
**PANORAMA TRANSCRIPCIONAL DE
MEIOCITOS DE GIRASOL (*Helianthus
annuus* L.), CON DIFERENCIAS EN LA TASA
DE RECOMBINACIÓN HOMÓLOGA**



Cinvestav

Tesis Doctoral

Nathalia Maria Vanessa Flórez Zapata

Unidad de Genómica Avanzada (LANGEBIO), CINVESTAV-
IPN. Irapuato Guanajuato, México.

Agosto de 2016.

**PANORAMA TRANSCRIPCIONAL DE MEIOCITOS DE GIRASOL
(*Helianthus annuus* L.), CON DIFERENCIAS EN LA TASA DE
RECOMBINACIÓN HOMÓLOGA**

Tesis que para obtener el grado de:

Doctor en Ciencias en Biotecnología en Plantas

Presenta

MSc. Nathalia Maria Vanessa Flórez Zapata

Director de Tesis:

Dr. Octavio Martínez de la Vega

Co-director de Tesis:

Dr. Manuel Humberto Reyes-Valdés

Comité Tutorial:

Dr. Alfredo Herrera Estrella

Dr. Cei Abreu-Goodger

Dr. Ruairidh Sawers

Dr. Jean-Philippe Vielle-Calzada

Unidad de Genómica Avanzada (LANGEBIO), CINVESTAV-IPN.
Irapuato Guanajuato, México.

Agosto de 2016.

El presente trabajo fue realizado en el Laboratorio de Biología Computacional de la Unidad de Genómica Avanzada (Langebio) del Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional (Cinvestav, Irapuato), bajo la dirección del Dr. Octavio Martínez de la Vega y la co-dirección del Dr. Manuel Humberto Reyes-Valdés. Para la realización del trabajo se contó con el apoyo económico del Consejo Nacional de Ciencia y Tecnología (Conacyt) por medio del proyecto 165778 (OMV) y de la beca número 262855 (NMVFZ).



A mi familia, en especial a mamá y papá,
quienes han apoyado cada uno de mis proyectos.
A Andrés por su amor incondicional.

AGRADECIMIENTOS

Seguramente las palabras no serán suficientes para expresar toda mi gratitud a todos aquellos que han hecho parte de este proyecto y de esta etapa de mi vida. Así que haré mi mejor esfuerzo para manifestar mi más profundo agradecimiento a:

El Consejo Nacional de Ciencia y Tecnología (Conacyt) por su apoyo económico para la realización de este trabajo, a través del del proyecto 165778 y de la beca de doctorado número 262855.

Al Dr. Octavio Martínez, por abrirme las puertas de su laboratorio y brindarme todo su apoyo tanto académico como personal a lo largo de estos 5 años. Gracias Dr. por el tiempo y paciencia que ha tenido en este largo camino. Gracias por darme tantas libertades, las cuales no solo me han enseñado a ser autónoma, sino que también me han presionado a dar lo mejor de mí.

Al Dr. Humberto Reyes por aceptarme como parte de este proyecto. Gracias por sus discusiones académicas, y por apoyar cada decisión relacionada con el proyecto y mi formación.

A los miembros de mi comité tutorial, los doctores: Alfredo Herrera, Ceil Abreu, Ruairidh Sawers y Jean-Philippe Vielle-Calzada. Gracias por sus críticas, ideas y reflexiones que no sólo han enriquecido este trabajo sino también han contribuido a mi formación científica. Gracias por aguantar las extenuantes reuniones tutoriales y participar en esta maratón de titulación con su mejor disposición.

Gracias también a todos aquellos profesores del Cinvestav-Irapuato han contribuido a mi formación, a través de diferentes actividades académicas (e.g. clases y seminarios).

A Fernando Hernández (Güero), por todo su apoyo y compromiso con este proyecto, por las largas jornadas de invernadero, los trámites administrativos y su voluntad de siempre ayudar en todo lo posible.

Gracias a todos aquellos que actúan “detrás de cámaras”, es decir al personal administrativo del Cinvestav y Langebio que hacen que todo funcione. Gracias a Dora y Diana por estar al pendiente de todos los trámites y sacarme de uno que otro enredo administrativo. Gracias al equipo de Servicios Genómicos, en especial a Lupita Mireles por su excelente trabajo en la manipulación y secuenciación de las muestras. Gracias a Alejandro Hernández, por su eficiencia en la administración del proyecto.

Es tiempo de agradecer a mi segundo hogar académico en este doctorado. Gracias a los doctores Juan Luis Santos y Mónica Pradillo por acogerme en su laboratorio de Cito A en el Departamento de Genética de la Universidad Complutense de Madrid. JL gracias por tus consejos, enseñanzas y largas discusiones meióticas; gracias también por las amenas pláticas de la *mili*, la taxonomía de peces y la gastronomía española. Gracias Mónica por acompañarme y guiarme durante mi estancia en Madrid, gracias por cada uno de los *papers de domingo*, y por compartir todos tus conocimientos. Además, gracias Moni por el ejemplo

que me diste de mujer científica trabajadora, integra y apasionada con su trabajo (cuando sea grande quiero ser como tú, pero con más vacaciones y menos preocupaciones!).

A los miembros del laboratorio de Cito A. Gracias Bianca, por cubrirme la espalda con todo el trabajo de laboratorio; gracias por enseñarme a “hacer el mal”, pero sobre todo gracias por tu amistad y compañía en el viejo continente. Gracias a los *pequeños*: Cris, Miguel H, Nadia, Gonzalo y Pachón, quienes hicieron muy ameno mi trabajo en el laboratorio. Gracias a Pablo por su tiempo para enseñarme muchas de las técnicas del laboratorio y a Maricarmen por apoyarnos en tener listo todo el material necesario para los experimentos.

Regresando a los agradecimientos en México, quiero darle las gracias a mi equipo triple A: Addy, Abraham y Alan (Sergio). Gracias por su amistad, por estar ahí en los buenos y malos momentos, por hacer de Irapuato un mejor lugar para vivir. Sin ustedes estoy segura que no lo hubiera logrado. Los quiero infinitamente. Gracias también a Rosy, quien ya hace parte del club y aporta una cuarta A. Gracias Rosy y Sergio por todo este tiempo que me han dejado estar de intrusa en su casa, por hacerme sentir parte de su familia y compartir cada momento de emoción de la llegada de Matías conmigo.

Gracias a mis otros amigos mexicanos por estar ahí: Lili, Manu, Luis Yañez, Luis Fer y Víctor V. A estos dos últimos gracias además por sacarme de todos mis atascos computacionales, pero sobre todo gracias por las largas horas de conversación y momentos de “divagación existencial”. Además, quiero agradecer a América, gracias “peque” porque en entrega te sacas 10, eres una de las personas más incondicionales que he conozco, por eso y más...a ti también te quiero un montón.

Gracias a cada uno de los miembros de la familia Guerrero Torres. En especial a Mamá Chuy, quien me acogió en su casa muchos fines de semana, me cuidó y *chiquió* cuando estuve enferma. Gracias también a Yola, por sus consejos de mamá, por su cariño y por esos apapachos que me componían todo. Gracias Eri por tu amistad y cariño, por toda la ternura que me das y por enseñar a tus pequeños a quererme tanto como yo a ellos. Gracias More por tus detalles, por tu buena energía. Gracias Gaby porque eres mi amiga, mi compinche, mi apoyo, mi válvula de escape; gracias porque siempre estás ahí, pero sobre todo gracias por dejarme estar ahí. Quisiera seguir uno por uno, pero son tantos que no acabo, así que una vez más como diría Pepe: gracias a toooodos! todos!.

Obviamente, también tengo varios agradecimientos colombianos. Gracias al Profe Daniel por contagiarme del amor por la ciencia. Gracias “Profe” porque sin duda has sido una de las personas con mayor influencia en mi vida, gracias por todo tu apoyo y amistad. Gracias a esos amigos de toda la vida Roberto y Sandra, ustedes siempre son un cable a tierra y una recarga de buena energía. Gracias también Diana P, Diana L y a los exintegrantes del DUV Lab.

Gracias a Nubia Maria y Carlos Iván, mis papás. Gracias a ustedes es que hoy esta tesis es una realidad. Gracias por todo su amor y apoyo, por dejarme elegir y seguir mi propio camino. Ustedes dos son mi motivo para volar cada vez más alto. Gracias papá por ser un trabajador incansable y enseñarme a ser disciplinada, organizada y muy estudiosa. Gracias mamá por tu ejemplo de lucha y actitud positiva ante las adversidades, gracias por cada uno de tus esfuerzos por hacerme sentir amada. Gracias a mi hermano Christian, por ser mi

amigo y apoyo; gracias también por ser mi patrocinador de viajes y por cuidar de los viejos todo este tiempo. Gracias a Dani, porque tú y tú familia (Yola y Manuela), siempre han estado ahí con todo su amor. Gracias Manuela, por recordarme cada día que la felicidad está también en las pequeñas cosas.

Gracias a la familia García Suaza, doña Gladis y don Fabían, gracias por quererme como una hija y rodearme siempre con sus oraciones. Gracias Juan Pablo por ser mi amigo, por confiar en mí y también por esas *chuladas* de figuras que me has ayudado a hacer. Finalmente, pero no lo menos importante, quiero darle las gracias a Andrés. Gracias porque siempre has estado a mi lado apoyándome en todo, gracias por elegirme, por tu amor incondicional, por dejarme ser parte del mismo equipo, e incluso por resolver mis dudas estadísticas. Gracias por hacerme creer que sí se puede, y enseñarme a ser una “cholista” cuando estaba perdiendo la fuerza. Gracias al *partido a partido* y al *nunca dejes de creer, nunca dejes de soñar*, es que por fin vamos a levantar nuestra segunda copa.

RESUMEN

Después de décadas de estudio la regulación de la expresión meiótica en plantas no está completamente entendida. Sin embargo, con los recientes avances en tecnologías de secuenciación, el análisis del transcriptoma meiótico podría dar nuevas pistas acerca del panorama transcripcional de meiocitos y podría ser útil para identificar nuevos elementos regulatorios. El girasol (*Helianthus annuus* L.) presenta ventajas para el estudio de la meiosis, dada la facilidad técnica de aislar poblaciones puras de meiocitos masculinos en etapas de la meiosis bien definidas. En este estudio, secuenciamos el transcriptoma de meiocitos en profase I y una mezcla de RNA de diferentes partes somáticas de la planta (transcriptoma somático). Encontramos que los meiocitos tienen una alta actividad transcripcional que se caracteriza por la expresión de genes no anotados. Asimismo, los genes meióticos y genes relacionados con rutas de modificación de la cromatina y silenciamiento génico mostraron mayor expresión en meiocitos que en el transcriptoma somático. También se identificaron factores de transcripción con expresión preferencial conservada en girasol, *Arabidopsis* y maíz. Posteriormente se analizó el transcriptoma de dos genotipos de girasol (silvestre y domesticado), los cuales exhibían diferencias claras en tasa de recombinación homóloga meiótica, así como el de meiocitos de la F1 obtenida del cruce entre estos dos genotipos. Encontramos que la mayor proporción de genes diferencialmente expresados (DEG) fueron genes no anotados. De estos, el 73.8% pasaron filtros estrictos y fueron identificados como lncRNAs. Los lncRNAs que mostraron expresión preferencial en meiocitos tenían mayor similitud de secuencia con sRNAs de 24-nt y mayor contenido de elementos transponibles que los lncRNAs de expresión no-específica en meiosis, sugiriendo que estos podrían estar involucrados con procesos de remodelamiento de la cromatina.

Por otro lado, cuando agrupamos los genes en patrones por niveles de expresión, observamos que los genes meióticos parecen estar más representados en patrones en los cuales la expresión es significativamente mayor ($P < 0.01$) en la F1, comparado con el resto de genes expresados (44.79% vs. 32.47%). Los cambios observados en genes meióticos no permitieron explicar completamente las diferencias de recombinación entre los genotipos de girasol analizados. Sin embargo, observamos que algunos genes meióticos DE pertenecientes al mismo patrón de expresión también habían sido previamente descritos en redes de co-expresión en *Arabidopsis*. Asimismo, encontramos evidencia que sugiere que la expresión de genes meióticos en plantas parece ser coordinada por etapas. Finalmente, llevamos a cabo la caracterización funcional de algunos candidatos interesantes, que podrían ser importantes para la meiosis y/o su regulación en *Arabidopsis*. En conjunto, nuestros resultados sugieren que la expresión de genes está altamente regulada para asegurar la expresión por etapas, para lo que el control de la estructura de la cromatina y la participación de factores de transcripción específicos parecen ser importantes. Además, el presente estudio brinda información valiosa acerca de los patrones de expresión de genes meióticos, lo cual podría ser importante para la posterior identificación de redes de regulación en los meiocitos masculinos de plantas, así como para la caracterización de nuevos genes meióticos.

ABSTRACT

After decades of study, the regulation of meiotic gene expression in plants is not completely understood. However, with the recent advances in sequencing technologies, the analysis of meiotic transcriptomes could give new insights about meicytes transcriptome landscape and could be helpful to identify new regulatory elements. Sunflower (*Helianthus annuus* L.) presents advantages for the study of meiosis, because the technical feasibility of isolating almost pure male meicytes populations in well-defined meiosis stages. In this study we sequenced the transcriptome of meicytes at prophase I and a "RNA pool" of different somatic parts of the plant (somatic transcriptome). We found that meicytes have a high transcriptional activity, characterized by the expression of unannotated genes. Besides, known-meiotic genes, chromatin modification and gene silencing pathway genes showed higher expression in meicytes than in the somatic transcriptome. We also identified transcription factors with conserved preferentially expression in sunflower, *Arabidopsis* and maize meicytes. Later, we analyzed the transcriptome of two sunflower genotypes (wild and domesticated), which exhibit clear differences in meiotic homologous recombination, as well as the one of an F1 obtained from a cross between these two genotypes. We found that the greater proportion (63,7%) of differentially expressed genes (DEG) were unannotated genes. Of these, 73,8% passed very stringent filters and were identified as lncRNAs. The lncRNAs that showed preferential expression in meicytes have more sequence similarity with 24-nt sRNAs and more transposable elements content than non-meicytes exclusively expressed lncRNAs, which suggest that some of these could be involved in chromatin remodeling processes.

On the other hand, when we grouped the genes by their expression level pattern, we observed that known-meiotic genes seem to be more represented in patterns in which the expression is significantly higher ($P < 0.01$) in the F1, compared with the rest of expressed genes (44.79% vs. 32.47%). The transcriptomic changes observed in meiotic genes could not completely explain the differences in recombination between the sunflower genotypes analyzed. However, we observed that some known-meiotic genes DE with the same expression level patterns were also previously described in co-expression networks in *Arabidopsis*. We also found evidence which suggest that meiotic gene expression in plants seems to be stage-wise coordinated. Finally we conducted the functional characterization of some interesting genes that could be important for meiosis and/or its regulation in *Arabidopsis*. Altogether, our results suggest that meiotic gene expression is tightly regulated to ensure stage-wise gene expression, for which the chromatin structure control and the participation of specific transcription factors could be important. Likewise, the present study gives valuable information on the expression patterns of known-meiotic genes, which could be helpful for an ulterior identification of gene expression regulatory networks in plant male meicytes, as well as for the characterization of new meiotic genes.

INTRODUCCIÓN

La meiosis es un proceso fundamental para los organismos que tienen reproducción sexual, ya que permite la formación de gametos haploides. Además, este proceso ha sido señalado como responsable de la generación de diversidad genética, dado que durante la primera etapa de este (Profase I), ocurre la recombinación entre cromosomas homólogos, lo cual trae como consecuencia la generación de nuevas combinaciones de alelos. Esta característica particular, la hace de gran interés para las personas que trabajan en el proceso de fitomejoramiento, puesto que la posibilidad de incrementar la tasa de recombinación homóloga significaría un aumento en la velocidad del proceso de mejoramiento. Aunque en los últimos años se han realizado grandes avances en el entendimiento de los genes y procesos biológicos relacionados con la meiosis y recombinación homóloga, la regulación de la expresión génica durante esta aún no ha sido completamente elucidada.

Previamente, se describió que plantas domesticadas presentaban mayores tasas de recombinación que sus parentales silvestres (Ross-Ibarra, 2004). Lo anterior, también se cumple para girasol (*Helianthus annuus* L), en donde se describió que plantas de un genotipo domesticado presentaban niveles de recombinación significativamente mayores a los de un genotipo silvestre (Rodríguez et al., 2007). Por otro lado, debido a su arquitectura floral, este es un excelente modelo para el estudio de la meiosis, puesto que permite el aislamiento de poblaciones de células (meiocitos) en etapas específicas del proceso. Dadas las dos características mencionadas (i.e diferencias significativas en la tasa de recombinación entre genotipos y posibilidad de aislar poblaciones puras de meiocitos), el objetivo de este trabajo es describir las diferencias transcripcionales de meiocitos masculinos de girasol en Profase I, de los genotipos domesticado y silvestre que previamente mostraron un fenotipo contrastante de recombinación.

El documento actual presenta los resultados de dicho análisis de expresión diferencial, organizado por capítulos que responden (a excepción del Capítulo 1), a cada uno de los objetivos específicos fijados para el cumplimiento del objetivo principal. De esta manera, en el Capítulo 1, se realiza una contextualización al trabajo, presentando una breve introducción a la meiosis y su regulación, así como a las metodologías que han sido empleadas para su estudio en plantas. En el Capítulo 2, se describen los resultados del análisis del transcriptoma de los meiocitos de girasol en comparación con el transcriptoma somático de esta planta, el cual fue realizado con el propósito de identificar rasgos particulares del panorama transcripcional de meiocitos de girasol. Por su parte en el Capítulo 3, se presentan los resultados del análisis de expresión diferencial entre los genotipos domesticado y silvestre, donde se destacan los lncRNAs como los mayores contribuidores de cambios transcriptómicos. Por tanto, en este capítulo se fijó el objetivo de describir algunas de las características de los lncRNAs identificados, como una primera aproximación para entender su función en meiosis. La descripción de cambios en elementos regulatorios de la

transcripción (e.g. factores de transcripción) y genes previamente descritos con función en meiosis, es presentada en el Capítulo 4, complementando el análisis de las diferencias transcripcionales en los meiocitos de girasol. Asimismo, en este capítulo se documentan algunos rasgos comunes entre los patrones de expresión de genes meióticos de *Arabidopsis* y girasol, por lo que se proponen algunos rasgos y elementos comunes de la regulación transcripcional que ocurre en la meiosis de plantas. Finalmente, en el Capítulo 5, se presenta una primera aproximación a la caracterización funcional de algunos genes diferencialmente expresados entre meiocitos de los genotipos de girasol (candidatos a participar en la regulación de la meiosis), a través del análisis del fenotipo de mutantes insercionales de sus ortólogos en *Arabidopsis thaliana*.

CONTENIDO

LA MEIOSIS EN PLANTAS	15
1.1 MEIOSIS: LA GENERACIÓN DE GAMETOS HAPLOIDES.	15
1.2 RECOMBINACIÓN HOMÓLOGA DURANTE LA MEIOSIS.	18
1.2.1 FORMACIÓN Y PROCESAMIENTO DE LAS DOBLES ROTURAS.	18
1.2.2 INTERCAMBIO DE LA CADENA SENCILLA DE DNA.	20
1.2.3 VÍAS PARA LA FORMACIÓN DE COS Y NCOS.	21
1.3 MECANISMOS DE REGULACIÓN DE LA RECOMBINACIÓN HOMÓLOGA Y FORMACIÓN DE ENTRECruzAMIENTOS EN MEIOSIS.	23
1.4 VARIABILIDAD DE LA RECOMBINACIÓN HOMÓLOGA EN MEIOSIS.	26
1.5 APROXIMACIONES METODOLÓGICAS DEL ESTUDIO DE LA MEIOSIS EN PLANTAS.	27
1.3 EL GIRASOL COMO MODELO DE ESTUDIO DE LA MEIOSIS EN PLANTAS.	28
EL TRANSCRIPTOMA DE LOS MEIOCITOS DE GIRASOL.	30
2.1 INTRODUCCIÓN	30
2.2 MATERIALES Y MÉTODOS	31
2.2.1 AISLAMIENTO DE MEIOCITOS, EXTRACCIÓN DE RNA Y SECUENCIACIÓN.	31
2.2.2 FILTRADO DE CALIDAD DE SECUENCIAS, ENSAMBLAJE Y RE-MAPEO.	32
2.2.3 IDENTIFICACIÓN DE TRANSCRITOS.	32
2.2.4 QRT-PCR.	32
2.2.5 DISEÑO Y ANÁLISIS ESTADÍSTICO.	33
2.3 RESULTADOS Y DISCUSIÓN	33
2.3.1 VISIÓN GENERAL DEL TRANSCRIPTOMA DE MEIOCITOS DE GIRASOL.	33
2.3.2 PERFIL DE EXPRESIÓN DE GENES CON FUNCIÓN EN MEIOSIS.	36
2.3.2 CATEGORÍAS FUNCIONALES SOBREENSPRESADAS EN MEIOCITOS	38
2.3.3 NIVELES DE EXPRESIÓN DE REGULADORES DE LA EXPRESIÓN GÉNICA	42
2.3.4 ANÁLISIS DE EXPRESIÓN DE ALGUNOS GENES POR LA TÉCNICA DE QRT-PCR.	43
2.4 CONCLUSIONES	44
IDENTIFICACIÓN DE LNCRNAs Y SUS CAMBIOS TRANSCRIPCIONALES EN LOS MEIOCITOS DE GIRASOL.	45
3.1 INTRODUCCIÓN	45
3.2 MATERIALES Y MÉTODOS	46
3.2.1 EXTRACCIÓN DE RNA, SECUENCIACIÓN Y ENSAMBLADO	46
3.2.2 IDENTIFICACIÓN DE LNCRNA.	46
3.2.3 SECUENCIACIÓN Y MAPEO DE RNAs PEQUEÑOS.	47
3.2.4 DESCRIPCIÓN DE ALGUNAS CARACTERÍSTICAS DE LOS LNCRNAs.	47
3.2.5 QRT-PCR.	47
3.2.6 DISEÑO Y ANÁLISIS ESTADÍSTICO.	48
3.3 RESULTADOS Y DISCUSIÓN	48
3.3.1 GENES DIFERENCIALMENTE EXPRESADOS ENTRE LOS MEIOCITOS DE GIRASOL.	48
3.3.2 IDENTIFICACIÓN DE LNCRNAs EXPRESADOS EN LOS MEIOCITOS DE GIRASOL.	49
3.3.3 EXPRESIÓN DE LNCRNAs EN MEIOCITOS Y EL TRANSCRIPTOMA SOMÁTICO DE GIRASOL.	53
3.3.4 POBLACIONES DE RNAs PEQUEÑOS EN MEIOCITOS DE GIRASOL Y SU RELACIÓN CON LOS LNCRNAs.	54

3.3.5	<i>ELEMENTOS REPETITIVOS PRESENTES EN LOS LNCRNAs DE GIRASOL.</i>	58
3.3.6	<i>VALIDACIÓN DE LAS DIFERENCIAS DE EXPRESIÓN POR LA TÉCNICA DE RT Q-PCR.</i>	60
3.3.7	<i>COMENTARIOS FINALES ACERCA DE LOS LNCRNAs MEIÓTICOS IDENTIFICADOS EN GIRASOL.</i>	61
3.4	CONCLUSIONES	63

CAMBIOS TRANSCRIPCIONALES DE GENES MEIÓTICOS Y ELEMENTOS REGULATORIOS.

4.1	INTRODUCCIÓN	66
4.2	MATERIALES Y MÉTODOS	67
4.1	<i>DELIMITACIÓN DE GENES CON FUNCIÓN MEIÓTICA Y AGRUPAMIENTO DE ESTOS EN CATEGORÍAS FUNCIONALES.</i>	67
4.2	<i>BÚSQUEDA DE GENES CO-EXPRESADOS Y AGRUPAMIENTO EN PATRONES DE EXPRESIÓN.</i>	67
4.2	<i>CONSTRUCCIÓN DE UN MODELO DE PROBABILIDAD DE CO-EXPRESIÓN DE GENES MEIÓTICOS DADA SU CATEGORÍA FUNCIONAL.</i>	68
4.3	RESULTADOS Y DISCUSIÓN	68
4.3.1	<i>DIFERENCIAS EN CATEGORÍAS FUNCIONALES (TÉRMINOS GO) ENTRE MEIOCITOS.</i>	68
4.3.2	<i>PATRONES DE EXPRESIÓN DE GENES MEIÓTICOS.</i>	70
4.3.3	<i>GENES MEIÓTICOS DIFERENCIALMENTE EXPRESADOS.</i>	72
4.3.4	<i>COORDINACIÓN POR ETAPAS DE LA EXPRESIÓN DE GENES MEIÓTICOS.</i>	75
4.3.5	<i>DIFERENCIAS EN ELEMENTOS REGULATORIOS DE LA EXPRESIÓN DE GENES (FACTORES DE TRANSCRIPCIÓN Y ELEMENTOS EN CIS).</i>	77
4.3.6	<i>ANÁLISIS DE CO-EXPRESIÓN DE GENES MEIÓTICOS Y FACTORES DE TRANSCRIPCIÓN</i>	80
4.4	CONCLUSIONES	82

DEL ANÁLISIS TRANSCRIPCIONAL HACIA LA IDENTIFICACIÓN DE NUEVOS GENES IMPORTANTES EN LA MEIOSIS DE PLANTAS.

5.1	INTRODUCCIÓN	83
5.2	MATERIALES Y MÉTODOS	84
5.2.1	<i>SELECCIÓN DE CANDIDATOS Y ADQUISICIÓN DE MUTANTES DE INSERCIÓN DE T-DNA.</i>	84
5.3	RESULTADOS Y DISCUSIÓN	85
5.3.1	<i>CRECIMIENTO Y GENOTIPADO DE LÍNEAS MUTANTES</i>	85
5.3.2	<i>CARACTERIZACIÓN CITOLÓGICA DE MUTANTES POR LA TÉCNICA DE "SQUASH".</i>	87
5.3.2	<i>CARACTERIZACIÓN CITOLÓGICA DE MUTANTES POR LA TÉCNICA DE "SPREADING".</i>	88
5.4	CONCLUSIONES	91

CONCLUSIONES FINALES Y PERSPECTIVAS

REFERENCIAS

MATERIAL SUPLEMENTARIO

Capítulo 1: La meiosis en plantas

1.1 Meiosis: la generación de gametos haploides.

La meiosis es un proceso de división celular, durante el cual luego de una única ronda de duplicación del DNA (ocurrida en la fase S anterior), se producen dos rondas de división celular, resultando en la formación de gametos haploides (Harrison et al., 2010). Las células que llevan a cabo este proceso de diferenciación en gametos a través de la meiosis son denominadas meiocitos. En plantas, estos últimos se encuentran localizados en órganos especializados i.e. anteras y óvulo, destinados a la formación de gametos masculinos y femeninos, respectivamente (Wilson and Yang, 2004). Cada ronda de división celular en la meiosis corresponde con cada una de sus dos etapas, durante las cuales ocurre la segregación de cromosomas homólogos (meiosis I), y la segregación de cromátidas hermanas (meiosis II) (Ma, 2006).

El éxito de la segregación ocurrida durante la meiosis I, depende en parte de que cada par de cromosomas homólogos establezcan una conexión física entre ellos (bivalente), lo que a su vez, generalmente requiere del intercambio recíproco de información entre las cromátidas no hermanas de cada par de homólogos (entrecruzamiento o CO, por sus siglas en inglés), a través del proceso de recombinación homóloga (Mercier et al., 2015). Como resultado de que al menos un evento de entrecruzamiento ocurrió durante la meiosis I, a nivel citogenético se puede observar la formación de una estructura denominada quiasma (Carpenter, 1994). Cabe resaltar que una de las principales consecuencias de dicho intercambio recíproco entre cromosomas homólogos es la creación de nuevas combinaciones de alelos, lo cual, sumado a la segregación independiente de cromosomas no homólogos, permite la generación de variabilidad genética (Cnudde et al., 2006; Hamant et al., 2006a; Wijnker and de Jong, 2008). En la meiosis II por su parte, la segregación de las cromátidas hermanas, se realiza de manera muy similar a como sucede en la mitosis.

Cada una de estas dos grandes etapas meióticas puede ser dividida en cuatro fases: profase, metafase, anafase y telofase (Ma, 2006). La profase I, es la primera y más extensa de las fases; en *A. thaliana* esta fase puede durar hasta 30.3 horas de las 33 horas totales que toma la meiosis (Armstrong et al., 2003). Asimismo, la profase I puede ser dividida en cinco sub-fases (leptoteno, cigoteno, paquiteno, diploteno y diacinesis). Es durante la primera sub-fase, **leptoteno** (Figura 1a), que se da inicio al proceso de recombinación homóloga, dado que es aquí cuando producen las primeras roturas en el DNA. Posteriormente, en **cigoteno** (Figura 1b), tiene lugar la asociación estable entre los cromosomas homólogos (sinapsis), mediada por un complejo protéico denominado complejo sinaptonémico (SC, por siglas en inglés). El SC entonces, permite a los homólogos permanecer unidos durante todo el **paquiteno** (Figura 1c), formando los denominados bivalentes. Asimismo, durante paquiteno, se puede apreciar la aparición de complejos nucleoproteícos, denominados nódulos de recombinación, los

cuales contienen algunas proteínas necesarias para el proceso de recombinación. Aunque el SC se desintegra en la transición de paquiteno a **diploteno** (Figura 1d), los homólogos permanecen unidos, gracias a la estructura denominada quiasma. Finalmente, en **diacinesis** (Figura 1e), la última sub-fase de la profase I, los cromosomas aumentan su condensación y desaparece la envoltura nuclear (Anderson et al., 2001; Hamant et al., 2006a; Pawlowski and Cande, 2005; Zickler and Kleckner, 1998).

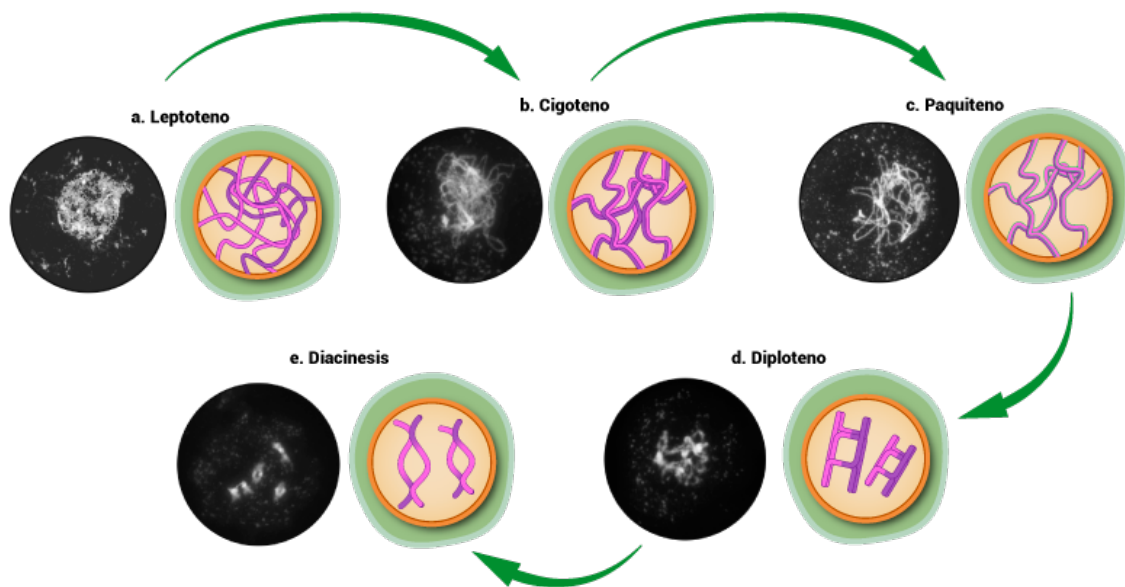


Figura 1. Profase I en *Arabidopsis thaliana*. Se presentan fotografías de cada una de las sub-etapas normales de Profase I de *Arabidopsis thaliana* ($n=5$), acompañadas de una representación esquemática ($n=2$) inspirada en la presentada en (Mercier et al., 2015).

Una vez terminada la profase, los cromosomas encuentran su punto máximo de condensación en la **metafase I** (Figura 2a), durante la cual además estos son orientados hacia el mismo polo, con ayuda de los microtúbulos del huso. En la **anafase I** (Figura 2b), la conexión física entre los homólogos se rompe, permitiendo que cada uno de los cromosomas homólogos migre a polos opuestos de la célula, finalizando así la primera división meiótica, también conocida como segregación reduccional (Zickler and Kleckner, 1998). Durante la meiosis II, los organelos del citoplasma se organizan de manera tal que forman una banda que separa los dos núcleos, los cuales durante la **profase II** (Figura 2c) se observan ligeramente descondensados. Sin embargo, en la **metafase II** (Figura 2d), estos se observan nuevamente condensados. Es en la **anafase II** (Figura 2e), donde finalmente se lleva a cabo la separación de las cromátidas de cada cromosoma (segregación ecuacional). Posteriormente, en la **telofase II** (Figura 2f) los cromosomas migran a cada uno de los polos de las células hijas y la barrera de orgánulos se dispone en forma de cruz, facilitando la división equilibrada del citoplasma (citocinesis); mientras que la cromatina se descondensa y se forman las cuatro envolturas nucleares, dando origen a la **tétrada** (Figura 2g) (Hamant et

al., 2006a; Mercier et al., 2015; Zickler and Kleckner, 1998). De esta manera, al final de la meiosis masculina y una vez ocurrida la citocinesis y dos divisiones mitóticas, cada producto meiótico dará origen a un gránulo de polen (Wilson and Yang, 2004).

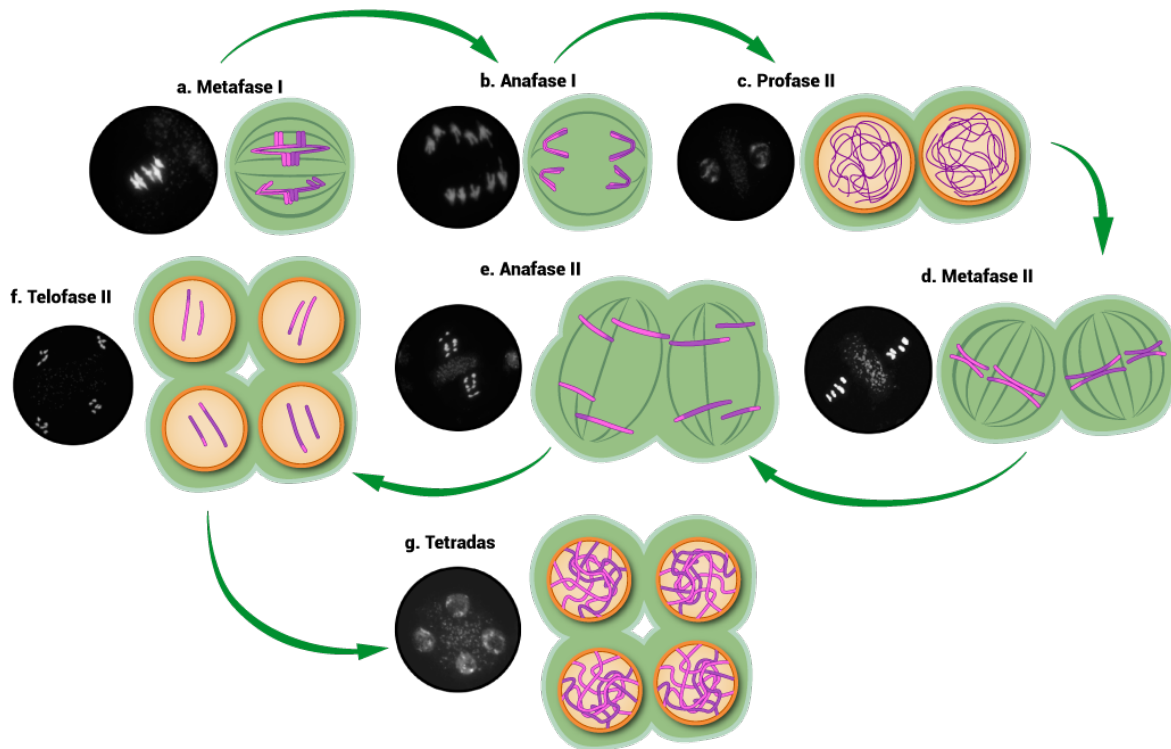


Figura 2. Fases de la meiosis en *Arabidopsis thaliana* posteriores a Profase I. Se presentan fotografías de cada una de las etapas posteriores a Profase I de *Arabidopsis thaliana* (n=5), acompañadas de una representación esquemática (n=2) inspirada en la presentada en (Mercier et al., 2015).

Son varias las razones que han hecho que las plantas sean excelentes modelos para el estudio de la meiosis. En primer lugar, estas suelen tener genomas grandes, lo que se traduce en grandes cromosomas, fácilmente visibles en el microscopio con técnicas citogenéticas, relativamente simples (Mercier et al., 2015). Asimismo, a diferencia de algunos modelos animales, en plantas la pérdida de función de varios genes involucrados en las rutas de reparación y recombinación homóloga no es letal, lo que permite estudiar a mayor detalle su función (Schuermann et al., 2005). Por otro lado, la amplia disponibilidad de colecciones de mutantes en plantas modelo, así como el hecho de que la meiosis ocurre tarde en el desarrollo, aumentando la probabilidad de que los cambios durante este puedan ser transmitidos y observados en la siguiente generación, son algunos de los motivos para los cuales estas son especialmente adecuadas para este tipo de análisis genéticos (Mercier et al., 2015; Schuermann et al., 2005). Finalmente, el estudio de la meiosis en plantas podría tener aplicaciones agronómicas relacionadas con el mejoramiento de la fertilidad de cultivo, la producción de semillas, así como en el incremento de la eficiencia del proceso de

fitomejoramiento (Mercier et al., 2015; Ronceret and Vielle-Calzada, 2015; Wijnker and de Jong, 2008; Wijnker and Schnittger, 2013; Wijnker et al., 2012). Vale la pena mencionar, que dada su alta disponibilidad y la facilidad de acceder a ellos, la mayoría de estudios de meiosis en plantas, han sido realizados en meiocitos masculinos (Hamant et al., 2006a).

1.2 Recombinación homóloga durante la meiosis.

La recombinación homóloga (HR), cumple importantes funciones biológicas como la reparación del DNA, ante la aparición de dobles roturas (“double strand breaks”, DSBs) u otros tipos de daño; el mantenimiento de los telómeros, el rescate de la horquilla de replicación del DNA, y la correcta segregación de los cromosomas durante la meiosis (San Filippo et al., 2008; Sung and Klein, 2006). Aunque el mecanismo y proteínas que intervienen en la recombinación homóloga somática y la meiótica son muy similares, esta última se diferencia de la primera, en que la aparición de DSBs en todos los cromosomas es programada, y en que en la meiótica se favorece el intercambio de secuencias entre cromosomas homólogos más que entre cromátidas hermanas (Sung and Klein, 2006).

Como resultado del proceso de HR pueden producirse entrecruzamientos y no-entrecruzamientos (“crossover” –CO- y “non-crossover” –NCO-, respectivamente), dependiendo de si existe intercambio recíproco (CO) o no (NCO) entre los cromosomas. De estos dos, sólo los CO se manifiestan como quiasmas. Además, la HR puede resultar en fenómenos de conversión génica (GC, por sus siglas en inglés), la cual se puede definir como el fenómeno en que uno o más marcadores genéticos, presentan patrones de segregación mendeliana entre los productos de una única célula meiótica (Bishop and Zickler, 2004).

Aunque aún existen debates y dudas sobre algunos puntos, la Figura 3, presenta un modelo de cómo sucede la recombinación homóloga durante la meiosis, y a continuación se describirán brevemente algunos de los subprocesos que esta implica.

1.2.1 Formación y procesamiento de las dobles roturas.

La HR en meiosis, empieza con la generación de dobles roturas por parte de la proteína Spo11 (Figura 3a), una topoisomerasa ampliamente conservada (Keeney et al., 1997). En *A. thaliana*, se han descrito tres homólogos para esta proteína, de los cuales solo dos de ellos (Spo11-1 y Spo11-2) tienen una función en meiosis (Grelon et al., 2001; Hartung and Puchta, 2000). En plantas, cinco genes adicionales a Spo11, han sido identificados como importantes para la formación de las DSBs (Mercier et al., 2015). ATPRD1 quien parece ser un homólogo funcional de la proteína MEI1 de ratón, la cual es necesaria para la iniciación de la meiosis (De Muyt et al., 2007); ATPRD2 y ATPRD3 las cuales están poco conservadas y cuya función no está completamente elucidada (De Muyt et al., 2009); DFO una proteína específica de plantas, sin ningún dominio funcional conocido (Zhang et al., 2012) y finalmente la proteína CRC1 de arroz, la cual al parecer actúa como un complejo para

favorecer la formación de DSBs, e interviene en el reclutamiento de proteínas necesarias para el apareamiento de los homólogos (Miao et al., 2013).

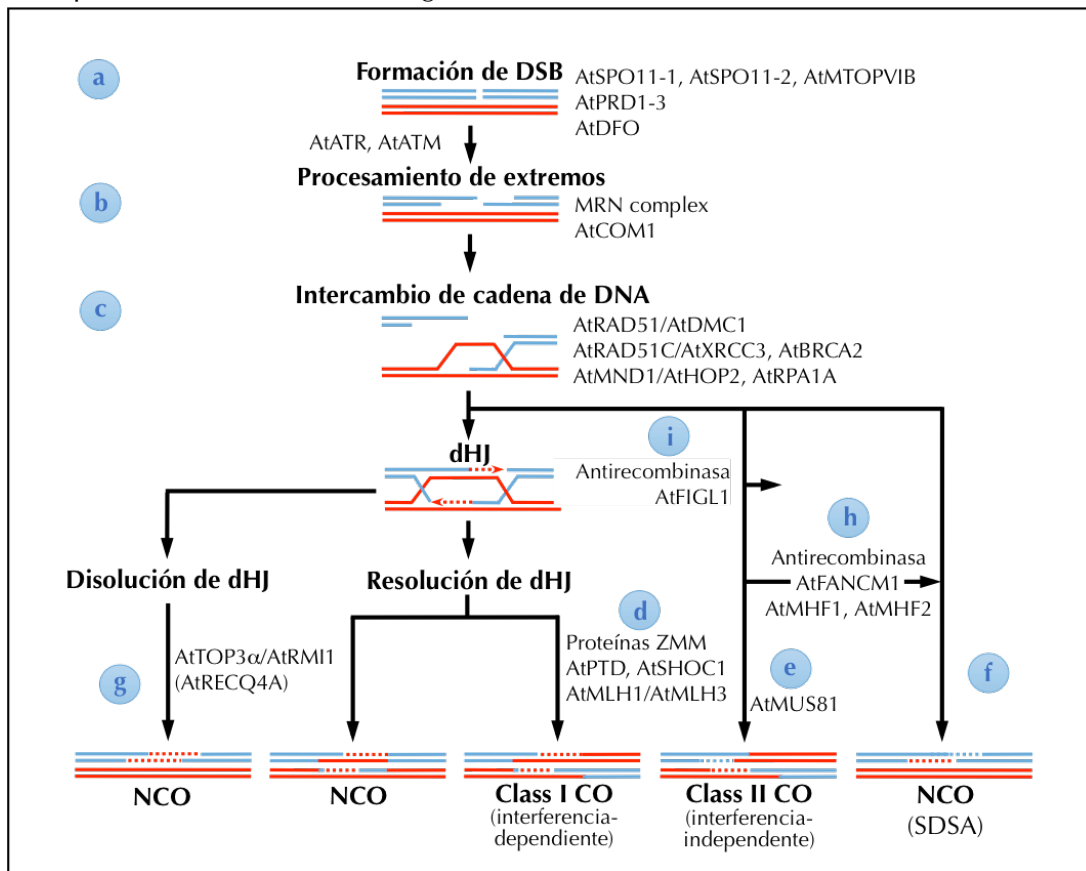


Figura 3. Modelo de recombinación meiótica en *Arabidopsis thaliana*.

Modificado a partir del de (Pradillo et al., 2014).

Se ha identificado que, posterior a las dobles roturas, ocurren modificaciones en la histona H2AX, la cual es fosforilada en un proceso donde parecen estar involucradas las proteínas ATM y ATR (Culligan and Britt, 2008). Esta versión fosforilada de la histona (γ -H2AX), al parecer es importante para el reclutamiento de factores necesarios para la reparación de la DSB (Turinetto and Giachino, 2015). Posteriormente, los extremos 5' resultantes de la DSB, son procesados (Figura 3b), gracias a la participación de los complejos Mre11-Rad50-Xrs2/Nbs1 y Com1/Sae2/CtIP (Osman et al., 2011). Dichos complejos no son específicos de meiosis, puesto que también participan en la reparación de DSBs somáticas por HR (Gallego et al., 2001; Waterworth et al., 2011). El procesamiento de los extremos 5', permite la liberación de la proteína Spo11, junto con un pequeño oligo de DNA (Pan et al., 2011). Recientemente, la utilización de técnicas que permiten la inmunoprecipitación de SPO11, y posterior secuenciación masiva de estos oligos unidos (Spo11 oligos), ha permitido la identificación de las zonas donde tienen lugar las dobles roturas en el genoma (Hwang and Hunter, 2011).

1.2.2 Intercambio de la cadena sencilla de DNA.

Como consecuencia del procesamiento de las DSBs, se generan extremos de DNA de cadena sencilla (ssDNA) 3', sobre los cuales se realiza el ensamblado de complejos nucleoproteicos presinápticos que incluyen las recombinasas de DNA, DMC1 y RAD51. DMC1 es la única que parece ser exclusiva de meiosis, mientras que RAD51 también participa en eventos de reparación somática del DNA a través de HR (Osman et al., 2011). Se ha propuesto que estas dos recombinasas interactúan de forma asimétrica, facilitando que una de las cadenas realice la invasión inicial sobre la doble hélice de la cromátida del cromosoma homólogo (Figura 3c), mientras que la segunda es recapturada posteriormente (Pradillo et al., 2014; Shinohara and Shinohara, 2004). Sin embargo, aún hay controversia sobre la contribución real de cada una de estas proteínas al proceso de recombinación.

Por un lado, se ha sugerido que DMC1 se une al ssDNA antes que RAD51 encargándose de la búsqueda inicial de homología, con la posterior colaboración de RAD51. Esta primera recombinasa, sería a su vez la principal responsable de la reparación de la DSB, utilizando el cromosoma homólogo como molde, mientras que RAD51 además podría reparar las DSBs en ausencia de DMC1, utilizando como molde la cromátida hermana (Pradillo et al., 2012, 2014). En un modelo alternativo, se propone que gracias a la actividad de los genes RPA1 y ATR, se impide la unión de DMC1 a la molécula de ssDNA, pero no la de RAD51, la cual se uniría a uno de los extremos de ssDNA. Posteriormente, gracias a la desaparición de ATR, se realizaría la unión de DMC1 en el filamento opuesto, al mismo tiempo que ocurren modificaciones en las proteínas del complejo sinaptonémico, que dificultarían que la invasión del ssDNA tenga lugar hacia la cromátida hermana (Kurzbaue et al., 2012). Finalmente, Da Ines et al. (Da Ines et al., 2013), afirman que la actividad de RAD51 no sería necesaria para la HR en meiosis.

En Arabidopsis, se ha identificado un único gen que codifica para DMC1, sin embargo para RAD51 se han descrito seis parálogos (RAD51, RAD51B, RAD51C, RAD51D, XRCC2, XRCC3) (Osman et al., 2011). Se piensa que estos parálogos de RAD51, fueron originados por eventos de duplicación y adquirieron nuevas funciones durante la evolución, de manera que actualmente estos no son funcionalmente redundantes y actúan en conjunto en la reparación de las DSBs (Bleuyard et al., 2006; Pradillo et al., 2014). El producto de la interacción de estos seis parálogos son dos complejos: BCDX2 (RAD51B/RAD51C/RAD51D/XRCC2) y CX3 (RAD51C/XRCC2). Aunque ambos complejos están involucrados con la reparación del DNA, sólo CX3 interviene en la recombinación meiótica, presuntamente mediando la unión y/o actividad de RAD51 (Bleuyard and White, 2004; Bleuyard et al., 2005; Pradillo et al., 2014).

Durante esta etapa del proceso de recombinación, también es necesaria la actividad de algunas proteínas accesorias. MND1 y AHP2, participan formando un complejo necesario para dirigir la reparación del DNA mediada por RAD51, utilizando como molde la

cromátida hermana (Uanschou et al., 2013). Asimismo, BRCA2A y BRCA2B, interaccionan físicamente con RAD51 y DMC1, mediando su reclutamiento (Dray et al., 2006; Seeliger et al., 2012; Siaud et al., 2004). También RPA ha sido identificada como importante durante esta etapa, ya que participa eliminando posibles estructuras secundarias que forma el ssDNA (Osman et al., 2009; Ribeiro et al., 2015).

1.2.3 Vías para la formación de COs y NCOs.

Una vez establecido el intercambio de cadenas, se forma una estructura intermediaria de DNA denominada “Holliday junction” (Hj) a partir de la cual el proceso de recombinación homóloga continúa, dando origen a CO y NCO (Mercier et al., 2015; Osman et al., 2011). Hasta el momento, se han identificado dos rutas que dan origen a dos tipos COs: CO de tipo I y CO de tipo II. Para la formación de CO de tipo I (Figura 3d), es necesaria la participación de un gran número de proteínas llamadas colectivamente ZMM (ZIP1, ZIP2, ZIP3, ZIP4, MSH4, MSH5 y MER3). En Arabidopsis se ha observado que la mutación en los ortólogos de dichas proteínas conlleva una disminución dramática del número de COs, a tal punto que se estima que esta ruta de formación de COs es responsable de entre el 85% y el 90% de los COs en esa planta modelo (Mercier et al., 2015). Existe además una proteína ZMM específica de plantas, necesaria para la formación de CO tipo I, denominada PTD (Lu et al., 2014; Ma et al., 2006). Asimismo, dos proteínas adicionales no-ZMM, intervienen en la formación de COs por esta primera ruta, MLH1 y MLH3. Según Dion et al. (Dion et al., 2007), MLH1, además de intervenir en el proceso de HR en meiosis, también es importante para llevar a cabo la recombinación mitótica. Respecto a MLH3, se ha propuesto que su papel principal es en la resolución de la Hj hacia CO (Jackson et al., 2006).

La segunda vía para la formación de CO (Figura 3e), aún no está completamente entendida, pero se sabe que necesita de la participación de la endonucleasa MUS81. Las plantas de Arabidopsis con mutaciones de pérdida de función de MUS81, presentan una reducción de aproximadamente el 10% de COs. Sin embargo, en plantas dobles mutantes para MUS81 y proteínas de la ruta de ZMM aún se observa formación de COs, lo cual sugiere que puede existir otra vía y/o genes responsables de la formación de COs aún no caracterizada (Berchowitz et al., 2007; Higgins et al., 2008; Mercier et al., 2015).

Existen diferencias importantes entre las dos vías de generación de COs. Si bien hay distintas proteínas implicadas en cada una de ellas, también se ha observado que para la formación de COs de tipo II no se precisa la formación de la estructura intermediaria Hj (Berchowitz and Copenhaver, 2010; Pradillo et al., 2014). Asimismo, estos dos tipos de COs, difieren en la manera como se ubican a lo largo del cromosoma. Mientras que los CO de tipo II se distribuyen aleatoriamente a lo largo del cromosoma (Berchowitz et al., 2007), los CO de tipo I, se ubican a una distancia entre sí mayor a la esperada por el azar (Osman et al., 2011). Lo anterior sugiere que por algún mecanismo, aún no completamente elucidado, la ocurrencia de un CO de tipo I interfiere con la ocurrencia de un CO (tipo I o tipo II), en su cercanía (Anderson et al., 2014). Así pues, se cree que los COs de tipo I, generados a través

de la vía que incluye las proteínas ZMM, son sensibles a interferencia; mientras que, los CO de tipo II, producidos con la participación de MUS81, son insensibles a interferencia (Berchowitz and Copenhaver, 2010; Mercier and Grelon, 2008; Osman et al., 2011).

Como se menciono anteriormente, la evidencia citogenética de que al menos un evento de entrecruzamiento (CO) tuvo lugar, es la formación de un quiasma. En *A. thaliana*, se estima que la media de quiasmas por célula se encuentra entre 7 y 14, dependiendo del ecotipo analizado (López et al., 2012; Sanchez-Moran et al., 2002). No obstante, al inicio del proceso de recombinación se producen alrededor de 200 DSBs (Higgins et al., 2008; Kurzbauer et al., 2012; Mercier et al., 2015), lo cual sugiere que la mayoría de DSBs deberían ser reparadas por mecanismos que conllevan a la formación de NCO (Sanchez-Moran et al., 2007). Sin embargo, el número de NCO detectados por meiosis en esta planta, se encuentra alrededor de 1 y 6 (Lu et al., 2012; Sun et al., 2012; Wijnker et al., 2013). Según Mercier et al. (Mercier et al., 2015), la baja detección de NCO, podría ser consecuencia de que los eventos de reparación de las DBSs sean realizados utilizando como molde la cromátida hermana (lo que no dejaría rasgo genético); o bien porque las metodologías actuales de detección de NCO tienen un bajo nivel de resolución.

Gracias a estudios realizados en levadura, se estima que la ruta principal para la generación de NCO en meiosis es la de “hibridación dependiente de síntesis de hebra” (SDSA, por sus siglas en inglés) (McMahill et al., 2007). En esta vía de reparación de DSBs no hay formación de H_j (Figura 3f). En cambio, la cadena de ssDNA invasora es disociada antes de la segunda captura y luego re-hibridada en el extremo 3' del otro extremo de la rotura, donde posteriormente se produce la síntesis y ligación del DNA (Sun et al., 2012). Sin embargo, existen otros mecanismos de formación de NCO en meiosis a partir de la resolución o disolución de la H_j intermediaria (Knoll and Puchta, 2011) (Figura 3g). Por ejemplo, para la generación de NCO por disolución de la H_j, es necesaria la participación de un complejo proteico (denominado RTR), constituido por las proteínas RECQ4A, TOP3 α y BLAP75 (también llamada RMI1) (Chelysheva et al., 2008; Hartung et al., 2008; Knoll et al., 2014). Asimismo, la helicasa RECQ4B también ha sido identificada como importante para la formación de NCO (Séguéla-Arnaud et al., 2015). La helicasa FANCM y sus cofactores MHF1 y MHF2, actúan previniendo la formación de CO, a través de un mecanismo diferente al propuesto para los genes anteriormente mencionados (Crismani et al., 2012; Girard et al., 2014) (Figura 3h). Por último, la AAA-ATPase FIDGTIN-L, que se propone actúa en el paso de invasión de cadena mediado por RAD51/DMC1, también ha sido descrita con actividad anti-CO (Girard et al., 2015) (Figura 3i). En plantas, estos últimos genes que favorecen la formación de NCOs (anti-CO), parecen ser blancos promisorios para incrementar la frecuencia de recombinación en programas de fitomejoramiento (Séguéla-Arnaud et al., 2015).

1.3 Mecanismos de regulación de la recombinación homóloga y formación de entrecruzamientos en meiosis.

La formación de CO y su control involucra varias rutas que presentan variaciones de un organismo modelo a otro (Cromie and Smith, 2007). Desde los años 70s, se tenía evidencia de que los pasos necesarios para llevar a cabo la recombinación se encontraban bajo un control genético de naturaleza compleja (Simchen and Stamberg, 1969). Es quizás dicha complejidad la que ha hecho que alrededor de cinco décadas después la regulación de la HR no esté completamente entendida (Hamant et al., 2006b; Ma, 2006; Zhou and Pawlowski, 2014a). De hecho, la diversidad de mecanismos es tal, que se han descrito tanto elementos que actúan de forma *cis* como *trans*, alterando la recombinación de manera global o regional (Esch and Horn, 2008).

Una hipótesis es que sitios discretos en el DNA regulan la distribución de los eventos de recombinación meiótica a lo largo del genoma. Lo anterior gracias a que se han identificado puntos donde los eventos de recombinación ocurren a tasas más altas que las de todo el genoma (“hotspots” de recombinación), los cuales a su vez parecen tener motivos conservados y marcas epigenéticas características (Wahls and Davidson, 2010). Por otro lado, se ha observado que la estructura de la cromatina también influye de manera importante en la localización y número de DSBs. En levadura, una cromatina más “abierta” (descondensada) y libre de nucleosomas, como la presente en regiones promotoras, parece ser más propicia para la actividad de Spo11 (Berchowitz et al., 2009; Pan et al., 2011). Asimismo, en plantas, también se ha descrito que cambios en la estructura de la cromatina podrían alterar no sólo la formación de DSB, sino también el panorama de eventos de recombinación (Mirouze et al., 2012; Rosa et al., 2013; Shilo et al., 2015).

El fenómeno de interferencia entre CO, también puede ser considerado como un mecanismo de regulación de la recombinación, puesto que limita el número posible de eventos que tendrían lugar. Aunque no es claro porqué se establece dicha interferencia, un modelo aceptado es el de “estrés mecánico”. Este modelo propone que los cromosomas sufren rondas programadas de expansión y contracción, lo cual genera cierto tipo de estrés mecánico que sería aliviado con la ocurrencia de un CO, inhibiendo la ocurrencia de otro CO en las regiones de DNA adyacentes (Kleckner et al., 2004; Zhang et al., 2014a). Sin embargo, se ha observado que al realizar mutaciones simultáneas de genes con actividad anti-CO (descritos anteriormente), se puede incrementar hasta nueve veces la frecuencia de CO sin afectar la segregación cromosómica. Lo anterior supone que más allá de las restricciones mecánicas, la limitación en el número de CO podría estar definida por los costos a largo plazo de la recombinación (Séguéla-Arnaud et al., 2015).

El número de CO que ocurren en la HR, también parece estar determinado por un proceso que ha sido denominado “homeostasis de CO”. Este fenómeno supone que el número de COs que tienen lugar durante la meiosis se mantiene constante a expensas de cambios en la

relación CO/NCO (Phadnis et al., 2011). Dicho fenómeno fue descrito por primera vez en levadura, en donde se analizaron mutantes hipomorfos de Spo11, los cuales mostraban una reducción variable en el número de DSBs pero un número constante de CO (Martini et al., 2006). Dado que en mutantes para los cuales se afecta la interferencia de CO (e.g. Zip2 y Zip4) la homeostasis también se reduce, se plantea la hipótesis de que ambos fenómenos podrían tener una vía de regulación común (Martinez-Perez and Colaiácovo, 2009).

La HR también parece ser regulada tanto a nivel transcripcional como traduccional. Se han descrito modificaciones post-traduccionales reversibles (ubiquitinación, fosforilación y sumolización) de proteínas importantes para la HR, que delimitan su actividad (Heyer et al., 2010). Por otro lado, al monitorear la actividad traduccional de *S. cerevisiae* durante la meiosis, se observó que durante este proceso existe un control traduccional generalizado, que a su vez refleja la existencia de regulación a nivel transcripcional (Brar et al., 2012). Además, en este organismo se ha evidenciado que los genes necesarios para la meiosis y HR, son expresados en puntos específicos en una vía temporalmente muy regulada (Mata et al., 2002; Vershon and Pierce, 2000). Sin embargo, los factores que median esta cascada de transcripción temporalmente regulada, se encuentran muy poco conservados (Andreuzza et al., 2015). Así pues, el entendimiento de cómo es regulada la expresión génica durante la meiosis en plantas es limitado (Andreuzza et al., 2015; Zhou and Pawlowski, 2014a). Por tanto, la regulación transcripcional de la HR en meiosis es aún más limitada.

Es importante resaltar que la regulación transcripcional durante la meiosis presenta una limitación adicional impuesta por el estado de condensación de los cromosomas, lo cual limita el acceso a la cromatina. Por lo anterior, la red de regulación debe coordinar entonces la ocurrencia de los cambios en estados de la cromatina propios de la meiosis, así como los necesarios para permitir la expresión de genes. En esto último, se propone que podría ser importante la participación de RNAs no codificantes tanto largos como pequeños (Zhou and Pawlowski, 2014a).

Los estudios transcriptómicos son un buen punto de partida para elucidar los mecanismos que intervienen en la regulación de la expresión génica en meiosis (Zhou and Pawlowski, 2014a). En los últimos años, ha sido posible obtener transcriptomas de meiocitos masculinos aislados de las plantas modelo *Arabidopsis* (Chen et al., 2010a; Yang et al., 2011) y maíz (Dukowic-Schulze et al., 2014). Dichos estudios han permitido evidenciar que los meiocitos tienen una alta actividad transcripcional; alrededor del 50 al 60% del total de los genes de la planta se expresan durante la meiosis. Además, esta se caracteriza por contener un alto número de secuencias no identificadas, elementos transponibles y genes de origen mitocondrial (Chen et al., 2010a; Dukowic-Schulze et al., 2013, 2014; Zhou and Pawlowski, 2014a). Además, estos estudios han permitido identificar promotores con actividad preferencial en meiocitos (Li et al., 2012), así como nuevos genes candidatos importantes para la meiosis, e.g. FEHLSTART (Li et al., 2015).

Son pocos los genes de plantas que se han descrito con una posible función regulatoria de la expresión de genes relacionados a la meiosis. En Lirio, una proteína de la familia GRAS llamada LISCL, ha sido asociada a la regulación de la expresión de genes en meiosis, aunque su papel aún no ha sido completamente elucidado (Morohashi et al., 2003). Mientras tanto, en maíz se ha observado que alelos mutantes del gen AMEIOTIC1, presentan diferencias en el perfil transcripcional de genes meióticos, lo cual podría sugerir que este actúa como un modulador de la expresión de genes importantes para la meiosis y progresión de la profase I (Nan et al., 2011; Pawlowski et al., 2009). Por otro lado, el gen meiótico MMD1 de *A. thaliana* posee un dominio PHD, comúnmente encontrado en proteínas con actividad en regulación transcripcional y proteínas asociadas a complejos de remodelamiento de cromatina, por lo cual se propone que podría ser un regulador transcripcional (Yang et al., 2003). Posteriormente, Andreuzza et al. (Andreuzza et al., 2015), demostraron que, en efecto, MMD1 participa en la regulación de la expresión de los genes TDM1 y JAS, importantes para las transiciones del ciclo celular y la organización del huso en meiosis. Además, gracias a que la expresión de este gen está limitada a una etapa específica de la meiosis (diploteno), los autores proponen que la transcripción de genes meióticos en plantas también podría estar temporalmente regulada.

La Figura 4 presenta un resumen de los mecanismos descritos anteriormente, que podrían estar involucrados en la regulación de la recombinación homóloga y formación de entrecruzamientos en meiosis.

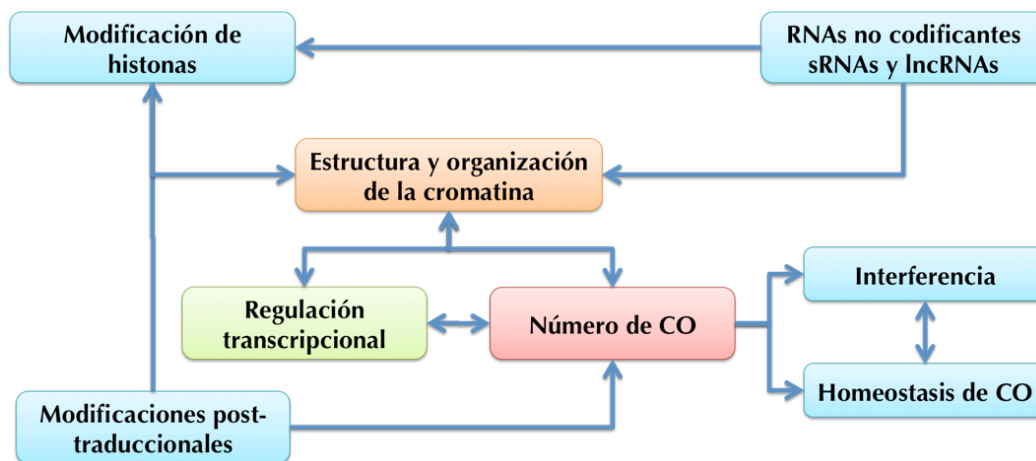


Figura 4. Propuesta de la relación de mecanismos relacionados con la regulación de la recombinación homóloga en meiosis. El presente diagrama retoma puntos de la visión de Zhou y Pawlowski (Zhou and Pawlowski, 2014a) sobre factores que podrían afectar al transcriptoma meiótico y que consideramos también son importantes para la regulación de la HR. El número de CO que tienen lugar durante la meiosis puede estar afectado por los fenómenos de interferencia y homeostasis, los cuales a su vez parecen tener una vía de regulación común. Además, se han descrito modificaciones postraduccionales de proteínas involucradas en la HR, como un mecanismo de regulación de la misma. Dentro de estas

modificaciones se destacan las que sufren las histonas, lo cual puede tener un efecto sobre la estructura y organización de la cromatina a través de una vía en la que además participarían los RNAs no codificantes. Finalmente, la estructura de la cromatina puede a su vez tener un efecto sobre la transcripción de genes en meiosis, o bien sufrir cambios gracias a la transcripción de elementos que median dicha estructura, como por ejemplo, los RNAs no codificantes.

1.4 Variabilidad de la recombinación homóloga en meiosis.

La variabilidad de la tasa de recombinación y de la frecuencia de quiasmas es influenciada tanto por factores genéticos como ambientales (Esch and Horn, 2008). Dentro de los factores ambientales que pueden afectarla, la temperatura parece ser uno de los más importantes. Se han descrito fallos en la meiosis tanto a altas como a bajas temperaturas en distintos procesos que van desde la sinapsis y formación de CO, hasta la formación de gametos no reducidos (Bomblies et al., 2015). El estatus nutricional, el ataque de patógenos y la aplicación de agroquímicos, también han sido descritos como factores que afectan la meiosis y formación de CO (Bomblies et al., 2015; Esch and Horn, 2008; De Storme and Geelen, 2014). Una hipótesis de cómo esta variedad de estímulos afectan la tasa de recombinación es planteada alrededor de la producción de ácido abscísico, pues esta hormona se produce en respuesta al estrés de la planta y se ha encontrado que influencia la tasa de recombinación (De Storme and Geelen, 2014).

Algunos rasgos de la estructura cromosómica son también responsables de las diferencias en la tasa de formación de CO, tanto a nivel local (regiones del cromosoma), como global. Por ejemplo, la región centromérica presenta una baja formación de CO y NCO, mientras que en la región telomérica aunque existe reducción del número de CO, los niveles de NCO son equivalentes a los de otras regiones cromosómicas (Serrentino and Borde, 2012). Asimismo, algunos cambios en la estructura del cromosoma, tales como: deleciones, formación de bloques de heterocromatina y presencia de segmentos homeólogos o de introgresión; pueden llevar a cambios en la formación de COs (Wijnker and de Jong, 2008).

La tasa de recombinación homóloga también puede variar de acuerdo al genotipo. En *A. thaliana* se ha descrito que la variación natural en el número de quiasmas puede implicar diferencias de hasta un 20% más (o menos) quiasmas, entre las accesiones más contrastantes para este rasgo (López et al., 2012; Sanchez-Moran et al., 2002). Asimismo, en esta planta se observaron diferencias en la tasa de recombinación en función de la heterocigosidad y el fondo genético (Barth et al., 2001). En maíz, gracias al análisis de mapas genéticos de alta densidad de marcadores, se describieron diferencias intraespecíficas en la tasa de recombinación a nivel genómico, cromosómico e intracromosomal (Bauer et al., 2013). A su vez, en soya, se encontró que cultivares altamente seleccionados presentaban una mayor tasa de recombinación que líneas recombinantes no seleccionadas de la misma población (Stefaniak et al., 2006).

El proceso de domesticación, también es un factor que influye en la tasa de recombinación en plantas (Esch and Horn, 2008). Realizando comparaciones pareadas de un total de 196 especies de plantas vasculares, que incluían 46 especies domesticadas con su progenitor, se identificó que el proceso de domesticación generalmente incrementa la tasa de recombinación (Ross-Ibarra, 2004). Este incremento, podría ser una estrategia para aumentar la diversidad genética en las plantas domesticadas, la cual se ve limitada por los cuellos de botella o deriva que tienen lugar durante la domesticación. Además, dicho incremento podría favorecer la aparición de relaciones epistáticas entre dos alelos benéficos (Otto and Barton, 2001; Ross-Ibarra, 2004)

1.5 Aproximaciones metodológicas del estudio de la meiosis en plantas.

Tradicionalmente, la descripción de la función de genes relacionados con meiosis en plantas ha sido realizada a través de aproximaciones citogenéticas y genéticas (Mercier and Grelon, 2008). Dentro de estas últimas, se incluyen tanto métodos de genética directa (*forward genetics*) como de genética reversa (*reverse genetics*). Son muchos los genes meióticos que han sido caracterizados en plantas a través de genética directa, realizando el tamizaje de colecciones de mutantes cuyo fenotipo evidencia problemas de fertilidad (e.g. reducción en el tamaño de la siliqua en *A. thaliana*) (Mercier et al., 2001). Por ejemplo, esta estrategia ha permitido la caracterización del gen MEL1 en *Oryza sativa* (Nonomura et al., 2007), SDS en *A. thaliana* (Azumi et al., 2002) o PHS1 en *Zea mays* (Pawlowski et al., 2004). En la genética reversa se realiza la selección previa de genes candidatos, y posteriormente se aíslan los mutantes para dichos genes y se caracterizan sus fenotipos. Estos genes candidatos pueden ser seleccionados por homología con genes descritos en otros organismos con función en meiosis, o bien, revisando la expresión diferencial de estos en órganos sexuales (Hamant et al., 2006a; Mercier and Grelon, 2008). CDC45 de *Arabidopsis thaliana* (Stevens et al., 2004) y RAD51A1 de *Zea mays* (Franklin et al., 1999), son ejemplos de genes que han sido caracterizados a través de este tipo de aproximación. Para plantas no modelo, en las cuales no existe una colección de mutantes disponible o suficientemente caracterizada, las aproximaciones de genética reversa parecen ser la mejor manera para realizar la búsqueda de nuevos genes, y este es uno de los puntos que resalta la importancia del análisis de expresión diferencial en células meióticas.

Los niveles de expresión génica de varios genes a la vez pueden ser medidos fundamentalmente de dos formas: microarreglos o a través de la secuenciación de las moléculas de RNA (RNA-Seq). Esta última metodología, tiene mayor sensibilidad, rango dinámico y menor variación que la técnica de microarreglos (Oshlack et al., 2010). Para el RNA-seq, es necesario el aislamiento de RNA (mRNA purificado) y su conversión a cDNA por transcripción reversa. Posteriormente, dicho cDNA es secuenciado e identificado en relación a su gen de origen, para finalmente contar el número de secuencias que pertenecen

a este e inferir su nivel relativo de expresión (Wang et al., 2009). Si el genoma del organismo no se encuentra disponible, una fase intermedia del análisis de RNA-seq consiste en el ensamblaje *de novo* del transcriptoma y la identificación de los genes reconstruidos (Strickler et al., 2012). Aunque la medición de los niveles de expresión de genes relacionados con meiosis ha sido realizada exitosamente a través de microarreglos (Crismani et al., 2006; Deveshwar et al., 2011) y RNA-seq (Chen et al., 2010a; Dukowic-Schulze et al., 2014; Yang et al., 2011), esta última parece más adecuada si se pretende la descripción de genes nuevos.

1.3 El girasol como modelo de estudio de la meiosis en plantas.

El girasol (*Helianthus annuus* L.), es una especie diploide ($2n=34$) de interés agronómico, ya sea para la producción de aceite y proteína de origen vegetal, o como flor ornamental (Seiler and Jan, 2010). Esta planta además ha sido utilizada como organismo modelo para el estudio de procesos de especiación por hibridación, introgresión y arquitectura genética (Rieseberg et al., 1993, 1995). Para el estudio de la meiosis, *H. annuus* presenta una ventaja fundamental, determinada por la posibilidad de aislar poblaciones puras de meiocitos masculinos en etapas claramente diferenciadas. Lo anterior, gracias a que la arquitectura de su inflorescencia permite el desarrollo de varias flores con diferentes edades, las cuales se organizan, de la más joven a la más vieja, desde el centro a la periferia del capítulo (Jovanka, 2004).

Un estudio previo, identificó diferencias significativas en la tasa de recombinación homóloga, determinada mediante el análisis de las configuraciones cromosómicas en diacinesis, entre plantas de girasol domesticado de la línea pública HA89 (*Helianthus annuus* L. var. *macrocarpus*) y plantas de girasol silvestre Ac-8 (*H. annuus* L ssp. *texanus*). Además, se observaron diferencias en la tasa de recombinación de estos dos genotipos de girasol, respecto a plantas de la F1 de su cruce, por lo cual se plantea que dichas diferencias podrían ser el resultado de mecanismos de regulación multigénicos (Rodríguez et al., 2007). La Figura 5, presenta algunas de las principales características de estos tres genotipos, los cuales serán el material de estudio del presente trabajo

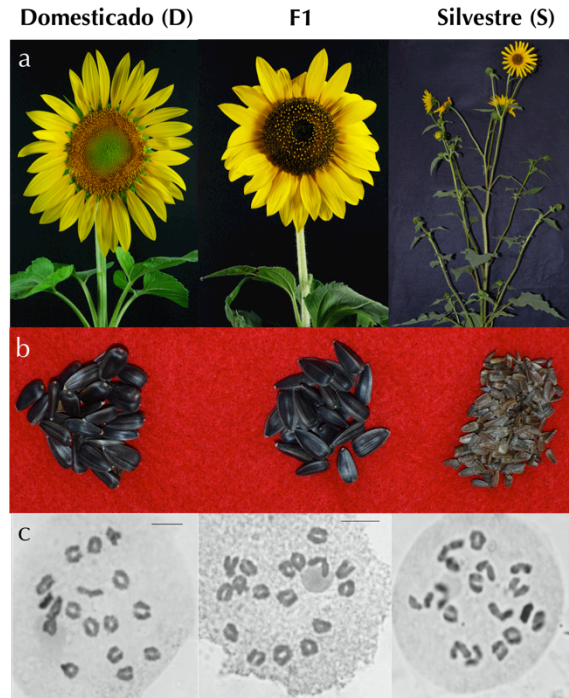


Figura 5. Diferencias entre las líneas de girasol utilizadas en el desarrollo de este estudio.

(a) Inflorescencias de los 3 genotipos de girasol: mientras que los genotipos Domesticado y F1 se caracterizan por tener un solo capítulo, el genotipo Silvestre presenta múltiples ramificaciones con varios capítulos. (b) Diferencias en el tamaño del aquenio de los 3 genotipos. (c) Configuraciones cromosómicas en Diacinesis, tomado de (Rodríguez et al., 2007). Se observan 17 bivalentes, donde el girasol Domesticado tiene mayor proporción de bivalentes en anillo, y el Silvestre en cadena, indicativos de la formación de al menos dos y un quiasma respectivamente. Para dar un total de brazos cromosómicos con quiasmas de 29.70, 28.10 y 25.60 para los genotipos Domesticado, F1 y Silvestre, respectivamente.

Por tanto, el presente trabajo propone estudiar el panorama transcripcional de meiocitos masculinos de girasol de los tres genotipos anteriormente mencionados, con el propósito de contribuir a la comprensión de la meiosis en plantas. Los capítulos siguientes presentan y discuten los resultados obtenidos, así como algunas de las preguntas que aún quedan por resolver.

CAPÍTULO 2: El transcriptoma de los meiocitos de girasol.

2.1 Introducción

Como se menciona en el capítulo anterior, una estrategia para aproximarse al entendimiento de los procesos que tienen lugar durante la meiosis en plantas, así como para identificar nuevos genes importantes para esta, es el análisis del perfil transcripcional de anteras y meiocitos. En trigo, el perfil transcripcional permitió la identificación de 1,350 transcritos regulados temporalmente durante la meiosis, muchos de los cuales parecen estar relacionados con funciones propias de la meiosis, e.g. formación del complejo sinaptonémico y recombinación. Sin embargo, el 50% de aquellos genes que presentaban cambios dramáticos en su expresión no mostraron similitud con secuencias disponibles en las bases de datos de nucleótidos del NCBI y TIGR (Crismani et al., 2006). De igual forma, en arroz, cerca del 50% de los genes que tenían expresión específica en la antera no pudieron ser anotados (Deveshwar et al., 2011). El análisis del perfil transcripcional de anteras de petunia, arrojó que de 293 fragmentos de genes que exhibían alta expresión en la meiosis I, sólo 90 pudieron ser caracterizados (Cnudde et al., 2006). De lo anterior, podemos resaltar que aunque estos estudios permitieron la identificación de ortólogos de genes meióticos, así como la descripción de procesos y funciones representadas en la meiosis, dichos estudios ponen en evidencia que una proporción considerable de genes específicos de anteras tienen una identidad y función desconocida.

El análisis de transcriptomas de anteras, presenta una desventaja importante: las anteras además de meiocitos contienen células somáticas. Lo anterior implica que algunos genes o los patrones de expresión descritos no sean específicos de meiosis, o bien, que dicha mezcla celular “enmascare” algunos genes o procesos importantes para esta. Sin embargo, el desarrollo del método de colección capilar de meiocitos (CCM, por sus siglas en inglés) (Chen and Retzel, 2013), permitió la descripción del panorama transcripcional de meiocitos aislados de *Arabidopsis* y maíz a través de la técnica de RNA-Seq. Estos estudios evidenciaron nuevas características del transcriptoma de los meiocitos, tales como el alto contenido de elementos transponibles activos, o la presencia de genes de origen mitocondrial (Chen et al., 2010a; Dukowic-Schulze et al., 2014; Yang et al., 2011). Sin embargo, una vez más se observó que un número importante de lecturas mapean a segmentos no anotados del genoma. Asimismo, se notó que aunque parecen existir similitudes en la regulación transcripcional entre estas dos plantas modelo, existen también particularidades en los patrones de expresión en los meiocitos de cada especie (Dukowic-Schulze et al., 2013).

Dadas las ventajas que presenta el girasol para el estudio de la meiosis, el análisis del transcriptoma de sus meiocitos en una sub-etapa específica (Profase I), podría ayudar a ratificar algunos puntos en común de la regulación transcripcional de la meiosis en plantas,

además de revelar algunas características propias de esta especie. Por tanto, este capítulo presenta un análisis general de cuáles y a qué nivel se expresan los genes en los meiocitos de girasol en profase I, respecto a la expresión de estos en un transcriptoma somático del mismo. Los resultados del presente capítulo, fueron previamente publicados (Flórez-Zapata et al., 2014).

2.2 Materiales y Métodos

2.2.1 Aislamiento de meiocitos, extracción de RNA y secuenciación.

Las plantas de girasol del genotipo HA89 fueron crecidas, bajo condiciones de invernadero, en macetas que contenían una mezcla de hojarasca, arena, arcilla y perlita (2:1:1:1). Las plantas fueron fertilizadas semanalmente con la solución nutritiva Long Ashton (Phillips and Jennings, 1976), hasta el inicio de la etapa R2 del desarrollo (Schneiter et al., 1998). En esta etapa, cuando los botones florales tenían un diámetro aproximado de 2.0 cm (etapas tempranas de la meiosis, de acuerdo a observaciones previas), alrededor de 10 flores en disco fueron colectadas de la inflorescencia y colocadas en un portaobjetos cóncavo con 80 μ L de agua destilada estéril, donde posteriormente fueron aplastadas con agujas de disección para liberar los meiocitos. Los meiocitos colectados fueron observados al microscopio, sin ninguna tinción especial, con el propósito de verificar que estos se encontraban aún asociados unos a otros, formando la estructura de “gusano”, característica de meiocitos en profase I (Yang et al., 2011) (ver Figura 6). En los casos en que se observaron tétradas o gránulos de polen, los botones florales se descartaron. Los meiocitos de los botones florales que se encontraban en profase I fueron colectados en 80 μ L de la solución de RNAlater® (Qiagen, Valencia, CA) y almacenados a -70°C hasta la extracción de RNA. Simultáneamente, una muestra adicional de flores en disco de estos botones en profase I se colectó en una solución de etanol absoluto : ácido acético, en proporción 3:1 por 24 horas. Los meiocitos de estas muestras fijadas, fueron observados utilizando la técnica de tinción con acetocarmín, con el propósito de confirmar la etapa de la meiosis de estos meiocitos. Si la etapa de la meiosis (profase I) de las muestras de dicho botón floral era confirmada por este segundo control, dicha muestra era utilizada para la extracción de RNA.

El RNA total de los meiocitos en profase I fue extraído utilizando el kit ZR RNA MicroPrep® (Zymo Research, Orange, CA), siguiendo las indicaciones del fabricante, y posteriormente almacenado a -70°C . El RNA para la construcción del transcriptoma somático fue extraído de forma independiente, de 6 partes diferentes de la planta (hoja, tallo, raíz, bráctea, corola y receptáculo) de este mismo genotipo, utilizando el kit NucleoSpin RNA Plant® (Macherey-Nagel, Düren, Germany), y posteriormente mezclado en partes equimolares. En total se construyeron 2 genotecas (réplicas biológicas) de meiocitos y una de tejidos somáticos, utilizando los kits estándar para la preparación de genotecas de Illumina TruSeq RNA. Dichas genotecas, fueron posteriormente secuenciadas en un equipo Illumina HiSeq 2500 para obtener lecturas pareadas de 100 pares de bases.

2.2.2 Filtrado de calidad de secuencias, ensamblaje y re-mapeo.

Una vez obtenidas las secuencias, estas fueron sometidas a una limpieza y filtrado de calidad. Primero se realizó la remoción de adaptadores de secuenciación utilizando el software cutadapt v. 1.3 (Martin, 2011). Posteriormente, el filtrado de calidad se realizó utilizando el software PRINSEQ v 0.20.4 (Schmieder and Edwards, 2011), permitiendo un puntaje mínimo de calidad de 20 y no más de dos bases ambiguas por lectura. El ensamble *de novo* de las lecturas limpias (tanto del transcriptoma somático como el de meiocitos), se realizó utilizando Trinity (versión 20130225) (Grabherr et al., 2011), con los parámetros por defecto, a excepción del argumento *-min_kmer_cov*, el cual fue fijado a 2. El resultado del ensamblador Trinity, es un conjunto de “componentes” los cuales corresponden a secuencias que el algoritmo considera son producto de genes diferentes. Cada componente es representado a su vez por una o más secuencias, que podrían ser variantes de empalme de los transcritos.

La cuantificación de la expresión relativa de cada componente se realizó re-mapeando las lecturas crudas a un conjunto de secuencias representativas de cada componente (la más larga), de cada uno de los componentes ensamblados (genes), utilizando el programa Bowtie2 2.1.0 (Langmead and Salzberg, 2012), con los parámetros *-a -rdg 6, 5 -rfg 6, 5 -score-min L, -0.6, -0.4*. El número único de lecturas para cada componente, es decir aquellas lecturas que mapearon exclusivamente a un componente, fue estimado utilizando eXpress 1.4.1 (Roberts and Pachter, 2013) con los parámetros por defecto. Dichos conteos fueron arreglados en una matriz de genes vs. cuentas en cada muestra, para los análisis posteriores.

2.2.3 Identificación de transcritos.

Las secuencias de este transcriptoma ensamblado, fueron identificados con el resultado del mejor *hit* del programa blastx (Altschul et al., 1997), contra cuatro bases de datos diferentes: péptidos de la versión TAIR10 de *A. thaliana*, péptidos de la base de datos de RefSeq (versión del 24/07/2013) y péptidos de girasol de las variedades HA412 y HaXRQ disponibles en el sitio web de HeliaOrg (<https://www.heliagene.org>). Únicamente aquellos hits con un puntaje en bits (bitscore) ≥ 90 y un valor esperado $E \leq 1 \times 10^{-6}$ fueron considerados significativos. Las secuencias que no pudieron ser identificadas con ninguna de estas cuatro bases de datos fueron catalogadas como no-identificadas. Adicionalmente los genes fueron anotados con términos de ontología génica (GO) (Harris et al., 2004) y asociados a rutas metabólicas de Aracyc (Mueller, 2003), utilizando el paquete de R *org.At.tair.db* y *GO.db* (Pages et al., 2008). La identificación de factores de transcripción, se llevo a cabo a partir del gen ortólogo de *A. thaliana* en la base de datos Plant TFDB v. 3.0 (Jin et al., 2014).

2.2.4 qRT-PCR.

Para los análisis de qRT-PCR, se realizaron nuevas extracciones de meiocitos aislados en profase I, y se realizó una nueva mezcla equimolar de tejidos somáticos, como se describió anteriormente. El cDNA fue obtenido de a partir del RNA total (1.5 μ g por reacción) empleando el kit High capacity RNA-to-cDNA[®] (Applied Biosystems, Foster City, CA). Las reacciones de PCR en tiempo real fueron realizadas en un volumen final de 20 μ L, que

contenía: 10 μL de 2x SYBR Green PCR Master Mix de Applied Biosystems, 0.4 μL de cada primer (10 μmol), 1.5 μL de cDNA (150 ng/ μL), y 7.7 μL de agua destilada. El programa de PCR fue corrido en un equipo StepOne (Applied Biosystems, Foster City, CA), de la siguiente forma: una desnaturalización inicial de 10 min. a 95°C, seguida de 40 ciclos a 90°C por 15 segundos y 1 minuto a 58°C. Tres réplicas técnicas fueron realizadas para la cuantificación de cada gen. Un análisis de curva de desnaturalización fue realizado al final de cada corrida, para verificar que sólo un producto de PCR estaba siendo amplificado. Las diferencias relativas en la expresión de los genes, fueron detectadas por el método de $\Delta\Delta\text{Ct}$, utilizando el gen de la proteína ribosomal S5 como normalizador.

2.2.5 Diseño y análisis estadístico.

Para la construcción de cada una de las dos genotecas de meiocitos se utilizaron diferentes conjuntos de plantas y se realizaron extracciones independientes de RNA total. Por tanto, la variación en la expresión encontrada en cada uno de los genes entre las dos réplicas biológicas da un estimado del error estadístico (variación no explicada), la cual incluye variación tanto biológica como técnica. Dado que no realizamos réplicas de la genoteca somática, asumimos que el error estadístico en esta es del mismo tamaño que el detectado entre las réplicas de la genoteca de meiocitos.

Como se describió anteriormente, el nivel de expresión de cada componente fue evaluado a partir del número único de lecturas mapeadas a este. Sin embargo, para aquellos que compartían el mismo identificador de TAIR10, los datos de expresión fueron colapsados en un único componente, sumando el número de lecturas que habían mapeado a estos. La prueba estadística aplicada para el análisis de expresión diferencial se realizó utilizando el paquete edgeR (Robinson et al., 2010), en R v 2.15.3 (R Core Team, 2013). Los valores de P resultantes de la función exactTest de edgeR, fueron posteriormente analizados con la función q-value (Storey and Tibshirani, 2003), con los parámetros por defecto, ajustando el FDR a un 1%. Las diferencias en las categorías funcionales de términos GO, así como de rutas metabólicas, fueron analizadas como se describió previamente por Martínez-López et al. (Martínez-López et al., 2014), con una ligera modificación; únicamente los términos GO, que pertenecían al nivel 4, de acuerdo con la función getLevel del paquete goProfiles de R (Sanchez et al., 2013), fueron considerados.

2.3 Resultados y Discusión

2.3.1 Visión general del transcriptoma de meiocitos de girasol.

En total, de las tres genotecas secuenciadas, se obtuvieron 491,701,991 lecturas crudas pareadas. 387,767,222 (78.86%) de estas fueron lecturas de alta calidad, después del proceso de limpieza (Tabla S1). Estas lecturas fueron ensambladas *de novo*, produciendo un total de 47,295 transcritos distintos (“componentes”). De las lecturas de alta calidad, 278,997,557 (71.94%) lecturas mapearon a un único transcrito, dando evidencia de la expresión de 46,386 transcritos. En otras palabras, el 98% de todos los componentes

reconstruidos en el ensamblado, mostró expresión en estas genotecas. El 2% restante de los componentes a los cuales no alinearon lecturas de manera exclusiva, fueron excluidos de los análisis posteriores. 32,303 (69.64%) de los componentes pudieron ser anotados *vía* BLAST, con al menos un hit con una proteína de las bases de datos utilizadas. El 30.36% (14,087) restante, no pudo ser identificado a través de esta aproximación. Por otro lado, los transcritos identificados que compartían el mismo identificador de BLAST, fueron considerados como productos del mismo *locus* de girasol o transcritos derivados de un gen cercanamente relacionado, e.g. parálogo. Por tanto, para cuantificar la expresión de estos transcritos, las lecturas que alinearon a estos *loci*, fueron sumadas y los componentes fueron “colapsados”, de manera que estos componentes relacionados fueron tratados como un único gen. Nuestro conjunto de datos final, estaba compuesto por un total de 36,304 transcritos, que fueron utilizados para los análisis posteriores.

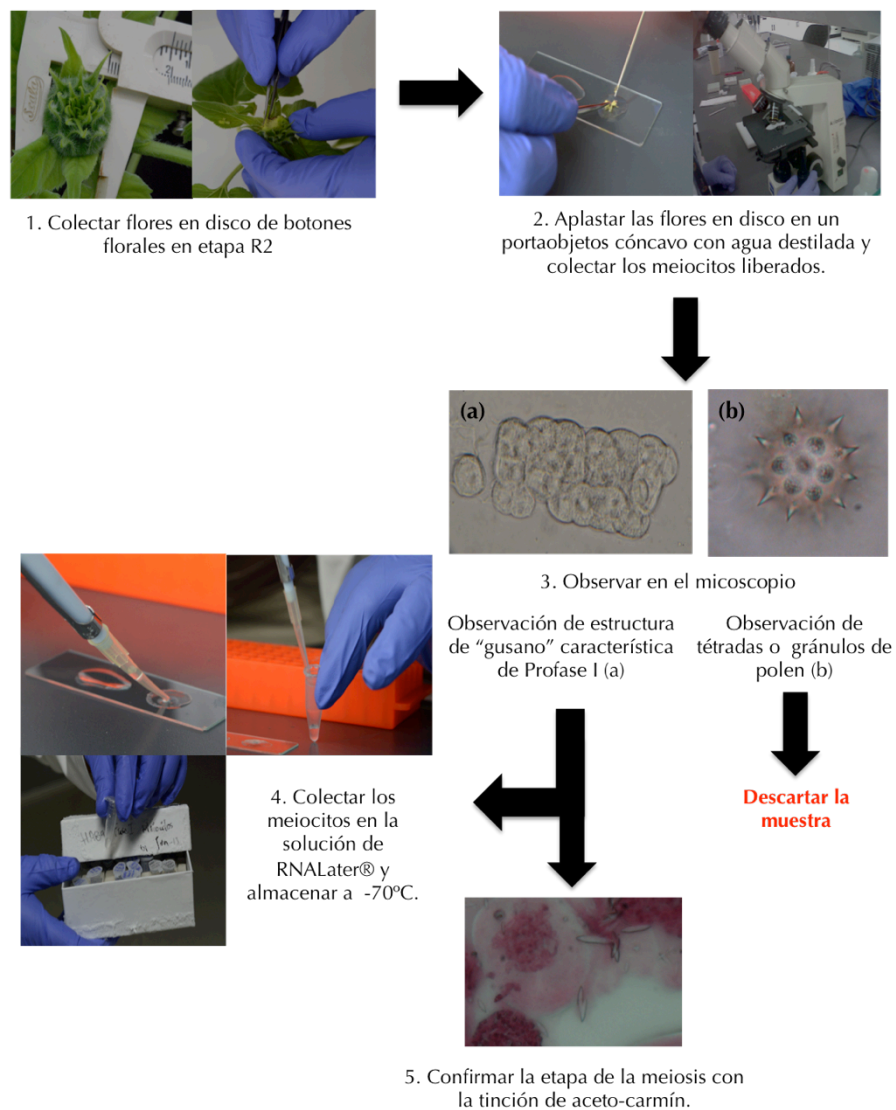


Figura 6. Proceso de recolección de meiocitos de girasol.

19,574 (54%) de los genes fueron encontrados diferencialmente expresados (DE) entre las genotecas somática y de meiocitos. De estos, 7,755 genes mostraron mayor expresión en meiocitos y 11,819 mostraron mayor expresión en tejido somático. Aproximadamente el 80% de los genes DE mostraron un cambio relativo (“*fold change*”) mayor o igual a 2 entre las muestras. Interesantemente, el 38.37% y el 39.05% de los genes DE identificados para las genotecas somática y de meiocitos, respectivamente, fueron genes que no pudieron ser identificados (Figura 7A). Adicionalmente, la mayoría de genes que mostraron evidencia de expresión “exclusiva” en alguna de las dos genotecas (somática o meiocitos), fueron también genes no-identificados (Figura 7B). El número de dichos genes de expresión “exclusiva”, no-identificados es ligeramente mayor en el transcriptoma de meiocitos. Lo anterior está de acuerdo con las observaciones realizadas en el transcriptoma de meiocitos de Arabidopsis y maíz (Dukowic-Schulze et al., 2013), lo cual sugiere que muchos de los genes que le confieren la identidad celular a los meiocitos siguen sin ser caracterizados.

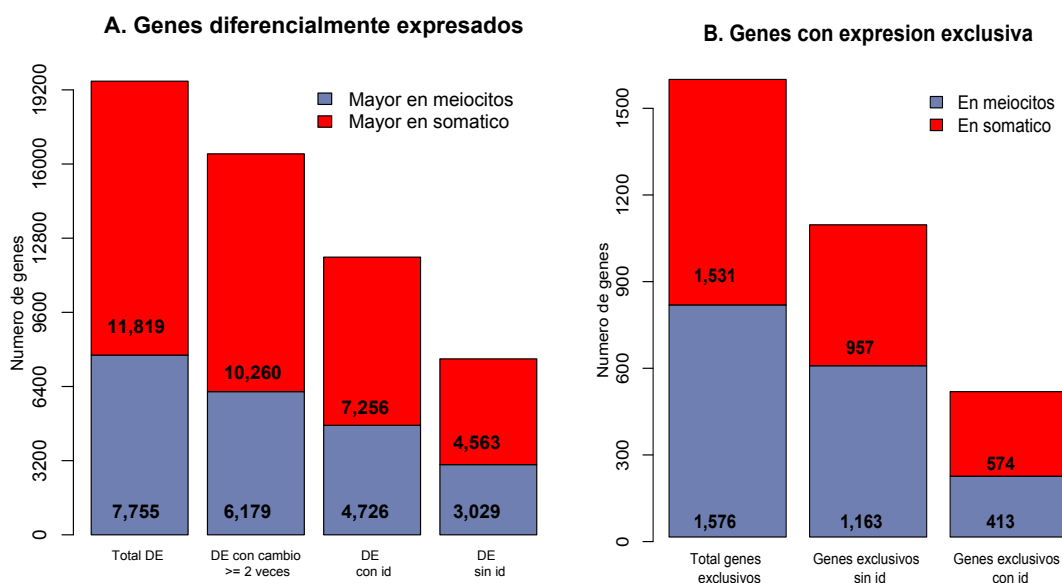


Figura 7. Número de genes de girasol diferencial y exclusivamente expresados entre el transcriptoma de meiocitos y el somático. A. Número de genes DE (FDR = 1%) en los dos transcriptomas; "Con id" son los genes identificados como bases de datos de péptidos y "sin id" son los genes no identificados. B. Número de genes con expresión detectada solamente en un tipo de transcriptoma (genes "exclusivos").

Es importante resaltar que mezclar los diferentes tejidos u órganos somáticos en una sola genoteca permite realizar comparaciones globales entre el transcriptoma de meiocitos y otros órganos o tejidos. Sin embargo, no permite discriminar si dicho gen se encuentra DE únicamente en uno o varios de los tejidos que conforman dicha mezcla. Lo anterior es importante porque se ha visto que es en tejido meristemático de rápida división en donde es posible que existan daños espontáneos en el DNA, y muchos genes meióticos pueden actuar

mediando la reparación del DNA (Zhou and Pawlowski, 2014b). Por otro lado, el análisis de expresión diferencial mostró que en general la mayoría de los genes DE presentaban mayores niveles de expresión en el transcriptoma somático en comparación con el de meiocitos. Es probable que esto se deba a que un gen puede cumplir una función en más de un tejido, y por tanto al mezclarlos exista un efecto aditivo sobre el nivel de expresión.

2.3.2 Perfil de expresión de genes con función en meiosis.

En el análisis del transcriptoma de meiocitos pudimos identificar 63 ortólogos de un total de 84 genes previamente asociados con meiosis en *A. thaliana* (Chen et al., 2010a; Yang et al., 2011). 51 de estos, fueron encontrados diferencialmente expresados (todos con una tasa de expresión mayor o igual al doble) entre la genoteca de meiocitos y la somática. De estos, sólo 2 mostraron mayor expresión en esta última (Figura 8). Asimismo, encontramos que el gen MMD1/DUET, que es necesario para la meiosis masculina y se expresa en una sub-etapa de la profase I (diploteno) (Andreuzza et al., 2015), tiene un nivel de expresión alrededor de 1000 veces mayor en meiocitos que en el transcriptoma somático, lo cual confirma la confiabilidad de nuestros datos.

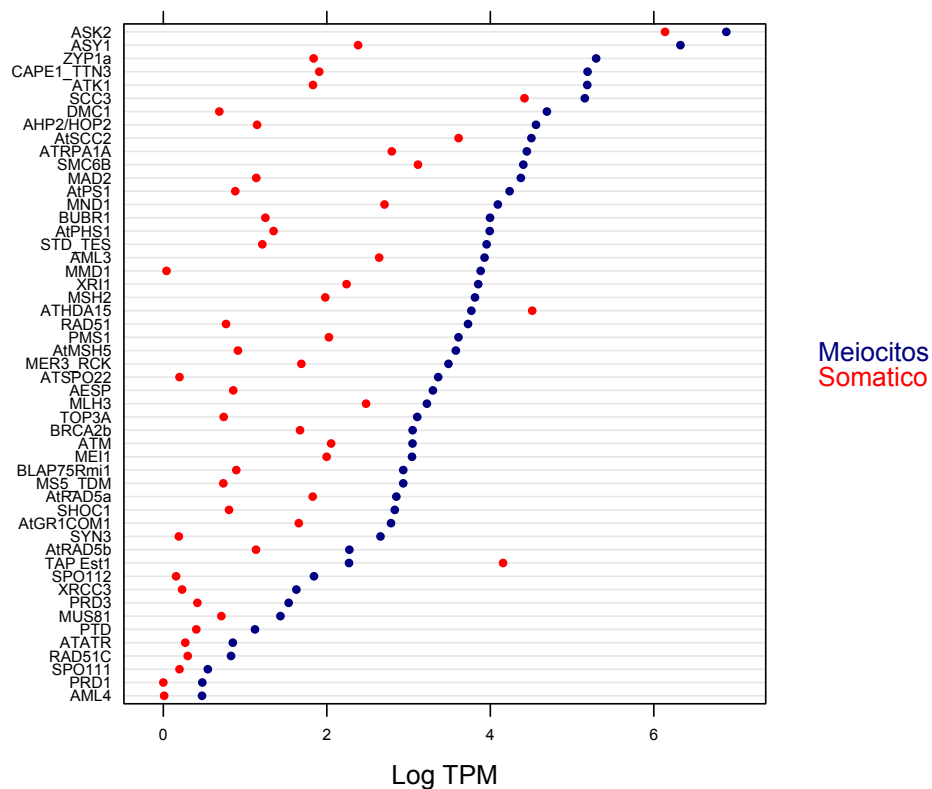


Figura 8. Niveles de expresión de genes con función en meiosis previamente descrita que se encontraron diferencialmente expresados (DE).

Dado que sólo fue posible la identificación de 63 de los 84 genes con función meiótica reportados en Arabidopsis, decidimos evaluar si los 21 genes restantes, se encontraban

conservados en otras especies de plantas (Figura S1). Para esto, buscamos los ortólogos de estos 21 genes, en el genoma de 25 especies diferentes de plantas disponibles en la base de datos de Plaza v 2.5 (Proost et al., 2009). El número de ortólogos de estos, así como su distribución entre las distintas especies de plantas varía considerablemente. Por ejemplo, el gen ATRSP3 (AT1G64030) aparece sólo conservado en el género *Arabidopsis*; mientras que, el gen AHP2 (AT3G29350), que posee ortólogos en casi todas las especies de plantas analizadas, no pudo ser identificado en el género *Oryza*. Entonces, no es sorprendente que en el presente estudio no se haya podido detectar la expresión de algunos genes asociados a meiosis en nuestro transcriptoma. Por un lado, es posible que estos no se encuentren presentes en el genoma de girasol o, alternativamente, las secuencias de los genes de girasol podrían haber divergido significativamente, de manera que ya no pudieron ser identificados como ortólogos con la estrategia empleada.

Para ocho de los 21 genes meióticos cuya expresión no fue detectada en nuestro transcriptoma (ZYPb, ASK1, AML2, BUB3.1, ATSMC2, BRCA2A, ATK5, y SMC6A), detectamos la expresión de al menos un ortólogo de girasol perteneciente a la misma familia de genes. Lo anterior sugiere, que aunque no podemos excluir totalmente la posibilidad de problemas técnicos para detectar genes de muy baja expresión, o para encontrar el ortólogo correspondiente, es posible que los ortólogos de estos genes estén ausentes en girasol, o que el número de genes parálogos que participan en meiosis, pertenecientes a esta familia de genes sea diferente a los reportados en *Arabidopsis*. Para confirmar esta observación es necesario realizar estos análisis en girasol a nivel genómico.

El análisis de expresión diferencial arrojó que 51 de los 82 genes meióticos considerados se encontraban DE y de estos 49 presentaron mayor expresión en meiocitos. Si bien la expresión preferencial de estos genes en el meiocito no asegura su función en meiosis, podría indicar que en girasol un amplio número de genes meióticos se encuentran conservados, tal como se observó previamente en *Arabidopsis* (Yang et al., 2011). Los dos genes meióticos, con mayor expresión en tejido somático codifican para las proteínas Est1 y la deacetilasa de histonas ATHDA15. De acuerdo con Alinsug y colaboradores (Alinsug et al., 2009), el gen ATHDA15 se encuentra significativamente expresado en la punta de la raíz y el pólen, mientras que Liu et al. (Liu et al., 2013) encontraron que esta deacetilasa de histonas es un regulador transcripcional negativo de genes importantes para la biosíntesis de clorofila en plántulas etioladas. Estos estudios, confirman que este gen también tienen una función importante en células no-meióticas, por tanto su mayor expresión en el transcriptoma somático no es sorprendente. La proteína Est1 de *Arabidopsis* es el ortólogo de la proteína SMG7, encargada de la degradación antisentido de RNA (Riehs et al., 2008). Este gen fue encontrado sobreexpresado en respuesta a un exceso de agua en la raíz en *Arabidopsis* (Hsu et al., 2011), sugiriendo que esta no tiene una expresión exclusiva de meiocitos.

Un total de 12 (19.04%) de los ortólogos de girasol para genes con función en meiosis, no mostraron expresión diferencial entre el transcriptoma somático y de meiocitos. Este resultado, esta de acuerdo con los datos de expresión para estos genes en Arabidopsis y maíz. En Arabidopsis, sólo el 72.05% de los genes meióticos presentaron una diferencia de expresión mayor o igual a el doble, entre meiocitos y plántulas (Chen et al., 2010a). En maíz, sólo el 57.31% de los genes presentaron niveles de expresión mayores o iguales a el doble con respecto a plántulas. Además, entre los genes meióticos no DE, nosotros encontramos que los ortólogos de maíz para ZIP4, MPA1, CDC45 u SMC1_TTN8 tampoco demostraron una alta expresión en los meiocitos de maíz (Dukowic-Schulze et al., 2014). Algunos genes como MLH1, MSH4, RAD50, y AML5, además de su papel en meiosis, tienen funciones en mitosis o en la reparación del DNA (Chen et al., 2010a; Osman et al., 2011; Yang et al., 2011). Por tanto, la falta de expresión diferencial de estos genes con funciones en meiosis, entre meiocitos y otros tejidos de la planta, no es de extrañar.

2.3.2 Categorías funcionales sobreexpresadas en meiocitos

Con el propósito de identificar diferencias biológicas globales entre los dos transcriptomas analizados, los ortólogos de girasol identificados con la base de péptidos de Arabidopsis fueron categorizados con las categorías de ontología génica (GO), dentro de las cuales se incluían: Componente Celular (CC), Proceso Biológico (BP) y Función Molecular (MF). Utilizando la metodología previamente reportada en Martínez-López et al. (Martínez-López et al., 2014), encontramos en el transcriptoma 160 términos GO de CC (de un total de 189 considerados), de los cuales 40 estaban representados al menos al doble en meiocitos (Tabla S2). Como se esperaba, dentro de estos términos se encuentran algunos relacionados con procesos de división celular y organización de cromatina, tales como: “condensin complex,” “kinetochore,” “chromatin assembly complex,” “nucleosome,” “microtubule associated complex”; así como términos relacionados a transcripción y actividad mitocondrial (“mitochondrial outer membrane translocase complex”, “mitochondrial intermembrane space” y “mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)"). En el transcriptoma somático, se observó una mayor representación de términos de CC relacionados con fotosíntesis (Table S3).

Nuestro conjunto de genes fue agrupado en 386 términos GO de BP, de los cuales 338 fueron encontrados diferencialmente representados entre los dos transcriptomas (somático y de meiocitos). De estos, los términos que mostraron un mayor cambio (79 términos GO de BP) en meiocitos, fueron los términos relacionados con reproducción como “pollination”, “pollen sperm cell differentiation” y “sex determination”, así como términos relacionados con la meiosis y el ciclo celular, incluyendo: “meiotic chromosome segregation”, “chiasma assembly”, “reciprocal meiotic recombination” y “meiotic DNA double-strand break formation”. Además, categorías relacionados con el silenciamiento génico y la regulación de la expresión génica (“negative regulation of transposition”, “chromatin silencing” y “gene silencing by RNA”), también estuvieron sobrerepresentadas (Figure 9, Table S4). Por otro lado, un amplio rango de términos GO de BP fueron encontrados con expresión

significativamente mayor en el transcriptoma somático comparado con el de meiocitos. Muchos de estos términos se encontraban relacionados con “response to stimulus” y “photosynthesis” (Table S5).

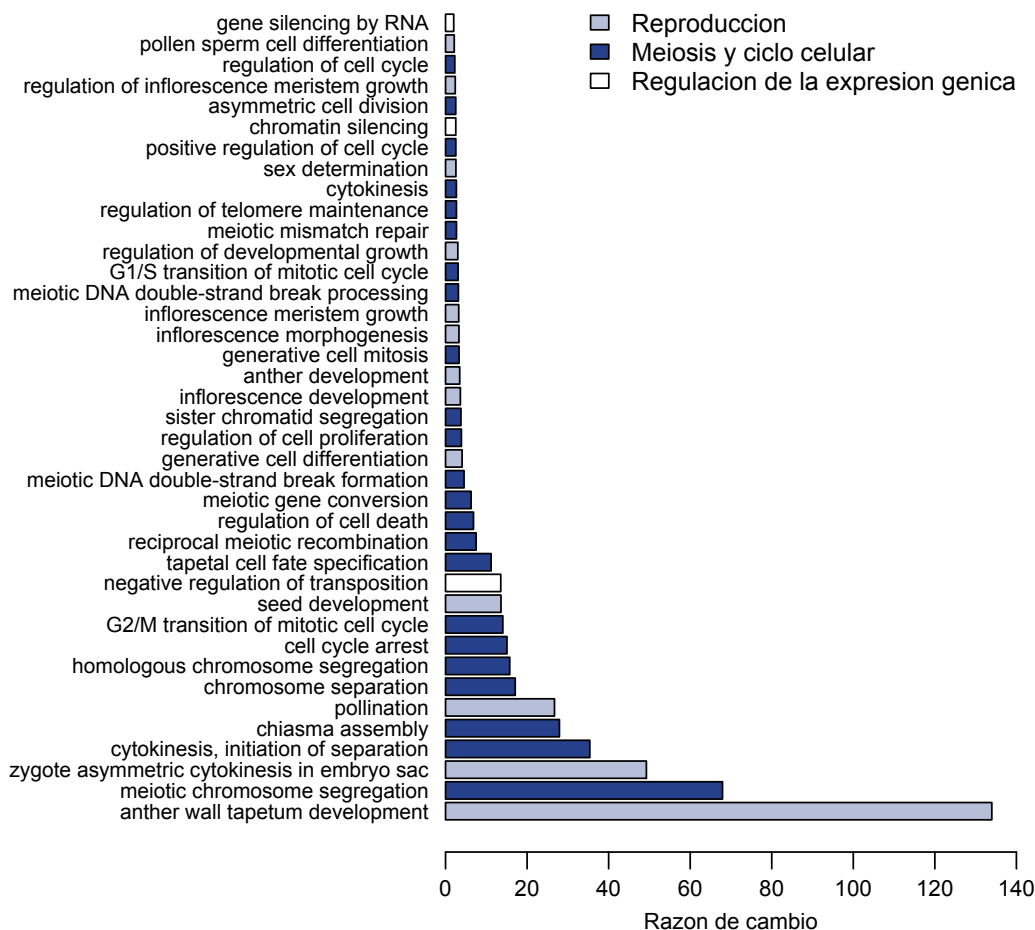


Figura 9. Diferencias en la expresión de términos GO de BP entre los transcriptomas de meiocitos y somático. Se presentan los términos BP con mayor expresión en meiocitos y una razón de cambio mayor a 2.

Cuando analizamos las diferencias de términos GO de la categoría MF, encontramos que aunque 139 de estos se observaron diferencialmente representados, sólo 8 tenían una representación dos veces mayor en meiocitos (Table S6). De estos, “histone binding”, “cyclin binding”, “nucleosome binding” y “ATPase activator activity” se encontraban relacionados con términos de las categorías CC y BP, previamente descritos con alta representación en meiocitos. De manera similar, los términos con mayor representación en el transcriptoma somático, se encontraron relacionados con otros términos de las categorías de CC y BP, e.g. “ribulose- 1,5-bisphosphate carboxylase/oxygenase activator activity” y “chlorophyll binding” (Table S7).

Como resultado del análisis de términos GO determinamos que la mayoría de términos de la categoría BP con mayor representación en meiocitos también fueron reportados previamente en los transcriptomas de maíz y Arabidopsis (Dukowic-Schulze et al., 2013). Sin embargo, en maíz fue reportado que la mayoría de términos sobre representados en estos, se encontraban relacionados con procesos biosintéticos. Aunque, los autores de dicho análisis, argumentan que esto podría ser resultado de la pobre anotación de los genes de maíz a términos GO (Dukowic-Schulze et al., 2014). Por tanto, aunque utilizar sólo los genes que tuvieron ortólogos con Arabidopsis para llevar a cabo este análisis en girasol puede ser limitado, los resultados que obtuvimos son consistentes con la función y especificidad de los meiocitos.

Entre los términos de las categorías CC y MF diferencialmente representados entre los transcriptomas somático y de meiocitos también destacan funciones relacionadas con la actividad energética del meiocito. Nuevamente, esto es consistente con los reportes previos de transcriptomas de Arabidopsis y maíz (Chen et al., 2010a; Dukowic-Schulze et al., 2014), en los cuales genes importantes para la función mitocondrial se encontraron sobre expresados en meiocitos. Otros estudios, también han discutido que algunas proteínas necesarias para la recombinación homóloga como Rad50, DMC1, y Rad51 tienen una actividad dependiente de ATP (Hopfner et al., 2000; Nara et al., 2001), reforzando la idea de la importancia del metabolismo energético en estas células especializadas.

Una aproximación adicional para determinar la función de los genes fue la descripción de nuestros transcriptomas en términos de rutas metabólicas. Con este objetivo categorizamos nuestro conjunto de genes en 370 rutas metabólicas distintas, de las cuales 323 presentan representación diferencial. De estas, 22 mostraron un incremento mayor al doble en meiocitos (Table S8), mientras que 67 mostraron un incremento mayor en el transcriptoma somático (Table S9). Además, considerando la jerarquía de rutas metabólicas previamente establecida en Aracyc, calculamos la frecuencia relativa a la cual las rutas metabólicas pertenecientes a una jerarquía particular se encontraban diferencialmente representados con un nivel mayor al doble (Figura 10). De este análisis, concluimos que la mayoría de rutas metabólicas tenían mayor representación en el transcriptoma somático, a excepción de rutas de las jerarquías “Cell Structures Biosynthesis”, “Nucleosides and Nucleotides Biosynthesis” y “Other Biosynthesis”, las cuales se encontraron mas representadas en meiocitos.

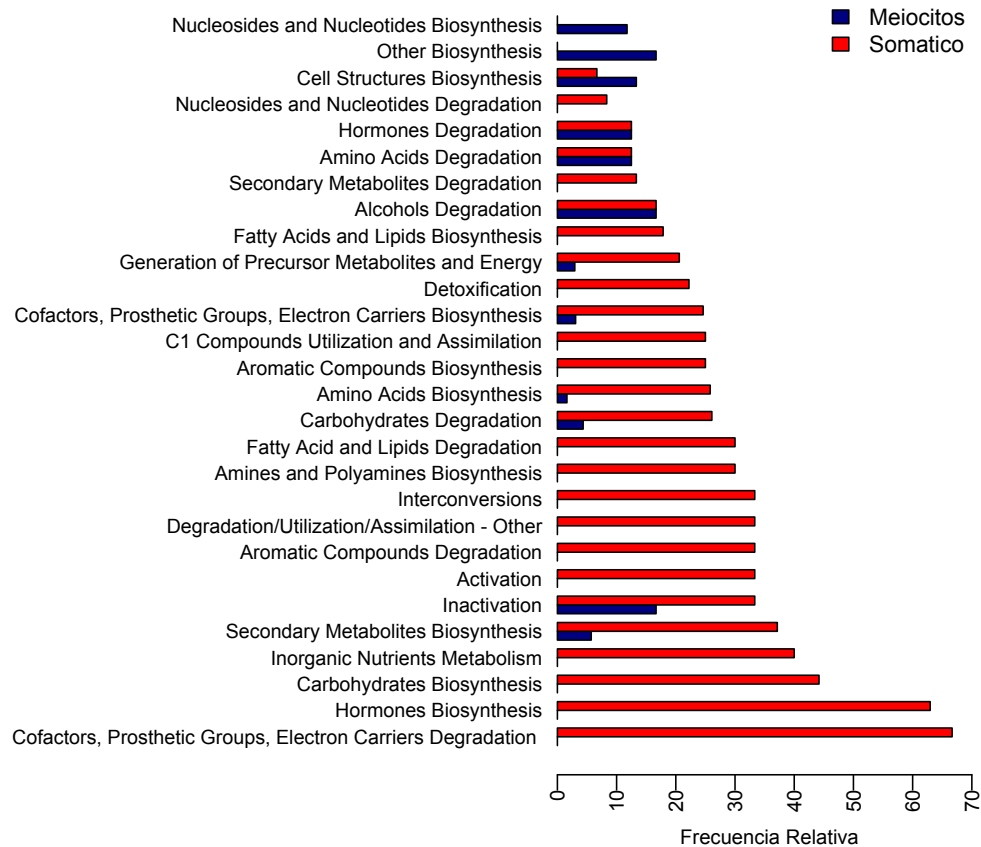


Figura 10. Frecuencia relativa de rutas metabólicas encontradas diferencialmente expresadas en cada jerarquía de ruta metabólica.

La mayoría de rutas metabólicas mostraron mayor representación en el transcriptoma somático en comparación con el de meiocitos. Lo anterior era de esperarse, puesto que los distintos órganos y tejidos representados en el transcriptoma somático representan una mayor cantidad de funciones y procesos metabólicos con respecto a los meiocitos, los cuales tienen una función especializada. En relación a las rutas con mayor representación en meiocitos, se resalta que las incluidas en la jerarquía “nucleosides and nucleotides biosynthesis”, son esperadas con mayor representación en este transcriptoma. De acuerdo a estudios previos en levadura, la síntesis de macromoléculas incluyendo nucleótidos, son observados como parte de los cambios que ocurren en este organismo durante la transición de mitosis a meiosis (Ray and Ye, 2013). Asimismo, Moss y colaboradores (Moss et al., 2010), establecieron que la síntesis de nucleótidos inducida como respuesta al daño celular es importante para la reparación del DNA a través de la recombinación homóloga en levadura. Downs (Downs, 1997), demostró que las rutas de generación de nucleótidos de tipo purinas están involucradas con la regulación inducida por hormonas en la maduración meiótica de oocitos de ratón. Por otro lado, la segunda jerarquía con mayor representación en meiocitos fue “Cell structures biosynthesis”. De dicha jerarquía, la ruta metabólica

“sporopollenin precursor biosynthesis”, fue en la que se encontraron las mayores diferencias entre transcriptomas. La esporopolenina es el mayor componente de la exina, la membrana exterior de la pared del pólen (Domínguez and Mercado, 1999), por lo que este no es un resultado inesperado. Finalmente, la ruta “PRPP biosynthesis I”, es una de las que pertenece a la jerarquía de “other biosynthesis”, la cual también se encontró con mayor representación en meiocitos. El PRPP (Fosforribosil pirofosfato), es un pentafofosfato que fue previamente relacionado con la inducción meiótica en oocitos de ratón (Downs et al., 1998), por lo que su síntesis también podría ser importante para la meiosis en girasol.

2.3.3 Niveles de expresión de reguladores de la expresión génica

Con la intención de encontrar reguladores transcripcionales preferencial o exclusivamente expresados en meiocitos, analizamos los patrones de expresión de genes que codifican para factores de transcripción (TF), presentes en nuestro transcriptoma. Encontramos 489 TF diferencialmente expresados, 188 de ellos con mayor expresión en meiocitos y 301 con mayor expresión en el transcriptoma somático. Estos TF pertenecen a 54 familias diferentes (Jin et al., 2014) (Figure S2). Sin embargo, el número de familias distintas con mayor expresión en meiocitos, es mayor que el número observado en el transcriptoma somático (49 vs. 38 familias de TF, respectivamente). Adicionalmente, observamos que 14 TF para los que su expresión fue detectada sólo en meiocitos (Table S10). La distribución del número de TF por familia, también fue significativamente diferente entre los dos transcriptomas ($P \approx 6.1e^{-13}$).

Los TF de las familias bHLH y MYB, son dos de los más expresados en meiocitos, lo cual está de acuerdo con las observaciones de Dukowic-Schulze et al. (Dukowic-Schulze et al., 2013) en meiocitos de maíz y *Arabidopsis*. Sin embargo, estas familias también se encuentran altamente representadas en el transcriptoma somático. Algunos TF de la familia C2H2 mostraron un alto nivel de expresión en meiocitos, que a su vez es mayor respecto al observado en el transcriptoma somático. Interesantemente, en *C. elegans*, cuatro proteínas C2H2 con dedos de zinc tienen un papel central en el apareamiento y sinapsis de los cromosomas homólogos durante la meiosis (Phillips and Dernburg, 2006). En *Drosophila* un TF con dedos de zinc de tipo C2H2, denominado Grauzone, es requerido para llevar a cabo la meiosis completa en oocitos (Chen et al., 2000). Por otro lado se observó que TFs de la familia GATA, también presentaron mayor expresión en meiocitos. Recientemente un factor GATA, Gaf1, fue descrito como un regulador transcripcional de desarrollo sexual (mediando la transición de mitosis a meiosis) en levaduras (Kim et al., 2012), sugiriendo que TF de este tipo también podrían tener un papel importante durante la meiosis. La familia de TFs FAR1 fue encontrada exclusivamente expresada en meiocitos, y es también una de las que mayor expresión presenta en estos. Miembros de esta familia de TFs son proteínas relacionadas con transposasas mutator-like (Lin et al., 2007). Se necesitan estudios adicionales para establecer si este resultado podría estar relacionado con observaciones previas de la alta actividad de elementos transponibles en los meiocitos de *Arabidopsis* (Chen and Retzel, 2013; Yang et al., 2011).

En comparación con los TFs sobreexpresados en meiocitos de Arabidopsis y maíz descritos en Dukowic-Schulze et al. (2013), nosotros no detectamos evidencia de la expresión de seis ortólogos de Arabidopsis en nuestros transcriptomas (AT1G02030, AT1G06170, AT2G36270, AT2G42830, AT5G17800, y AT5G19790). Asimismo, ortólogos de tres TF no mostraron evidencia de expresión diferencial entre nuestros transcriptomas (AT3G54340, AT5G20240, AT5G60910). Sin embargo, algunos factores de transcripción descritos en estas plantas, necesarios para el desarrollo floral, como CRAB CLAW, AMS y AGAMOUS, también mostraron una alta expresión en los meiocitos de girasol. En estas tres especies de plantas fue posible detectar la expresión preferencial en meiocitos de dos TF, los cuales no se encuentran anotados y podrían ser genes candidatos interesantes. Este análisis de la expresión de TF en meiocitos de girasol sugiere la actividad de un número diverso de reguladores transcripcionales durante la meiosis, tal como fue propuesto por Yang et al. (2011).

El silenciamiento génico por RNA, fue uno de los términos GO de BP, con una representación significativa mayor en meiocitos. Por tanto, nosotros observamos las diferencias en expresión de 36 genes que se encontraban previamente asociados a rutas de silenciamiento (Table S11). En total, 29 ortólogos para estos genes fueron encontrados en nuestro conjunto de genes de girasol. Dieciocho de estos se encontraron DE entre los transcriptomas, de los cuales 17 presentaban una expresión 2 veces mayor en meiocitos, comparada con el transcriptoma somático (Figure S3). Esta maquinaria de silenciamiento génico es claramente importante para el silenciamiento de transposones (Eamens et al., 2008), y cómo fue propuesto anteriormente, la alta actividad de estas vías de silenciamiento podría ser un mecanismo importante para prevenir mutaciones que pueden resultar del movimiento de estos elementos en la línea germinal (Yang et al., 2011).

2.3.4 Análisis de expresión de algunos genes por la técnica de qRT-PCR.

Los niveles de expresión de 5 genes, fueron cuantificados por el método de qRT-PCR, con el propósito de validar los patrones de expresión descritos en nuestro experimento de RNA-seq. De acuerdo con el análisis del transcriptoma por RNA-seq, cuatro de estos genes presentaron una mayor expresión en meiocitos y el quinto en el transcriptoma somático. Aquellos que presentan una mayor expresión en meiocitos fueron: INO, que codifica para un factor previamente relacionado con el desarrollo del óvulo (Baker et al., 1997); RBR1, que codifica para una proteína relacionada con la proteína de retinoblastoma, la cual es necesaria para eventos tempranos de la meiosis en Arabidopsis (Chen et al., 2011) y finalmente, dos genes meióticos previamente descritos, MMD1 y ATK1. Como ejemplo de genes con mayor expresión en la genoteca somática, nosotros seleccionamos el gen ARGAH2 (que codifica para la arginina amidohidrolasa 2), la cual esta implicada en la defensa contra patógenos (Gravot et al., 2012). Los cambios en los niveles de expresión de estos genes entre los dos transcriptomas, evaluados por las dos metodologías (RNA-seq y qRT-PCR), son presentados en la Figura 11. Incluso cuando, como se esperaba, los cambios en expresión son altamente

variables entre los dos métodos (Marioni et al., 2008), para cada uno de estos 5 genes, las diferencias de expresión son consistentes entre las dos metodologías evaluadas.

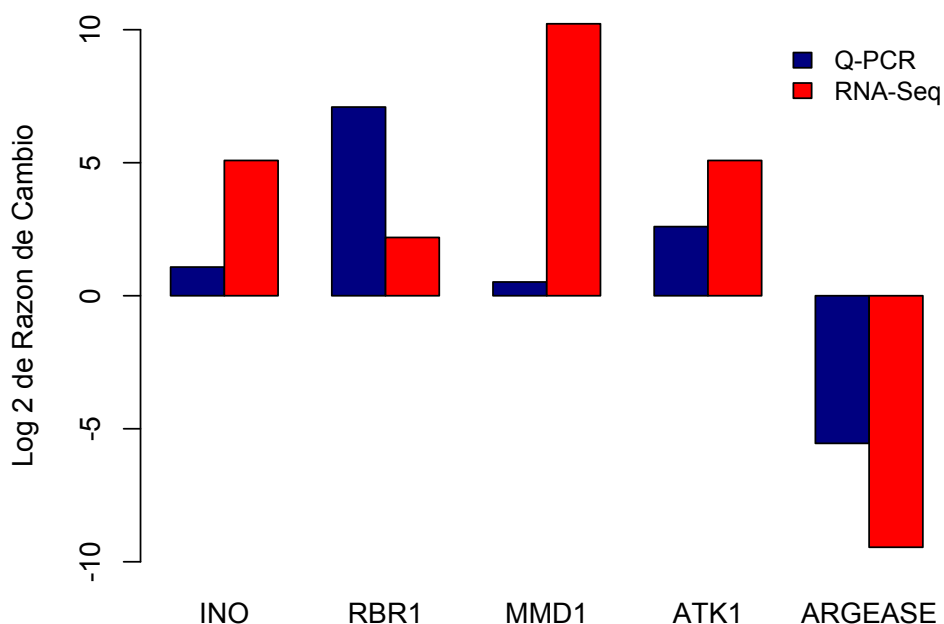


Figura 11. Comparación de los niveles de expresión de algunos genes, estimados por las metodologías de RNA-seq y RT Q-PCR.

2.4 Conclusiones

El análisis del transcriptoma de meiocitos de girasol nos permitió encontrar diferencias y similitudes entre los perfiles de expresión de genes que participan en meiosis, así como en los mecanismos y procesos biológicos con respecto a los trabajos previos realizados en las plantas modelo *Arabidopsis* y maíz. Aunque algunos mecanismos globales de regulación parecen estar conservados entre estas plantas modelo, también fue posible identificar algunos factores de transcripción, así como rasgos de funciones biológicas y rutas metabólicas importantes para la meiosis que no habían sido previamente descritos. Lo anterior, resalta la importancia del estudio de algunos procesos biológicos en plantas no-modelo.

CAPÍTULO 3: Identificación de lncRNAs y sus cambios transcripcionales en los meiocitos de girasol.

3.1 Introducción

El análisis del transcriptoma de meiocitos ha permitido evidenciar que la alta actividad transcripcional de estos, se caracteriza además porque existe una alta expresión de rasgos no anotados del genoma (Dukowic-Schulze et al., 2013), o de transcritos que no pudieron ser identificados con un ortólogo de proteína (Flórez-Zapata et al., 2014). Por tanto, es de esperar que algunos de estos transcritos pudieran ser RNAs no codificantes, dado que se ha descrito que la mayoría de genes transcritos en eucariotas son de este tipo (Liu et al., 2015b). Los RNAs no-codificantes (ncRNAs), son un grupo diverso de transcritos que incluyen RNAs constitutivos (e.g. RNA ribosomal, RNA de transferencia) y ncRNA regulatorios (Ariel et al., 2015). Dentro del grupo de ncRNAs regulatorios se encuentran los RNAs largos no-codificantes (lncRNA, por sus siglas en inglés), los cuales son transcritos de una longitud mayor a 200 nucleótidos (nt), que no codifican para una proteína y pueden actuar como reguladores de la transcripción de genes en forma *cis* o *trans*, o como andamios para proteínas en complejos modificadores de cromatina (Bonasio and Shiekhattar, 2014; Cech and Steitz, 2014; Liu et al., 2015b; Rinn and Chang, 2012). Por otro lado, los RNAs pequeños (sRNA) son ncRNA regulatorios, generalmente de una longitud de 20 a 27 nt, que participan en la regulación post-transcripcional de genes y en el mantenimiento de la estabilidad genómica (Axtell, 2013; Vaucheret, 2006). Recientemente, algunos sRNAs y lncRNAs han sido asociados con la regulación de la meiosis y la fertilidad, aunque su función específica aún debe ser clarificada (Axtell, 2015; Grant-Downton and Rodriguez-Enriquez, 2012; Zhai et al., 2015; Zhang et al., 2014b).

Dadas las ventajas (descritas en los capítulos anteriores) que tiene el girasol como modelo para el estudio de la meiosis en plantas, así como las diferencias en la tasa de recombinación homóloga existentes en los genotipos utilizados para este estudio, decidimos llevar a cabo el análisis de expresión diferencial del transcriptoma de meiocitos de girasol de estos genotipos. Interesantemente, la mayor proporción (~64%) de los genes diferencialmente expresados (DEG) fueron genes no codificantes para proteínas, de los cuales el 70% pasó filtros astringentes para ser clasificados como lncRNAs. Pudimos identificar lncRNAs específicos de meiosis, de los cuales además algunos parecen tener funciones asociadas a sRNAs. Sin embargo, su participación en otros mecanismos regulatorios no se puede descartar. Nuestra propuesta es que los lncRNAs pueden tener un papel importante en la regulación de la expresión de genes o en los cambios de los estados de la cromatina durante la meiosis, lo cual podría también estar relacionada con las diferencias de recombinación homóloga entre los genotipos analizados. Los resultados presentados a continuación fueron publicados recientemente por Flórez-Zapata et al. (Flórez-Zapata et al., 2016).

3.2 Materiales y Métodos

3.2.1 Extracción de RNA, secuenciación y ensamblado

El aislamiento de los meiocitos, extracción de RNA, secuenciación, filtrado de calidad de secuencias y re-mapeo, se realizó tal como fue descrito en el Capítulo 2, secciones 2.2.1 y 2.2.2. Por otro lado, se realizó un nuevo ensamblado para llevar a cabo el análisis de expresión diferencial, en el cual se consideraron únicamente las secuencias provenientes del genotipo F1, bajo el supuesto de que en este ensamblado podrían encontrarse transcritos provenientes de los dos genotipos parentales. Este ensamblado *de novo* fue realizado con una versión actualizada del software Trinity (release 20140413) (Grabherr et al., 2011), con los parámetros de defecto. Esta nueva versión del ensamblador, clasifica los transcritos reconstruidos en dos categorías “genes”, correspondientes a secuencias que el algoritmo considera son producto de diferentes genes, y “transcritos”, para los que las secuencias contienen pequeñas diferencias entre ellas, pero no las suficientes para ser catalogadas como un gen diferente, por lo que estos pueden ser variantes de empalme o alelos distintos de un mismo gen.

Cabe resaltar que en una comparación de los ensamblados del transcriptoma de meiocitos obtenidos a partir de cada uno de tres genotipos de girasol, se encontró que el riesgo de ignorar durante el análisis de expresión diferencial a un gen altamente conservado, que tenga expresión exclusiva en alguno de los dos genotipos parentales (Silvestre o Domesticado) es muy baja. Asimismo, no se encontró evidencia que sugiera que el análisis de expresión diferencial pudiera verse distorsionado a favor del genotipo F1 (para información más detallada, puede consultarse el reporte interno del Laboratorio de Biología Computacional, titulado: “*Evaluating putative biases due to differences in assemblies*”). La identificación de estos transcritos producto del ensamblado del genotipo F1, fue realizada de la misma forma, y con las mismas bases de datos de péptidos descritas previamente en la sección 2.2.3.

3.2.2 Identificación de lncRNA.

Dado que todos los transcritos identificados tenían una longitud mayor a 200 nt, no fue necesario realizar un filtrado por longitud. Se utilizaron dos algoritmos diferentes para determinar si un transcrito podría ser el producto de un lncRNA: el análisis del cálculo de potencial codificante en la web del CPC (<http://cpc.cbi.pku.edu.cn>) (Kong et al., 2007), y el análisis con el algoritmo CPAT (Coding-Potential Assessment Tool) (Wang et al., 2013); los dos con los parámetros por defecto. Para el análisis con CPAT, los modelos de genes codificantes para proteína, fueron construidos utilizando el genoma de *A. thaliana* (TAIR10). Solo los genes que pasaron el umbral para los dos análisis (CPC score ≤ -1 y CPAT score ≤ 0.3) fueron catalogados como lncRNAs. Además, realizamos un blastn de los lncRNAs identificados contra el borrador del genoma de la línea HA412 de girasol (Celera_14libs_sspace2_ext.final.scaffolds.split.fasta), depositado en la base de datos del proyecto de secuenciación del genoma de girasol (<http://www.sunflowergenome.org>), al cual tuvimos acceso gracias a la autorización del profesor Loren Rieseberg (University of British

Columbia). Un hit con el genoma fue considerado significativo si el resultado tenía un bitscore ≥ 90 y un valor esperado $E \leq 10^{-6}$.

3.2.3 Secuenciación y mapeo de RNAs pequeños.

El RNA total de meiocitos de los genotipos Silvestre y Domesticado fue extraído utilizando el kit ZR RNA MicroPrep (Zymo Research, Orange, CA), siguiendo las instrucciones del fabricante, con las modificaciones sugeridas para la extracción de sRNAs. Posteriormente, el RNA extraído fue almacenado a -70°C . En total se prepararon dos genotecas (una por cada genotipo), utilizando los kits de preparación de genotecas de Illumina TruSeq de Small RNA, las cuales fueron secuenciadas utilizando el equipo Illumina HiSeq 2500 para obtener lecturas de un sólo extremo de 37 pares de bases de longitud. Las lecturas fueron filtradas y procesadas con el set de herramientas de Kraken (Davis et al., 2013), y aquellas de buena calidad con una longitud entre 20-25 nt fueron seleccionadas para análisis posteriores. Las secuencias de sRNA fueron mapeadas con el programa Bowtie versión 1.1.2 (Langmead et al., 2009) a una referencia mixta de los *contigs* del borrador del genoma (descrito anteriormente) y el transcriptoma de meiocitos del presente estudio, con los siguientes parámetros: `-v 1 --best --strata -a -f --chunkmbs 512`.

3.2.4 Descripción de algunas características de los lncRNAs.

Con el objetivo de determinar si los lncRNAs eran transcritos naturales antisentido (NATs), se descargó la base de datos de NATs predichos para *Helianthus annuus* del PlantNATsDB (<http://bis.zju.edu.cn/pnatdb/>) (Chen et al., 2012), y se realizó la búsqueda de similitud de estos con nuestras secuencias de lncRNAs utilizando el programa blastn. La descripción de transposones (TEs) o elementos repetitivos en los lncRNAs se realizó con una búsqueda de blastn de estos, contra la base de datos de secuencias repetitivas de girasol SUNREP (Natali et al., 2013), RepBase (Versión 21.02) (Jurka, 2000) y PGSB Repeat Element Database (PGSB-REdat) (Nussbaumer et al., 2013). Los valores de corte para estos dos análisis de BLAST fueron iguales: bitscore ≥ 70 y valor esperado $E \leq 10^{-6}$. Adicionalmente, realizamos una búsqueda de elementos repetitivos en nuestros lncRNAs, utilizando la herramienta web de RepeatMasker (<http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker>) (Smit et al., 2013-2015), con el set de opciones "slow", lo que permitió tener una mayor sensibilidad de identificación.

3.2.5 qRT-PCR.

Para los análisis de qRT-PCR, se realizaron nuevas extracciones de meiocitos aislados en profase I. Posteriormente, las muestras de RNA total fueron enviadas a analizar al Laboratorio de Servicios Genómicos, donde se realizó el proceso de síntesis de cDNA y Q-PCR. Un análisis de curva de desnaturalización, fue realizado al final de cada corrida, para verificar que sólo un producto de PCR estaba siendo amplificado. Las diferencias relativas en la expresión de los genes, fueron detectadas por el método de $\Delta\Delta\text{Ct}$, utilizando dos lncRNAs que presentaban baja variación en su expresión como normalizadores (Con1 y G1).

3.2.6 Diseño y análisis estadístico.

El análisis de expresión diferencial fue realizado de la misma manera cómo se describo anteriormente (Sección 2.2.5), con algunas diferencias. En esta oportunidad se realizó la comparación entre los transcriptomas de meiocitos de los genotipos Silvestre y Domesticado. Como se describió anteriormente, sólo fueron consideradas aquellas lecturas que mapeaban a un único transcrito. Sin embargo, para los genes que compartían el mismo hit con un péptido (identificador), los datos de expresión fueron colapsados a un único transcrito, sumando el número de lecturas que mapeaban a cada uno de estos. Las cuentas fueron colapsadas de acuerdo a su identificador en el siguiente orden: TAIR10, HA412, HAXRQ y Refseq. Dado que este proceso de colapsar en un solo transcrito todas las cuentas de aquellos que podrían ser producto del mismo gen sólo fue realizado con los genes codificantes para proteína, decidimos evaluar si en los genes que no pudieron ser identificados la redundancia era alta, como consecuencia de esta falta de agrupamiento. En el reporte interno del Laboratorio de Biología Computacional “Evaluating redundancy of sunflower lncRNA”, se describen con mayor detalle estos resultados, gracias a los cuales se concluyó que a excepción de un par de secuencias que podrían ser alelos de un mismo gen, el set de secuencias no identificadas es no-redundante.

Para calcular el número de genes faltantes en nuestro transcriptoma se utilizó el método descrito por García-Ortega y Martínez (García-Ortega and Martínez, 2015). Las diferencias entre proporciones fueron probadas con una prueba de dos colas para proporciones a un nivel de significancia $P < 0.01$. Todos los análisis estadísticos fueron realizados in R versión 2.15.3 (R Core Team, 2013).

3.3 Resultados y Discusión

3.3.1 Genes diferencialmente expresados entre los meiocitos de girasol.

Se obtuvieron alrededor de 8.6×10^8 lecturas pareadas de los meiocitos de girasol (Tabla S12). Las lecturas del genotipo F1 fueron filtradas por calidad y usadas para el ensamble *de novo* del transcriptoma que fue utilizado como referencia para los análisis posteriores. Cerca del 78% de estas lecturas mapearon a uno sólo de los 73,658 transcritos distintos “genes” de este transcriptoma. Más de la mitad de los transcritos (39,354, 53.42%) fueron anotados via BLAST con una de las bases de datos utilizadas, mientras que los 34,304 (46.58%) transcritos restantes no pudieron ser identificadas a través de esta aproximación. Los transcritos identificados que compartían el mismo identificador de BLAST fueron considerados como productos generados de un mismo locus de girasol, o derivados de un gen parálogo cercano. Para cuantificar la expresión de estos transcritos las lecturas alineadas a estos loci fueron sumadas y los transcritos fueron “colapsados” para tratar cada uno de estos transcritos relacionados como un único gen. De esta manera el set de datos final incluye en total 59.085 genes.

La estimación de genes faltantes en nuestro transcriptoma, con el método de García-Ortega y Martínez (García-Ortega and Martínez, 2015), arrojo que el número de genes faltantes es

igual a cero (con un intervalo de confianza del 95% que va de 0 a 3), indicando que nuestro experimento de RNA-seq detecto prácticamente todos los genes expresados y por tanto no se necesita un muestreo adicional para detectar genes faltantes. Este resultado implica además, que los genes detectados como expresados en meiocitos y no en el transcriptoma somático, pueden ser considerados efectivamente como genes de expresión exclusiva, dado que la profundidad de secuenciación (tamaño efectivo de muestra) fue más que suficiente para este fin.

Encontramos 29,469 (49.87%) genes diferencialmente expresados (DEGs) entre el genotipo Domesticado y Silvestre, utilizando una tasa de descubrimientos falsos esperados (FDR) del 1%. La mayoría de estos genes (63.62%) no pudieron ser identificados via BLAST (Figure 12), lo cual es consistente con nuestros análisis previos en los que los transcritos que no pudieron ser identificados eran más abundantes en meiocitos de girasol con respecto a el transcriptoma somático, y además exhibían una expresión más tejido-específica (Flórez-Zapata et al., 2014). Por tanto, decidimos centrarnos en estos transcritos no identificados, con el propósito de investigar cuantos de ellos podrían ser considerados lncRNAs con un posible papel regulatorio de la recombinación homóloga durante la meiosis.

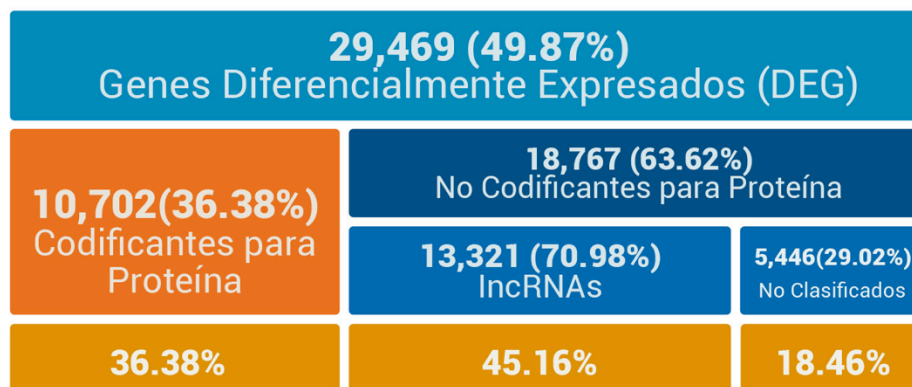


Figura 12. Números y proporciones de genes diferencialmente expresados entre los genotipos Silvestre y Domesticado agrupados por su categoría. El porcentaje global de DEG, fue calculado respecto al número total de genes. Los porcentajes de los genes con y sin potencial de codificar para proteína son presentados respecto al total de DEG. Los porcentajes de lncRNAs y genes no clasificados fueron calculados respecto al total de genes no-codificantes para proteína. Los porcentajes de la última fila, representan los porcentajes de cada categoría, con referencia al número total de DEG.

3.3.2 Identificación de lncRNAs expresados en los meiocitos de girasol.

Para probar si algunos de los transcritos no identificados eran lncRNAs, seguimos el flujo de trabajo descrito en la Figura 13. Primero excluimos todos los 39,354 transcritos los cuales tuvieron un hit significativo con péptidos según nuestro análisis de BLAST. Con los 34,304 transcritos restantes no identificados de realizó un análisis de blastn, contra el borrador del genoma de girasol (línea mejorada HA412). El 90.88% de estos transcritos mostraron un hit

con un porcentaje de identidad mayor al 90% con el borrador del genoma. Aunque este resultado sugiere que la mayoría de estas secuencias son en realidad transcritos de girasol y no artefactos del ensamblado, descartar en este punto aquellas secuencias sin hit con el genoma podría ser inapropiado, puesto que estas pudieran ser variantes producto del procesamiento de RNA o secuencias específicas del genotipo. Entonces decidimos evaluar el potencial codificante de ambos transcritos no-identificados (con y sin hit con el genoma), utilizando dos algoritmos distintos: CPC (Kong et al., 2007) y CPAT (Wang et al., 2013), considerando que sólo aquellos transcritos que pasaran el umbral de los dos algoritmos, fueran clasificados como lncRNAs. Dado que los dos métodos utilizados son complementarios (CPC utiliza un clasificador basado en máquinas de vectores de soporte, mientras que CPAT emplea un modelo logístico de regresión), la clasificación de lncRNAs sólo cuando se cumplan los filtros de los dos algoritmos puede considerarse altamente confiable y fidedigna.

De los transcritos no-identificados con y sin hit con el genoma el 74.17% (23,960) y 68.28% (1,367), respectivamente, fueron clasificados como lncRNAs (Figura 13). Los transcritos no identificados que no pasaron uno o los dos de los filtros de los algoritmos (CPC y CPAT) fueron designados como no-clasificados. Al examinar la expresión de los transcritos clasificados como lncRNAs en los tres genotipos de girasol, encontramos que el 75.65% (18,128) de los lncRNAs con hit en el genoma también mostraron evidencia de expresión en los tres genotipos de girasol evaluados, por lo que estos pueden ser considerados lncRNAs altamente fiables. Por otro lado, el 32.26% (441) de los lncRNAs sin hit en el genoma mostraron expresión en los tres genotipos de girasol (Figura 13). En contraste, la proporción de lncRNAs con hit en el genoma detectados sólo en los genotipos F1 y silvestre fue del 11.03% (2,644 de 23,960), mientras que aquellos sin hit en el genoma representaban aproximadamente el 53.62% (733 de 1,367) del total. Este enriquecimiento en lncRNAs de "origen silvestre" en los lncRNAs sin hit en el genoma podría ser debido a que el genoma que se está utilizando como referencia es de un genotipo Domesticado (línea HA412).

El diagrama de Venn del total de lncRNAs, muestra que todos (a excepción de dos lncRNAs, que sólo fueron detectados en el genotipo Domesticado) fueron detectados en el híbrido F1 (Figura 13). La mayoría (73.31%; 18,569) de los lncRNAs fueron detectados en la F1 y los dos genotipos parentales (intersección interior), mientras que proporciones similares fueron detectadas en la F1 y uno de los genotipos parentales (12.08% (3,062) y 13.33% (3,377) para las comparaciones F1-Domesticado y F1-silvestre, respectivamente). Interesantemente, 317 (1.25%) de los lncRNAs fueron exclusivamente detectados en la F1. Este patrón de expresión podría ser evidencia de la existencia de algún tipo de procesamiento para la maduración del lncRNA genotipo-dependiente. Después de la transcripción, la mayoría de lncRNAs son procesados de manera similar a los RNAs de genes codificantes para proteína, incluyendo la adición de la cubierta en el extremo 5', la poliadenilación del extremo 3' y las modificaciones de empalme (Yoon et al., 2013). Aunque el empalme alternativo de lncRNAs

parece ser menos común en plantas que en animales, en plantas de vid se evidenció la expresión de formas alternativas de empalme en genotipos diferentes (Vitulo et al., 2014).

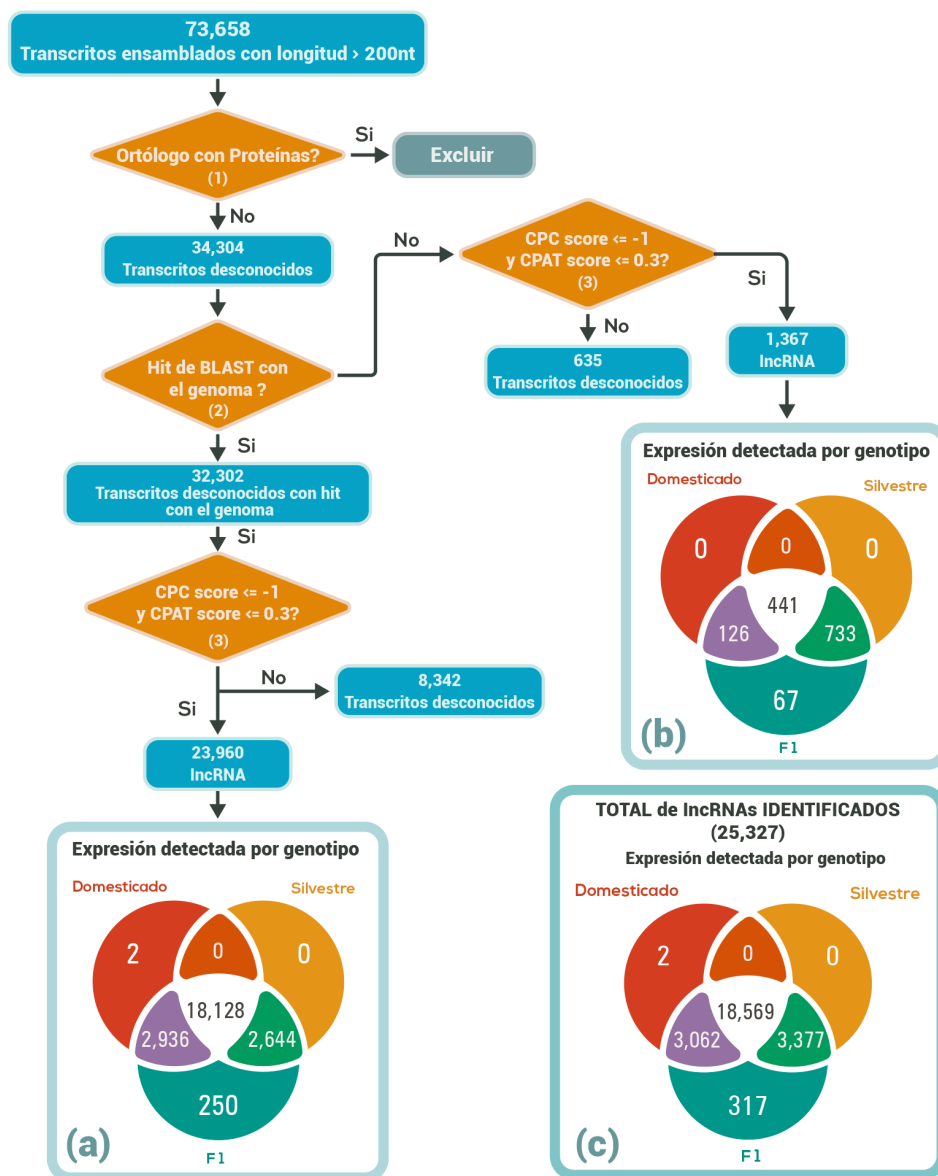


Figura 13. Flujo de trabajo para determinar la naturaleza de lncRNA de los transcritos y diagramas de Venn con el número de lncRNAs expresados por genotipo.

Entre los 25,327 lncRNAs identificados, 13,321 mostraron expresión diferencial entre los genotipos Domesticado y silvestre. La proporción de lncRNAs diferenciales y de transcritos no clasificados fue significativamente mayor que la proporción de genes codificantes para proteínas diferencialmente expresados (45.16% y 36.38%, respectivamente, $P < 0.01$) (Figura 12). Este resultado indica que la mayoría de cambios observados a nivel transcriptómico entre los meiocitos de origen silvestre y Domesticado son debidos a cambios en la expresión

de lncRNAs, lo cual a su vez, podría sugerir que estos están relacionados con las diferencias en la tasa de recombinación entre estos genotipos.

En un estudio reciente, Ding y colaboradores (Ding et al., 2012a) resaltaron la importancia de los lncRNAs en la meiosis de levadura, demostrando que el gen *sme2* que codifica para un lncRNA específico de meiosis, es importante para el reconocimiento y apareamiento de los cromosomas homólogos. Aunque aún no se ha descrito algún lncRNA específico de plantas involucrado en meiosis (Zhou and Pawlowski, 2014a), muchos lncRNAs han sido asociados con reproducción sexual y fertilidad. Por ejemplo, en maíz el lncRNA *Zm401* se cree que es esencial para el desarrollo del tapete y las microesporas (Ma et al., 2008), mientras que en arroz el lncRNA *LDMAR* regula la esterilidad masculina sensitiva al fotoperiodo y el desarrollo normal del polen (Ding et al., 2012b). De hecho, un gran número de lncRNAs importantes para la reproducción sexual, han sido identificados en arroz a través de una estrategia de cribado del genoma completo (Zhang et al., 2014b).

El análisis del transcriptoma de testículos de mamífero mostró que los niveles de expresión de los genes en este órgano eran relativamente mayores a los de otros órganos (e.g. cerebro, corazón, hígado, riñón), y que esta diferencia era más pronunciada para los lncRNAs, los cuales tienen mayor expresión en los testículos que en los otros órganos (Soumillon et al., 2013). Concordantemente, nosotros previamente describimos que en humanos los testículos tenían el transcriptoma más diverso (Martínez and Reyes-Valdés, 2008). Asimismo, el análisis de los patrones de expresión de lncRNAs en tetrápodos mostró que los lncRNAs están preferencialmente expresados en testículos y que su expresión está activamente regulada, lo cual sugiere que su expresión no es debida únicamente a transcripción no-específica en regiones de cromatina abierta (Necsulea et al., 2014). Las redes de co-expresión de lncRNAs con genes que codifican para proteína mostraron que los clusters de genes con mayores proporciones de lncRNAs estaban enriquecidos en funciones relacionadas con espermatogénesis, lo cual concuerda con las altas proporciones de lncRNAs en testículos, así como la contribución sustancial que tienen los espermatoцитos en paquitenio a el transcriptoma del testículo completo (Necsulea et al., 2014; Soumillon et al., 2013). Por otro lado, la caracterización de todos los lncRNAs presentes en el genoma de maíz, puso en evidencia que los tejidos reproductivos masculinos, tales como la borla inmadura, las anteras y el polen, tenían mayores niveles de expresión de lncRNAs que otros tejidos (Li et al., 2014). Dado lo anterior, es posible pensar que este enriquecimiento de lncRNAs en las estructuras y tejidos reproductivos de plantas y animales durante la meiosis y gametogénesis podría ser el resultado de una vía de regulación bien estructurada y conservada de expresión de genes, en la cual participan los lncRNAs.

Actualmente los mecanismos moleculares a través de los cuales los lncRNAs podrían participar en la meiosis de plantas son poco comprendidos. De acuerdo con su función en otros procesos biológicos, estos podrían participar controlando la expresión de genes, influenciando factores epigenéticos, manteniendo las características de la heterocromatina o

controlando elementos transponibles (Zhou and Pawlowski, 2014a). En arroz, más de 700 lncRNAs parecen ser importantes para la biogénesis de siRNAs en fase (phasiRNAs) de 21 nt, en un proceso asociado con la proteína argonauta MEL1 específica de la línea germinal (Komiya et al., 2014), lo cual podría implicar que los lncRNAs en meiosis también podrían actuar como precursores o imitadores de blancos de sRNAs (Wu et al., 2013). Además, los lncRNAs podrían tener un papel estructural directo, similar al del lncRNA DDSR1 de fibroblastos en humanos, el cual interactúa con el gen BRCA2 para modular la reparación del DNA por recombinación homóloga (Sharma et al., 2015); una vía de reparación que también es importante para llevar a cabo correctamente la meiosis en Arabidopsis (Seeliger et al., 2012).

Diferencias en la expresión de lncRNAs entre genotipos y los consecuentes efectos fenotípicos de estas diferencias han sido observados en trigo, en donde la expresión diferencial de dos lncRNAs (TalncRNA73 y TalncRNA108) está cercanamente relacionada con la susceptibilidad a la roya en tres genotipos distintos (Zhang et al., 2013). Además, dos lncRNAs porcinos (linc-ssc2561 y Dnmt3a) mostraron niveles diferentes de expresión entre cerdos domesticados y jabalíes, sugiriendo un posible papel en la domesticación del cerdo (Zhou et al., 2014). En maíz, por ejemplo, ha sido reportada una reconfiguración del transcriptoma como producto de la domesticación (Swanson-Wagner et al., 2012). Lo cual, sumado al efecto previamente reportado de la domesticación sobre la recombinación homóloga (Ross-Ibarra, 2004), podría sugerir que algunos de los lncRNAs diferencialmente expresados identificados en este estudio podrían haber sido blancos de selección durante la domesticación y estar relacionados con las diferencias de recombinación entre los genotipos Silvestre y Domesticado.

3.3.3 Expresión de lncRNAs en meiocitos y el transcriptoma somático de girasol.

Para acercarnos al entendimiento de la función de los lncRNAs de girasol, analizamos algunas características de la expresión de estos en el transcriptoma somático y de meiocitos del genotipo Domesticado (HA89), previamente descrito (Flórez-Zapata et al., 2014). En los meiocitos, la proporción de lncRNAs expresados fue significativamente mayor ($P < 0.01$), respecto al transcriptoma somático (Figure 14A). Igualmente, como se menciono anteriormente, esto corresponde con los altos niveles de expresión de los lncRNAs en estructuras reproductivas descritas en plantas (Li et al., 2014) y animales (Necsulea et al., 2014; Soumillon et al., 2013).

Además, encontramos que la mayoría de transcritos que muestran expresión en uno de los dos transcriptomas (genes exclusivamente expresados), fueron lncRNAs, tanto para el transcriptoma somático como para el de meiocitos (Figure 14B), aunque el número de lncRNAs fue mayor en el transcriptoma de meiocitos. Por otro lado, el nivel de expresión (en transcritos por millón; TPM) de estos lncRNAs fue menor que el observado para los genes que codifican para proteína (Figure 14C), sugiriendo que los lncRNAs de girasol también

tiene mayor tejido-especificidad y menores niveles de expresión que los genes que codifican para proteínas. Estas dos características de los lncRNAs son consistentes con hallazgos previos en otros organismos (Li et al., 2014; Liu et al., 2015b; Zhang et al., 2014b).

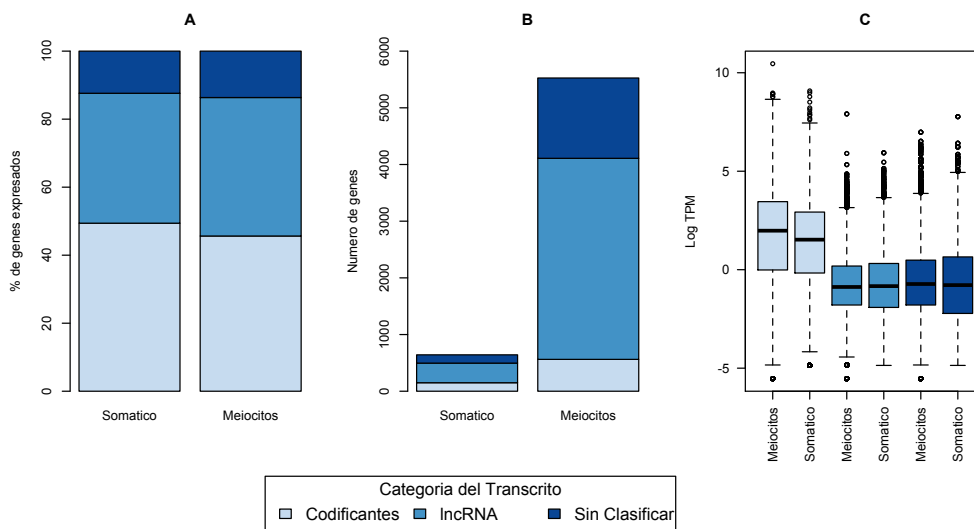


Figura 14. Resumen de la expresión en los transcriptomas somático y de meiocitos, de genes agrupados en las distintas categorías según su potencial codificante. A. Proporciones de genes expresados en cada categoría. **B.** Número de genes expresados exclusivamente. **C.** Distribuciones del logaritmo de la expresión de los genes en cada categoría.

3.3.4 Poblaciones de RNAs pequeños en meiocitos de girasol y su relación con los lncRNAs.

De acuerdo con Li et al. (Li et al., 2014), más del 90% de los lncRNAs de maíz tienen similitud de secuencia con RNAs pequeños (sRNAs), por lo cual ellos los clasificaron como pre-lncRNAs. Adicionalmente en otras especies modelo como arroz, *Arabidopsis* y *Populus trichocarpa*, se ha descrito que los lncRNAs pueden actuar como precursores (Komiya et al., 2014; Ma et al., 2014) o imitadores de blancos de sRNAs (Shuai et al., 2014; Wu et al., 2013). Entonces, para describir la conexión entre los RNAs no codificantes largos y pequeños, secuenciamos el transcriptoma de RNAs pequeños de meiocitos en profase I de los genotipos Silvestre y Domesticado.

Se obtuvieron alrededor de 5 millones de lecturas limpias (Tabla S13) de una longitud de 20 a 25 nt, donde la mayoría correspondían con lecturas de 24 nt (Figure 15). Estos sRNAs de 24 nt son típicamente siRNAs endógenos (Axtell, 2013), y constituyen el mayor componente de poblaciones de RNAs pequeños en plantas, que participan en el silenciamiento de genes basado en cromatina mediado por RNA (Matzke et al., 2009). Durante la meiosis de maíz, los phasiRNAs de 24 nt se acumulan (Zhai et al., 2015) de forma similar a la manera como

son acumulados durante la espermatogénesis de ratón (Axtell, 2015). Aunque la función de estos sRNAs de 24 nt no está completamente entendida, estos pueden participar en la supervisión del mantenimiento de la integridad del genoma (e.g. rutas de silenciamiento de TE), o pueden actuar como señales móviles y/o modificadores de cromatina (Axtell, 2015; Zhai et al., 2015).

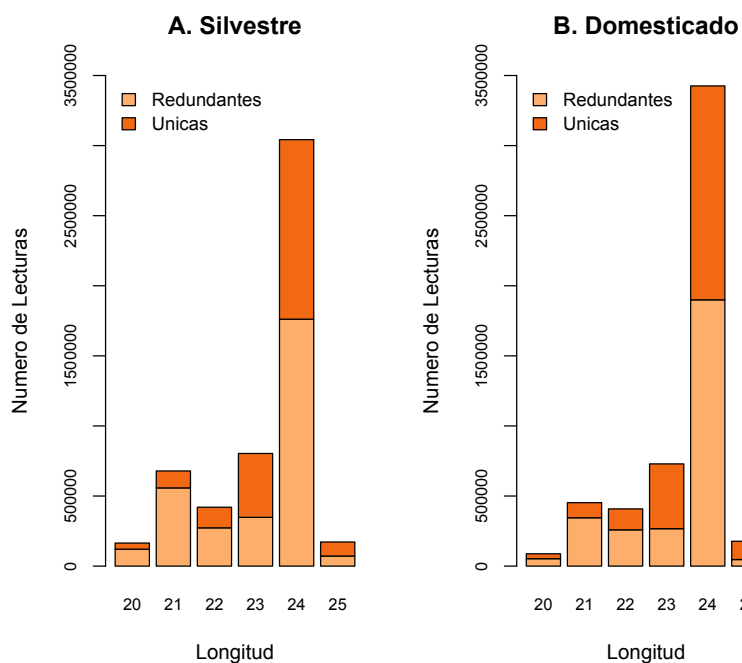


Figura 15. Número de lecturas de cada longitud (pb) de poblaciones de sRNAs de meiocitos de los genotipos de girasol Silvestre (A) y Domesticado (B).

Nosotros previamente encontramos que genes de las vías de síntesis de sRNAs se encontraban altamente expresados en meiocitos, especialmente el ortólogo de la proteína de Arabidopsis AGO4 (2.9 veces más expresado en meiocitos que en el transcriptoma somático) (Flórez-Zapata et al., 2014). Dado que la proteína AGO4 de Arabidopsis una preferencialmente siRNAs de 24 nt con una adenosina en el 5' (Mi et al., 2008), evaluamos la existencia de un sesgo en el nucleótido 5' terminal en los sRNAs de 24 nt de meiocitos de girasol. Encontramos que la mayoría de sRNAs de 24 nt, tenían una adenosina en el 5' terminal (Figura S4), sugiriendo que estos sRNAs de 24 nt podrían ser cargados en el complejo de AGO4 ortólogo de girasol, y así participar en el mantenimiento del estado silenciado de loci repetidos, transposones y heterocromatina, de manera similar a como sucede en Arabidopsis (Vaucheret, 2008). La descripción de esta abundante proporción de sRNAs de 24 nt en los meiocitos de girasol ayuda a soportar la idea de una evolución convergente de los sistemas de sRNA que regulan el desarrollo reproductivo masculino (Axtell, 2015; Zhai et al., 2015).

Consistente con las observaciones en otras especies de plantas (Han et al., 2014; Lelandais-Brière et al., 2010; Pantaleo et al., 2010), la población de sRNAs de 24 nt de girasol es menos redundante que la de 21 nt. Los sRNAs de 24 nt tienen un gran número de lecturas únicas o de poca abundancia, mientras que la población de 21 nt incluye múltiples copias de un mismo sRNA. Por otro lado, dado que la población de sRNAs de 21 nt de plantas está compuesta usualmente por miRNAs, estos podrían ayudar a regular la expresión de genes a través de mecanismos de silenciamiento post-transcripcional, similares a los descritos anteriormente (Montes et al., 2014).

La expresión de 32 familias diferentes de miRNAs fue detectada en los meiocitos tanto del genotipo Domesticado como del genotipo Silvestre (Figura S5). Sin embargo, el 92.69% y el 85.87% de las lecturas en el genotipo Domesticado y Silvestre, respectivamente, correspondían a lecturas de 3 familias de miRNAs (miR166, miR396, and miR319), las cuales están altamente conservadas entre distintas especies de plantas (Montes et al., 2014). Aunque la función de estos tres miRNAs en meiosis es desconocida, en tomate miR396 y miR166 se encontraron diferencialmente expresados entre las anteras de una planta no mutante y de una mutante con esterilidad masculina (7B-1), sugiriendo que estos podrían tener un papel en el desarrollo de la antera y la fertilidad masculina (Omidvar et al., 2015), además de que miR319 fue previamente encontrado con expresión en la línea germinal de *Arabidopsis* (Borges et al., 2011). La abundancia relativa de estas familias de miRNAs fue mas o menos la misma entre los genotipos Silvestre y Domesticado (Figura S5), con excepción de miR398, la cual fue encontrada 7 veces más abundante en los meiocitos del genotipo silvestre. En algodón, miR398 también fue encontrado diferencialmente expresado durante las fases de tétradas en anteras de plantas no mutantes y otras con esterilidad masculina genética (GMS) (Wei et al., 2013). Los hallazgos anteriores indican que el posible rol de miR398 en meiosis merece estudios adicionales.

Para establecer la relación entre poblaciones de sRNAs y lncRNAs, nosotros mapeamos los sRNAs contra una referencia mixta (*contigs* del borrador del genoma y el transcriptoma ensamblado en este estudio), lo cual nos permitió evitar el sesgo de selección, es decir, designar un transcrito como precursor o blanco de un sRNA, cuando el mejor hit para ese sRNA se encuentra en una región intergénica no transcrita. La mayoría de sRNAs de 21 nt mapearon a transcritos que codifican para proteína, mientras que la proporción de estos de 24 nt que mapearon a transcritos que codifican para proteína, es similar a la de aquellos que mapean a lncRNA (Figure 16). Las lecturas de sRNAs del genotipo silvestre mapearon a 9,370 lncRNAs, mientras que las sRNAs del genotipo Domesticado mapearon a 8,852 lncRNAs (42.69% y 40.91%, relativo a la expresión total de lncRNAs expresados en cada genotipo). Las lecturas de sRNAs de ambos genotipos mapearon en común a 8,852 lncRNAs. No obstante, aunque la proporción de lncRNAs que tuvieron similitud de secuencia con miRNAs es notable, esta no es tan alta como la reportada para maíz (Li et al., 2014)

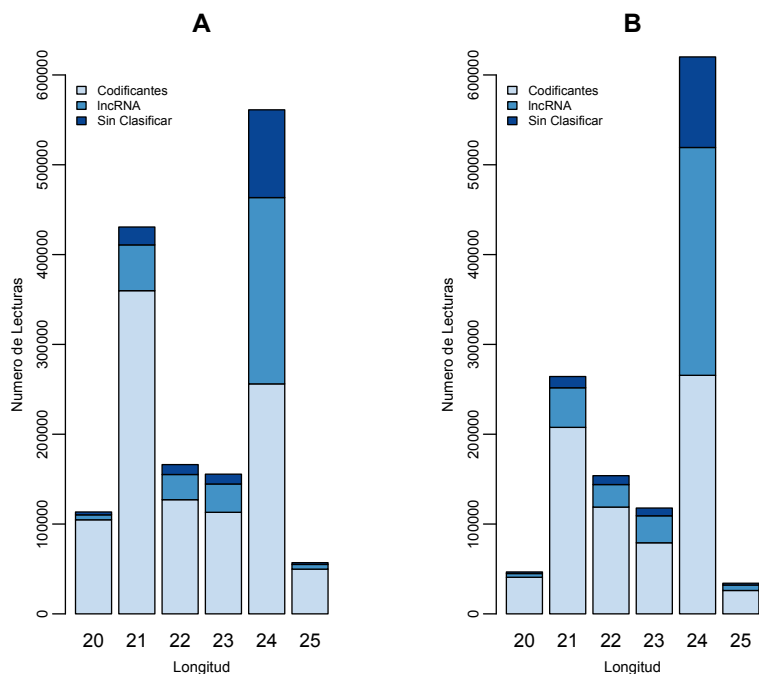


Figura 16. Número de lecturas de cada longitud (pb) de poblaciones de sRNAs de meiocitos de los genotipos de girasol Silvestre (A) y Domesticado (B), que mapearon a las diferentes categorías de transcritos.

De acuerdo con Axtell (Axtell, 2013), los patrones de alineado a la referencia de los sRNAs, proveen información importante acerca de la biogénesis de los mismos. Nosotros encontramos que los sRNAs de 20 y 21 nt mostraron mayores diferencias en la proporción de lecturas que mapearon a los lncRNA de la cadena “+” o “-” (~15% más lecturas mapearon a la cadena “+”; Figura S6). Este resultado es esperado, considerando que los sRNA de este tamaño han sido asociados con miRNA, los cuales son generados a través de la fragmentación de un precursor de cadena sencilla, mientras que los de 24 nt (principalmente siRNAs) provienen de un precursor de doble cadena de RNA, por lo cual puede esperarse que estos últimos mapeen en proporciones cercanas a una u otra cadena.

Por otro lado, de acuerdo a las búsquedas de BLAST contra las bases de datos de transcritos naturales antisentido (NATs) de girasol de la base de datos pública PlantNATs, determinamos que 388 lncRNAs eran transcritos antisentido (trans-NATs). De estos trans-NATs, 198 y 215 para los genotipos Domesticado y silvestre, respectivamente, fueron catalogados como precursores de sRNA, dado que lecturas de sRNAs mapearon a estos. Además, el 68.37% de los NATs para el genotipo silvestre y 68.68% para el genotipo Domesticado estuvieron relacionados con sRNAs de 24 nt. Aunque sRNAs de otras longitudes también parecen ser producto de estos transcritos antisentido, las proporciones son menores (Figura S7). Por tanto, algunos de los lncRNA identificados en los meiocitos parecen ser NAT-siRNAs, los cuales pueden formar parte de mecanismos regulatorios de la expresión génica que

involucran la escisión de transcritos, tal como fue previamente descrito para este tipo de sRNAs (Borsani et al., 2005; Chen et al., 2010b).

En conjunto, nuestros resultados indican que hay una red de sRNAs-lncRNAs trabajando a nivel transcriptómico en los meiocitos de girasol. Además, considerando las diferencias en expresión de los lncRNAs, es posible que los sRNAs que se derivan de estos también estén relacionados con las diferencias fenotípicas a nivel de recombinación homóloga entre estos. Sin embargo, dado que algunos lncRNAs parecen no estar relacionados con rutas de sRNAs, deben existir otros mecanismos mediante los cuales estos participan en la regulación de la recombinación en meiocitos.

3.3.5 Elementos repetitivos presentes en los lncRNAs de girasol.

Se ha propuesto que los elementos transponibles (TE) contribuyen importantemente al origen, evolución y función de los lncRNAs (Johnson and Guigó, 2014). A través del análisis de diferentes regiones de RNAs no codificantes largos intergénicos (lincRNA) en ratón y humanos, Kannan y colaboradores (Kannan et al., 2015) encontraron que el contenido de TEs en los lincRNA es mayor que el de los genes que codifican para proteína, además de que la mayoría de TEs se localizan en los exones y regiones promotoras de los lincRNAs. Estos autores además observaron una correlación entre la inserción de TEs y la tasa de evolución de los lincRNAs (i.e. había mayor fijación de TEs en los genes de lincRNA que evolucionaban más rápido). Por otro lado, en tomate, la inserción de un retrotransposon de tipo LTR fue importante para el origen de un lncRNA específico de fruto (lncRNA-314), el cual presumiblemente fue generado durante la domesticación del tomate dada su relación potencial con la maduración del fruto (Wang et al., 2015).

Nosotros encontramos que 2,326 (9.18%) de todos los lncRNA identificados contenían TEs (Figura 17). De estos, el 91.01% contenían retrotransposones, mientras que el 8.99% contenían transposones de DNA (Figura 17A). Los TEs más frecuentemente encontrados, pertenecían a las familias de retrotransposones LTR Gypsy y Copia (Figura 17B). Estos resultados son consistentes con observaciones previas que mostraron que estas dos familias de retrotransposones también son de las más abundantes en el genoma de girasol (Giordani et al., 2014; Natali et al., 2013). También se identificaron otros elementos repetitivos, tales como repeticiones en tándem o repeticiones no clasificadas (desconocidas) (Figura S8). El porcentaje de lncRNAs que contenían elementos repetitivos fue de 14.31%, el cual es mucho menor al reportado previamente para los lncRNAs de maíz (Li et al., 2014). Aunque las diferencias en el contenido de TEs en los lncRNAs de maíz y girasol pudieran ser debidas a diferencias en las estrategias metodológicas, también podrían estar relacionadas con diferencias reales en el contenido de TE a nivel genómico, diferencias en el tejido de origen de los lncRNAs muestreados (i.e. lncRNAs expresados en meiocitos en girasol vs lncRNAs expresados en toda la planta de maíz), o características intrínsecas de los lncRNAs. Incluso en maíz, la mayoría de los llamados HC-lncRNAs (aquellos que no son precursores de

sRNAs) no contenían secuencias repetitivas (Li et al., 2014), lo cual sugiere que la diversidad de lncRNAs ocurre tanto a nivel funcional como evolutivo, incluso entre la misma especie. Por tanto, se pueden esperar diferencias entre especies en los lncRNAs.

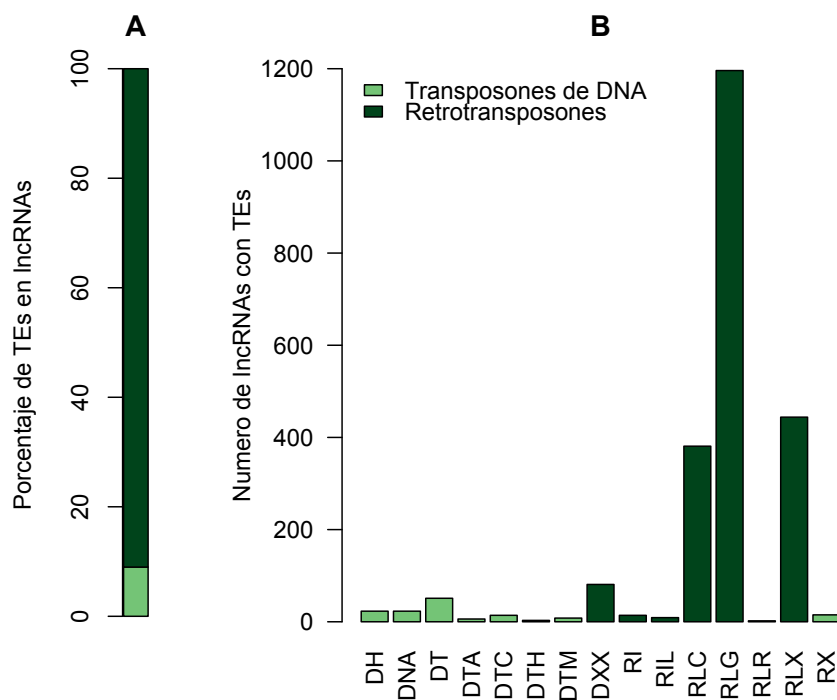


Figura 17. Elementos transponibles (TEs) en los lncRNAs de girasol. **A.** Porcentaje relativo de transposones de DNA y retrotransposones. **B.** Número de lncRNAs que con tienen TEs por familia de TEs de girasol según (Natali et al., 2013). **Transposones de DNA:** DH = Helitron, DT = Unknown DNA transposon, DTA = hAT, DTC = CACTA, DTH = Harbinger, DTM = Mutator, DXX = MITE. **Retrotransposones:** RI = unknown non-LTR retrotransposons, RIL = LINEs, RLC = Copia retrotransposons, RLG = Gypsy retrotransposons, RLR = retroviruses, RLX = Unknown LTR retrotransposons, RX = Unknown retrotransposons.

La presencia de TEs en los lncRNAs de girasol es un resultado interesante, dado que estos elementos podrían ayudar a entender su biología y función. Por ejemplo, la variabilidad de TEs entre plantas domesticadas de girasol es menor que la observada en accesiones silvestres (Giordani et al., 2014). Asimismo, la relación de los TE con la domesticación y el mejoramiento ha sido también reportada en varias plantas, incluyendo maíz y trigo (Olsen and Wendel, 2013), indicando que el contenido de TE entre plantas silvestres y domesticadas es dinámico. Entonces, el proceso de domesticación podría crear un ambiente permisivo para la generación de lncRNAs nuevos o alterados *via* TEs, que a su vez podrían afectar las vías de regulación que lleven a cambios fenotípicos (e.g. diferencias en la tasa de recombinación) en un lapso de tiempo relativamente corto.

Un hallazgo sorprendente del estudio del transcriptoma de meiocitos de *Arabidopsis* (Chen et al., 2010a; Yang et al., 2011) fue la alta expresión de TEs, dado que estos podrían generar mutaciones o tener efectos dañinos sobre el genoma o la línea germinal, pudiendo tener efecto en las generaciones futuras. Aunque la expresión de transposones podría ser un resultado no intencionado de la reorganización de la cromatina durante la meiosis (Zhou and Pawlowski, 2014a), su expresión también podría tener un papel funcional, especialmente dado que estos pueden participar en el remodelamiento de la cromatina a través del reclutamiento de la maquinaria de silenciamiento (Slotkin and Martienssen, 2007). Por tanto, nosotros especulamos que aunque la expresión de algunos TE en meiocitos puede ser genuina, lo cual justificaría la expresión de proteínas como AGO9 que participan en los mecanismos de silenciamiento durante la formación de gametos femeninos (Olmedo-Monfil et al., 2010) y de vías de silenciamiento por siRNAs descritas anteriormente (Axtell, 2015; Zhai et al., 2015), algunos de los TEs encontrados expresados en meiocitos podrían ser en realidad dominios funcionales de los lncRNAs (Johnson and Guigó, 2014). Esta posibilidad implica que los TEs podrían desempeñar su papel en el remodelamiento de la cromatina durante la meiosis, pero su integración en lncRNAs no comprometería la integridad del genoma. Aunque se necesitan más análisis para probar esta hipótesis, la presencia de TEs en los lncRNAs que están altamente expresados en meiocitos abre una nueva perspectiva que podría brindar pistas acerca de su alta y contra intuitiva expresión en la línea germinal.

3.3.6 Validación de las diferencias de expresión por la técnica de RT Q-PCR.

Como se describió en la Sección 2.3.4, la medición de los cambios de expresión por la metodología de RT Q-PCR, permite validar las diferencias de expresión detectadas por técnica de RNA-Seq. Previamente, mostramos que los cambios de expresión por estas dos metodologías eran congruentes para genes meióticos y no meióticos diferencialmente expresados entre el transcriptoma somático y de meiocitos. Aquí, presentamos el análisis de 5 lncRNAs (Figura 18), de los cuales se obtuvieron productos específicos en la PCR, todos ellos diferencialmente expresados entre los genotipos Domesticado y Silvestre, los cuales a su vez presentan similitud con secuencias de sRNAs y hit con el genoma. Uno de ellos, G6, además presenta similitud con un retrotransposon de la familia Copia. Se utilizaron dos lncRNAs con expresión constante como normalizadores (Con1 y G1), lo cual permitió identificar que los lncRNA G4, G8 y G10 presentan mayor expresión en el genotipo Domesticado, mientras que los lncRNAs G5 y G6 tienen mayor expresión en el genotipo Silvestre, consistente con lo observado en el análisis del transcriptoma por el método de RNA-Seq. Además, se determinó que Con1 podría ser mejor control, dado que al calcular la variancia entre genotipos de los cambios de expresión esta fue de 1.7, mientras que para G1 fue de 0.25. Por otro lado, se ve que al normalizar la expresión con Con1 (Figura 18A) la razón de cambio estimada entre las dos metodologías es más similar que al realizar la estimación con G1 (Figura 18B).

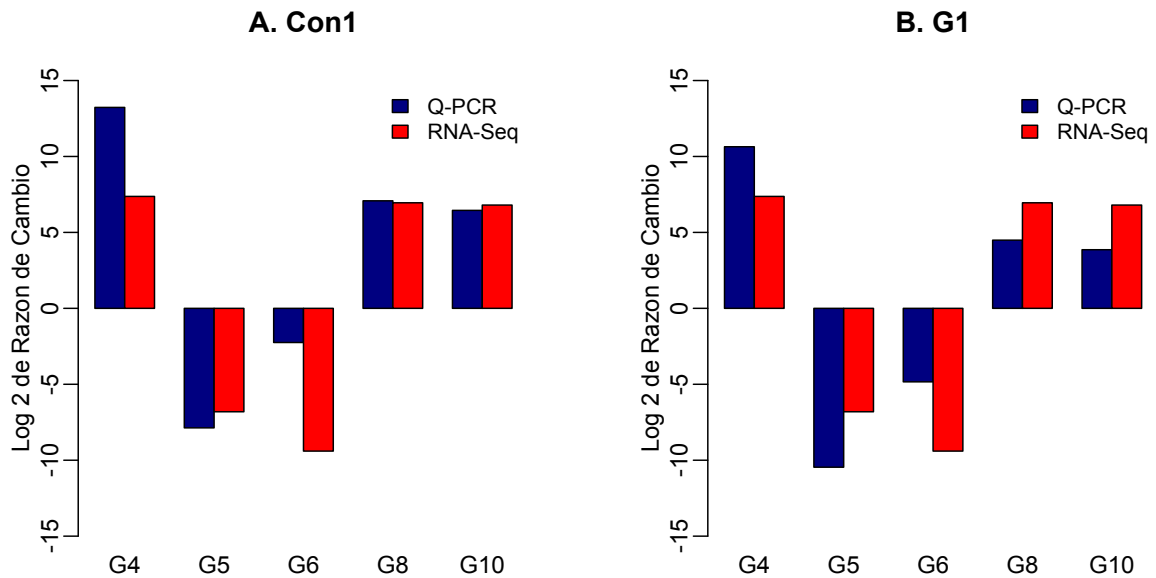


Figura 18. Comparación de los niveles de expresión de algunos genes, estimados por las metodologías de RNA-seq y RT Q-PCR. Para el análisis de la Q-PCR por el método de $\Delta\Delta Ct$ se utilizaron los normalizadores Con1 (A) y G1 (B).

3.3.7 Comentarios finales acerca de los lncRNAs meióticos identificados en girasol.

La idea de que los lncRNAs tienen un papel clave como reguladores es soportada por varios de sus rasgos como: i) funcionalidad inmediata después de la transcripción (i.e. la no necesidad de traducción a proteína); ii) su versatilidad estructural o interacción secuencia-específica con proteínas o ácidos nucleicos; iii) la flexibilidad evolutiva que promueve tolerancia a inserciones o deleciones, dado que su función es independiente de un marco estricto de secuencia como en los genes codificantes para proteína (Johnson and Guigó, 2014). Estos rasgos son consistentes con la evidencia creciente que sugiere que los lncRNAs están involucrados en muchos procesos biológicos, y han ampliado nuestro conocimiento de la regulación de estos y/o cambiado algunos paradigmas (Cech and Steitz, 2014; Liu et al., 2015b). Un punto de encuentro de los estudios que han realizado la identificación a nivel de todo el genoma de lncRNAs, es que estos se encuentran altamente expresados en las estructuras reproductivas de animales (Necsulea et al., 2014) y plantas (Li et al., 2014). Aunque la relación de los lncRNAs con la fertilidad de las plantas también ha sido establecida (Ding et al., 2012b; Ma et al., 2008; Zhang et al., 2014b), y se había propuesto la relación de estos como precursores de phasiRNAs necesarios para la meiosis de arroz (Komiya et al., 2014), ningún lncRNA específico de meiosis ha sido descrito aún en plantas (Zhou and Pawlowski, 2014a). Por tanto, hasta donde sabemos, este es el primer reporte de lncRNAs meióticos específicos de plantas, lo cual plantea cuestionamientos acerca de su función e importancia en la regulación de este proceso fundamental para la reproducción sexual.

Para explorar el posible papel de estos lncRNAs meiosis-específicos expresados en los meiocitos de girasol, comparamos algunos rasgos de estos con los de los meiocitos que además tienen expresión en el transcriptoma somático (Tabla 1). Los lncRNAs con expresión exclusiva en el genoma tienen significativamente mayor número de hits con el genoma de girasol (87.18% vs. 64.77%, $P < 0.01$), mayor expresión diferencial (58.22% vs. 50.49%), mayor similitud con sRNAs (36.77% vs. 23.33%), y mayor contenido de TEs (8.22% vs. 6.46%) que los lncRNAs expresados tanto en meiocitos como en el transcriptoma somático. Estas diferencias de los lncRNAs específicos de meiosis dan algunas pistas sobre su función. Primero, el alto número de hits con el genoma sugieren una mayor conservación, lo cual está de acuerdo con las observaciones previas en tetrápodos, donde los lncRNAs relacionados con espermatogénesis presentan un mayor nivel de conservación (Necsulea et al., 2014). Segundo, los lncRNAs meiosis-específicos parecen estar más involucrados con las diferencias fenotípicas de recombinación con respecto a los no específicos, por lo cual se encuentran diferencialmente expresados en mayor proporción. Finalmente, su alta similitud con sRNAs y alto contenido de TE sugiere que muchos de ellos podrían promover cambios en el estado de la cromatina, lo cual es fundamental dados los procesos de reorganización de esta que tienen lugar durante la meiosis (Dawe et al., 1994; Oliver et al., 2013), así como el papel regulatorio que se piensa tienen los sRNAs y TEs en la modificación de la cromatina (Amaral et al., 2008; Slotkin and Martienssen, 2007).

En este estudio también observamos algunas diferencias en las características de lncRNAs diferencial y no diferencialmente expresados. En los lncRNAs con expresión exclusiva, la proporción de lncRNAs no diferencialmente expresados con hit en el genoma fue menor que la de los lncRNAs diferenciales (43.34% vs. 56.66%), mientras que para los no-exclusivos se observó la tendencia opuesta (mayor proporción en los lncRNA no diferencialmente expresados, 52.94% vs. 47.06%). Estas diferencias soportan la idea de que los lncRNAs específicos de meiocitos podrían estar más cercanamente relacionados con las diferencias en la tasa de recombinación que los lncRNAs no específicos, además, que la mayoría de estos están más conservados o no tienen modificaciones de empalme (lo cual explica que sea mayor el número con hit con el genoma). También encontramos, que los lncRNAs con expresión diferencial, tienen mayor similitud con sRNAs, tanto para lncRNAs meiocito-específicos como no-meiocito-específicos, aunque esta diferencia es más notable para los lncRNAs específicos (57.36% vs. 42.64% y 54.56% vs. 45.44% para lncRNAs meiocito-específicos y no específicos, respectivamente). Dado que los lncRNAs con expresión más variable están más cercanamente relacionados con sRNAs, la función de modificación de la cromatina asociada a la relación lncRNAs-sRNAs podría tener una contribución significativa a las diferencias en la tasa de recombinación. Finalmente, también se observaron diferencias en la proporción del contenido de TEs entre lncRNAs diferenciales y no diferenciales. En los lncRNAs específicos de meiocitos la proporción de lncRNAs no diferenciales con TEs fue ligeramente menor que aquella para los diferencialmente expresados (49.21% vs. 50.79%). Sin embargo, para los lncRNAs no específicos, observamos que dicha diferencia es mayor

(46.31% vs. 53.69% para no-diferenciales y diferenciales, respectivamente). Se necesitan estudios adicionales para clarificar el papel de estos lncRNAs que contienen TE en la meiosis, así como para aclarar si su función está relacionada con la alta frecuencia de TE expresados en meiocitos (Chen et al., 2010a; Yang et al., 2011).

Tabla 1. Resumen de las características observadas de los lncRNAs de girasol. Se presenta el número total de lncRNAs con cada rasgo. Los porcentajes de lncRNAs diferencial y no diferencialmente expresados fueron calculados de acuerdo al total de lncRNAs en cada rasgo (valores en la columna de "Total"). Los porcentajes en el total de cada rasgo (columna "Total") fueron calculados con respecto al número total de lncRNAs independientemente del rasgo (valor en la fila y columna "Total").

	No Diferencialmente Expresados	Diferencialmente Expresados	Total
Expresados en transcriptoma de meiocitos y somático			
Hit con el Genoma	8,925 (49.72%)	9,024 (50.28%)	17,949 (70.87%)
Expresión en los 3 genotipos	8,250 (52.65%)	7,421 (47.35%)	15,671 (61.87%)
Expresión en los 3 genotipos y con hit con el genoma	8,083 (52.80%)	7,227 (47.40%)	15,310 (60.45%)
Similitud con sRNAs	3,035 (44.40%)	3,800 (55.60%)	6,835 (26.99%)
Contiene Transposones	831 (47.25%)	928 (52.75%)	1,759 (6.94%)
Total de lncRNAs identificados	12,006 (47.40%)	13,321 (52.60%)	25,327
Exclusivos de meiocitos			
Genome hit	2,605 (43.34%)	3,406 (56.66%)	6,011 (87.18%)
Expression in 3 genotypes	1,564 (53.97%)	1,334 (46.03%)	2,898 (42.03%)
Expression in 3 genotypes and genome hit	1,520 (53.94%)	1,298 (46.06%)	2,818 (40.87%)
sRNAs similarity	1,081 (42.64%)	1,454 (57.36%)	2,535 (36.77%)
Contains transposons	279 (49.21%)	288 (50.79%)	567 (8.22%)
Total de expresados exclusivamente en meiocitos	2,881 (41.78%)	4,014 (58.22%)	6,895
Expresados en meiocitos y somático (excluyendo exclusivos de meiocitos)			
Genome hit	6,320 (52.94%)	5,618 (47.06%)	11,938 (64.77%)
Expression in 3 genotypes	6,686 (52.34%)	6,087 (47.66%)	12,773 (69.33%)
Expression in 3 genotypes and genome hit	6,563 (52.54%)	5,929 (47.46%)	12,492 (67.77%)
sRNAs similarity	1,954 (45.44%)	2,346 (54.56%)	4,300 (23.33%)
Contains transposons	552 (46.31%)	640 (53.69%)	1,192 (6.46%)
Total de lncRNAs expresados en meiocitos y somático, excluyendo los exclusivos de meiocitos	9,125 (49.51%)	9,307 (50.49%)	18,432

3.4 Conclusiones

En años recientes, cada vez más evidencia sugiere que los lncRNAs tienen un mayor número de importantes funciones regulatorias (Ariel et al., 2015; Bonasio and Shiekhhattar, 2014; Liu et al., 2015b; Yoon et al., 2013). En particular, la identificación y caracterización de lncRNAs en diferentes especies de plantas y animales ha destacado que la expresión de estos es abundante en estructuras reproductivas como testículos y anteras (Li et al., 2014;

Necsulea et al., 2014; Soumillon et al., 2013; Zhang et al., 2014c), y que estos además tienen un alto grado de conservación, sugiriendo que su papel es importante (Necsulea et al., 2014). Aunque en plantas hay evidencia de su relación con la reproducción sexual y la fertilidad (Ding et al., 2012b; Ma et al., 2008; Zhang et al., 2014b), la función que pueden tener estos en meiosis es desconocida, y no se ha habido descrito ningún lncRNA meiótico específico de plantas (Zhou and Pawlowski, 2014a). En este trabajo, gracias al análisis transcriptómico de meiocitos en profase I de tres genotipos de girasol con diferencias en la tasa de recombinación, pudimos observar que la mayor cantidad de genes que cambian su expresión son lncRNAs, de los cuales algunos presentan expresión exclusiva en meiocitos.

Nuestros datos sugieren que, relativo a los lncRNAs no específicos de meiosis, los lncRNAs meióticos están más conservados entre los genotipos relacionados (una mayor proporción tiene hit con el genoma), tienen mayor variabilidad de expresión entre los genotipos contrastantes y además tienen una relación más cercana con elementos relacionados con el remodelamiento de la cromatina como siRNAs y TEs. Estos resultados resaltan el papel fundamental de estos lncRNAs en meiosis, y sugieren una conexión entre dos rasgos previamente observados de los meiocitos y/o anteras: una alta actividad de TEs y una alta frecuencia de siRNAs. Así pues este trabajo presenta evidencia que soporta la idea de posibles funciones meióticas de los lncRNAs, como el mantenimiento de la heterocromatina o la influencia sobre factores epigenéticos, como se había propuesto previamente (Zhou and Pawlowski, 2014a).

Es poco claro si los lncRNAs afectan directa o indirectamente la recombinación homóloga de la meiosis, o si otros factores podrían tener una mayor contribución a las diferencias en la tasa de recombinación en los genotipos de girasol analizados. Sin embargo, la expresión diferencial de lncRNAs específicos de meiosis, sugiere que si bien estos pudieran no ser totalmente responsables de las diferencias en recombinación, sí están relacionados con procesos que conllevan a estas diferencias fenotípicas. Por ejemplo, algunos lncRNAs han sido asociados con la aparición de características de domesticación (Zhou et al., 2014) o fueron modificados durante la domesticación (Wang et al., 2015), sugiriendo que la fuerte selección artificial que ocurre durante la domesticación, puede afectar también a los lncRNAs. Del mismo modo, los lncRNAs son elementos que pueden evolucionar rápidamente, pero algunos de ellos también pueden mantenerse ampliamente conservados (Derrien et al., 2012; Li et al., 2014; Necsulea et al., 2014), por lo que pueden ser fuente importante de innovaciones evolutivas (Amaral et al., 2008) que permitan una mayor flexibilidad en el proceso de selección sin comprometer la reproducción sexual, e.g. permitiendo cambios en la regulación de la meiosis que no afecten a los elementos genéticos esenciales de esta.

Se requieren esfuerzos futuros que se centren en determinar si los lncRNAs específicos o preferencialmente expresados de girasol están conservados en otras especies de plantas, así como para realizar la caracterización de su actividad a nivel funcional. Para esto se requiere

la generación y re-análisis de los datos de expresión en meiocitos y/o estructuras reproductivas de otras plantas, así como la finalización de la secuenciación del genoma del girasol y el desarrollo de herramientas de genómica funcional en esta planta.

CAPÍTULO 4: Cambios transcripcionales de genes meióticos y elementos regulatorios.

4.1 Introducción

El análisis de los patrones de expresión de genes con función conocida en meiosis puede ser útil para elucidar redes de regulación de genes que controlan la expresión de genes meióticos (Zhou and Pawlowski, 2014a). Sin embargo, dos factores han dificultado la realización de este tipo de análisis: en primer lugar, hasta el momento sólo se han descrito alrededor de 90 genes que juegan un papel en meiosis en plantas (Mercier et al., 2015), y segundo muchos de los genes que tienen función en meiosis también participan en otros procesos, e.g. reparación del DNA, por lo que su expresión no es exclusiva de meiosis (Yang et al., 2011). Hasta el momento, se han analizado transcriptomas de meiocitos respecto al transcriptoma somático (Chen et al., 2010a; Dukowic-Schulze et al., 2014; Yang et al., 2011), transcriptomas de anteras en diferentes etapas del desarrollo (Crismani et al., 2006; Deveshwar et al., 2011) y transcriptomas entre anteras mutadas y no mutadas (Wijeratne et al., 2007), por lo que el conocimiento generado de los patrones de expresión de los genes con función meiótica se ha basado principalmente en su expresión preferencial en meiosis y/o co-expresión con genes meióticos conocidos. En este sentido, el presente estudio presenta una ventaja importante puesto que además de la expresión preferencial en los meiocitos, tenemos el criterio adicional de su expresión diferencial entre genotipos de un fenotipo de recombinación meiótica contrastante.

Como se menciono anteriormente, es poco lo que se conoce de la regulación de expresión de genes en la meiosis de plantas. Sin embargo, en otros modelos como levadura este proceso está relativamente bien entendido. En ese organismo la coordinación de la expresión de genes es necesaria para llevar a cabo exitosamente la generación de gametos (Andreuzza et al., 2015). Dos reguladores maestros que median la expresión de genes tempranos (que participan en la fase S y profase) y genes medios (necesarios para la salida de paquiteno, división nuclear y formación de esporas), son los responsables de tal coordinación en la expresión. Estos reguladores son los genes *Ime1* y *Ndt80* en *S. cerevisiae*, y *Ste11* y *Mei4* en *S. pombe* (Vershon and Pierce, 2000; Winter, 2012). Cabe destacar que aunque estas parejas de genes cumplen funciones equivalentes en estas dos especie de levadura, son genes no conservados entre sí (van Werven and Amon, 2011). Lo anterior también sucede en otras especies de hongos en los que los ortólogos de estos genes desempeñan una gran variedad de funciones no relacionadas con su papel en meiosis (Winter, 2012). La falta de conservación de estos reguladores es quizás uno de los factores que ha hecho que los mecanismos que controlan la expresión de genes en plantas estén poco entendidos (Andreuzza et al., 2015).

Aunque el análisis de las diferencias transcripcionales entre los meiocitos de girasol con diferencias en la tasa de recombinación arrojó que el mayor número de genes diferencialmente expresados eran lncRNAs, alrededor del 36.38% eran genes codificantes para proteína. Por tanto, el análisis de este conjunto de genes diferencialmente expresados, la descripción de los patrones de expresión de los genes con función conocida en meiosis, y la búsqueda de elementos regulatorios potenciales, tales como factores de transcripción, pueden dar algunas pistas sobre la regulación transcripcional de la meiosis en plantas. Lo anterior se aborda en el presente capítulo.

4.2 Materiales y Métodos

Los análisis presentados en el presente capítulo fueron realizados con base en el transcriptoma y el análisis de expresión diferencial, descritos en el capítulo anterior. Por tanto, para información detallada al respecto puede consultarse la Sección 3.2. Asimismo la anotación de los genes de este nuevo ensamble en categorías funcionales de términos GO y la identificación de factores de transcripción se realizó de la forma previamente descrita en la Sección 2.2.3, mientras que el análisis de expresión diferencial de estos términos se realizó según lo previamente descrito en la Sección 2.2.5.

4.1 Delimitación de genes con función meiótica y agrupamiento de estos en categorías funcionales.

En el Capítulo 1, se realizó la búsqueda de ortólogos en girasol de genes de Arabidopsis con función en meiosis, basados en lo reportado por Chen et al. y Yang et al. (Chen et al., 2010a; Yang et al., 2011). Sin embargo, para llevar a cabo un análisis más preciso de los patrones y cambios de expresión de estos genes entre los transcriptomas de meiocitos de los diferentes genotipos de girasol, decidimos actualizar dicha lista de genes de Arabidopsis con función conocida en meiosis. Para esto, tomamos la lista reciente de proteínas que han mostrado función en meiosis de plantas descrita por Mercier et al. (Mercier et al., 2015), la contrastamos con la lista de genes previamente considerada, descartamos genes cuya evidencia no sugiera un papel principal en meiosis (e.g. genes cuyo fenotipo mutante no implicaba reducción en la fertilidad y/o alteraciones meióticas), e incluimos algunos nuevos genes cuyo mutante está afectado en meiosis (e.g. HVE). La Tabla S14, presenta la descripción detallada de los genes meióticos que fueron incluidos. Una vez definida esta lista de genes, estos se agruparon en categorías funcionales, tomando como referencia la clasificación previa realizada por Mercier et al. (Mercier et al., 2015), haciendo los ajustes necesarios basados en nuestra propia revisión de literatura.

4.2 Búsqueda de genes co-expresados y agrupamiento en patrones de expresión.

Los genes fueron agrupados en seis patrones de expresión según su nivel de expresión (alto, medio o bajo) en los tres genotipos de girasol analizados (Tabla 3). Según Ballouz et al. (Ballouz et al., 2015), para la construcción de redes de co-expresión a partir de datos de RNA-Seq se requieren como mínimo 20 transcriptomas de al menos 100 millones de lecturas cada uno. Por tanto, dado que no tenemos los transcriptomas necesarios para llevar a cabo

la construcción de redes de co-expresión con los datos generados de transcriptomas de girasol, decidimos basarnos en la información disponible de co-expresión para *Arabidopsis thaliana*. Primero, buscamos los genes que co-expresaban con cada uno de los genes meióticos diferencialmente expresados (entre los genotipos parentales y la F1 respecto a cada uno de ellos) con la aplicación ThaleMine del portal de información de Arabidopsis, Araport (Krishnakumar et al., 2015). Asimismo, utilizamos la herramienta CoExSearch de la base de datos de co-expresión de Arabidopsis ATTED-II versión 8.0 (Obayashi et al., 2007, 2011) para identificar los genes que co-expresaban con los factores de transcripción diferencialmente expresados entre los genotipos Domesticado y Silvestre. Los genes co-expresados en cada uno de estos dos casos fueron anotados funcionalmente con los términos GO, cómo se describió en la Sección 2.2.3. De esta última base de datos, se extrajo además la información de los elementos cis regulatorios predichos para los genes diferencialmente expresados, así como de los factores de transcripción cuyos perfiles de mRNA estaban relacionados con dichos elementos.

4.2 Construcción de un modelo de probabilidad de co-expresión de genes meióticos dada su categoría funcional.

Con el objetivo de evaluar si la probabilidad de co-expresión de cada par de genes meióticos considerados, estaba determinada por la función que desempeña (i.e. a la categoría funcional a la que pertenece), se construyó un modelo de probabilidad lineal. Para esto, primero se buscaron los datos de co-expresión para todos los posibles pares de genes meióticos de Arabidopsis utilizando la herramienta EdgeAnnotation de la base de datos ATTED-II versión 8.0 (Obayashi et al., 2007, 2011). Posteriormente, para cada par de genes, se definió como variable a explicar una variable dicotómica que toma el valor de 1 si se observa co-expresión y 0 en otro caso. Finalmente se estimó el modelo controlando por la categoría funcional a la que pertenecen los genes. Así, los coeficientes obtenidos pueden interpretarse como la probabilidad de que dos genes se co-expresen dada la categoría funcional a la que pertenecen. Del mismo modo se realizó la estimación de este modelo para aquellos genes co-expresados en Arabidopsis, que además presentaban el mismo patrón de expresión en girasol.

4.3 Resultados y Discusión

4.3.1 Diferencias en categorías funcionales (términos GO) entre meiocitos.

Con el propósito de reportar cambios globales de los genes codificantes de proteínas, entre los genotipos de girasol Domesticado y Silvestre decidimos realizar un análisis de expresión diferencial de estos agrupándolos en categorías funcionales de términos GO, tal como se describió previamente (Sección 2.2.5). En total se encontraron 103, 46 y 38 términos diferenciales con un cambio mayor al doble, en las categorías de proceso biológico (BP), componente celular (CC) y función molecular (MF), respectivamente (Tabla 2, S15, S16 y S17). Entre el 89% y 92% de los términos en las distintas categorías (89.32% BP, 91.30% CC y 92.11% MF) mostraron mayor representación en el genotipo Domesticado. Llama la

atención como algunos términos de BP directamente relacionados con HR se encuentran más representados en el transcriptoma del genotipo Domesticado (el cual presenta mayores niveles de HR), tales como: “chromosome separation”, “sister chromatid segregation” y “meiotic DNA double-strand break processing” (Tabla 2), lo cual sugiere que la mayor expresión de genes agrupados en estas funciones, podría estar relacionada con el fenotipo de recombinación observado.

Tabla 2. Términos GO que pueden estar relacionados con meiosis, diferenciales entre los genotipos Domesticado y Silvestre. Los términos subrayados también fueron encontrados diferenciales con mayor expresión en meiocitos (respecto al transcriptoma somático, descrito en el Capítulo 2).

Identificador	Término	Razón cambio: D/S
Términos GO de Proceso Biológico		
<u>GO:0051304</u>	chromosome separation	4.11
<u>GO:0010234</u>	anther wall tapetum cell fate specification	4.05
GO:0007076	mitotic chromosome condensation	3.28
GO:0048444	floral organ morphogenesis	3.03
<u>GO:0000819</u>	sister chromatid segregation	3.01
GO:0060968	regulation of gene silencing	2.70
GO:0060148	positive regulation of posttranscriptional gene silencing	2.57
GO:0042753	positive regulation of circadian rhythm	2.55
GO:0007276	gamete generation	2.52
<u>GO:0000706</u>	meiotic DNA double-strand break processing	2.35
GO:0010480	microsporocyte differentiation	2.24
<u>GO:0006446</u>	regulation of translational initiation	2.19
GO:0060184	cell cycle switching	2.17
GO:0007018	microtubule-based movement	2.08
GO:0032197	transposition, RNA-mediated	2.06
<u>GO:0045787</u>	positive regulation of cell cycle	2.01
<u>GO:0045143</u>	homologous chromosome segregation	0.43
GO:0048235	pollen sperm cell differentiation	0.33
Términos GO de Componente Celular		
GO:0043626	PCNA complex	3.18
<u>GO:0009330</u>	DNA topoisomerase complex (ATP-hydrolyzing)	2.86
<u>GO:0008278</u>	cohesin complex	2.62
GO:0045298	tubulin complex	2.56
<u>GO:0000796</u>	condensin complex	2.31
GO:0031304	intrinsic component of mitochondrial inner membrane	2.20
GO:0005643	nuclear pore	2.10
<u>GO:0016591</u>	DNA-directed RNA polymerase II, holoenzyme	2.02
<u>GO:0032300</u>	mismatch repair complex	0.45
Términos GO de Función Molecular		
GO:0032947	protein complex scaffold	3.65
GO:0019210	kinase inhibitor activity	2.76
GO:0003916	DNA topoisomerase activity	2.30
<u>GO:0001671</u>	ATPase activator activity	2.20
GO:0003712	transcription cofactor activity	2.18
GO:0019948	SUMO activating enzyme activity	2.13
GO:0031490	chromatin DNA binding	2.02

Otro grupo de términos funcionales que destacan son los relacionados con cambios en la estructura de la cromatina, puesto que como se mencionó anteriormente durante la meiosis ocurren procesos de reorganización de esta (Dawe et al., 1994; Oliver et al., 2013). En este conjunto de términos se destacan: “DNA topoisomerase complex (ATP-hydrolyzing)”, “cohesin complex” y “condensin complex” para términos GO de CC; y “DNA topoisomerase activity” para términos GO de MF. Lo anterior, está de acuerdo con los hallazgos descritos en el Capítulo 3, en donde se encontró que la mayoría de los genes diferencialmente expresados eran lncRNAs, los cuales a su vez parecen estar relacionados con procesos que median los cambios estructurales de la cromatina (Flórez-Zapata et al., 2016).

Por otro lado, este análisis diferencial de términos GO pone en evidencia diferencias entre los dos genotipos que pueden deberse a un mecanismo adicional de regulación de la HR: las modificaciones post-traduccionales reversibles. Algunos términos que indican cambios en dichas modificaciones son: “SUMO activating enzyme activity” y “kinase inhibitor activity”, relacionados con los procesos de sumolización y fosforilación, dos modificaciones previamente descritas que pueden afectar la actividad de proteínas relacionadas con la HR (Heyer et al., 2010). Finalmente, se observaron cambios en términos que habían sido previamente encontrados con mayor representación en meiocitos (Flórez-Zapata et al., 2014), tales como los relacionados con el metabolismo energético (“ATPase activator activity”), y el silenciamiento génico (“positive regulation of posttranscriptional gene silencing” y “regulation of gene silencing”). Lo anterior reafirma la importancia de estas funciones en el meiocito, y sugiere que estas pueden verse afectadas y/o estar relacionadas con las diferencias de recombinación de los genotipos.

4.3.2 Patrones de expresión de genes meióticos.

La lista final de genes meióticos actualizada incluida en esta etapa del estudio, esta compuesta por 109 genes con función en meiosis (Tabla S14), para los cuales se encontraron 96 (88.07%) ortólogos en girasol. No es sorprendente que no haya sido posible identificar ortólogos de girasol para todos los genes con función en meiosis descritos en *A. Thaliana*, pues como se describió anteriormente, la mayoría de los genes meióticos para los cuales no fue posible encontrar ortólogo en girasol, también se encuentran ausentes en otras especies de plantas (Flórez-Zapata et al., 2014), lo cual es evidencia de que pueden existir diferencias particulares de este proceso entre especies de plantas.

Con el propósito de entender el comportamiento global de la expresión de los genes meióticos clasificamos su expresión en 6 patrones distintos (descritos en la Tabla 3), basados en el nivel en que se expresan (alto, medio o bajo) en cada uno de los tres genotipos de girasol. Se observó que estos genes están significativamente más representados ($P < 0.01$) en patrones en los que la expresión es mayor en la F1 (5 y 6), comparado con el resto de genes expresados, 44.79% vs 32.47% (Figura 19D y Figura 19A). Esta mayor proporción de genes meióticos en los patrones 5 y 6, también fue observada al compararse con los genes codificantes para proteína y lncRNAs (36.31% y 27.74%, respectivamente) (Figura 19B y Figura 19C). Asimismo, al analizar este mismo comportamiento utilizando como referencia

el ensamble *de novo* construido con secuencias de meiocitos y el transcriptoma somático del genotipo Domesticado (Flórez-Zapata et al., 2014), los genes meióticos presentan también una proporción significativamente mayor ($P < 0.01$) en los patrones con mayor expresión en la F1 (Figura S9), comparado con el total de genes (36.31% vs. 23.34%). Esto es importante, puesto que significa que este patrón particular de expresión de los genes meióticos es consistente, independientemente del transcriptoma de referencia (Domesticado vs F1), o con la naturaleza de genes (codificantes y lncRNAs) respecto a la cual sea comparada.

Tabla 3. Patrones de expresión definidos en función del nivel de expresión de los genes en cada uno de los genotipos de girasol.

Patrón	Nivel de Expresión		
	Alto	Medio	Bajo
1	Domesticado	F1	Silvestre
2	Domesticado	Silvestre	F1
3	Silvestre	Domesticado	F1
4	Silvestre	F1	Domesticado
5	F1	Silvestre	Domesticado
6	F1	Domesticado	Silvestre

La observación de mayores niveles de expresión de genes meióticos en los meiocitos de la F1, es una observación interesante, puesto que previamente se habían descrito efectos de la heterocigosidad sobre la recombinación de *A. thaliana* (Barth et al., 2001; Ziolkowski et al., 2015). Por ejemplo, Barth et al. (Barth et al., 2001) observaron que en promedio las frecuencias de recombinación en las cruza F1 fueron sustancialmente mayores que los ecotipos parentales homocigotos. En girasol no se observó que la recombinación fuera mayor en la F1 respecto a sus parentales (Rodríguez et al., 2007). Sin embargo, el número de brazos cromosómicos con quiasmas previamente estimado de este genotipo (28.10) es mayor que el esperado si la tasa de recombinación de este fuera la mitad (27.65), de la de los dos parentales (29.70 y 25.60, para el genotipo Domesticado y Silvestre, respectivamente). Aunque se necesitan estudios posteriores para determinar si estos efectos de la heterocigosidad sobre la recombinación podrían ser causa o consecuencia de mayores niveles de expresión de genes meióticos en la F1, este resultado es importante puesto que resalta cierta particularidad de la expresión de estos genes que podría impactar en el fenotipo de recombinación. Por otro lado, la heterocigosidad de genes relacionados con el ciclo celular como p53 y la reparación del DNA como ATM, BRCA1 y BRCA2, ha sido descrita como causa de cambios de expresión de varios genes, incluyendo genes blanco y relacionados con su función en ratón, humano y líneas celulares (Boley et al., 2002; Salmon et al., 2013; Watts et al., 2002).

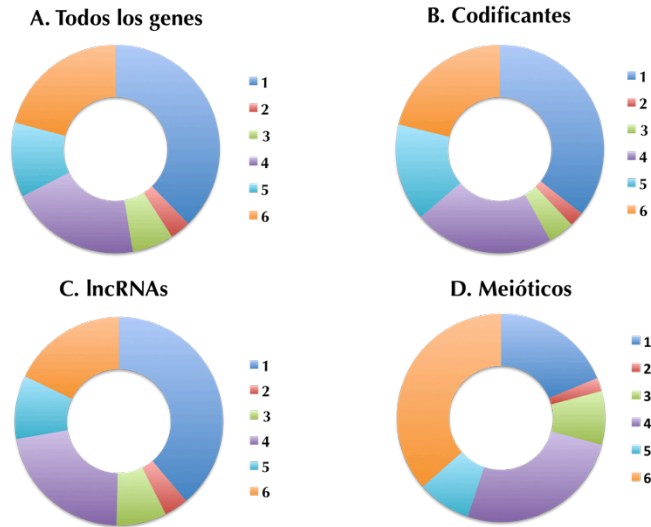


Figura 19. Patrones de expresión de los genes Totales (A), Codificantes (B), lncRNAs (C) y Meióticos (D).

4.3.3 Genes meióticos diferencialmente expresados.

Luego de describir los patrones generales de expresión de los genes meióticos, procedimos a observar los cambios de expresión de estos entre los genotipos de girasol. A un valor de significancia $P < 0.01$, se observaron 34 genes meióticos diferencialmente expresados (33 con cambios mayores al doble) en el contraste Domesticado-Silvestre (D-S), 23 para el contraste F1-Silvestre (F1-S, 20 de los cuales también se encontraban diferenciales entre D-S) y 0 genes meióticos diferenciales entre los genotipos F1-Domesticado (F1-D) (Figura 20A). Estos resultados son congruentes con las diferencias fenotípicas de recombinación previamente reportadas para estos genotipos por Rodríguez et al. (Rodríguez et al., 2007), en los cuales las mayores diferencias del número de brazos cromosómicos con quiasmas se observaron en la comparación D-S (4.1), seguidas de las diferencias entre F1-S (2.5) y F1-D (1.6). Para facilitar el análisis, los genes con función en meiosis fueron además agrupados en 13 categorías funcionales (Figura 20B), de acuerdo a la función que desempeñan. La mayor cantidad de genes diferencialmente expresados fueron los relacionados con el control del ciclo celular, la formación del huso y la citocinesis, seguido de genes que están relacionados con la entrada en meiosis, con 7 y 5 genes diferenciales en la comparación D-S , respectivamente.

Recientemente, se ha descrito la función de dos genes regulatorios de la meiosis de *A. thaliana* que parece pueden modular la expresión génica. Estos genes son: MMD/DUET, una proteína esencial para la meiosis masculina y FEHLSTART (FST), una proteína específica de plantas que controla la entrada y sincronización de la meiosis. Curiosamente, los fenotipos mutantes de estos dos reguladores se caracterizan por problemas en la formación del huso (Andreuzza et al., 2015; Li et al., 2015), lo cual podría sugerir que una de las funciones

meióticas que más se ve afectada por cambios en la regulación de la expresión es la formación del huso. Ortólogos de girasol para estos dos genes fueron identificados, y ambos genes se encontraron diferencialmente expresados en la comparación D-S (Figura 20C). FST presenta mayor expresión en el genotipo Domesticado y MMD/DUET presenta mayor expresión en el genotipo Silvestre. Asimismo, también se detectó expresión diferencial de algunos de los genes regulados por estos genes. El mutante *fst-1* presenta una sobreexpresión de los genes MSH5, ZIP4 y SHOC1 (Li et al., 2015). Esto concuerda con nuestro análisis de expresión diferencial, en el que estos genes presentan mayor expresión en el genotipo Silvestre (Figura 20C, Figura S10C) y un patrón de expresión opuesto (1 y 4, Tabla S14). Es decir, que a menor expresión de FST mayor expresión de MSH5, ZIP4 y SHOC1 en el genotipo Silvestre, o viceversa en el genotipo Domesticado. De manera similar, MMD/DUET ha sido descrito como un regulador de la expresión de TDM1(MS5), por lo que el mutante de este primer gen presenta un descenso considerable de la expresión del segundo (Andreuzza et al., 2015). En nuestro análisis, estos dos genes presentan mayor expresión en el genotipo Silvestre (Figura 20C) y el mismo patrón de expresión (Tabla S14), sugiriendo una vez más que hay correspondencia entre nuestros resultados y los reportes previos de expresión de estos genes.

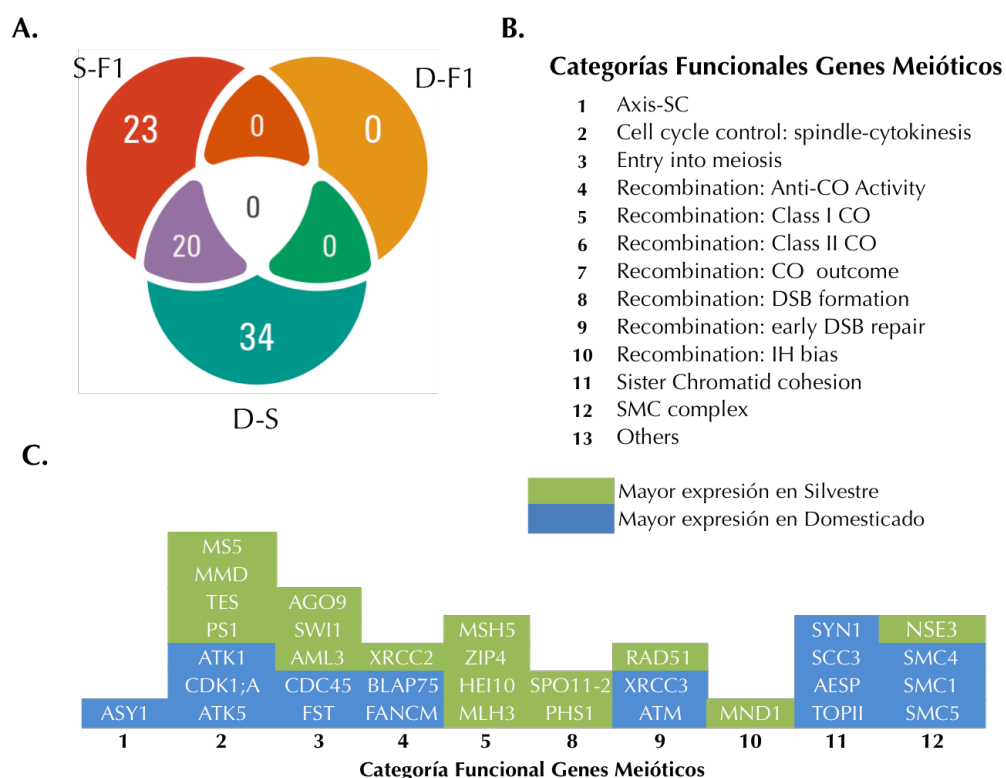


Figura 20. Resultados de expresión diferencial de los genes meióticos entre los genotipos de girasol. (A) Número de genes diferencialmente expresados en cada uno de los contrastes. (B) Categorías funcionales en las que fueron agrupados los genes meióticos. (C) Genes diferencialmente expresados en cada una de las categorías funcionales.

Se podría esperar que el genotipo Domesticado, que presenta mayor tasa de recombinación, también presente mayores niveles de expresión de genes relacionados con la formación de CO. Sin embargo, los genes relacionados con la formación de CO (categorías funcionales 5 y 6), se encuentran principalmente más expresados en el genotipo Silvestre (Figura 20C y Tabla S14). Pueden hacerse algunas reflexiones alrededor de este resultado inesperado. La primera es que en girasol pueden existir otras vías y/o genes que participan en la formación de COs, las cuales podrían estar más directamente relacionadas con las diferencias en la tasa de recombinación observadas. Esta posibilidad surge de observaciones previas, en las cuales las vías de generación de COs pueden variar de un organismo a otro, e.g. en *S. pombe* todos son generados a partir de la ruta de generación de CO de clase II (Mus81-Mms4), mientras que esta ruta es responsable de sólo un subconjunto de los CO (alrededor del 20%) en *S.cerevisiae* (Kohl and Sekelsky, 2013). Asimismo, la existencia de una ruta alternativa de la generación de COs ha sido propuesta en *S.cerevisiae* en donde después de mutar genes esenciales de cada una de las dos rutas para la generación de CO (Msh4-Msh5 y Mus81-Mms4), se observaron algunos COs residuales (De los Santos et al., 2003).

Por otro lado, estos resultados inesperados de expresión podrían sugerir que las diferencias en el número de COs entre los genotipos de girasol, estarían relacionadas con otros mecanismos de control del número de COs, tales como la interferencia. Al estudiar las adaptaciones meióticas en *A. arenosa*, posteriores a un evento de duplicación de genoma completo (WGD), Yant et al. (Yant et al., 2013) observaron que 8 genes meióticos no relacionados entre sí mostraron marcas de selección. De estos genes se destaca que varios de ellos son cohesinas y proteínas del eje cromosómico, tales como ASY1, ASY3, SMC3, SYN1 y PDS5. Por lo cual, se propone que algunas de estas proteínas seleccionadas podrían afectar la longitud del eje y el tamaño de los báculos de cromatina, lo que a su vez modificaría las tasas de recombinación (Bomblies et al., 2015). En nuestro transcriptoma, encontramos que genes del complejo de cohesinas y relacionados con el eje como SMC1, SCC3, ASY1 y SYN1, se encuentran diferencialmente expresadas en la comparación D-S, todos con mayor expresión en el genotipo Domesticado. Esto, sugiere diferencias en la estructura del eje, que a su vez traerían consecuencias en la recombinación, tal como se mencionó anteriormente. A la anterior observación, se suma el hecho de que se observa mayor expresión de antirecombinasas y genes de reparación temprana de DSBs en el genotipo Domesticado (Figura 20C y Figura S10C), dos procesos que también han sido asociados al fenómeno de interferencia (Chapman et al., 2012; Youds et al., 2011).

Finalmente, es importante resaltar que es probable que existan otros mecanismos de regulación posteriores a la transcripción (e.g. modificaciones post-traduccionales), que contribuyan al fenotipo observado y generen que no haya correspondencia entre el perfil transcripcional observado y el fenotipo. Por tanto, el valor de este análisis de expresión diferencial radica en la generación de hipótesis sobre los mecanismos que pueden estar relacionados con las diferencias de recombinación (e.g. cambios transcripcionales de

elementos reguladores como MMD y FST, o cambios en la interferencia), más que en la explicación del porqué de dichas diferencias. Por otro lado, este análisis deja en evidencia algunas particularidades de la expresión de estos genes meióticos. Por ejemplo, MLH3 (AT4G35520) se encontró diferencialmente expresado en el contraste D-S, con mayor expresión en el genotipo Silvestre. Este mismo gen también fue encontrado con marcas de selección y diferencialmente expresado entre teosintle y maíz, con mayor expresión en teosintle (Swanson-Wagner et al., 2012). Se necesitarán estudios posteriores para aclarar si esta observación es solo una coincidencia, o podría estar relacionada con algún mecanismo conservado propenso a ser seleccionado, que contribuya a las diferencias de recombinación previamente observadas entre plantas silvestres y domesticadas (Ross-Ibarra, 2004).

4.3.4 Coordinación por etapas de la expresión de genes meióticos.

Una observación que llama la atención de la expresión diferencial de los genes meióticos, es que los cambios en expresión no ocurren en la misma dirección para genes que forman parte de la misma categoría funcional. Es decir, pueden existir genes diferencialmente expresados de una categoría, que tengan mayor expresión en el genotipo Domesticado y otros de la misma categoría que tengan mayor expresión en el genotipo Silvestre (Figura 20C). Por tanto, decidimos evaluar el patrón de expresión de los genes meióticos en cada una de las categorías funcionales en la que estos fueron agrupados. Para todas las categorías funcionales con más de un gen (i.e. a excepción de la categoría “Recombination: Class II CO”), se observa más de un patrón de expresión (Figura 21). Las categorías que mayor diversidad de patrones de expresión mostraron, fueron “Cell cycle control: spindle-cytokinesis” y “Recombination: early DSB repair”, entre las cuales se encuentran genes del ciclo celular y reparación del DNA, que no cumplen funciones exclusivas de meiosis (Yang et al., 2011; Zhou and Pawlowski, 2014b), lo que podría tener un efecto sobre la coordinación de su expresión. Por tanto, las diferencias de expresión de genes con una misma función, implica que aunque dos genes tengan una función similar, su patrón de expresión no está determinado por esta, sugiriendo, que no hay coordinación o congruencia funcional de la expresión, durante la meiosis de girasol.

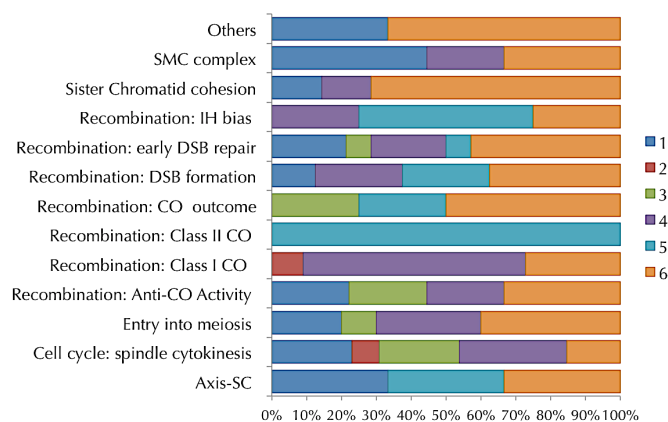
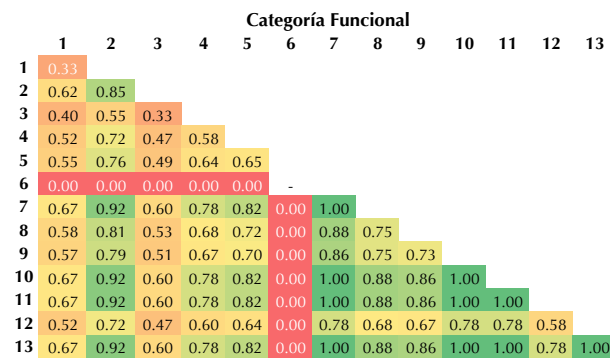


Figura 21. Patrones de expresión de genes meióticos en cada una de las categorías funcionales consideradas.

Esto último nos motivó a evaluar si existía coordinación de la expresión durante la meiosis de plantas, a través de la estimación de la probabilidad de co-expresión de cada par de genes meióticos, dada la categoría funcional a la que pertenecen. En Arabidopsis (Figura 22A), los resultados sugieren la existencia de cierto nivel de coordinación de la expresión, puesto que pares de genes de categorías que dependen secuencialmente unas de otras tienen mayor probabilidad de co-expresión, que con genes que cumplen funciones estructurales. e.g. “Recombination: CO outcome” y “Recombination: Class I CO” (categorías 7 y 5, $P=0.82$) vs “Axis-SC” y “Recombination: Class I CO” (categorías 1 y 5, $P=0.55$), o “SMC complex” y “Recombination: Class I CO” (categorías 12 y 5, $P=0.64$). Los resultados del ejemplo anterior, son esperados si se considera que en esta planta la formación de CO de tipo I, ocurre en etapas posteriores al ensamblaje del complejo sinaptonémico (Mercier et al., 2015; Osman et al., 2011). Nuestros resultados aportan evidencia adicional de que en plantas la transcripción de genes meióticos es controlada etapa por etapa, de manera similar a como ocurre en levadura, tal como lo sugirieron Andreuzza et al. (Andreuzza et al., 2015).

A. Arabidopsis



B. Girasol

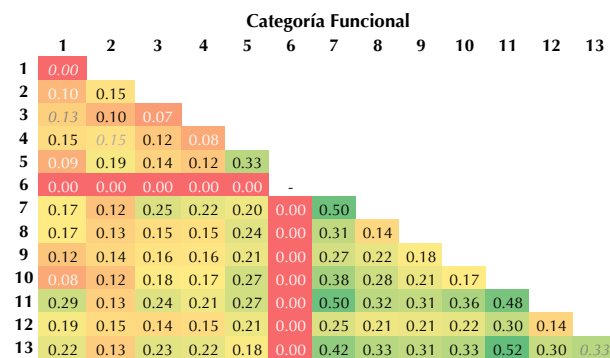


Figura 22. Probabilidades estimadas de co-expresión de genes meióticos dada la categoría funcional a la que pertenece en Arabidopsis (A) y girasol (B). Categorías funcionales: 1: Axis-SC, 2: Cell Cycle Control: Spindle-Cytokinesis, 3: Entry into meiosis, 4: Recombination-AntiCO Activity, 5: Recombination-Class I CO, 6: Recombination-Class II CO, 7: Recombination-CO outcome, 8: Recombination-DSB formation, 9: Recombination-Early

DSB repair, 10: Recombination-IH bias, 11: Sister chromatid cohesion, 12: SMC complex, 13: Others.

Este modelo de probabilidad de co-expresión dada la categoría funcional de los genes meióticos también fue estimado para girasol. Dado que no hay información extensa de datos de co-expresión en esta planta, para considerar un gen como “co-expresado”, se consideró que este tuviera reportes de co-expresión en Arabidopsis y además presentará el mismo patrón de expresión entre los meiocitos de los diferentes genotipos de girasol. El resultado de este análisis se presenta en la Figura 22B, en donde se observa que el resultado es muy similar a lo observado en Arabidopsis. Los genes de categorías relacionadas funcionalmente en una etapa cercana presentan mayor probabilidad de co-expresión que genes pertenecientes a categorías funcionales de etapas diferentes. De hecho, la correlación entre las probabilidades estimadas para Arabidopsis y girasol es de 0.79. Este resultado, junto con la consistencia observada en girasol de la relación entre genes reguladores de expresión previamente descritos en Arabidopsis (i.e. MMD/DUET y FST) y los genes que regulan, sugiere la existencia de reguladores transcripcionales clave, los cuales además podrían estar conservados entre diferentes especies.

4.3.5 Diferencias en elementos regulatorios de la expresión de genes (factores de transcripción y elementos en cis).

El transcriptoma de los meiocitos de plantas se ha caracterizado por una alta actividad transcripcional y sobre-representación de términos GO relacionados con la regulación de la actividad transcripcional (Dukowic-Schulze et al., 2013; Flórez-Zapata et al., 2014). Por otro lado, muchos de los cambios fenotípicos de la domesticación han sido debido a cambios adaptativos en factores de transcripción (Doebley et al., 2006; Swanson-Wagner et al., 2012). Estos dos precedentes motivaron el análisis detallado de los cambios de expresión de factores de transcripción entre los meiocitos de girasol Domesticado y Silvestre. En total, 255 TF fueron encontrados diferencialmente expresados en el contraste D-S. La mayor proporción de estos (168, 65.88%), presentó mayor expresión en el genotipo Domesticado, de los cuales el 85.11% (143) tiene diferencias de expresión mayores o iguales al doble. Por otro lado, 169 (66.27%) de estos presentan mayor expresión en meiocitos, respecto al transcriptoma somático previamente descrito (Capítulo 2).

La mayoría de TF diferencialmente expresados pertenecen a la familia bHLH (Figura 23), la cual fue previamente descrita como la más abundante en meiocitos de girasol (Flórez-Zapata et al., 2014), y de la cual la mayor cantidad de miembros se encuentran sobreexpresados en el transcriptoma de meiocitos de Arabidopsis y maíz (Dukowic-Schulze et al., 2013). A diferencia de los TF de la familia WRKY, las familias a las que pertenecen los TF diferenciales con mayor representación (MYB, bZIP, NAC, C2H2, entre otros) también fueron previamente descritas como abundantemente expresadas en meiocitos (Dukowic-Schulze et al., 2013; Flórez-Zapata et al., 2014). Por tanto, es posible que las variaciones entre los genotipos de esta familia de TF (WRKY), no estén directamente relacionadas con las diferencias en recombinación meiótica.

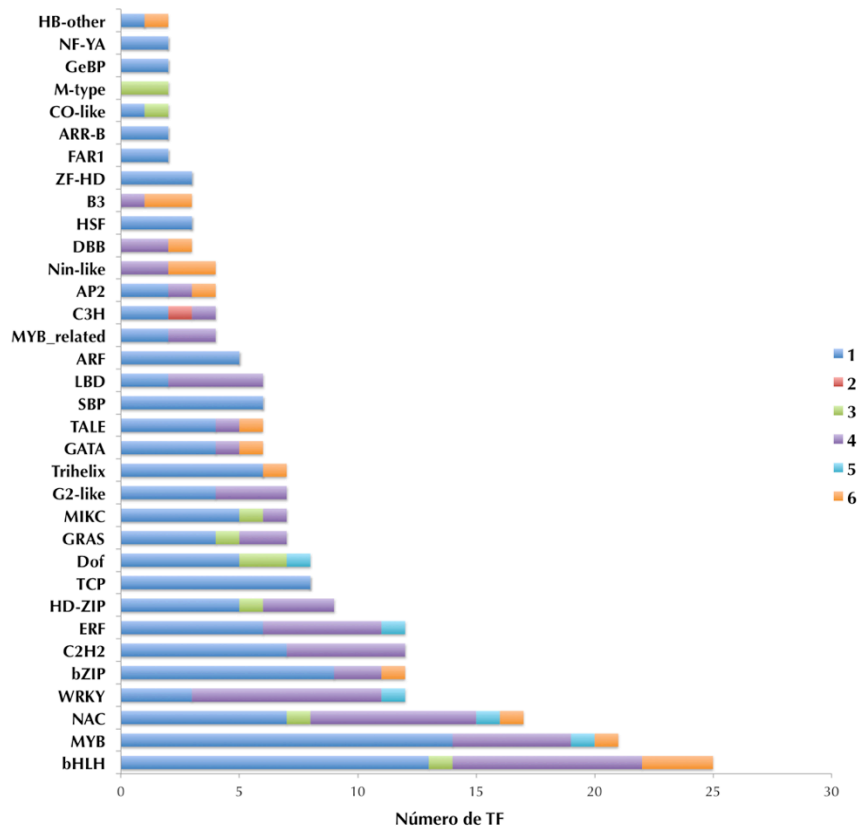


Figura 23. Familias y patrones de expresión de factores de transcripción diferencialmente expresados, con más de un representante por familia.

Por otro lado, se observó que alrededor del 93% de los TF diferenciales pueden ser agrupados en los patrones de expresión 1 (57.65%), 4 (27.45%) y 6 (7.84%). Estos resultados son congruentes con los patrones de expresión observados para todos los genes, genes codificantes y lncRNAs, en los que la mayor proporción de genes es agrupada en estos tres patrones (Figura 19). Sin embargo, como se describió anteriormente, la mayoría de los genes meióticos se expresan siguiendo el patrón de expresión 6, por lo cual revisamos con mayor detalle TF que presentan este mismo patrón de expresión. De los 20 TF que se ajustan a este patrón de expresión, 9 tienen un patrón de expresión preferencial en meiocitos (dos veces más expresión en comparación con el transcriptoma somático). Estos 9 TFs, pertenecen a 7 familias distintas (Tabla S18), de las cuales se resalta una vez más la familia bHLH, para la cual se observaron 3 TF que cumplían las características anteriormente mencionadas.

Interesantemente, dos de estas bHLH (AtbHLH091 y AtbHLH010), forman parte del mismo grupo, según sus similitudes estructurales (Heim et al., 2003). Además, ambos TF fueron encontrados consistentemente sobreexpresados en los meiocitos de Arabidopsis y maíz

(Dukowic-Schulze et al., 2013). Al analizar anteras normales y mutantes de *Arabidopsis* se detectó una reducción significativa de la expresión de AtbHLH091 en las anteras mutantes (Wijeratne et al., 2007). La función de este gen también ha sido relacionada con la del gen DYT1, una proteína bHLH cuyo mutante presenta esterilidad masculina y deficiencias en la diferenciación y función del tapetum (Feng et al., 2012; Zhang et al., 2006). Lo anterior es importante puesto que un tapetum funcional es necesario para el desarrollo normal del polen y la conclusión de la meiosis (Zhang et al., 2006). Finalmente, con base en lo anterior, Wijeratne et al. (Wijeratne et al., 2007), propusieron que AtbHLH091 (y posiblemente AtbHLH010, dada su similitud estructural), podría actuar como un regulador de la expresión de genes meióticos. Lo anterior resalta la importancia de nuestros resultados, los cuales sugieren nuevamente que estos dos genes podrían tener relación con la regulación de la expresión de genes meióticos en girasol.

La tercera proteína bHLH, que se ajusta al patrón 6 de expresión, previamente fue encontrada expresada en el gametofito femenino de *Arabidopsis* (Wang et al., 2010), sugiriendo que esta también podría estar relacionado con meiosis. Finalmente, otros dos de los FT de transcripción diferenciales, preferenciales de meiocitos y con el patrón de expresión 6 cuya función previa descrita podría sugerir algún papel en meiosis son BME3 y HAT3.1. El primero se encontró sobreexpresado en *Arabidopsis* en respuesta al daño inducido en el DNA con un agente químico (por lo que podría ser importante para la reparación del DNA) (Liu et al., 2015a); mientras que para el segundo se detectó expresión en la MMC (Megaspore Mother Cell), sugiriendo un papel en el desarrollo de la línea germinal (Schmidt et al., 2011).

De la lista de TF sobreexpresados en meiocitos de *Arabidopsis* y maíz (Dukowic-Schulze et al., 2013) que además presentan expresión preferencial en meiocitos de girasol, fueron encontrados diferencialmente expresados 6 en la comparación D-S (Tabla S19). Dos de estos, como se menciona anteriormente, corresponden a las proteínas AtbHLH091 y AtbHLH010. Otros dos, pertenecen a la familia MIKC (un tipo MADS), los cuales presentan patrones de expresión opuestos (1 y 6). Por último, los dos restantes, tienen el patrón de expresión 1 y pertenecen a las familias bZIP y ERF. No existen reportes adicionales que asocien a estos elementos regulatorios con meiosis. Sin embargo, AG (MIKC) y TGA10 (bZIP) han sido asociados con el desarrollo del estambre y tapetum, respectivamente (Ito et al., 2007; Murmu et al., 2010), sugiriendo un papel importante para el apropiado desarrollo floral. Por tanto, estos TF son candidatos regulatorios interesantes, dada su consistente expresión preferencial en meiocitos y los cambios de expresión observados en el presente estudio. Son necesarios estudios posteriores para esclarecer la relación de estos TF con meiosis.

Otros elementos que pueden regular la transcripción de genes son los elementos en *cis*; pequeños segmentos del DNA que regulan la transcripción de genes cercanos al funcionar, generalmente, como sitios de unión a TF. Nosotros realizamos una búsqueda de estos

elementos en los genes meióticos diferencialmente expresados (en el contraste D-S), utilizando la información disponible de elementos *cis* en la base de datos ATTED-II versión 8.0 (Obayashi et al., 2007, 2011). Para 11 de los 34 genes meióticos diferenciales se encontraron elementos *cis* predichos (Tabla S20). La mayoría de estos genes, cumplen funciones de control del ciclo celular, organización del huso y citocinesis (“Cell cycle control: spindle and cytokinesis”). Este resultado llama la atención, puesto que la mayor cantidad de genes diferenciales se observó en esta categoría. Además, como se describió anteriormente, los reguladores transcripcionales MMD/DUET y FST parecen tener un efecto sobre la correcta organización del huso (Andreuzza et al., 2015; Li et al., 2015). No obstante, en la reconstrucción de las probabilidades de co-expresión de pares de genes meióticos en Arabidopsis (Figura 22A), se observó que los genes que pertenecen a esta categoría presentan probabilidades de co-expresión altas con genes pertenecientes a casi todas las demás categorías funcionales meióticas. En conjunto esta evidencia podría sugerir que estos genes que participan en el control del ciclo celular y organización de huso, pues presentan una marcada regulación transcripcional.

Posteriormente colectamos la información de los TF cuya expresión estaba correlacionada con estos elementos en *cis*, según la información disponible en la base de datos utilizada. De estos, 13 TF distintos son preferencialmente expresados en meiocitos y presentan expresión diferencial en los meiocitos de los genotipos Domesticado y Silvestre. Además, al contrastar la información del gen meiótico cuyo elemento en *cis* regulatorio correlaciona con la expresión de estos TF (Tabla 4), observamos diferentes relaciones entre los genes meióticos, los elementos en *cis* y los TF. En un primer caso, un mismo TF puede estar relacionado con elementos en *cis* diferentes, presentes en genes meióticos distintos (e.g. AT1G08320, tiene relación con elementos en *cis* presentes en ATK1 y ATK5, entre otros). Además, un mismo elemento en *cis* puede estar presente en diferentes genes meióticos y estar relacionado con el mismo TF (e.g. GGGTTTT, presente en SMC4 y CDKA;1, cuya expresión correlaciona con AT3G52910). Así como un mismo elemento en *cis* puede tener relación con TFs distintos (e.g. GCCCATA, relacionado con AT1G08320 y AT3G52910). En otras palabras, estos resultados sugieren que un mismo TF podría regular la expresión de distintos genes meióticos *via* elementos en *cis* regulatorios iguales o diferentes, y que un mismo elemento en *cis* podría regular la expresión de genes meióticos *via* TF diferentes. Este resultado, refleja la complejidad de la red de regulación que puede operar en meiosis, y soporta la posibilidad de una expresión coordinada por etapas, al sugerir que un mismo gen puede tener efectos simultáneos (o coordinados) en distintos genes meióticos, tal como ha sido descrito para los reguladores transcripcionales MMD/DUET y FST (Andreuzza et al., 2015; Li et al., 2015).

4.3.6 Análisis de co-expresión de genes meióticos y factores de transcripción

Con el propósito de entender mejor las diferencias transcripcionales observadas en girasol decidimos observar, a partir de reportes de co-expresión en Arabidopsis, que genes se encontraban co-expresados con genes meióticos y factores de transcripción diferencialmente expresados en el contraste D-S. Como era de esperarse, entre los genes que co-expresan con

genes meióticos diferenciales se encuentran algunos genes meióticos. La categoría funcional “Cell cycle control: spindle and cytokinesis” fue la más representada cuando se buscaron genes meióticos diferenciales que co-expresaban con genes meióticos que pertenecían a la misma categoría funcional (Table S21). Por otro lado, al buscar genes meióticos diferenciales y meióticos co-expresados relacionados con categorías funcionales de recombinación (Tabla S22), se observó que la mayoría de estos genes pertenecían a las categorías “Recombination: Anti-CO Activity” y “Recombination: Class I CO”. Incluso, en algunos casos se observó co-expresión de genes pertenecientes a estas dos categorías, aunque parecen opuestas. Lo anterior sugiere la existencia no sólo de coordinación temporal de la expresión (i.e. genes que participan en un mismo proceso, se co-expresan), sino también de genes de rutas de generación CO y NCO, lo cual parece indicar que a nivel transcripcional también hay mecanismos que promueven el fenómeno de “homeostasis de CO”, el cual fue previamente descrito como un mecanismo de control de la recombinación (Phadnis et al., 2011).

Tabla 4. Elementos en *cis* predichos, presentes en genes meióticos diferencialmente expresados y TF diferenciales con los que ha sido correlacionada su expresión.

Locus Meiótico	Gen	Categoría funcional meiótica	Locus TF	Familia TF	Elemento <i>cis</i>
AT3G54670	SMC1	SMC complex	AT1G01060	MYB_related	AAATATC
AT3G54670	SMC1	SMC complex	AT1G08320	bZIP	GCCCATC
AT4G05190	ATK5	Cell cycle control: spindle and cytokinesis	AT1G08320	bZIP	TGGGCCT
AT4G21270	ATK1	Cell cycle control: spindle and cytokinesis	AT1G08320	bZIP	GCCCATC
AT5G57450	XRCC3	Recombination: early DSB repair	AT1G08320	bZIP	GCCCAAT
AT5G62410	SMC4	SMC complex	AT1G08320	bZIP	TAGGGTT
AT5G65460	CDKA;1	Cell cycle control: spindle and cytokinesis	AT1G08320	bZIP	TAGGGTT
AT3G54670	SMC1	SMC complex	AT1G27370	SBP	AAAATAT
AT3G54670	SMC1	SMC complex	AT1G28050	CO-like	AAATATC
AT4G21270	ATK1	Cell cycle control: spindle and cytokinesis	AT2G26150	HSF	TCCAGAA
AT5G65460	CDKA;1	Cell cycle control: spindle and cytokinesis	AT3G01330	E2F/DP	TTAGGGT
AT3G54670	SMC1	SMC complex	AT3G46640	G2-like	AAATATC
AT3G54670	SMC1	SMC complex	AT3G52910	GRF	GCCCATC
AT4G05190	ATK5	Cell cycle control: spindle and cytokinesis	AT3G52910	GRF	TGGGCCT
AT4G18120	AML3	Entry into meiosis	AT3G52910	GRF	AGGGTTT
AT4G21270	ATK1	Cell cycle control: spindle and cytokinesis	AT3G52910	GRF	GCCCATC
AT5G57450	XRCC3	Recombination: early DSB repair	AT3G52910	GRF	GCCCAAT
AT5G62410	SMC4	SMC complex	AT3G52910	GRF	GGGTTTT
AT5G65460	CDKA;1	Cell cycle control: spindle and cytokinesis	AT3G52910	GRF	GGGTTTT
AT4G21270	ATK1	Cell cycle control: spindle and cytokinesis	AT4G17230	GRAS	TTGGGCC
AT5G57450	XRCC3	Recombination: early DSB repair	AT4G17230	GRAS	AGCCCAA
AT3G43210	TES	Cell cycle control: spindle and cytokinesis	AT4G21440	MYB	ACACGTG
AT4G21270	ATK1	Cell cycle control: spindle and cytokinesis	AT4G21440	MYB	CACGTGG
AT4G18120	AML3	Entry into meiosis	AT4G37750	AP2	AGGGTTT
AT5G62410	SMC4	SMC complex	AT4G37750	AP2	GGGTTTT
AT5G65460	CDKA;1	Cell cycle control: spindle and cytokinesis	AT4G37750	AP2	GGGTTTT
AT3G43210	TES	Cell cycle control: spindle and cytokinesis	AT5G12840	NF-YA	ACACGTG
AT4G21270	ATK1	Cell cycle control: spindle and cytokinesis	AT5G12840	NF-YA	CACGTGG
AT3G54670	SMC1	SMC complex	AT5G48250	CO-like	AAATATC

Finalmente, para tener una visión general de los genes que co-expresaban tanto con meiocitos como con TF, estos fueron agrupados en términos GO. Tanto para los términos de GO de BP (Tabla S24), como para los de MF (Tabla S23), se observaron varias similitudes entre los genes co-expresados con genes meióticos (coMei) y los co-expresados con TF (coTF). Dentro de los términos GO de BP comunes para los dos tipos de genes co-

expresados se destacan los términos 'gene silencing', 'chromosome segregation', y 'microgametogenesis', los cuales fueron previamente encontrados sobrerrepresentados en los meiocitos de girasol (Flórez-Zapata et al., 2014). Asimismo, se resalta la presencia del término 'circadian rhythm', el cual fue más frecuentemente encontrado en genes co-expresados con TF, ya que la expresión de un gran número de genes relacionados con ciclos circadianos fue encontrada afectada en el mutante del gen regulador FST (Li et al., 2015). La correspondencia de términos GO de MF para genes coMei y coTF fue mucho mayor que la observada para genes de la categoría BP (Figura S24). Términos GO de MF como 'protein binding' y 'chromatin binding', que fueron encontrados representados en los genes co-expresados, también fueron previamente reportados como sobrerrepresentados en meiocitos de girasol (Flórez-Zapata et al., 2014). Estos resultados, sugieren que los cambios transcripcionales observados entre los meiocitos de girasol, en términos de genes meióticos y factores de transcripción, podrían a su vez inducir cambios (dado que están co-expresados) en genes con funciones importantes para el meiocito, previamente descritas.

4.4 Conclusiones

Aunque el análisis de las diferencias en los transcriptomas de meiocitos de girasol con genotipo de recombinación contrastante no permite explicar claramente el porque de estas diferencias, aporta información invaluable para la comprensión general de la regulación de la expresión meiótica en plantas. En primer lugar, este análisis deja en evidencia que hay un efecto particular de la heterocigosidad sobre la expresión de los genes meióticos. Asimismo, provee evidencia clara de que durante la meiosis en plantas existe una coordinación temporal de la expresión de genes, en la cual pueden intervenir TF y elementos en *cis*. Finalmente, genera gran cantidad de información que, analizada en detalle, podría ser útil para describir nuevos genes con función en meiosis o en su regulación.

CAPÍTULO 5: Del análisis transcripcional hacia la identificación de nuevos genes importantes en la meiosis de plantas.

5.1 Introducción

Entre otros factores, el número limitado de genes que hasta el momento han sido descritos con función meiótica en plantas, ha dificultado la comprensión de este importante proceso (Zhou and Pawlowski, 2014b). Una aproximación para llevar a cabo la identificación de nuevos genes importantes para la meiosis es la selección de genes candidatos basados en su perfil de expresión diferencial en órganos sexuales (Hamant et al., 2006a; Mercier and Grelon, 2008). Por ejemplo, el gen regulador FST en *Arabidopsis*, fue identificado gracias a que este se encontraba preferencialmente expresado en meiocitos, en los cuales además se encontró una alta actividad de su promotor (Li et al., 2015). Por tanto, nuestro transcriptoma podría ser muy útil para la búsqueda de nuevos genes candidatos, puesto que además de la expresión preferencial en meiocitos, tenemos el criterio de expresión diferencial entre los genotipos de girasol con un fenotipo de frecuencia de quiasmas contrastante.

Una de las formas para poder llevar a cabo la descripción de la función de un gen candidato es realizando la descripción de los efectos que trae consigo su mutación. Lo anterior no es una opción directa para girasol, puesto que aunque existe una colección de mutantes puntuales obtenida por el tratamiento con EMS, esta se encuentra muy poco caracterizada (Sabetta et al., 2011). Sin embargo, podríamos llevar a cabo la caracterización funcional de los ortólogos de girasol en *A. thaliana*, puesto que como ya se menciono anteriormente una estrategia que ha sido exitosa para la identificación de nuevos genes meióticos es la búsqueda de homólogos en organismos modelo (Mercier et al., 2001), dado que muchos de estos genes se encuentran ampliamente conservados.

Por tanto, el objetivo del presente capítulo es presentar el análisis funcional de algunos genes candidatos de girasol, a través de la descripción de su fenotipo mutante en *A. thaliana*. Para esta última existen gran cantidad de colecciones de mutantes, además de información genómica y de expresión, lo cual sumado a su tamaño reducido, ciclo corto de vida y descendencia numerosa, la convierten en un buen organismo modelo (Meinke et al., 1998). Los resultados presentados en este capítulo, son el producto de la estancia de investigación realizada en el laboratorio de Citogenética A, del departamento de Genética de la Universidad Complutense de Madrid, bajo la coordinación de los doctores Juan Luis Santos y Mónica Pradillo.

5.2 Materiales y Métodos

5.2.1 Selección de candidatos y adquisición de mutantes de inserción de T-DNA.

En total se seleccionaron 19 genes candidatos ortólogos de girasol para caracterizar en *A. thaliana*. Algunos de los criterios que se tuvieron en cuenta para su selección fueron: i). Expresión diferencial, con cambios superiores al doble entre genotipos de girasol. ii). Expresión preferencial en meiocitos. iii). Anotación con términos GO relacionados con regulación de la transcripción, metabolismo de DNA, interacciones con proteínas y/o DNA. iv) Evidencia de co-expresión con otros genes que cumplen función en meiosis. Una vez seleccionados los genes, se solicitaron 3 líneas diferentes de inserción distintas para cada gen al Nottingham Arabidopsis Stock Centre (NASC).

5.2.1 Crecimiento de las plantas

Las semillas fueron sembradas bandejas de crecimiento para 51 plantas, que contenían una mezcla estéril (3:1) de suelo abonado y vermiculita. Posteriormente, fueron colocadas en una cámara de crecimiento de condiciones controladas de luz (16 horas luz y 8 horas oscuridad), temperatura (18°C) y humedad (60% de humedad relativa). Las plantas fueron mantenidas bajo estas condiciones, a lo largo del desarrollo, sin realizar fertilización adicional, y realizando riegos periódicos según las necesidades.

5.2.2. Caracterización molecular de líneas mutantes.

Se verificó la presencia de inserciones de T-DNA en los genes de interés mediante PCR. Para ello se amplificó el DNA extraído de hoja, con primers específicos (Tabla S25) diseñados para cada línea con el programa "T-DNA primer desing tool" del Instituto Salk (<http://signal.salk.edu/tdnaprimers.2.html>). Las reacciones de PCR fueron realizadas utilizando la mezcla de reacción BioMix™ Red (Bioline , London, UK), y 0.2 µL de cebador a una concentración de 100 µmol, y 0.6 µL de DNA extraído, para un volumen final de reacción de 10 µL. El programa de amplificación utilizado, consistió en un paso inicial de denaturación a 94°C por 10 minutos, seguida de 35 ciclos de: 1 min 94°C, 1 min 55°C y 2 min a 72°C, seguidos de una extensión final de 10 min a 72°C. Los productos de PCR fueron visualizados en geles de agarosa al 1% (p/v), teñidos con SYBR Safe (Molecular Probes, Eugene, OR, USA).

5.2.3. Análisis citológico.

Se realizaron dos metodologías distintas, para el análisis citológico: la técnica de aplastado "squash" y la técnica de esparcido "spreading".

5.2.3.1 Técnica de aplastado "squash".

Para esta técnica, se colectaron botones de gran tamaño (con polen maduro), se retiraron los sépalos y se colocó el botón sobre un portaobjetos, al cual se le añadió una gota de colorante carmín-acético. La preparación fue calentada por 5 segundos en la llama de un mechero. Posteriormente se colocó un cubreobjetos, e interponiendo un papel de filtro, se

realizó el aplastado del botón con el dedo pulgar. En seguida, esta preparación fue observada en el microscopio de contraste de fases, donde se verificó la homogeneidad en forma y tamaño de los gránulos de pólen.

5.2.3.1 Técnica de esparcido “spreading”.

Para poder realizar esta técnica los botones florales fueron fijados en solución Carnoy (etanol absoluto, cloroformo y ácido acético en proporciones 6:3:1). Transcurridas 24 horas a temperatura ambiente, se retiró y agregó nuevamente fijador, y los botones florales se almacenaron a -20°C. Posteriormente, los botones fijados fueron lavados tres veces con una solución de etanol absoluto: ácido acético (3:1), y tres veces con tampón citrato (10mM; pH4.5). Los botones limpios, fueron sometidos a digestión enzimática (celulasa, pectinasa y citohelicasa, cada una al 3% peso/volumen), por 2 horas a 37°C. La reacción enzimática fue detenida agregando tampón citrato frío. Cada botón fue macerado sobre un portaobjetos, al que luego se agregaron 10 µL de ácido acético frío al 60%. Esta preparación fue calentada por 1 minuto durante 45°C, luego de los cuales se agregaron otros 10 µL de ácido acético frío y se realizaron lavados con la solución etanol-acético (3:1). Las preparaciones se dejaron secar y fueron teñidas con 8 µL de DAPI (10 µL/mL) disuelto en Vectashield (Vector Laboratories, Burlingame, CA). Las preparaciones teñidas fueron observadas en el microscopio de fluorescencia, donde se analizaron cada una de las fases de la meiosis.

5.3 Resultados y Discusión

5.3.1 Crecimiento y genotipado de líneas mutantes

De los 19 genes candidatos, a excepción de REV7 y AHL5 (Tabla 5), se solicitaron 3 líneas insercionales al NASC. Posteriormente, las plantas fueron crecidas bajo condiciones controladas, para su posterior genotipado. Para el gen T7B11.5 no fue posible obtener plantas de ninguna de las 3 líneas, puesto que las semillas enviadas no germinaron después de dos intentos de siembra, por lo cual decidimos retirarlo del estudio. El genotipado fue realizado por PCR, utilizando 3 primers en la reacción (Figura 24B), de los cuales, LP y RP, son específicos del gen que se quiere analizar, mientras que el primer Lb es específico del T-DNA que se tiene en la inserción y es diferente según la naturaleza de cada una de las líneas (i.e Salk, Sail, Gabi o Wisc).

Tabla 5. Lista de genes candidatos y las líneas mutantes correspondientes, solicitadas al NASC.

Locus	Gen	Línea 1	Línea 2	Línea 3
AT1G16590	REV7	SALK_014571 (Intrón). N682024.	SAIL_556_A02 (3'UTR). N823521.	-
AT1G75430	BLH11	SALK_067443 (Intrón). N672601	SALK_067982 (Intrón). N567982	GABI_183F05 (Exón). N320041
AT2G32350		SALK_144037 (Exón). N644037	SALK_032843 (3'UTR). N672534	SALK_104448 (promotor). N681556
AT1G37140	MCT1	SALK_014311 (3'UTR). N514311	WiscDsLoxHs032_08A (Exón). N903033	SALK_075516 (5'UTR). N654722
AT1G10610	BHLH90	SALK_000633 (Exón). N500633	SALK_022615 (Intrón). N674076	SALK_052405 (Intrón). N685937
AT1G72570	AIL1	SALK_003704 (Exón). N685062	SAIL_69_E10 (Intrón). N860230	SAIL_109_A03 (Intrón). N860362
AT4G12240	T4C9.80	SALK_208312 (promotor). N698486	SAIL_41_H09 (promotor). N862425	SAIL_31_H09 (promotor). N801537
AT1G17040	SHA	GABI_367A12 (Exón). N321445	GABI_511A08 (Exón). N322116	GABI_379C07 (Exón). N321309
AT5G01310	APTX	SALK_024760 (Exón). N657900	SALK_081627 (Exón). N681374	SALK_125225 (Exón). N625225
AT4G01790	T7B11.5	GABI_440B08 (Exón). N305357	SALK_070883 (5'UTR). N672674	SALK_095758 (5'UTR). N680022
AT3G50690		SALK_033316 (Exón). N533316	SALK_133139 (3'UTR). N664040	Hay una GK 5'UTR
AT1G13790	FDM4	SALK_205923 (Exón). N695284	SALK_206362 (Exón). N695930	WiscDsLox426D07 (Intrón). N855334
AT2G37880		SALK_204855 (3'UTR). N693772	GABI_931B01 (Exón). N769468	Hay SALK de promotor y FLAG 3'UTR
AT1G05950		SAIL_69_E08 (Intrón). N860229	SAIL_409_D03 (Exón). N861854	SAIL_360_A05 (Exón). N816727
AT2G47680		SALK_127155 (Intrón). N669007	GABI_015F05 (Exón). N798004	SAIL_877_H01 (Intrón). N877618
AT4G08310		GABI_413F08 (Exón). N787590	SALK_146231 (5'UTR). N658456	WiscDsLox367H4_041 (3'UTR). N853113
AT4G11130	RDR2	SALK_207201 (Exón). N697132	SALK_206644 (Exón). N696374	SALK_203854 (Exón). N692343
AT1G63470	AHL5	SALK_123590 (Exón). N668976	GABI_106G12 (Exón). N739952	-
AT3G04620	F7O18.10	SAIL_676_G06 (Exón). N829632	SAIL_649_E11 (Intrón). N828069	SAIL_548_H07 (Intrón). N823256

Así, para cada una de las plantas crecidas por línea (alrededor de 17 en la mayoría de los casos en los que la germinación fue del 100%), se realizó una PCR con los tres primers juntos en la reacción (LP y RP específicos para la región del gen cercana a la inserción y el Lb acorde al tipo de T-DNA) (Figura 24A). Se observó el patrón de bandas, y se realizó una confirmación de 1 a 3 muestras por cada patrón, con los primers por separado (LP-RP para la primera reacción, y Lb-RP para la segunda). En caso de la presencia de T-DNA insertado en la región de interés no se observó producto en la reacción LP-RP (la presencia del T-DNA genera que entre los dos primers haya una distancia superior a la que puede ser amplificada en la PCR), pero sí para la reacción Lb-RP. En caso de plantas heterocigotas, se observó producto de amplificación tanto en la reacción LP-RP, como en la Lb-RP. En el ejemplo presentado en la Figura 24A se evidencia cómo las plantas 1, 3, 4, 8 y 9 tienen el mismo patrón de bandas que una planta silvestre (*wild-type* -*Wt*-), mientras que las plantas 2, 5, 6, 11 y 15 presentan una única banda (de tamaño diferente a la del *Wt*), lo que sugiere que se trata de plantas homocigotas para el alelo mutante. Por otro lado, aquellas que poseen dos bandas podrían ser plantas heretocigotas para la inserción, que tendrían un alelo *Wt* y uno mutante. Al realizar la amplificación con primers por separado para las plantas 2, 5, 6, 7 y 3 (Figura 24B), se confirmó que el patrón de las plantas 2, 5 y 6 corresponde con el de plantas homocigotas mutantes, el de la planta 7 con el de una planta heterocigota para la mutación

y el de la planta 3 con el de una planta Wt. Como resultado del genotipado, fue necesario excluir los genes MCT1, APTX, AT1G05950 y AHL5, para los cuales no se encontraron plantas homocigotas y/o no pudo llevarse a cabo la amplificación.

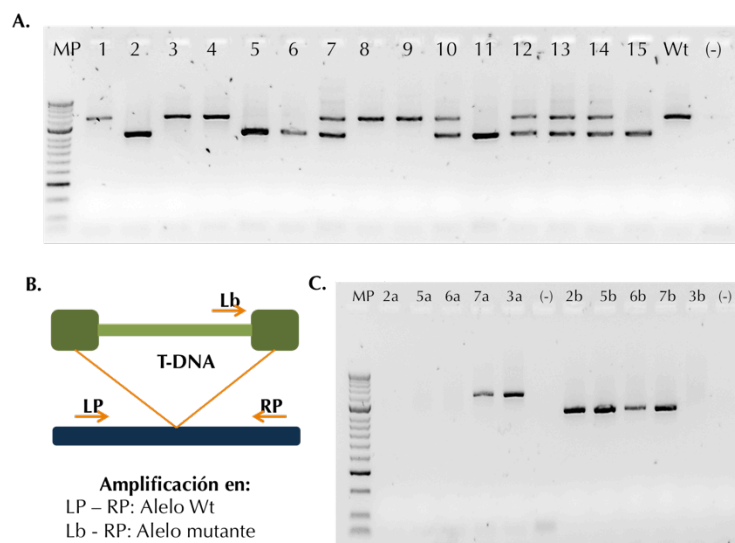


Figura 24. Genotipado de líneas de inserción de T-DNA. (A) Resultado de la reacción de amplificación con los 3 primers juntos (LP, RP y Lb). (B) Representación de la localización e interpretación de amplificación de los primers. (C) Resultado de amplificación de la reacción con primers por separados (se representa la misma muestra amplificada con los primers LP-RP (a), y Lb-RP (b)).

5.3.2 Caracterización citológica de mutantes por la técnica de "squash".

La obtención de plantas homocigotas es necesaria para llevar a cabo las etapas posteriores de caracterización fenotípica de los mutantes, puesto que en plantas heterocigotas la función del alelo mutante puede ser suplida por el alelo silvestre. Por ejemplo, al caracterizar plantas heterocigotas mutantes para el gen MND1, en ninguna de estas se observó efecto fenotípico, lo cual si se logró en plantas homocigotas (Panoli et al., 2006). Así, para asegurarnos de observar un efecto fenotípico en el caso de que los genes candidatos estuvieran relacionados con meiosis en *Arabidopsis* colectamos botones florales de 3 plantas para cada línea y realizamos la caracterización citológica por la técnica de squash. Esta es fácil de realizar y permite evidenciar defectos en los gránulos de polen, los cuales pueden sugerir la presencia de defectos en la meiosis que llevan a la segregación desbalanceada. La Figura 25, presenta un ejemplo de cómo se observan por esta técnica el polen normal de una planta silvestre (Col-0), y el polen de una planta mutante del gen *Zip4*, fundamental para la meiosis en *A. thaliana*, en el que su mutación genera, entre otros efectos, células hijas con un número cromosómico aberrante (Chelysheva et al., 2007).

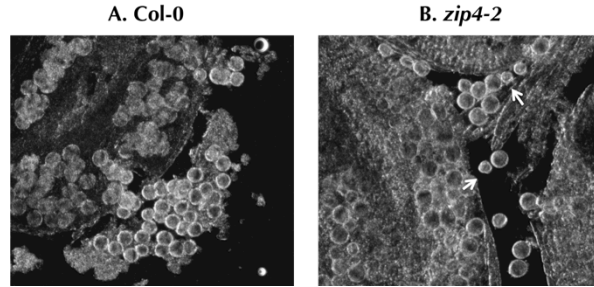


Figura 25. Ejemplo de un resultado de la caracterización por “squash”. (A) Gránulos de polen normales de una planta Wt (Col-0). (B) Gránulos de polen de un mutante para Zip4, una proteína esencial para la meiosis. Las flechas blancas, señalan gránulos de tamaño anormalmente reducido.

Ninguna de las líneas mutantes evaluadas mostró defectos en el tamaño o forma del polen, lo cual podría sugerir que la mutación de los genes candidatos no lleva consigo un efecto fenotípico en meiosis, y por tanto que estos no son importantes para la misma. Sin embargo, para algunos genes importantes para la meiosis, e.g. MUS81, su mutación no conlleva a efectos fenotípicos en meiosis, hasta que no se realizan dobles mutantes con otros genes meióticos (MSH4 en el caso de MUS81) (Higgins et al., 2008). Una segunda posibilidad es que el efecto fenotípico de los genes no sea tan drástico, de forma que los defectos en alguna etapa de la meiosis puedan ser remontados y la segregación sea balanceada, llevando a que no se observen diferencias en los gránulos de polen.

5.3.2 Caracterización citológica de mutantes por la técnica de “spreading”.

Dado que el primer análisis de las líneas mutantes por la técnica de squash no arrojó resultados que sugieren una alteración en meiosis, decidimos realizar una descripción más detallada a través de la técnica de “spreading”. Esta técnica permite hacer un seguimiento detallado de todas las fases de la meiosis, en preparaciones teñidas con DAPI. Sin embargo, es laboriosa y la probabilidad de encontrar todas las fases es limitada (considerando que la duración de cada fase es distinta), lo cual requiere la realización de al menos 20 preparaciones de botones florales por línea y su posterior rastreo en el microscopio de fluorescencia. Debido a lo anterior, se decidió analizar por esta técnica, una sola línea (seleccionada al azar), por gen candidato. En la Figura 26, se presentan fotografías de 3 etapas de la meiosis (Profase I, Metafase I y Tétrada), para 5 de los genes caracterizados, las demás, son presentadas en el material suplementario (Figura S11).

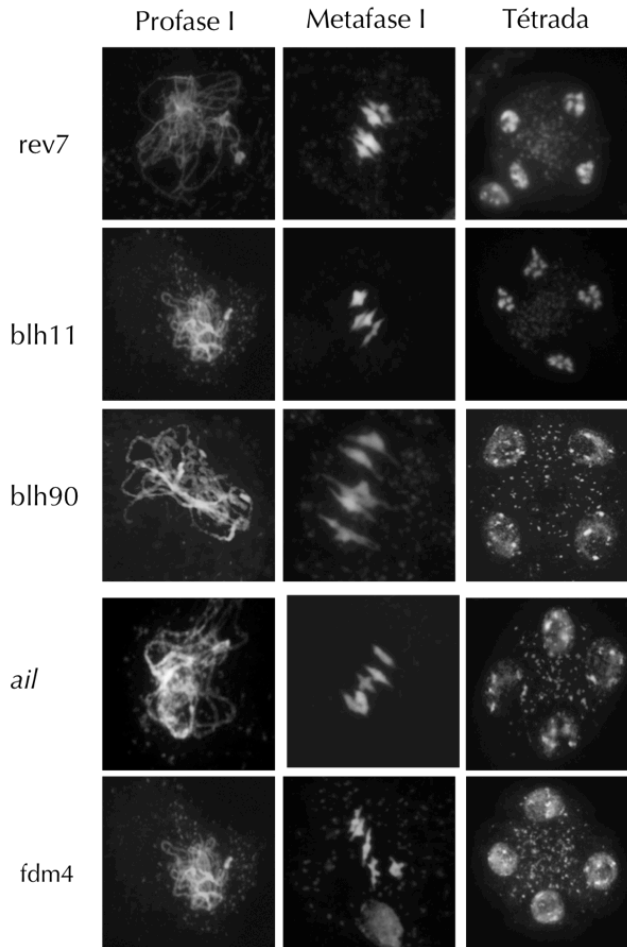


Figura 26. Resultados de la técnica de esparcido “spreading”, para líneas mutantes de algunos de los genes candidatos estudiados.

Al monitorear la Profase I se pueden observar problemas en la sinapsis, fragmentación o condensación de la cromatina, mientras que en Metafase I es posible estimar el número de quiasmas (normalmente en *Arabidopsis* se observan 5 bivalentes “cerrados”, con dos quiasmas, para dar una media de 10 quiasmas por célula); además, en esta etapa se pueden distinguir fragmentos y asociaciones multivalentes (asociación entre cromosomas no homólogos). Finalmente, en la etapa de Tétrada, se verifica que el resultado de la meiosis haya sido equilibrado, con la formación de 4 células hijas iguales. Al analizar las líneas mutantes de los genes candidatos en ninguna de las líneas analizadas se observaron anomalías meióticas. Lo anterior era esperado, dado que tampoco se observaron diferencias en el polen (técnica de squash) o en el tamaño de las siliquas.

La meiosis, es un proceso fundamental para los organismos que se reproducen sexualmente, puesto que gracias a esta se lleva a cabo la generación de gametos que darán origen a su descendencia. Por tanto, es de esperarse que este proceso se encuentre altamente controlado y sujeto a procesos evolutivos que le permiten adaptarse gradualmente a las perturbaciones

(Bomblies et al., 2015). Por tanto, es posible que existan mecanismos redundantes, que compensen las alteraciones que puedan ocurrir (e.g. la mutación de un gen), y garanticen la correcta formación de los gametos. No obstante, los mutantes simples de algunos de los genes importantes para la meiosis no presentan un fenotipo meiótico alterado, tales como MUS81 (Higgins et al., 2008), FANCM (Crismani et al., 2012) y SUN1/SUN2 (Varas et al., 2015). Así pues, no se puede descartar totalmente que los genes candidatos estudiados desempeñen una función en meiosis.

Una alternativa a la ausencia de fenotipo de los genes candidatos en meiosis es que estos pertenezcan a una familia de parálogos que desempeñen funciones redundantes. Este fenómeno ha sido previamente descrito para algunos genes que participan en meiosis como RAD51 (Wang et al., 2014), miembros del complejo APC/C (Stein et al., 2010) y proteínas de la envoltura nuclear como SUN1 y SUN2 (Varas et al., 2015). Por tanto, decidimos buscar si alguno de estos genes candidatos tenían parálogos, con funciones relacionadas con meiosis, que pudieran tener redundancia funcional y amortiguar el efecto fenotípico de la mutación. Para esto, utilizamos la herramienta pATsi (Ambrosino et al., 2015), la cual permitió identificar que para todos, a excepción de REV7, hay parálogos reportados (Tabla 6). Además, algunos de los parálogos parecen cumplir funciones similares a las reportadas para el gen candidato, y estar relacionados con el desarrollo del óvulo o del polen, o ser controlados por factores de transcripción altamente expresados en meiocitos (e.g. AGAMOUS). Por otro lado también es posible, que como se describió anteriormente para los reguladores transcripcionales de genes meióticos en levadura (Winter, 2012), los ortólogos de girasol en Arabidopsis tengan homología de secuencia, pero diversidad de función, por lo cual no se observó fenotipo en meiosis de las líneas mutantes.

Tabla 6. Número y función relacionada de genes parálogos de los candidatos estudiados.

Locus	Gen	Número de Parálogos	Función de algunos parálogos
AT1G16590	REV7	Singleton	
AT1G75430	BLH11	16	Organización del meristema, activados por AGAMOUS
AT2G32350		Unassigned	
AT1G10610	BHLH90	16	ABORTED MICROSPORES (AMS)
AT1G72570	AIL1	47	Desarrollo del óvulo
AT4G12240	T4C9.80	1	
AT1G17040	SHA	1	
AT3G50690		Unassigned	
AT1G13790	FDM4	9	
AT2G37880		11	
AT2G47680		21	DNA/RNA helicasas
AT4G08310		1	
AT4G11130	RDR2	4	Otras proteínas RDR
AT3G04620	F7O18.10	2	Proteínas de unión a DNA

5.4 Conclusiones

El análisis transcriptómico de meiocitos de girasol con diferencias en la tasa de recombinación es una fuente de información valiosa para la descripción de nuevos genes importantes para la recombinación y meiosis. Sin embargo, existe gran dificultad de llevar a cabo la caracterización funcional de estos en girasol. Por tanto, se realizó una caracterización de un subconjunto de estos candidatos interesantes en Arabidopsis, cuyos resultados preliminares sugieren que estos no están relacionados con meiosis. Sin embargo, se necesitan estudios posteriores (e.g. generación de dobles mutantes), para esclarecer si es completamente cierto que estos no tienen una asociación importante con meiosis. Además, el avance en metodologías de caracterización funcional (por ejemplo, mutagénesis dirigida) en plantas no modelo como girasol, permitiría realizar la caracterización directamente en esta planta, describir la función de candidatos y contrastarla con las observaciones realizadas en Arabidopsis.

CONCLUSIONES FINALES Y PERSPECTIVAS

El desarrollo del presente trabajo ha permitido realizar contribuciones importantes al entendimiento de la regulación transcripcional en plantas. La primera descripción de rasgos característicos de la transcripción en meiocitos, reveló que uno de estos es la expresión de un número alto de genes, muchos de ellos no anotados o de función desconocida. Lo anterior fue posteriormente destacado en el análisis de expresión diferencial entre los genotipos domesticado y silvestre, donde luego de análisis adicionales se esclareció que dichos genes eran en su mayoría lncRNAs. Los lncRNAs específicos de meiosis que se identificaron parecen tener mayor relación con sRNAs de 24-nt y elementos transponibles, por lo que proponemos que una de sus funciones principales es participar en el remodelamiento de la cromatina. Sin embargo, dada la versatilidad funcional de este tipo de RNAs no codificantes, es posible que estos desempeñen funciones adicionales que pueden ser investigadas a futuro.

Por otro lado, se logró la identificación de factores de transcripción con expresión preferencial conservada en meiocitos de *Arabidopsis*, maíz y girasol. Algunos de estos, además mostraron expresión diferencial entre los genotipos de girasol domesticado y silvestre, reforzando su papel en meiosis. Lo anterior, abre la posibilidad a investigaciones futuras, donde se establezca el papel de estos elementos regulatorios en la meiosis. Asimismo, se identificaron elementos *cis* regulatorios en genes meióticos de *Arabidopsis*, que correlacionan con la expresión de factores de transcripción encontrados con expresión preferencial en meiocitos. De manera interesante, se destacó la presencia de elementos *cis* comunes en distintos genes meióticos, sugiriendo que un mismo factor de transcripción podría controlar la expresión de varios genes meióticos. Esta observación, puede indicar que de cierta forma la expresión en plantas se encuentra coordinada, como se ha descrito en otros organismos modelo, como levadura.

La descripción de los patrones de expresión de los genes con función conocida en meiosis, reveló que la heterocigosidad parece tener un efecto sobre su expresión, lo cual es interesante, dado que previamente se describió que esta condición favorece la recombinación. Además, este agrupamiento en patrones de expresión, en conjunto con el conocimiento previo en *Arabidopsis*, permitió proponer relaciones de “co-expresión” de genes meióticos en girasol las cuales parecen ser similares a los de esta planta modelo, y aportan evidencia a la propuesta de que en plantas existe una regulación transcripcional coordinada por etapas. Por otro lado, aunque el análisis de expresión diferencial no arrojó diferencias que puedan ser claramente asociadas al fenotipo de recombinación observado en cada genotipo, nos permitió plantearnos algunas hipótesis, tales como la participación de una ruta adicional de formación de entrecruzamientos o el efecto de la interferencia, las cuales pueden ser abordadas en estudios posteriores.

Finalmente, la caracterización funcional de genes candidatos interesantes diferencialmente expresados en los meiocitos de girasol, no arrojó resultados concluyentes de su función en meiosis. Sin embargo, aún se pueden realizar otro tipo de análisis, e.g. generación de dobles mutantes o realización de pruebas de reparación de DNA, para poder discernir si cumplen o no funciones meióticas. Además, nuevos genes candidatos pueden ser analizados y a medida que avancen técnicas de caracterización funcional en plantas no modelo (e.g. mutagénesis dirigida con el sistema CRISPR-Cas9), es posible realizar la caracterización funcional de estos directamente en girasol.

REFERENCIAS

- Alinsug, M. V., Yu, C.-W., and Wu, K. (2009). Phylogenetic analysis, subcellular localization, and expression patterns of RPD3/HDA1 family histone deacetylases in plants. *BMC Plant Biol.* 9, 37. doi:10.1186/1471-2229-9-37.
- Altschul, S. F., Madden, T. L., Schäffer, a a, Zhang, J., Zhang, Z., Miller, W., et al. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25, 3389–402. Available at: doi: 10.1093/nar/25.17.3389.
- Amaral, P. P., Dinger, M. E., Mercer, T. R., and Mattick, J. S. (2008). The eukaryotic genome as an RNA machine. *Science* 319, 1787–9. doi:10.1126/science.1155472.
- Ambrosino, L., Bostan, H., di Salle, P., Sangiovanni, M., Vigilante, A., and Chiusano, M. L. (2015). pATsi: Paralogs and singleton genes from *Arabidopsis thaliana*. *Evol. Bioinforma.* 12, 1–7. doi:10.4137/EBO.S32536.
- Anderson, L. K., Hooker, K. D., and Stack, S. M. (2001). The distribution of early recombination nodules on zygotene bivalents from plants. *Genetics* 159, 1259–1269.
- Anderson, L. K., Lohmiller, L. D., Tang, X., Hammond, D. B., Javernick, L., Shearer, L., et al. (2014). Combined fluorescent and electron microscopic imaging unveils the specific properties of two classes of meiotic crossovers. *Proc. Natl. Acad. Sci. U. S. A.* 111, 13415–13420. doi:10.1073/pnas.1406846111.
- Andreuzza, S., Nishal, B., Singh, A., and Siddiqi, I. (2015). The Chromatin Protein DUET/MMD1 Controls Expression of the Meiotic Gene TDM1 during Male Meiosis in *Arabidopsis*. *PLoS Genet.* 11, e1005396. doi:10.1371/journal.pgen.1005396.
- Ariel, F., Romero-Barrios, N., Jégu, T., Benhamed, M., and Crespi, M. (2015). Battles and hijacks: noncoding transcription in plants. *Trends Plant Sci.* 20, 362–371. doi:10.1016/j.tplants.2015.03.003.
- Armstrong, S. J., Franklin, F. C. H., and Jones, G. H. (2003). A meiotic time-course for *Arabidopsis thaliana*. *Sex. Plant Reprod.* 16, 141–149. doi:10.1007/s00497-003-0186-4.
- Axtell, M. J. (2013). ShortStack: comprehensive annotation and quantification of small RNA genes. *RNA* 19, 740–51. doi:10.1261/rna.035279.112.
- Axtell, M. J. (2015). Non-coding RNAs: The small mysteries of males. *Nat. Plants* 1, 15055. doi:10.1038/nplants.2015.55.
- Azumi, Y., Liu, D., Zhao, D., Li, W., Wang, G., Hu, Y., et al. (2002). Homolog interaction during meiotic prophase I in *Arabidopsis* requires the SOLO DANCERS gene encoding a novel cyclin-like protein. *EMBO J.* 21, 3081–95. doi:10.1093/emboj/cdf285.
- Baker, S. C., Robinson-Beers, K., Villanueva, J. M., Gaiser, J. C., and Gasser, C. S. (1997). Interactions among genes regulating ovule development in *Arabidopsis thaliana*. *Genetics* 145, 1109–1124.
- Ballouz, S., Verleyen, W., and Gillis, J. (2015). Guidance for RNA-seq co-expression network construction and analysis: safety in numbers. *Bioinformatics* 31, 2123–2130. doi:10.1093/bioinformatics/btv118.
- Barth, S., Melchinger, a E., Devezi-Savula, B., and Lübberstedt, T. (2001). Influence of genetic background and heterozygosity on meiotic recombination in *Arabidopsis thaliana*. *Genome* 44, 971–978. doi:10.1139/gen-44-6-971.
- Bauer, E., Falque, M., Walter, H., Bauland, C., Camisan, C., Campo, L., et al. (2013). Intraspecific variation of recombination rate in maize. *Genome Biol.* 14, R103. doi:10.1186/gb-2013-14-9-r103.
- Berchowitz, L. E., and Copenhaver, G. P. (2010). Genetic interference: don't stand so close to me. *Curr. Genomics* 11, 91–102. doi:10.2174/138920210790886835.
- Berchowitz, L. E., Francis, K. E., Bey, A. L., and Copenhaver, G. P. (2007). The role of AtMUS81 in interference-insensitive crossovers in *A. thaliana*. *PLoS Genet.* 3, 1355–1364. doi:10.1371/journal.pgen.0030132.
- Berchowitz, L. E., Hanlon, S. E., Lieb, J. D., and Copenhaver, G. P. (2009). A positive but complex association between meiotic double-strand break hotspots and open chromatin in

- Saccharomyces cerevisiae*. *Genome Res.* 19, 2245–2257. doi:10.1101/gr.096297.109.
- Bishop, D. K., and Zickler, D. (2004). Early Decision. *Cell* 117, 9–15. doi:10.1016/S0092-8674(04)00297-1.
- Bleuyard, J.-Y., Gallego, M. E., Savigny, F., and White, C. I. (2005). Differing requirements for the *Arabidopsis* Rad51 paralogs in meiosis and DNA repair. *Plant J.* 41, 533–45. doi:10.1111/j.1365-3113X.2004.02318.x.
- Bleuyard, J.-Y., Gallego, M. E., and White, C. I. (2006). Recent advances in understanding of the DNA double-strand break repair machinery of plants. *DNA Repair (Amst)*. 5, 1–12. doi:10.1016/j.dnarep.2005.08.017.
- Bleuyard, J.-Y., and White, C. I. (2004). The *Arabidopsis* homologue of Xrcc3 plays an essential role in meiosis. *EMBO J.* 23, 439–49. doi:10.1038/sj.emboj.7600055.
- Boley, S. E., Wong, V. A., French, J. E., and Recio, L. (2002). p53 heterozygosity alters the mRNA expression of p53 target genes in the bone marrow in response to inhaled benzene. *Toxicol. Sci.* 66, 209–215. doi:10.1093/toxsci/66.2.209.
- Bombliès, K., Higgins, J. D., and Yant, L. (2015). Meiosis evolves: Adaptation to external and internal environments. *New Phytol.* 208, 306–323. doi:10.1111/nph.13499.
- Bonasio, R., and Shiekhattar, R. (2014). Regulation of transcription by long noncoding RNAs. *Annu. Rev. Genet.* 48, 433–55. doi:10.1146/annurev-genet-120213-092323.
- Borges, F., Pereira, P. a., Slotkin, R. K., Martienssen, R. a., and Becker, J. D. (2011). MicroRNA activity in the *Arabidopsis* male germline. *J. Exp. Bot.* 62, 1611–1620. doi:10.1093/jxb/erq452.
- Borsani, O., Zhu, J., Verslues, P. E., Sunkar, R., and Zhu, J.-K. (2005). Endogenous siRNAs Derived from a Pair of Natural cis-Antisense Transcripts Regulate Salt Tolerance in *Arabidopsis*. *Cell* 123, 1279–1291. doi:10.1016/j.cell.2005.11.035.
- Brar, G. a, Yassour, M., Friedman, N., Regev, A., Ingolia, N. T., and Weissman, J. S. (2012). High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335, 552–7. doi:10.1126/science.1215110.
- Carpenter, A. (1994). Chiasma Function. *Cell* 77, 959–962. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/7766240> [Accessed August 9, 2012].
- Cech, T. R., and Steitz, J. A. (2014). The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* 157, 77–94. doi:10.1016/j.cell.2014.03.008.
- Chapman, J. R., Taylor, M. R. G., and Boulton, S. J. (2012). Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell* 47, 497–510. doi:10.1016/j.molcel.2012.07.029.
- Chelysheva, L., Gendrot, G., Vezon, D., Doutriaux, M. P., Mercier, R., and Grelon, M. (2007). Zip4/Spo22 is required for class I CO formation but not for synapsis completion in *Arabidopsis thaliana*. *PLoS Genet.* 3, 802–813. doi:10.1371/journal.pgen.0030083.
- Chelysheva, L., Vezon, D., Belcram, K., Gendrot, G., and Grelon, M. (2008). The *Arabidopsis* BLAP75/Rmi1 homologue plays crucial roles in meiotic double-strand break repair. *PLoS Genet.* 4. doi:10.1371/journal.pgen.1000309.
- Chen, B., Harms, E., Chu, T., Henrion, G., and Strickland, S. (2000). Completion of meiosis in *Drosophila* oocytes requires transcriptional control by grauzone, a new zinc finger protein. *Development* 127, 1243–51. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10683177>.
- Chen, C., Farmer, A. D., Langley, R. J., Mudge, J., Crow, J. a, May, G. D., et al. (2010a). Meiosis-specific gene discovery in plants: RNA-Seq applied to isolated *Arabidopsis* male meiocytes. *BMC Plant Biol.* 10, 280. doi:10.1186/1471-2229-10-280.
- Chen, C., and Retzel, E. F. (2013). “Analyzing the Meiotic Transcriptome Using Isolated Meiocytes of *Arabidopsis thaliana*,” in *Plant Meiosis: Methods and Protocols (Methods Mol. Biol.)*, eds. W. Pawlowski, M. Grelon, and S. Armstrong (Humana Press), 203–213. doi:10.1007/978-1-62703-333-620.
- Chen, D., Meng, Y., Ma, X., Mao, C., Bai, Y., Cao, J., et al. (2010b). Small RNAs in angiosperms: Sequence characteristics, distribution and generation. *Bioinformatics* 26, 1391–1394. doi:10.1093/bioinformatics/btq150.
- Chen, D., Yuan, C., Zhang, J., Zhang, Z., Bai, L., Meng, Y., et al. (2012). PlantNATsDB: a comprehensive database of plant natural antisense transcripts. *Nucleic Acids Res.* 40, 1187–93.

- doi:10.1093/nar/gkr823.
- Chen, Z., Higgins, J. D., Hui, J. T. L., Li, J., Franklin, F. C. H., and Berger, F. (2011). Retinoblastoma protein is essential for early meiotic events in Arabidopsis. *EMBO J.* 30, 744–55. doi:10.1038/emboj.2010.344.
- Cnudde, F., Hedat, V., de Jong, H., Pierson, E. S., Rainey, D. Y., Zabeau, M., et al. (2006). Changes in gene expression during male meiosis in *Petunia hybrida*. *Chromosome Res.* 14, 919–32. doi:10.1007/s10577-006-1099-5.
- Crismani, W., Baumann, U., Sutton, T., Shirley, N., Webster, T., Spangenberg, G., et al. (2006). Microarray expression analysis of meiosis and microsporogenesis in hexaploid bread wheat. *BMC Genomics* 7, 267. doi:10.1186/1471-2164-7-267.
- Crismani, W., Girard, C., Froger, N., Pradillo, M., Santos, J. L., Chelysheva, L., et al. (2012). FANCM Limits Meiotic Crossovers. Sup. material. *Science (80-)*. 336, 1588–1590. doi:10.1126/science.1220381.
- Cromie, G. a, and Smith, G. R. (2007). Branching out: meiotic recombination and its regulation. *Trends Cell Biol.* 17, 448–55. doi:10.1016/j.tcb.2007.07.007.
- Culligan, K. M., and Britt, A. B. (2008). Both ATM and ATR promote the efficient and accurate processing of programmed meiotic double-strand breaks. *Plant J.* 55, 629–638. doi:10.1111/j.1365-313X.2008.03530.x.
- Davis, M. P. a, van Dongen, S., Abreu-Goodger, C., Bartonicek, N., and Enright, A. J. (2013). Kraken: A set of tools for quality control and analysis of high-throughput sequence data. *Methods* 63, 41–49. doi:10.1016/j.ymeth.2013.06.027.
- Dawe, R. K., Sedat, J. W., Agard, D. a, and Cande, W. Z. (1994). Meiotic chromosome pairing in maize is associated with a novel chromatin organization. *Cell* 76, 901–912. doi:0092-8674(94)90364-6 [pii].
- Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res.* 22, 1775–89. doi:10.1101/gr.132159.111.
- Deveshwar, P., Bovill, W. D., Sharma, R., Able, J. A., and Kapoor, S. (2011). Analysis of anther transcriptomes to identify genes contributing to meiosis and male gametophyte development in rice. *BMC Plant Biol.* 11, 78. doi:10.1186/1471-2229-11-78.
- Ding, D.-Q., Okamasa, K., Yamane, M., Tsutsumi, C., Haraguchi, T., Yamamoto, M., et al. (2012a). Meiosis-specific noncoding RNA mediates robust pairing of homologous chromosomes in meiosis. *Science* 336, 732–6. doi:10.1126/science.1219518.
- Ding, J., Lu, Q., and Ouyang, Y. (2012b). A long noncoding RNA regulates photoperiod-sensitive male sterility, an essential component of hybrid rice. *Proc. ...* 109, 2654–2659. doi:10.1073/pnas.1121374109/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1121374109.
- Dion, É., Li, L., Jean, M., and Belzile, F. (2007). An Arabidopsis MLH1 mutant exhibits reproductive defects and reveals a dual role for this gene in mitotic recombination. *Plant J.* 51, 431–440. doi:10.1111/j.1365-313X.2007.03145.x.
- Doebley, J. F., Gaut, B. S., and Smith, B. D. (2006). The Molecular Genetics of Crop Domestication. *Cell* 127, 1309–1321. doi:10.1016/j.cell.2006.12.006.
- Domínguez, E., and Mercado, J. (1999). Pollen sporopollenin: degradation and structural elucidation. *Sex. plant ...* 12, 171–178. Available at: <http://link.springer.com/article/10.1007/s004970050189> [Accessed March 26, 2014].
- Downs, S. M. (1997). Involvement of purine nucleotide synthetic pathways in gonadotropin-induced meiotic maturation in mouse cumulus cell-enclosed oocytes. *Mol. Reprod. Dev.* 46, 155–67. doi:10.1002/(SICI)1098-2795(199702)46:2<155::AID-MRD6>3.0.CO;2-P.
- Downs, S. M., Humpherson, P. G., and Leese, H. J. (1998). Meiotic induction in cumulus cell-enclosed mouse oocytes: involvement of the pentose phosphate pathway. *Biol. Reprod.* 58, 1084–94. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9546744>.
- Dray, E., Siaud, N., Dubois, E., and Doutriaux, M.-P. (2006). Interaction between Arabidopsis Brca2 and its partners Rad51, Dmc1, and Dss1. *Plant Physiol.* 140, 1059–1069.

- doi:10.1104/pp.105.075838.
- Dukowic-Schulze, S., Harris, A., Li, J., Sundararajan, A., Mudge, J., Retzel, E. F., et al. (2013). Comparative Transcriptomics of Early Meiosis in Arabidopsis and Maize. *J. Genet. Genomics*. doi:10.1016/j.jgg.2013.11.007.
- Dukowic-Schulze, S., Sundararajan, A., Mudge, J., Ramaraj, T., Farmer, A. D., Wang, M., et al. (2014). The transcriptome landscape of early maize meiosis. *BMC Plant Biol.* 14, 118. doi:10.1186/1471-2229-14-118.
- Eamens, A., Wang, M.-B., Smith, N. a, and Waterhouse, P. M. (2008). RNA silencing in plants: yesterday, today, and tomorrow. *Plant Physiol.* 147, 456–68. doi:10.1104/pp.108.117275.
- Esch, E., and Horn, R. (2008). Variability of Recombination Rates in Higher Plants. *Prog. Bot.*, 37–60. Available at: <http://www.springerlink.com/index/r681270l382v8181.pdf> [Accessed August 9, 2012].
- Feng, B., Lu, D., Ma, X., Peng, Y., Sun, Y., Ning, G., et al. (2012). Regulation of the Arabidopsis anther transcriptome by DYT1 for pollen development. *Plant J.* 72, 612–624. doi:10.1111/j.1365-313X.2012.05104.x.
- Flórez-Zapata, N. M. V., Reyes-Valdés, M. H., Hernandez-Godínez, F., and Martínez, O. (2014). Transcriptomic landscape of prophase I sunflower male meiocytes. *Front. Plant Sci.* 5, 277. doi:10.3389/fpls.2014.00277.
- Flórez-Zapata, N. M. V., Reyes-Valdés, M. H., and Martínez, O. (2016). Long non-coding RNAs are major contributors to transcriptome changes in sunflower meiocytes with different recombination rates. *BMC Genomics*, 1–16. doi:10.1186/s12864-016-2776-1.
- Franklin, a E., McElver, J., Sunjevaric, I., Rothstein, R., Bowen, B., and Cande, W. Z. (1999). Three-dimensional microscopy of the Rad51 recombination protein during meiotic prophase. *Plant Cell* 11, 809–24. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=144225&tool=pmcentrez&rendertype=abstract>.
- Gallego, M. E., Jeanneau, M., Granier, F., Bouchez, D., Bechtold, N., and White, C. I. (2001). Disruption of the Arabidopsis RAD50 gene leads to plant sterility and MMS sensitivity. *Plant J.* 25, 31–41. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11169180>.
- García-Ortega, L. F., and Martínez, O. (2015). How Many Genes Are Expressed in a Transcriptome? Estimation and Results for RNA-Seq. *PLoS One* 10, e0130262. doi:10.1371/journal.pone.0130262.
- Giordani, T., Cavallini, a., and Natali, L. (2014). The repetitive component of the sunflower genome. *Curr. Plant Biol.* 1, 45–54. doi:10.1016/j.cpb.2014.05.001.
- Girard, C., Chelysheva, L., Choinard, S., Froger, N., Macaisne, N., Lemhemdi, A., et al. (2015). AAA-ATPase FIDGETIN-LIKE 1 and Helicase FANCM Antagonize Meiotic Crossovers by Distinct Mechanisms. *PLoS Genet.* 11, e1005369. doi:10.1371/journal.pgen.1005369.
- Girard, C., Crismani, W., Froger, N., Mazel, J., Lemhemdi, A., Horlow, C., et al. (2014). FANCM-associated proteins MHF1 and MHF2, but not the other Fanconi anemia factors, limit meiotic crossovers. *Nucleic Acids Res.* 42, 9087–9095. doi:10.1093/nar/gku614.
- Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. a, Amit, I., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* 29, 644–52. doi:10.1038/nbt.1883.
- Grant-Downton, R., and Rodriguez-Enriquez, J. (2012). Emerging Roles for Non-Coding RNAs in Male Reproductive Development in Flowering Plants. *Biomolecules* 2, 608–621. doi:10.3390/biom2040608.
- Gravot, A., Deleu, C., Wagner, G., Lariagon, C., Lugan, R., Todd, C., et al. (2012). Arginase induction represses gall development during clubroot infection in arabidopsis. *Plant Cell Physiol.* 53, 901–911. doi:10.1093/pcp/pcs037.
- Grelon, M., Vezon, D., Gendrot, G., and Pelletier, G. (2001). AtSPO11-1 is necessary for efficient meiotic recombination in plants. *EMBO J.* 20, 589–600. doi:10.1093/emboj/20.3.589.
- Hamant, O., Ma, H., and Cande, W. Z. (2006a). Genetics of meiotic prophase I in plants. *Annu. Rev. Plant Biol.* 57, 267–302. doi:10.1146/annurev.arplant.57.032905.105255.

- Hamant, O., Ma, H., and Cande, W. Z. (2006b). Genetics of meiotic prophase I in plants. *Annu. Rev. Plant Biol.* 57, 267–302. doi:10.1146/annurev.arplant.57.032905.105255.
- Han, R., Jian, C., Lv, J., Yan, Y., Chi, Q., Li, Z., et al. (2014). Identification and characterization of microRNAs in the flag leaf and developing seed of wheat (*Triticum aestivum* L.). *BMC Genomics* 15, 289. doi:10.1186/1471-2164-15-289.
- Harris, M. a, Clark, J., Ireland, a, Lomax, J., Ashburner, M., Foulger, R., et al. (2004). The Gene Ontology (GO) database and informatics resource. *Nucleic Acids Res.* 32, D258–61. doi:10.1093/nar/gkh036.
- Harrison, C. J., Alvey, E., and Henderson, I. R. (2010). Meiosis in flowering plants and other green organisms. *J. Exp. Bot.* 61, 2863–75. doi:10.1093/jxb/erq191.
- Hartung, F., and Puchta, H. (2000). Molecular characterisation of two paralogous SPO11 homologues in *Arabidopsis thaliana*. *Nucleic Acids Res.* 28, 1548–54. Available at: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=102794&tool=pmcentrez&rendertype=abstract>.
- Hartung, F., Suer, S., Knoll, A., Wurz-Wildersinn, R., and Puchta, H. (2008). Topoisomerase 3?? and RMI1 suppress somatic crossovers and are essential for resolution of meiotic recombination intermediates in *Arabidopsis thaliana*. *PLoS Genet.* 4. doi:10.1371/journal.pgen.1000285.
- Heim, M. A., Jakoby, M., Werber, M., Martin, C., Weisshaar, B., and Bailey, P. C. (2003). The basic helix-loop-helix transcription factor family in plants: A genome-wide study of protein structure and functional diversity. *Mol. Biol. Evol.* 20, 735–747. doi:10.1093/molbev/msg088.
- Heyer, W.-D., Ehmsen, K. T., and Liu, J. (2010). Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* 44, 113–39. doi:10.1146/annurev-genet-051710-150955.
- Higgins, J. D., Buckling, E. F., Franklin, F. C. H., and Jones, G. H. (2008). Expression and functional analysis of AtMUS81 in *Arabidopsis* meiosis reveals a role in the second pathway of crossing-over. *Plant J.* 54, 152–162. doi:10.1111/j.1365-313X.2008.03403.x.
- Hopfner, K. P., Karcher, a, Shin, D. S., Craig, L., Arthur, L. M., Carney, J. P., et al. (2000). Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* 101, 789–800. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/10892749>.
- Hsu, F.-C., Chou, M.-Y., Peng, H.-P., Chou, S.-J., and Shih, M.-C. (2011). Insights into hypoxic systemic responses based on analyses of transcriptional regulation in *Arabidopsis*. *PLoS One* 6, e28888. doi:10.1371/journal.pone.0028888.
- Hwang, P. Y.-H., and Hunter, N. (2011). Mapping meiotic breaks: Spo11 oligonucleotides precisely mark the spots. *Genome Biol.* 12, 111. doi:10.1186/gb-2011-12-4-111.
- Da Ines, O., Degroote, F., Goubely, C., Amiard, S., Gallego, M. E., and White, C. I. (2013). Meiotic Recombination in *Arabidopsis* Is Catalysed by DMC1, with RAD51 Playing a Supporting Role. *PLoS Genet.* 9. doi:10.1371/journal.pgen.1003787.
- Ito, T., Ng, K.-H., Lim, T.-S., Yu, H., and Meyerowitz, E. M. (2007). The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in *Arabidopsis*. *Plant Cell* 19, 3516–29. doi:10.1105/tpc.107.055467.
- Jackson, N., Sanchez-Moran, E., Buckling, E., Armstrong, S. J., Jones, G. H., and Franklin, F. C. H. (2006). Reduced meiotic crossovers and delayed prophase I progression in AtMLH3-deficient *Arabidopsis*. *EMBO J.* 25, 1315–1323. doi:10.1038/sj.emboj.7600992.
- Jin, J., Zhang, H., Kong, L., Gao, G., and Luo, J. (2014). PlantTFDB 3.0: a portal for the functional and evolutionary study of plant transcription factors. *Nucleic Acids Res.* 42, D1182–7. doi:10.1093/nar/gkt1016.
- Johnson, R., and Guigó, R. (2014). The RIDL hypothesis : transposable elements as functional domains of long noncoding RNAs. *RNA* 20, 959–976. doi:10.1261/rna.044560.114.
- Jovanka, A. (2004). Roles of Interspecific Hybridization and Cytogenetic Studies in Sunflower. *Helia* 27, 1–24.
- Jurka, J. (2000). Repbase Update: A database and an electronic journal of repetitive elements. *Trends Genet.* 16, 418–420. doi:10.1016/S0168-9525(00)02093-X.
- Kannan, S., Chernikova, D., Rogozin, I. B., Poliakov, E., Managadze, D., Koonin, E. V., et al. (2015).

- Transposable Element Insertions in Long Intergenic Non-Coding RNA Genes. *Front Bioeng Biotechnol* 3, 71. doi:10.3389/fbioe.2015.00071.
- Keeney, S., Giroux, C. N., and Kleckner, N. (1997). Meiosis-specific DNA double-strand breaks are catalyzed by Spo11, a member of a widely conserved protein family. *Cell* 88, 375–84. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/9039264>.
- Kim, L., Hoe, K.-L., Yu, Y. M., Yeon, J.-H., and Maeng, P. J. (2012). The fission yeast GATA factor, Gaf1, modulates sexual development via direct down-regulation of *ste11+* expression in response to nitrogen starvation. *PLoS One* 7, e42409. doi:10.1371/journal.pone.0042409.
- Kleckner, N., Zickler, D., Jones, G. H., Dekker, J., Padmore, R., Henle, J., et al. (2004). A mechanical basis for chromosome function. *Proc. Natl. Acad. Sci. U. S. A.* 101, 12592–12597. doi:10.1073/pnas.0402724101.
- Knoll, A., and Puchta, H. (2011). The role of DNA helicases and their interaction partners in genome stability and meiotic recombination in plants. *J. Exp. Bot.* 62, 1565–1579. doi:10.1093/jxb/erq357.
- Knoll, A., Schröpfer, S., and Puchta, H. (2014). The RTR complex as caretaker of genome stability and its unique meiotic function in plants. *Front. Plant Sci.* 5, 33. doi:10.3389/fpls.2014.00033.
- Kohl, K. P., and Sekelsky, J. (2013). Meiotic and mitotic recombination in meiosis. *Genetics* 194, 327–334. doi:10.1534/genetics.113.150581.
- Komiya, R., Ohyanagi, H., Niihama, M., Watanabe, T., Nakano, M., Kurata, N., et al. (2014). Rice Germline-specific Argonaute MEL1 protein binds to phasiRNAs generated from more than 700 lincRNAs. *Plant J.* 78, 385–397. doi:10.1111/tpj.12483.
- Kong, L., Zhang, Y., Ye, Z.-Q., Liu, X.-Q., Zhao, S.-Q., Wei, L., et al. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. *Nucleic Acids Res.* 35, W345–9. doi:10.1093/nar/gkm391.
- Krishnakumar, V., Hanlon, M. R., Contrino, S., Ferlanti, E. S., Karamycheva, S., Kim, M., et al. (2015). Araport: The Arabidopsis Information Portal. *Nucleic Acids Res.* 43, D1003–D1009. doi:10.1093/nar/gku1200.
- Kurzbauer, M.-T., Uanschou, C., Chen, D., and Schlogelhofer, P. (2012). The Recombinases DMC1 and RAD51 Are Functionally and Spatially Separated during Meiosis in Arabidopsis. *Plant Cell* 24, 2058–2070. doi:10.1105/tpc.112.098459.
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–9. doi:10.1038/nmeth.1923.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25. doi:10.1186/gb-2009-10-3-r25 [pii]\n10.1186/gb-2009-10-3-r25.
- Lelandais-Brière, C., Sorin, C., Declerck, M., Benslimane, A., Crespi, M., and Hartmann, C. (2010). Small RNA diversity in plants and its impact in development. *Curr. Genomics* 11, 14–23. doi:10.2174/138920210790217918.
- Li, J., Dukowic-Schulze, S., Lindquist, I. E., Farmer, A. D., Kelly, B., Li, T., et al. (2015). The plant-specific protein FEHLSTART controls male meiotic entry, initializing meiotic synchronization in Arabidopsis. *Plant J.* 84, 659–671. doi:10.1111/tpj.13026.
- Li, J., Farmer, A. D., Lindquist, I. E., Dukowic-Schulze, S., Mudge, J., Li, T., et al. (2012). Characterization of a set of novel meiotically-active promoters in Arabidopsis. *BMC Plant Biol.* 12, 104. doi:10.1186/1471-2229-12-104.
- Li, L., Eichten, S. R., Shimizu, R., Petsch, K., Yeh, C.-T., Wu, W., et al. (2014). Genome-wide discovery and characterization of maize long non-coding RNAs. *Genome Biol.* 15, R40. doi:10.1186/gb-2014-15-2-r40.
- Lin, R., Ding, L., Casola, C., and Ripoll, D. (2007). Transposase-Derived Transcription Factors Regulate Light Signaling in Arabidopsis. *Science (80-.)*. 318, 1302–1305. Available at: <http://www.sciencemag.org/content/318/5854/1302.short> [Accessed March 23, 2014].
- Liu, C.-H., Finke, A., Díaz, M., Rozhon, W., Poppenberger, B., Baubec, T., et al. (2015a). Repair of DNA Damage Induced by the Cytidine Analog Zebularine Requires ATR and ATM in Arabidopsis. *Plant Cell* 27, 1788–1800. doi:10.1105/tpc.114.135467.

- Liu, X., Chen, C.-Y., Wang, K.-C., Luo, M., Tai, R., Yuan, L., et al. (2013). PHYTOCHROME INTERACTING FACTOR3 associates with the histone deacetylase HDA15 in repression of chlorophyll biosynthesis and photosynthesis in etiolated Arabidopsis seedlings. *Plant Cell* 25, 1258–73. doi:10.1105/tpc.113.109710.
- Liu, X., Hao, L., Li, D., Zhu, L., and Hu, S. (2015b). Long Non-coding RNAs and Their Biological Roles in Plants. *Genomics. Proteomics Bioinformatics* 13, 137–147. doi:10.1016/j.gpb.2015.02.003.
- López, E., Pradillo, M., Oliver, C., Romero, C., Cuñado, N., and Santos, J. L. (2012). Looking for natural variation in chiasma frequency in Arabidopsis thaliana. *J. Exp. Bot.* 63, 887–894. doi:10.1093/jxb/err319.
- De los Santos, T., Hunter, N., Lee, C., Larkin, B., Loidl, J., and Hollingsworth, N. M. (2003). The MUS81/MMS4 endonuclease acts independently of double-holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics* 164, 81–94.
- Lu, P., Han, X., Qi, J., Yang, J., Wijeratne, A. J., Li, T., et al. (2012). Analysis of Arabidopsis genome-wide variations before and after meiosis and meiotic recombination by resequencing Landsberg erecta and all four products of a single meiosis. *Genome Res.* 22, 508–518. doi:10.1101/gr.127522.111.
- Lu, P., Wijeratne, A. J., Wang, Z., Copenhaver, G. P., and Ma, H. (2014). Arabidopsis PTD Is Required for Type I Crossover Formation and Affects Recombination Frequency in Two Different Chromosomal Regions. *J. Genet. Genomics* 41, 165–175. doi:10.1016/j.jgg.2014.02.001.
- Ma, H. (2006). A molecular portrait of Arabidopsis meiosis. *Arabidopsis Book* 4, e0095. doi:10.1199/tab.0095.
- Ma, H., Wijeratne, A. J., Chen, C., Zhang, W., and Timofejeva, L. (2006). The Arabidopsis thaliana PARTING DANCERS gene encoding a novel protein is required for normal meiotic homologous recombination. *Mol. Biol. Cell* 17, 1331–43. doi:10.1091/mbc.E05-09-0902.
- Ma, J., Yan, B., Qu, Y., Qin, F., Yang, Y., Hao, X., et al. (2008). Zm401, a short-open reading-frame mRNA or noncoding RNA, is essential for tapetum and microspore development and can regulate the floret formation in maize. *J. Cell. Biochem.* 105, 136–46. doi:10.1002/jcb.21807.
- Ma, X., Shao, C., Jin, Y., Wang, H., and Meng, Y. (2014). Long non-coding RNAs: A novel endogenous source for the generation of Dicer-like 1-dependent small RNAs in Arabidopsis thaliana. *RNA Biol* 11, 373–390. doi:10.4161/rna.28725.
- Marioni, J. C., Mason, C. E., Mane, S. M., Stephens, M., and Gilad, Y. (2008). comparison with gene expression arrays RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. 1509–1517. doi:10.1101/gr.079558.108.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17, 10. doi:10.14806/ej.17.1.200.
- Martínez-López, L., Ochoa-Alejo, N., and Martínez, O. (2014). Dynamics of the chili pepper transcriptome during fruit development. *BMC Genomics* 15, 143. doi:10.1186/1471-2164-15-143.
- Martinez-Perez, E., and Colaiácovo, M. P. (2009). Distribution of meiotic recombination events: talking to your neighbors. *Curr. Opin. Genet. Dev.* 19, 105–12. doi:10.1016/j.gde.2009.02.005.
- Martínez, O., and Reyes-Valdés, M. H. (2008). Defining diversity, specialization, and gene specificity in transcriptomes through information theory. *Proc. Natl. Acad. Sci. U. S. A.* 105, 9709–14. doi:10.1073/pnas.0803479105.
- Martini, E., Diaz, R. L., Hunter, N., and Keeney, S. (2006). Crossover Homeostasis in Yeast Meiosis. *Cell* 126, 285–295. doi:10.1016/j.cell.2006.05.044.
- Mata, J., Lyne, R., Burns, G., and Bähler, J. (2002). The transcriptional program of meiosis and sporulation in fission yeast. *Nat. Genet.* 32, 143–7. doi:10.1038/ng951.
- Matzke, M., Kanno, T., Daxinger, L., Huettel, B., and Matzke, A. J. (2009). RNA-mediated chromatin-based silencing in plants. *Curr. Opin. Cell Biol.* 21, 367–376. doi:10.1016/j.ceb.2009.01.025.
- McMahill, M. S., Sham, C. W., and Bishop, D. K. (2007). Synthesis-dependent strand annealing in meiosis. *PLoS Biol.* 5, e299. doi:10.1371/journal.pbio.0050299.
- Meinke, D. W., Cherry, J. M., Dean, C., Rounsley, S. D., and Koornneef, M. (1998). Arabidopsis

- thaliana: A Model Plant for Genome Analysis. *Science* (80-). 282, 662–682. doi:10.1126/science.282.5389.662.
- Mercier, R., and Grelon, M. (2008). Meiosis in plants: ten years of gene discovery. *Cytogenet. Genome Res.* 120, 281–90. doi:10.1159/000121077.
- Mercier, R., Grelon, M., Vezon, D., Horlow, C., and Pelletier, G. (2001). How to characterize meiotic functions in plants? *Biochimie* 83, 1023–8. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/11879730>.
- Mercier, R., Mézard, C., Jenczewski, E., Macaisne, N., and Grelon, M. (2015). The Molecular Biology of Meiosis in Plants. *Annu. Rev. Plant Biol.* 66, 141210140145001–. doi:10.1146/annurev-arplant-050213-035923.
- Mi, S., Cai, T., Hu, Y., Chen, Y., Hodges, E., Ni, F., et al. (2008). Sorting of Small RNAs into Arabidopsis Argonaute Complexes Is Directed by the 5' Terminal Nucleotide. *Cell* 133, 116–127. doi:10.1016/j.cell.2008.02.034.
- Miao, C., Tang, D., Zhang, H., Wang, M., Li, Y., Tang, S., et al. (2013). CENTRAL REGION COMPONENT1, a Novel Synaptonemal Complex Component, Is Essential for Meiotic Recombination Initiation in Rice. *Plant Cell* 25, 2998–3009. doi:10.1105/tpc.113.113175.
- Mirouze, M., Lieberman-Lazarovich, M., Aversano, R., Bucher, E., Nicolet, J., Reinders, J., et al. (2012). Loss of DNA methylation affects the recombination landscape in Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 5880–5. doi:10.1073/pnas.1120841109.
- Montes, R. a. C., de Fátima Rosas-Cárdenas, F., De Paoli, E., Accerbi, M., Rymarquis, L. a., Mahalingam, G., et al. (2014). Sample sequencing of vascular plants demonstrates widespread conservation and divergence of microRNAs. *Nat. Commun.* 5. doi:10.1038/ncomms4722.
- Morohashi, K., Minami, M., Takase, H., Hotta, Y., and Hiratsuka, K. (2003). Isolation and characterization of a novel GRAS gene that regulates meiosis-associated gene expression. *J. Biol. Chem.* 278, 20865–731. Morohashi, K., Minami, M., Takase, H., . doi:10.1074/jbc.M301712200.
- Moss, J., Tinline-Purvis, H., Walker, C. a, Folkes, L. K., Stratford, M. R., Hayles, J., et al. (2010). Break-induced ATR and Ddb1-Cul4(Cdt)² ubiquitin ligase-dependent nucleotide synthesis promotes homologous recombination repair in fission yeast. *Genes Dev.* 24, 2705–16. doi:10.1101/gad.1970810.
- Mueller, L. A. (2003). AraCyc: A Biochemical Pathway Database for Arabidopsis. *PLANT Physiol.* 132, 453–460. doi:10.1104/pp.102.017236.
- Murmu, J., Bush, M. J., DeLong, C., Li, S., Xu, M., Khan, M., et al. (2010). Arabidopsis basic leucine-zipper transcription factors TGA9 and TGA10 interact with floral glutaredoxins ROXY1 and ROXY2 and are redundantly required for anther development. *Plant Physiol.* 154, 1492–1504. doi:10.1104/pp.110.159111.
- De Muyt, A., Pereira, L., Vezon, D., Chelysheva, L., Gendrot, G., Chambon, A., et al. (2009). A high throughput genetic screen identifies new early meiotic recombination functions in Arabidopsis thaliana. *PLoS Genet.* 5, e1000654. doi:10.1371/journal.pgen.1000654.
- De Muyt, A., Vezon, D., Gendrot, G., Gallois, J.-L., Stevens, R., and Grelon, M. (2007). AtPRD1 is required for meiotic double strand break formation in Arabidopsis thaliana. *EMBO J.* 26, 4126–4137. doi:10.1038/.
- Nan, G.-L., Ronceret, A., Wang, R. C., Fernandes, J. F., Cande, W. Z., and Walbot, V. (2011). Global transcriptome analysis of two ameiotic1 alleles in maize anthers: defining steps in meiotic entry and progression through prophase I. *BMC Plant Biol.* 11, 120. doi:10.1186/1471-2229-11-120.
- Nara, T., Hamada, F., Namekawa, S., and Sakaguchi, K. (2001). Strand exchange reaction in vitro and DNA-dependent ATPase activity of recombinant LIM15/DMC1 and RAD51 proteins from *Coprinus cinereus*. *Biochem. Biophys. Res. Commun.* 285, 92–7. doi:10.1006/bbrc.2001.5095.
- Natali, L., Cossu, R., Barghini, E., Giordani, T., Buti, M., Mascagni, F., et al. (2013). The repetitive component of the sunflower genome as shown by different procedures for assembling next generation sequencing reads. *BMC Genomics* 14, 686. doi:10.1186/1471-2164-14-686.
- Necsulea, A., Soumillon, M., Warnefors, M., Liechti, A., Daish, T., Zeller, U., et al. (2014). The evolution of lncRNA repertoires and expression patterns in tetrapods. *Nature* 505, 635–640.

- doi:10.1038/nature12943.
- Nonomura, K.-I., Morohoshi, A., Nakano, M., Eiguchi, M., Miyao, A., Hirochika, H., et al. (2007). A germ cell specific gene of the ARGONAUTE family is essential for the progression of premeiotic mitosis and meiosis during sporogenesis in rice. *Plant Cell* 19, 2583–94. doi:10.1105/tpc.107.053199.
- Nussbaumer, T., Martis, M. M., Roessner, S. K., Pfeifer, M., Bader, K. C., Sharma, S., et al. (2013). MIPS PlantsDB: A database framework for comparative plant genome research. *Nucleic Acids Res.* 41, 1144–1151. doi:10.1093/nar/gks1153.
- Obayashi, T., Kinoshita, K., Nakai, K., Shibaoka, M., Hayashi, S., Saeki, M., et al. (2007). ATTED-II: A database of co-expressed genes and cis elements for identifying co-regulated gene groups in Arabidopsis. *Nucleic Acids Res.* 35, 4–6. doi:10.1093/nar/gkl783.
- Obayashi, T., Nishida, K., Kasahara, K., and Kinoshita, K. (2011). ATTED-II updates: Condition-specific gene coexpression to extend coexpression analyses and applications to a broad range of flowering plants. *Plant Cell Physiol.* 52, 213–219. doi:10.1093/pcp/pcq203.
- Oliver, C., Pradillo, M., Corredor, E., and Cuñado, N. (2013). The dynamics of histone H3 modifications is species-specific in plant meiosis. *Planta* 238, 23–33. doi:10.1007/s00425-013-1885-1.
- Olmedo-Monfil, V., Durán-Figueroa, N., Arteaga-Vázquez, M., Demesa-Arévalo, E., Autran, D., Grimanelli, D., et al. (2010). Control of female gamete formation by a small RNA pathway in Arabidopsis. *Nature* 464, 628–632. doi:10.1038/nature08828.
- Olsen, K. M., and Wendel, J. F. (2013). A Bountiful Harvest: Genomic Insights into Crop Domestication Phenotypes. *Annu. Rev. Plant Biol.* 64, 47–70. doi:10.1146/annurev-arplant-050312-120048.
- Omidvar, V., Mohorianu, I., Dalmay, T., and Fellner, M. (2015). Identification of miRNAs with potential roles in regulation of anther development and male-sterility in 7B-1 male-sterile tomato mutant. *BMC Genomics* 16, 878. doi:10.1186/s12864-015-2077-0.
- Oshlack, A., Robinson, M. D., and Young, M. D. (2010). From RNA-seq reads to differential expression results. *Genome Biol.* 11, 220. doi:10.1186/gb-2010-11-12-220.
- Osman, K., Higgins, J. D., Sanchez-Moran, E., Armstrong, S. J., and Franklin, F. C. H. (2011). Pathways to meiotic recombination in Arabidopsis thaliana. *New Phytol.* 190, 523–44. doi:10.1111/j.1469-8137.2011.03665.x.
- Osman, K., Sanchez-Moran, E., Mann, S. C., Jones, G. H., and Franklin, F. C. H. (2009). Replication protein A (AtRPA1a) is required for class I crossover formation but is dispensable for meiotic DNA break repair. *EMBO J.* 28, 394–404. doi:10.1038/emboj.2008.295.
- Otto, S. P., and Barton, N. H. (2001). Selection for recombination in small populations. *Evolution* 55, 1921–1931. doi:10.2307/2680441.
- Pan, J., Sasaki, M., Kniewel, R., Murakami, H., Blitzblau, H. G., Tischfield, S. E., et al. (2011). A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* 144, 719–31. doi:10.1016/j.cell.2011.02.009.
- Panoli, A. P., Ravi, M., Sebastian, J., Nishal, B., Reddy, T. V., Marimuthu, M. P. a, et al. (2006). AtMND1 is required for homologous pairing during meiosis in Arabidopsis. *BMC Mol. Biol.* 7, 24. doi:10.1186/1471-2199-7-24.
- Pantaleo, V., Szittyá, G., Moxon, S., Miozzi, L., Moulton, V., Dalmay, T., et al. (2010). Identification of grapevine microRNAs and their targets using high-throughput sequencing and degradome analysis. *Plant J.* 62, 960–976. doi:10.1111/j.1365-313X.2010.04208.x.
- Pawlowski, W. P., and Cande, W. Z. (2005). Coordinating the events of the meiotic prophase. *Trends Cell Biol.* 15, 674–81. doi:10.1016/j.tcb.2005.10.005.
- Pawlowski, W. P., Golubovskaya, I. N., Timofejeva, L., Meeley, R. B., Sheridan, W. F., and Cande, W. Z. (2004). Coordination of meiotic recombination, pairing, and synapsis by PHS1. *Science* 303, 89–92. doi:10.1126/science.1091110.
- Pawlowski, W. P., Wang, C.-J. R., Golubovskaya, I. N., Szymaniak, J. M., Shi, L., Hamant, O., et al. (2009). Maize AMELOTIC1 is essential for multiple early meiotic processes and likely required for the initiation of meiosis. *Proc. Natl. Acad. Sci. U. S. A.* 106, 3603–8.

- doi:10.1073/pnas.0810115106.
- Phadnis, N., Hyppa, R. W., and Smith, G. R. (2011). New and old ways to control meiotic recombination. *Trends Genet.* 27, 411–21. doi:10.1016/j.tig.2011.06.007.
- Phillips, C. M., and Dernburg, A. F. (2006). A family of zinc-finger proteins is required for chromosome-specific pairing and synapsis during meiosis in *C. elegans*. *Dev. Cell* 11, 817–29. doi:10.1016/j.devcel.2006.09.020.
- Phillips, R. D., and Jennings, D. H. (1976). Succulence, Cations and Organic Acids in Leaves of *Kalanchoe Daigremontiana* Grown in Long and Short Days in Soil and Water Culture. *New Phytol.* 77, 599–611. doi:10.1111/j.1469-8137.1976.tb04652.x.
- Pradillo, M., López, E., Linacero, R., Romero, C., Cuadado, N., Sánchez-Morán, E., et al. (2012). Together yes, but not coupled: New insights into the roles of RAD51 and DMC1 in plant meiotic recombination. *Plant J.* 69, 921–933. doi:10.1111/j.1365-3113.2011.04845.x.
- Pradillo, M., Varas, J., Oliver, C., and Santos, J. L. (2014). On the role of AtDMC1, AtRAD51 and its paralogs during Arabidopsis meiosis. *Front. Plant Sci.* 5, 23. doi:10.3389/fpls.2014.00023.
- Proost, S., Van Bel, M., Sterck, L., Billiau, K., Van Parys, T., Van de Peer, Y., et al. (2009). PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* 21, 3718–31. doi:10.1105/tpc.109.071506.
- R Core Team (2013). R: A Language and Environment for Statistical Computing. Available at: <http://www.r-project.org/>.
- Ray, D., and Ye, P. (2013). Characterization of the metabolic requirements in yeast meiosis. *PLoS One* 8, e63707. doi:10.1371/journal.pone.0063707.
- Ribeiro, J., Abby, E., Livera, G., and Martini, E. (2015). RPA homologs and ssDNA processing during meiotic recombination. *Chromosoma*, 265–276. doi:10.1007/s00412-015-0552-7.
- Riehs, N., Akimcheva, S., Puizina, J., Bulankova, P., Idol, R. A., Siroky, J., et al. (2008). Arabidopsis SMG7 protein is required for exit from meiosis. *J. Cell Sci.* 121, 2208–16. doi:10.1242/jcs.027862.
- Rieseberg, L. H., Choi, H., Chan, R., and Spore, C. (1993). Genomic map of a diploid hybrid species. *Heredity (Edinb.)*. 70, 285 – 293. Available at: <http://dx.doi.org/10.1038/hdy.1993.41> [Accessed August 9, 2012].
- Rieseberg, L. H., Van Fossen, C., and Desrochers, A. M. (1995). Hybrid speciation accompanied by genomic reorganization in wild sunflowers. *Nature* 375, 313– 316. Available at: <http://dx.doi.org/10.1038/375313a0> [Accessed August 9, 2012].
- Rinn, J. L., and Chang, H. Y. (2012). Genome Regulation by Long Noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166. doi:10.1146/annurev-biochem-051410-092902.
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–40. doi:10.1093/bioinformatics/btp616.
- Rodríguez, J., Gómez, M., and Reyes-Valdés, M. H. (2007). Análisis meiótico de una cruce entre girasol cultivado (*Helianthus annuus* L. var. macrocarpus) y girasol silvestre (*Helianthus annuus* L. ssp. texanus Heiser). *Acta Botánica Mex.* 80, 7–20. Available at: <http://redalyc.uaemex.mx/redalyc/pdf/574/57408002.pdf> [Accessed August 9, 2012].
- Ronceret, A., and Vielle-Calzada, J. P. (2015). Meiosis, unreduced gametes, and parthenogenesis: implications for engineering clonal seed formation in crops. *Plant Reprod.* 28, 91–102. doi:10.1007/s00497-015-0262-6.
- Rosa, M., Von Harder, M., Aiese Cigliano, R., Schlogelhofer, P., and Mittelsten Scheid, O. (2013). The Arabidopsis SWR1 Chromatin-Remodeling Complex Is Important for DNA Repair, Somatic Recombination, and Meiosis. *Plant Cell* 25, 1990–2001. doi:10.1105/tpc.112.104067.
- Ross-Ibarra, J. (2004). The evolution of recombination under domestication: a test of two hypotheses. *Am. Nat.* 163, 105–112. Available at: <http://www.jstor.org/stable/10.1086/380606> [Accessed August 9, 2012].
- Sabetta, W., Alba, V., Blanco, A., and Montemurro, C. (2011). sunTILL: a TILLING resource for gene function analysis in sunflower. *Plant Methods* 7, 20. doi:10.1186/1746-4811-7-20.
- Salmon, A. Y., Salmon-Divon, M., Zahavi, T., Barash, Y., Levy-Drummer, R. S., Jacob-Hirsch, J., et al.

- (2013). Determination of molecular markers for BRCA1 and BRCA2 heterozygosity using gene expression profiling. *Cancer Prev. Res. (Phila)*. 6, 82–90. doi:10.1158/1940-6207.CAPR-12-0105.
- San Filippo, J., Sung, P., and Klein, H. (2008). Mechanism of eukaryotic homologous recombination. *Annu. Rev. Biochem.* 77, 229–57. doi:10.1146/annurev.biochem.77.061306.125255.
- Sanchez-Moran, E., Armstrong, S. J., Santos, J. L., Franklin, F. C. H., and Jones, G. H. (2002). Variation in chiasma frequency among eight accessions of *Arabidopsis thaliana*. *Genetics* 162, 1415–1422.
- Sanchez-Moran, E., Santos, J. L., Jones, G. H., and Franklin, F. C. H. (2007). ASY1 mediates AtDMC1-dependent interhomolog recombination during meiosis in *Arabidopsis*. *Genes Dev.* 21, 2220–2233. doi:10.1101/gad.439007.
- Sanchez, A., Ocana, J., and Salicru, M. goProfiles: goProfiles: an R package for the statistical analysis of functional profiles.
- Schmidt, A., Wuest, S. E., Vijverberg, K., Baroux, C., Kleen, D., and Grossniklaus, U. (2011). Transcriptome analysis of the *Arabidopsis* megaspore mother cell uncovers the importance of RNA helicases for plant germline development. *PLoS Biol.* 9. doi:10.1371/journal.pbio.1001155.
- Schmieder, R., and Edwards, R. (2011). Quality control and preprocessing of metagenomic datasets. *Bioinformatics* 27, 863–4. doi:10.1093/bioinformatics/btr026.
- Schneiter, A. A., F, M. J., and Berglund, D. R. (1998). Stages of Sunflower Development. *NDSU Ext. Serv.* 21, 7–8. Available at: doi:10.2135/cropsci1981.0011183X002100060024x [Accessed March 20, 2014].
- Schuermann, D., Molinier, J., Fritsch, O., and Hohn, B. (2005). The dual nature of homologous recombination in plants. *Trends Genet.* 21, 172–181. doi:10.1016/j.tig.2005.01.002.
- Seeliger, K., Dukowic-Schulze, S., Wurz-Wildersinn, R., Pacher, M., and Puchta, H. (2012). BRCA2 is a mediator of RAD51- and DMC1-facilitated homologous recombination in *Arabidopsis thaliana*. *New Phytol.* 193, 364–375. doi:10.1111/j.1469-8137.2011.03947.x.
- Séguéla-Arnaud, M., Crismani, W., Larchevêque, C., Mazel, J., Froger, N., Choinard, S., et al. (2015). Multiple mechanisms limit meiotic crossovers: TOP3 α and two BLM homologs antagonize crossovers in parallel to FANCM. *Proc. Natl. Acad. Sci. U. S. A.* 112, 4713–8. doi:10.1073/pnas.1423107112.
- Seiler, G., and Jan, C. (2010). “Basic information,” in *Genetics, Genomics and Breeding of Sunflower*, eds. J. Hu, G. Seiler, and C. Kole (CRC Press), 1–40.
- Serrentino, M.-E., and Borde, V. (2012). The spatial regulation of meiotic recombination hotspots: are all DSB hotspots crossover hotspots? *Exp. Cell Res.* 318, 1347–52. doi:10.1016/j.yexcr.2012.03.025.
- Sharma, V., Khurana, S., Kubben, N., Abdelmohsen, K., Oberdoerffer, P., Gorospe, M., et al. (2015). A BRCA 1 -interacting lncRNA regulates homologous recombination. *EMBO Rep.* 16, 1520–1534.
- Shilo, S., Melamed-Bessudo, C., Dorone, Y., Barkai, N., and Levy, A. A. (2015). DNA Crossover Motifs Associated with Epigenetic Modifications Delineate Open Chromatin Regions in *Arabidopsis*. *Plant Cell* 27, tpc.15.00391. doi:10.1105/tpc.15.00391.
- Shinohara, A., and Shinohara, M. (2004). Roles of RecA homologues Rad51 and Dmc1 during meiotic recombination. *Cytogenet. Genome Res.* 107, 201–207. doi:10.1159/000080598.
- Shuai, P., Liang, D., Tang, S., Zhang, Z., Ye, C.-Y., Su, Y., et al. (2014). Genome-wide identification and functional prediction of novel and drought-responsive lincRNAs in *Populus trichocarpa*. *J. Exp. Bot.* 65, 4975–83. doi:10.1093/jxb/eru256.
- Siaud, N., Dray, E., Gy, I., Gérard, E., Takvorian, N., and Doutriaux, M.-P. (2004). Brca2 is involved in meiosis in *Arabidopsis thaliana* as suggested by its interaction with Dmc1. *EMBO J.* 23, 1392–1401. doi:10.1038/sj.emboj.7600146.
- Simchen, G., and Stamberg, J. (1969). Fine and coarse controls of genetic recombination. *Nature* 222, 329–332. Available at: <http://adsabs.harvard.edu/abs/1969Natur.222..329S> [Accessed August 9, 2012].
- Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–85. doi:10.1038/nrg2072.

- Soumillon, M., Necsulea, A., Weier, M., Brawand, D., Zhang, X., Gu, H., et al. (2013). Cellular source and mechanisms of high transcriptome complexity in the mammalian testis. *Cell Rep.* 3, 2179–90. doi:10.1016/j.celrep.2013.05.031.
- Stefaniak, T. R., Hyten, D. L., Pantalone, V. R., Klarer, a., and Pfeiffer, T. W. (2006). Soybean Cultivars Resulted from More Recombination Events Than Unselected Lines in the Same Population. *Crop Sci.* 46, 43. doi:10.2135/cropsci2005.0016.
- Stein, K. K., Nesmith, J. E., Ross, B. D., and Golden, A. (2010). Functional redundancy of paralogs of an anaphase promoting complex/cyclosome subunit in *Caenorhabditis elegans* meiosis. *Genetics* 186, 1285–1293. doi:10.1534/genetics.110.123463.
- Stevens, R., Grelon, M., Vezon, D., Oh, J., Meyer, P., Perennes, C., et al. (2004). A CDC45 homolog in *Arabidopsis* is essential for meiosis, as shown by RNA interference-induced gene silencing. *Plant Cell* 16, 99–113. doi:10.1105/tpc.016865.
- Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100, 9440–5. doi:10.1073/pnas.1530509100.
- De Storme, N., and Geelen, D. (2014). The impact of environmental stress on male reproductive development in plants: biological processes and molecular mechanisms. *Plant. Cell Environ.* 37, 1–18. doi:10.1111/pce.12142.
- Strickler, S. R., Bombarely, A., and Mueller, L. a (2012). Designing a transcriptome next-generation sequencing project for a nonmodel plant species1. *Am. J. Bot.* 99, 257–66. doi:10.3732/ajb.1100292.
- Sun, Y., Ambrose, J. H., Haughey, B. S., Webster, T. D., Pierrie, S. N., Mu??oz, D. F., et al. (2012). Deep Genome-Wide Measurement of Meiotic Gene Conversion Using Tetrad Analysis in *Arabidopsis thaliana*. *PLoS Genet.* 8. doi:10.1371/journal.pgen.1002968.
- Sung, P., and Klein, H. (2006). Mechanism of homologous recombination: mediators and helicases take on regulatory functions. *Nat. Rev. Mol. Cell Biol.* 7, 739–751. doi:10.1038/nrm2008.
- Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M. B., Ross-Ibarra, J., Myers, C. L., et al. (2012). Reshaping of the maize transcriptome by domestication. *Proc. Natl. Acad. Sci. U. S. A.* 109, 11878–83. doi:10.1073/pnas.1201961109.
- Turinetto, V., and Giachino, C. (2015). Multiple facets of histone variant H2AX: a DNA double-strand-break marker with several biological functions. *Nucleic Acids Res.* 43, 2489–98. doi:10.1093/nar/gkv061.
- Uanschou, C., Ronceret, A., Von Harder, M., De Muyt, A., Vezon, D., Pereira, L., et al. (2013). Sufficient amounts of functional HOP2/MND1 complex promote interhomolog DNA repair but are dispensable for intersister DNA repair during meiosis in *Arabidopsis*. *Plant Cell* 25, 4924–40. doi:10.1105/tpc.113.118521.
- Varas, J., Graumann, K., Osman, K., Pradillo, M., Evans, D. E., Santos, J. L., et al. (2015). Absence of SUN1 and SUN2 proteins in *Arabidopsis thaliana* leads to a delay in meiotic progression and defects in synapsis and recombination. *Plant J.* 81, 329–346. doi:10.1111/tpj.12730.
- Vaucheret, H. (2006). Post-transcriptional small RNA pathways in plants: Mechanisms and regulations. *Genes Dev.* 20, 759–771. doi:10.1101/gad.1410506.
- Vaucheret, H. (2008). Plant ARGONAUTES. *Trends Plant Sci.* 13, 350–8. doi:10.1016/j.tplants.2008.04.007.
- Vershon, A. K., and Pierce, M. (2000). Transcriptional regulation of meiosis in yeast. *Curr. Opin. Cell Biol.* 12, 334–339. doi:10.1016/S0955-0674(00)00104-6.
- Vitulo, N., Forcato, C., Carpinelli, E., Telatin, A., Campagna, D., D’Angelo, M., et al. (2014). A deep survey of alternative splicing in grape reveals changes in the splicing machinery related to tissue, stress condition and genotype. *BMC Plant Biol.* 14, 99. doi:10.1186/1471-2229-14-99.
- Wahls, W. P., and Davidson, M. K. (2010). Discrete DNA sites regulate global distribution of meiotic recombination. *Trends Genet.* 26, 202–8. doi:10.1016/j.tig.2010.02.003.
- Wang, D., Zhang, C., Hearn, D. J., Kang, I.-H., Punwani, J. A., Skaggs, M. I., et al. (2010). Identification of transcription-factor genes expressed in the *Arabidopsis* female gametophyte. *BMC Plant Biol.* 10, 1–15. doi:10.1186/1471-2229-10-110.
- Wang, L., Park, H. J., Dasari, S., Wang, S., Kocher, J.-P., and Li, W. (2013). CPAT: Coding-Potential

- Assessment Tool using an alignment-free logistic regression model. *Nucleic Acids Res.* 41, e74–e74. doi:10.1093/nar/gkt006.
- Wang, X., Ai, G., Zhang, C., Cui, L., Wang, J., Li, H., et al. (2015). Expression and diversification analysis reveals transposable elements play important roles in the origin of Lycopersicon-specific lncRNAs in tomato. *New Phytol.* doi:10.1111/nph.13718.
- Wang, Y., Xiao, R., Wang, H., Cheng, Z., Li, W., Zhu, G., et al. (2014). The Arabidopsis RAD51 paralogs RAD51B, RAD51D and XRCC2 play partially redundant roles in somatic DNA repair and gene regulation. *New Phytol.* 201, 292–304. doi:10.1111/nph.12498.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi:10.1038/nrg2484.
- Waterworth, W. M., Drury, G. E., Bray, C. M., and West, C. E. (2011). Repairing breaks in the plant genome: The importance of keeping it together. *New Phytol.* 192, 805–822. doi:10.1111/j.1469-8137.2011.03926.x.
- Watts, J. a, Morley, M., Burdick, J. T., Fiori, J. L., Ewens, W. J., Spielman, R. S., et al. (2002). Gene expression phenotype in heterozygous carriers of ataxia telangiectasia. *Am. J. Hum. Genet.* 71, 791–800. doi:10.1086/342974.
- Wei, M., Wei, H., Wu, M., Song, M., Zhang, J., Yu, J., et al. (2013). Comparative expression profiling of miRNA during anther development in genetic male sterile and wild type cotton. *BMC Plant Biol.* 19, 66. doi:10.1186/1471-2229-13-66.
- van Werven, F. J., and Amon, a. (2011). Regulation of entry into gametogenesis. *Philos. Trans. R. Soc. B Biol. Sci.* 366, 3521–3531. doi:10.1098/rstb.2011.0081.
- Wijeratne, A. J., Zhang, W., Sun, Y., Liu, W., Albert, R., Zheng, Z., et al. (2007). Differential gene expression in Arabidopsis wild-type and mutant anthers: insights into anther cell differentiation and regulatory networks. *Plant J.* 52, 14–29. doi:10.1111/j.1365-313X.2007.03217.x.
- Wijnker, E., van Dun, K., de Snoo, C. B., Lelivelt, C. L. C., Keurentjes, J. J. B., Naharudin, N. S., et al. (2012). Reverse breeding in Arabidopsis thaliana generates homozygous parental lines from a heterozygous plant. *Nat. Genet.* 44, 467–70. doi:10.1038/ng.2203.
- Wijnker, E., James, G. V., Ding, J., Becker, F., Klasen, J. R., Rawat, V., et al. (2013). The genomic landscape of meiotic crossovers and gene conversions in Arabidopsis thaliana. *Elife* 2013, 1–22. doi:10.7554/eLife.01426.
- Wijnker, E., and de Jong, H. (2008). Managing meiotic recombination in plant breeding. *Trends Plant Sci.* 13, 640–6. doi:10.1016/j.tplants.2008.09.004.
- Wijnker, E., and Schnittger, A. (2013). Control of the meiotic cell division program in plants. *Plant Reprod.* 26, 143–158. doi:10.1007/s00497-013-0223-x.
- Wilson, Z. a., and Yang, C. (2004). Plant gametogenesis: Conservation and contrasts in development. *Reproduction* 128, 483–492. doi:10.1530/rep.1.00306.
- Winter, E. (2012). The Sum1/Ndt80 transcriptional switch and commitment to meiosis in *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* 76, 1–15. doi:10.1128/MMBR.05010-11.
- Wu, H.-J., Wang, Z.-M., Wang, M., and Wang, X.-J. (2013). Widespread long noncoding RNAs as endogenous target mimics for microRNAs in plants. *Plant Physiol.* 161, 1875–84. doi:10.1104/pp.113.215962.
- Yang, H., Lu, P., Wang, Y., and Ma, H. (2011). The transcriptome landscape of Arabidopsis male meiocytes from high-throughput sequencing: the complexity and evolution of the meiotic process. *Plant J.* 65, 503–16. doi:10.1111/j.1365-313X.2010.04439.x.
- Yang, X., Makaroff, C. A., and Ma, H. (2003). The Arabidopsis MALE MEIOCYTE DEATH1 gene encodes a PHD-finger protein that is required for male meiosis. *Plant Cell Online* 15, 1281–1295. doi:10.1105/tpc.010447.Considerable.
- Yant, L., Hollister, J., Wright, K., Arnold, B., Higgins, J., Franklin, F. ??Chris??H, et al. (2013). Meiotic Adaptation to Genome Duplication in Arabidopsis arenosa. *Curr. Biol.* 23, 2151–2156. doi:10.1016/j.cub.2013.08.059.
- Yoon, J.-H., Abdelmohsen, K., and Gorospe, M. (2013). Posttranscriptional Gene Regulation by Long Noncoding RNA. *J. Mol. Biol.* 425, 3723–3730. doi:10.1016/j.jmb.2012.11.024.
- Youds, J. L., Boulton, S. J., Adamo, A., Montemauri, P., Silva, N., Ward, J. D., et al. (2011). The choice

- in meiosis - defining the factors that influence crossover or non-crossover formation. *J. Cell Sci.* 124, 501–13. doi:10.1242/jcs.074427.
- Zhai, J., Zhang, H., Arikiti, S., Huang, K., Nan, G.-L., Walbot, V., et al. (2015). Spatiotemporally dynamic, cell-type-dependent premeiotic and meiotic phasiRNAs in maize anthers. *Proc. Natl. Acad. Sci.* 112, 3146–3151. doi:10.1073/pnas.1418918112.
- Zhang, C., Song, Y., Cheng, Z. H., Wang, Y. X., Zhu, J., Ma, H., et al. (2012). The Arabidopsis thaliana DSB formation (AtDFO) gene is required for meiotic double-strand break formation. *Plant J.* 72, 271–281. doi:10.1111/j.1365-313X.2012.05075.x.
- Zhang, H., Chen, X., Wang, C., Xu, Z., Wang, Y., Liu, X., et al. (2013). Long non-coding genes implicated in response to stripe rust pathogen stress in wheat (*Triticum aestivum* L.). *Mol. Biol. Rep.* 40, 6245–6253. doi:10.1007/s11033-013-2736-7.
- Zhang, L., Wang, S., Yin, S., Hong, S., Kim, K. P., and Kleckner, N. (2014a). Topoisomerase II mediates meiotic crossover interference. *Nature*, 1–21. doi:10.1038/nature13442.
- Zhang, W., Sun, Y., Timofejeva, L., Chen, C., Grossniklaus, U., and Ma, H. (2006). Regulation of Arabidopsis tapetum development and function by DYSFUNCTIONAL TAPETUM1 (DYT1) encoding a putative bHLH transcription factor. *Development* 133, 3085–3095. doi:10.1242/dev.02463.
- Zhang, Y.-C., Liao, J.-Y., Li, Z.-Y., Yu, Y., Zhang, J.-P., Li, Q.-F., et al. (2014b). Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 15, 512. doi:10.1186/s13059-014-0512-1.
- Zhang, Y.-C., Liao, J.-Y., Li, Z.-Y., Yu, Y., Zhang, J.-P., Li, Q.-F., et al. (2014c). Genome-wide screening and functional analysis identify a large number of long noncoding RNAs involved in the sexual reproduction of rice. *Genome Biol.* 15, 512. doi:10.1186/s13059-014-0512-1.
- Zhou, A., and Pawlowski, W. P. (2014a). Regulation of meiotic gene expression in plants. *Front. Plant Sci.* 5, 413. doi:10.3389/fpls.2014.00413.
- Zhou, A., and Pawlowski, W. P. (2014b). Regulation of meiotic gene expression in plants. *Front. Plant Sci.* 5, 413. doi:10.3389/fpls.2014.00413.
- Zhou, Z.-Y., Li, A.-M., Adeola, A. C., Liu, Y.-H., Irwin, D. M., Xie, H.-B., et al. (2014). Genome-wide identification of long intergenic noncoding RNA genes and their potential association with domestication in pigs. *Genome Biol. Evol.* 6, 1387–92. doi:10.1093/gbe/evu113.
- Zickler, D., and Kleckner, N. (1998). the Leptotene-Zygotene Transition of Meiosis. *Annu. Rev. Genet.* 32, 619–697. doi:10.1146/annurev.genet.32.1.619.
- Ziolkowski, P. A., Berchowitz, L. E., Lambing, C., Yelina, N. E., Zhao, X., Kelly, K. A., et al. (2015). Juxtaposition of heterozygosity and homozygosity during meiosis causes reciprocal crossover remodeling via interference. *Elife* 2015, 1–29. doi:10.7554/eLife.03708.

MATERIAL SUPLEMENTARIO

Tabla S1. Número de lecturas obtenidas (total), limpias y que mapearon a un solo solo transcrito. El porcentaje de secuencias limpias está calculado respecto al total de secuencias, mientras que el de mapeo esta calculado con el de secuencias limpias.

Genoteca	Total	Secuencias Limpias	Mapeo único
Somático	173,458,302	133,126,131 (76.75%)	100,607,075 (75.57%)
Meiocitos R1	175,789,421	139,227,278 (79.20%)	98,778,565 (70.95%)
Meiocitos R2	142,454,268	115,413,813 (81.02%)	79,611,917 (68.98%)
Total	491,701,991	387,767,222 (78.86%)	278,997,557 (71.94%)

Figura S1. Conservación de familia de ortólogos de genes meióticos no detectados en girasol. Los números representan los miembros de cada familia por especie.

		Green Plants																											
		Land Plants																				Chlorophyta							
		Vascular Plants																		Ostreococcus		Chlamydomonadales							
		Angiosperms																Ppa		Olu		Ota		M. sp.		Vca		Cre	
		Eudicots												Monocots															
		Fabids						Malvids						BEP Clade				PACCMADClade											
		N2 - Fixing Clade																											
		Papilionoideae			Rosaceae			Euphorbiaceae			Arabidopsis			Brassicales			Oryza												
Locus	Gen	Galegoids	Lja	Mtr	Gma	Mdo	Fve	Mes	Rco	Pir	Ath	Aly	Cpa	Tca	Vvi	Osa	Osain	Bdi	Sbi	Zma	Smo	Ppa	Olu	Ota	M. sp.	Vca	Cre		
AT1G14750	SDS	1		1	2	1	1	1	1	1	1	1	1	1	1	1	2	1	1	2									
AT1G22275	ZYP1b	1	1	2			5	1	1	1	2	2	1	1	1	1	1	1	2	1	1								
AT1G64030	ATSRP3										1	1																	
AT1G75950	ASK1	1	2	3	1	2	2	3	3	2	1	2	1	2	2	2	2	2	2	2	1	2				1	1		
AT1G77390	CYCA1	1	1	4	2	1	2	1	2	1	1	1	1	1	1	1	2	2	3		2	8				1	1		
AT2G14540	ATSRP2										1	1																	
AT2G27170	SMC3	2	1	2	2	1	1	1	1	1	1	1	1	1	2	3	1	2	1				1	1	1	1	1		
AT2G42890	AML2	1	1	4		2	2	2	2	3	3	2	2	3	2	2	2	1	2	1	2	1	3			1	2		
AT3G19590	BUB3.1	1	1	3	2	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1		1			1	1	1		
AT3G29350	AHP2	1	4	4	2		2	2	3	2	2	2	2	2				2	1				3	1	1	1	1		
AT3G47460	ATSMC2	2		2	4	3	1	1	2	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1		
AT3G57860	OSD1										1	1																	
AT4G00020	BRCA2A	1	2	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1			1	1	1	1	1		
AT4G05190	ATK5	3	4	8	5	2	2	2	4	4	4	2	2	2	2	3	2	2	1	2	2	2	1	1	1	1	1		
AT4G27170	SESA4										1	5																	
AT5G05490	SYN1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1								
AT5G07660	SMC6A	1	1	1	1	1	1	1	3	2	2	3	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1		
AT5G15920	SMC5	2	4	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1		
AT5G40840	SYN2	2	1		1	1	1	1	1	1	1	1	1	1															
AT5G51330	SWI1		1	2	3	1	1	1	2	1	1	1	1	2	1	1	1	1	1	1	2	3							
AT5G54260	MRE11	2	1	1	2	1	1	1	1	1	1	1	1	1	2	2	2	1	1	1	1	1	1	1	1	1	1		

Lja: Lotus japonicus, Mtr: Medicago truncatula, Gma: Glycine max, Mdo: Malus domestica, Fve: Fragaria vesca, Mes: Manihot esculenta, Rco: Ricinus communis, Pir: Populus trichocarpa, Ath: Arabidopsis thaliana, Aly: Arabidopsis lyrata, Cpa: Carica papaya, Tca: Theobroma cacao, Vvi: Vitis vinifera, Osa: Oryza sativa ssp. japonica, Osain: Oryza sativa ssp. indica, Bdi: Brachypodium distachyon, Sbi: Sorghum bicolor, Zma: Zea mays, Smo: Selaginella moellendorffii, Ppa: Physcomitrella patens, Olu: Ostreococcus lucimarinus, Ota: Ostreococcus tauri, M. sp.: Micromonas sp. RCC299, Vca: Volvox carteri, Cre: Chlamydomonas reinhardtii.

Tabla S2. Términos GO de la categoría CC con mayor expresión en meiocitos.

GO id	Término GO	Fold Change
GO:0000796	condensin complex	32.70
GO:0000808	origin recognition complex	29.18
GO:0000776	kinetochore	25.38
GO:0042555	MCM complex	23.06
GO:0048196	middle lamella-containing extracellular matrix	16.10
GO:0032300	mismatch repair complex	12.41
GO:0044459	plasma membrane part	12.20
GO:0005971	ribonucleoside-diphosphate reductase complex	10.54
GO:0000811	GIN5 complex	7.37
GO:0000786	nucleosome	6.17
GO:0005663	DNA replication factor C complex	5.68
GO:0009331	glycerol-3-phosphate dehydrogenase complex	4.13
GO:0005640	nuclear outer membrane	3.93
GO:0033597	mitotic checkpoint complex	3.66
GO:0005678	chromatin assembly complex	3.58
GO:0033588	Elongator holoenzyme complex	3.57
GO:0048500	signal recognition particle	3.56
GO:0030880	RNA polymerase complex	3.52
GO:0010317	pyrophosphate-dependent phosphofructokinase complex, alpha-subunit complex	3.41
GO:0005731	nucleolus organizer region	3.00
GO:0005732	small nucleolar ribonucleoprotein complex	2.67
GO:0005658	alpha DNA polymerase:primase complex	2.67
GO:0008180	signalosome	2.60
GO:0005787	signal peptidase complex	2.55
GO:0005844	polysome	2.54
GO:0005838	proteasome regulatory particle	2.52
GO:0015934	large ribosomal subunit	2.51
GO:0005742	mitochondrial outer membrane translocase complex	2.44
GO:0005839	proteasome core complex	2.40
GO:0016272	prefoldin complex	2.33
GO:0019866	organelle inner membrane	2.30
GO:0035145	exon-exon junction complex	2.25
GO:0005875	microtubule associated complex	2.23
GO:0005758	mitochondrial intermembrane space	2.19
GO:0016591	DNA-directed RNA polymerase II, holoenzyme	2.18
GO:0009330	DNA topoisomerase complex (ATP-hydrolyzing)	2.16
GO:0005840	ribosome	2.15
GO:0000276	mitochondrial proton-transporting ATP synthase complex, coupling factor F(o)	2.07
GO:0005852	eukaryotic translation initiation factor 3 complex	2.05
GO:0005667	transcription factor complex	2.00

Tabla S3. Términos GO de la categoría CC con mayor expresión en transcriptoma somático.

Goid	Término GO	Fold Change
GO:0070825	micropyle	187.39
GO:0009782	photosystem I antenna complex	55.52
GO:0031012	extracellular matrix	54.90
GO:0012506	vesicle membrane	47.64
GO:0043674	columella	27.78
GO:0009512	cytochrome b6f complex	19.69
GO:0016328	lateral plasma membrane	16.93
GO:0009523	photosystem II	15.54
GO:0031226	intrinsic to plasma membrane	15.25
GO:0009522	photosystem I	14.46
GO:0009538	photosystem I reaction center	13.65
GO:0009654	oxygen evolving complex	13.64
GO:0005955	calcineurin complex	11.82
GO:0030076	light-harvesting complex	10.88
GO:0030118	clathrin coat	10.62
GO:0010598	NAD(P)H dehydrogenase complex (plastoquinone)	10.31
GO:0009544	chloroplast ATP synthase complex	9.77
GO:0009344	nitrite reductase complex [NAD(P)H]	9.40
GO:0009930	longitudinal side of cell surface	8.44
GO:0009783	photosystem II antenna complex	6.88
GO:0009517	PSII associated light-harvesting complex II	6.37
GO:0005945	6-phosphofructokinase complex	5.59
GO:0005754	mitochondrial proton-transporting ATP synthase, catalytic core	5.47
GO:0042170	plastid membrane	5.24
GO:0045259	proton-transporting ATP synthase complex	4.80
GO:0019898	extrinsic to membrane	4.17
GO:0031519	PcG protein complex	4.00
GO:0009579	thylakoid	3.79
GO:0009925	basal plasma membrane	3.77
GO:0045177	apical part of cell	3.18
GO:0009539	photosystem II reaction center	3.08
GO:0009986	cell surface	3.02
GO:0031224	intrinsic to membrane	3.01
GO:0009329	acetate CoA-transferase complex	3.01
GO:0031897	Tic complex	2.97
GO:0009898	internal side of plasma membrane	2.85
GO:0042175	nuclear outer membrane-endoplasmic reticulum membrane network	2.85
GO:0005615	extracellular space	2.83
GO:0031227	intrinsic to endoplasmic reticulum membrane	2.69
GO:0005835	fatty acid synthase complex	2.68
GO:0005618	cell wall	2.61
GO:0071944	cell periphery	2.37
GO:0010007	magnesium chelatase complex	2.33
GO:0043233	organelle lumen	2.33
GO:0090406	pollen tube	2.23
GO:0005960	glycine cleavage complex	2.22
GO:0090404	pollen tube tip	2.12
GO:0005625	soluble fraction	2.08
GO:0000813	ESCRT I complex	2.07

Tabla S4. Términos GO de la categoría BP con mayor expresión en meiocitos.

GOid	Término GO	Fold Change
GO:0048658	tapetal layer development	133.97
GO:0010865	stipule development	108.00
GO:0045132	meiotic chromosome segregation	67.93
GO:0010069	zygote asymmetric cytokinesis in embryo sac	49.26
GO:0007108	cytokinesis, initiation of separation	35.38
GO:0051026	chiasma assembly	27.93
GO:0009856	pollination	26.72
GO:0030155	regulation of cell adhesion	24.58
GO:0051304	chromosome separation	17.09
GO:0045143	homologous chromosome segregation	15.75
GO:0007050	cell cycle arrest	15.07
GO:0048479	style development	14.08
GO:0000086	G2/M transition of mitotic cell cycle	14.05
GO:0048316	seed development	13.59
GO:0010529	negative regulation of transposition	13.55
GO:0010376	stomatal complex formation	11.46
GO:0010234	tapetal cell fate specification	11.18
GO:0051171	regulation of nitrogen compound metabolic process	9.85
GO:0019079	viral genome replication	9.62
GO:0007131	reciprocal meiotic recombination	7.50
GO:0010440	stomatal lineage progression	7.20
GO:0009755	hormone-mediated signaling pathway	6.97
GO:0010941	regulation of cell death	6.85
GO:0006311	meiotic gene conversion	6.27
GO:0010254	nectary development	6.08
GO:0009413	response to flooding	5.99
GO:0042138	meiotic DNA double-strand break formation	4.54
GO:0042754	negative regulation of circadian rhythm	4.21
GO:0002679	respiratory burst involved in defense response	4.08
GO:0022619	generative cell differentiation	4.06
GO:2000036	regulation of stem cell maintenance	4.06
GO:0010344	seed oilbody biogenesis	3.96
GO:0042127	regulation of cell proliferation	3.89
GO:0006072	glycerol-3-phosphate metabolic process	3.84
GO:0000819	sister chromatid segregation	3.78
GO:0010229	inflorescence development	3.64
GO:0048653	anther development	3.48
GO:0055047	generative cell mitosis	3.31
GO:0048281	inflorescence morphogenesis	3.29
GO:0009850	auxin metabolic process	3.26
GO:0044249	cellular biosynthetic process	3.26
GO:0010450	inflorescence meristem growth	3.24
GO:0009314	response to radiation	3.20
GO:0000706	meiotic DNA double-strand break processing	3.16
GO:0000082	G1/S transition of mitotic cell cycle	3.09
GO:0048638	regulation of developmental growth	3.01

GO:0080157	regulation of plant-type cell wall organization or biogenesis	2.70
GO:0010358	leaf shaping	2.68
GO:0000710	meiotic mismatch repair	2.68
GO:0032204	regulation of telomere maintenance	2.68
GO:0000910	cytokinesis	2.66
GO:0048513	organ development	2.64
GO:0016197	endosome transport	2.63
GO:0000917	barrier septum assembly	2.59
GO:2000026	regulation of multicellular organismal development	2.55
GO:0007530	sex determination	2.55
GO:0045787	positive regulation of cell cycle	2.53
GO:0009799	specification of symmetry	2.52
GO:0006342	chromatin silencing	2.51
GO:0008356	asymmetric cell division	2.50
GO:0006446	regulation of translational initiation	2.49
GO:0010220	positive regulation of vernalization response	2.44
GO:0010081	regulation of inflorescence meristem growth	2.39
GO:0010022	meristem determinacy	2.39
GO:0035266	meristem growth	2.31
GO:0019048	virus-host interaction	2.27
GO:0051726	regulation of cell cycle	2.27
GO:0010015	root morphogenesis	2.27
GO:0048827	phyllome development	2.25
GO:0048577	negative regulation of short-day photoperiodism, flowering	2.16
GO:0050792	regulation of viral reproduction	2.15
GO:0048235	pollen sperm cell differentiation	2.14
GO:0060548	negative regulation of cell death	2.07
GO:0009652	thigmotropism	2.04
GO:0048528	post-embryonic root development	2.03
GO:0006793	phosphorus metabolic process	2.02
GO:0006897	endocytosis	2.01
GO:0031047	gene silencing by RNA	2.01

Tabla S5. Términos GO de la categoría BP con mayor expresión en transcriptoma somático

GOid	Término GO	Fold Change
GO:0010647	positive regulation of cell communication	142.00
GO:0048838	release of seed from dormancy	138.69
GO:0080090	regulation of primary metabolic process	76.00
GO:0009835	ripening	74.73
GO:0010221	negative regulation of vernalization response	74.29
GO:0022603	regulation of anatomical structure morphogenesis	70.60
GO:0009606	tropism	47.81
GO:0015977	carbon fixation	30.69
GO:0010623	developmental programmed cell death	29.76
GO:0009830	cell wall modification involved in abscission	27.70
GO:0006066	alcohol metabolic process	19.28
GO:0048657	tapetal cell differentiation	17.84
GO:0080027	response to herbivore	17.47
GO:0006723	cuticle hydrocarbon biosynthetic process	14.50
GO:0045848	positive regulation of nitrogen utilization	14.00
GO:0034059	response to anoxia	13.15
GO:2000072	regulation of defense response to fungus, incompatible interaction	12.25
GO:0007584	response to nutrient	10.22
GO:0023014	signal transduction by phosphorylation	9.85
GO:0080086	stamen filament development	9.78
GO:0010241	ent-kaurene oxidation to kaurenoic acid	8.78
GO:0030397	membrane disassembly	8.56
GO:0060862	negative regulation of floral organ abscission	8.19
GO:0010942	positive regulation of cell death	8.19
GO:0006949	syncytium formation	7.96
GO:0009638	phototropism	7.91
GO:0010483	pollen tube reception	7.88
GO:0051211	anisotropic cell growth	7.77
GO:0080170	hydrogen peroxide transmembrane transport	7.67
GO:0050891	multicellular organismal water homeostasis	7.50
GO:2000068	regulation of defense response to insect	6.77
GO:0051090	regulation of sequence-specific DNA binding transcription factor activity	6.77
GO:0017145	stem cell division	6.35
GO:0015833	peptide transport	6.22
GO:0009629	response to gravity	5.99
GO:0010274	hydrotropism	5.94
GO:0048359	mucilage metabolic process involved seed coat development	5.87
GO:0048544	recognition of pollen	5.65
GO:0015931	nucleobase-containing compound transport	5.53
GO:2000652	regulation of secondary cell wall biogenesis	5.48
GO:0015979	photosynthesis	5.45
GO:0001666	response to hypoxia	5.39
GO:0045926	negative regulation of growth	5.39
GO:0019419	sulfate reduction	5.26

GO:1900056	negative regulation of leaf senescence	4.97
GO:0048444	floral organ morphogenesis	4.83
GO:0080126	ovary septum development	4.72
GO:0051607	defense response to virus	4.58
GO:0010618	aerenchyma formation	4.41
GO:0009864	induced systemic resistance, jasmonic acid mediated signaling pathway	4.40
GO:0009414	response to water deprivation	4.33
GO:0010260	organ senescence	4.11
GO:0009866	induced systemic resistance, ethylene mediated signaling pathway	3.95
GO:0009268	response to pH	3.68
GO:0006855	drug transmembrane transport	3.64
GO:0010256	endomembrane system organization	3.58
GO:0048480	stigma development	3.45
GO:0031349	positive regulation of defense response	3.43
GO:0009612	response to mechanical stimulus	3.26
GO:0009620	response to fungus	3.18
GO:0008610	lipid biosynthetic process	3.17
GO:0006792	regulation of sulfur utilization	3.05
GO:0009625	response to insect	3.02
GO:0007389	pattern specification process	2.96
GO:0072593	reactive oxygen species metabolic process	2.95
GO:0046283	anthocyanin metabolic process	2.86
GO:0080167	response to karrikin	2.77
GO:0009624	response to nematode	2.67
GO:0060148	positive regulation of posttranscriptional gene silencing	2.64
GO:0019538	protein metabolic process	2.60
GO:0005976	polysaccharide metabolic process	2.60
GO:0009698	phenylpropanoid metabolic process	2.60
GO:0045165	cell fate commitment	2.51
GO:0006644	phospholipid metabolic process	2.49
GO:0030036	actin cytoskeleton organization	2.44
GO:0001887	selenium compound metabolic process	2.43
GO:0006730	one-carbon metabolic process	2.43
GO:0009636	response to toxin	2.40
GO:0010017	red or far-red light signaling pathway	2.32
GO:0009825	multidimensional cell growth	2.32
GO:0010617	circadian regulation of calcium ion oscillation	2.32
GO:0051410	detoxification of nitrogen compound	2.30
GO:0048317	seed morphogenesis	2.28
GO:0009611	response to wounding	2.27
GO:0051707	response to other organism	2.26
GO:0016998	cell wall macromolecule catabolic process	2.26
GO:0010112	regulation of systemic acquired resistance	2.26
GO:0032922	circadian regulation of gene expression	2.25
GO:0032940	secretion by cell	2.24
GO:0009308	amine metabolic process	2.20

GO:0007267	cell-cell signaling	2.15
GO:0009653	anatomical structure morphogenesis	2.14
GO:0007009	plasma membrane organization	2.13
GO:0006629	lipid metabolic process	2.12
GO:0006952	defense response	2.12
GO:0008104	protein localization	2.10
GO:0048833	specification of floral organ number	2.02
GO:0009266	response to temperature stimulus	2.01

Tabla S6. Términos GO de la categoría MF con mayor expresión en meiocitos

Goid	Término GO	Fold Change
GO:0050662	coenzyme binding	6.01
GO:0016018	cyclosporin A binding	4.15
	polygalacturonase inhibitor	
GO:0090353	activity	4.13
GO:0016830	carbon-carbon lyase activity	3.73
GO:0042393	histone binding	3.04
GO:0030332	cyclin binding	2.98
GO:0031491	nucleosome binding	2.60
GO:0001671	ATPase activator activity	2.10

Tabla S7. Términos GO de la categoría MF con mayor expresión en transcriptoma somático

Goid	Términos GO	Fold Change
GO:0051738	xanthophyll binding	279.04
GO:0051213	dioxygenase activity	180.30
GO:0008379	thioredoxin peroxidase activity	106.29
GO:0080015	sabinene synthase activity	97.62
GO:0046863	ribulose-1,5-bisphosphate carboxylase/oxygenase activator activity	91.16
GO:0010313	phytochrome binding	30.59
GO:0080016	(-)-E-beta-caryophyllene synthase activity	18.31
GO:0080017	alpha-humulene synthase activity	18.31
GO:0010242	oxygen evolving activity	14.63
GO:0045435	lycopene epsilon cyclase activity	12.12
GO:0016168	chlorophyll binding	11.52
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	9.77

GO:0000155	two-component sensor activity	9.60
GO:0044183	protein binding involved in protein folding	7.50
GO:0051740	ethylene binding	7.13
GO:0000170	sphingosine hydroxylase activity	6.96
GO:2001147	camalexin binding	6.43
GO:2001227	quercitrin binding	6.43
GO:0016798	hydrolase activity, acting on glycosyl bonds	6.37
GO:0008559	xenobiotic-transporting ATPase activity	6.35
GO:0004602	glutathione peroxidase activity	6.23
GO:0015197	peptide transporter activity	6.22
GO:0030674	protein binding, bridging	5.90
GO:0004096	catalase activity	5.49
GO:0008517	folic acid transporter activity	5.34
GO:0043495	protein anchor	5.32
GO:0000900	translation repressor activity, nucleic acid binding	5.04
GO:0008428	ribonuclease inhibitor activity	4.83
GO:0005529	sugar binding	4.54
GO:0051861	glycolipid binding	4.33
GO:0045436	lycopene beta cyclase activity	4.30
GO:0045550	geranylgeranyl reductase activity	4.05
GO:0046422	violaxanthin de-epoxidase activity	3.97
GO:0005344	oxygen transporter activity	3.78
GO:0008398	sterol 14-demethylase activity	3.75
GO:0046423	allene-oxide cyclase activity	3.53
GO:0015238	drug transmembrane transporter activity	3.50
GO:0005522	profilin binding	3.34
GO:0046910	pectinesterase inhibitor activity	3.27
GO:0071992	phytochelatin transmembrane transporter activity	2.91
GO:0031210	phosphatidylcholine binding	2.90
GO:0004134	4-alpha-glucanotransferase activity	2.87
GO:0016866	intramolecular transferase activity	2.79
GO:0017077	oxidative phosphorylation uncoupler activity	2.67
GO:0016846	carbon-sulfur lyase activity	2.66
GO:0080132	fatty acid alpha-hydroxylase activity	2.53
GO:0005496	steroid binding	2.50
GO:0042802	identical protein binding	2.40
GO:0016667	oxidoreductase activity, acting on a sulfur group of donors	2.35
GO:0010331	gibberellin binding	2.35
GO:0004888	transmembrane signaling receptor activity	2.33
GO:0016788	hydrolase activity, acting on ester bonds	2.24
GO:0030247	polysaccharide binding	2.20
GO:0016757	transferase activity, transferring glycosyl groups	2.15
GO:0003876	AMP deaminase activity	2.11
GO:0005516	calmodulin binding	2.08
GO:0005543	phospholipid binding	2.06
GO:0031418	L-ascorbic acid binding	2.05

Tabla S8. Rutas metabólicas de Aracyc con mayor expresión en meiocitos.

Ruta	Fold Change
sporopollenin precursor biosynthesis	360.44
anthocyanin biosynthesis (pelargonidin 3-O-glucoside, cyanidin 3-O-glucoside)	42.61
superpathway of anthocyanin biosynthesis (from cyanidin and cyanidin 3-O-glucoside)	39.90
IAA degradation IV	12.70
indole-3-acetyl-amino acid biosynthesis	12.70
alpha-amyrin biosynthesis	11.46
proanthocyanidin biosynthesis from flavanols	10.17
lysine degradation I	9.62
galactose degradation I (Leloir pathway)	5.07
pyrimidine deoxyribonucleotides de novo biosynthesis	3.57
glycerol-3-phosphate shuttle	2.99
cutin biosynthesis	2.77
guanine and guanosine salvage II	2.66
ubiquinone-9 biosynthesis (eukaryotic)	2.44
PRPP biosynthesis I	2.34
1,4-dihydroxy-2-naphthoate biosynthesis II (plants)	2.32
glutamate degradation I	2.24
hydroxycinnamic acid tyramine amides biosynthesis	2.10
methionine degradation II	2.09
oxidative ethanol degradation I	2.08
alanine biosynthesis III	2.06
ferulate and sinapate biosynthesis	2.02

Tabla S9. Rutas metabólicas de Aracyc con mayor expresión en transcriptoma somático.

Ruta	Fold Change
lanosterol biosynthesis	89.20
nicotianamine biosynthesis	84.67
fatty acid alpha-oxidation	73.36
xanthophyll cycle	68.70
antheraxanthin and violaxanthin biosynthesis	68.70
arginine degradation VI (arginase 2 pathway)	64.82
camalexin biosynthesis	64.00
farnesene biosynthesis	55.50
pyruvate fermentation to ethanol II	47.76
acetaldehyde biosynthesis I	47.76
sinapate ester biosynthesis	40.77
ent-kaurene biosynthesis	28.93
zeaxanthin biosynthesis	26.10
methylquercetin biosynthesis	26.01
homogalacturonan degradation	25.27
nitrate reduction II (assimilatory)	23.52
gibberellin inactivation	18.41

beta-caryophyllene biosynthesis	18.31
phaseic acid biosynthesis	17.38
abscisic acid glucose ester biosynthesis	16.32
asparagine biosynthesis I	16.13
photosynthesis light reactions	15.89
UDP-D-galacturonate biosynthesis I (from UDP-D-glucuronate)	14.70
2,4,6-trinitrotoluene degradation	14.55
(E,E)-4,8,12-trimethyltrideca-1,3,7,11-tetraene biosynthesis	14.34
(3E)-4,8-dimethylnona-1,3,7-triene biosynthesis	14.12
ethanol degradation	14.10
lutein biosynthesis	14.00
monoterpene biosynthesis	13.83
UDP-glucose biosynthesis (from sucrose)	12.97
delta-carotene biosynthesis	12.12
monolignol glucosides biosynthesis	11.35
coniferin metabolism	11.35
salicylate glucosides biosynthesis IV	10.81
putrescine biosynthesis IV	10.58
putrescine biosynthesis I	10.58
cytokinins-O-glucoside biosynthesis	9.74
hydroxyjasmonate sulfate biosynthesis	9.65
indole glucosinolate breakdown (active in intact plant cell)	9.38
melibiose degradation	8.86
GA12 biosynthesis	8.79
glutamate degradation IV	8.71
phytol salvage pathway	8.43
phospholipid desaturation	8.43
glucosinolate biosynthesis from phenylalanine	8.22
glucosinolate biosynthesis from tryptophan	8.15
13-LOX and 13-HPL pathway	8.12
trans-lycopene biosynthesis II (plants)	8.09
glucosinolate biosynthesis from pentahomomethionine	7.75
glucosinolate biosynthesis from hexahomomethionine	7.75
poly-hydroxy fatty acids biosynthesis	7.66
glutamate biosynthesis IV	7.66
ajugose biosynthesis II (galactinol-independent)	7.57
galactosylcyclitol biosynthesis	7.57
thiamine biosynthesis II	7.41
quercetinsulphates biosynthesis	7.29
methyl indole-3-acetate interconversion	7.05
citrulline biosynthesis	6.90
cyanide degradation	6.53
cytokinins degradation	6.51
sorbitol degradation I	6.42
ammonia assimilation cycle II	6.33
stachyose biosynthesis	6.21

sulfate activation for sulfonation	5.40
ribose degradation	5.33
serine biosynthesis	5.32
oxidized GTP and dGTP detoxification	5.31
glucosinolate biosynthesis from tetrahomomethionine	5.23
glucosinolate biosynthesis from trihomomethionine	5.23
glucosinolate biosynthesis from homomethionine	5.23
Fe(III)-reduction and Fe(II) transport	5.10
gibberellin biosynthesis I (non C-3, non C-13 hydroxylation)	5.06
gibberellin biosynthesis II (early C-3 hydroxylation)	5.06
gibberellin biosynthesis III (early C-13 hydroxylation)	5.06
alanine degradation III	5.04
alanine biosynthesis II	5.04
ajugose biosynthesis I (galactinol-dependent)	4.96
vitamin E biosynthesis	4.92
nonaprenyl diphosphate biosynthesis II	4.85
linoleate biosynthesis I (plants)	4.79
sulfate reduction II (assimilatory)	4.76
flavonoid biosynthesis	4.73
phenylpropanoid biosynthesis, initial reactions	4.65
trans-cinnamoyl-CoA biosynthesis	4.63
benzoate biosynthesis II (CoA-independent, non- β -oxidative)	4.63
cellulose biosynthesis	4.46
photorespiration	4.43
ethylene biosynthesis from methionine	4.31
beta-carotene biosynthesis	4.30
urea cycle	4.18
cuticular wax biosynthesis	4.11
phytyl diphosphate biosynthesis	4.05
uracil degradation (reductive)	3.86
choline biosynthesis I	3.80
pentose phosphate pathway (non-oxidative branch)	3.78
proline biosynthesis III	3.74
free phenylpropanoid acid biosynthesis	3.70
phenylethanol biosynthesis	3.66
nonaprenyl diphosphate biosynthesis III	3.66
geranylgeranyldiphosphate biosynthesis	3.66
methylerythritol phosphate pathway	3.60
chlorophyll a degradation	3.54
mannose degradation	3.49
aldehyde oxidation I	3.39
chlorophyll a biosynthesis II	3.39
glucosinolate biosynthesis from dihomomethionine	3.36
glycolipid desaturation	3.35
trans-zeatin biosynthesis	3.30
simple coumarins biosynthesis	3.24

Calvin-Benson-Bassham cycle	3.10
UDP-L-rhamnose biosynthesis	3.03
chlorophyllide a biosynthesis I	3.02
phospholipases	2.99
cytokinins 7-N-glucoside biosynthesis	2.93
cytokinins 9-N-glucoside biosynthesis	2.93
glycine betaine biosynthesis III (plants)	2.92
triacylglycerol degradation	2.90
jasmonic acid biosynthesis	2.89
UDP-L-arabinose biosynthesis I (from UDP-xylose)	2.88
siroheme biosynthesis	2.87
homomethionine biosynthesis	2.83
choline biosynthesis III	2.83
lupeol biosynthesis	2.81
S-adenosylmethionine biosynthesis	2.76
pentose phosphate pathway (oxidative branch)	2.67
pyridine nucleotide cycling (plants)	2.64
cyanide detoxification	2.54
folate polyglutamylation II	2.52
UDP-D-xylose biosynthesis	2.51
farnesylcysteine salvage pathway	2.51
cysteine biosynthesis I	2.47
oleate biosynthesis I (plants)	2.47
S-adenosyl-L-methionine cycle II	2.46
aliphatic glucosinolate biosynthesis, side chain elongation cycle	2.43
brassinosteroid biosynthesis III	2.41
sulfite oxidation IV	2.41
glutathione degradation	2.40
sucrose biosynthesis	2.38
xyloglucan biosynthesis	2.36
4-hydroxyphenylpyruvate biosynthesis	2.35
chlorophyll cycle	2.31
plastoquinone-9 biosynthesis	2.28
GDP-L-fucose biosynthesis II (from L-fucose)	2.26
beta-alanine biosynthesis I	2.25
leucine degradation I	2.24
quercetin glucoside biosynthesis (Arabidopsis)	2.23
sphingolipid biosynthesis (plants)	2.22
pyruvate fermentation to lactate	2.18
very long chain fatty acid biosynthesis	2.13
dTDP-L-rhamnose biosynthesis II	2.12
gluconeogenesis	2.12
sucrose degradation III	2.12
1D-myo-inositol hexakisphosphate biosynthesis III (Spirodela polyrrhiza)	2.10
glycolysis I (plastidic)	2.10

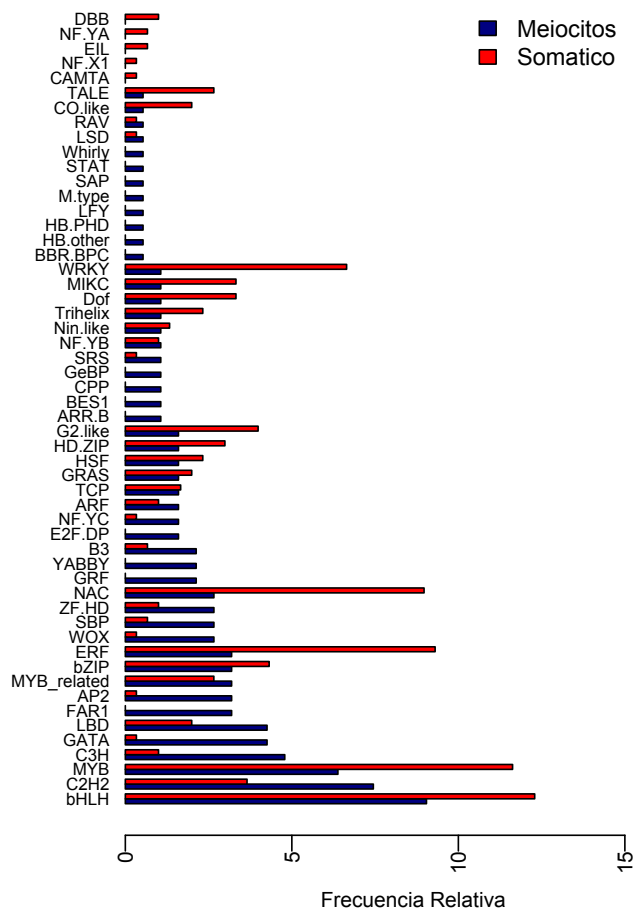


Figura S1. Familias de factores de transcripción expresadas en meiocitos y el transcriptoma somático.

Tabla S10. Ortólogos de factores de transcripción con expresión exclusiva en meiocitos

Locus	TPM Meiocitos	Familia TF
AT2G16910	214.62	bHLH
AT1G18960	97.90	MYB_related
AT2G17950	9.52	WOX
AT3G12680	6.22	C3H
AT1G75430	4.25	TALE
AT3G04070	2.38	NAC
AT5G53210	2.00	bHLH
AT2G45410	1.89	LBD
AT1G06280	1.33	LBD
AT1G18790	1.11	Nin-like
AT3G13890	0.58	MYB
AT2G28810	0.42	Dof
AT3G18010	0.38	WOX
AT1G51120	0.17	RAV

Tabla S11. Genes de Arabidopsis relacionados con silenciamiento, incluidos en el presente estudio

Gen	Locus	Descripción
AGO01	AT1G48410	RNA slicer/core component of plant RISC
AGO02	AT1G31280	RNA slicer/involved in antiviral defense response
AGO03	AT1G31290	RNA slicer
AGO04	AT2G27040	RNA slicer/involved in the establishment phase of RdDM
AGO05	AT2G27880	RNA slicer
AGO06	AT2G32940	RNA slicer/rasiRNA-directed heterochromatin formation
AGO07	AT1G69440	RNA slicer/tasiRNA biogenesis and juvenile-to-adult transition
AGO08	AT5G21030	RNA slicer
AGO09	AT5G21150	RNA slicer/silencing the activity of TEs activity in the female gametophyte
AGO10	AT5G43810	RNA slicer
CMT3	AT1G69770	Methyltransferase/maintenance phase of RdDM
DCL1	AT1G01040	RNase III/miRNA, natsiRNA, and tasiRNA biogenesis
DCL2	AT3G03300	RNase III/natsiRNA biogenesis and viral defense
DCL3	AT3G43920	RNase III/rasiRNA biogenesis and establishment phase of RdDM
DCL4	AT5G20320	RNase III/tasiRNA biogenesis and viral defense
DRM2	AT3G17310	Methyltransferase/establishment phase of RdDM
HYL1	AT1G09700	dsRBP/miRNA and tasiRNA biogenesis
DRB4	AT3G62800	dsRBP/miRNA and tasiRNA biogenesis
DRD1	AT2G16390	SNF2-like chromatin-remodeling factor/establishment phase of RdDM
HEN1	AT4G29160	sRNA-specific methyltransferase/sRNA biogenesis
HST	AT3G05040	Exportin-5 ortholog/miRNA exportation from nucleus
MET1	AT5G49160	Methyltransferase/maintenance phase of RdDM methylation
NRPD1a	AT1G63020	DNA-dependent RNA polymerase/establishment phase of RdDM
NRPD1b	AT2G40030	DNA-dependent RNA polymerase/establishment phase of RdDM
NRPD2	AT3G23780	DNA-dependent RNA polymerase/establishment phase of RdDM
RDR1	AT1G14790	RNA-dependent RNA polymerase/viral defense
RDR2	AT4G11130	RNA-dependent RNA polymerase/rasiRNA biogenesis
RDR6	AT3G49500	RNA-dependent RNA polymerase/tasiRNA and natsiRNA biogenesis
SGS3	AT5G23570	Coiled-coil protein/tasiRNA and natsiRNA biogenesis
SDE3	AT1G05460	Silencing defective/Encodes a protein with similarity to RNA helicases
SDE5	AT3G15390	Similar to PRL1 interacting factor and is involved in virus induced silencing
DDM1	AT5G66750	Similar to SWI2/SNF2 chromatin remodeling proteins
DRD2	AT3G23780	Chromatin-remodeling factor
DMS3	AT3G49250	Involving secondary siRNA and spreading of DNA methylation
DMS4	AT2G30280	Transcriptional regulator functioning in RNA-directed DNA methylation and plant development
HDA6	AT5G63110	RPD3-like histone deacetylase/RNA-Mediated transcriptional silencing
MET1	AT5G49160	Encodes a cytosine methyltransferase
MOM1	AT1G08060	Encodes a transcriptional silencer/maintenance of methylation

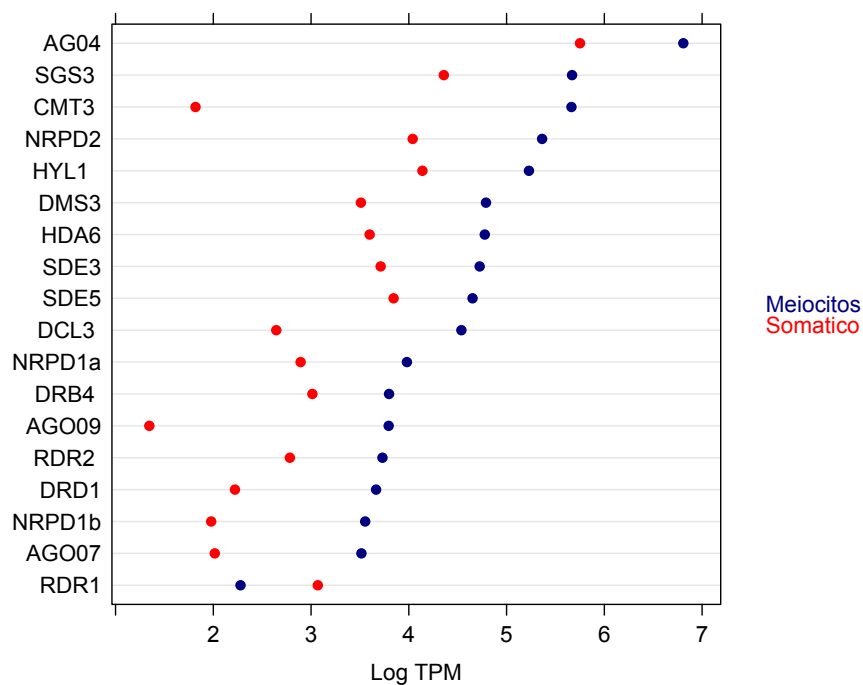


Figura S3. Niveles de expresión de genes relacionados con silenciamiento.

Tabla S12. Número total de lecturas y con mapeo único obtenidas para los tres genotipos de girasol.

Genoteca	Total	Lecturas mapeadas
Meiocitos		
Domesticado R1	179,789,421	139,532,735 (77.61%)
Domesticado R2	142,454,268	112,861,139 (79.23%)
F1 R1	189,180,804	151,174,022 (79.91%)
F1 R2	93,914,482	74,366,257 (79.19%)
Silvestre R1	123,666,567	94,260,868 (76.22%)
Silvestre R2	132,921,292	100,034,300 (75.26%)
Total	861,926,834	672,229,321 (77.99%)
Transcriptoma Somático		
Somático	173,458,030	128,984,174 (74.36%)

Tabla S13. Número total de lecturas de RNA pequeños obtenidas para los genotipos silvestre y domesticado.

Longitud	Silvestre	Domesticado
20	87,931 (1.66%)	164,495 (3.11%)
21	452,676 (8.57%)	678,774 (12.85%)
22	408,185 (7.73%)	420,184 (7.96%)
23	729,207 (13.81%)	803,569 (15.22%)
24	3,425,818 (64.87%)	3,042,461 (57.61%)
25	177,383 (3.36%)	171,924 (3.26%)
Total	5,281,200	5,281,407

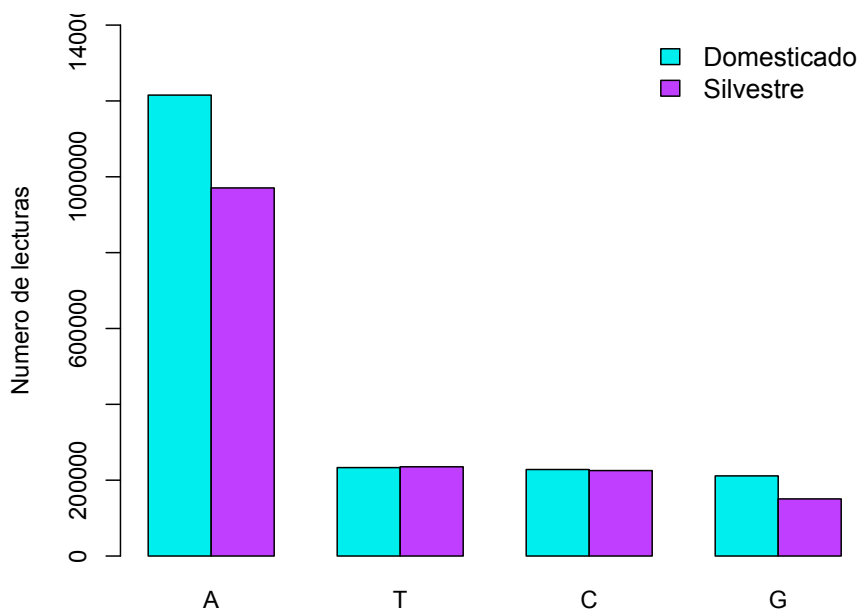


Figura S4. Nucleótido del extremo 5' de las secuencias de 24 nt.

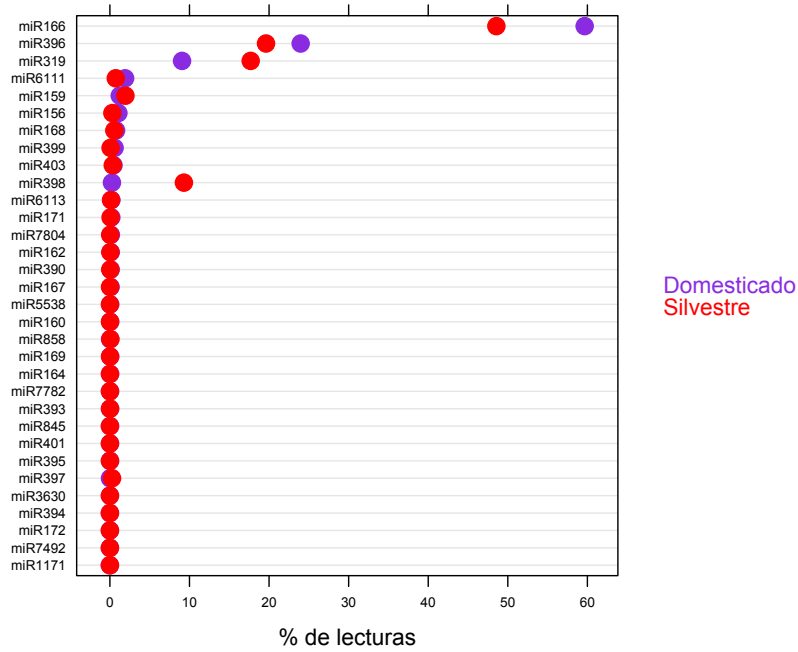


Figura S5. Proporción de lecturas encontradas para cada una de las familias de miRNAs encontradas en los genotipos silvestre y domesticado.

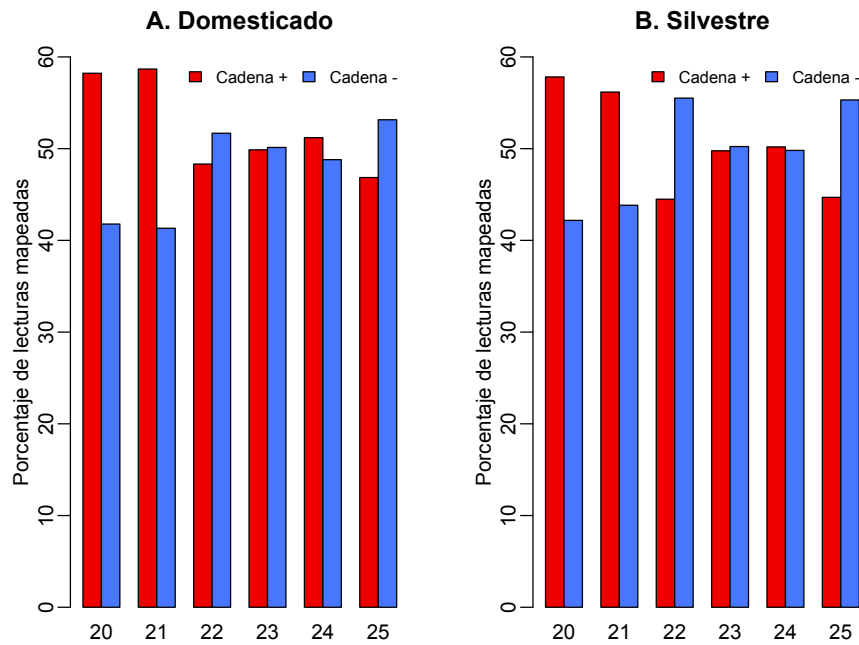


Figura S6. Proporción de lecturas de sRNAs que mapean a cada una de las cadenas para el genotipo domesticado (A) y silvestre (B).

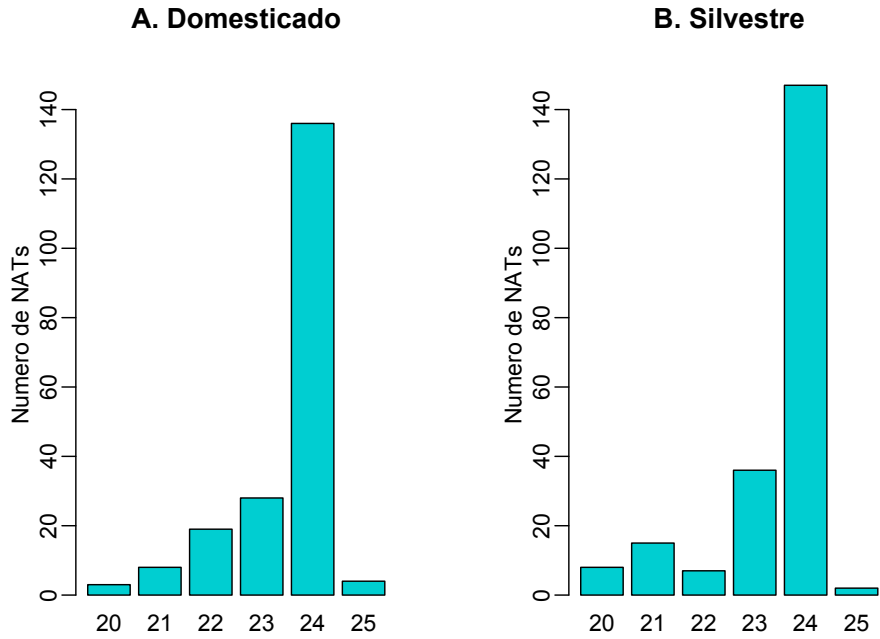


Figura S7. Proporción de lecturas de sRNAs que mapean a transcritos NATs identificados.

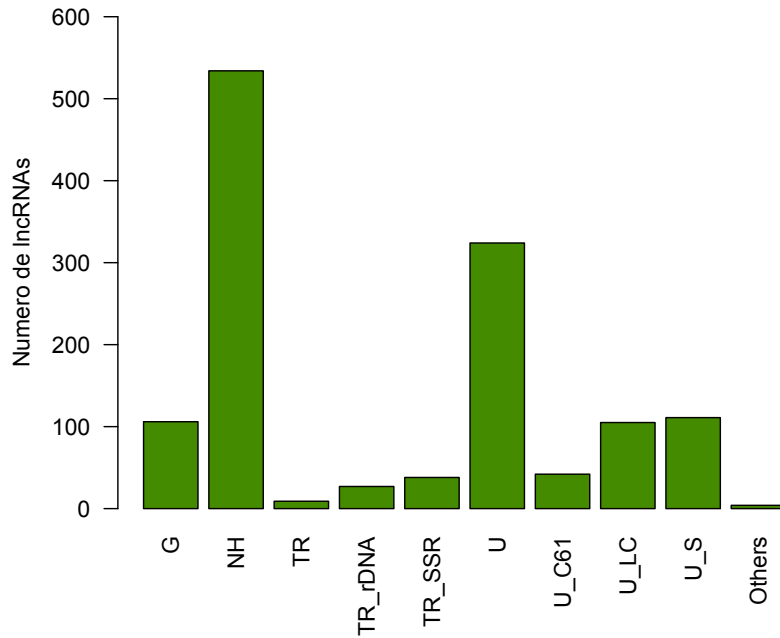


Figura S8. Número de otros elementos repetitivos (no-TEs) encontrados en los lincRNAs de girasol.

Tabla S14. Listado de genes meióticos incluidos. Se presenta la categoría funcional, el número de transcritos por millón (TPM), para cada uno de los genotipos, el q-value para los diferentes contrastes y el patrón de expresión.

Locus	Gen	Categoría Funcional	TPM D	TPM F1	TPM S	TPM Som	D-Sqv	D-F1qv	S-F1qv	Patrón
AT1G67370	ASY1	1	394.180	339.830	173.850	7.668	0.002	0.768	0.012	1
AT2G46980	ASY3	1	21.220	26.330	25.620	0.109	0.257	0.576	0.377	5
AT1G22260	ZYP1	1	132.480	193.130	91.300	2.900	0.081	0.255	0.002	6
AT5G65460	CDKA;1	2	54.610	49.250	22.070	66.946	0.000	0.769	0.002	1
AT4G20900	MS5	2	12.600	16.650	84.180	0.853	0.000	0.681	0.000	4
AT5G19400	SMG7	2	134.490	137.690	108.540	98.198	0.202	0.876	0.216	6
AT1G06660	JASON	2	66.560	72.300	53.650	0.248	0.217	0.741	0.114	6
AT1G34355	PS1	2	42.260	54.420	86.110	0.977	0.003	0.586	0.044	4
AT4G21270	ATK1	2	115.410	106.340	32.960	2.969	0.000	0.789	0.000	1
AT1G66170	MMD	2	33.580	55.310	182.220	0.023	0.000	0.465	0.000	4
AT4G01370	MPK4	2	30.290	21.190	21.900	117.588	0.104	0.357	0.430	2
AT3G43210	TES	2	36.150	46.010	96.820	1.837	0.000	0.674	0.005	4
AT2G33560	BUBR1	2	37.770	33.910	44.430	1.961	0.258	0.747	0.146	3
AT3G25980	MAD2	2	51.760	43.190	53.040	0.643	0.431	0.664	0.230	3
AT4G05190	ATK5	2	11.720	6.170	2.730	18.374	0.001	0.299	0.054	1
AT5G42190	ASK2	2	832.330	707.950	844.690	94.578	0.437	0.705	0.262	3
AT5G21150	AGO9	3	43.230	36.140	99.800	0.000	0.001	0.614	0.000	3
AT5G51330	SWI1	3	27.870	37.820	68.930	2.458	0.000	0.460	0.017	4
AT1G77320	MEI1	3	14.120	16.130	13.740	4.977	0.425	0.787	0.315	6
AT3G25100	CDC45	3	51.230	61.290	26.750	2.186	0.008	0.687	0.002	6
AT5G48720	XRI1	3	31.870	32.930	25.040	6.264	0.166	0.873	0.158	6
AT5G61960	AML1	3	116.630	115.240	111.020	1.861	0.397	0.865	0.444	1
AT4G18120	AML3	3	0.090	0.120	0.330	0.000	0.002	0.703	0.015	4
AT5G07290	AML4	3	45.860	52.180	45.440	5.636	0.439	0.768	0.315	6
AT1G29400	AML5	3	36.630	38.430	44.100	4.411	0.243	0.876	0.260	4
AT1G64625	FST	3	14.430	7.980	4.320	1.760	0.000	0.154	0.026	1
AT3G27120	FIGL1	4	23.130	26.080	13.560	1.551	0.022	0.786	0.011	6
AT5G64520	XRCC2	4	8.230	8.440	19.950	4.551	0.000	0.876	0.001	4
AT5G63540	BLAP75/RMI1	4	12.960	12.900	6.090	1.194	0.002	0.876	0.004	1
AT5G63920	Top3_	4	19.630	22.160	31.610	0.954	0.033	0.790	0.087	4
AT1G10930	RECQ4A	4	34.940	35.430	29.750	0.605	0.276	0.875	0.322	6
AT1G35530	FANCM	4	11.200	12.560	5.100	5.977	0.003	0.836	0.002	6
AT5G50930	MHF1	4	51.850	51.810	56.440	0.721	0.362	0.860	0.418	3
AT1G78790	MHF2	4	12.230	10.920	17.770	7.730	0.082	0.806	0.052	3
AT3G18524	MSH2	4	34.890	33.740	26.270	6.280	0.151	0.822	0.239	1
AT4G17380	MSH4	5	18.180	22.480	26.350	3.497	0.071	0.576	0.289	4
AT3G20475	MSH5	5	32.540	51.950	124.460	1.109	0.000	0.198	0.004	4
AT3G24495	MSH7	5	33.000	37.370	18.470	1.016	0.015	0.770	0.007	6
AT1G10970	ZIP4	5	26.090	53.050	71.370	8.629	0.000	0.012	0.148	4
AT3G27730	MER3	5	28.940	31.140	16.410	1.675	0.012	0.804	0.007	6
At1g12790	PTD	5	12.530	16.700	19.660	2.535	0.053	0.433	0.332	4
AT5G52290	SHOC1	5	11.300	17.220	18.440	0.977	0.029	0.198	0.407	4
AT1G53490	HEI10	5	37.580	54.320	86.990	0.760	0.001	0.296	0.061	4
AT4G09140	MLH1	5	22.850	29.300	18.360	12.033	0.230	0.650	0.086	6

AT3G13170	SPO11-1	8	8.260	7.110	5.120	0.775	0.055	0.787	0.129	1
AT1G63990	SPO11-2	8	31.830	43.880	73.030	0.923	0.001	0.370	0.046	4
AT4G14180	PRD1	8	9.710	12.180	8.590	0.124	0.309	0.537	0.089	6
AT5G57880	PRD2	8	11.930	16.080	11.000	4.489	0.350	0.424	0.075	6
AT1G01690	PRD3	8	21.010	28.960	27.740	2.923	0.145	0.359	0.372	5
AT1G07060	DFO	8	14.600	22.010	19.360	1.566	0.135	0.254	0.316	5
AT1G10710	PHS1	8	37.540	61.700	106.890	2.217	0.000	0.123	0.060	4
AT4G24710	Pch2	8	45.370	50.510	43.370	3.093	0.399	0.777	0.284	6
AT5G54260	MRE11	9	25.920	29.970	21.510	21.902	0.244	0.779	0.155	6
AT2G31970	RAD50	9	58.090	77.250	56.410	32.415	0.418	0.460	0.120	6
AT3G02680	NBS1	9	12.030	11.940	13.310	11.467	0.326	0.876	0.330	3
AT3G52115	ATCOM1	9	10.740	10.120	5.780	3.303	0.010	0.829	0.026	1
AT2G06510	RPA1A	9	58.040	73.630	47.640	11.645	0.222	0.538	0.052	6
AT5G20850	RAD51	9	29.050	43.370	132.670	0.915	0.000	0.376	0.000	4
AT2G28560	RAD51B	9	7.100	4.830	4.280	1.341	0.041	0.432	0.296	1
AT2G45280	RAD51C	9	6.990	9.360	10.570	2.101	0.077	0.606	0.284	4
AT1G07745	RAD51D	9	5.670	6.620	4.820	0.892	0.264	0.693	0.121	6
AT5G01630	BRCA2(V)	9	14.190	15.970	9.260	3.373	0.059	0.806	0.037	6
AT5G57450	XRCC3	9	2.630	2.470	0.870	0.186	0.000	0.850	0.000	1
AT3G48190	ATM	9	43.860	46.750	18.880	3.155	0.003	0.876	0.004	6
AT5G40820	ATR	9	5.540	8.200	7.550	0.403	0.127	0.393	0.418	5
AT3G09660	MCM8	9	12.510	19.810	21.100	1.140	0.025	0.200	0.389	4
AT3G22880	DMC1	10	74.800	125.760	119.840	0.752	0.052	0.113	0.344	5
AT1G13330	AHP2	10	66.760	75.350	39.050	1.675	0.034	0.748	0.014	6
AT4G29170	MND1	10	41.620	52.000	85.850	10.893	0.003	0.537	0.051	4
AT1G14750	SDS	10	36.250	72.550	62.690	3.404	0.022	0.020	0.255	5
AT5G05490	SYN1	11	38.500	43.030	14.740	0.426	0.001	0.751	0.000	6
AT3G59550	SYN3	11	23.390	28.200	29.170	0.504	0.188	0.662	0.410	4
AT2G47980	SCC3	11	121.960	126.390	55.320	63.892	0.001	0.862	0.001	6
AT5G15540	SCC2	11	62.620	65.890	31.840	27.941	0.010	0.876	0.016	6
AT5G04320	SGO2	11	26.070	28.100	18.630	0.411	0.091	0.831	0.068	6
AT4G22970	AESP	11	19.940	23.070	6.280	1.822	0.000	0.788	0.000	6
AT3G23890	TOP2A	11	274.500	241.140	118.520	0.558	0.002	0.689	0.018	1
AT2G27170	SMC3	12	204.890	196.140	136.010	92.314	0.058	0.832	0.107	1
AT5G62410	SMC4	12	135.150	135.790	54.440	0.000	0.001	0.876	0.002	6
AT3G54670	SMC1	12	115.560	108.670	50.380	74.118	0.002	0.801	0.008	1
AT5G15920	SMC5	12	71.920	73.170	29.360	8.660	0.001	0.876	0.001	6
AT5G61460	SMC6B	12	57.060	59.130	48.050	16.708	0.253	0.876	0.275	6
AT5G21140	NSE1	12	28.260	25.050	17.160	2.287	0.028	0.772	0.083	1
AT3G15150	NSE2	12	9.070	9.320	15.790	3.202	0.016	0.860	0.030	4
AT1G51130	NSE4a	12	46.430	46.210	27.330	0.000	0.029	0.876	0.037	1
AT1G34770	NSE3	12	27.860	46.440	86.700	12.157	0.000	0.239	0.012	4
AT1G63770	MPA1	13	190.760	193.360	122.360	5.024	0.052	0.876	0.065	6
AT2G02560	HVE	13	104.600	125.150	79.000	84.119	0.143	0.682	0.052	6
AT5G46210	CUL4	13	133.270	133.250	96.900	3.590	0.110	0.871	0.149	1

****Genes meióticos para los cuales no se encontró ortólogo en girasol: AML2, RPA1C, BRCA2(IV), RECQ4B, CDKG1, ASY2, ZYP1b, CTF7, SGO1, PANS1, SMC2, SMC6A, NSE4b, TAM, OSD1, ASK1, BUB3.1.**

Tabla S15. Términos GO de categoría BP diferenciales entre meiocitos de los genotipos silvestre y domesticado.

GOid	Término GO	Fold Change D/S	qvalue
GO:2000604	negative regulation of secondary growth	10.13	2.98E-20
GO:0051511	negative regulation of unidimensional cell growth	7.50	0.001350908
GO:0009838	abscission	6.56	1.63663E-10
GO:0080126	ovary septum development	6.17	6.98361E-15
GO:0006792	regulation of sulfur utilization	5.12	7.30176E-13
GO:0080181	lateral root branching	5.06	8.86402E-10
GO:0009606	tropism	4.09	3.27699E-10
GO:0030155	regulation of cell adhesion	3.91	2.38E-17
GO:0006091	generation of precursor metabolites and energy	3.52	1.16E-26
GO:0051211	anisotropic cell growth	3.48	4.69903E-14
GO:0044403	symbiosis, encompassing mutualism through parasitism	3.40	0.002784881
GO:0030307	positive regulation of cell growth	3.26	1.41E-19
GO:0006081	cellular aldehyde metabolic process	3.25	9.89E-19
GO:0008285	negative regulation of cell proliferation	3.19	3.87E-25
GO:0048359	mucilage metabolic process involved in seed coat development	3.17	1.24E-20
GO:0006900	membrane budding	3.09	1.67893E-14
GO:0012501	programmed cell death	3.04	1.43436E-14
GO:0001881	receptor recycling	3.00	4.62666E-11
GO:0010098	suspensor development	2.99	3.95E-16
GO:0008104	protein localization	2.99	5.08882E-10
GO:2000693	positive regulation of seed maturation	2.97	3.84597E-13
GO:0019419	sulfate reduction	2.93	2.60273E-06
GO:0009856	pollination	2.92	7.70967E-13
GO:0010324	membrane invagination	2.92	3.61452E-08
GO:0080006	internode patterning	2.66	5.08882E-10
GO:0019079	viral genome replication	2.64	6.47805E-09
GO:0019048	modulation by virus of host morphology or physiology	2.60	2.03E-16
GO:0010074	maintenance of meristem identity	2.60	7.12877E-13
GO:0007009	plasma membrane organization	2.54	7.26656E-12
GO:0010256	endomembrane system organization	2.52	2.07525E-11
GO:0018920	glyphosate metabolic process	2.50	2.99703E-12
GO:0009653	anatomical structure morphogenesis	2.49	2.57278E-12
GO:0051014	actin filament severing	2.48	1.4001E-13
GO:0040034	regulation of development, heterochronic	2.48	7.52039E-08
GO:0008356	asymmetric cell division	2.48	2.36114E-14
GO:0043412	macromolecule modification	2.46	1.39113E-12
GO:0010434	bract formation	2.44	1.30831E-06
GO:0032955	regulation of barrier septum assembly	2.43	1.67441E-10
GO:0010015	root morphogenesis	2.43	2.38849E-09
GO:0009875	pollen-pistil interaction	2.42	5.34479E-07
GO:0080027	response to herbivore	2.42	4.27025E-07
GO:0010941	regulation of cell death	2.40	3.57149E-05
GO:0010358	leaf shaping	2.39	2.54776E-06

GO:0009799	specification of symmetry	2.30	0.001691089
GO:0008360	regulation of cell shape	2.29	8.97732E-11
GO:0040019	positive regulation of embryonic development	2.28	5.0558E-06
GO:0055047	generative cell mitosis	2.28	4.59682E-08
GO:0040014	regulation of multicellular organism growth	2.27	3.27804E-11
GO:0048530	fruit morphogenesis	2.25	1.04844E-09
GO:1900056	negative regulation of leaf senescence	2.24	0.000138762
GO:0010071	root meristem specification	2.23	1.24174E-09
GO:0000918	barrier septum site selection	2.18	2.73361E-09
GO:0010112	regulation of systemic acquired resistance	2.18	8.88418E-06
GO:0043697	cell dedifferentiation	2.17	2.30319E-07
GO:0010187	negative regulation of seed germination	2.15	1.06092E-07
GO:0010483	pollen tube reception	2.15	5.38035E-07
GO:0001558	regulation of cell growth	2.13	2.53071E-09
GO:0007267	cell-cell signaling	2.13	4.54629E-08
GO:0035019	somatic stem cell maintenance	2.12	8.75055E-06
GO:0051775	response to redox state	2.09	4.01335E-07
GO:0048827	phyllome development	2.09	6.02803E-08
GO:2000026	regulation of multicellular organismal development	2.08	1.7321E-06
GO:0051205	protein insertion into membrane	2.06	3.84767E-07
GO:0010497	plasmodesmata-mediated intercellular transport	2.06	1.05201E-08
GO:0048519	negative regulation of biological process	2.05	3.16516E-07
GO:0003002	regionalization	2.03	2.16269E-06
GO:0033043	regulation of organelle organization	2.03	2.16269E-06
GO:2000068	regulation of defense response to insect	2.03	0.000983949
GO:0010017	red or far-red light signaling pathway	2.02	6.8563E-07
GO:0006887	exocytosis	2.01	9.54548E-08
GO:0006066	alcohol metabolic process	0.48	3.50E-29
GO:0060862	negative regulation of floral organ abscission	0.45	3.22E-20
GO:0060867	fruit abscission	0.45	1.31473E-14
GO:2000692	negative regulation of seed maturation	0.45	1.31473E-14
GO:0022603	regulation of anatomical structure morphogenesis	0.45	4.96489E-11
GO:0009314	response to radiation	0.40	9.46E-28
GO:0009830	cell wall modification involved in abscission	0.32	4.30E-32
GO:0045848	positive regulation of nitrogen utilization	0.14	5.62E-48
GO:0010344	seed oilbody biogenesis	0.14	5.62E-48

Tabla S16. Términos GO de categoría CC diferenciales entre meiocitos de los genotipos silvestre y domesticado.

GOid	Término GO	Fold Change D/S	qvalue
GO:0010170	glucose-1-phosphate adenylyltransferase complex	8.116	3.08455E-09
GO:0000307	cyclin-dependent protein kinase holoenzyme complex	6.848	1.60E-16
GO:0009344	nitrite reductase complex [NAD(P)H]	4.143	2.81621E-07
GO:0012506	vesicle membrane	4.067	6.39598E-13
GO:0043674	columella	3.560	1.56449E-08
GO:0016328	lateral plasma membrane	3.259	3.96935E-08
GO:0010330	cellulose synthase complex	3.024	9.33272E-13
GO:0031520	plasma membrane of cell tip	3.008	1.51481E-15
GO:0043235	receptor complex	2.871	2.4528E-13
GO:0009897	external side of plasma membrane	2.866	2.55E-16
GO:0010009	cytoplasmic side of endosome membrane	2.697	2.92032E-09
GO:0009986	cell surface	2.637	5.05448E-10
GO:0010006	Toc complex	2.558	1.99331E-11
GO:0009521	photosystem	2.555	2.69991E-07
GO:0030132	clathrin coat of coated pit	2.541	1.79024E-11
GO:0005952	cAMP-dependent protein kinase complex	2.413	2.39335E-07
GO:0031965	nuclear membrane	2.408	1.87767E-11
GO:0045252	oxoglutarate dehydrogenase complex	2.398	2.42383E-07
GO:0005971	ribonucleoside-diphosphate reductase complex	2.349	4.67411E-10
GO:0005965	protein farnesyltransferase complex	2.270	3.07669E-09
GO:0005844	polysome	2.237	2.5819E-09
GO:0005951	carbamoyl-phosphate synthase complex	2.227	3.58727E-08
GO:0030427	site of polarized growth	2.210	1.79712E-08
GO:0015935	small ribosomal subunit	2.208	7.08207E-07
GO:0030089	phycobilisome	2.181	3.19648E-08
GO:0035145	exon-exon junction complex	2.172	1.23413E-06
GO:0035618	root hair	2.163	6.70489E-08
GO:0033588	Elongator holoenzyme complex	2.157	5.61984E-07
GO:0042555	MCM complex	2.106	1.96624E-07
GO:0005852	eukaryotic translation initiation factor 3 complex	2.073	2.42383E-07
GO:0031897	Tic complex	2.025	5.11725E-06
GO:0015934	large ribosomal subunit	2.014	2.60969E-05
GO:0000814	ESCRT II complex	2.011	2.32849E-06
GO:0009349	riboflavin synthase complex	2.007	7.08207E-07
GO:0000214	tRNA-intron endonuclease complex	0.416	4.22E-34
GO:0070825	micropyle	0.338	4.76202E-08
GO:0031012	extracellular matrix	0.204	3.78E-82

Tabla S17. Términos GO de categoría MF diferenciales entre meiocitos de los genotipos silvestre y domesticado.

GOid	Término GO	Fold Change D/S	qvalue
GO:0090353	polygalacturonase inhibitor activity	10.974	6.57E-42
GO:0005344	oxygen transporter activity	4.120	1.51E-24
GO:0004126	cytidine deaminase activity	4.097	6.02E-23
GO:0016879	ligase activity, forming carbon-nitrogen bonds	3.840	4.05548E-14
GO:0003963	RNA-3'-phosphate cyclase activity	3.521	5.71E-22
GO:0080016	(-)-E-beta-caryophyllene synthase activity	3.338	5.65769E-08
GO:0080017	alpha-humulene synthase activity	3.338	5.65769E-08
GO:0010013	N-1-naphthylphthalamic acid binding	3.325	3.42E-21
GO:0016846	carbon-sulfur lyase activity	3.313	8.63482E-13
GO:0000900	translation repressor activity, nucleic acid binding	3.138	3.21E-18
GO:0009881	photoreceptor activity	3.108	7.11254E-15
GO:0003876	AMP deaminase activity	3.008	2.14E-16
GO:0004016	adenylate cyclase activity	2.942	4.62664E-14
GO:0008835	diaminohydroxyphosphoribosylaminopyrimidine deaminase activity	2.854	3.60228E-15
GO:0019210	kinase inhibitor activity	2.762	2.08787E-06
GO:0045550	geranylgeranyl reductase activity	2.689	4.41377E-14
GO:0016830	carbon-carbon lyase activity	2.317	7.21899E-08
GO:0005096	GTPase activator activity	2.282	9.01286E-08
GO:0043022	ribosome binding	2.271	9.51262E-07
GO:0080132	fatty acid alpha-hydroxylase activity	2.244	0.000454083
GO:0046422	violaxanthin de-epoxidase activity	2.189	1.29402E-05
GO:0030611	arsenate reductase activity	2.183	9.39248E-06
GO:0005516	calmodulin binding	2.171	4.20311E-07
GO:0010011	auxin binding	2.092	4.40676E-07
GO:0008660	1-aminocyclopropane-1-carboxylate deaminase activity	2.085	3.81394E-06
GO:0003923	GPI-anchor transamidase activity	2.074	1.78803E-05
GO:0015225	biotin transporter activity	2.062	0.001130616
GO:0030742	GTP-dependent protein binding	2.021	7.04025E-06
GO:0016627	oxidoreductase activity, acting on the CH-CH group of donors	2.012	8.78105E-06
GO:0016701	oxidoreductase activity, acting on single donors with incorporation of molecular oxygen	0.479	2.80E-19
GO:0046910	pectinesterase inhibitor activity	0.373	1.89E-32
GO:0051738	xanthophyll binding	0.178	3.92E-39

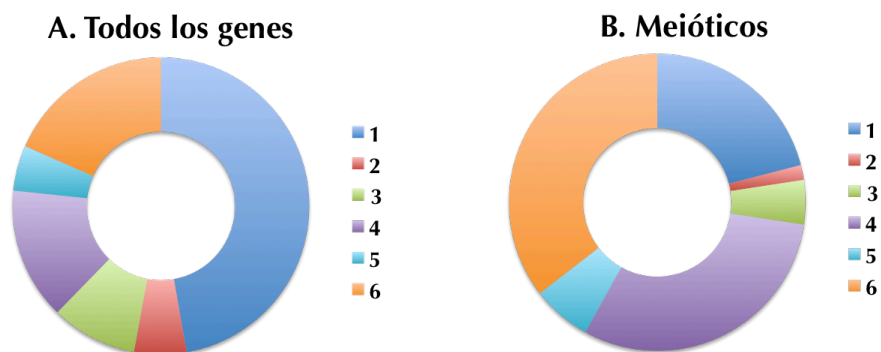


Figura S9. Patrones de expresión de todos los genes (A) y de los genes meióticos (B).
Estimados con los resultados del ensamble de los transcriptomas de meiocitos domesticado y somático (Flórez-Zapata et al., 2014).

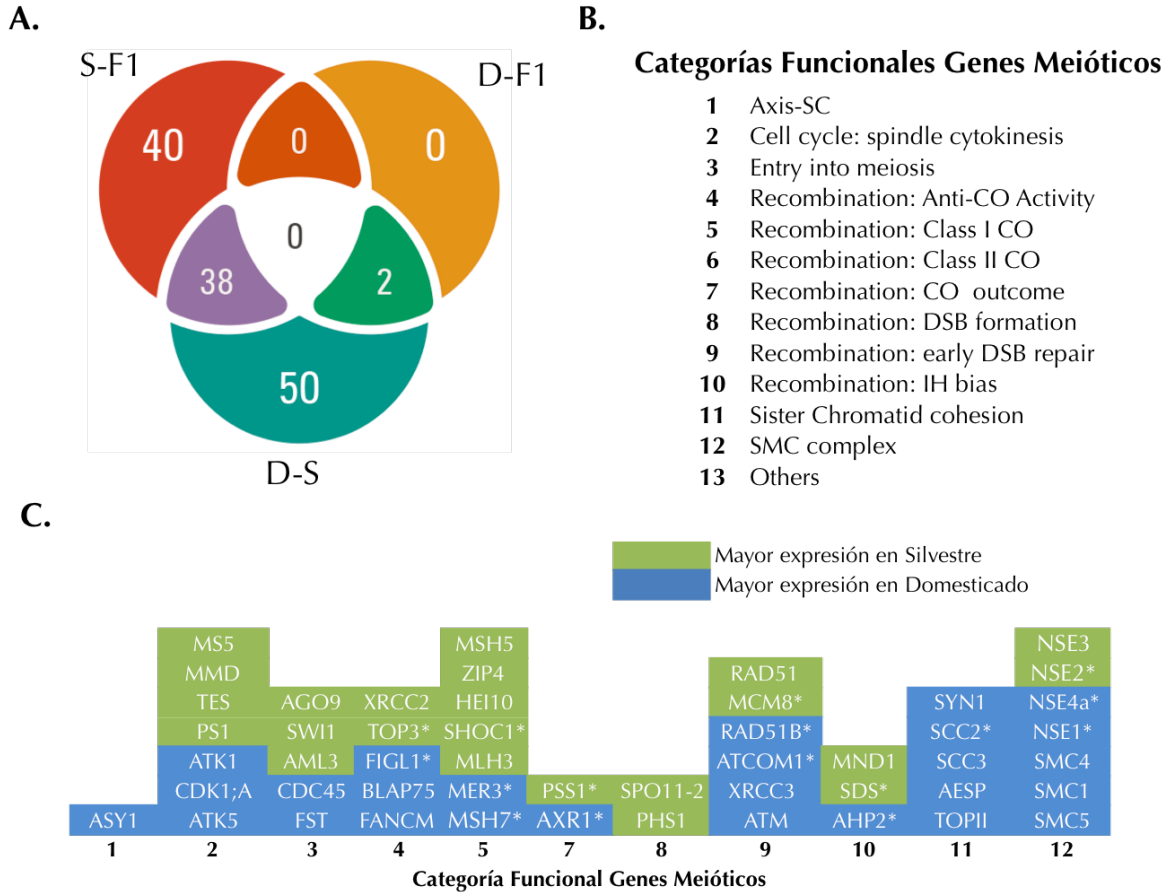


Figura S10. Resultados de expresión diferencial de los genes meióticos entre los genotipos de girasol a un valor $P < 0.05$. (A) Número de genes diferencialmente expresados en cada uno de los contrastes. (B) Categorías funcionales en las que fueron agrupados los genes meióticos. (C) Genes diferencialmente expresados en cada una de las categorías funcionales.

Tabla S18. Factores de transcripción que tienen el patrón de expresión 6

Locus	Familia TF	D-Sqv	TPM D	TPM F1	TPM Sil	TPM Som
AT1G28300	B3	0.002345493	0.20	0.23	0.05	0.02
AT2G13570	NF-YB	1.26861E-05	0.25	0.34	0.03	0.02
AT2G31210	bHLH	0.004368595	1187.21	1386.84	608.29	8.42
AT2G31220	bHLH	0.003213692	796.61	1022.99	399.30	0.81
AT3G19510	HB-PHD	0.001045617	25.98	29.26	10.73	1.57
AT3G54810	GATA	0.009639214	44.04	44.66	24.37	16.54
AT4G24020	Nin-like	0.007821154	15.54	16.18	8.42	4.68
AT4G38070	bHLH	0.000380513	41.04	42.58	17.29	0.11
AT5G44180	HB-other	0.003962151	48.81	50.98	23.20	2.52

Tabla S19. Factores de transcripción diferenciales previamente encontrados con mayor expresión en meiocitos de Arabidopsis y maíz.

Locus	Familia TF	D-Sqv	TPM D	TPM F1	TPM Sil	TPM Som	Patrón
AT2G31210	bHLH	0.004368595	1187.21	1386.84	608.29	8.42	6
AT2G31220	bHLH	0.003213692	796.61	1022.99	399.30	0.81	6
AT4G18960	MIKC	0.005308977	89.12	70.46	42.90	0.68	1
AT5G06839	bZIP	1.00582E-06	31.56	15.20	8.71	1.60	1
AT5G19790	ERF	2.83E-34	27.57	19.66	0.39	0.03	1
AT5G60910	MIKC	6.88191E-06	49.20	86.74	199.40	5.40	4

Tabla S20. Genes meióticos diferencialmente expresados para los cuales se identificaron elementos *cis* regulatorios.

Locus	Gen	Categoría Funcional	Patrón
AT4G18120	AML3	Entry into meiosis	4
AT5G57450	XRCC3	Recombination: early DSB repair	1
AT3G48190	ATM	Recombination: early DSB repair	6
AT4G29170	MND1	Recombination: IH bias	4
AT1G10970	ZIP4	Recombination: Class I CO	4
AT5G62410	SMC4	SMC complex	6
AT3G54670	SMC1	SMC complex	1
AT5G65460	CDKA;1	Cell cycle control: spindle and cytokinesis	1
AT4G21270	ATK1	Cell cycle control: spindle and cytokinesis	1
AT3G43210	TES	Cell cycle control: spindle and cytokinesis	4
AT4G05190	ATK5	Cell cycle control: spindle and cytokinesis	1

Tabla S21. Genes meióticos diferenciales que co-expresan con genes meióticos que pertenecen a la misma categoría funcional.

Gen 1	Patrón 1	Gen 2	Patrón 2	Categoría Funcional
PS1	4	TES	4	Cell cycle control: spindle and cytokinesis
PS1	4	ATK1	1	Cell cycle control: spindle and cytokinesis
FANCM	6	BLAP75/RMI1	1	Recombination: Anti-CO Activity
MMD	4	MS5	4	Cell cycle control: spindle and cytokinesis
MER3	6	MLH3	4	Recombination: Class I CO
PS1	4	ATK5	1	Cell cycle control: spindle and cytokinesis
ATK5	1	ATK1	1	Cell cycle control: spindle and cytokinesis
NSE3	4	SMC4	6	SMC complex

Tabla S22. Genes meióticos diferenciales que co-expresan con genes meióticos que pertenecen categorías relacionadas con recombinación.

Gen 1	Patrón 1	Categoría Funcional 1	Gen 2	Patrón 2	Categoría Funcional 2
PHS1	4	Recombination: DSB formation	MER3	6	Recombination: Class I CO
PHS1	4	Recombination: DSB formation	MLH3	4	Recombination: Class I CO
FANCM	6	Recombination: Anti-CO Activity	HEI10	4	Recombination: Class I CO
FANCM	6	Recombination: Anti-CO Activity	MER3	6	Recombination: Class I CO
FANCM	6	Recombination: Anti-CO Activity	MLH3	4	Recombination: Class I CO
FANCM	6	Recombination: Anti-CO Activity	RAD51	4	Recombination: early DSB repair
FANCM	6	Recombination: Anti-CO Activity	BLAP75/RMI1	1	Recombination: Anti-CO Activity
SPO11-2	4	Recombination: DSB formation	MLH3	4	Recombination: Class I CO
MSH7	6	Recombination: Class I CO	BLAP75/RMI1	1	Recombination: Anti-CO Activity
MER3	6	Recombination: Class I CO	MLH3	4	Recombination: Class I CO
MER3	6	Recombination: Class I CO	BLAP75/RMI1	1	Recombination: Anti-CO Activity
HEI10	4	Recombination: Class I CO	MND1	4	Recombination: IH bias
MND1	4	Recombination: IH bias	RAD51	4	Recombination: early DSB repair
MLH3	4	Recombination: Class I CO	RAD51	4	Recombination: early DSB repair
RAD51	4	Recombination: early DSB repair	BLAP75/RMI1	1	Recombination: Anti-CO Activity
MLH3	4	Recombination: Class I CO	XRCC3	1	Recombination: early DSB repair
MLH3	4	Recombination: Class I CO	BLAP75/RMI1	1	Recombination: Anti-CO Activity

Tabla S23. Frecuencia relativa de términos GO de categoría MF de genes co-expresados con genes meióticos y factores de transcripción (TF).

Término GO	Meióticos	TF
protein binding	0.427	0.308
nucleic acid binding	0.355	0.105
transcription factor activity, sequence-specific DNA binding	0.113	0.497
signal transducer activity	0.024	0.028
carbohydrate binding	0.016	
chromatin binding	0.016	0.035
hydrolase activity	0.016	0.007
oxidoreductase activity	0.016	0.007
peroxidase activity	0.008	
transferase activity	0.008	0.007
binding, bridging		0.007

Tabla S24. Frecuencia relativa de términos GO de categoría BP de genes co-expresados con genes meióticos y factores de transcripción (TF).

Término GO	Meióticos	TF
gene silencing	0.282	0.027
chromosome segregation	0.085	0.054
cell-cell signaling	0.079	-
cytokinesis	0.073	-
cell adhesion	0.068	0.135
oxidation-reduction process	0.040	-
regulation of cell proliferation	0.040	-
embryo development	0.034	-
seed development	0.028	-
signal transduction	0.028	-
transmembrane transport	0.028	-
negative regulation of biological process	0.023	-
seed germination	0.023	0.054
megagametogenesis	0.017	-
microgametogenesis	0.017	0.054
multicellular organism reproduction	0.017	0.027
positive regulation of cell proliferation	0.017	-
biosynthetic process	0.011	-
cell division	0.011	0.054
circadian rhythm	0.011	0.108
meristem initiation	0.011	0.108
multicellular organismal development	0.011	0.054
pollen germination	0.011	-
translational initiation	0.011	-
transport	0.011	-
cell growth	0.006	0.027
respiratory gaseous exchange	0.006	-
aging	-	0.027
cell communication	-	0.027
cell death	-	0.027
developmental growth	-	0.054
meristem maintenance	-	0.081
secondary growth	-	0.027
sexual reproduction	-	0.027
stomatal movement	-	0.027

Tabla S25. Secuencias de primers utilizados para el genotipado de líneas de los genes candidatos estudiados.

Gen	Primer	Secuencia	Gen	Primer	Secuencia
BLH11	RP_067982	TGCAAACATGCCACACATAAC	AT2G47680	RP_127155	GGCTCATGTTCTTCAAGCTTG
BLH11	LP_067982	TGGAGAGGAGTGTGAGAACG	AT2G47680	LP_127155	ACAGGGTGGTCTTGATGTCTG
BLH11	RP_GK_183705_01	TTCTGGTTTCATGTCITTTGCC	RDR2	RP_207201	TCCAAACTCATACCACGAAGG
BLH11	LP_GK_1837F05.01	TCATCATTGGCTTCTCTTTGC	RDR2	LP_207201	AGCAAGCTCCTTTTCAAAAACC
BLH11	RP_067443	CGTTCAACCACACAATGATTG	RDR2	RP_203854	TTCTGAACCTCGGTGCATATCC
BLH11	LP_067443	GGTAGGGACGTACGTAAGGATG	RDR2	LP_203854	TCCTTCGTGGTATGAGTTTGG
BLH90	RP_022615	CAATATTCACATGCAATCCC	RDR2	RP_206644	CGAGAAGCAGTAATCAGCACC
BLH90	LP_022615	TGAACCATTCCCTGAACAAAG	RDR2	LP_206644	CAGCTGAAATATGGACCGAAG
BLH90	RP_000633	TTAGGACGACGACCCITTTAG	AT1G05950	RP_SAIL_409_D03	CTTTGTTAGCAAGCAAGCTGC
BLH90	LP_000633	TGTGTGGAAATTCTCAGGGAG	AT1G05950	LP_SAIL_409_D03	TCACCCTTTTGTAAATGCAGG
BLH90	RP_052405	TCGTGGACTTCAATTTTCACC	AT1G05950	RP_SAIL_69_E08	CTTTGTTAGCAAGCAAGCTGC
BLH90	LP_052405	GTTTTGGCAGGAAATGTTTCAG	AT1G05950	LP_SAIL_69_E08	TCACCCTTTTGTAAATGCAGG
SHA	RP_GK-511A08.01	TCCTCGAAACTGTCACCAAC	APTC	RP_081627	RP_081627
SHA	LP_GK_511A08.01	ATGGCCGACATAAGTCACAAC	APTC	LP_081627	TTTTCGTCTGCACCTTCAGTCC
SHA	RP_GK-367A12.01	GAAATGTCACCACGTTTTC	APTC	RP_024760	CCCAAAATTTACATAAAACGCC
SHA	LP_GK-367A12.01	CAGAACGTTTCGCTTCAGAATC	APTC	LP_024760	GAAGACATGGATTGGGACAAG
SHA	RP_GK-379C07.01	TTCTTCTTCATCTCTAGCCGG	APTC	RP_125225	TGGATAATCACTTGGAAACACG
SHA	LP_GK-379C07.01	TAGTTCCTACACCGCAGCAG	APTC	LP_125225	AAGGGTATCCATTGGACCAAG
T7B11.05	RP_GK_440B08.01	TCCAAGCTTTCGAATGTGAC	FDM4	RP_206362	TTGGTTTCTTCTGTTTCCACG
T7B11.05	LP_GK_440B08.01	CACCAGGGCTAATGAAAGATG	FDM4	LP_206362	GATGCTGAGCAGGCATTAGAG
T7B11.05	RP_095758	ACAAGACGAACCGGAGAAATC	T4C980	RP_208312	GTTTAATCCCAAAACCCAAACC
T7B11.05	LP_095758	AGAAGTGGTAAGACCGCCTTC	T4C980	LP_208312	TTAACTGTGGATGCTGCTTCC
T7B11.05	RP_070883	ATCAATGGGAAAAGCGAAAC	T4C980	RP_SAIL_41_H09	TCAAACCGGTTTTATTGCTTG
T7B11.05	LP_070883	CTCGAGTGACTTCGTTGATCC	T4C980	LP_SAIL_41_H09	TAGATCCCACCAAACCCCTAC
AT3G50690	RP_033316	TCCGATGAGTAATGGCCATAG	AIL1	RP_SAIL_109_A03	TCAGGTAAGGAATCCGTTCC
AT3G50690	LP_033316	ACCCACAGCATCATGATCTTC	AIL1	LP_SAIL_109_A03	CTATACACGGAAAGCTCCCCTC
AT3G50690	RP_133139	TTGCAACAACCTACCTAACG	AIL1	RP_SAIL_69_E10	TCAGGTAAGGAATCCGTTCC
AT3G50690	LP_133139	CTTGCAATTGTTTTGTTGCTG	AIL1	LP_SAIL_69_E10	AACATGGCAACAAATCTTGC
MCT1	RP_014311	TCCATTACAATATCATTACCGCC	AIL1	RP_003704	CTCCTTGCTGTCCCTCCTTC
MCT1	LP_014311	TGTGCGATTATTAATTTTGGG	AIL1	LP_003704	GCCACCTGTTTTGTCTCAAATC
MCT1	RP_075516	AAGGCTTCTTCAAAGCCTTG	AT2G37880	RP_204855	AAGTCAAGAGGCTTCAAAGGG
MCT1	LP_075516	AACGGACACATTCACCTCCAC	AT2G37880	LP_204855	AAGACGATGGTATTGGTACG
AT4G08310	RP_146231	ACGGGTTGATAAGGATCCATC	AT2G37880	RP_GK_391B01	TTCAAAACAGAGTCTTGGATGAG
AT4G08310	LP_146231	AAGTCAGTCCATCGTTGTTGC	AT2G37880	LP_GK_931B01	TCAAACCCAGTCAAAGCTCAC
AHL5	RP_123590	TCTGGTGAATGAAGAAGGTGG	AT2G32350	RP_032843	AACACACCGATAAAGCGAATG
AHL5	LP_123590	AGAGCCATGACATTGACTTGG	AT2G32350	LP_032843	TCTCCGTCATTGAGAATTGG
AHL5	RP_GK_106G12.01	TAGCCATCAAAAACACAACC	REV7	RP_014571	AAGAAGGACCTAAGCGCAAAC
AHL5	LP_GK_106G12.01	CTGGTGCTTGAATCTGGTAGC	REV7	LP_014571	CATCGTTCTATGTCCGAGCTC
AT2G47680	RP_SAIL_877_H01	TAGTGCATTCAACCCCTTGTG			
AT2G47680	LP_SAIL_877_H01	ACCCGCACACATTTGACTAG			
AT2G47680	RP_GK_015F05.01	ATGCCGTTTGCAAAACTAATG			
AT2G47680	LP_GK_015F05.01	AAAGATATGGAACGTGGAGGC			

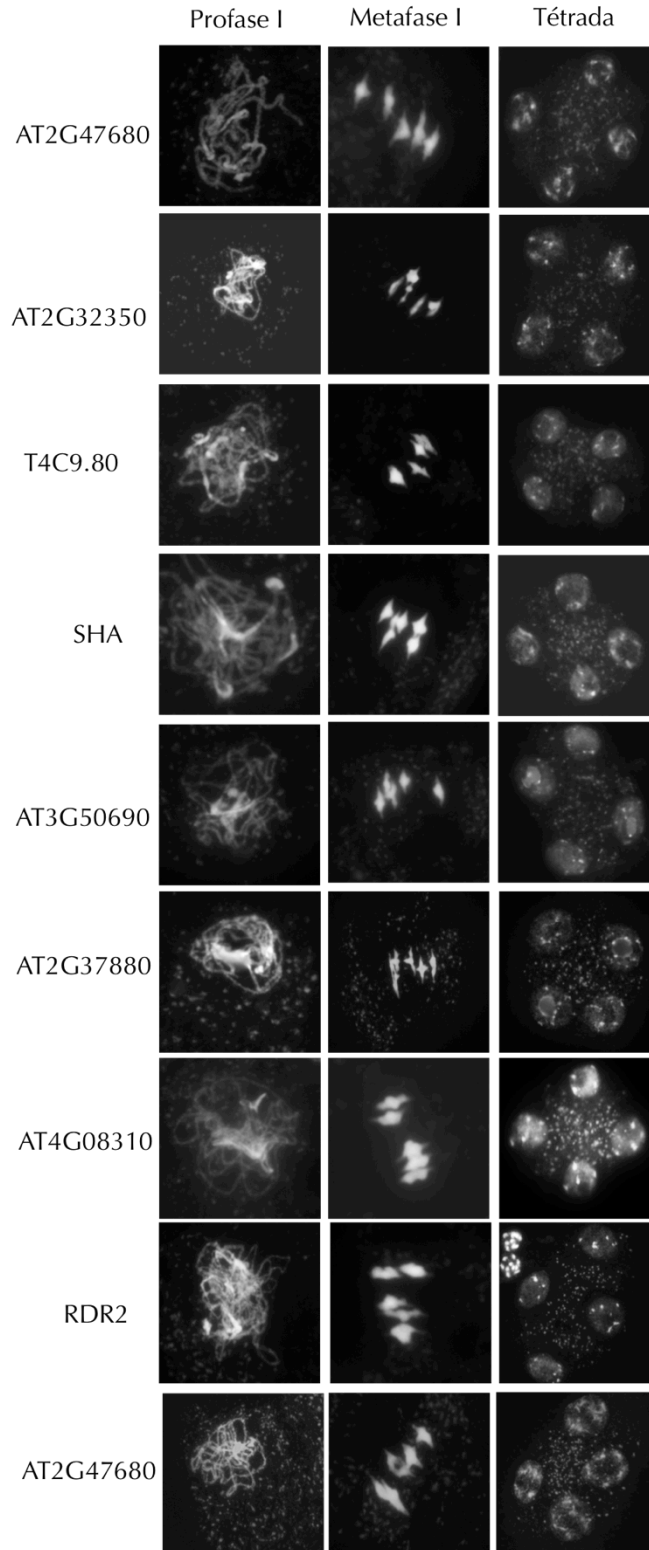


Figura S11. Resultados de la técnica de esparcido “spreading”, para líneas mutantes de algunos de los genes candidatos estudiados.