

Cassina, M.H.

Facultad de Agronomía, Universidad de Buenos Aires, Av. San Martín 4453, C1417DSE. Ciudad Autónoma de Buenos Aires, Argentina

E-mail: mcassina@agro.uba.ar

INTRODUCCIÓN

La podredumbre carbonosa es una de las enfermedades cuya presencia ha tenido un sostenido aumento en el país, principalmente en ambientes cálidos y secos. Las estrategias de manejo incluyen métodos culturales y fungicidas, pero han proporcionado un control limitado (Mengistu et al., 2015). En este escenario, la resistencia genética puede ser el método más factible y sostenible para manejar la enfermedad. Con el uso de herramientas biotecnológicas es posible comprender la respuesta vegetal, las vías metabólicas y genes asociados, para posteriormente desarrollar marcadores moleculares que guíen el mejoramiento hacia la obtención de variedades con mejor desempeño ante dicho patógeno. Como primer paso resulta indispensable la caracterización de la respuesta frente a la enfermedad en genotipos de comportamiento contrastante frente a dicha adversidad. Posteriormente, es indispensable la secuenciación y un análisis transcriptómico que permite dilucidar cuáles son los genes y las rutas metabólicas responsables del diferente comportamiento de los genotipos observado previamente.

MATERIALES Y MÉTODOS

1. Ensayo de selección de genotipos para el estudio RNA-seq

Contemplando el desempeño histórico de un panel de genotipos iniciales en situación de campo y las descripciones bibliográficas, se eligieron los genotipos D86 -PI518772- considerado como resistente, al igual que Toyoshirome (GR) -PI 594302- (Mengistu et al., 2013) y por otra parte los genotipos DM4800 y Williams 82, considerados como genotipos susceptibles (Pearson et al., 1984), (EEAOC., 2014). El objetivo de dicho ensayo es la caracterización a partir de variables epidemiológicas, para poder dilucidar genotipos de comportamiento contrastante frente a la adversidad.

Condiciones del ensayo

El ensayo se realizó en invernáculo de la Facultad de Agronomía. Se utilizaron macetas de 3l de capacidad, y se llenaron con sustrato Growmix (turba de musgo,

spragnum de fibras medias, perlita, compost de corteza, cal calcita, caldolomita, agentes humectantes, fertilizante). El sustrato se esterilizó en autoclave a 121°C durante 1 hora. Se cultivó el patógeno in vitro en medio agar papa glucosado y se inoculó vía la colocación del micelio directamente en el sustrato (Cassina et al., 2017). Se utilizaron semillas con PG del 95%. Previo a la siembra se calculó la capacidad de retención hídrica, resultante de la diferencia entre el peso a capacidad de campo (CC) y el peso seco. Para calcular el peso a CC se regaron las macetas a saturación y luego se dejaron secar durante la noche. El peso seco se registró luego de dejar deshidratar el sustrato hasta lograr peso constante en estufa (80°C). El tratamiento diferencial de riego se aplicó a partir de V1. El criterio fue el del rango de CC fijo (Mian et al., 1996). Para los tratamientos con sequía, se mantuvo la condición hídrica en 20%, mientras que los tratamientos control un 80%. Los tratamientos ejecutados combinan ambos factores: presencia de inóculo y sequía (H+S+), presencia inóculo y situación hídrica en capacidad de campo (H+S-), sin presencia inóculo y sequía (H-S+) y tratamiento control – sin sequía ni inóculo (H-S-). El diseño experimental consistió en un DCA y se realizaron 5 repeticiones.

2. Estudio RNA-seq

A partir de los genotipos contrastantes obtenidos del ensayo (1), se procedió a realizar la secuenciación y el RNA-seq. Se eligieron los tejidos radicales dado que son el principal órgano de afectación del patógeno.

Condiciones del estudio

Las plántulas se cosecharon a los 21 días desde V1, donde el ensayo preliminar permitió obtener las diferencias significativas entre los genotipos. Se mantuvieron los tejidos en N líquido y posteriormente a freezer -80°C para su conservación. La extracción del ARN se llevó a cabo con el kit Qiagen (RNeasy Plant Mini Kit) y se controló que cumpla con las condiciones de masa, pureza e integridad requeridas para la secuenciación. Se realizó en los laboratorios de Novogene en Sacramento, California. Para la secuenciación, en primer lugar, se fragmentó químicamente el ARN para luego sintetizar el ADNc doble hebra. Luego se realizó la reparación y adenilación de los

extremos. Por último, se ligó los adaptadores y se procedió con el enriquecimiento de los fragmentos de la biblioteca de acuerdo con el protocolo descrito por en TruSeq® ARN Sample Preparation Guide Illumina. Una vez que sea realiza la secuenciación, se obtiene lo que se conoce como información cruda (raw data) a partir de la cual se realiza el control de la calidad de la información y el posterior análisis bioinformático. Posteriormente, se realiza el mapeado al genoma de soja y la cuantificación del nivel de expresión génica. A partir de allí, es posible la realización de un análisis de expresión diferencial (DEGs) y su enriquecimiento funcional a través de la utilización de mapas metabólicos de procesos puntuales.

Evaluaciones realizadas

1. Ensayo de selección de genotipos para el estudio RNA-seq

Severidad y tasa de infección: A partir de los 0,7,14 y 21 días posteriores a V1, la severidad se midió utilizando escala (Abawi y Pastor Corrales, 1990) (**Tabla 1**)

. La severidad de *Macrophomina phaseolina* fue calculada asignándole valores de 0% al 100% dentro de la escala propuesta. El valor final para cada tratamiento resultó del promedio de los valores estimados para cada una de las cinco plántulas (**Figura 1**).



Figura 1. Valores registrados de la escala de Pastor Corrales y Abawi, 1990, desde el de menor severidad (1) hasta el valor máximo registrado (7).

A partir de los datos obtenidos de severidad, fue posible obtener la tasa de infección (r) de la enfermedad. Su importancia radica en que representa el cambio en los niveles de enfermedad por unidad de tiempo. Asimismo, representa la pendiente que se obtiene al linealizar los datos de progreso de la enfermedad (en este caso la severidad), mediante transformaciones logarítmicas.

Tabla 1. Descripción de la escala propuesta por Abawi y Pastor – Corrales, 1990

INFECCIONES EN <i>MACROPHOMINA</i> (Abawi y Pastor Corrales, 1990)
(1) No hay síntomas visibles
(3) Las lesiones se encuentran limitadas a tejidos cotiledonares y/o hay una pequeña decoloración en la raíz (hasta un 10% de lesiones en raíces e hipocótilo), pero sin evidenciar necrosis
(5) Las lesiones han progresado desde los cotiledones hasta 2cm de los tejidos corticales o del tallo y/o hay pequeños daños a nivel radical (hasta 25% de lesiones en raíces e hipocótilo), pero los tejidos radicales permanecen firmes.
(7) Las lesiones son extensivas al tallo y las ramas. El follaje exhibe necrosis y clorosis y/o el sistema radical sufre daños considerables (hasta 50% de lesiones en raíces e hipocótilo). Se pueden observar fructificaciones del hongo.
(9) La mayoría de los tallos, peciolo y puntos de crecimiento están infectados y/o hay un avanzado deterioro del sistema radical (aproximado 75% de lesiones en raíces e hipocótilo). Una considerable cantidad de picnidios y microesclerocios se han producido.
Nota: Los valores (2; 4; 6; 8 y 10) corresponden a estadios visuales intermedios y quedan a criterio del observador.

ABCPE: Por otra parte, se procedió al cálculo del ABCPE (área bajo la curva del progreso de la enfermedad). El área bajo la curva de progreso de la enfermedad representa la cantidad acumulada de enfermedad. Una de las mayores utilidades del cálculo de ABCPE radica en la posibilidad de efectuar comparaciones entre cultivares. A modo de ejemplo, un cultivar más susceptible dará un

ABCPE más alto y este análisis de ABCPE es una manera de justificarlo. Como fue explicado anteriormente, esta estrategia sirve para poder comparar entre diferentes epifitias o dentro de una misma, en diferentes situaciones.

2. Estudio RNA-seq y enriquecimiento funcional

DEGs: se contabilizan la cantidad de genes sobre y subexpresados entre el genotipo tolerante y el susceptible en condiciones de estrés hídrico e infección simultánea (H+S+)

Genes enriquecidos: A partir de las bibliotecas conocidas de GO y KEGG se analizan los genes resultantes diferencialmente expresados (DEGs). Posteriormente, se realiza el análisis de enriquecimiento, que consiste en un método para identificar clases de genes o proteínas que están sobrerrepresentadas en un gran conjunto de genes y proteínas, y pueden tener una asociación con rasgos fenotípicos. El método utiliza enfoques estadísticos para identificar grupos de genes significativamente enriquecidos.

RESULTADOS

1. Ensayo de selección de genotipos para el estudio RNA-seq

Severidad y tasa de infección: hubo diferencia a los 21 días desde la siembra. El genotipo D86 resultó ser el de menor severidad, tanto en sequía como en CC. Los genotipos GR y Williams tuvieron desempeños regulares dependiendo el tratamiento. DM 4800 fue el de peor comportamiento. El mismo patrón de comportamiento se registró para la tasa de infección (r) (**Figuras 2 y 3**).

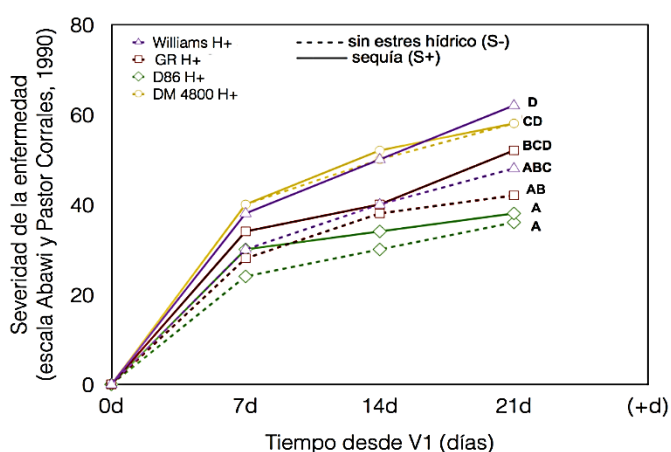


Figura 2. Severidad de la enfermedad según escala (Abawi y Pastor Corrales, 1990) desde los 0 a los 21 días desde V1, para cada uno de los genotipos.

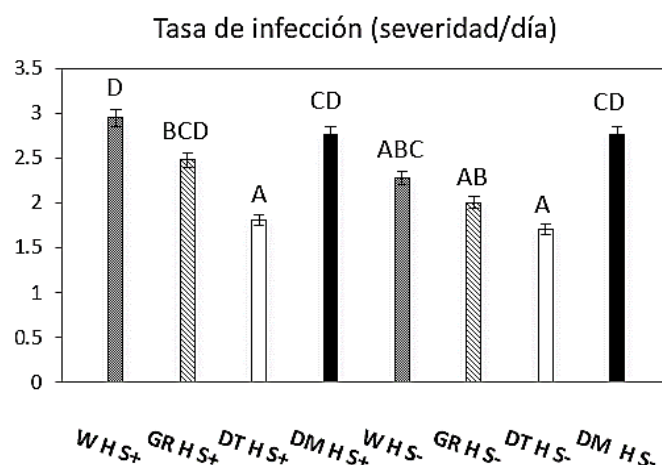


Figura 3. Tasa de infección obtenidas a los 21 días desde V1

ABCPE: El genotipo D86 resultó ser el de menor severidad, tanto en sequía como en CC. Los genotipos GR y Williams tuvieron desempeños regulares dependiendo el tratamiento. DM 4800 fue el de peor comportamiento (**Figura 4**). Se repite el mismo patrón observado con la severidad y la tasa de infección.

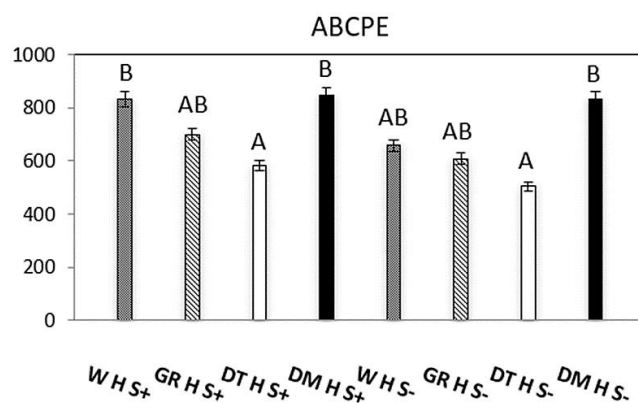


Figura 4- ABCPE a los 21 días desde V1 para cada uno de los tratamientos

2. Estudio RNA-seq y enriquecimiento funcional

DEGs: Análisis de expresión diferencial entre genotipos contrastantes.

El recuento de lecturas obtenido del análisis de expresión génica se utiliza para el análisis de expresión diferencial. El diseño experimental contó de ambos genotipos (DM y DT(D86)) en los 4 tratamientos descritos anteriormente en el primer ensayo (condiciones combinadas de factor hídrico e inóculo). Al tratarse de muestras con repeticiones biológicas, se realizó un análisis de expresión diferencial de dos condiciones /grupos utilizando el paquete DESeq2 R (Anders *et al.*, 2010).

A partir del uso de dicho paquete se pudo obtener el listado de todos los genes que se expresan diferencialmente en la comparación mencionada anteriormente en evaluaciones (DT_HS vs DM_HS). El

análisis permitió dilucidar un total de 686 genes que cambiaron su expresión, siendo 373 los genes que aumentaron su expresión en DT_HS con respecto a DM_HS, y 313 disminuyeron su expresión.

Contemplando la totalidad de los genes de dicha lista, se construyó un gráfico de volcán (Figura 5), donde se puede ver claramente la distribución de todos los 686 genes (en rojo, aquellos que aumentaron su expresión, y en verde los que disminuyeron su expresión). Dicho gráfico muestra en simultáneo la significancia de cada uno de los genes (a través del logaritmo negativo base 10 del valor p ajustado), con la magnitud de la relación cambio de expresión de cada uno de los genes en el genotipo DT con respecto a DM (expresado a través del logaritmo base 2 del fold change).

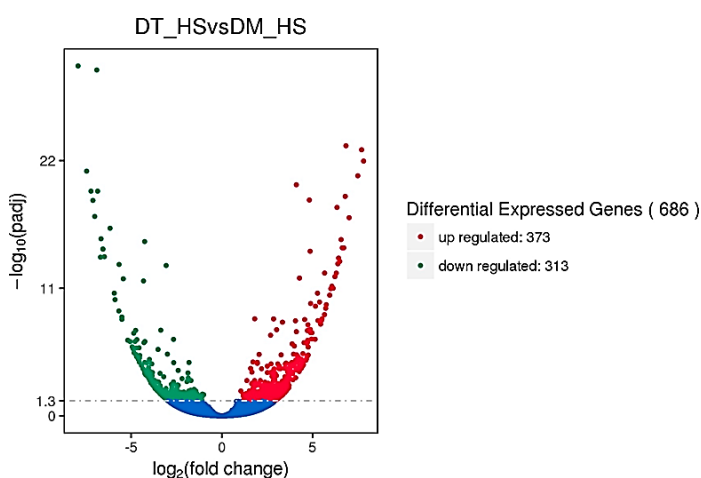


Figura 5. Volcanoplot de la expresión diferencial de genes del genotipo tolerante DT, con respecto al genotipo susceptible DM, en condiciones de infección y sequía simultánea.

Hay que tener presente que el valor clave descripto en el gráfico de volcán es el umbral de significancia, representado con el $-\log_{10}(\text{padj}) = 1,3$. Este valor surge de considerar el valor p ajustado de 0,05 que es ampliamente utilizado como valor de significancia en los análisis estadísticos.

En este sentido, el conjunto de puntos superpuestos en azul representa todo el conjunto de genes que no resultaron significativos, y por ende no están incluidos en los 686 genes totales. Asimismo, se puede observar en rojo los 373 genes que aumentaron su expresión en el genotipo tolerante DT con respecto al genotipo DM. Por otra parte, los puntos verdes representan los 313 genes que disminuyeron su expresión en DT con respecto a DM. Como se puede observar, el análisis de expresión diferencial cobra una gran relevancia dado que permite filtrar el comportamiento de la totalidad de genes presentes, aportando un panorama más simplificado acerca de qué genes cambian su expresión (aumentando o disminuyendo) en la comparación analizada. No

obstante, el paso posterior corresponde al análisis de enriquecimiento, que permite extraer mayores conclusiones a partir de vincular la expresión de los genes a categorías funcionales y vías metabólicas.

3. Análisis de enriquecimiento (GO/KEGG)

El análisis de enriquecimiento consiste en un método para identificar clases de genes o proteínas que están sobrerrepresentadas en un gran conjunto de genes y proteínas, y pueden tener una asociación con rasgos fenotípicos.

Análisis de Enriquecimiento utilizando la herramienta Gene Ontology (GO)

A través de gene ontology (GO), utilizando los diferentes genes sobre (+) y sub- expresados (-), se los puede asociar a diferentes categorías GO, agrupados en 3 ramas principales: función molecular (MF), componente celular (CC) y procesos biológicos (BP). La siguiente tabla 2 resume las categorías GO que resultaron significativas, es decir con un $\text{padj} < 0,05$, junto con el número de genes involucrados en dichas categorías (Tabla 2).

Tabla 2. Categorías enriquecidas en genes sobre expresados (GO).

CATEGORIAS ENRIQUECIDAS GENES SOBREENEXPRESADOS (DT_HS vs DM_HS)				
Categoría	GO - ID	Descripción	padj	Cantidad genes
BP	GO:0006457	protein folding	4,68E-07	10
MF	GO:0051082	unfolded protein binding	5,95E-06	10
MF	GO:0045735	nutrient reservoir activity	0.00637	5

CATEGORIAS ENRIQUECIDAS EN GENES SUBEXPRESADOS (DT_HS vs DM_HS)				
Categoría	GO - ID	Descripción	padj	Cantidad de genes
CC	GO:0071944	cell periphery	0.02794	5
CC	GO:0005618	cell wall	0.02794	3
CC	GO:0048046	apoplast	0.02794	3
MF	GO:0043531	ADP binding	7,97E-08	13

Como se puede observar, en el genotipo tolerante DT, considerando la rama de MF, hay una sobreexpresión de un grupo de 10 genes asociados al término GO de unión de proteínas no plegadas. Asimismo, dentro de esta interpretación se suma el grupo de 5 genes representados dentro de la categoría de actividad asociada a reserva de nutrientes. Como contrapartida, se puede ver un grupo de 13 genes sub-expresados asociados a la categoría unión al ADP. Cuando se analiza el CC, se puede observar únicamente una subexpresión génica en el genotipo tolerante DT compuesta por pocos genes (3-5), asociados a todas categorías GO muy similares entre sí, como lo son los de pared celular, apoplasto y periferia celular. Por último, en la rama BP, podemos observar un grupo de 10 genes sobre expresados en el genotipo tolerante, asociados a la categoría GO de plegado de proteínas. Los genes resultantes del análisis totales y únicos son 33: los genes de periferia celular, pared celular

y apoplasto son los mismos (5), lo mismo ocurre con los genes asociados al plegado de proteínas y la unión de proteínas no plegadas (10), sumado a la actividad de reserva de nutrientes (5) y la unión al ADP (13).

La **Figura 6** muestra todas las categorías resultantes de los genes sub- expresados en el genotipo tolerante con respecto al susceptible.

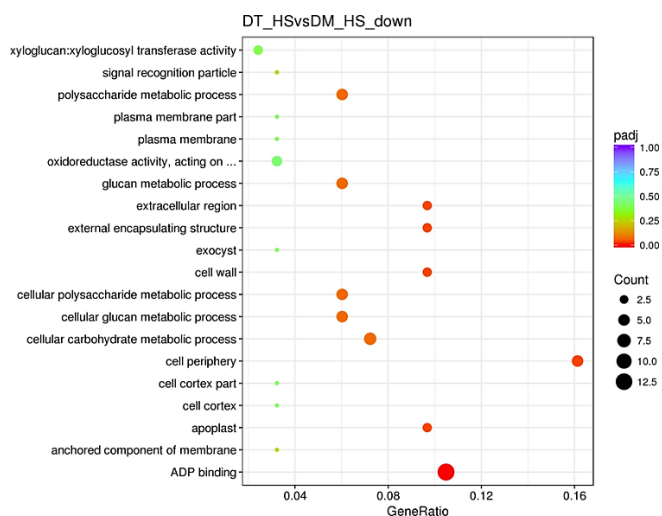


Figura 6. Categorías de la plataforma GO asociados al conjunto de genes sub- expresados diferencialmente en la comparación DT_HS vs DM_HS.

El gráfico tiene muchas variables involucradas. Para cada una de las categorías GO, se pueden observar puntos en diferentes posiciones. El tamaño de los puntos indica el conteo de genes, es decir el número de genes involucrados dentro de dicho término. En simultáneo, cada uno de los puntos tiene una gama de colores diferente. Dichos colores representan la significancia de cada una de las categorías. Para que una categoría sea significativa, el $p_{adj} < 0,05$. Los puntos que se asemejen al color rojizo, en base a la paleta de colores de la leyenda del gráfico, representan categorías significativas.

El gráfico resume los términos ejemplificados en la tabla 2. Se puede observar las categorías significativas de periferia celular, apoplasto, unión al ADP, pared celular y otros términos asociados a componente celular, tales como región extracelular. Los tamaños de los puntos correlacionan con el número de genes de cada una de las categorías descriptos anteriormente en la tabla 2. En la **Figura 7** se muestran todas las categorías resultantes de los genes sobre- expresados en el genotipo tolerante con respecto al susceptible.

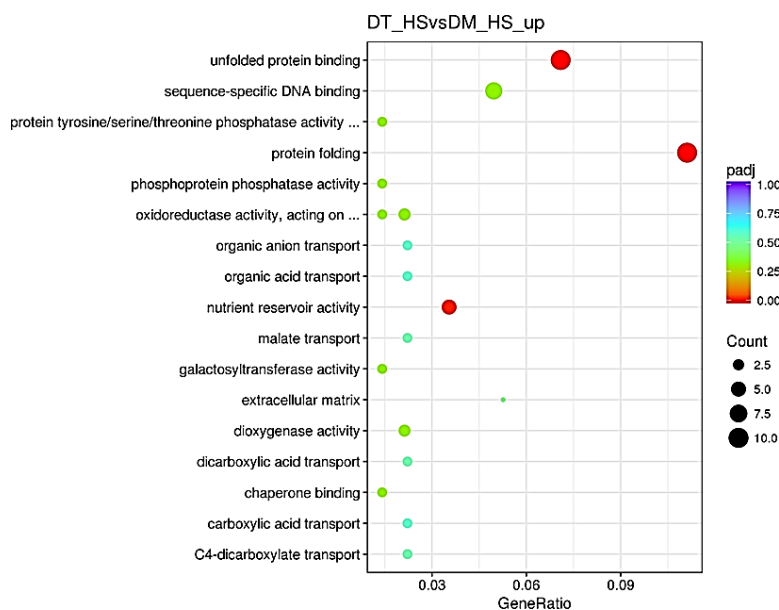


Figura 7. Categorías de la plataforma GO asociados al conjunto de genes sobre- expresados diferencialmente en la comparación DT_HS vs DM_HS

Se pueden observar las categorías estadísticamente significativas (asociadas a genes sobre expresados) en la comparación. La unión de proteínas no plegadas se destaca como una de las categorías significativas, representada por un conjunto de 10 genes. Asimismo, el plegado de proteínas también resulta como una de las categorías significativas, también representado por 10 genes. La última categoría significativa es la actividad de reserva de nutrientes, representada por 5 genes.

Como una primera aproximación que se puede realizar a partir del análisis GO, es evidente que hay un conjunto de genes sobre expresados en el genotipo tolerante DT involucrados en diferentes procesos asociados a proteínas, como se observa en las categorías de unión de proteínas no plegadas y el plegado de proteínas. Es posible que dichos mecanismos contribuyan en mayor o en menor medida a la tolerancia observada en el genotipo DT.

Análisis de Enriquecimiento utilizando la herramienta Kyoto Encyclopedia of Genes and Genomes (KEGG)
KEGG (Enciclopedia de Genes y Genomas de Kioto, <http://www.kegg.jp/>) es una colección de bases de datos

curadas manualmente que contienen recursos de información genómica, de vías biológicas y de enfermedades (Kanehisa, 2008). El análisis de enriquecimiento identifica rutas metabólicas significativamente enriquecidas o rutas de transducción de señales asociadas con genes expresados diferencialmente, en comparación con todo el fondo del genoma. La siguiente tabla 3 resume las categorías KEGG estadísticamente significativas (es decir, con $\text{padj} < 0.05$) (**Tabla 3**).

Tabla 3. Categorías enriquecidas en genes sobre expresados (KEGG)

CATEGORIAS ENRIQUECIDAS GENES SOBREENSPRESADOS (DT_HS vs DM_HS)			
KEGG - ID	Descripción	padj	Cantidad de genes
gmx04141	Protein processing in endoplasmic reticulum	2,19E-10	34
gmx04626	Plant-pathogen interaction	0.041	9

La **Figura 8** muestra las categorías de enriquecimiento KEGG de la comparación DT_HS vs DM_HS para el caso de subexpresión génica.

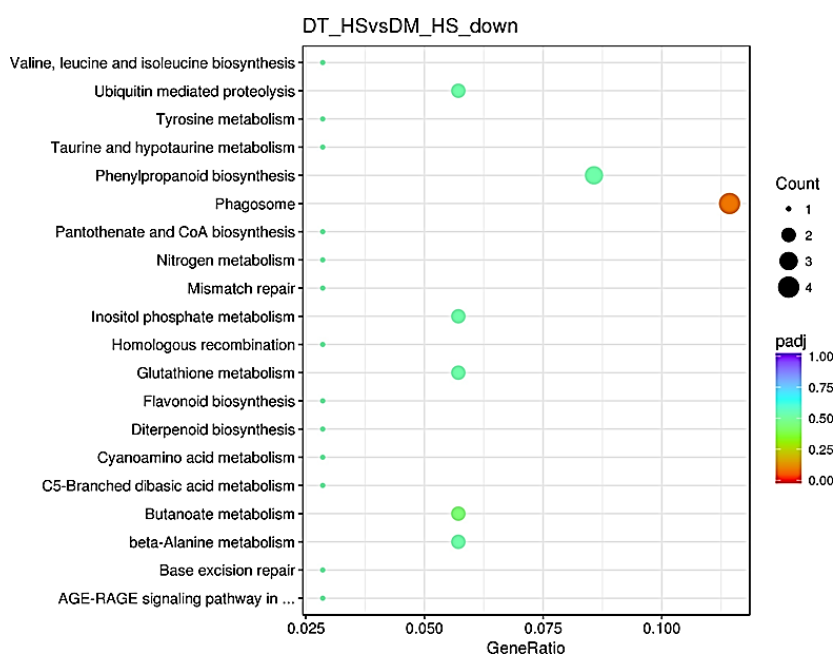


Figura 8. Categorías de la plataforma KEGG asociados al conjunto de genes sub expresados diferencialmente en la comparación DT_HS vs DM_HS.

Interpretando la **Tabla 3** y la **Figura 8** se puede extraer como conclusión que no hubo categorías KEGG asociadas a disminuciones de expresión en el genotipo tolerante DT con respecto al genotipo DM, como resultado del análisis. Dentro de las 20 categorías de mayor enriquecimiento evidenciados en la figura 8, se puede observar que

ninguno resultó significativo (valor $\text{padj} < 0,05$), para el caso de categorías KEGG asociadas a subexpresión de genes en el genotipo tolerante DT.

La **Figura 9** muestra las categorías de enriquecimiento KEGG de la comparación DT_HS vs DM_HS para el caso de sobreexpresión génica:

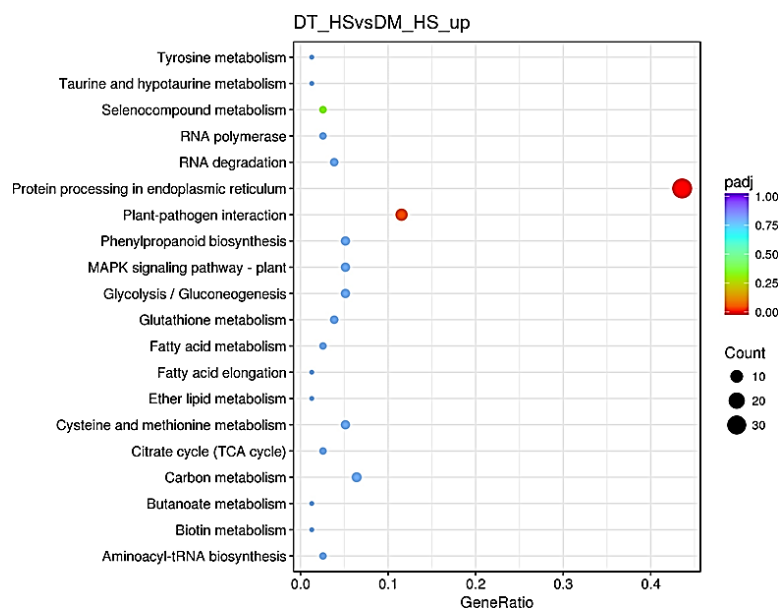


Figura 9. Categorías de la plataforma KEGG asociados al conjunto de genes sobre expresados diferencialmente en la comparación DT_HS vs DM_HS.

Como se puede observar en el cuadro 3, y se corrobora en la figura 9, solamente se observan 2 categorías KEGG que resultaron significativas (valor padj < 0,05) en el análisis, asociadas a sobreexpresión génica en el genotipo tolerante DT con respecto al genotipo susceptible DM.

Por un lado, hay una sobreexpresión de un grupo de 34 genes asociados al procesamiento de proteínas a nivel de retículo endoplasmático. Dentro de esta interpretación se suma el grupo de 9 genes representados dentro del término interacción planta – patógeno. Estas vías metabólicas sobrerrepresentadas sugieren mecanismos responsables de la tolerancia observada en el genotipo DT frente a la enfermedad.

CONCLUSIONES

Se pudo caracterizar la respuesta de los genotipos a la enfermedad, viendo el contraste en la respuesta entre D86 y DM4800, como resultado de las evaluaciones epidemiológicas. La selección de dichos genotipos contrastantes permitió posteriormente avanzar con la secuenciación. A partir de estos datos resultantes de dicho proceso, se pudo analizar los genes, su enriquecimiento y las categorías funcionales que participan de dichas respuestas. La respuesta tolerante observada en el genotipo D86 está relacionado a la sobreexpresión de genes involucrados en las vías de procesamiento de proteínas en retículo endoplasmático y en la vía de interacción planta-patógeno, lo cual se pudo dilucidar mediante el enriquecimiento KEGG. Los genes sobre expresados en el enriquecimiento GO refuerzan este abordaje, principalmente a través de la actividad incrementada de genes asociadas al plegamiento de proteínas y la unión de proteínas no plegadas. Esta

información es sumamente útil, dado que se pueden elaborar SNPs que formen parte de un plan de mejoramiento para la soja contra dicha enfermedad.

AGRADECIMIENTOS

Quiero agradecer a mi director, el Dr. Eduardo Antonio Pagano por el acompañamiento y soporte brindado a lo largo del doctorado. También, a mi familia y amigos que siempre están presentes y son fundamentales en el día a día.

BIBLIOGRAFÍA

- Abawi, G. S. and Pastor-Corrales, M. A. 1990. Root rots of beans in Latin America and Africa: diagnosis, research methodologies, and management strategies. Centro Internacional de Agricultura Tropical (CIAT). Cali, Colombia. 114 p
- Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology*, doi:10.1186/gb-2010-11-10-r106. (DESeq2)
- Cassina, Mariano. 2017. Utilización de fosfito de manganeso y potasio como tratamiento químico alternativo de semillas de soja tanto in vivo como in vitro para el control de tizón de plántulas: causadas por diferentes especies de Pythium. [Tesis de Grado, Universidad de Buenos Aires. Facultad de Agronomía].
- EEAOC, Informe Anual 2014 <http://www.eeaoc.org.ar/upload/publicaciones/archivos>
- Kanehisa M, Goto S. 2008. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research*. 2008, 28(1): 27-30.

KEGG: Enciclopedia de Genes y Genomas de Kioto, <http://www.kegg.jp>

Mengistu, A., Bond, J., Nelson, R.L., Rupe, J., Shannon, G., Arelli, P.R., Wrather, A. 2013. Identification of soybean accessions resistant to *Macrophomina phaseolina* by field screening and laboratory validation. Online. Plant Health Progress. 10.1094/PHP-2013-0318-01-RS.

Mengistu, A., Wrather, A., and Rupe, J. C. 2015. "Charcoal rot," in Compendium of Soybean Diseases and Pests, eds G. L. Hartman, J. C. Rupe, E. F. Sikora, L. L. Domier, J. A. Davis, and K. L. Steffey (St. Paul, MN: American Phytopathological Society)

Mian, M. A. R., Bailey, M. A., Ashley, D. A., Wells, R.,

Carter, T. E., Parrott, W. A., & Boerma, H. R. 1996. Molecular Markers Associated with Water Use Efficiency and Leaf Ash in Soybean. *Crop Science*, 36(5), 1252.

Pearson C.A.S, F. W. Schwenk, and Crowe F.J. 1984. Colonization of Soybean Roots by *Macrophomina phaseolina*. Department of Plant Pathology, Kansas State University, Manhattan 66506; and K. KELLEY, Southeast Kansas Branch Station, Kansas State University, Parsons 67357