



Tesis de Maestría

**PATRONES DE TRANSCRIPCIÓN Y METABOLITOS DE ÓVULOS
FECUNDADOS DE MAÍZ (*Zea mays*) INVOLUCRADOS EN LOS MECANISMOS
DE ABORTO EN FUNCIÓN DE LA DENSIDAD DEL CULTIVO**

TESISTA: Biol. SILVIA PIEROTTI

DIRECTOR: Dr. JULIO ALEJANDRO DI RIENZO

CO-DIRECTOR: Dr. JUAN IGNACIO CAGNOLA

Febrero 2021



PATRONES DE TRANSCRIPCIÓN Y METABOLITOS DE ÓVULOS FECUNDADOS DE MAÍZ (Zea mays) INVOLUCRADOS EN LOS MECANISMOS DE ABORTO EN FUNCIÓN DE LA DENSIDAD DEL CULTIVO por SILVIA PIEROTTI se distribuye bajo una [Licencia Creative Commons Atribución-NoComercial-SinDerivadas 4.0 Internacional](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Agradecimientos

A mi familia por su presencia en todo momento

Resumen

La expresión génica es el proceso mediante el cual las células son capaces de decodificar la información contenida dentro del material genético (ácidos nucleicos) para la elaboración del producto génico necesario para el buen funcionamiento del organismo. El proceso más utilizado para revelar estos productos génicos es la secuenciación de ARN. Mediante técnicas de microarreglos de ARN, se puede estudiar el nivel de expresión génica (transcriptoma) de un gran conjunto de genes al mismo tiempo. Esta técnica, genera un gran volumen de datos, y con ello, la necesidad de desarrollar nuevas herramientas, tanto informáticas como estadísticas, para su procesamiento, análisis e interpretación.

En el presente trabajo, se usaron dos estrategias para analizar los patrones de transcripción y metabolitos de óvulos fecundados de maíz. Redes de co-expresión, que tienen como objetivo identificar grupos de genes o metabolitos altamente correlacionados entre sí (módulos). Y un modelo logístico, para realizar un análisis de ontología-funcional e identificar aquellos términos de ontologías génicas que fueron modificadas en las distintas prácticas.

Luego de la construcción de la red de expresión génica se determinaron nueve módulos, cuatro de los cuales presentaban perfiles de expresión de interés, ya que los cambios observados en estos perfiles eran más destacados en comparación con los restantes y debidos a las diferentes zonas de la espiga o las densidades de siembra distintas. Además, se detectaron como centros de estos módulos genes que cumplen funciones relacionadas a “metabolismo de carbohidratos”, “pared celular” y “señalización celular”, que son funciones que se identificaron correlacionadas con aborto de granos. En uno de los cuatro módulos de interés, el gen más conectado fue un factor de transcripción perteneciente a la familia MYB reportada ampliamente como sobre expresado en distintos tipos de estrés (bióticos y abióticos).

Los modelos mixtos ajustados para los genes identificados como centros de modulo en los cuatro módulos de interés, mostraron que sólo el gen más conectado para uno de estos módulos fue sensible tanto al efecto principal de zona como de densidad, esto sugiere que este gen puede ser propuesto y estudiado como un biomarcador asociado a procesos de aborto de granos desencadenados por densidad de siembra. Los restantes, sólo fueron sensibles al efecto principal de la zona de la espiga, que se sabe es una región propensa al aborto de granos.

El análisis ontológico funcional mediante el ajuste de una regresión logística, permitió profundizar el análisis de los módulos encontrados. Mediante este análisis se detectaron los términos GO más representados dentro de cada módulo, y sus enriquecimientos bajo las distintas condiciones ensayadas.

PALABRAS CLAVE

Maíz, aborto, óvulos fecundados, densidad de siembra, redes de co-expresión, ontología funcional, modelo logístico.

Índice general

Índice de tablas	8
Índice de figuras	10
➤ Introducción	12
• Objetivos	19
➤ Materiales y métodos	21
• Ensayo experimental	21
• Polinización sincronizada	21
• Extracción de muestra	21
• Análisis metabólico y transcriptómicos	21
➤ Análisis estadístico	22
• Análisis de redes de co-expresión	22
• Selección de genes candidatos	22
• Construcción de redes de co-expresión	21
• Perfiles de expresión	23
• Análisis ontológico funcional	24
➤ Resultados y discusión	26
• Redes de co-expresión génica	27
• Redes de co-expresión de metabolitos	41
• Correlación entre genes y metabolitos más conectados	47
➤ Conclusiones	49
➤ Bibliografía	50

Índice de tablas

Valores de meda, errores estándar, p-valores y fold change obtenidos mediante el ajuste del modelo mixto	29
valores de pendientes, sus significancias y una descripción para las GO consideradas de interés en estudio de aborto por densidad de siembra en el módulo turquesa	30
Valores de media, errores estándar, fold change y p-valores para efectos principales de densidad, zona y su interacción para factores de transcripción identificados en el módulo turquesa	31
Pendientes estimadas, significancias y una descripción de las GO consideradas de interés en estudio de aborto en el módulo azul	32
Valores de medias, errores estándar, fold change y p-valores obtenidos en el ajuste del modelo mixto para los 2 genes identificados como factores de transcripción en el módulo azul.	32
Valores de pendientes, sus significancias y una descripción de las GO significativas presentes en el módulo marrón	34
valores de fold change, medias corregidas, errores estándar y p-valores obtenidos en el ajuste de un modelo mixto para cada gen identificado como factor de transcripción en el módulo marrón	33
valores de fold change, medias corregidas, errores estándar y p-valores obtenidos en el ajuste del modelo mixto para el gen identificado como factor de transcripción y centro de módulo	34
p-valores obtenidos mediante el ajuste de un modelo mixto para el gen identificado como factor de transcripción y centro de modulo negro	35
p-valores obtenidos mediante el ajuste de un modelo mixto para el gen con expresión más nítida del módulo turquesa	36
valores de fold change, medias y errores estándar, p-valores obtenidos mediante el ajuste de un modelo mixto para los genes de interés con expresión más nítida del módulo azul	38
Valores de medias, errores estándar, fold change y p-valores obtenidos mediante el ajuste de un modelo mixto para los genes de interés con expresión más nítida del módulo marrón	39
Valores de meda, errores estándar, p-valores y fold change obtenidos mediante el ajuste del modelo mixto para los metabolitos centrales de los módulos de interés.	42

Valores de meda, errores estándar, p-valores y fold change obtenidos mediante el ajuste del modelo mixto para los metabolitos con expresión más nítida 46

Índice de figuras

Perfiles de expresión génica promedio en los cuatro módulos obtenidos a partir de los perfiles centrados y escalados de expresión génica sobre los tratamientos 28

biplot correspondiente al análisis de componentes principales para la identificación de genes con niveles de expresión más nítida entre tratamientos 36

Perfiles de expresión metabólica promedio en los módulos obtenidos a partir de los perfiles centrados y escalados de la expresión metabólica sobre los doce tratamientos42

biplot correspondiente al análisis de componentes principales para la identificación de metabolitos con niveles de expresión más nítida entre tratamientos46

mapa de calor para las correlaciones entre genes y metabolitos más conectados de cada módulo48

INTRODUCCIÓN

El maíz (*Zea mays*) es uno de los principales cereales producidos en el mundo y en la Argentina (FAO-STAT, 2017). El rendimiento por área ($Tn\ ha^{-1}$) de cultivos de maíz depende del número de granos fijados durante el período crítico (intervalo de tiempo de 30 días alrededor de la floración, Andrade et al. 1999). Durante este período de tiempo los estigmas de la inflorescencia femenina (espiga) se elongan hasta quedar expuestos (fuera de las chalas) para recibir el polen de las inflorescencias masculinas (panoja) de plantas vecinas. A pesar de que los estigmas reciban polen y el óvulo sea fecundado, el desarrollo del grano se puede interrumpir a causa de un proceso denominado aborto que sucede principalmente en los ovarios apicales de las espigas (últimos en desarrollarse dentro de la inflorescencia). El desarrollo de los granos de la punta de la espiga es dominada por los granos de la base de la espiga que son fertilizados antes en el tiempo (Cárcova & Otegui 2007). La des-sincronización en la fecundación de los ovarios de la punta y la base de la espiga aumenta el proceso de aborto (Cárcova & Otegui 2001; Oury et al. 2016). El aborto puede ser grave particularmente bajo condiciones de estrés por sequía (Hall et al. 1981; Westgate y Boyer 1986) o alta densidad de siembra (Otegui 1997). A los 10 días después de la polinización (DDP) se puede evidenciar morfológicamente el aborto de los óvulos fecundados (Westgate & Boyer 1986), incluso Cárcova y Otegui (2001) sostienen que los primeros 8 DDP son altamente sensibles a condiciones de estrés causando abortos de granos.

Conocer cuáles son los mecanismos que disparan el aborto de óvulos en condiciones de estrés fue el objetivo de innumerables estudios en las últimas cuatro décadas, principalmente bajo estrés por sequía. Tradicionalmente, la reducción de la disponibilidad de carbohidratos en los ovarios se ha considerado el principal factor desencadenante del aborto del grano (Boyle et al. 1991; Boyer & Westgate 2004; Ruan et al. 2012). Los bajos potenciales de agua pueden reducir la acumulación de glucosa y almidón, así como la actividad invertasa de la pared celular en el pedicelo superior (McLaughlin & Boyer 2004) y condicionar al óvulo a abortar. En base a hallazgos similares, se propuso el aborto como resultado de bajos niveles de fotosíntesis en condiciones de estrés hídrico, lo que reduce las concentraciones de azúcar en el ovario y causa menor actividad de las enzimas del metabolismo del azúcar y la muerte celular programada (Boyer y McLaughlin 2007; Ruan et al. 2012). Actualmente existen trabajos en la literatura donde caracterizan el conjunto de todos los ARNm (transcriptoma) de ovarios de maíz durante los primeros DDP en condiciones de sequía (Oury et al. 2016; Cagnola et al. 2018). A partir de la expresión

genética Oury et al. (2016) sugieren que el aborto es principalmente causado por cambios asociadas a la expansión celular y no tanto a los cambios en los niveles de carbohidratos como reportaron otros autores previamente. En el mismo sentido, se han identificado genes de la familia de los arabino-galactanos como jugadores claves en el aborto de óvulos causados por sequía (Cagnola et al. 2018). Mucho se sabe de los efectos del estrés por sequía, sin embargo, es poco lo que se sabe del estrés por altas densidades de siembra. En valores de densidades de siembra por encima de la de la densidad de siembra óptima, la tasa de crecimiento por planta y el número de granos por planta son menores por una menor disponibilidad de recursos para cada una de ellas (Andrade et al. 1999) induciendo el aborto de los óvulos (Otegui 1997).

La genómica funcional en plantas tiene el objetivo de identificar las funciones de los genes y sus interacciones a partir de información de los cambios en el transcriptoma y su correlación a respuestas fisiológicas (fenotipo), en distintos procesos del crecimiento y desarrollo de las plantas para intentar comprender el funcionamiento de estos genomas en su conjunto (Xiao et al. 2017).

Mediante técnicas de microarreglos de ARN, se puede estudiar el transcriptoma de un gran conjunto de genes al mismo tiempo. Esta técnica, genera un gran volumen de datos, y con ello, la necesidad de desarrollar nuevas herramientas, tanto informáticas como estadísticas, para su procesamiento, análisis e interpretación.

En el presente trabajo, se usaron dos estrategias para analizar los patrones de transcripción y metabolitos de óvulos fecundados de maíz. Redes de co-expresión, que tienen como objetivo identificar grupos de genes o metabolitos altamente correlacionados entre sí (módulos). Y un modelo logístico, para realizar un análisis de ontología-funcional e identificar aquellas ontologías génicas que fueron modificadas en las distintas prácticas.

Las redes de co-expresión son redes que integran la información obtenida en experimentos de expresión génica y metabolómica. El análisis ponderado de redes de correlación de genes (*WGCNA*, de sus siglas en inglés: *Weighted Gene Correlation Network Analysis*) tiene como objetivo hallar grupos de nodos, denominados módulos, altamente correlacionados. Los nodos de una red pueden representar genes o metabolitos. A menudo, los módulos obtenidos, representan procesos biológicos específicos (Langfelder & Horvath 2008).

Construcción de la red de co-expresión

Estas redes, se construyen con base en las correlaciones entre los perfiles de expresión de los nodos. Los perfiles de expresión, se pueden representar en una matriz $\mathbf{P}_{n \times m}$, donde cada fila (1,2,...,n), corresponde a un nodo de la red (gen o metabolito), y cada columna (1,2,...,m) corresponde a una unidad de observación. Una unidad de observación se define como el elemento de la muestra sobre el cual se mide el perfil de expresión génico (microarreglo) y metabolómico. (Langfelder & Horvath 2008), de modo que:

$$\mathbf{P} = \begin{pmatrix} p_{11} & \cdots & p_{1m} \\ \vdots & \ddots & \vdots \\ p_{n1} & \cdots & p_{nm} \end{pmatrix}$$

Una red de co-expresión, queda completamente especificada por su matriz de adyacencia $\mathbf{A}_{n \times n} = \{a_{ij}\}$. (Langfelder & Horvath, 2008). Para calcular \mathbf{A} primero se define una cantidad intermedia llamada *similitud de co-expresión*. La matriz de similitud de co-expresión está definida como:

$$\mathbf{S} = \{s_{ij}\} = |\text{cor}(p_i, p_j)|$$

Donde $\text{cor}(p_i, p_j)$ es la correlación entre los perfiles de expresión del i-ésimo y el j-ésimo nodo de \mathbf{P} , e informa la fuerza de conexión entre dichos nodos (Langfelder & Horvath, 2008).

Usando un procedimiento de umbral, la matriz de similitud de co-expresión se transforma en la matriz de adyacencia \mathbf{A} .

Para redes no ponderadas, la adyacencia a_{ij} entre los perfiles de expresión génica p_i y p_j se define mediante un umbral duro como:

$$a_{ij} = \begin{cases} 1, & s_{ij} \geq \lambda \\ 0, & s_{ij} < \lambda \end{cases}$$

Siendo λ un parámetro de umbral duro. De esta manera, entre el perfil de expresión p_i del i-ésimo nodo y el perfil de expresión p_j del j-ésimo nodo habrá una adyacencia $a_{ij}=1$ si el valor absoluto de la correlación entre p_i y p_j es igual o superior al umbral λ .

Una red ponderada de co-expresión, es aquella que permite que las adyacencias tomen valores continuos entre 0 y 1. Para ello, los valores absolutos de las correlaciones entre las filas de \mathbf{P} son elevados a una potencia π (parámetro de umbral suave), y conforman la matriz de adyacencia $\mathbf{A}_{n \times n} = \{a_{ij}\}$, donde $\{a_{ij}\}$ se define como,

$$a_{ij} = (|\text{cor}(p_i, p_j)|^\pi)$$

a_{ij} es la adyacencia entre el i -ésimo nodo y j -ésimo nodo de \mathbf{P} . De esta manera, entre el perfil de expresión p_i del i -ésimo nodo y el perfil de expresión p_j del j -ésimo nodo habrá una adyacencia $a_{ij} \in (0,1)$.

La elección adecuada del valor π puede hacerse mediante el criterio de topología libre de escala aproximada (Zhang & Horvath 2005). Se define grado del i -ésimo nodo como el número de nodos vecinos a los que se conecta el nodo i . Una red de co-expresión es libre de escala aproximada cuando los grados de sus nodos tienen una distribución exponencial negativa. Muchas redes de interacciones biológicas, como las redes génicas, de metabolitos y otras presentan una topología libre de escala aproximada (Barabási et al. 2003).

En contraste con las redes no ponderadas, las redes ponderadas reflejan la naturaleza continua de los valores de correlación, evitando una posible pérdida de información.

Elección del parámetro de umbral suave π para redes ponderadas

Como se mencionó anteriormente, una red de co-expresión es libre de escala aproximada cuando la distribución de las conectividades de sus nodos (grados de los nodos) sigue una distribución exponencial negativa. Equivalentemente, la frecuencia η de la conectividad k , η_k , (el concepto de conectividad se describe en la siguiente sección), sigue una distribución exponencial negativa (Barabási et al., 2003),

$$\eta_k \sim k^{-\pi}$$

Un valor de π adecuado es aquel que logra un buen ajustar de la red de co-expresión a una topología libre de escala aproximada. Mediante un modelo de regresión lineal se puede verificar el ajuste adecuado de la red a una topología libre de escala (Zhang & Horvath, 2005). Para este modelo, la variable respuesta es el $\log_{10}\eta_k$ y el predictor es el $\log_{10}k$ (Zhang & Horvath, 2005). El modelo de regresión lineal se ajusta para distintos valores de potencia π , y el coeficiente de determinación R^2 de dichas regresiones proporciona un índice de ajuste a una topología libre de escala.

Detección de módulos en la red

Una vez definida la matriz de adyacencias \mathbf{A} , el paso siguiente es la detección de módulos mediante el agrupamiento de nodos. Para el agrupamiento se requiere definir una medida de similitud entre nodos. Una medida de similitud basada en lenguaje de redes es la superposición topológica, TO, de su sigla en inglés, *topological overlapping* (Zhang & Horvath 2005; Yip & Horvath, 2006). La superposición topológica entre dos nodos refleja su interconectividad (Zhang & Horvath, 2005). Intuitivamente, dos nodos tienen una superposición topológica alta cuando comparten el mismo grupo de nodos vecinos. Dos nodos que comparten el mismo grupo de nodos vecinos se consideran cercanos en la red.

La matriz de superposición topológica, $\mathbf{TOM}_{ij} = [w_{ij}]$, provee una medida de similaridad entre dos nodos basada en su superposición topológica definida como:

$$w_{ij} = \frac{l_{ij} + a_{ij}}{\min\{k_i; k_j\} + 1 - a_{ij}}$$

Dónde,

a_{ij} es la adyacencia entre el nodo i y el nodo j de \mathbf{P}

$$l_{ij} = \sum_u a_{iu} a_{uj}$$

$k_i; k_j$ es la conectividad del nodo i y j respectivamente, tal que $k_i = \sum_u a_{iu}$ y $k_j = \sum_u a_{uj}$

La matriz \mathbf{TOM} , se transformará en una matriz de disimilitud $\mathbf{D}_{n \times n} = \{d_{ij}\}$ dónde:

$$d_{ij} = 1 - w_{ij}$$

Ravasz (Ravasz et al., 2002.) define al módulo de una red como una agrupación de nodos con una superposición topológica alta. La detección de módulos con nodos altamente conectados puede hacerse mediante agrupamiento jerárquico utilizando enlace promedio (UPGMA) y una medida de distancia basada en la superposición topológica.

Típicamente, cada módulo tiene un nodo (gen o metabolito) que presenta la mayor cantidad de conexiones con sus nodos vecinos. Estos nodos con alta conectividad pueden ser propuestos y estudiados como posibles reguladores de la expresión de sus vecinos o como biomarcadores involucrados en procesos biológicos de interés (Langfelder & Horvath 2008).

Si bien existen numerosas técnicas estadísticas para el análisis de co-expresión, usar una técnica basada en el lenguaje de redes permite:

1. Encontrar grupos de nodos (genes o metabolitos) altamente correlacionados entre sí (módulos) que participen en el mismo proceso biológico.
2. Encontrar, para cada módulo, el nodo con la conectividad más alta. Este nodo puede ser propuesto y estudiado como un regulador de la expresión génica o biomarcador.

El análisis ontológico–funcional, es de sumo interés en estudios de genómica. Se lleva a cabo con el objetivo de relacionar una lista de genes con características biológicas de interés que pueden ser

modificadas (enriquecidas o agotadas) bajo las distintas condiciones experimentales a las que son sometidas.

La ontología génica (GO) se define como un vocabulario en el cual se guarda toda la información referida a un gen y a los atributos del producto génico. La ontología-funcional agrupa a los distintos genes en tres campos según tres aspectos diferentes de la biología celular: Función Molecular (FM), Proceso Biológico (PB) y Componente celular (CC).

La identificación de términos enriquecidos dentro de las GO se ha convertido en una parte rutinaria del análisis e interpretación de datos de micro arreglos (Tavazoie et al ., 1999 ;Curtis et al ., 2005). El enfoque más utilizado para identificar estos términos enriquecidos se basa en contar el número de genes que están anotados en cada termino que, además, fueron diferencialmente expresados. La significación estadística de esta superposición se establece utilizando las pruebas exactas de Fisher o χ^2 (Sartor M. A et al . 2008).

Se han desarrollado varias herramientas informáticas que utilizan estos métodos, como David / EASE (Dennis, Jr et al ., 2003 ; Hosack et al ., 2003), el paquete Gostats de Bioconductor (Gentleman, 2005), GOMiner (Zeeberg et al ., 2003 , 2005), entre otros. Khatri y Draghici (2005) proporcionaron una comparación de varios de estos programas y Rivals et al ., 2007 presentaron una revisión exhaustiva. La limitación inherente de los enfoques que se basan en los recuentos de genes expresados diferencialmente es el requisito de elegir un nivel de corte de significación específico para distinguir entre los genes que cambian y los que no (Sartor M. A et al . 2008). Las diferentes opciones de umbral pueden conducir a categorías enriquecidas muy diferentes y, por lo tanto, a diferentes conclusiones biológicas (Pan et al ., 2005).

Sartor M. A et al . 2008 presentó y validó un nuevo método basado en regresión logística, que relaciona funcionalmente la pertenencia al conjunto de genes anotados en un término GO particular (variable dependiente) con la expresión diferencial de dichos genes (variable independiente). De esta manera, dado un término GO y un conjunto $G=\{g_1,g_2,\dots,g_n\}$ de genes, se define la variable indicadora Y talque $Y_i=1$ si el i-ésimo gen de G, pertenece a GO, caso contrario $Y_i=0$. Así, Y indica las ocurrencias de GO en G. Por otra parte, sea x_i un score para cada gen, como, por ejemplo, un *fold change*, o el estadístico de un

contraste de hipótesis (tal como, el estadístico T para diferencias de medias) podemos definir el siguiente modelo para Y.

$$\text{Log} \left(\frac{p(Y_i = 1|x_i)}{1 - p(Y_i = 1|x_i)} \right) = \text{Logit}(p_i) = \beta_0 + \beta_1 x_i$$

Donde,

$p(Y_i = 1|x_i)$ es la probabilidad que $Y_i = 1$ cuando el valor de score es x_i ,

β_1 Indica cuanto aumenta el $\text{Logit}(p_i)$ por cada incremento unitario en x_i , y

β_0 El valor esperado del logit cuando $x_i = 0$

De esta manera, si $\text{Logit}(p_i)$ aumenta (o disminuye) a medida que aumenta x_i concluimos que la categoría de GO está asociada a expresión diferencial.

La regresión logística es una extensión natural de test χ^2 , que permite que los valores de expresión permanezcan en una escala continua sin requerir el uso de umbrales de significancia.

En base a estos antecedentes, en el presente trabajo nos proponemos examinar los patrones de transcripción y de metabolitos en óvulos fecundados de espigas de maíz posicionados en zona apical (propensa al aborto) en comparación a la zona media-baja (resguardada del aborto) para cultivos en dos condiciones de densidad de siembra: baja y alta. Mediante el uso de un análisis de redes ponderadas y un análisis ontológico funcional.

OBJETIVOS

Objetivo General

Analizar y comparar los patrones transcriptómicos y metabolómicos de óvulos fecundados de maíz ubicados en dos posiciones dentro de la espiga con distinta probabilidad de aborto, para altas y bajas densidades de siembra.

Objetivos Específicos

Obtener y analizar redes de correlación de genes para encontrar grupos (módulos) de genes y metabolitos altamente correlacionados.

Determinar para cada módulo aquellos nodos con la mayor conectividad, que puedan ser propuestos y estudiados como posibles reguladores y/o biomarcadores.

Realizar un análisis de enriquecimiento ontológico-funcional para determinar aquellas ontologías que son modificadas por las distintas condiciones experimentales.

MATERIALES Y MÉTODOS

ENSAYO EXPERIMENTAL

En la campaña 2009-2010 se sembró en el campo experimental de la Facultad de Agronomía de la Universidad de Buenos Aires (34° 35'S, 58° 29'O, suelo Argiudol Vértico, Soil Survey Staff, 2010) un cultivo de maíz a dos densidades de siembra contrastante: 6 y 10 Pl m⁻². El material utilizado fue el híbrido simple AX886 de la compañía Nidera Argentina en un diseño en bloques completos aleatorizados con 3 réplicas por cada tratamiento. Cada parcela contaba de 5 surcos de una longitud de 8 m y distanciados a 0.7 m. Se sembraron manualmente 3 semillas en cada posición, luego del estadio V3 (Ritchie et al. 1993) se ralearon dos para dejar una de ellas. Al mismo momento se instaló un sistema de riego por goteo para abastecer de agua al cultivo durante todo su ciclo de vida. Se fertilizó en V4 con urea a una dosis de 100 kg N ha⁻¹. El ensayo se mantuvo libre de enfermedades, malezas e insectos que lo perjudiquen.

POLINIZACIÓN SINCRONIZADA

En cada parcela se identificaron 3 plantas a las cuales se les cubrió la espiga principal con un sobre días antes de alcanzar el estadio de *silking* (R1, Ritchie et al. 1993). El mismo día que sucedía el *silking* (liberación del primer estigma), se registraba la fecha y se dejaban pasar cuatro días para luego descubrir la espiga para que recibiera polen de plantas vecinas. Con este procedimiento se lograba que todos los estigmas de la espiga salieran por fuera de las chalas y fueran polinizados al mismo momento (Carcova y Otegui 2001). Con la sincronización de la polinización minimizamos los disturbios ocasionados por óvulos fertilizados en distintos días de crecimiento y dejábamos en evidencia los disturbios ocasionados por la posición del ovario dentro de la espiga.

EXTRACCIÓN DE MUESTRAS

A los 6 días después de la polinización se cosechó la espiga principal, se disectaron las chalas dejando al descubierto los óvulos fecundados. Se congeló inmediatamente en Nitrógeno líquido la zona apical de la espiga y la zona media inferior. Ambas posiciones se desgranaron por separado quedando únicamente los óvulos fecundados. El material proveniente de las 3 *pseudoréplicas* se mezcló y molió. La mitad de la muestra se destinó para el análisis de metabolitos y otra para su extracción de ARN y análisis de expresión genética.

ANÁLISIS METABOLÓMICOS Y TRANSCRIPTÓMICOS

Para el análisis de expresión masiva de genes utilizamos la técnica de micro-arreglos. Se extrajo el ARN de 100 mg de muestra con Trizol (Life Technologies) siguiendo los protocolos adecuados. El análisis micro-arreglos se realizó en las instalaciones de Syngenta USA utilizando un equipo *Affymetrix MAS 5.0*

Hybridisation. El análisis de presencia de 250 metabolitos se realizó en Syngenta USA y para tal fin se utilizaron las muestras en frío y pulverizadas.

ANÁLISIS ESTADÍSTICO

Los datos correspondientes a la expresión de 46682 genes y 214 metabolitos de maíz fueron obtenidos mediante un experimento factorial de tipo 2^2 . Los 4 tratamientos fueron conducidos en bloques completos con 3 repeticiones. Los factores de tratamiento fueron **zona de la espiga (Z)** con dos niveles, *punta* y *medio*, identificada como “T” y “M” respectivamente y **densidad de siembra (D)** con dos niveles 6 plantas.m⁻²; y 10 plantas. m⁻², identificados, respectivamente, como “6” y “10”.

ANÁLISIS PONDERADO DE REDES DE CO-EXPRESIÓN

Se realizó un análisis de red ponderada de co-expresión, tanto para las expresiones de genes candidatos como para la de metabolitos, a los fines de encontrar grupos de genes y metabolitos altamente correlacionados y posibles candidatos reguladores de co-expresión o biomarcadores. Para ello, se usó el paquete de R WGCNA.

La función `blockwiseModules` realiza la construcción automática de la red y la detección de módulos. La función asigna un color para cada módulo, el color gris es reservado para identificar aquellos nodos (genes o metabolitos) que mostraron superposiciones topológicas muy diferentes a las de los nodos agrupados en los restantes módulos. Es decir, agrupa aquellos nodos que no muestran semejanzas con los demás. La identificación de los diferentes módulos de la red se realizó mediante técnica de agrupamiento jerárquico utilizando enlace promedio (UPGMA) y una medida de distancia basada en la superposición topológica. Las opciones usadas para los argumentos de la función `blockwiseModules` fueron las indicadas por defecto, excepto el argumento *potencia* y *TOMType* al que se le asignó el valor “unsigned”.

La elección del valor de potencia π se hizo utilizando la función `pickSoftThreshold`. Esta función, ajusta un modelo de regresión lineal para el $\log_{10}\eta_k$ y el predictor $\log_{10}k$ proporcionando un valor de R^2 que permite seleccionar el valor adecuado de π mediante el criterio de topología libre de escala aproximada. Se ensayaron valores de π comprendidos entre 1 y 20

Para identificar los genes (o metabolitos) con las conectividades más altas para cada módulo se usó la función `chooseTopHubInEachModule`.

Selección de genes candidatos

Los análisis de expresión masiva de genes mediante técnicas de microarreglos arrojan grandes volúmenes de datos. Pero no todos estos genes muestran una expresión diferencial entre las distintas condiciones estudiadas. Por esta razón, mediante el ajuste de un modelo lineal mixto para un experimento bifactorial conducido en bloques completos, se seleccionaron aquellos genes que mostraron una expresión significativamente diferente entre tratamientos.

Para cada gen, se ajustó el siguiente modelo:

$$Y_{ijklg} = \mu + D_i + Z_j + DZ_{ij} + b_k + p_{kl} + e_{ijklg}$$

El modelo describe un experimento bi-factorial donde:

Y_{ijklg} representa el valor de expresión observado para el g -ésimo gen en la i -ésima densidad, j -ésima zona, k -ésimo bloque.

D_i es el efecto correspondiente al i -ésimo nivel del factor densidad de siembra (con dos niveles, 6 y 10 pl m⁻²),

Z_j es el efecto correspondiente al j -ésimo nivel del factor zona de la espiga (con dos niveles medio y punta),

DZ_{ij} corresponde a las interacciones entre el i -ésimo nivel del factor densidad y el j -ésimo nivel del factor zona de la espiga,

b_k corresponde al efecto aleatorio del k -ésimo bloque, para el que se asume $b_k \sim N(0, \sigma_b^2)$,

p_{kl} corresponde al efecto de la l -ésima parcela dentro de b_k y se corresponde con las dos densidades ensayadas, para la que se asume $p_{kl} \sim N(0, \sigma_p^2)$

e_{ijklg} el término aleatorio, para el que se realizan los supuestos usuales de normalidad, independencia y homocedasticidad.

Se comenzó analizando las interacciones hasta llegar a los factores principales. En este esquema analítico se consideraron los genes cuyos p -valores fueron menor que 0.01. Los genes que fueron significativos para la interacción no se consideraron en el análisis de factores principales.

Para cada gen se calculó el valor de fold change en cada término del modelo lineal mixto. Para esto se usó la función `foldchange` del paquete `gtools` de R. Para el cálculo del fold change para los términos de densidad y zona se tomó como grupo de referencia la densidad de 6 y la zona T respectivamente. Para el cálculo del fold change para las interacciones se tomó como referencia el tratamiento 6T

Perfiles de expresión

Luego de la construcción de las redes y la identificación de los módulos, se analizó el perfil de expresión promedio de los genes y metabolitos para cada tratamiento dentro de cada módulo. Para ello, los valores de expresión fueron centrados por las medias de cada nodo (gen o metabolito) y escalados por el desvío estándar de los mismos.

Para identificar aquellos genes o metabolitos que mostraron una expresión diferencial más nítida, se realizó un análisis de componentes principales.

En el espacio de los n componentes principales (cp) que lograron retener un porcentaje de variabilidad total igual o mayor a 80% se definió la distancia al origen del i -ésimo gen o metabolito como sigue:

$$D_{\text{origen}} = \sqrt{cp1^2 + cp2^2 + cp3^2 + \dots + cpn^2}$$

Luego, a estos valores de distancia D_{origen} se les determino el valor de percentil 95 (D_{95}), finalmente, se identificaron los genes y metabolitos que tuvieron un valor de distancia al origen D_{origen} superior al valor de D_{95}

Identificación de Factores de transcripción

Para cada módulo se identificaron los factores de transcripción presentes y la familia a la que pertenecen. Para esto, se utilizó una lista de factores de transcripción de *Zea mays* reportada en <http://plntfdb.bio.uni-potsdam.de/v3.0/>

ANÁLISIS ONTOLÓGICO–FUNCIONAL

Se llevó a cabo con el objetivo de relacionar una lista de genes (nodos) con características biológicas de interés que pueden expresarse o reprimirse bajo las distintas condiciones experimentales. Para este objetivo, se usó una técnica basada regresión logística.

La regresión logística, en su forma más simple, modela el $\text{logit}(p)$ de la probabilidad p de un evento en función de una covariable.

Si se toma como x_i al estadístico T para la prueba de diferencia de medias entre dos condiciones experimentales, se puede determinar si la probabilidad de la ocurrencia de GO sigue un patrón aleatorio respecto de x_i o no. Un patrón aleatorio implicaría, en el modelo logístico anterior, una pendiente (β_1) igual a cero. Si la pendiente fuera positiva y estadísticamente diferente de cero, entonces el patrón no es aleatorio y se interpreta que la ontología está siendo modificada por las condiciones experimentales contrastadas y que esa modificación esta mediada por el incremento de la actividad génica. Si la pendiente es negativa entonces la ontología también está siendo modificada, pero por una disminución de la expresión génica. Existe un tercer escenario en el que, para una misma ontología, haya tanto genes sobre como sub expresados, y en tal caso, la pendiente podría resultar cercana a cero y no significativa. Para diferenciar esta situación de un patrón aleatorio, se ajusta el mismo modelo logístico, utilizando como covariable el valor absoluto de x_i , $\text{abs}(x_i)$, y el valor esperado de la pendiente (β_2), en este caso, sería positivo. De esta forma, se pueden identificar aquellas ontologías que fueron reguladas hacia arriba, hacia abajo o que simplemente están “activas” por el efecto de los distintos tratamientos.

Para realizar este análisis se utilizó como covariable x_i el estadístico T correspondiente a las pruebas realizadas sobre los coeficientes de los modelos mixtos entre las siguientes condiciones experimentales 6T-6M, 6T-10T, 6T-10M.

Resultados y discusión

Para todas las redes, se seleccionó el valor de potencia más bajo que mostro un buen ajuste a topología libre de escala aproximada ($R^2 \geq 0.60$).

Bajo los criterios utilizados para la selección de genes candidatos, de un total de 46.682 genes presentes en cada microarreglo se retuvieron 6.894 genes candidatos.

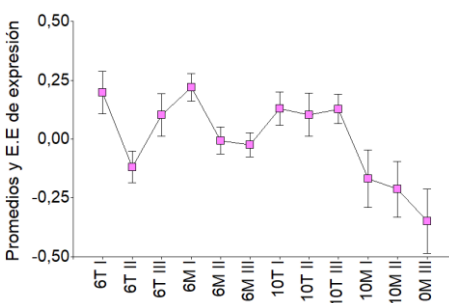
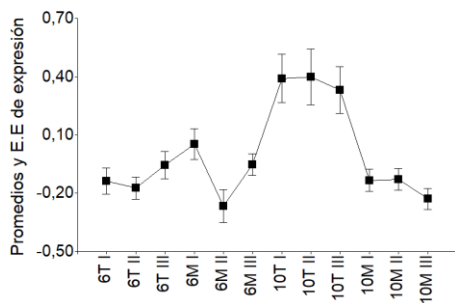
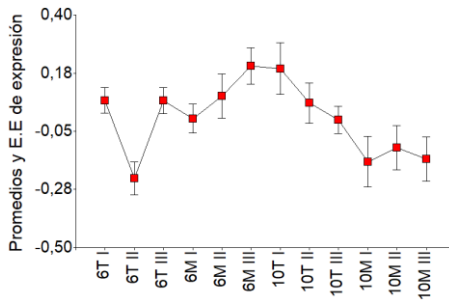
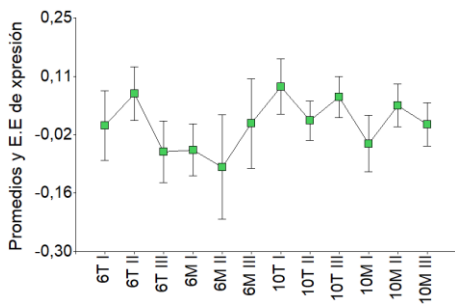
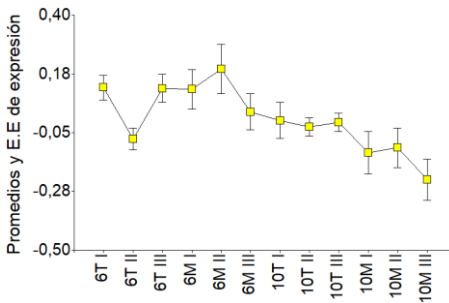
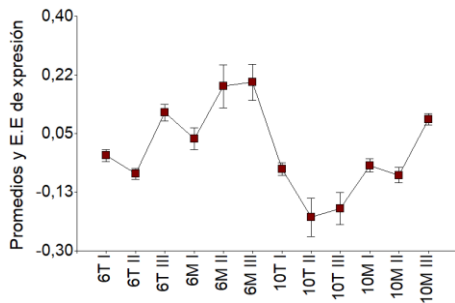
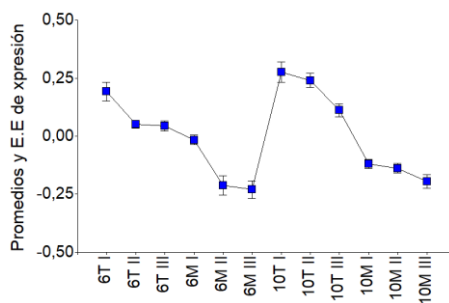
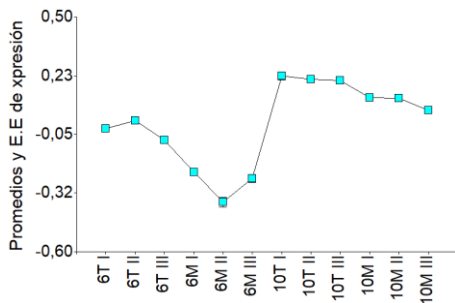
Red de co-expresión génica

Para esta red se trabajó con una potencia de umbral suave de 13, para la cual se observó un $R^2 = 0.60$, y un valor de pendiente de -0.97. Por lo tanto, la red de co-expresión génica tuvo un buen ajuste a una topología libre de escala aproximado.

El análisis agrupo los 6.894 genes en 10 módulos identificados por códigos de colores: 4.407 en el módulo *turquesa*, 968 en el módulo *azul*, 667 en el módulo *marrón*, 209 en el módulo *amarillo*, 172 en el módulo *verde*, 160 en el módulo *rojo*, 118 en el módulo *negro*, 113 en el módulo *rosa*, 69 en el módulo *magenta*, 11 genes fueron ubicado en el módulo *gris*.

Los perfiles de expresión que describen estos grupos se muestran en la Figura 1. Para los genes agrupados en el módulo *turquesa* los promedios de expresión más altos se observaron a densidad de 10 Pl.m⁻² (tanto en la zona media, M, como en la punta, T). Para los genes agrupados en el módulo *marrón* los promedios de expresión más altos se observaron a densidad 6 Pl.m⁻² (tanto para la zona T como M).

Para los genes agrupados en los módulos *azul* y *negro*, los promedios de expresión génica más altos se observaron en el tratamiento 10 Pl.m⁻² en la zona T. Para los genes agrupados en los módulos *rojo*, *rosa* y *amarillo*, los promedios de expresión más bajos se observaron a 10 Pl.m⁻² en la zona M. Por último, para los genes agrupados en el módulo verde los perfiles de expresión promedio no muestran diferencias claras entre tratamientos.



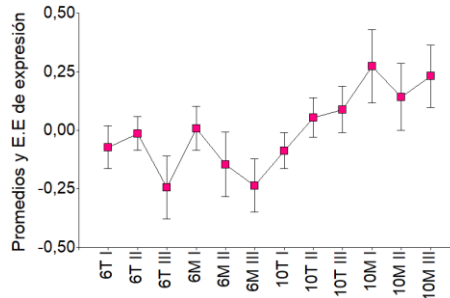


Figura 1: Perfiles de expresión génica promedio en los nueve módulos obtenidos a partir de los perfiles centrados y escalados de expresión génica sobre los tratamientos

El gen con la conectividad más alta para el módulo *turquesa* fue GRMZM2G039919_s_at. Este gen fue anotado en las ontologías GO:0004553; GO:0005618; GO:0005975; GO:0006073; GO:0016762; involucradas en actividades referidas a metabolismo de carbohidratos y en la ontología GO:0048046 involucrada en actividad de apoplastos (pared celular).

El gen GRMZM2G045430_01_at fue el más conectado en el módulo *marrón*, entre las ontologías en las cuales ha sido anotado este gen se mencionan GO:0004674; GO:0004713; GO:0005509. El gen GRMZM2G100288_01_at fue el más conectado en el módulo *magenta* y ha sido anotado en la ontología GO:0004672. Todas estas ontologías están asociada a actividad de proteína quinasa (señalización).

El gen GRMZM2G108959_01_at fue el más conectado en el módulo *negro*, este gen ha sido anotado en la ontología GO:0003677 involucrada en proteínas de unión a ADN. Además, este gen ha sido anotado como factor de transcripción perteneciente a la familia MYB.

GRMZM2G083725_01_at para el módulo *verde* que ha sido anotado en las ontologías GO:0006869; GO:0008289 involucradas en actividades de transporte y unión a lípidos.

GRMZM2G109526_s_at para el módulo *rojo*, este gen ha sido anotado en las ontologías GO:0006457; GO:0015031 involucradas en plegamiento y transporte de proteínas respectivamente.

Por último, el gen GRMZM2G144857_01_at fue el más conectado en el módulo *azul*, el gen GRMZM2G026013_s_at fue el más conectado en el módulo *amarillo* y el gen GRMZM2G101699_01_at fue el más conectado en el módulo *rosa*. Estos genes no fueron anotados en ninguna ontología génica.

Como se mencionó anteriormente, la zona de la punta de la espiga (T) es más propensa al aborto y la zona media-baja (M) de la espiga está más protegida del aborto. Es por esta razón que, los módulos *turquesa*, *azul*, *marrón* y *negro* fueron los de mayor interés. En estos módulos, no sólo se observaron patrones de interés en los perfiles de expresión génica (valores de expresión génica más altos a densidad de siembra 10 y en la zona T para los módulos *turquesa*, *azul* y *negro*. Mientras que para el módulo *marrón* se observaron patrones de expresión promedio más bajos a densidad de 10 en la zona T), sino que, además, los genes más conectados para cada uno de estos módulos (a excepción del módulo *azul*, cuyo gen más conectado no fue anotado en ninguna ontología génica) están asociados con procesos, como metabolismo de carbohidratos, componentes de pared celular, ambos correlacionados con aborto

de granos (Zhong et al. 2002), señalización celular (actividad de proteínas quinasa) y proteínas de unión a ADN. La tabla 1 muestra los valores de medias, errores estándares, fold changes y p-valores obtenidos mediante el ajuste del modelo lineal mixto para estos módulos de interés.

Tabla 1: Valores de meda, errores estándar, p-valores y fold change obtenidos mediante el ajuste del modelo mixto.

Módulo	Gen más conectado	Términos del modelo	Medias E.E	fold change	p-valor	
turquesa	GRMZM2G039919_s_at	Densidad	6	10.47±0.07	-1.07	<0.01
			10	11.25±0.07		
		Zona	T	11.10±0.07	1.05	<0.01
			M	10.61±0.07		
		Densidad*Zona	6T	10.81±0.09	...	0.28
			6M	10.12±0.09	1.07	
			10T	11.39±0.09	-1.05	
			10M	11.10±0.09	-1.03	
azul	GRMZM2G144857_01_at	Densidad	6	7.64±0.07	-1.03	0.17
			10	7.85±0.07		
		Zona	T	8.10±0.07	1.10	<0.01
			M	7.38±0.07		
		Densidad*Zona	6T	8.01±0.09	...	0.47
			6M	7.26±0.09	1.10	
			10T	8.19±0.09	-1.02	
			10M	7.50±0.09	1.07	
marrón	GRMZM2G045430_01_at	Densidad	6	9.89±0.07	1.05	0.02
			10	9.41±0.07		
		Zona	T	9.46±0.07	-1.04	<0.01
			M	9.84±0.07		
		Densidad*Zona	6T	9.67±0.08	...	0.59
			6M	10.11±0.08	-1.05	
			10T	9.25±0.08	1.05	
			10M	9.58±0.08	1.01	
negro	GRMZM2G108959_01_at	Densidad	6	5.03±0.07	-1.07	0.06
			10	5.36±0.07		
		Zona	T	5.41±0.07	1.09	<0.01
			M	4.98±0.07		
		Densidad*Zona	6T	5.15±0.09	...	0.02
			6M	4.90±0.09	1.05	
			10T	5.67±0.09	-1.10	
			10M	5.05±0.09	1.02	

Para los 3 módulos de interés, se realizó un análisis de enriquecimiento de ontologías génicas (GO) mediante un modelo de regresión logística.

La tabla 2 resumen los valores de pendientes, sus significancias y una descripción de las GO consideradas de interés en estudio de aborto que están más representadas en el módulo *turquesa*. Se pudo observar

que este módulo está enriquecido con genes que cumplen funciones en “procesos metabólicos de carbohidratos” (GO:0005975), “membrana” (GO:0016020) “integrales de membrana” (GO:0016021) y “regulación de transcripción” (GO:0006355).

Para GO:0005975 se observó una prevalencia de genes sub expresados en la zona de la punta a densidad de 10. Por otra parte, el gen más conectado de este módulo (GRMZM2G039919_s_at) fue anotado en GO:0005975. Para este gen se observaron niveles de expresión significativamente distintos para los efectos principales de zona y densidad. Los valores medios más altos de expresión se observaron a densidad de 10 y en la zona T (tabla1). Esto muestra la sensibilidad de este gen a las zonas de la espiga y a la densidad de siembra.

Para GO:0006355 (regulación de transcripción dependiente de plantillas de ADN) se observó activa entre los grupos 6T-10T.

Para GO:0016020 (membrana) y GO:0016021 (integrales de membrana), se observó una prevalencia de genes sub expresados en la zona T a densidad de 10.

Tabla 2: valores de pendientes, sus significancias y una descripción para las GO consideradas de interés en estudio de aborto por densidad de siembra en el módulo turquesa

División	Trar	GO	Genes	Beta1	p-value	q-beta	beta2	beta2.p-value	beta2.q-valor	Descripción
BP	6T-10T	GO:0006355	368	-0.08	0.07	0.23	-0.16	0.02	0.08	Regulación de la transcripción dependiente de plantilla de ADN
	6T-6M			-0.23	<0.0001	<0.0001	-0.07	0.24	0.32	
	6T-10M			-0.16	<0.0001	<0.0001	-0.2	0.02	0.08	
	6T-10T	GO:0005975	180	0.24	<0.0001	<0.0001	0.04	0.66	0.48	Procesos metabólicos de carbohidratos
	6T-6M			-0.17	<0.0001	<0.0001	0.07	0.3	0.34	
	6T-10M			-0.18	0.01	0.05	0.1	0.32	0.35	
CC	6T-10T	GO:0016020	195	0.29	<0.0001	<0.0001	-0.04	0.53	0.44	Membrana
	6T-6M			-0.23	<0.0001	<0.0001	-0.03	0.58	0.44	
	6T-10M			0.16	<0.0001	<0.0001	-0.19	0.02	0.08	
	6T-10T	GO:0016021	122	0.16	<0.0001	<0.0001	0.28	<0.0001	<0.0001	Integrales de membrana
	6T-6M			-0.19	<0.0001	<0.0001	-0.06	0.36	0.38	

Cagnola et al (2018) observó un aumento de expresión de genes asociados a ubiquitinas (descomponedores de proteínas) y un aumento de un gen que codifica para una Polygalacturonasa (degradación de la pared celular) en maíz bajo distintas condiciones de riego y densidad de siembra. Estos resultados sugieren que aumentan los genes asociados a la desintegración de los tejidos cuanto mayor es el riesgo de aborto. En el análisis ontológico realizado en el módulo *turquesa*, no se encontraron términos GO asociados a ubiquitinas y Polygalacturonasa, pero sí se observó una sub expresión de genes asociados a proteínas integrales de membranas, procesos metabólicos de

carbohidratos y membranas bajo las condiciones del tratamiento 10T, esto sugiere que, bajo estas condiciones, hubo un aumento en la desintegración de tejidos.

Un total de dos genes identificados como factores de transcripción estuvieron expresados diferencialmente entre las distintas condiciones. El gen GRMZM2G051528, pertenecientes a la familia MYB, estuvo expresados diferencialmente sólo para el efecto de la densidad. Estas familias se reportaron expresadas diferencialmente en condiciones de estrés (biótico y abiótico) en el organismo modelos *A. thaliana* (R. Sharma.; 2018). Esto sugiere que estos genes son sensibles a estrés por densidad de siembra. El gen GRMZM2G001223, perteneciente a la familia MYB, estuvo expresado diferencialmente entre las zonas de la espiga y por el efecto de la densidad de siembra. La Tabla 3 muestra valores de medias, errores estándar, fold change y p-valores obtenidos mediante el ajuste del modelo mixto para estos genes.

Tabla 3: Valores de media, errores estándar, fold change y p-valores para efectos principales de densidad, zona y su interacción para factores de transcripción identificados en el módulo turquesa

Genes	Términos del modelo		Medias E.E	fold change	p-valor
MYB_GRMZM2G001223	Densidad	6	6.23 ± 0.07	-1.12	<0.01
		10	6.98 ± 0.07		
	Zona	T	6.90 ± 0.07	1.09	<0.01
		M	6.30 ± 0.07		
	Densidad*Zona	6T	6.58 ± 0.09	...	0.3
		6M	5.88 ± 0.09	1.12	
		10T	7.23 ± 0.09	-1.1	
		10M	6.73 ± 0.09	-1.02	
MYB_GRMZM2G051528	Densidad	6	6.55 ± 0.05	-1.04	<0.01
		10	6.90 ± 0.05		
	Zona	T	8.95 ± 0.05	-1.03	0.12
		M	9.19 ± 0.05		
	Densidad*Zona	6T	6.62 ± 0.07	...	0.83
		6M	6.50 ± 0.07	1.02	
		10T	6.98 ± 0.07	-1.06	
		10M	6.83 ± 0.07	-1.03	

Para el módulo azul, si bien no se pudo asociar el gen más conectado del módulo con una función específica (GO), es de interés el perfil de expresión, ya que los valores promedios más altos se observaron en la zona T a densidad 10.

Las ontologías más representadas (aquellas que tienen una mayor cantidad de genes presentes en este módulo), están asociadas a procesos referidos a “membrana”, “integrales de membrana”, “plastidos” (mitocondrias y cloroplastos).

El gen más conectado de este módulo (GRMZM2G144857_01_at) mostro diferencias significativas en su expresión sólo para el efecto principal de zona, los valores medios de expresión más altos se observaron en la zona T (tabla1).

La Tabla 4 resumen los valores de pendientes, sus significancias y una descripción de las GO consideradas de interés en estudio de aborto presentes en el módulo azul.

Para GO:0016020 (membrana) se observó una prevalencia de genes sobre expresadas en la zona T a densidad de 10.

Para GO:0016021, GO:0009536 y GO:0005975 (integrales de membrana, plastidos y procesos metabólicos de carbohidratos), se observó una prevalencia de genes sub expresados en la zona T a densidad de 10.

Tabla 4: Pendientes estimadas, significancias y una descripción de las GO consideradas de interés en estudio de aborto en el módulo azul

División	Trat	GO	Genes	beta	p-valor	q-beta	beta2	beta2.p-valor	Beta2.q-valor	Descripción
MF	6T-10T	GO:0016020	49	-0.4	<0.001	<0.001	0.12	0.56	0.68	Membrana
	6T-6M			-0.32	<0.001	<0.001	0.13	0.37	0.63	
	6T-10T	GO:0016021	39	0.41	<0.001	<0.001	0.12	0.54	0.68	Integral de membrana
	6T-10M			-0.15	0.01	0.01	-0.12	0.46	0.65	
	6T-6M			-0.24	0.01	0.01	-0.12	0.59	0.69	
CC	6T-10T	GO:0009536	21	0.29	0.03	0.03	0.25	0.37	0.63	Plastidos
	6T-10M			-0.43	0.02	0.02	0.34	0.16	0.43	
BP	6T-10T	GO:0005975	10	1.15	<0.001	<0.001	1.00	<0.001	<0.001	Procesos metabólicos de carbohidratos

Dos genes pertenecientes a este módulo fueron identificados como factores de transcripción. GRMZM2G110402_s_at perteneciente a la familia C3H mostro expresión diferencial sólo entre las zonas de la espiga, de la misma manera ocurrió con los genes GRMZM2G056986_01_at perteneciente a la familia MYB y GRMZM2G127379_01_at perteneciente a la familia NAC. Esto sugiere que, estos genes son sensibles a la zona de la espiga, pero no a la densidad de siembra. La Tabla 5 muestra los valores de fold chan, medias corregidos, errores estándar y los p-valores obtenidos en el ajuste del modelo mixto para estos tres genes.

Tabla 5 Valores de medias, errores estándar, fold change y p-valores obtenidos en el ajuste del modelo mixto para los 3 genes identificados como factores de transcripción en el módulo azul.

Genes	Términos del modelo		Medias E.E	fold change	p-valor
C3H_GRMZM2G110402_s_at	Densidad	6	10.32± 0.04	-1.01	0.04
		10	10.38 ± 0.04		
	Zona	T	10.46 ± 0.05	1.02	<0.01
		M	10.23 ± 0.05		

	Densidad*Zona	6T	10.48 0.05	...	0.05
		6M	10.15 0.05	1.03	
		10T	10.44 0.05	1.00	
		10M	10.32 0.05	1.02	
MYB_GRMZM2G051528	Densidad	6	6.55 ± 0.05	-1.05	<0.01
		10	6.90 ± 0.05		
	Zona	T	6.80 ± 0.5	1.02	0.12
		M	6.66 ± 0.05		
	Densidad*Zona	6T	6.62 ± 0.07	...	0.83
		6M	6.50 ± 0.07	1.02	
		10T	6.98 ± 0.07	-1.06	
		10M	6.83 ± 0.07	-1.03	

El módulo *marrón* se vió enriquecido con genes que cumplen funciones en “actividad quinasa del sensor de fósforo”. La Tabla 7 resumen los valores de pendientes, sus significancias y una descripción para las GO significativas en módulo *marrón*.

Para la ontología GO:0000155 (actividad quinasa del sensor de foforo), se observó una prevalencia de genes sub expresado en la zona T a densidad de 10. Además, como se mencionó anteriormente, el gen más conectado para este módulo (GRMZM2G045430_01_at) cumple funciones asociadas a la actividad de proteínas quinasa, y mostro diferencias significativas en su expresión sólo para el efecto principal de zona, el valor medio de expresión más alto se observó en la zona M de la espiga (tabla 1). Esto muestra que este gen es sensible a la zona de la espiga, pero no a la densidad de siembra.

La ontología GO:0005975 (procesos catabólicos de carbohidratos) estuvo activa entre los grupos de interés 6T-10T.

Tabla 7: Valores de pendientes, sus significancias y una descripción de las GO significativas presentes en el módulo *marrón*

División	Trat	GO	Genes	beta	p-value	q-beta	beta2	beta2.p-value	beta2.q_valor	Descripción
MF	6T-10T	GO:0000155	17	0.71	<0.001	0.02	0.12	0.7	0.26	Actividad quinasa del sensor de fosforo
	6T-6M			-0.46	0.01	0.04	-0.12	0.68	0.26	
BP	6T-10T	GO:0005975	159	0.14	0.18	0.22	0.38	0.04	0.12	Procesos metabólicos de carbohidratos
	6T-6M			-0.02	0.84	0.31	0.58	<0.001	0.01	

Dos genes anotados como factores de transcripción estuvieron sobre expresados sólo en la zona T. Uno pertenecía a la familia bHLH y el otro a la familia TALE, lo que sugiere que estos genes son factores de transcripción sensibles a la zona de la espiga, pero no a la densidad de siembra. La familia de bHLH se reportó sobre expresada bajo condiciones de estrés por sequía y salinidad en *Arabidopsis thaliana* (Baillio & Kimotho 2019).

Un gen fue anotado en la familia de factores de transcripción NF-YC, este gen mostro cambios de expresión significativos tanto para densidad como para zona, el valor de media más alto se observó a densidad 10 en la zona T. La familia NF-YC se vio involucrada en la aceleración de la floración como respuesta a escape a estrés por sequía en *Arabidopsis thaliana* (Hwang & Susila 2019)

La Tabla 8 muestra los valores de fold change, medias corregidas, errores estándar y p-valores obtenidos en el ajuste del modelo mixto para estos 4 genes.

Tabla 8: valores de fold change, medias corregidas, errores estándar y p-valores obtenidos en el ajuste de un modelo mixto para cada gen identificado como factor de transcripción en el módulo marrón

Genes	Términos del modelo		Medias E.E	fold change	p-valor
bHLH_GRMZM2G385543_01_at	Densidad	6	6.63 ± 0.11	-1.07	0.04
		10	7.13 ± 0.11		
	Zona	T	7.15 ± 0.09	1.08	<0.01
		M	6.61 ± 0.09		
	Densidad*Zona	6T	6.92 ± 0.13	...	0.09
		6M	6.35 ± 0.13	1.09	
		10T	7.38 ± 0.13	-1.07	
		10M	6.88 ± 0.13	1.01	
TALE_GRMZM2G076272_s_at	Densidad	6	6.96 ± 0.09	-1.05	0.07
		10	7.30 ± 0.09		
	Zona	T	7.25 ± 0.07	1.04	<0.01
		M	7.00 ± 0.07		
	Densidad*Zona	6T	7.08 ± 0.09	...	0.3
		6M	6.84 ± 0.09	1.04	
		10T	7.42 ± 0.09	-1.05	
		10M	7.17 ± 0.09	-1.01	
NF-YC_GRMZM2G124421_01_at	Densidad	6	8.07 ± 0.04	-1.05	<0.01
		10	8.51 ± 0.04		
	Zona	T	8.43 ± 0.04	1.03	<0.01
		M	8.15 ± 0.04		
	Densidad*Zona	6T	8.26 ± 0.07	...	0.2
		6M	7.89 ± 0.07	1.05	
		10T	8.61 ± 0.07	-1.04	
		10M	8.41 ± 0.07	-1.02	

El modulo *negro*, de menor tamaño (118 genes), tuvo como gen central a GRMZM2G108959_01_at, anotado como factor de transcripción perteneciente a la familia MYB, este gen mostro diferencias significativas en su expresión sólo para el factor principal zona (tabla 1). El valor más alto de media de expresión se observó en la zona T (tabla 1). La Tabla 9 muestra los valores de fold change, medias

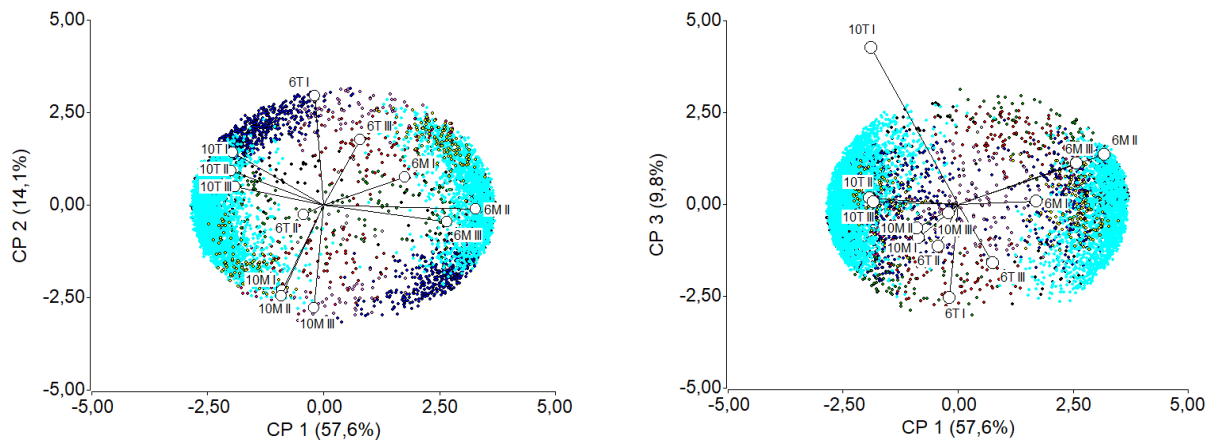
corregidas, errores estándar y p-valores obtenidos en el ajuste del modelo mixto para este factor de transcripción.

Tabla 9: valores de fold change, medias corregidas, errores estándar y p-valores obtenidos en el ajuste del modelo mixto para el gen identificado como factor de transcripción y centro de modulo

Genes	Términos del modelo		Medias E.E	fold change	p-valor
MYB_GRMZM2G108959_01_at	Densidad	6	5.03 ± 0.06	-1.07	0.02
		10	5.36 ± 0.06		
	Zona	T	5.41 ± 0.04	1.09	<0.01
		M	4.98 ± 0.04		
	Densidad*Zona	6T	5.15 ± 0.07	...	0.08
		6M	4.90 ± 0.07	1.05	
		10T	5.67 ± 0.07	-1.10	
		10M	5.05 ± 0.07	1.02	

Como se mencionó anteriormente, los genes con expresión diferencial más nítida fueron identificados mediante un análisis de componentes principales. La Figura 2 muestra el biplot correspondiente a dicho análisis. En el mismo, se puede observar que 86.7% de la variabilidad total ha sido retenida por los cuatro primeros componentes, por lo que la selección de los genes en este espacio estará basada en una representación confiable de su variabilidad.

En este espacio los puntos más extremos en cada eje de variación corresponden a los genes cuyas expresiones diferenciales son más nítidas. Como se recordará, los genes con niveles de expresión más nítida se identificaron mediante el cálculo de distancias al origen (D_i) y luego se seleccionaron aquellos genes cuyas distancias al origen superaron el valor de distancia definido por el percentil 95 (D_{95}).



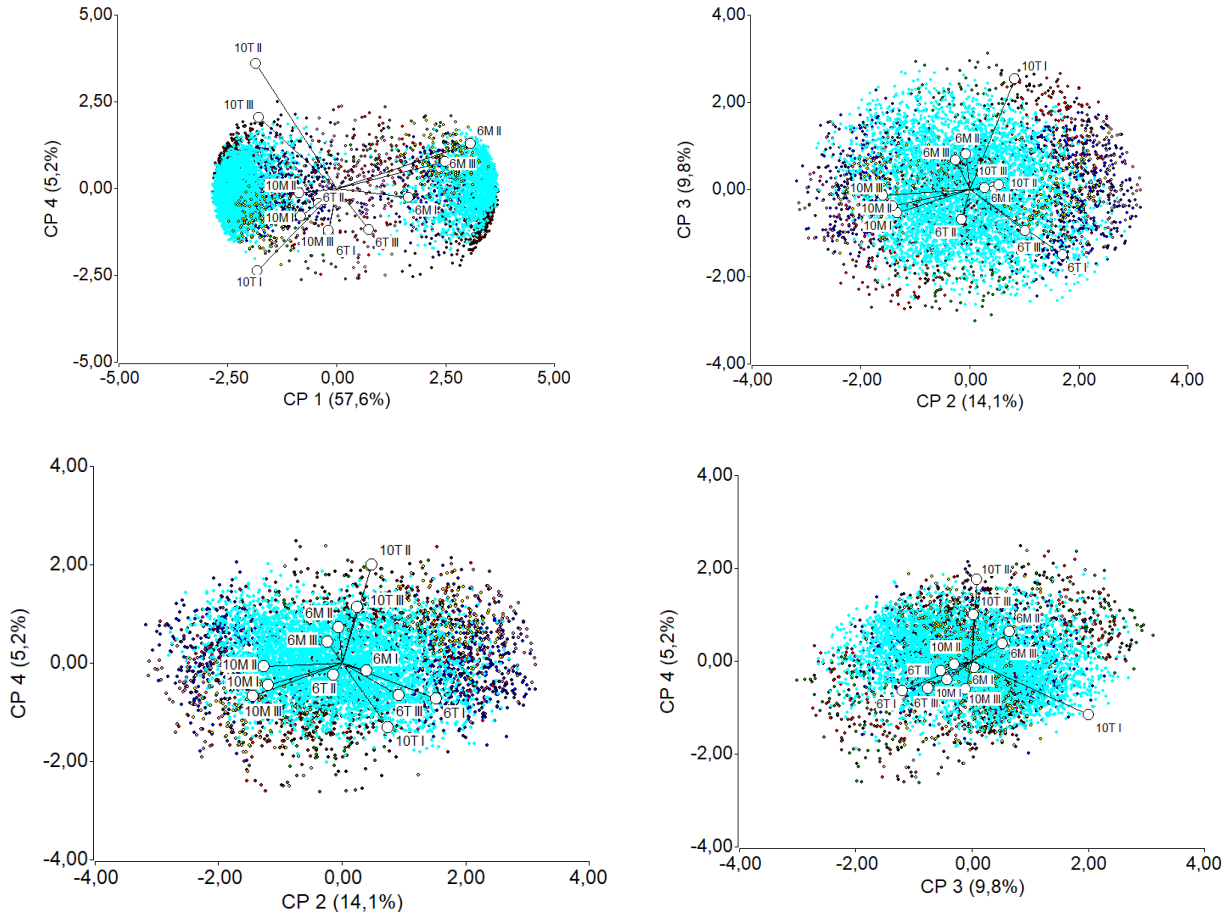


Figura 2: biplot correspondiente al análisis de componentes principales para la identificación de genes con niveles de expresión más nítida entre tratamientos

Para el módulo *turquesa* un total de 226 genes mostraron diferencias nítidas en sus niveles de expresión. 9 de estos genes presentan funciones de interés y diferencias significativas para densidad y/o zona. La Tabla 10 muestra los valores de fold change, medias, errores estándar y p-valores obtenidos en el ajuste del modelo mixto para estos genes, las GO en las que estos genes han sido anotados y una descripción de las mismas.

Tabla 10: p-valores obtenidos mediante el ajuste de un modelo mixto para el gen con expresión más nítida del módulo *turquesa*

Genes	Términos del modelo		Medias E.E	fold change	p-valor	GO	Descripción
GRMZM2G028353	Densidad	6	10.36 ± 0.07	1.08	<0.01	GO:0005515; GO:0008270; GO:0016020; GO:0016760; GO:0030244;	Unión a proteínas; unión de iones de zinc; membrana; Actividad celulosa sintasa (formadora de UDP); Proceso de biosíntesis de celulosa
		10	9.57 ± 0.07				
	Zona	T	9.70 ± 0.07	-1.06	<0.01		
		M	10.24 ± 0.07				
	Densidad*Zona	6T	10.04 ± 0.10	...	0.4		
		6M	10.68 ± 0.10	-1.06			
		10T	9.35 ± 0.10	1.07			
		10M	9.79 ± 0.10	1.03			

GRMZM2G039811	Densidad	6	8.87 ± 0.07	1.06	<0.01	GO:0016021	Integral de membrana
		10	8.34 ± 0.07				
	Zona	T	8.41 ± 0.07	-1.05	<0.01		
		M	8.80 ± 0.07				
	Densidad*Zona	6T	8.76 ± 0.07	...	0.9		
		6M	8.98 ± 0.07	-1.03			
		10T	8.06 ± 0.07	1.09			
10M		8.61 ± 0.07	1.02				
GRMZM2G128880	Densidad	6	9.34 ± 0.04	1.03	<0.01	GO:0003824; GO:0008152;	Actividad catalítica; Procesos metabólicos
		10	9.04 ± 0.04				
	Zona	T	9.13 ± 0.04	-1.01	0.06		
		M	9.25 ± 0.04				
	Densidad*Zona	6T	9.23 ± 0.05	...	0.14		
		6M	9.44 ± 0.05	-1.02			
		10T	9.02 ± 0.05	1.02			
10M		9.06 ± 0.05	1.02				
GRMZM2G090563	Densidad	6	8.79 ± 0.20	1.21	<0.01	GO:0016021	Integral de membrana
		10	7.26 ± 0.20				
	Zona	T	7.62 ± 0.20	-1.11	0.04		
		M	8.42 ± 0.20				
	Densidad*Zona	6T	8.31 ± 0.28	...	0.6		
		6M	9.26 ± 0.28	-1.11			
		10T	6.93 ± 0.28	1.20			
10M		7.58 ± 0.28	1.10				
GRMZM2G003252	Densidad	6	11.10 ± 0.09	1.07	<0.01	GO:0005199	Constituyente estructural de pared celular
		10	10.40 ± 0.09				
	Zona	T	10.45 ± 0.09	-1.06	<0.01		
		M	11.05 ± 0.09				
	Densidad*Zona	6T	10.77 ± 0.13	...	0.6		
		6M	11.43 ± 0.13	-1.06			
		10T	10.13 ± 0.13	1.06			
10M		10.67 ± 0.13	1.01				
GRMZM2G164821	Densidad	6	11.32 ± 0.07	1.03	<0.01	GO:0016021	Integral de membrana
		10	10.95 ± 0.07				
	Zona	T	10.98 ± 0.07	-1.03	<0.01		
		M	11.29 ± 0.07				
	Densidad*Zona	6T	11.19 ± 0.09	...	0.7		
		6M	11.45 ± 0.09	-1.02			
		10T	10.77 ± 0.09	1.04			
10M		11.13 ± 0.09	1.01				
GRMZM2G083058	Densidad	6	8.72 ± 0.07	1.06	<0.01	GO:0016021	Integral de

	Zona	10	8.25 ± 0.07	-1.05	0.02		membrana
		T	8.29 ± 0.08				
	M	8.68 ± 0.08	0.14				
	Densidad*Zona	6T		8.43 ± 0.07	...		
		6M		9.01 ± 0.07	-1.07		
		10T		8.14 ± 0.07	1.04		
10M		8.36 ± 0.07	1.01				
AF200533	Densidad	6	10.21 ± 0.08	1.05	<0.01	GO:0005515; GO:0008270; GO:0016020; GO:0016760; GO:0030244;	Unión a proteína; Membrana; Actividad celulosa sintasa (formadora de UDP); Procesos de biosíntesis de celulosa
		10	9.69 ± 0.08				
	Zona	T	9.68 ± 0.07	-1.06	<0.01		
		M	10.22 ± 0.07				
	Densidad*Zona	6T	10.02 ± 0.08	...	0.1		
		6M	10.40 ± 0.08	-1.04			
		10T	9.33 ± 0.08	1.07			
		10M	10.04 ± 0.08	-1.00			
	S48291434	Densidad	6	9.51 ± 0.09	1.02		
10			9.35 ± 0.09				
Zona		T	9.26 ± 0.09	-1.04	<0.01		
		M	9.61 ± 0.09				
Densidad*Zona		6T	9.34 ± 0.05	...	0.90		
		6M	9.68 ± 0.05	-1.04			
		10T	9.17 ± 0.05	1.02			
		10M	9.53 ± 0.05	-1.02			

Entre los genes del módulo azul sólo 9 de ellos mostraron diferencias nítidas, y 4 de ellos están asociados a funciones de interés en estudios de aborto. Los resultados obtenidos en los modelos mixtos ajustados para estos 4 genes mostraron que sólo tuvieron diferencias significativas para el efecto principal de zona. La Tabla 14 muestra valores de fold change, medias, errores estándar y los p-valores obtenidos en el ajuste del modelo mixto, GO y una descripción para estos 4 genes. La tabla 15 muestra los valores de medias corregidos y errores estándar.

Tabla 14: valores de fold change, medias y errores estándar, p-valores obtenidos mediante el ajuste de un modelo mixto para los genes de interés con expresión más nítida del módulo azul

Genes	Términos del modelo		Medias E.E	fold change	p-valor	GO	Descripción
GRMZM2G339728_01_at	Densidad	6	9.85 ± 0.11	1.04	0.04	GO:0016020; GO:0051260	Membrana Homooligomerización de proteína
		10	9.40 ± 0.11				
	Zona	T	9.32 ± 0.09	-1.07	<0.01		
		M	9.96 ± 0.09				
	Densidad*Zona	6T	9.45 ± 0.14	...	0.2		
		6M	10.25 ± 0.14	-1.08			
		10T	9.20 ± 0.14	1.06			

		10M	9.60 ± 0.14	-1.11			
S11418315_ug_at	Densidad	6	7.61 ± 0.06	1.03	0.2	GO:0016021; GO:0005249; GO:0006813; GO:0009536; GO:0055085	Canal de potasio dependiente de Voltaje; actividad de ion de potasio; transporte integral de plastidos; transporte transmembrana
		10	7.41 ± 0.06				
	Zona	T	7.16 ± 0.06	-1.10	<0.001		
		M	7.86 ± 0.06				
	Densidad*Zona	6T	7.31 ± 0.09	...	0.6		
		6M	7.92 ± 0.09	-1.08			
		10T	7.01 ± 0.09	-1.04			
10M		7.81 ± 0.09	-1.07				
S46909602_ug_at	Densidad	6	9.74 ± 0.09	1.02	0.01	GO:0016021; GO:0016023	Integral de membrana; Vesículas rodeadas de membrana
		10	9.62 ± 0.09				
	Zona	T	9.46 ± 0.09	-1.05	<0.001		
		M	9.90 ± 0.09				
	Densidad*Zona	6T	9.51 ± 0.08	...	0.05		
		6M	9.97 ± 0.08	-1.05			
		10T	9.41 ± 0.08	1.02			
10M		9.83 ± 0.08	-1.06				
S50707673_ug_at	Densidad	6	9.00 ± 0.05	1.06	0.	GO:0016020; GO:0016023; GO:0016740	Membrana citoplasmática; actividad de transferencia de vesícula rodeada por membrana
		10	8.50 ± 0.05				
	Zona	T	8.40 ± 0.05	-1.08	<0.01		
		M	9.10 ± 0.05				
	Densidad*Zona	6T	8.66 ± 0.10	...	0.3		
		6M	9.35 ± 0.10	-1.08			
		10T	8.14 ± 0.10	-1.06			
10M		8.86 ± 0.10	-1.02				

Un total de 100 genes pertenecientes al módulo *marrón* mostraron diferencias significativas nítidas. 5 de ellos están asociados a funciones de señalización celular. Para el caso de S48285162_ug_at, GRMZM2G375904_s_at, GRMZM2G155837_02_at y GRMZM2G011896_01_at, los modelos mixtos ajustados mostraron que sus cambios en los niveles de expresión eran significativos para los efectos principales de densidad y zona.

La Tabla 15 resume los valores de fold change, medias, errores estándar y p-valores obtenidos en el ajuste del modelo mixto para estos 5 genes y las GO en la que estos genes fueron anotados más una descripción de las mismos.

Tabla 15: Valores de medias, errores estándar, fold change y p-valores obtenidos mediante el ajuste de un modelo mixto para los genes de interés con expresión más nítida del módulo *marrón*

Genes	Términos del modelo	Medias E.E	fold change	p-valor	GO	Descripción	
S48285162_ug_at	Densidad	6	10.53 ± 0.18	1.15	<0.01	GO:0016301	Actividad quinasa
		10	9.17 ± 0.18				

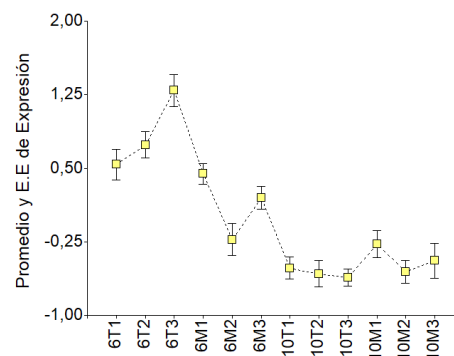
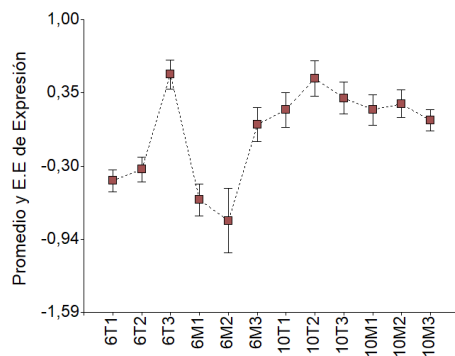
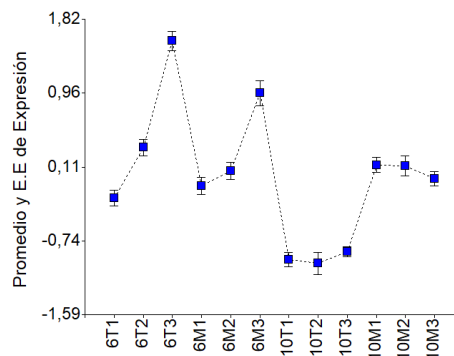
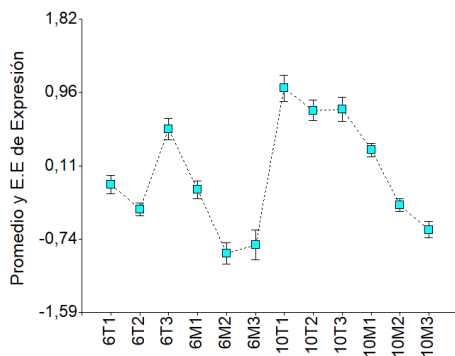
	Zona	T	9.56 ± 0.18	-1.06	0.1					
		M	10.13 ± 0.18							
	Densidad*Zona	6T	10.05 ± 0.25	...	0.2					
		6M	11.01 ± 0.25	-1.10						
		10T	9.08 ± 0.25	1.11						
		10M	9.25 ± 0.25	1.09						
GRMZM2G375904_s_at	Densidad	6	8.76 ± 0.06	1.05	<0.01	GO:0004672; GO:0004674; GO:0004713; GO:0005524; GO:0006468	Actividad de proteína quinasa; Actividad de Proteína tirosinaquinasa/ treonina quinasa; Actividad de proteína tirosina quinasa; unión a ATP; Fosforilación de proteína;			
		10	8.34 ± 0.06							
	Zona	T	8.37 ± 0.06	-1.04	0.02					
		M	8.74 ± 0.06							
	Densidad*Zona	6T	8.48 ± 0.09	-1.07	0.07					
		6M	9.05 ± 0.09							
		10T	8.26 ± 0.09					1.03		
		10M	8.42 ± 0.09					1.01		
	GRMZM2G155837_02_at	Densidad	6	9.04 ± 0.13	1.10			<0.01	GO:0004672; GO:0004674; GO:0004713; GO:0005524; GO:0006468;	Actividad de proteína quinasa; Actividad de Proteína tirosinaquinasa/ treonina quinasa; Actividad de proteína tirosina quinasa; unión a ATP; Fosforilación de proteína;
			10	8.25 ± 0.13						
Zona		T	8.34 ± 0.13	-1.07	0.02					
		M	8.95 ± 0.13							
Densidad*Zona		6T	8.68 ± 0.19	...	0.56					
		6M	9.41 ± 0.19			-1.08				
		10T	8.00 ± 0.19			1.09				
		10M	8.50 ± 0.19			1.02				
GRMZM2G011896_01_at		Densidad	6	7.68 ± 0.05	1.12	<0.01	GO:0004672; GO:0004674; GO:0004713, GO:0005515; GO:0005524; GO:0006468;	Actividad de proteína serina/ treonina quinasa; actividad proteína tirosina quinasa; unión a ATP; Fosforilación de proteína		
			10	6.88 ± 0.05						
	Zona	T	6.92 ± 0.05	-1.10	<0.01					
		M	7.64 ± 0.05							
	Densidad*Zona	6T	7.19 ± 0.06	...	0.12					
		6M	8.16 ± 0.06			-1.14				
		10T	6.66 ± 0.06			1.08				
		10M	7.11 ± 0.06			1.01				
	GRMZM2G045430_01_at	Densidad	6	9.89 ± 0.08	1.12	0.02			GO:0004674; GO:0004713; GO:0005509; GO:0005524; GO:0005529; GO:0006468; GO:0048544;	Actividad de proteína serina/treonina quinasa; Actividad proteína tirosina quinasa; Unión a ion calcio; unión a ATP; unión a azúcares; Fosforilación de proteína; Reconocimiento de polen;
			10	9.41 ± 0.08						
Zona		T	9.45 ± 0.08	-1.10	<0.01					
		M	9.84 ± 0.08							
Densidad*Zona		6T	9.67 ± 0.11	...	0.58					
		6M	10.11 ± 0.11			-1.14				
		10T	9.25 ± 0.11			1.08				
		10M	9.58 ± 0.11			1.01				

Red de co-expresión de Metabolitos

Para la construcción de esta red se trabajó con una potencia de 8. Se obtuvo un valor de $R^2 = 0.78$ y una pendiente -0.99 . Se verificó así un buen ajuste a una topología libre de escala aproximada para esta red.

De un total de 214 metabolitos, el análisis agrupó 39 metabolitos en el módulo *turquesa*, 39 en el módulo *azul*, 34 en el módulo *marrón*, 31 en el módulo *amarillo*, 21 en el *verde* y 16 en el *rojo*. 34 metabolitos no fueron asignados a ninguno de los módulos anteriores (módulo *gris*).

Los perfiles de expresión para cada módulo se muestran en la Figura . Los perfiles de expresión promedio más altos para los metabolitos agrupados en el módulo *turquesa* se observaron a densidad 10 en la zona T. En el módulo *azul* los promedios más altos de expresión se observaron a densidad de 6. En el módulo *marrón* los perfiles de expresión más altos se observaron a densidad de 10. Para el módulo *amarillo* los promedios de expresión más altos se observaron a densidad de 6. Para los metabolitos agrupados en los módulos *verde* y *rojo* no se observaron patrones de expresión de interés.



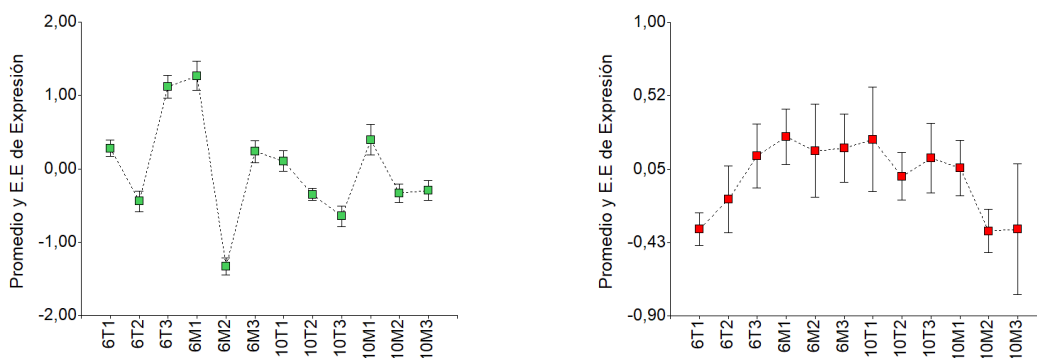


Figura 3: Perfiles de expresión metabólica promedio en los módulos obtenidos a partir de los perfiles centrados y escalados de la expresión metabólica sobre los doce tratamientos.

El metabolito con la mayor conectividad en el módulo *turquesa* fue Uridina (importante en el metabolismo de pirimidinas). Para el módulo *azul*, el metabolito con la conectividad más alta fue O-acetylserina (Derivado de fosfogliceridos). Para el módulo *marrón* fue salicilato (Fenilpropanoide). Para el modulo *verde* 1-palmitoilglicerophosphoinositol, para el caso del módulo *rojo* no fue identificado el nombre del metabolito, pero si se informó que cumple funciones en el metabolismo de purinas. Para el modulo amarillo el metabolito más conectado fue valina (Aminoácidos de cadena ramificada (derivados del piruvato)).

Los módulos *turquesa*, *azul*, *marrón* y *amarillo* mostraron patrones de interés en el estudio de aborto por densidad de siembra. La tabla 16 muestra los valores de medias, errores estándar, fold change y p-valores obtenidos mediante el ajuste de un modelo mixto para estos metabolitos.

Tabla 16: Valores de meda, errores estándar, p-valores y fold change obtenidos mediante el ajuste del modelo mixto para los metabolitos centrales de los módulos de interés.

Modulo	Metabolito	Términos del modelo	Medias E.E	fold change	p-valor	
<i>turquesa</i>	Uridina	Densidad	6	0.96 ± 0.05	-1.09	0.06
			10	1.04 ± 0.05		
		Zona	T	1.08 ± 0.09	1.16	<0.01
			M	0.92 ± 0.09		
		Densidad*Zona	6T	1.03 ± 0.05	...	0.32
			6M	0.89 ± 0.05	1.13	
			10T	1.13 ± 0.05	-1.13	
10M	0.95 ± 0.05		1.07			
<i>azul</i>	O-acetylserina	Densidad	6	1.06 ± 0.04	1.30	<0.01
			10	0.82 ± 0.04		
		Zona	T	0.85 ± 0.04	-1.21	0.08
			M	1.03 ± 0.04		
		Densidad*Zona	6T	1.03 ± 0.05	...	0.02
			6M	1.10 ± 0.05	-1.07	
10T	0.67 ± 0.05		1.53			

			10M	0.96 ± 0.05	1.07	
<i>marrón</i>	Salicilato	Densidad	6	0.67 ± 0.06	-1.95	<0.01
			10	1.31 ± 0.06		
		Zona	T	1.10 ± 0.06	1.25	0.29
			M	0.88 ± 0.06		
		Densidad*Zona	6T	0.88 ± 0.07	...	0.34
			6M	0.46 ± 0.07	1.91	
			10T	1.32 ± 0.07	-1.57	
10M	1.29 ± 0.07		-1.47			
<i>amarillo</i>	Valina	Densidad	6	1.21 ± 0.03	1.46	<0.01
			10	0.83 ± 0.03		
		Zona	T	1.05 ± 0.03	1.06	0.26
			M	0.99 ± 0.03		
		Densidad*Zona	6T	1.31 ± 0.04	...	0.05
			6M	1.12 ± 0.04	1.17	
			10T	0.80 ± 0.04	1.64	
10M	0.87 ± 0.04		1.51			

El metabolito más conectado en el módulo *turquesa* mostro diferencias significativas sólo para el efecto principal de zona. Los nucleótidos de pirimidina son esenciales para las plantas. No solo son los componentes básicos para la síntesis de ácidos nucleicos, sino también proporcionan precursores para una amplia gama de componentes celulares, como azúcar, polisacáridos, glicoproteínas, fosfolípidos, y como sustratos de enzimas, como el caso de Uridina difosfato glucosa (UDP-glucosa) sustrato de la glucosiltransferasa. En arroz de tipo aus (*Oryza sativa L.*) tolerante a estrés por sequía y frío, se encontró una acumulación mayor de este metabolito en tallo (Dong Q et al. 2019)

El metabolito más conectado del módulo *azul* mostro diferencias significativas solo para el efecto de Densidad. Los metabolitos pueden actuar como moléculas de señal que regulan las funciones de genes y proteínas, reflejando el estado del medio interno y externo (Watanabe M et al. 2021). La O-acetilserina cumple funciones reguladoras y es un metabolito clave que se relaciona con aminoácidos que contienen azufre (Watanabe M et al. 2021). El gen inducido por deficiencia de azufre (SDI) se identificó originalmente en trigo (*Triticum aestivum var. Hereward*). En el genoma de *Arabidopsis thaliana*, hay cinco genes SDI, de los cuales dos, SDI1 y SDI2, pertenecen al grupo de genes O-acetilserina. Experimentos adicionales utilizando ensayos de interacción proteína-proteína demostraron que SDI1 interactúa físicamente con MYB28, que es un regulador positivo de glucosinolatos, para reprimir la biosíntesis de glucosinolatos y utilizar el sulfato preferentemente para metabolitos primarios en condiciones de deficiencia de azufre (Watanabe M et al. 2021).

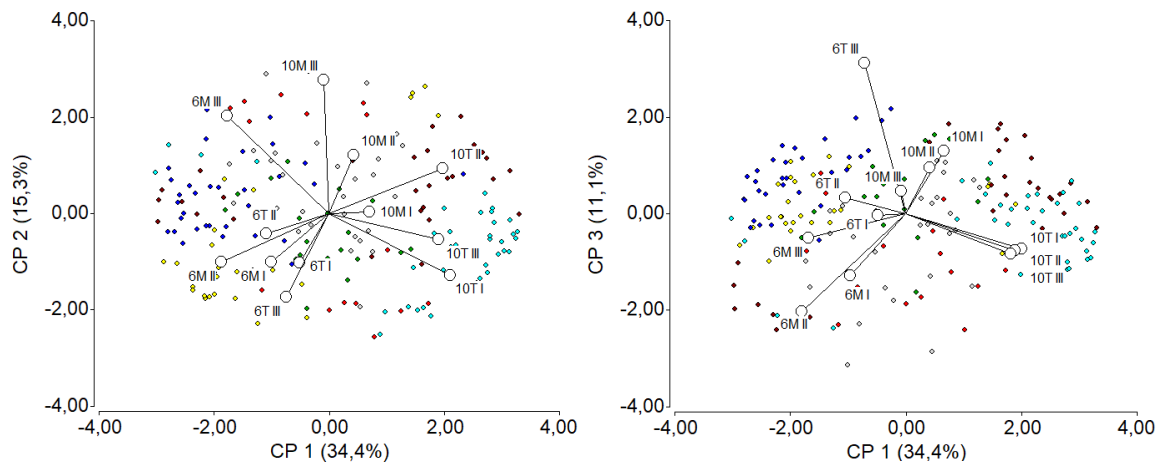
El metabolito más conectado del módulo *marrón* presento una diferencia significativa para el efecto principal de densidad. Una gran cantidad de estudios han demostrado que la vía metabólica de los fenilpropanoides responde a una variedad de estímulos ambientales, que incluyen infección por

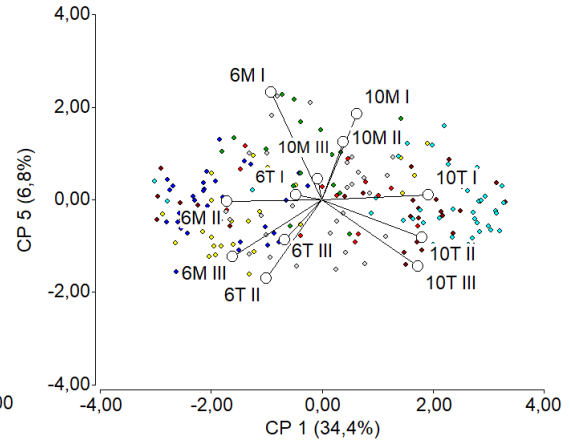
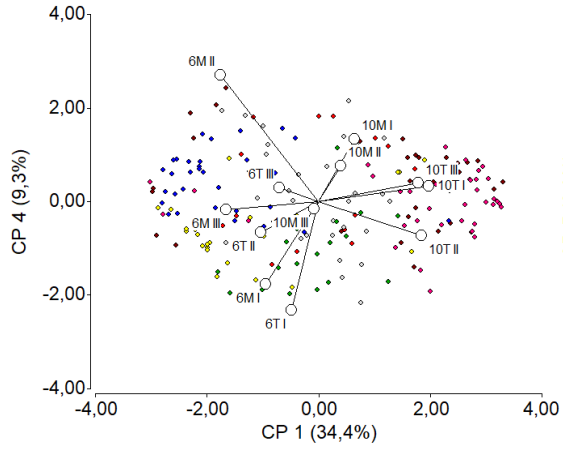
patógenos, heridas, agotamiento de nutrientes, irradiación UV, temperaturas extremas y otras condiciones de estrés. (Huan J et al. 2010)

Varios estudios han empleado enfoques genéticos y moleculares para silenciar y/o alterar la vía metabólica de los fenilpropanoides (Huan J et al. 2010). Las plantas de tabaco (*Nicotiana tabacum*) suprimidas epigenéticamente en la expresión de estas vías no desarrollaron resistencia sistémica adquirida en respuesta a la infección por el virus del mosaico del tabaco. Esta resistencia suprimida se asoció con una acumulación reducida de ácido salicílico. Análisis realizados en *Arabidopsis thaliana* mostraron una alteración en la ultra estructura de la pared celular secundaria. Además, estas plantas fueron completamente susceptibles al oomiceto de mildiú veloso (Huan J et al. 2010).

Por último, el metabolito más conectado del módulo *amarillo* mostro valores de medias con diferencias significativas para el efecto principal de densidad. Durante el estrés abiótico, la adaptación metabólica es crucial para la tolerancia a dicho estrés, y la acumulación de aminoácidos específicos como metabolitos secundarios derivados del metabolismo de los aminoácidos puede correlacionarse con una mayor tolerancia a condiciones ambientales adversas (Buffangni V et a. 2020). En estudios realizados en plantas de maíz (*Zea mays*) expuestas a sequía se observó una acumulación de aminoácidos de cadena ramificada, entre ellos valina, en la hoja. En estudios realizados en *Arabidopsis thaliana* bajo condiciones de sequía, también se observaron alteraciones a nivel transcripciones de estos aminoácidos (Obata T et al. 2015)

Los metabolitos con expresión diferencial más nítida fueron identificados mediante un análisis de componentes principales. La Figura 4 muestra el biplot correspondiente a dicho análisis. En el mismo, se puede observar que 83.1% de la variabilidad total ha sido retenida por los seis primeros componentes, por lo que la selección de los metabolitos en este espacio estará basada en una representación confiable de su variabilidad.





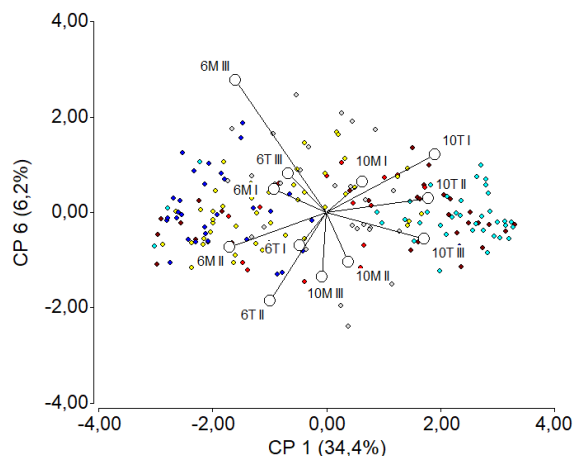


Figura 4: biplot correspondiente al análisis de componentes principales para la identificación de metabolitos con niveles de expresión más nítida entre tratamientos.

Sólo dos metabolitos mostraron diferenciación nítida, hesperetin (metabolito secundario de para flavonoides, fenilpropanoides) y ferulate (metabolito secundario para fenilpropanoides). Ambos presentaron diferencia significativa sólo para el efecto principal de densidad. La Tabla 17 muestra los valores de media, errores estándar, fold change y p-valores obtenidos en el ajuste del modelo mixto para estos metabolitos.

Tabla 17: valores de medias y E.E fold change y p-valores obtenidos mediante el ajuste de un modelo mixto para los metabolitos con expresión más nítida

Metabolitos	Términos del modelo		Medias E.E	fold change	p-valor
hesperetin	Densidad	6	2.03 ± 0.06	5.64	<0.01
		10	0.36 ± 0.06		
	Zona	T	0.96 ± 0.05	-1.50	0.04
		M	1.43 ± 0.05		
	Densidad*Zona	6T	1.64 ± 0.07	...	0.57
		6M	2.22 ± 0.07	-1.49	
		10T	0.08 ± 0.07	23	
10M		0.62 ± 0.07	2.88		
ferulate	Densidad	6	2.03 ± 0.05	4.17	<0.01
		10	0.36 ± 0.05		
	Zona	T	0.96 ± 0.07	-2.08	0.02
		M	1.43 ± 0.07		
	Densidad*Zona	6T	1.84 ± 0.06	...	0.20
		6M	2.22 ± 0.06	-1.79	
		10T	0.08 ± 0.06	8.16	
10M		0.64 ± 0.06	1.82		

Correlaciones entre metabolitos y genes más conectados en los módulos de interés

El metabolito más conectado del módulo *turquesa* (Uridina) mostro una correlación positiva con el gene más conectado de los módulos *turquesa* ($r=0.73$), que participa en procesos de metabolismo de carbohidratos y pared celular, con el gen más conectado del módulo *azul* ($r=0.86$) y con el gen central del módulo *negro* ($r=0.86$) que cumple funciones de unión a ADN. Estos genes tuvieron niveles de expresión más altos a densidad de 10 y en la zona T (figura 1), lo que sugiere que el metabolito Uridina se acumula en los granos a densidad de 10 y en la zona T susceptible al aborto. Para este metabolito se observó una correlación negativa con el gen más conectado del módulo *marrón* ($r=-0.76$) que cumple funciones de señalización celular (proteínas quinasa), esto sugiere que la acumulación de Uridina en los granos se asocia negativamente con las vías de señalización mediada por proteínas quinasa.

El metabolito más conectado del módulo *azul* (O-acetilserina) mostro una correlación positiva con el gen más conectado del módulo *marrón* ($r=0.70$), mientras que mostro una correlación significativa y negativa con el gen más conectado del módulo *negro* ($r=-0.71$) y el gene más conectado del módulo *turquesa* ($r=-0.64$). O-acetilserina tuvo una menor concentración a densidad de 10, esto sugiere que su baja acumulación en granos a esta densidad puede participar en la regulación de la expresión del gen GRMZM2G108959_01_at (factor de transcripción de la familia MYB, módulo *negro*) y/o del gen GRMZM2G039919_s_at (módulo *turqués*).

El metabolito más conectado del módulo *marrón* (salicilato) mostro una correlación positiva sólo con el gene más conectado de los módulos *turquesa* ($r=0.63$).

El metabolito más conectado del módulo *amarillo* (valina) no presento correlaciones significativas con los genes centrales de los módulos *turquesa*, *azul*, *marrón* y *negro*.

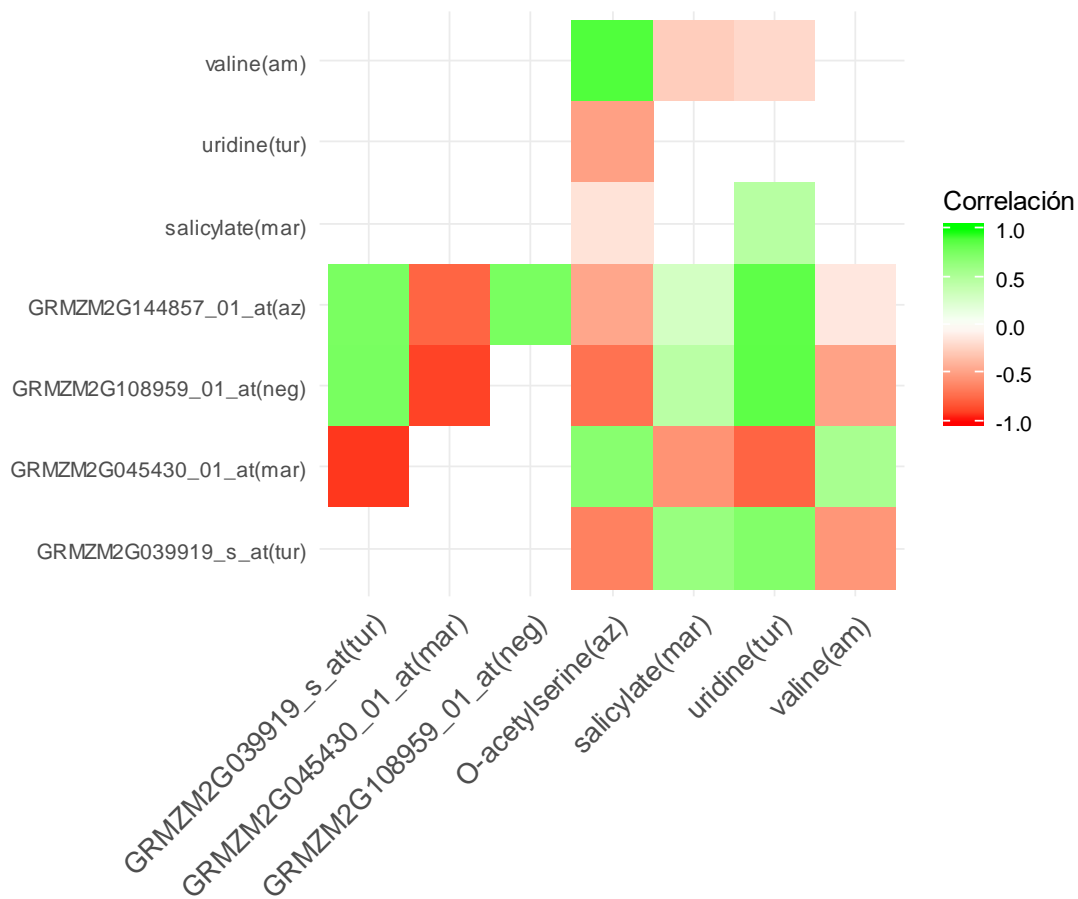


Figura 5: mapa de calor para las correlaciones entre genes y metabolitos más conectados de cada modulo

Conclusiones

El análisis de datos genómicos y metabolómicos mediante redes ponderadas de co-expresión génica es una estrategia ampliamente utilizada para detectar agrupaciones de genes (o metabolitos) potencialmente co-expresados y, además, permite detectar genes (o metabolitos) centrales en estos grupos que pueden pertenecer a una familia de factores de transcripción (que regulan la expresión aguas debajo de los restantes genes del módulo) o pueden ser propuestos como biomarcadores en procesos de interés. Sin embargo, los resultados obtenidos en estos análisis se ven afectados debido a la necesidad de una pre-selección de genes candidato, que, por lo general, excluye del análisis gran parte de los genes presentes en un microarreglo.

Para la red ponderada de co-expresión génica realizada en este trabajo, se utilizaron aquellos genes que tuvieron p-valores <0.001 para el efecto de interacción Densidad*zona o para los efectos principales de Densidad y Zona. Si bien se obtuvo una red con buen ajustarse a una topología libre de escala aproximada y, además, se pudieron detectar módulos de genes con patrones de interés en sus niveles de expresión y genes centrales en estos módulos con funciones que se vieron asociadas al aborto de granos, estos resultados pueden cambiar si se cambia el criterio de selección de genes candidatos, ya que esto podría quitar o agregar genes al grupo inicial.

Un modelo de regresión logística es una técnica sencilla y muy intuitiva para el análisis ontológico funcional, y no requiere de la instalación de paquetes ni del manejo de plataformas informáticas que fueron desarrolladas para estos fines. Si bien es una técnica que ha demostrado un buen desempeño (Sartor M. A et al. 2008), la distribución binomial se recomienda en grandes conjuntos de genes y no si se trabaja con un conjunto pequeño y específico para un fenómeno determinado.

Para este trabajo, el análisis ontológico funcional se utilizó para profundizar en la importancia de cada uno de los módulos de interés en procesos de aborto. Sin embargo, es de esperar que todos los genes anotados en una misma ontología, no hayan quedado agrupados dentro del mismo módulo. Además, algunos genes anotados en ontologías de consideración en procesos de aborto quedaron excluidos debido a la selección de genes candidatos.

Por lo expuesto en el párrafo anterior, surge el interés y la propuesta de realizar un análisis ontológico funcional independiente de redes de co-expresión génica y con la totalidad de los genes presentes en los microarreglos.

Sin embargo, a pesar de las limitaciones mencionadas para las dos estrategias propuestas, los resultados obtenidos en este trabajo apoyan los resultados publicados en la literatura que asocian los procesos referidos a metabolismo de carbohidratos, membranas celulares y pared celular con el aborto de granos bajo condiciones de estrés. Además, se propone a los genes agrupados en el módulo marrón (módulo asociado a funciones de señalización) como candidatos de interés para el estudio de aborto en granos por densidad de siembra. Destacando, no solo las interacciones entre los genes de dicho

módulo, sino también las asociaciones que hay entre estos y los pertenecientes a los módulos restantes.

Por último, los resultados obtenidos para metabolitos no solo sugieren patrones metabólicos de interés en el estudio de aborto sino también, correlaciones de interés entre estos y los genes centrales de cada módulo.

BIBLIOGRAFÍA

Baillo, EH, Kimotho, RN, Zhang, Z. y Xu, P. (2019). Factores de transcripción asociados con la tolerancia al estrés abiótico y biótico y su potencial para mejorar los cultivos. *Genes*, 10 (10), 771. <https://doi.org/10.3390/genes10100771>

Barabasi, Albert - Laszlo & Bonabeau (2003) "Scale-Free Networks", *Scientific American*: Vol. 50.

Basnet RK, Moreno-Pachon N, Lin K, Bucher J, Visser RGF, Maliepaard C, Bonnema G. Genome-wide analysis of coordinated transcript abundance during seed development in different Brassica rapa morphotypes. *BMC Genomics*. 2013;14:840.

Borges RM, Bessièrre J-M, Ranganathan Y. Diel variation in fig volatiles across syconium development: making sense of scents. *J Chem Ecol*. 2013;39:630–42.

Borrás L., Slafer G.A. & Otegui M.E. (2004) Seed dry weight response to source-sink manipulations in wheat, maize and soybean: A quantitative reappraisal. *Field Crops Research* 86, 131–146.

Boyer J.S. & Westgate M.E. (2004) Grain yields with limited water. *Journal of Experimental Botany* 55, 2385–2394.

Boyle M.G., Boyer J.S. & Morgan P.W. (1991) Stem infusion of liquid culture medium prevents reproductive failure of maize at low water potential. *Crop Science* 31, 1246–1252.

Broderick SR, Wijeratne S, Wijeratn AJ, Chapin LJ, Meulia T, Jones ML. RNA-sequencing reveals early, dynamic transcriptome changes in the corollas of pollinated petunias. *BMC Plant Biol*. 2014;14:307.

Buffagni V, Vurro F, Janni M, Gullì M, Keller AA, Marmiroli N. Shaping Durum Wheat for the Future: Análisis de expresión genética y perfiles de metabolitos apoyan la contribución de los genes BCAT a la respuesta al estrés por sequía. *Front Plant Sci* . 2020; 11: 891. Publicado el 3 de julio de 2020 doi: 10.3389 / fpls.2020.00891

Cagnola, JI, Dumont de Chassart, GJ, Ibarra, SE, et al. La expresión reducida de genes seleccionados de la proteína ARABINO GALACTAN SIMILAR A LA FASCICLINA se asocia con el aborto de granos en cultivos de campo de Zea mays (maíz) y de semillas de Arabidopsis. *Célula Vegetal Environ*. 2018; 41: 661–674 . <https://doi.org/10.1111/pce.13136>

- Casartelli A, Riewe D, Hubberten HM, Altmann T, Hoefgen R, Heuer S. Explorando el arroz tradicional tipo aus para los metabolitos que confieren tolerancia a la sequía. *Rice* (Nueva York) . 2018; 11 (1): 9. Publicado el 25 de enero de 2018. doi: 10.1186 / s12284-017-0189-7
- DiLeo MV, Strahan GD, den Bakker M, Hoekenga OA. Weighted correlation network analysis (WGCNA) applied to the tomato fruit metabolome. *PLoS One*. 2011;6:e26683.
- Dong Q, Zhang YX, Zhou Q y col. UMP Quinasa regula el desarrollo de cloroplasto y la respuesta al frío en el arroz. *Int J Mol Sci* . 2019; 20 (9): 2107. Publicado el 29 de abril de 2019. doi: 10.3390 / ijms20092107
- Hall A.J., Lemcoff J.H. & Trápani N. (1981) Water stress before and during flowering in maize and its effects on yield, its components, and their determinants. *Maydica* 26, 19–38.
- Huang J, Gu M, Lai Z y col. Análisis funcional de la familia de genes PAL de *Arabidopsis* en el crecimiento, desarrollo y respuesta de las plantas al estrés ambiental. *Plant Physiol* . 2010; 153 (4): 1526-1538. doi: 10.1104 / pp.110.157370
- Hwang K, Susila H, Nasim Z, Jung JY, Ahn JH. *Arabidopsis* ABF3 and ABF4 Transcription Factors Act with the NF-YC Complex to Regulate SOC1 Expression and Mediate Drought-Accelerated Flowering. *Mol Plant*. 2019 Apr 1;12(4):489-505. doi: 10.1016/j.molp.2019.01.002. Epub 2019 Jan 9. PMID: 30639313.
- Langfelder P & Horvath S, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* (2008), 9:559 doi:10.1186/1471-2105-9-559
- Langfelder P & Horvath S (2012). Fast R Functions for Robust Correlations and Hierarchical Clustering. *Journal of Statistical Software*, 46(11), 1-17. URL <http://www.jstatsoft.org/v46/i11/>
- McLaughlin J.E. & Boyer J.S. (2004) Glucose localization in maize ovaries when kernel number decreases at low water potential and sucrose is fed to the stems. *Annals of Botany* 94, 75–86.
- McLaughlin J.E. & Boyer J.S. (2004b) Sugar-responsive gene expression, invertase activity, and senescence in aborting maize ovaries at low water potentials. *Annals of Botany* 94, 675–689.
- Moschen S, Higgins J, Di Rienzo J, Heinz R, Paniago N and Fernandez P. Network and biosignature analysis for the integration of transcriptomic and metabolomic data to characterize leaf senescence process in sunflower
- Montaner D, Dopazo J (2010) Multidimensional Gene Set Analysis of Genomic Data. *PLoS ONE* 5(4): e10348. doi:10.1371/journal.pone.0010348
- Sharma R, Singh G, Comparative transcriptome meta-analysis of *Arabidopsis thaliana* under drought and cold stress. *PLoS One*. 2018; 13(9): e0203266.
- Obata T, Witt S, Lisek J y col. Los perfiles de metabolitos de las hojas de maíz en ensayos de campo de sequía, calor y estrés combinado revelan la relación entre el metabolismo y el rendimiento de grano. *Plant Physiol* . 2015; 169 (4): 2665-2683. doi: 10.1104 / pp.15.01164

- Otegui M.E. (1997) Kernel set and flower synchrony within the ear of maize: II. Plant population effects. *Crop Science* 37, 448–455.
- Rivals, I., Personnaz, L., Taing, L., y Potier, M.-C. (2007). Enrichment or depletion of a GO category within a class of genes: which test? *Bioinformatics*, 23(4):401– 407.
- Ruan Y.L., Jin Y., Yang Y.J., Li G.J. & Boyer J.S. (2010) Sugar input, metabolism, and signaling mediated by invertase: Roles in development, yield potential, and response to drought and heat. *Molecular Plant* 3, 942–955.
- Sartor MA, Leikauf GD, Medvedovic M, LRpath: a logistic regression approach for identifying enriched biological groups in gene expression data. *Bioinformatics* 2008; 25:211–217. URL <https://doi.org/10.1093/bioinformatics/btn592>
- Smita S, Katiyar A, Pandey DM, Chinnusamy V, Archak S, Bansal KC. Identification of conserved drought stress responsive gene-network across tissues and developmental stages in rice. *Bioinformation*. 2013;9:72–8.
- Stein O., Granot D. (2019) An Overview of Sucrose Synthases in Plants. *Frontiers in Plant Science*. 10. 95. <https://www.frontiersin.org/article/10.3389/fpls.2019.00095>
- Vega C.R.C., Andrade F.H., Sadras V.O., Uhart S.A. & Valentinuz O.R. (2001) Seed number as a function of growth. A comparative study in soybean, sunflower, and maize. *Crop Science* 41, 748–754.
- Watanabe M, Chiba Y, Hirai MY. Metabolismo y funciones reguladoras de O- acetilserina, S- adenosilmetionina, homocisteína y serina en el desarrollo de plantas y respuestas ambientales. *Front Plant Sci* . 2021; 12: 643403. Publicado el 7 de mayo de 2021. Doi: 10.3389 / fpls.2021.643403
- Westgate M.E. & Boyer J.S. (1986) Reproduction at low silk and pollen water potentials in maize. *Crop Science* 26, 951–956.
- Xiao et al., 2017: <https://www.sciencedirect.com/science/article/pii/S1674205216303082>