee una preferencia por lugares cálidos de baja elevación y húmedos, así como Ecuador. Por otro lado, la dispersión geográfica de *P. fijiensis* se da a lo largo de grandes áreas mediante vía aérea por la producción de ascosporas. Según Kimunye (2021), se la considera una de las especies más agresivas del género ya que posee mecanismos de evasión de control de enfermedades en el banano por su alta tasa de recombinación, gran población, alta diversidad y capacidad de dispersión

### **1.4.2.2 Manejo y control**

Actualmente la estrategia más eficaz de control empleada es evitar la entrada al país de material vegetal contaminado mediante vigilancia y diagnóstico molecular. La poda de hojas jóvenes enfermas es otra alternativa para controlar la dispersión del patógeno en el cultivo. Adicionalmente, se aplican fertilizantes y pesticidas, tanto protectores como

sistémicos que pueden ser una ventaja con un plan de manejo adecuado, no obstante, genera resistencia en el patógeno (Arango Isaza et al., 2016).

Se necesitan de 50 a 66 ciclos anuales de aplicación de pesticidas aumentando los costos de producción de 15 a 27 % (Israeli & Lahav, 2016). Los daños provocados por el hongo ascienden a un 76% de la producción total, por lo que varios autores (Alakonya et al., 2018; Churchill, 2011; Rodriguez et al., 2016) indican la inviabilidad de tener una sola variedad de consumo masivo (Williams). Nuevas variedades con resistencia poligénicas son imperativas para combatir la creciente epidemia de la Sigatoka Negra.

### **1.4.3 Genes relacionados a la infección**

Investigaciones de expresión diferencial en la relación planta-patógeno se han realizado con anterioridad en el caso de *M. acuminata* y *P. fijiensis* con diferentes estrategias, tales como RNA-seq, microarrays, SSH (suppressive subtractive hybridization). El descubrimiento de genes involucrados en el metabolismo de la infección es una herramienta útil en programas de selección genética ya que es importante la generación de variantes con resistencia o tolerancia poligénica. Según de Bellaire (2010) la resistencia poligénica es preferible ya que dificulta la fácil adaptación del patógeno.

Según un análisis usando RNA-seq, las variedades que muestran una mayor expresión diferencial durante el tiempo de infección son las que probablemente poseen algún grado de resistencia como DH Pahang y Calcutta 4. En un estudio de análisis comparativo de secuencias de ARN (RNA-seq) en DH Pahang, se han encontrado genes ortólogos en *A. thaliana*, que detectan patrones moleculares asociados a patógenos (PAMPs; por sus siglas en inglés) los cuales activan la inmunidad innata de la planta en presencia de *P. fijiensis*. También se han encontrado altamente expresados receptores citoplasmáticos asociados a quinasas que regulan señales intracelulares (Noar & Daub, 2016).

Otro acercamiento se ha dado para estudiar el transcriptoma mediante SSH (Portal et al., 2011) cuando hay interacción con *P. fijiensis*. Enzimas relacionadas a la ruta metabólica de fenilpropanoides se encontraron en cultivares susceptibles. En la etapa necrotrófica se encontraron rutas de señales de transcripción de resistencia de jasmonato y etileno.

Estas rutas en cultivares resistentes se activan antes por la acumulación de la toxina 2,4,8-trihidrotetralona. Por otro lado, en cultivares resistentes como Calcutta IV, en etapas iniciales de infección, genes que codifican dominios de dedos de zinc, metalotioneinas, involucrados en la glicolisis o glucogenólisis, así como la proteína putativa de resistencia a la enfermedad RGA1 se vieron expresados positivamente (Sánchez Timm et al., 2016).

De acuerdo con las investigaciones realizadas por Rodriguez (2016) usando la tecnología de microarray en la variedad resistente Calcutta se encontraron expresadas de manera positiva proteínas relacionadas a la patogénesis, genes codificantes de peroxidasa y fenilalanina amoniaco liasa. La expresión de estas enzimas se dio en la etapa temprana de infección luego de un día de inoculación. No obstante, no se vieron genes expresados en la variedad susceptible Williams luego de 144 horas post infección.

#### **1.4.4 RNA-seq y transcriptómica**

RNA-seq es una tecnología que se ha desarrollado en los últimos 10 años, con el objetivo de obtener la secuencia de ARN de una muestra, para cuantificar y luego explorar la información de acuerdo con el objetivo de la investigación, que puede ir desde el análisis diferencial de expresión génica (DGE; por sus siglas en inglés) a isoformas. La metodología consiste primero en aislar el ARN de una muestra y se seleccionan las moléculas principales a través de diversos métodos. Dependiendo de la tecnología de secuenciación que se vaya a utilizar, el siguiente paso es la fragmentación del ARN y la síntesis de ADNc de doble cadena que luego se ligará a adaptadores para que el equipo secuencie los fragmentos. El último paso es realizar una amplificación por PCR, para que la muestra sea suficiente para facilitar la secuenciación, seguida de la selección del tamaño y finalizando con la secuenciación (Stark et al., 2019).

Se han desarrollado tres tecnologías principales para RNA-seq. La secuenciación de fragmentos cortos, que usa la tecnología Illumina, necesita fragmentar la muestra para leerla correctamente y es útil principalmente para investigar la DGE. La secuenciación de fragmentos largos y la secuenciación directa de ARN no necesitan el paso de fragmentación; PacBio y Nanopore son las tecnologías utilizadas respectivamente,

además de que se utilizan principalmente para evaluar isoformas en una muestra porque mantienen la longitud del fragmento, lo que conserva información útil de la secuencia (Conesa et al., 2016).

Con los datos adquiridos, el siguiente paso es analizarlos. En primer lugar, los datos deben limpiarse eliminando lecturas incorrectas, verificar el contenido de GC, descartar lecturas de baja calidad y eliminar adaptadores. A continuación, se implementa la alineación con el transcriptoma de referencia o de novo para vincular las transcripciones a una coordenada de referencia del genoma. Posteriormente, los transcritos se cuantifican para medir la abundancia con normalización, vinculando los datos a una región específica del genoma. La información cuantitativa podría expresarse en una matriz que represente muestras y cada uno de los transcritos expresadas. El software y el análisis estadístico son necesarios en el paso de modelado, siendo este último el que va a ayudar a interpretar los resultados por visualización o resultados de diferentes modelos estadísticos (Stark et al., 2019).

# CAPÍTULO 2

## 2. METODOLOGÍA

### 2.1 Metodología preliminar

El proyecto tiene un carácter bioinformático partiendo de las secuencias de RNA-seq en formato fastq de un tratamiento entre cultivares con diferente susceptibilidad, por lo que se explica a breves rasgos el ensayo en el que se obtuvo la información.

Las variedades: susceptible (Williams) y resistente (Tuu Gia) fueron sometidas a un bioensayo con la inoculación del hongo patógeno *P. fijiensis* sin duplicados y con un control cada uno muestreados a las 6 y 12 horas. Los controles de 6 y 12 horas fueron agrupados como un solo control. Fueron un total de 6 muestreos en cada uno luego de diferentes tiempos post infección (tabla 2.1). Se extrajo el ARN y se prepararon librerías que fueron posteriormente secuenciadas “single-end” usando la plataforma Illumina HiSeq-2000. Resultados de la secuenciación se pueden observar en la tabla 2.2.

**Tabla 2.1 Esquema de ensayos y días cuando hubo recuperación de tejido vegetal luego de la infección.**

Ensayos	Tiempo post infección
Williams y Tuu Gia	6 horas
	12 horas
	24 horas
	3 días
	7 días
Control	6 horas
	12 horas

**Tabla 2.2 Información de los resultados de secuenciación en cada librería. Elaborado por: Efrén Santos Ph.D.**

# Índice	Index 1 (I7)	% Lecturas indentificadas (PF) Lane 1	% Lecutras indentificadas (PF) Lane 2	% Lecturas indentificadas (PF) Lane 3	% Lecturas indentificadas (PF) Lane 4	Muestra
1	TGACC A	9,71	9,75	9,70	9,73	Williams Ctrl 6, 12 hpi
2	ACAGT G	10,67	10,55	10,69	10,58	Williams M. fijiensis 6 hpi. R1, R2
3	GCCAA T	6,04	6,07	6,04	6,07	Williams M. fijiensis 12 hpi. R1, R2
4	CAGAT C	7,52	7,53	7,52	7,53	Williams M. fijiensis 24 hpi. R1, R2
5	CTTGT A	9,58	9,47	9,59	9,51	Williams M. fijiensis 3 dpi. R1, R2
6	AGTCA A	8,39	8,43	8,39	8,43	Williams M. fijiensis 7 dpi. R1, R2
7	AGTTC C	9,09	9,06	9,09	9,08	Tuu Gia Ctrl 6, 12 hpi
8	ATGTC A	8,58	8,60	8,58	8,61	Tuu Gia M. fijiensis 6 hpi. R1, R2
9	CCGTC C	5,98	5,87	5,98	5,91	Tuu Gia M. fijiensis 12 hpi. R1, R2
10	GTCCG C	9,66	9,53	9,66	9,58	Tuu Gia M. fijiensis 24 hpi. R1, R2

11	GTGAA A	6,06	6,08	6,06	6,09	Tuu Gia M. fijiensis 3 dpi. R1, R2
12	CGTAC G	5,25	5,24	5,24	5,25	Tuu Gia M. fijiensis 7 dpi. R1, R2

## 2.2 Corte de secuencias

Los 12 archivos fastq con tratamientos diferentes en las diferentes variedades fueron cortados a un millón de lecturas para proseguir con el proyecto. El propósito de acortar los archivos fue de facilitar los procesos subsecuentes en una computadora de escritorio.

## 2.3 Limpieza de secuencias

Las secuencias obtenidas en bruto fueron corregidas con el programa fastp (Chen et al., 2018). El programa realiza el control de calidad de las lecturas mediante la filtración, corte de secciones a los extremos de baja calidad, corte de adaptadores y colas de polyG/polyX. Se generaron tres archivos: un fastq corregido y los reportes de las estadísticas de calidad en formato html y json que contienen el resumen de modificaciones, el contraste de calidad antes y después, contenido de bases y conteo de k-meros. El código utilizado se observa en anexos.

Un chequeo más detallado de la calidad de los archivos fastq filtrados previamente fue realizado usando FastQC (Andrews, 2010). Este programa mostró las estadísticas por calidad de bases, competencia de adaptadores, contenido GC, entre otras. Según los resultados y los parámetros que indica el programa se seleccionaron los que tienen una calidad suficiente para proceder a los otros análisis. Todas las secuencias fueron seleccionadas.

## 2.4 Mapeo

Se utilizó el genoma más reciente de *M. acuminata* variedad malaccensis o DH Pahang versión 4 junto con archivo de anotación obtenido de Banana Genome Hub (<https://banana-genome-hub.southgreen.fr/organism/1>). En primer lugar, el archivo de anotación al estar en formato gff, fue convertido a gtf para ser procesado, lo cual se realizó con el paquete AGAT (Dainat, 2022). Por otro lado, se utilizó el programa STAR (Dobin et al., 2013) para efectuar el mapeo o alineación de las lecturas con el genoma de referencia. Luego se creó el índice con el genoma descargado y el documento de anotación en formato gtf, lo que generó un directorio con varios archivos importantes para el paso de mapeo. Posteriormente se llevó a cabo el alineamiento por cada archivo fastq. Como resultado se obtuvieron el mismo número de archivos en formato SAM y otros que muestran las estadísticas como el número de lecturas procesadas y el éxito que tuvieron en el mapeo.

## 2.5 Cuantificación de genes

Luego del mapeo, con los archivos en formato SAM que se obtuvieron y el de anotación se estimó la abundancia de transcritos mediante el programa featureCounts (Liao et al., 2014). Se removi6 el encabezado y se creó un nuevo archivo solamente con la primera columna de genes y las últimas doce columnas que representan a cada muestra.

## 2.6 Expresión diferencial

Para el análisis de expresión diferencial, se instaló el programa R versión 4.1.1 junto con el paquete edgeR (Robinson et al., 2010). Con los datos del archivo de texto se creó una matriz y se asignaron diferentes grupos a cada variedad y tiempo post infección. Posteriormente se elaboró una lista DGE y una matriz adicional que refería cada variedad con su correspondiente tiempo post infección en horas lo cual facilitó procedimientos posteriores. Se calcularon los factores de normalización y se elaboró una matriz de modelo de diseño considerando todos los grupos sin la intercepción. Se decidió un valor de coeficiente de variación biológico de 0,2 debido a la falta de replicados para realizar un análisis de variabilidad. Por otro lado, diferentes contrastes fueron realizados tomando

en cuenta el modelo lineal generalizado dentro y entre los cultivares, considerando el tiempo post infección, debido a que hay 12 condiciones diferentes tratadas en el estudio, así como la prueba de razón de verosimilitud (likelihood ratio test) que es utilizada cuando no hay réplicas. El estudio se centró en evaluar los genes en la variedad Tuu Gia por su resistencia, ya que, en la comparación, como se puede observar en el anexo, se restaron los genes encontrados en el control de ambas además de los de la variedad susceptible Williams y se obtuvo finalmente los genes netamente expresados en esa variedad. En la tabla 2.3 se pueden observar los contrastes realizados. Como resultado principal se tiene una tabla por cada comparación donde se encuentran los genes expresados diferencialmente clasificados por varios estadísticos. Posteriormente la tabla se filtró siguiendo  $\log_2fc > |1|$ , que excluye genes que no sufren cambios significativos y FDR menor a 0,5 que indica la tasa de errores tipo I. Posteriormente los nombres de los genes se extrajeron en un archivo csv.

**Tabla 2.3 Tratamientos usados en el análisis de expresión diferencial en R. Los tratamientos entre cultivares Williams vs Tuu Gia (12, 24, 72, 144, 168 hpi) tienen restado el control (6 hpi).**

Código	Cultivares	hpi		hpi	
W.12vs6	Williams	12	<b>vs</b>	6 (control)	
W.24vs6		24			
W.72vs6		72			
W.144vs6		144			
W.168vs6		168			
T.12vs6	Tuu Gia	12		<b>vs</b>	6 (control)
T.24vs6		24			
T.72vs6		72			
T.144vs6		144			
T.168vs6		168			
TvsW.6h	Tuu Gia vs Williams	6	<b>vs</b>		6
TvsW.12h		12			12
TvsW.24h		24			24
TvsW.72h		72			72
TvsW.144h		144			144
TvsW.168h		168		168	

## 2.7 Análisis comparativo

El análisis de expresión diferencial en R dio como resultado listas de genes de banano para cada tratamiento realizado. Luego se efectuó la asignación de esos ID de las listas con su correspondiente región codificante (CDS), para lo cual se descargó el archivo de CDS en formato fasta de la variedad DH Pahang (obtenida de <https://banana-genome-hub.southgreen.fr/node/50/7720975>). La asignación se dio mediante la utilización del paquete biokit (Bedre, 2020). Luego de obtenidas las secuencias para cada ID de las listas se procedió a realizar el análisis comparativo.

El análisis comparativo tiene como objetivo relacionar los genes que se obtuvieron en el análisis de expresión diferencial con genes de especies que son organismos modelos, es decir, buscar una homología. Los organismos modelos poseen una anotación exhaustiva de sus secuencias, por lo que poseen más información, en contraste con las anotaciones de las secuencias de *M. acuminata*. Los genes homólogos fueron encontrados mediante la herramienta BLASTx de NCBI (Altschul et al., 1990) que busca nucleótidos en una base de datos de proteínas.

Se realizaron dos BLASTx respectivamente con la especie modelo *Arabidopsis thaliana* y la gramínea de importancia comercial *Oryza sativa* o arroz. En la tabla 2.4 se muestran los hits de genes expresados de manera positiva y negativa en cada comparación que se hizo en el análisis diferencial por cada especie. Se puede observar que existen más ocurrencias con respecto a la gramínea, por lo que se la escogió para el posterior análisis.

**Tabla 2.4 Genes homólogos de *M. acuminata* encontrados por BLASTx con las especies *A. thaliana*, *O. sativa* y *Z. mays*.**

Código	Número genes expresados diferencialmente		Número de genes encontrados por BLASTx			
			<i>Arabidopsis thaliana</i>		<i>Oryza sativa</i>	
	+	-	+	-	+	-
W.12vs6	204	431	203	419	204	428
W.24vs6	751	804	741	796	751	799
W.72vs6	1055	1061	1041	1042	1053	1058
W.144vs6	199	167	194	160	197	167
W.168vs6	1090	1088	1075	1072	1086	1081
T.12vs6	160	357	158	345	160	354
T.24vs6	986	705	976	699	983	703
T.72vs6	1103	825	1087	814	1099	822
T.144vs6	198	312	195	302	197	312
T.168vs6	1283	993	1268	983	1278	990
TvsW.6h	100	210	97	203	99	209
TvsW.12h	19	6	19	5	19	6
TvsW.24h	14	3	14	3	14	3
TvsW.72h	28	13	28	12	28	13
TvsW.144h	6	7	6	7	6	7
TvsW.168h	45	13	44	13	45	13

## **2.8 Análisis funcional y de enriquecimiento**

El análisis funcional se realiza para poder dar un sentido biológico a las listas de genes. Esto se da mediante la asignación de familias de genes, términos de ontología genética (GO) o de la identificación de rutas metabólicas sobrerrepresentados, lo que da más precisión para descubrir el involucramiento que tienen los genes en la relación planta-patógeno.

Para realizar este proceso se crearon nuevas listas de genes expresados diferencialmente a partir de las derivadas del análisis comparativo con arroz, pero clasificándolos entre expresión positiva y negativa. Estas fueron utilizadas para realizar el análisis funcional en el cual se utilizaron tres herramientas diferentes. Esta separación permitió efectuar un contraste entre los procesos que están más presentes (expresión positiva) y los que disminuyen (expresión negativa).

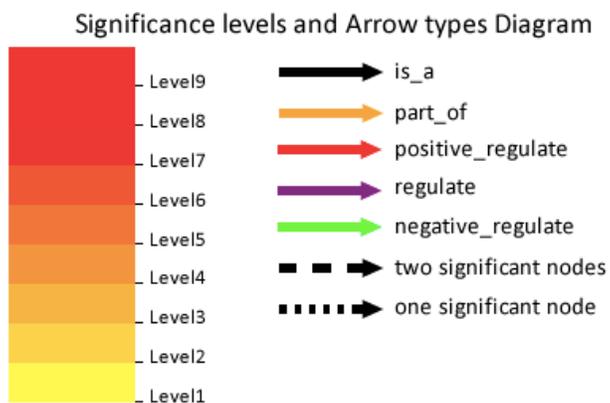
### **2.8.1 GenFam: Identificación de familias de genes enriquecidas**

GenFam es una herramienta web o de línea de comando que permite la anotación, clasificación y enriquecimiento de genes significativos según familias sobrerrepresentadas (Bedre & Mandadi, 2019). En un ciclo se procesó cada lista de genes obtenidas de las comparaciones entre diferentes tratamientos. Tres archivos fueron generados en este proceso: uno que muestra el resumen, otro las estadísticas de enriquecimiento y finalmente un gráfico de barras de la familia de genes encontrados con su valor p.

### **2.8.2 AgriGO: Identificación de términos GO enriquecidos**

AgriGO es una herramienta web que realiza un análisis de Ontología Genética en plantas comerciales (Tian et al., 2017). La Ontología Genética es una iniciativa que compendia una representación de los atributos de genes y productos de estos, de una manera estructurada, denominándolos 'términos' (Carbon et al., 2019). Se clasifican en: funciones moleculares, procesos biológicos y componentes celulares.

Para realizar el análisis se seleccionó en plantas y posteriormente en Poáceas, que es la familia a la cual pertenece el arroz. Luego se elige SEA (Singular Enrichment Analysis) como herramienta de análisis y se selecciona la especie *Oryza sativa* japónica. En el recuadro se ingresaron las listas de genes. Los resultados que se obtuvieron fueron: Un resumen de todos los ID genéticos con sus respectivos términos GO, los genes con sus términos GO más significativos y gráficos DAG (Direct Acyclic Graph) que muestran jerárquicamente con códigos de color las relaciones entre los términos GO ya sean de funciones moleculares, procesos biológicos y componentes celulares. Se puede observar los elementos que componen este diagrama en la figura 2.1.



**Figura 2.1 Elementos que componen los diagramas DAG. Mientras más intenso es el color del recuadro, mayor es el nivel por lo que es una relación más significativa. Tomado de AgriGO.**

### 2.8.3 KOBAS: Identificación de rutas metabólicas enriquecidas.

Las rutas metabólicas significativas en cada lista de genes fueron exploradas mediante el programa web KOBAS. Este permite anotar una lista de genes con rutas putativas y relaciones patogénicas sobrerrepresentadas ya que tiene a KEGG como base de datos principal (Xie et al., 2011).

Antes de ingresar la lista de genes se tuvo que convertir el formato de los ID de genes de MSU a RAP, dependiendo de la base de datos que haya sido obtenido, ya que este último formato es el aceptado por KOBAS. La herramienta de Rice Genome Hub fue utilizada (<https://rice-genome-hub.southgreen.fr/content/id-converter>) para la conversión.

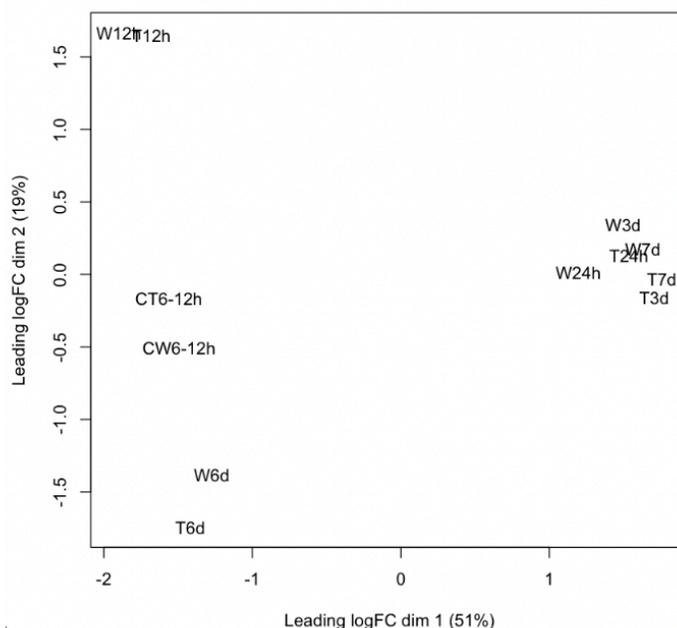
Luego de ya convertidos los ID se ingresó a la página de análisis de rutas metabólicas. Como resultado se obtuvieron: un archivo con las rutas metabólicas con un valor p menor a 0.05 y gráficos de las rutas enriquecidas según el valor p y de enriquecimiento.

# CAPÍTULO 3

## 3. RESULTADOS Y ANÁLISIS

### 3.1 Análisis de expresión diferencial

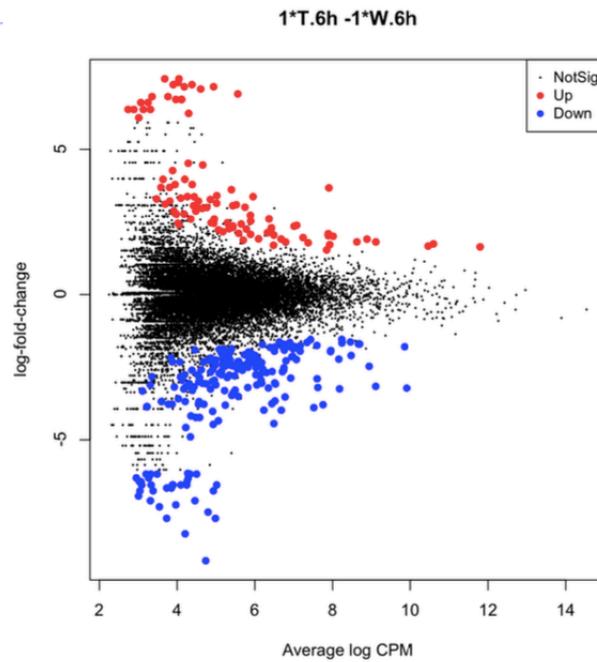
En primer lugar, se realizó una exploración de las matrices de conteo que se tenían de cada tratamiento principal mediante una gráfica multidimensional basadas en fold of change (FC) que permite visualizar diferencias entre los perfiles de expresión. Esta gráfica (figura 3.1) demuestra que la expresión de ambos cultivares en los mismo tiempos post infección tienden a ser similares, debido a las agrupaciones que se pueden observar.



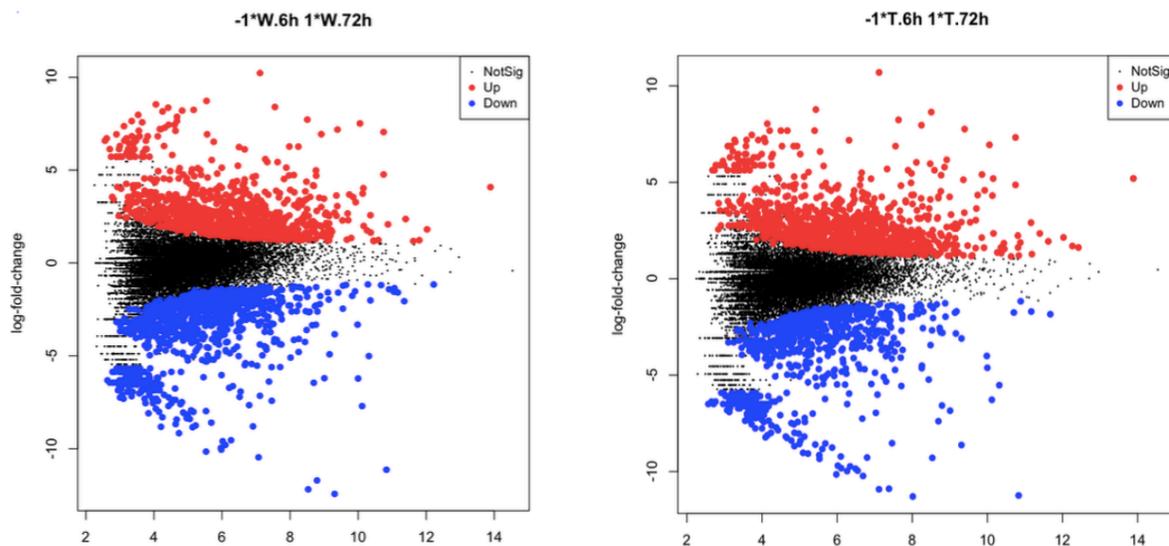
**Figura 3.1 Gráfica de exploración de datos de manera multi-dimensional entre los grupos de tratamiento.**

Luego de realizar el análisis de expresión diferencial se obtuvo el número de genes expresados de manera positiva y negativa, como se puede observar en las primeras dos columnas de la tabla 2.4. Además, gráficos que comparan FC, que se refiere al nivel de expresión, ya sea negativa o positiva, con CPM (counts per million) que es la cantidad

de copias en millones. Posteriormente se mostrarán algunas de las gráficas que se realizaron dentro del estudio.

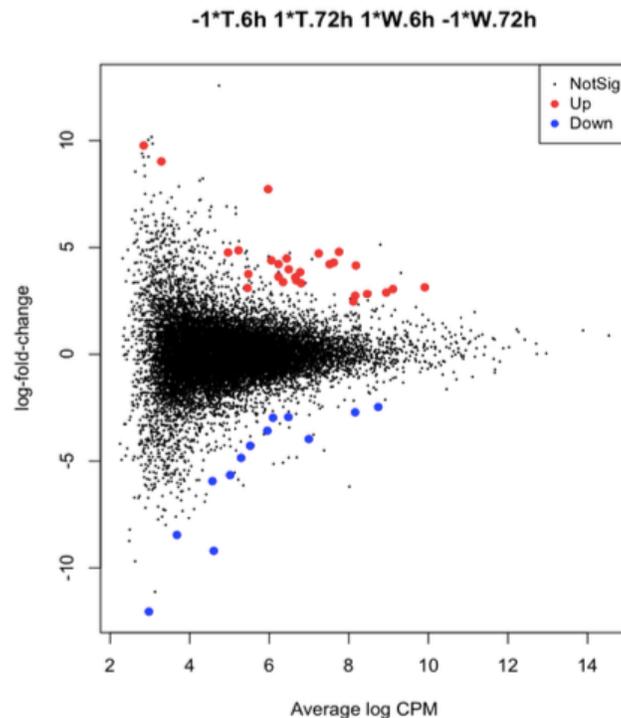


**Figura 3.2 Gráfico de expresión diferencial en la comparación entre los controles de Williams y Tuu Gia.**



**Figura 3.3 Las gráficas representan la expresión diferencial entre el control vs 3 días post-infección. A la derecha se puede observar la comparación en Tuu Gia y a la izquierda en Williams**

En la figura 3.2 podemos observar la comparación que se realizó entre los controles de las variedades. La información que brinda el gráfico revela diferencias en la expresión normal de la planta sin ser infectada. También, en la figura 3.3 se observan dos gráficas que representan los contrastes entre el control y 3 días post-infección, en Williams (izquierda) y Tuu Gia (derecha). Las gráficas se ven similares, no obstante, el gráfico 3.4 demuestra que hay genes que se expresan de manera diferente entre las dos variedades.



**Figura 3.4 Gráfico de expresión diferencial que representa la comparación entre ambos cultivares luego de 3 días post-infección.**

### 3.2 Análisis comparativo

Se encontró que el banano tiene mejores hits en arroz que en Arabidopsis. La cantidad de genes totales de banano al realizar el blastx fue de 14.236 con 7.241 genes expresados de manera positiva y 6.995 genes expresados negativamente. Los hits con la planta modelo *A. thaliana* fueron en total de 98,49%, mientras que, en arroz, al ser una planta modelo cultivable monocotiledónea, posee más similitud con banano con 99,63% de hits.

### 3.3 Análisis funcional

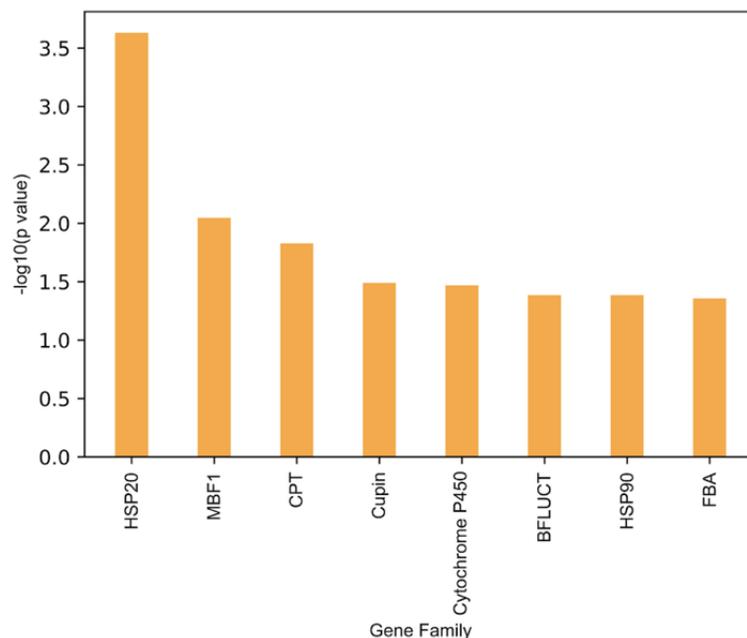
Para hacer el análisis funcional se consideraron las comparaciones entre las variedades en el estadístico R. Algunas de estas no pudieron ser analizadas de manera funcional ya que no poseen una cantidad de genes suficientes para ser procesados por los programas utilizados en esta sección.

#### **Control**

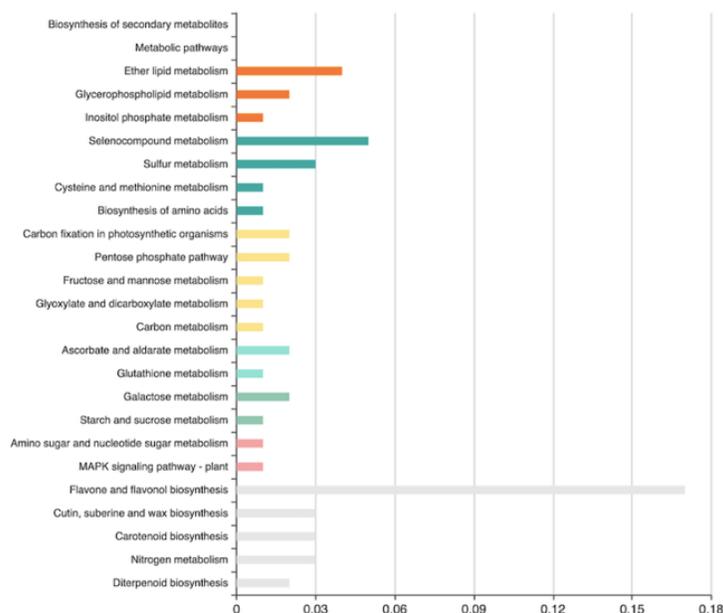
Se realizó un análisis de expresión diferencial base para medir el efecto inicial que tiene Tuu Gia. En primer lugar, analizando los genes positivos expresado, de manera significativa, en la figura 3.5, ocho familias genéticas fueron identificadas mediante GenFam. HSP20 (Heat Shock Protein) es una de las familias genéticas más sobrerrepresentadas. Esta familia actúa como chaperona y confiere tolerancia a estrés biótico y abiótico, así como molécula de control de calidad proteico (Lee et al., 2012). Estudios han demostrado que juega un rol en la inmunidad innata (Park & Seo, 2015) por lo que inicialmente puede ser considerado como de interés en la interacción planta patógeno. Se pudo identificar el término GO 0006952 de respuesta a las defensas con actividad en el retículo endoplasmático, además del gen responsable LOC\_OS04G01740.1. De igual manera ocurre con HSP90 que tiene la misma funcionalidad de la familia proteica anterior pero también interactúa con otras proteínas que dan estabilidad a la proteína R que reconoce patrones moleculares de patógenos.

La familia de genes MBF1 (Multipleprotein Bridging Factor 1) fue la segunda con mayor significancia. Se ha identificado esta familia en *A. thaliana* como responsable de la inducción de estrés biótico y abiótico, mostrando una mayor tolerancia a la salinidad y acción de patógenos cuando se encuentra expresada de manera positiva (Zhang et al., 2019). Además, según los términos GO, se encuentra involucrado en metabolismo del ácido abscisico y el etileno, lo que concuerda con estudios previamente realizados (Sánchez Timm et al., 2016). El gen involucrado en esta familia es LOC\_OS08G27850.1.

Estos términos corresponden también a la familia genética FBA (Aldolasa fructosa-1, 6-bifosfato) y a Beta-fructofuranosidasa (BFLUCT).



**Figura 3.5 Familias genética expresadas positivamente sobrerrepresentadas en la comparación base entre los controles de ambas variedades. Obtenida de GenFam.**



**Figura 3.6. Gráfico de enriquecimiento de rutas metabólicas expresadas positivamente generado en KOBAS de la comparación entre los controles de ambas variedades.**

En KOBAS cada cluster se refiere a un grupo de metabolismo que tiene actividad similar, en este caso en la figura 3.6 se muestran 7 clusters principales según el radio de enriquecimiento, es decir, la relación entre la cantidad de genes ingresados que pertenecen al metabolismo y la totalidad de genes de la planta. No hay rutas metabólicas enriquecidas significativas.

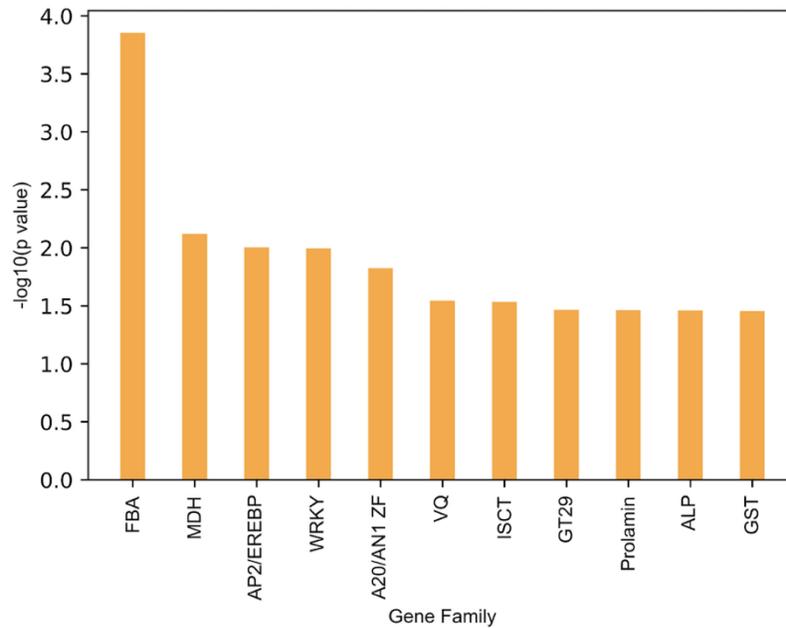
Con respecto a la expresión negativa en esta comparación, se puede observar en la figura 3.7 que en estas condiciones los grupos familiares están mucho más sobrerrepresentados. La familia FBA se encuentra en mayor proporción, en contraste con lo observado en el gráfico positivo. FBA regula las respuestas de la planta a fitohormonas como ácido abscísico, cierre de estomas y es antagonista de rutas de etileno y ácido salicílico. El etileno es una hormona del estrés para que activen señales de defensa de evasión de patógenos

Esto indica que los genes de susceptibilidad van a estar en menor proporción ya que se ha encontrado que reacciona contra el ácido abscísico (Lievens et al., 2017). Se ha encontrado que puede suprimir la acción de respuesta de la planta contra el patógeno, siendo utilizada como una molécula efectora. Los genes involucrados son LOC\_OS11G07020.1, LOC\_OS05G33380.1, LOC\_OS01G02880.1.

Se encontró también la familia AP2/EREBP es la responsable de responder en estrés biótico con los términos GO asociados ruta de señalamiento activada por auxinas (GO:0009734), ruta de señalización mediada por ácido jasmónico (GO:0009867), regulación de señalización de ácido abscísico (GO:0009788), regulación etileno (GO:0010105). Los genes involucrados son: LOC\_OS06G09390.2, LOC\_OS02G43790.1, LOC\_OS01G12440.1, LOC\_OS04G46250.1, LOC\_OS01G58420.1. Por otro lado, estudios han demostrado que los genes RSV y RTSV se encuentran expresados positivamente cuando hay estrés biótico (Sharoni et al., 2011).

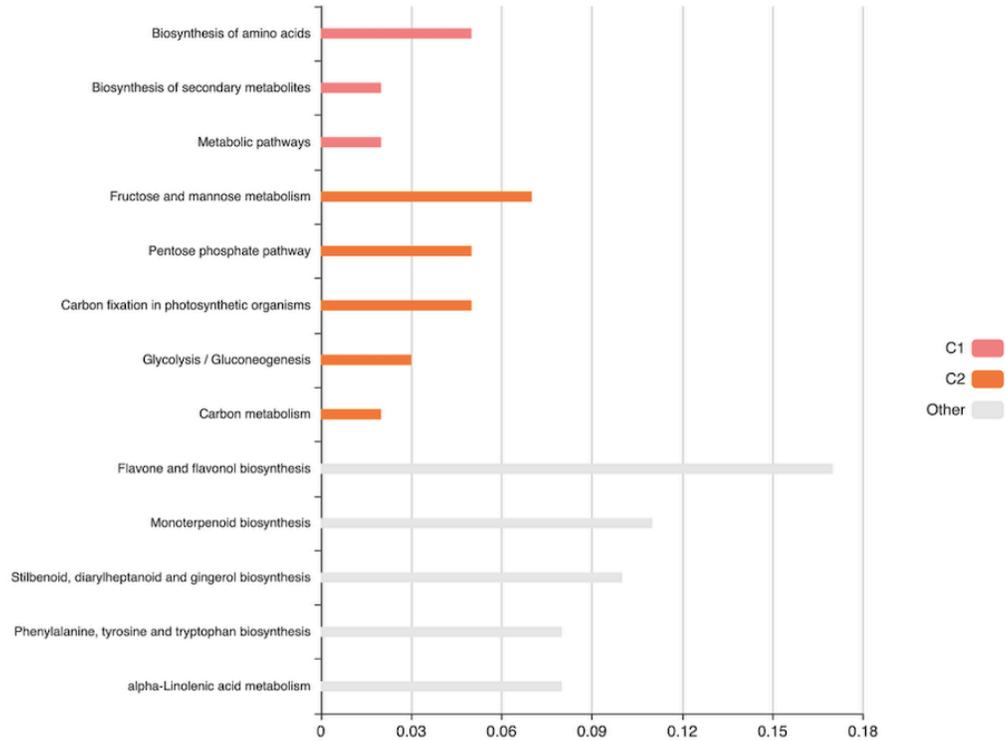
Además, WRKY tiene en su familia proteínas que juegan un papel importante en regular las respuestas de defensa inducidas por el patógeno, lo que demuestra potencial como candidato de resistencia ya que se posee el término GO de respuesta de defensa contra

hongos (GO:0050832), pero al no necesitarse aun en esta etapa, no está expresada de manera positiva. El término GO mencionado anteriormente también se lo encuentra en metabolismo de lípido acil (ALP), Glutation s-transferasa (GST), transporte de solutos inoránicos (ISCT). Genes que pertenezcan a esta familia tienen un potencial de hacerle frente a patógenos, por lo que se los consideraría genes candidatos. No obstante, se encuentra en baja proporción ya que aún no existe infección por parte del patógeno.

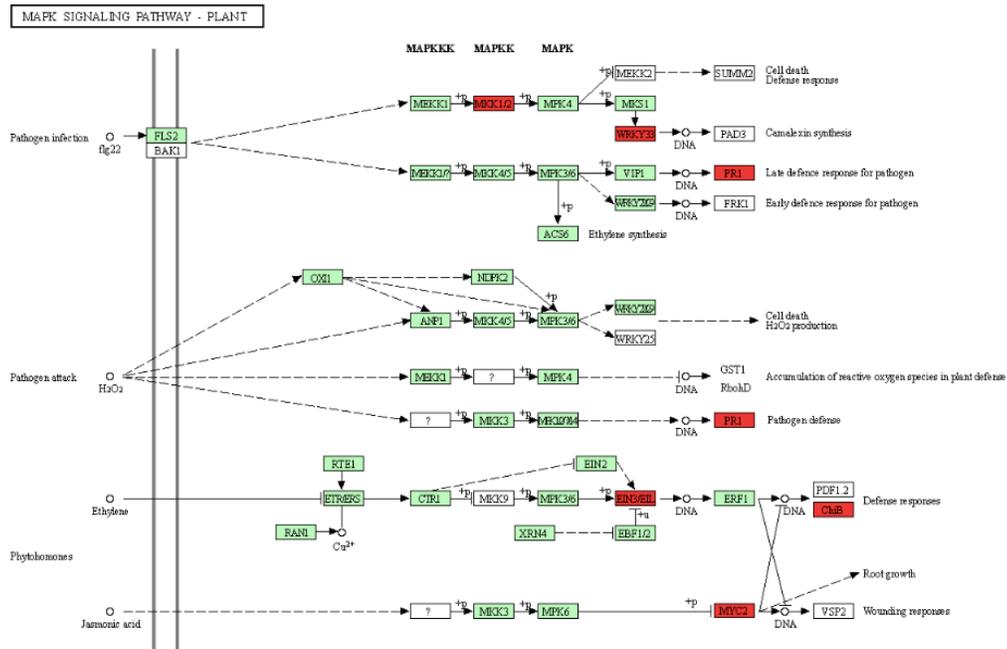


**Figura 3.7 Familias genética expresadas negativamente sobrerrepresentadas en la comparación base entre los controles de ambas variedades. Obtenida de GenFam.**

Con respecto a las rutas metabólicas identificadas, en la figura 3.8, las más destacadas que se pueden corroborar con estudios previamente realizados son la de: ruta de señalización MAPK (LOC\_Os01g28500.1, LOC\_Os01g61080.1, LOC\_Os03g20790.1, LOC\_Os06g05520.1, LOC\_Os06g51060.1, LOC\_Os10g42430.1), transducción de señales hormonales y síntesis de flavonoides. En la figura 3.9 se muestra un mapa de ruta metabólica obtenida de KEGG de la señalización MAPK. Esta ruta de señalización se encuentra involucrada en la activación de la inmunidad de la planta, incluso, se ha encontrado que esta ruta tanto de la planta como del hongo patógeno interactúan generando un diálogo de señales moleculares (Hamel et al., 2012). También, provoca mecanismos de defensa dirigidos directamente a las células fúngicas.



**Figura 3.9** Gráfico de enriquecimiento de rutas metabólicas expresadas negativamente generado en KOBAS de la comparación entre los controles de ambas variedades. No se encontraron rutas significativas.



**Figura 3.8** Ruta metabólica de señalización de MAPK obtenido de KEGG.

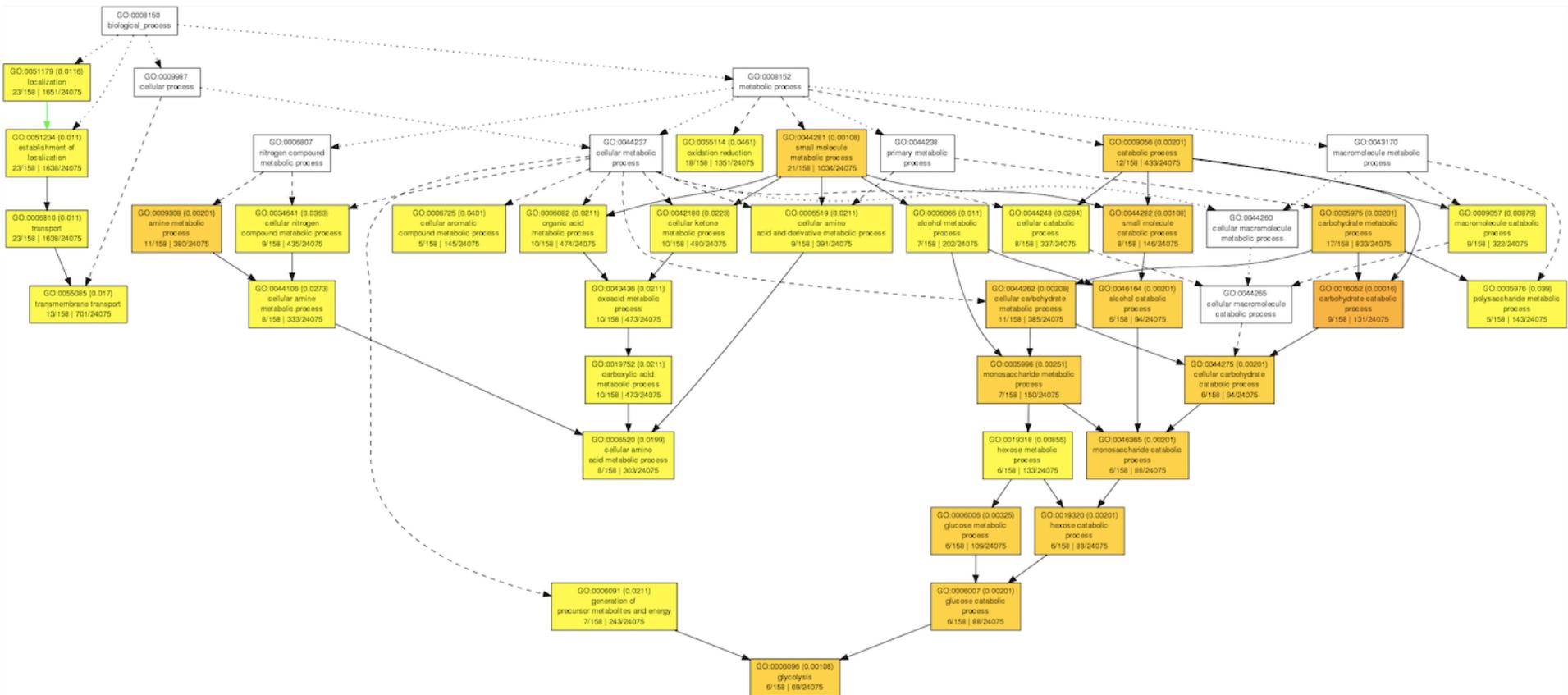


Figura 3.10 Mapa de ontología genética obtenido de AgriGO. Se muestran términos GO relevantes en el contraste entre los controles de Williams y Tuu Gia.

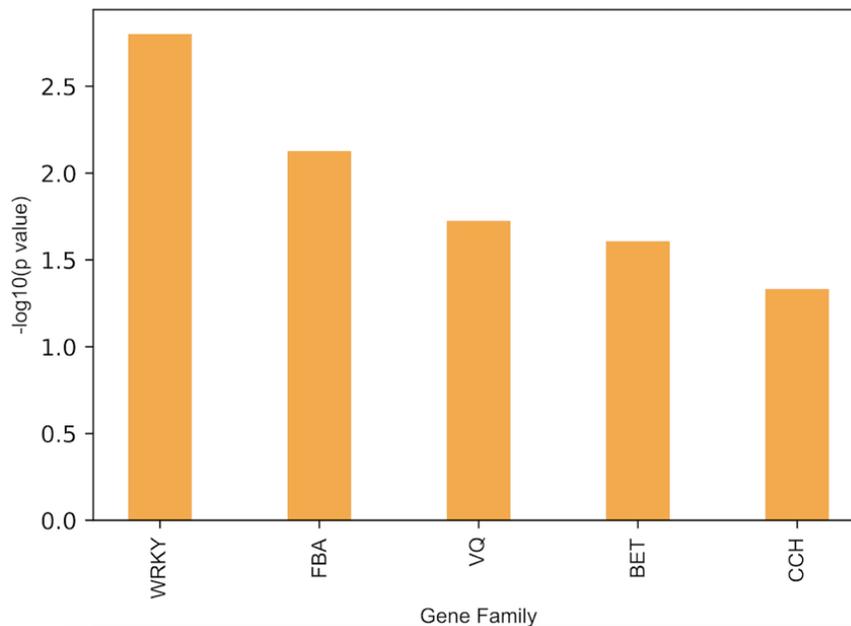
En la figura 3.10 se puede observar el gráfico DAG de procesos biológicos de términos GO. Los recuadros más oscuros son los más significativos. En este caso, los procesos catabólicos de carbohidratos (GO:0016052).

### ***12 horas post infección***

La comparación entre Williams y Tuu Gia luego de 12 horas de infección dio como resultado cinco familias de genes sobrerrepresentadas positivamente, como se puede observar en la figura 3.11. La familia WRKY es la más significativa, a la cual se le asocia según términos GO: ruta de señalamiento de ácido abscísico (GO:0009738), como se mencionó en la comparación anterior con los controles, siendo los genes importantes: LOC\_OS05G46020.1, LOC\_OS05G25770.1. Al encontrarse en este caso regulado de manera positiva y no negativa como en los controles, se puede identificar como un gen candidato de tolerancia ante el patógeno.

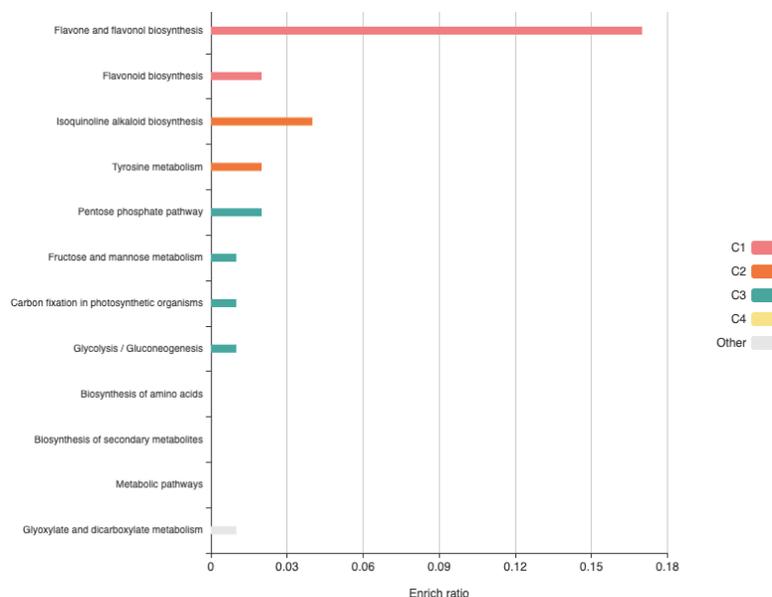
Además, se puede observar que la familia FBA es la segundo más sobrerrepresentada, lo que indica que se generará una acción contra el ácido abscísico, que, según investigaciones, este compuesto puede coadyuvar en las infecciones, siendo un objetivo para los patógenos (Hu & Bidochka, 2021). Esta familia está expresada de manera negativa en el control. Gen de interés para esta familia es LOC\_OS05G33380.1.

La familia de genes VQ, son reguladores transcripcionales que juegan diversos roles en respuesta a defensas y tolerancia hacia estrés (Jing & Lin, 2015). El ácido salicílico (GO:0010337) y ácido jasmónico fueron identificados como términos importantes en procesos biológicos del análisis de la ontología genética dentro de esta familia. Se ha demostrado que la acumulación de estos compuestos indica que ha habido una acción por parte del patógeno para dar pie a las respuestas de defensa. Esta familia tiene el mismo valor p que en las identificadas como expresión positiva en los controles. El gen involucrado aquí es LOC\_OS01G59410.1.



**Figura 3.11 Contraste entre Williams y Tuu Gia de genes expresados positivamente luego de 12 horas de infección. Obtenida de GenFam.**

También se identificó la familia de genes pertenecientes a bromodominio y dominios extra terminales (BET) que regula la expresión de genes por modificaciones en su cromatina y se encuentra involucrada en rutas metabólicas activadas por ácido abscísico y a respuestas de defensa. Por otro lado, se tiene a la familia chaperona del cobre (CCH) que se expresa cuando secciones de la hoja están senesciendo. Se encontró además asociado el término GO de respuesta de defensa hacia hongos (GO:0050832), siendo el gen LOC\_OS12G23280.4 uno de los involucrados. La ruta metabólica identificada mediante KOBAS en la figura 3.12, con valor p corregido menor a 0.05 es la de biosíntesis de flavon y flavonol con el gen LOC\_Os03g25150.1. Esta ruta puede ser considerada en la relación planta patógeno ya que es una especie de repelente hacia estos.



**Figura 3.12 Análisis de enriquecimiento de rutas metabólicas en KOBAS contrastando Williams y Tuu Gia luego de 12 horas de infección.**

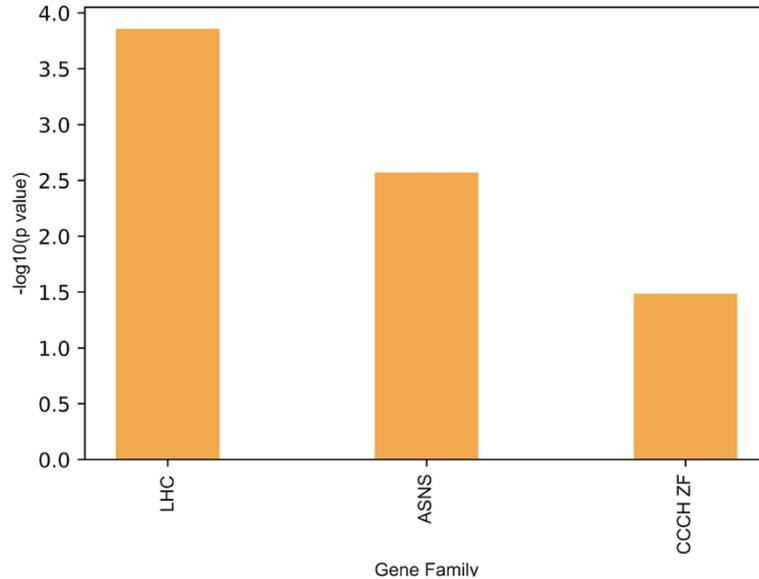
Con respecto a los resultados de expresión negativa, no hay una cantidad suficiente para determinar genes enriquecidos.

### ***24 horas post infección***

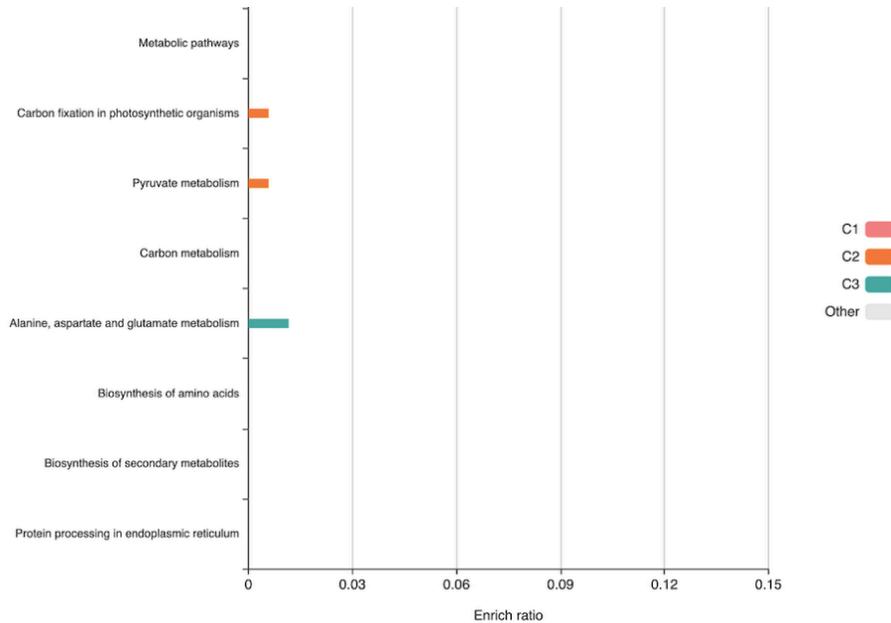
Los genes expresados diferencialmente, en mayor proporción fueron positivos, por lo que solo de estos se pudo realizar el análisis de enriquecimiento. La familia que se encuentra mayormente representada es la de unión a la clorofila A/B (LHC) (figura 3.13). Según sus términos GO, se encuentra involucrado en el proceso biológico de respuesta molecular a estímulo de ácido abscísico, además de la mayoría de los componentes que se asociación con la fotosíntesis. Incluso, la ruta metabólica de fotosíntesis se encuentra mayormente sobrerrepresentada en el análisis en KOBAS como se puede observar en la figura 3.14.

La otra familia, asparagina sintasa (ASNS) ha sido caracterizada anteriormente en la planta de tabaco, confiriendo resistencia ante patógenos virales (Hwang et al., 2011). Aun no se la ha caracterizado en infecciones fúngicas, por lo que explorar su rol en este tipo de interacción planta patógeno es de utilidad para identificar genes involucrados. Finalmente, CCCH familia genética de dedos de zinc (CCCH ZF) se encuentra ligada a

modificaciones post-transcripcionales y a procesos de estrés, que incrementan la tolerancia.



**Figura 3.13** Análisis de enriquecimiento de familias genéticas luego de 24 hpi. Se muestran las familias sobrerrepresentadas positivamente. Obtenido de GenFam.

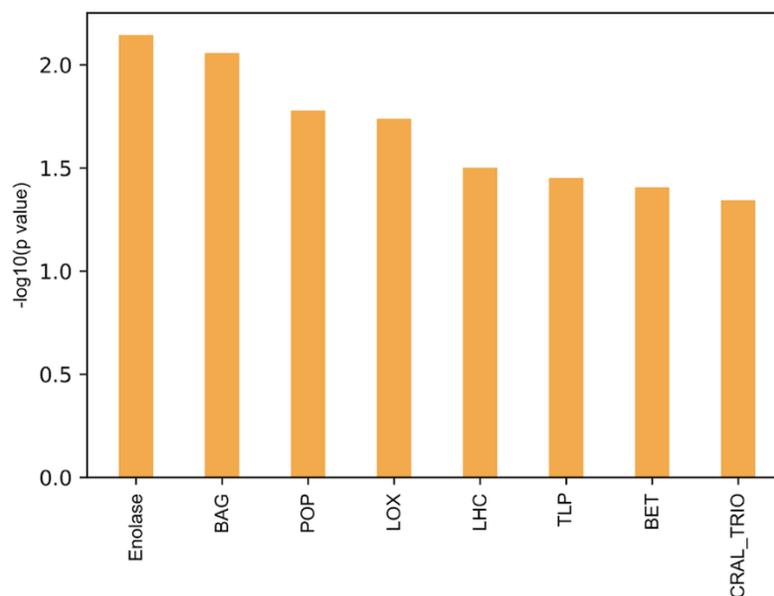


**Figura 3.14** Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 24 hpi.

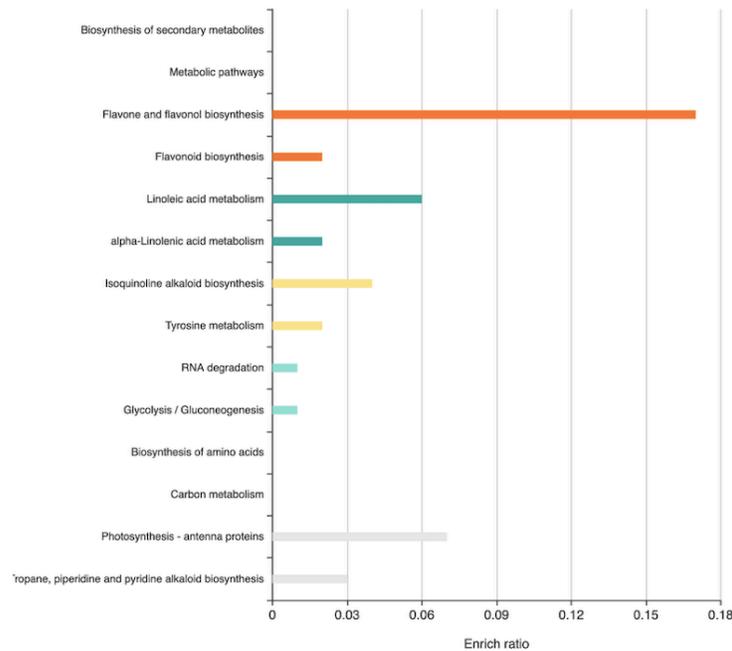
## 72 horas post infección

Tomando en cuenta las familias genéticas que se involucran en la relación planta-patógeno. La familia Bcl-2-asociada a atánogeno (BAG) confiere resistencia basal contra patógenos fúngicos necrofíticos, como se demostró en una investigación con *A. thaliana* y el patógeno *Botrytis cinerea* (Kabbage et al., 2016). Además, términos GO asociados a esta familia son la de respuesta a patógeno fúngico, a estrés biótico, especies reactivas de oxígeno, con el gen involucrado LOC\_OS05G43670.1, que es de interés para conferir resistencia al cultivo. La segunda familia que cumple un rol importante es la de lipo oxigenasas (LOX) que se acumula cuando existe estrés biológico de la planta, senescencia y por la acumulación de ácido jasmónico y abscísico, con modificaciones en la pared celular. Siendo el gen de interés LOC\_OS02G10120.1 que brinda resistencia o tolerancia al presentarse expresado de manera positiva.

La familia de LHC redujo su expresión en comparación a la etapa de infección anterior, que incluso fue un metabolismo significativo en esta etapa como se puede observar en la figura 3.16 y BET se expresó con la misma frecuencia que luego de 12 hpi.

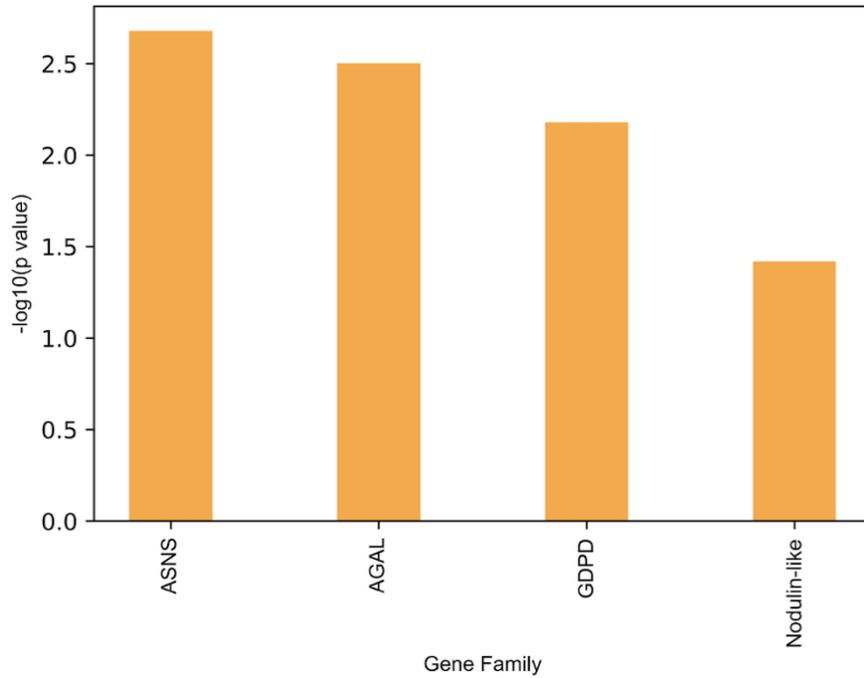


**Figura 3.15 Familias de genes expresadas positivamente de manera diferencial luego de 72 hpi. Generado en GenFam.**

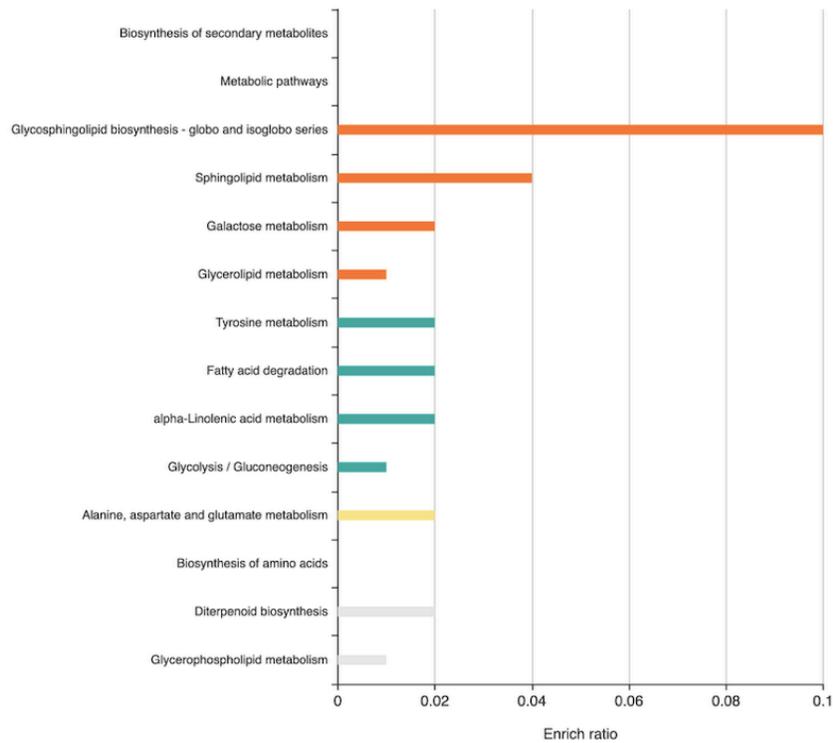


**Figura 3.16 Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 72 hpi**

Con respecto a los resultados de la expresión negativa, se pudieron obtener cuatro familias genéticas sobrerrepresentadas como se puede observar en la figura 3.17. La familia genética ASNS que confiere resistencia ha disminuido su expresión lo que puede significar la superación de la infección (LOC\_OS06G15420.1). Así como la familia de galactosidasa (AGAL) que tiene un rol similar sobre el patógeno con el gen LOC\_OS10G35110.2; y la familia genética de glicerofosfodiéster fosfodiesterasa (GDPD) con gen LOC\_OS02G31030.2, que se puede ver su sobrerrepresentación en el análisis en KOBAS (figura 3.18) en la ruta de biosíntesis de glico esfingolípidos.



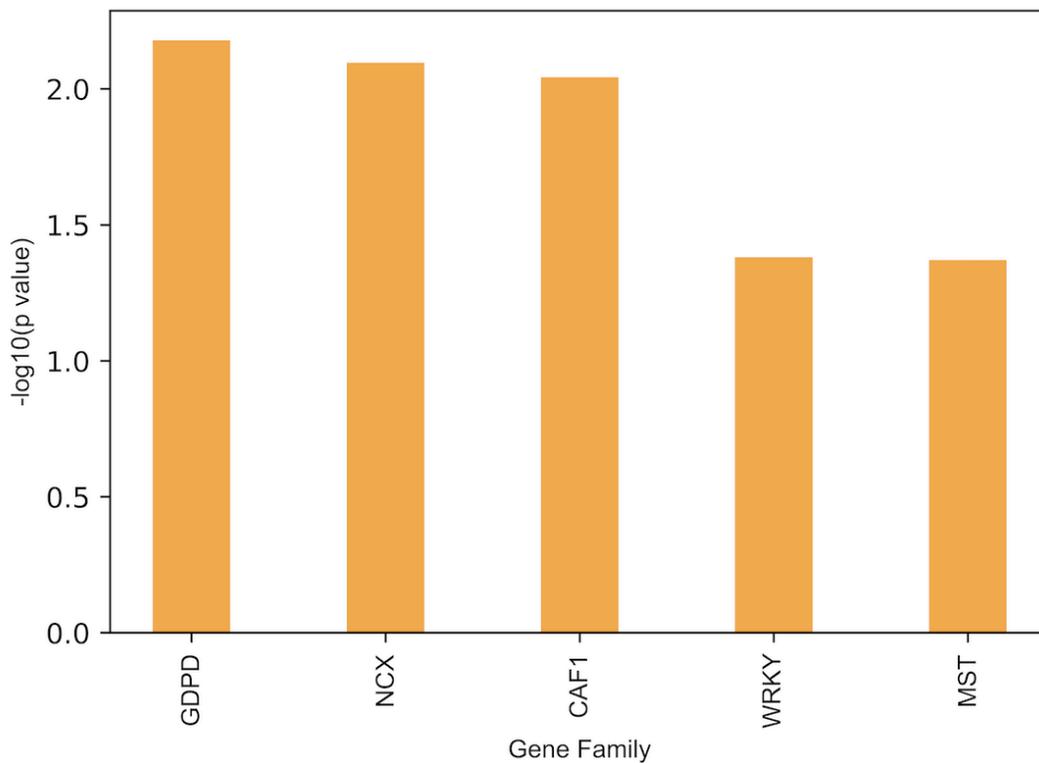
**Figura 3.17 Familias de genes expresadas negativamente de manera diferencial 72 hpi. Generado en GenFam**



**Figura 3.18 Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 72 hpi**

### 144 horas post infección

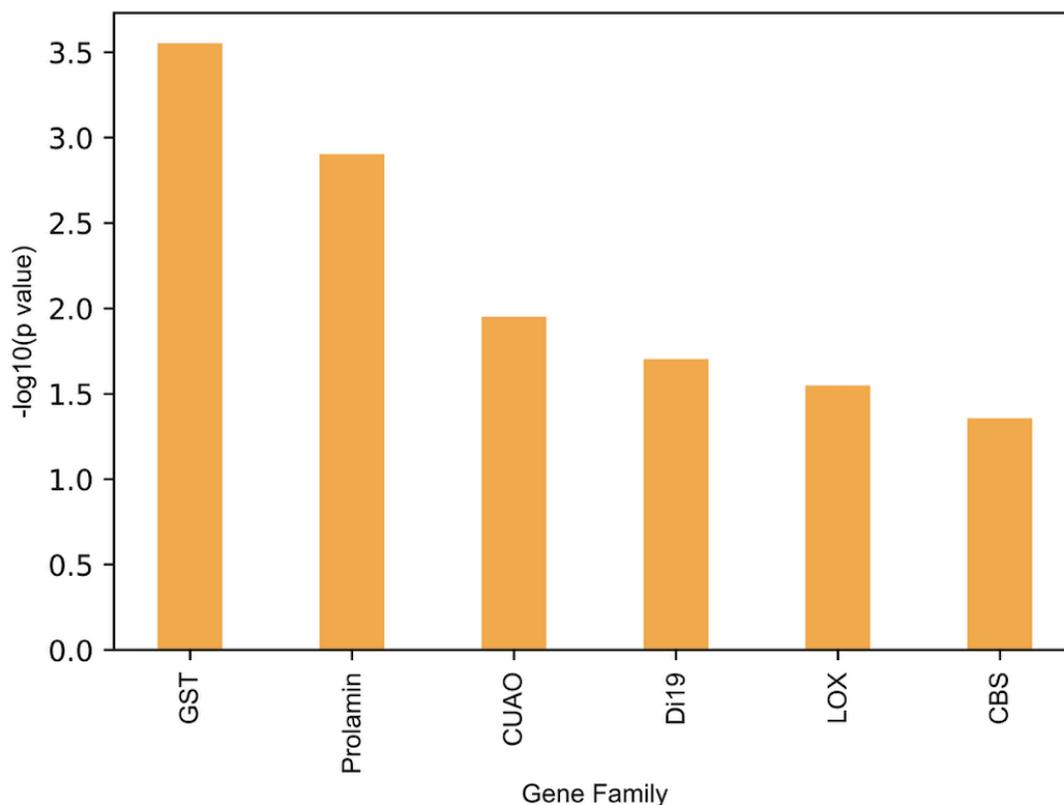
En esta etapa solo se analizó la expresión negativa, ya que hay una mayor cantidad de transcritos que puede ser procesada por los programas. La familia de GDPD se expresa con la misma frecuencia de la etapa anterior, lo cual corresponde además con la ruta metabólica significativa de glicerolípidos. Incluso se puede observar como la familia que le confería resistencia al estar regulada positivamente (WRKY) luego de 12 hpi, se expresa de manera negativa como el control, lo que indicaría que la planta está volviendo a un estado post infección.



**Figura 3.19 Familias de genes expresadas negativamente de manera diferencial luego de 144 hpi. Generado en GenFam**

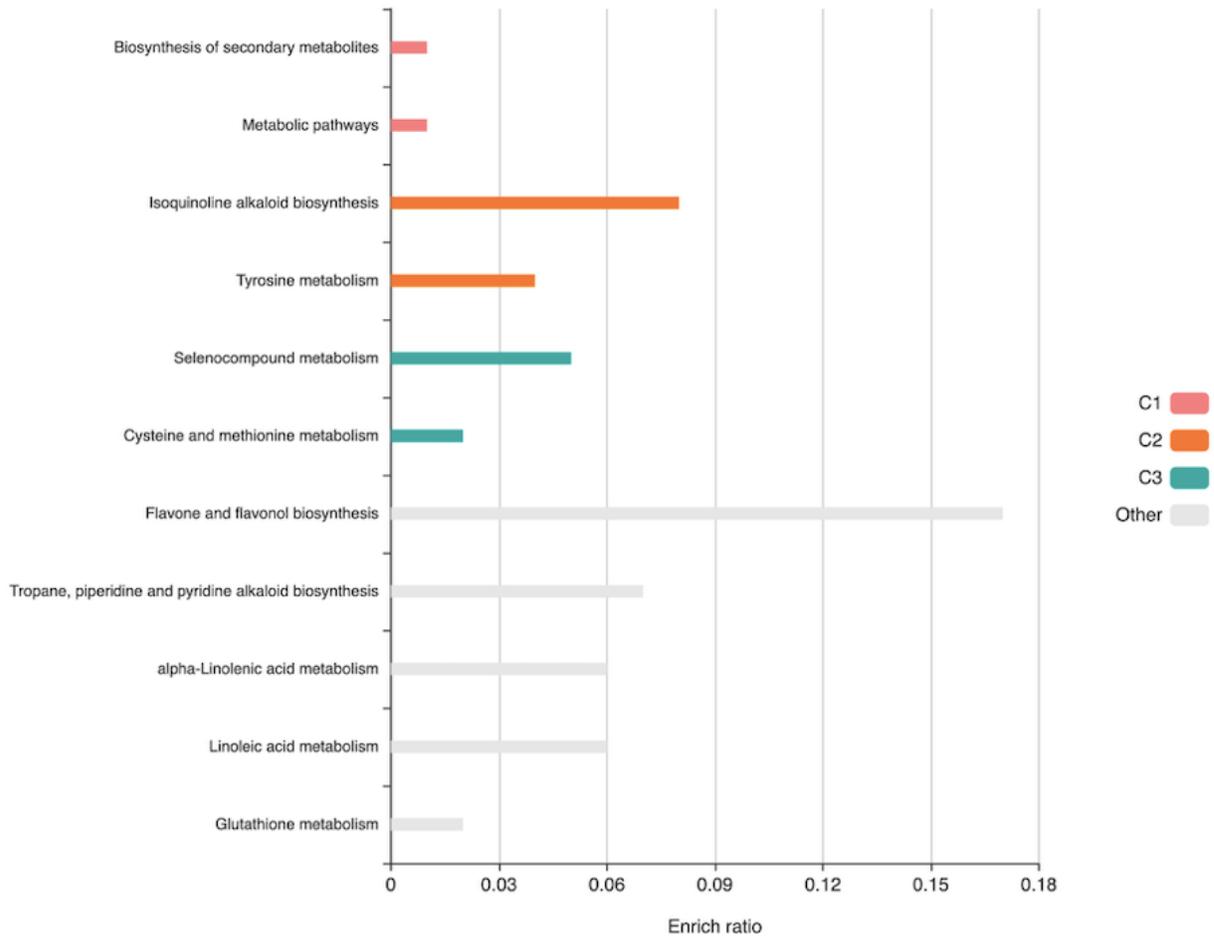
### 168 horas post infección

En la figura 3.20 se puede observar la familia GST que en contraste a la comparación entre controles, va a estar expresada diferencialmente de manera positiva y estar en una etapa de desintoxicación. Se han encontrado genes de esta familia en la especie *Lilium regale* que fue inducible por ácido salicílico y etileno con *Fusarium oxysporum* que tiene la característica de también ser hemibiotrófico que, además, indica que atenúa el estrés oxidativo (Gullner et al., 2018). La expresión de estos genes dependerá de la planta en la que se exprese porque puede generar tanto susceptibilidad como tolerancia. En el caso del banano se ha demostrado (Wang et al., 2013) que aumenta la expresión de estos genes cuando se encuentran infectados por patógenos fúngicos, tratando de mitigarla. Los genes involucrados en el análisis son LOC\_OS05G34150.1, LOC\_OS10G38710.1, LOC\_OS10G38140.1.



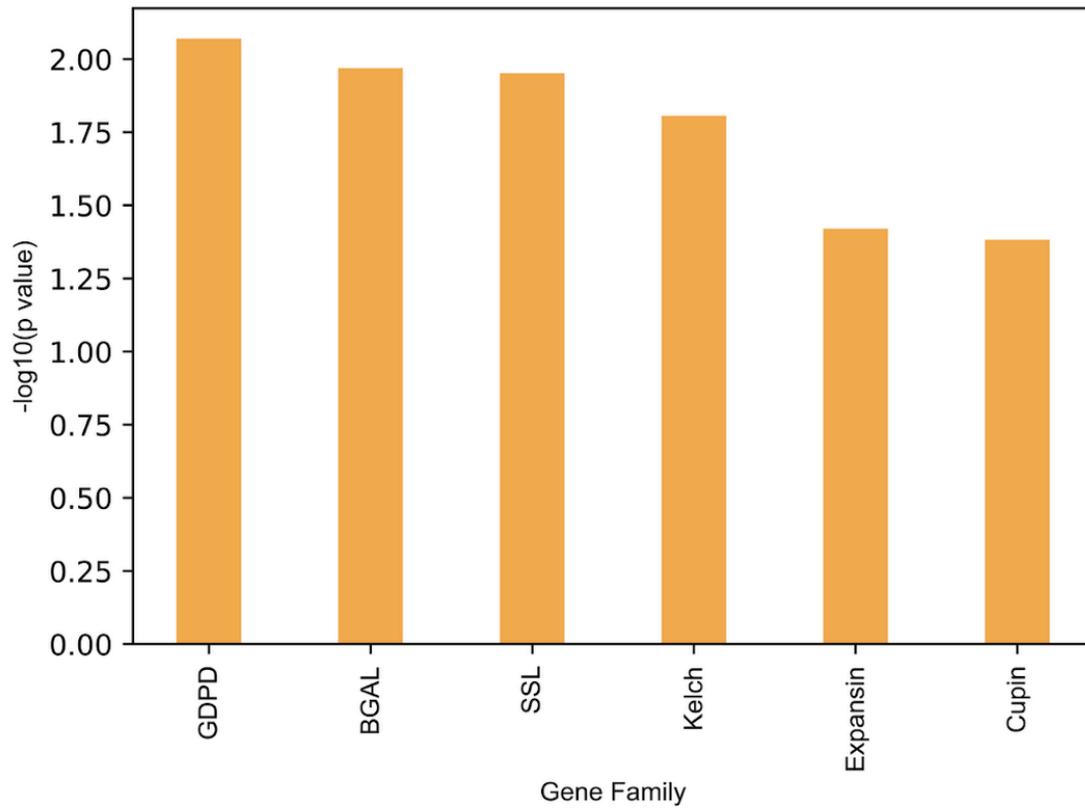
**Figura 3.20 Familias de genes expresadas positivamente de manera diferencial luego de 168 hpi. Generado en GenFam**

Las rutas metabólicas enriquecidas en esta etapa (figura 3.21) fue la de biosíntesis de metabolitos secundarios, así como el de flavonol que se ha mantenido en todos los análisis.



**Figura 3.21 Análisis de enriquecimiento de rutas metabólicas mediante KOBAS en la expresión diferencial positiva 72 hpi**

La familia GDPD fue una de las mayormente expresadas negativamente, la cual además se ve involucrada en la ruta metabólica de glicerofosfolípidos. Se conoce que tienen una actividad protectora, por lo que, al estar regulada negativamente, aumenta la susceptibilidad de la planta con el gen LOC\_OS02G31030.2.



**Figura 3.22 Familias de genes expresadas positivamente de manera diferencial luego de 168 hpi. Generado en GenFam**

### ***Lista de genes candidatos de resistencia y/o tolerancia***

Finalmente, con la información brindada en cada etapa de la infección se pueden discriminar los genes de resistencia y/o tolerancia según las familias de genes identificadas. Tomando como referencia información bibliográfica sobre estas familias genéticas y su involucramiento en la interacción planta-patógeno y especialmente en las respuestas de defensa de la planta.

**Tabla 3.1 Genes seleccionados como candidatos de resistencia y/o tolerancia con sus respectivas familias genéticas o rutas metabólicas.**

<b>Genes</b>	<b>Familia genética involucrada o metabolismo</b>
LOC_OS04G01740.1	HSP20
LOC_OS08G27850.1	MBF1
LOC_OS11G07020.1 LOC_OS05G33380.1 LOC_OS01G02880.1.	FBA
LOC_OS06G09390.2 LOC_OS02G43790.1 LOC_OS01G12440.1 LOC_OS04G46250.1 LOC_OS01G58420.1	AP2/EREBP
LOC_OS05G46020.1 LOC_OS05G25770.1	WRKY

LOC_Os01g28500.1	Ruta metabólica MAPK
LOC_Os01g61080.1	
LOC_Os03g20790.1	
LOC_Os06g05520.1	
LOC_Os06g51060.1	
LOC_Os10g42430.1	
LOC_OS01G59410.1	VQ
LOC_OS12G23280.4	BET
LOC_OS05G43670.1	BAG
LOC_OS06G15420.1	ASNS
LOC_OS05G34150.1	GST
LOC_OS10G38710.1	
LOC_OS10G38140.1	

# CAPÍTULO 4

## 4. CONCLUSIONES Y RECOMENDACIONES

### Conclusiones

- Los objetivos específicos se cumplieron para poder desarrollar la base de datos de genes de resistencia y/o tolerancia del banano ante la infección mediante la exploración de las diferencias en la expresión de la variedad Williams y Tuu Gia. La exploración se realizó mediante el uso de métodos y herramientas bioinformáticas.
- El análisis comparativo demostró que posee una mayor semejanza con el arroz que con *A. thaliana*, con un 99,63% de hits, lo que permitió descubrir una mayor cantidad de genes homólogos.
- Las comparaciones de las dos variedades en un mismo tiempo de infección resultaron en una cantidad de genes expresados diferencialmente en la variedad resistente. Algunas etapas, al no tener mucha diferencia, no se pudo obtener una cantidad suficiente para poder evaluarlas.
- Se encontró que la variedad resistente desde el primer momento, incluso antes de la infección, tiene una expresión diferencial con respecto a la susceptible. Esta expresión se verá modificada, por lo que en diferentes tiempos van a preponderar diferentes familias de genes en diferente proporción.
- Además, se encontraron 25 genes de tolerancia o resistencia. Estos ayudarán a que los métodos en programas de mejoramiento genético tengan un direccionamiento certero.
- Los genes asociados a tolerancia o resistencia sin infección se encuentran mayormente involucrados en rutas metabólicas de MAPK, y familias genéticas como HSP20 implicadas en producción de etileno, ácido abscísico y jasmonato. Estos últimos productos se encuentran fuertemente regulados en todos los tiempos post infección muestreados en el ensayo.
- Las familias genéticas y rutas metabólicas encontradas corresponden a la relación planta-patógeno según bibliografía de estudios similares.

- Es importante destacar que la lista de genes candidatos pueden actuar como resistentes o susceptibles, dependiendo del nivel de expresión genética. Por ello, estudiar su comportamiento más de cerca daría más pistas sobre la resistencia de la planta hacia patógenos.

### **Recomendaciones**

- El proyecto tuvo la limitante de utilizar solamente con un millón de lecturas en el archivo fastq ya que para trabajar con la totalidad de las lecturas se necesita de un gran poder informático de una super computadora o servidor con grandes capacidades. Es por esto por lo que los resultados obtenidos solo muestran parte de la expresión genética del ensayo. A su vez limita el análisis de enriquecimiento al haber muy pocos genes, lo que ocurrió en el análisis ontológico usando AgriGO.
- El ensayo biológico que se realizó no tuvo replicados en cada tiempo post infección. Es por esto por lo que en el análisis de expresión diferencial se realizó una estimación de la varianza. La carencia de estos datos pudo resultar en un falso hallazgo de genes, por lo que futuras investigaciones se deberían de tener al menos tres replicados biológicos.
- En el proyecto se consideró un solo control tomado luego de 6 y 12 horas post infección como uno solo. Lo óptimo para realizar un análisis transcriptómico es de tener controles por triplicado en cada etapa de la interacción planta-patógeno.
- Luego del hallazgo de los genes es importante verificar por RT-qPCR la presencia y el número de copias en variedades sanas e infectadas.
- Ampliar los muestreos en tiempos donde la infección *P. fijiensis* muestra síntomas visibles ya que los genes involucrados podrían dar nuevas pistas sobre la resistencia de sus genes.

# BIBLIOGRAFÍA

- Alakonya, A. E., Kimunye, J., Mahuku, G., Amah, D., Uwimana, B., Brown, A., & Swennen, R. (2018). Progress in understanding *Pseudocercospora* banana pathogens and the development of resistant *Musa* germplasm. *Plant Pathology*, 67(4), 759–770. <https://doi.org/10.1111/PPA.12824>
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3), 403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2)
- Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. 2010. En <http://www.bioinformatics.babraham.ac.uk/projects/>. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Apolo Aguilar, D. A., Cevallos Vite, H., & Carvajal Romero, H. (2021). Análisis de la producción bananera pre y pos pandemia de la “Asociación “Asocobaoro” periodo 2019-2020. *Revista Metropolitana de Ciencias Aplicadas*, 128–135. <https://orcid.org/0000-0001-8818-1042>
- Arango Isaza, R. E., Diaz-Trujillo, C., Dhillon, B., Aerts, A., Carlier, J., Crane, C. F., v. de Jong, T., de Vries, I., Dietrich, R., Farmer, A. D., Fortes Ferreira, C., Garcia, S., Guzman, M., Hamelin, R. C., Lindquist, E. A., Mehrabi, R., Quiros, O., Schmutz, J., Shapiro, H., ... Kema, G. H. J. (2016a). Combating a Global Threat to a Clonal Crop: Banana Black Sigatoka Pathogen *Pseudocercospora fijiensis* (Synonym *Mycosphaerella fijiensis*) Genomes Reveal Clues for Disease Control. *PLOS Genetics*, 12(8), e1005876. <https://doi.org/10.1371/JOURNAL.PGEN.1005876>
- Arango Isaza, R. E., Diaz-Trujillo, C., Dhillon, B., Aerts, A., Carlier, J., Crane, C. F., v. de Jong, T., de Vries, I., Dietrich, R., Farmer, A. D., Fortes Ferreira, C., Garcia, S., Guzman, M., Hamelin, R. C., Lindquist, E. A., Mehrabi, R., Quiros, O., Schmutz, J., Shapiro, H., ... Kema, G. H. J. (2016b). Combating a Global Threat to a Clonal Crop: Banana Black Sigatoka Pathogen *Pseudocercospora fijiensis* (Synonym

- Mycosphaerella fijiensis) Genomes Reveal Clues for Disease Control. *PLOS Genetics*, 12(8), e1005876. <https://doi.org/10.1371/JOURNAL.PGEN.1005876>
- Bebber, D. P. (2019). Climate change effects on Black Sigatoka disease of banana. *Philosophical Transactions of the Royal Society B*, 374(1775). <https://doi.org/10.1098/RSTB.2018.0269>
- Bedre, R. (2020). *reneshbedre/bioinfokit: Bioinformatics data analysis and visualization toolkit*. <https://doi.org/10.5281/ZENODO.3841708>
- Bedre, R., & Mandadi, K. (2019). GenFam: A web application and database for gene family-based classification and functional enrichment analysis. *Plant Direct*, 3(12), 1–7. <https://doi.org/10.1002/PLD3.191>
- Betancourt, M., Dita, M., Saini, E., & Salazar, L. (2020). *Agenda para la prevención y el manejo de brotes de la raza 4 tropical de Fusarium (R4T) en el cultivo de musáceas en América Latina y el Caribe (ALC)*. <https://doi.org/10.18235/0002346>
- Carbon, S., Douglass, E., Dunn, N., Good, B., Harris, N. L., Lewis, S. E., Mungall, C. J., Basu, S., Chisholm, R. L., Dodson, R. J., Hartline, E., Fey, P., Thomas, P. D., Albou, L. P., Ebert, D., Kesling, M. J., Mi, H., Muruganujan, A., Huang, X., ... Westerfield, M. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, 47(D1), D330–D338. <https://doi.org/10.1093/NAR/GKY1055>
- Castelan, F. P., Saraiva, L. A., Lange, F., de Lapeyre de Bellaire, L., Cordenunsi, B. R., & Chillet, M. (2012). Effects of Black Leaf Streak Disease and Sigatoka Disease on fruit quality and maturation process of bananas produced in the subtropical conditions of southern Brazil. *Crop Protection*, 35, 127–131. <https://doi.org/10.1016/J.CROPRO.2011.08.002>
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884–i890. <https://doi.org/10.1093/BIOINFORMATICS/BTY560>
- Churchill, A. C. L. (2011). *Mycosphaerella fijiensis*, the black leaf streak pathogen of banana: progress towards understanding pathogen biology and detection, disease development, and the challenges of control. *Molecular Plant Pathology*, 12(4), 307–328. <https://doi.org/10.1111/J.1364-3703.2010.00672.X>

- 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* 2016 17:1, 17(1), 1–19. <https://doi.org/10.1186/S13059-016-0881-8>
- Dainat, J. (2022, junio 7). *AGAT: Another GFF Analysis Toolkit to handle annotations in any GTF/GFF format*. Zenodo. <https://doi.org/10.5281/ZENODO.6621429>
- de Bellaire, L., Fouré, E., Abadie, C., & Carlier, J. (2010). *Black Leaf Streak Disease is challenging the banana industry*. 65(6), 327–342. <https://doi.org/10.1051/fruits/2010034>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., & Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1), 15. <https://doi.org/10.1093/BIOINFORMATICS/BTS635>
- FAO. (2020). *Sector Bananero Ecuatoriano*. <https://blogs.cedia.org.ec/obest/wp-content/uploads/sites/7/2020/08/Sector-bananero-ecuatoriano-final.pdf>
- Gullner, G., Komives, T., Király, L., & Schröder, P. (2018). Glutathione S-Transferase Enzymes in Plant-Pathogen Interactions. *Frontiers in Plant Science*, 9, 1836. <https://doi.org/10.3389/FPLS.2018.01836>
- Hamel, L. P., Nicole, M. C., Duplessis, S., & Ellis, B. E. (2012). Mitogen-activated protein kinase signaling in plant-interacting fungi: Distinct messages from conserved messengers. *Plant Cell*, 24(4), 1327–1351. <https://doi.org/10.1105/TPC.112.096156/DC1>
- Hu, S., & Bidochka, M. J. (2021). Abscisic acid implicated in differential plant responses of *Phaseolus vulgaris* during endophytic colonization by *Metarhizium* and pathogenic colonization by *Fusarium*. *Scientific Reports* 2021 11:1, 11(1), 1–12. <https://doi.org/10.1038/s41598-021-90232-4>
- Hwang, I. S., An, S. H., & Hwang, B. K. (2011). Pepper asparagine synthetase 1 (CaAS1) is required for plant nitrogen assimilation and defense responses to microbial pathogens. *The Plant Journal: For Cell and Molecular Biology*, 67(5), 749–762. <https://doi.org/10.1111/J.1365-313X.2011.04622.X>
- Israeli, Y., & Lahav, E. (2016). Banana. *Encyclopedia of Applied Plant Sciences*, 3, 363–381. <https://doi.org/10.1016/B978-0-12-394807-6.00072-1>

- Jing, Y., & Lin, R. (2015). The VQ Motif-Containing Protein Family of Plant-Specific Transcriptional Regulators. *Plant Physiology*, 169(1), 371–378. <https://doi.org/10.1104/PP.15.00788>
- Kabbage, M., Kessens, R., & Dickman, M. B. (2016). A plant Bcl-2-associated athanogene is proteolytically activated to confer fungal resistance. *Microbial Cell*, 3(5), 224. <https://doi.org/10.15698/MIC2016.05.501>
- Kimunye, J., Were, E., Swennen, R., Viljoen, A., & Mahuku, G. (2021). Sources of resistance to *Pseudocercospora fijiensis*, the cause of black Sigatoka in banana. *Plant Pathology*, 70(7), 1651–1664. <https://doi.org/10.1111/PPA.13408>
- Lee, J. H., Yun, H. S., & Kwon, C. (2012). Molecular Communications between Plant Heat Shock Responses and Disease Resistance. *Molecules and Cells*, 34(2), 109. <https://doi.org/10.1007/S10059-012-0121-3>
- Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics (Oxford, England)*, 30(7), 923–930. <https://doi.org/10.1093/BIOINFORMATICS/BTT656>
- Lievens, L., Pollier, J., Goossens, A., Beyaert, R., & Staal, J. (2017). Abscisic acid as pathogen effector and immune regulator. *Frontiers in Plant Science*, 8, 587. <https://doi.org/10.3389/FPLS.2017.00587/BIBTEX>
- Marguerat, S., & Bähler, J. (2009). RNA-seq: from technology to biology. *Cellular and Molecular Life Sciences* 2009 67:4, 67(4), 569–579. <https://doi.org/10.1007/S00018-009-0180-6>
- Mulder, J. L., & Stover, R. H. (1976). *Mycosphaerella* species causing banana leaf spot. *Transactions of the British Mycological Society*, 67(1), 77–82. [https://doi.org/10.1016/S0007-1536\(76\)80010-1](https://doi.org/10.1016/S0007-1536(76)80010-1)
- Noar, R. D., & Daub, M. E. (2016). Transcriptome sequencing of *Mycosphaerella fijiensis* during association with *Musa acuminata* reveals candidate pathogenicity genes. *BMC Genomics* 2016 17:1, 17(1), 1–25. <https://doi.org/10.1186/S12864-016-3031-5>
- Park, C. J., & Seo, Y. S. (2015). Heat Shock Proteins: A Review of the Molecular Chaperones for Plant Immunity. *The Plant Pathology Journal*, 31(4), 323. <https://doi.org/10.5423/PPJ.RW.08.2015.0150>

- Passos, M. A. N., de Oliveira Cruz, V., Emediato, F. L., de Camargo Teixeira, C., Souza, M. T., Matsumoto, T., Rennó Azevedo, V. C., Ferreira, C. F., Amorim, E. P., de Alencar Figueiredo, L. F., Martins, N. F., de Jesus Barbosa Cavalcante, M., Baurens, F.-C., da Silva, O. B., Pappas, G. J., Pignolet, L., Abadie, C., Ciampi, A. Y., Piffanelli, P., & Miller, R. N. G. (2012). Development of expressed sequence tag and expressed sequence tag–simple sequence repeat marker resources for *Musa acuminata*. *AoB PLANTS*, 2012. <https://doi.org/10.1093/aobpla/pls030>
- Portal, O., Izquierdo, Y., de Vleeschauwer, D., Sánchez-Rodríguez, A., Mendoza-Rodríguez, M., Acosta-Suárez, M., Ocaña, B., Jiménez, E., & Höfte, M. (2011). Analysis of expressed sequence tags derived from a compatible *Mycosphaerella fijiensis*-banana interaction. *Plant Cell Reports*, 30(5), 913–928. <https://doi.org/10.1007/S00299-011-1008-Z>
- ProMusa. (2020). *Williams | Improving the understanding of banana*. <https://www.promusa.org/Williams>
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139. <https://doi.org/10.1093/BIOINFORMATICS/BTP616>
- Rodriguez, H. A., Rodriguez-Arango, E., Morales, J. G., Kema, G., & Arango, R. E. (2016). Defense gene expression associated with biotrophic phase of *Mycosphaerella fijiensis* M. Morelet infection in banana. *Plant Disease*, 100(6), 1170–1175. [https://doi.org/10.1094/PDIS-08-15-0950-RE/ASSET/IMAGES/LARGE/PDIS-08-15-0950-RE\\_F4.JPEG](https://doi.org/10.1094/PDIS-08-15-0950-RE/ASSET/IMAGES/LARGE/PDIS-08-15-0950-RE_F4.JPEG)
- Sánchez Timm, E., Hidalgo Pardo, L., Pacheco Coello, R., Chávez Navarrete, T., Navarrete Villegas, O., & Santos Ordóñez, E. (2016). Identification of Differentially-Expressed Genes in Response to *Mycosphaerella fijiensis* in the Resistant *Musa* Accession ‘Calcutta-4’ Using Suppression Subtractive Hybridization. *PLOS ONE*, 11(8), e0160083. <https://doi.org/10.1371/JOURNAL.PONE.0160083>
- Sharoni, A. M., Nuruzzaman, M., Satoh, K., Shimizu, T., Kondoh, H., Sasaya, T., Choi, I. R., Omura, T., & Kikuchi, S. (2011). Gene Structures, Classification and Expression Models of the AP2/EREBP Transcription Factor Family in Rice. *Plant and Cell Physiology*, 52(2), 344–360. <https://doi.org/10.1093/PCP/PCQ196>

- Singh, B., Singh, J. P., Kaur, A., & Singh, N. (2016). Bioactive compounds in banana and their associated health benefits – A review. *Food Chemistry*, 206, 1–11. <https://doi.org/10.1016/J.FOODCHEM.2016.03.033>
- Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics* 2019 20:11, 20(11), 631–656. <https://doi.org/10.1038/s41576-019-0150-2>
- Swarupa, V., Ravishankar, K. v., & Rekha, A. (2013). Characterization of tolerance to *Fusarium oxysporum* f.sp., cubense infection in banana using suppression subtractive hybridization and gene expression analysis. *Physiological and Molecular Plant Pathology*, 83, 1–7. <https://doi.org/10.1016/J.PMPP.2013.02.003>
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., Xu, W., & Su, Z. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community, 2017 update. *Nucleic Acids Research*, 45(W1), W122–W129. <https://doi.org/10.1093/NAR/GKX382>
- Voora, V., Larrea, C., & Bermudez, S. (2020). *Global Market Report: Bananas SUSTAINABLE COMMODITIES MARKETPLACE SERIES 2019*. <https://www.iisd.org/system/files/publications/ssi-global-market-report-banana.pdf>
- Wang, Z., Huang, S., Jia, C., Liu, J., Zhang, J., Xu, B., & Jin, Z. (2013). Molecular cloning and expression of five glutathione S-transferase (GST) genes from Banana (*Musa acuminata* L. AAA group, cv. Cavendish). *Plant Cell Reports*, 32(9), 1373–1380. <https://doi.org/10.1007/S00299-013-1449-7>
- Xie, C., Mao, X., Huang, J., Ding, Y., Wu, J., Dong, S., Kong, L., Gao, G., Li, C. Y., & Wei, L. (2011). KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. *Nucleic Acids Research*, 39(suppl\_2), W316–W322. <https://doi.org/10.1093/NAR/GKR483>
- Zhang, X., Xu, Z., Chen, L., & Ren, Z. (2019). Comprehensive analysis of multiprotein bridging factor 1 family genes and SIMBF1c negatively regulate the resistance to *Botrytis cinerea* in tomato. *BMC Plant Biology*, 19(1). <https://doi.org/10.1186/S12870-019-2029-Y>



# APÉNDICES

## 1. Limpieza de secuencias con fastp y fastqc

```
conda install -c bioconda fastp
```

```
for x in *.fastq;
```

```
do
```

```
fastp -i $x -o "out-$x" -h "out-h-$x.html" -j "out-j-$x.json";
```

```
done
```

```
conda install -c bioconda fastqc
```

```
fastqc -o /Volumes/MelissaCRP/Transcriptomics-RNA-seq/fastQC-filt *
```

## 2. Alineamiento con STAR

```
conda install -c bioconda star
```

```
conda install -c bioconda agat
```

```
#Convertir el archivo de anotaciones de gff a gtf
```

```
agat_convert_sp_gff2gtf.pl -gff Musa_acuminata_pahang_v4.gff3 \
```

```
-o Musa_acuminata_pahang_v4.gtf
```

```

#Creación del índice

STAR --runThreadN 64 --runMode genomeGenerate --genomeDir index \

--genomeFastaFiles /home/melissarp_mr/hisat/Musa_acuminata_pahang_v4.fasta \

--sjdbGTFfile /home/melissarp_mr/hisat/Musa_acuminata_pahang_v4.gtf

#Alineamiento

for x in *.fastq;

do

STAR --runThreadN 32 --genomeDir /home/melissarp_mr/index \

--outFileNamePrefix $x \

--readFilesIn $x;

done

```

### 3. Cuantificación de genes con featureCounts

```

conda install -c bioconda subread

mkdir counts

featureCounts -T 4 -s 2 -a

/home/melissarp_mr/hisat/Musa_acuminata_pahang_v4.gtf \

-o /home/melissarp_mr/counts/banana_featurecounts.txt \

```

```
/home/melissarp_mr/mapping/sam-file-map/*sam
```

```
#Remover la primera columna
```

```
sed '1d' banana_featurecounts.txt > tmpfile; mv tmpfile  
banana_featurecounts.txt
```

```
#Seleccionar la primera y las últimas doce columnas
```

```
awk '{print $1,$7,$8,$9,$10,$11,$12,$13,$14,$15,$16,$17,$18}' \
```

```
banana_featurecounts.txt > clean_banana_featurecounts.txt
```

#### 4. Análisis de expression diferencial en R

```
if (!requireNamespace("BiocManager", quietly = TRUE))
```

```
  install.packages("BiocManager")
```

```
BiocManager::install(version = "3.12")
```

```
if (!requireNamespace("BiocManager", quietly = TRUE))
```

```
  install.packages("BiocManager")
```

```
BiocManager::install()
```

```
BiocManager::install("edgeR")
```

```
BiocManager::install("limma")
```

```

install.packages("statmod")

library(limma)

library(edgeR)

library(gplots)

getwd()

#Información común

matrixDGA <- read.delim(file =
"/Users/melissa/Desktop/Transcriptomics&Bioinformatics |
Thesis/counts/clean_banana_featurecounts.txt", header = TRUE,
                        sep=" ")

Cultivar <- factor(c("T","T","T","T","T","T","W","W","W","W","W","W"))

tpi <-
factor(c("6h","12h","24h","72h","144h","168h","6h","12h","24h","72h","144h
","168h")) #Control es 6-12h con mock

colnames(matrixDGA) = c("Gene ID","CT6-
12h","T12h","T24h","T3d","T6d","T7d","CW6-
12h","W12h","W24h","W3d","W6d","W7d")

types <- factor(c("CT6-12h","T12h","T24h","T3d","T6d","T7d","CW6-
12h","W12h","W24h","W3d","W6d","W7d"))

#teniendo en cuenta que no hay réplicas

y <- DGEList(counts=matrixDGA[,2:13], genes=matrixDGA[,1],group = types)

```

```

#crear matriz de condiciones

targets <- data.frame(Sample = colnames(y),Cultivar,tpi)

Group <- factor(paste(targets$Cultivar,targets$tpi,sep="."))

cbind(targets,Group=Group)

keep <- filterByExpr(y)

table(keep)

y <- y[keep, , keep.lib.sizes=FALSE]

#filtrar los que no tienen un nivel significativo de expresión

y <- calcNormFactors(y)

y$samples

plotMDS(y) #data exploration

#create a design matrix

design <- model.matrix(~0+Group)

colnames(design) <- levels(Group)

#Asignación un valor de variación biológica ya que no hay réplicas

bcv <- 0.2

fit <- glmFit(y, design, dispersion = bcv^2)

```

```

#likelihood

lrt <- glmLRT(fit,coef=2)

topTags(lrt)

my.contrasts <- makeContrasts(

    T.12vs6 = T.12h-T.6h,

    T.24vs6 = T.24h-T.6h,

    T.72vs6 = T.72h-T.6h,

    T.144vs6 = T.144h-T.6h,

    T.168vs6 = T.168h-T.6h,

    W.12vs6 = W.12h-W.6h,

    W.24vs6 = W.24h-W.6h,

    W.72vs6 = W.72h-W.6h,

    W.144vs6 = W.144h-W.6h,

    W.168vs6 = W.168h-W.6h,

    TvsW.6h = T.6h-W.6h,

    TvsW.12h = (T.12h-T.6h)-(W.12h-W.6h),

    TvsW.24h = (T.24h-T.6h)-(W.24h-W.6h),

    TvsW.72h = (T.72h-T.6h)-(W.72h-W.6h),

    TvsW.144h = (T.144h-T.6h)-(W.144h-W.6h),

    TvsW.168h = (T.168h-T.6h)-(W.168h-W.6h),

    levels=design)

```

```

fit.col.names <- colnames(fit)

fit.col.names[1:6]

keep <- out1$table$FDR < 0.05 & abs(out1$table$logFC) > 1

filt <- out1[keep,]

#genes respondiendo diferente ante la infección en Tuu gia vs the control

qlf1 <- glmLRT(fit, contrast=my.contrasts[,"T.12vs6"])

plot1 <-plotMD(qlf1)

out1 <- topTags(qlf1)

summary(decideTests(qlf1))

#Por cada tratamiento se realiza

qlf <- glmLRT(fit, contrast=my.contrasts[,"T.24vs6"])

plotMD(qlf)

topTags(qlf)

summary(decideTests(qlf))

```

## 5. Extraer CDS de genes expresados diferencialmente con bioinfokit

```
import bioinfokit
```

```
from bioinfokit.analys import Fasta
```

```
#Dentro de la carpeta donde se encuentran las listas de genes
```

```

import os

list = os.listdir() #retorna lista de direcotrio

genes = [item for item in list if item.startswith('genes')]

for gene in genes:

    bioinfokit.analys.Fasta.extract_seq(file =
"Musa_acuminata_pahang_v4.cds.fa" , id = gene)

    os.rename("output.fasta", "CDS"+str(gene[8:-4])+".fasta")

```

## 6. Análisis comparativo

### a. Blastx

```
conda install -c bioconda blast
```

```
makeblastdb -in Athaliana_447_Araport11.cds_primaryTranscriptOnly.fa -dbtype
nucl
```

```
cd /Users/melissa/Desktop/DGE-Analysis/Names-ID
```

```
makeblastdb -in RicePept2011.pep -dbtype prot -parse_seqids -out
ArabidpsisProt
```

```

for x in CDS*;
do
blastx -db /Users/melissa/Desktop/DGE-Analysis/Protein/Database/RiceProt -
query $x -max_target_seqs 1 -outfmt "10 qseqid sseqid" -out "prot-$x"
mv "prot-$x" /Users/melissa/Desktop/DGE-Analysis/Protein/Database/Pept-DEG;
done

```

## b. GeneFam:

```
import os

import bioinfokit

from bioinfokit.analys import genfam

res = genfam()

list_negs = os.listdir() #returns list of things in directory

arab_genes = [item for item in list_negs if item.startswith('x-ARAB')]

fail = 'FailGenFam.txt'

for file in arab_genes:

    try:

        res.fam_enrich(id_file = file, species = 'atha', id_type = 2)

        os.rename("fam_all_out.txt", "fam_all_"+str(file[10:-11])+".txt")

        os.rename("fam_enrich_out.txt", "fam_enrich_"+str(file[10:-11])+".txt")

        os.rename("genfam_enrich.png", str(file[10:-11])+".png")

    except Exception:

        with open(fail, 'a') as x:
```

```
x.write(str(file))
```

```
x.write('\n')
```

```
x.close
```

