

# **EXPLORACIÓN FUNCIONAL Y ALERGÉNICA DE LOS TEJIDOS REPRODUCTIVOS DEL OLIVO (*OLEA EUROPAEA*)**



Amanda Bullones Pendón  
Tutor: M. Gonzalo Claros Díaz  
Cotutor: Noé Fernández Pozo





### DECLARACIÓN DE ORIGINALIDAD DEL TFM

D./Dña. Amanda Bullones Pendón.....

con DNI (NIE o pasaporte) .....

DECLARA que es autor/a del presente Trabajo de Fin de Máster, titulado: **Exploración funcional y alergénica de los tejidos reproductivos del olivo (Olea europaea)**

Que ha cumplido con las obligaciones legales sobre propiedad intelectual e industrial y que además, es un trabajo original, no habiéndose utilizado fuentes bibliográficas sin ser citadas debidamente. De no cumplir con este compromiso, el abajo firmante es consciente de que, de acuerdo con la Normativa de Evaluación y de Calificación de los estudiantes de la Universidad de Málaga de 14 de marzo de 2018, esto conllevará automáticamente la calificación numérica de cero, independientemente del resto de las calificaciones que hubiera obtenido. Esta consecuencia debe entenderse sin perjuicio de las responsabilidades disciplinarias en las que se pudieran incurrir.

En Málaga, a 28 de Junio de 2022

Firmado por BULLONES



Fdo. Amanda Bullones Pendón

### VISTO BUENO DEL TUTOR/COTUTOR

D./Dña. M. Gonzalo Claros Díaz..... en calidad de Tutor/a

con la Categoría Profesional de Catedrático....., en la Institución UMA.....,

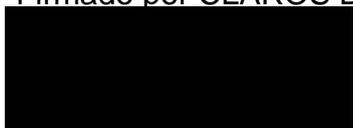
y D./Dña. Noé Fernández Pozo..... en calidad de Cotutor/a

con la Categoría Profesional de Investigador RYC....., en la Institución CSIC.....,

INFORMAN que el presente Trabajo Fin de Máster ha sido realizado bajo nuestra supervisión y cumple con el Reglamento relativo a la propiedad intelectual e industrial.

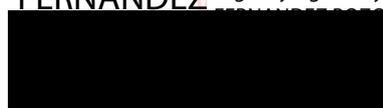
En Málaga, a 28 de Junio de 2022

Firmado por CLAROS DIAZ



Fdo. M. Gonzalo Claros Díaz.....

FERNANDEZ Digitally signed by



Fdo. Noé Fernández Pozo.....

## **Agradecimientos**

A Gonzalo y Noé por su infinita paciencia y por enseñarme todo lo que sé sobre Bioinformática. Gracias por darme esta oportunidad y por guiarme a lo largo de este año. Cada día estoy un poco menos perdida, aunque todavía me queda mucho por aprender. Espero que podamos seguir colaborando durante una buena temporada, ya que no se me ocurren ningunos tutores mejores que vosotros.

A Pedro y Darío, por ayudarme a aclararme con *Picasso* el día que fuera y a la hora que fuera, incluso en vacaciones.

A Juande y Rocío, por responder todos los correos que les mandaba con mil dudas.

Al equipo de la Estación Experimental del Zaidín por recoger el material vegetal y hacer posible este TFM.

A mis padres, por permitirme avanzar a mi propio ritmo y por apoyarme incondicionalmente.

A mi hermana, por estar siempre ahí.

A Juanjo, por recorrer este camino a mi lado y hacer que las cosas fueran un poquito menos feas. Has sido mi punto de apoyo en los momentos más difíciles y este último año ha sido bastante duro a nivel personal y académico, pero me alivia saber que estás ahí.

# Índice

Acrónimos y abreviaturas.....	1
Resumen.....	2
1. Introducción.....	3
1.1. El olivo: un modelo de interés.....	3
1.2. Genómica del olivo.....	4
1.3. Sistemas de autoincompatibilidad.....	5
1.4. Reproducción del olivo.....	7
1.5. Desarrollo del tubo polínico.....	8
1.6. Alérgenos del olivo.....	12
1.7. La semilla.....	13
2. Objetivos.....	15
3. Material y Métodos.....	16
3.1. Polen de Picual.....	16
3.2. Semilla de Picual.....	16
3.3. Extracción de proteínas.....	17
3.4. Separación electroforética.....	17
3.5. Análisis bioinformáticos.....	18
3.5.1. Análisis de calidad.....	18
3.5.2. Preprocesamiento de las lecturas.....	18
3.5.3. Mapeo y cuantificación.....	18
3.5.4. Determinación de ortologías.....	19
3.5.5. Expresión diferencial.....	19
3.5.6. Análisis funcionales.....	19
3.5.7. Genomas y proteomas.....	19
3.5.8. Agrupamiento de genes.....	20
3.5.9. Construcción de redes.....	21
3.5.10. Construcción del atlas de expresión del olivo (OliveAtlas).....	21
3.5.11. Identificación de alérgenos en Picual.....	22
3.5.12. Genes de la autoincompatibilidad esporofítica.....	22
3.5.13. Genes imprescindibles para el desarrollo del tubo polínico.....	22
4. Resultados.....	23
4.1. Mapeo de las lecturas.....	23

4.2. Análisis del desarrollo del tubo polínico.....	24
4.2.1. Filtrado de genes.....	24
4.2.2. Congruencia de las réplicas.....	25
4.2.3. Funciones celulares del polen.....	26
4.2.4. Expresión diferencial del desarrollo del tubo polínico.....	28
4.2.5. Coexpresión de genes durante el desarrollo del tubo polínico.....	31
4.3. Autoincompatibilidad esporofítica en el olivo.....	35
4.4. Genes imprescindibles en el desarrollo del tubo polínico.....	36
4.5. Alérgenos.....	39
4.5.1. Ortología.....	39
4.5.2. Perfil de expresión.....	40
4.6. Caracterización transcripcional de la semilla de olivo.....	40
4.6.1. Perfil proteico.....	40
4.6.2. Filtrado de los genes.....	41
4.6.3. Congruencia de las réplicas.....	42
4.6.4. Funciones celulares de la semilla.....	43
4.6.5. Expresión diferencial de los tejidos de la semilla.....	44
Coexpresión de genes de los tejidos de la semilla.....	46
4.7. El atlas de expresión del olivo.....	48
4.7.1. Construcción del atlas.....	49
4.7.2. Genes que más varían durante el desarrollo del tubo polínico.....	50
5. Discusión.....	54
6. Limitaciones y líneas futuras.....	60
7. Conclusiones.....	61
Material complementario.....	62
8. Bibliografía.....	63

## **Acrónimos y abreviaturas**

**CPM:** Cuentas por millón.

**CV:** Coeficiente de variación.

**DEG:** Gen expresado diferencialmente.

**DSI:** Autoincompatibilidad dialélica (*Diallelic Self-Incompatibility*).

**FDR:** Tasa de falsos positivos.

**GO:** Ontología génica (*Gene Ontology*).

**GSI:** Autoincompatibilidad gametofítica (*Gametophytic Self-Incompatibility*).

**NGS:** Secuenciación de nueva generación (*Next Generation Sequencing*).

**PCA:** Análisis de componentes principales.

**PSV:** Vacuola de almacenamiento de proteínas.

**SI:** Autoincompatibilidad (*Self-Incompatibility*).

**SSI:** Autoincompatibilidad esporofítica (*Sporophytic Self-Incompatibility*).

**SSP:** Proteínas de almacenamiento de la semilla.

## Resumen/Abstract

El olivo (*Olea europaea*) tiene una gran importancia socioeconómica debido a la producción de aceite y aceitunas de mesa. España es el primer productor mundial y, de las 272 variedades presentes en nuestro país, Picual es la más cultivada. Pese a su relevancia, aún hay aspectos de su biología que no se conocen en profundidad y que son importantes, por ejemplo, para mejorar la producción y aprovechar los huesos desechados en la extracción del aceite. En este trabajo, hemos demostrado que las variedades se parecen más entre sí que al acebuche. También caracterizamos el transcriptoma del polen de Picual incluido el desarrollo del tubo polínico, así como los procesos biológicos implicados en ello. Hemos determinado dos genes que podrían coordinar el crecimiento del tubo polínico y la respuesta al estrés abiótico, y otro que parece tener claras implicaciones reguladoras durante todo del desarrollo del tubo polínico. Además, identificamos y analizamos los perfiles de expresión de los genes de Picual que codifican los alérgenos de olivo conocidos hasta la fecha. También caracterizamos el transcriptoma de la semilla madura de Picual, donde se ven los procesos biológicos del endospermo que nutren el embrión, y el establecimiento de la polaridad embrionaria. Por último, creamos un atlas de olivo (OliveAtlas) donde se podrán consultar los datos de expresión de Picual de este trabajo.

The olive tree (*Olea europaea*) is of great socioeconomic importance due to the production of oil and table olives. Spain is the world's leading producer and, of the 272 varieties present in our country, Picual is the most cultivated. Despite its relevance, there are still aspects of its biology that are not fully understood and that are important for agriculture, for example, to improve yield and take advantage of pits discarded in oil extraction. In this work, we have shown that the varieties are more similar to each other than to the oleaster. We also characterized the transcriptome of Picual pollen including the pollen tube development, as well as the biological processes involved. We have found two genes that may regulate pollen tube growth and response to abiotic stress, and another that seems to have clear implications in the process of pollen tube development. In addition, we identified and analyzed the expression profiles of the Picual genes that encode the olive allergens known to date. We also characterized the transcriptome of the mature Picual seed, where the biological processes of the endosperm that nourish the embryo, and the establishment of embryonic polarity are seen. Finally, we created an olive atlas (OliveAtlas) where the Picual expression data from this work can be consulted.

## 1. Introducción

### 1.1. El olivo: un modelo de interés

El olivo (*Olea europaea*) es una planta leñosa dicotiledónea de hoja perenne que pertenece a la familia *Oleaceae*, que comprende entre 24 y 27 géneros (<https://eol.org>) y pertenece a las angiospermas (plantas con flores). El género *Olea* incluye alrededor de 40 especies, la mayoría arbustos o árboles. De todas ellas, *Olea europaea* es la única con frutos comestibles (Consejo Oleícola Internacional 1996). Dentro de esta especie se reconocen seis subespecies: *O. europaea* subsp. *laperrinei*, presente en las montañas del Sáhara; *O. europaea* subsp. *cuspidata*, proceden del sur y el este de África, Península Arábiga y Asia Central; *O. europaea* subsp. *guanchica*, de las Canarias; *O. europaea* subsp. *cerasiformis*, de Madeira; *O. europaea* subsp. *maroccana*, localizada en las montañas de Agadir (Marruecos) y *O. europaea* subsp. *europaea*, que se trata de la planta cultivada y se distribuye principalmente por la Cuenca del Mediterráneo y algunos puntos del suroeste de Europa (Talavera et al. 2017). Esta última no se conoce en la naturaleza y se le considera un antiguo cultigen (Green 2002).

El olivo es autóctono de la Cuenca del Mediterráneo y constituye un complejo de formas silvestres (*O. europaea* var. *sylvestris* o también conocido como *Olea oleaster* o acebuche) y cultivares clasificados como *O. europaea* var. *europaea* (*Olea sativa*) (Terral et al. 2004; Baldoni et al. 2006). Se da por hecho que ha sido explotado durante al menos 19.000 años (Besnard et al. 2001), ya que se ha descrito el uso del olivo silvestre en diferentes partes de la Cuenca del Mediterráneo durante el Neolítico. Se cree que su domesticación ocurrió hace al menos unos 7.000 años en Oriente Próximo a partir del acebuche (Diez et al. 2015; Langgut and Garfinkel 2022) y que se extendió desde el este hasta el oeste de la cuenca mediterránea con las migraciones humanas (Breton et al. 2009).

Se estima que en el mundo hay más de 2.000 variedades de olivo (Consejo Oleícola Internacional 1996; Breton et al. 2009). El cultivo se propaga por esquejes o injertos, por lo que los cultivares son clones (Breton et al. 2009). En España existen 272 variedades diferentes, pero las más extendidas se limitan a 24, que abarcan más de 2 millones de hectáreas. De todas ellas, la variedad Picual es la más cultivada con 900.000 hectáreas (Talavera et al. 2017) entre las provincias de Jaén, con presencia en el 97 % de su superficie olivarera, Granada y Córdoba (Consejo Oleícola Internacional 2000). También son relevantes las variedades Cornicabra, Hojiblanca y Lechín de Sevilla (Sanz-Cortés et al.

2002).

En el área mediterránea tiene una gran importancia a nivel social y económico (Carmona 2020) debido a su cultivo extendido y a que de su fruto derivan dos productos principales: las aceitunas de mesa y el aceite de oliva (Rallo et al. 2018). El olivo es un cultivo oleícola muy importante por su alta productividad y por la calidad de su aceite (Alché, et al. 2006), que tiene potenciales beneficios para la salud humana (Jiménez-Ruiz et al. 2020). Además, sus ramas pueden ser utilizadas para alimentar al ganado y su madera es de interés como leña y en ebanistería.

En la actualidad, su cultivo ocupa más de 11 millones de hectáreas en todo el mundo, de las que un 98 % se localizan en la Cuenca del Mediterráneo (Rallo et al. 2018) y un 21 % en España, lo que convierte a nuestro país en el primer productor mundial. Por lo tanto, dadas las repercusiones socioeconómicas que tiene el olivo, está claro que es un modelo de estudio en el que debemos profundizar.

## 1.2. Genómica del olivo

En 2016 se publicó la primera secuencia de un genoma completo de olivo (Cruz et al. 2016): un individuo con más de 1.200 años de la variedad Farga (subsp. *europaea*) cuyo interés económico es escaso, aunque tiene interés científico por su antigüedad. Al año siguiente, el consorcio internacional del genoma de olivo (IOGC) publicó la secuencia del genoma de un acebuche (subsp. *sylvestris*) de Oriente Medio (Unver et al. 2017). Desde la Universidad de Jaén se ha coordinado la secuenciación de la variedad Picual (subsp. *europaea*), cuyo borrador apareció en 2020 (Jiménez-Ruiz et al. 2020), y un año después ha aparecido el genoma de la variedad Arbequina (subsp. *europaea*) secuenciado desde China (Rao et al. 2021). Este mismo año se ha publicado un quinto genoma de la primera variedad de la subespecie *cuspidata* (Wang et al. 2022).

El genoma haploide del olivo se ha organizado en 23 cromosomas, con un tamaño que oscila entre las 1,31 Gb de Farga, y las 1,68 Gb de Picual, con un 51-63 % de secuencias repetitivas. Se han descrito entre 50.000 y 80.000 modelos de gen, de los que solo 47.124 tienen respaldo por RNA-seq. Antes de que aparecieran estos genomas ya se habían hecho análisis de expresión masiva de distintos tejidos, interacciones con patógenos, síntesis de aceite, maduración del fruto e incluso el desarrollo de la flor (Alagna et al. 2016, 2019), donde se describen genes que parece intervenir incluso en la interacción entre el polen y el pistilo de olivo. Pero queremos destacar el transcriptoma reproductivo de la va-

riedad Picual publicado por nuestro laboratorio (Carmona et al. 2015). Debido a estos avances ya tenemos referencias fiables para hacer análisis bioinformáticos con el olivo.

### 1.3. Sistemas de autoincompatibilidad

En las angiospermas, grupo al que pertenece el olivo, uno de los mecanismos que ayuda al mantenimiento de la diversidad genética es la autoincompatibilidad que impide la autofecundación mediante la discriminación entre el polen propio y el no propio (Muñoz-Sanz et al. 2020). Este mecanismo se ha descrito para las gramíneas, donde se basa en dos locus, S y Z, que son polialélicos, lo que implica que la reacción incompatible solo se da si los dos alelos del grano de polen coinciden con los del pistilo receptor (Klaas et al. 2011). En cambio, en el caso de las Brasicáceas, que es el modelo de autoincompatibilidad por antonomasia, está controlada solo por el locus S (Yamamoto and Nishio 2014). Además, el polen en su superficie, sobre la exina, lleva restos del tapete de la antera pertenecientes a la planta que lo formó, lo que también influye a la hora de que la planta receptora acepte o rechace el polen en función de su genotipo. Al primer mecanismo se le denomina sistema de autoincompatibilidad gametofítica (GSI) y el segundo se conoce como sistema de autoincompatibilidad esporofítica (SSI). No hay consenso sobre el tipo de autoincompatibilidad que presenta el olivo, pero se ha visto que suele permitir la germinación de su propio polen para acabar rechazándolo más adelante si no llega el polen de otra variedad (Cuevas González 2017).

Un gran número de estudios han intentado caracterizar el sistema de reproducción del olivo, sin que se haya llegado a ningún modelo de consenso al respecto. En esta especie, incluso los detalles biológicos más básicos de la autoincompatibilidad están sin resolver y no sabemos si está bajo control GSI o SSI (Saumitou-Laprade et al. 2017a). Algunos estudios afirman que en el olivo predomina la SSI (Breton and Bervillé 2012; Breton et al. 2014; Koubouris et al. 2014), lo que significaría que el grano de polen llevaría los dos determinantes especificados por el árbol del que procede el polen y el árbol polinizado no llevaría ninguno de esos (a menos que exista dominancia entre los dos alelos S en el macho). Otros, consideran que es de tipo GSI (Irene et al. 2010), que implicaría que el grano de polen albergaría uno de los dos determinantes de la planta del que procede el polen y si se posa en el estigma de una flor de un individuo portador del mismo determinante, no se logra la fecundación porque se desencadena una reacción autoincompatible que provoca el bloqueo del crecimiento del tubo polínico (Breton et al. 2014).

La tendencia más apoyada hasta el momento era la de tipo SSI, pero, recientemente,

un grupo de autores (Saumitou-Laprade et al. 2017b) propuso una nueva visión que parece cada vez más acertada para la autoincompatibilidad del olivo: el sistema de autoincompatibilidad dialéctica (DSI por sus siglas en inglés, *diallelic self-incompatibility*) homomórfico (Besnard et al. 2020; Mariotti et al. 2021). Parece que esta DSI no es frecuente en otras plantas (Alagna et al. 2019), y que en el olivo solo existirían dos genotipos de incompatibilidad, siendo todos los individuos de un grupo determinado incompatibles entre sí y totalmente compatibles con todos los individuos del otro grupo. Se cree que este sistema DSI se originó en un ancestro de las Oleáceas. Se ha estudiado en *Phillyrea angustifolia* y *Fraxinus ornus*, que pertenecen a la misma subtribu (Oleinae) que el olivo. Estos mismos autores confirmaron que los tres géneros (*Olea*, *Phillyrea* y *Fraxinus*) comparten el mismo sistema de autoincompatibilidad. No obstante, existen variedades de olivo que sí son autofertilizables y tendrían un sistema de pseudo-auto-incompatibilidad (Alagna et al. 2019).

A falta de caracterización de genes que intervienen en el DSI del olivo, el sistema de SSI del género *Brassica* está bien estudiado y se han descrito seis genes (tabla 1) importantes en este proceso (Yamamoto and Nishio 2014). La SSI está controlada por el locus S y, en el caso de las Brasicáceas, se han descrito los genes *S-locus receptor kinase* (SRK) en el estigma y su ligando en la exina del polen *S-locus cysteine rich protein/S-locus protein 11* (SCR/SP11), aunque se sintetiza en la antera. En la autopolinización, la SRK del estigma interactúa con la SCR de la exina y da lugar a la autoincompatibilidad y se activa una señalización en la que intervienen el gen *M-locus protein kinase* (MLPK), el *Arm repeat-containing* (ARC1) y la *Thioredoxin H-Like* (THL1). Por último, se ha descrito un tercer gen, el *S-locus glycoprotein* (SLG) en el locus S, pero su papel en la autoincompatibilidad no está claro (Yamamoto and Nishio 2014).

**Tabla 1.** Genes que regulan la autoincompatibilidad esporofítica en *Brassica*.

Gen	Función	Organismo en el que se ha descrito	UniProt
SRK	<i>S-locus receptor kinase</i>	<i>Brassica oleracea</i>	Q93YK0
SCR	<i>S-locus cysteine rich protein</i>	<i>Brassica oleracea</i> (en las anteras)	Q9SE19
MLPK	<i>M-locus protein kinase</i>	<i>Brassica campestris</i>	Q75W78
ARC1	<i>Arm repeat-containing</i>	<i>Brassica oleracea</i>	Q84WX0
THL-1	<i>Thioredoxin H-Like</i>	<i>Brassica napus</i>	P68177
SLG	<i>S-locus glycoprotein</i>	<i>Brassica napus</i>	Q6SV92

Otro mecanismo descrito en las plantas que dificulta la autofecundación es la dicogamia, o sea, la maduración separada en el tiempo de los órganos sexuales masculinos y femeninos de la flor. Se puede dar una protoginia, si precede la función femenina, o una protandria si lo hace la masculina. En el olivo se da la segunda, por lo que las anteras se abren y liberan su polen antes de que el estigma sea receptivo (Cuevas González 2017).

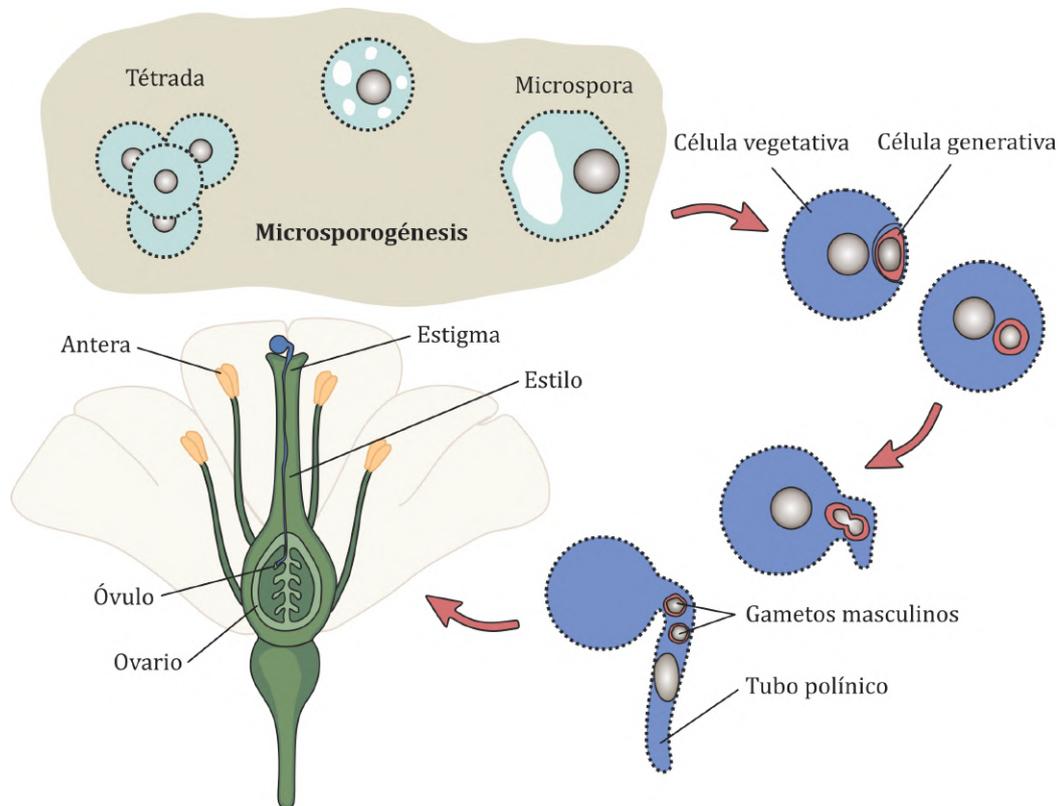
El olivo es capaz de formar flores hermafroditas, pero también flores estaminadas (masculinas) incapaces de fructificar. Su polinización es anemófila, lo que explica la producción de tanto polen (una media de 116.240 granos de polen por flor) y su difusión a grandes distancias (Cuevas González 2017).

#### **1.4. Reproducción del olivo**

El olivo, como las demás angiospermas, tiene un ciclo de vida haplodiplonte donde se alternan las fases gametofítica haploide y esporofítica diploide. Los gametofitos masculinos (granos de polen) y femeninos (óvulos) se desarrollan a través de una serie de divisiones mitóticas de las esporas haploides, que se producen cuando las células esporofíticas dentro de la antera (macho) y la ovocélula (hembra) sufren meiosis (Qin et al. 2009) (figura 1). Las cuatro microsporas haploides del polen se asocian en una tétrada, pero se separan y experimentan vacuolación y expansión, y el núcleo de la microspora migra hacia la pared celular (Rutley and Twell 2015). En esta disposición, sufre una mitosis asimétrica (mitosis polínica I) para producir la célula vegetativa que envuelve a una célula generativa más pequeña (Qin et al. 2009). La célula vegetativa no vuelve a dividirse, pero formará el tubo polínico. Sin embargo, la célula generativa sufre una segunda mitosis (mitosis polínica II) que da lugar a dos gametos masculinos. El momento de esta mitosis varía en las diferentes familias de plantas: lo más habitual es que suceda durante el crecimiento del tubo polínico (McCormick 2004), aunque en el caso de las gramíneas se da dentro de la antera. Está claro que el estudio detallado del mecanismo de autoincompatibilidad que rige la reproducción del olivo es interesante para optimizar el rendimiento de producción de frutos.

Las angiospermas han desarrollado dos estrategias principales para la interacción polen-pistilo: especies con estigmas secos, cuyas plantas modelos son *Arabidopsis* y el arroz, y especies con estigmas húmedos entre las que se encuentran las solanáceas y el olivo (Quiapim et al. 2009). Para que haya fecundación, el polen deshidratado debe llegar a un estigma compatible, que tiene una alta actividad de peroxidasas y esterases (Rejón et al. 2012), y entra en contacto con el exudado estigmático, se hidrata y se esti-

mula la emergencia del tubo polínico (Kessler et al. 2015). A este proceso de fusión de los gametos se le denomina doble fecundación porque uno de los gametos masculinos se fusionan con la ovocélula para dar lugar el embrión, mientras que el otro gameto masculino se fusiona con los dos núcleos polares del saco embrionario para generar el endospermo (Cuevas González 2017).



**Figura 1.** Etapas citológicas del desarrollo del gametofito masculino en *Arabidopsis*. Las células madre del polen diploide sufren una división meiótica para producir una tétrada de microsporas haploides. Las microsporas liberadas de la tétrada sufren una división celular asimétrica (mitosis polínica I) para producir un grano de polen bicelular con una célula generativa dentro del citoplasma de la célula vegetativa. La célula generativa se divide de nuevo (mitosis polínica II) para producir un par de gametos masculinos. (Adaptada de (Rutley and Twell 2015; da Costa-Nunes and Grossniklaus 2004) y Biorender.com 2022).

### 1.5. Desarrollo del tubo polínico

En la mayoría de las plantas, los granos de polen maduros se liberan de las anteras en un estado parcialmente deshidratado (McCormick 2004). Se estima que el polen maduro contiene unos 20.000 transcritos diferentes, de los que el 10 % se expresan específicamente en el genoma haploide. Varios de estos genes han sido aislados en diferentes especies vegetales y muchos comparten un patrón común "tardío" de expresión durante el

desarrollo del polen (Rogers et al. 2001).

Cuando el polen se encuentra con un estigma compatible, emerge el tubo polínico, que se elonga mediante un mecanismo basado en filamentos de actina y miosina, que transportan vesículas cargadas de material para la nueva pared celular hasta el ápice de la extensión (Qin et al. 2009). A este proceso se le conoce como "crecimiento de la punta" (*tip growth* en inglés). La pared celular que recubre la región de la punta del tubo polínico tiene una composición especial que permite la expansión continua, mientras que la del eje está reforzada y no es extensible (Beuder et al. 2022). Las peroxidasas intervienen en la formación y el crecimiento de la pared celular del tubo polínico (Bredemeijer 1984) y en el polen de *Arabidopsis* tenemos dos genes que codifican una peroxidasa clase III. A medida que el tubo polínico se desarrolla, el núcleo de la célula vegetativa y los dos gametos masculinos no móviles se desplazan por su interior. El desarrollo del polen se completa cuando los dos gametos masculinos se liberan en el saco embrionario (McCormick 2004) donde los esperan cuatro óvulos disponibles para la fecundación, pero solo uno de ellos formará la semilla (Cuevas González 2017).

Un grano de polen maduro contiene enzimas que se liberan al entrar en contacto con la superficie estigmática, entre las que podemos destacar las esterases que en el polen de *Arabidopsis* están reguladas por tres genes: uno de ellos codifican fosfotriesterasas dependientes de calcio y los otros dos codifican esterases de tipo GDSL. Dentro de este grupo encontramos las hidrolasas de ésteres carboxílicos; las cutinasas, que se han descrito en el polen de *Tropaeolum majus*, *Brassica napus* y *Arabidopsis thaliana* y se encargan de romper la cutina presente en la cutícula del estigma; las carboxilesterasas, que son necesarias para la penetración del tubo polínico en el estigma (descrita en *Brassica napus*); y las lipasas, que podrían estar implicadas en la degradación de estructuras lipídicas del estigma y que se han descrito en la cubierta del polen de *Arabidopsis* (donde están involucradas en el desarrollo y la germinación del polen y el crecimiento del tubo polínico) y el girasol (Rejón et al. 2012).

El tubo polínico es una estructura tubular cuya pared celular suele estar formada por dos capas: la capa fibrilar externa, compuesta por pectina, hemicelulosa y celulosa; y la capa interna de callosa (Yang et al. 2021). Por eso, el crecimiento del tubo polínico requiere pectina para la extensión de la punta. Además, la callosa está regulada por dos genes: uno de ellos es necesario para la viabilidad del polen y se expresa más en el polen maduro y el otro codifica una proteína de la superfamilia UDP-glucosiltransferasa cuya ausencia conlleva la rotura del tubo polínico tras la germinación. Unos genes "tardíos" codifican pro-

ductos con homología a las enzimas que metabolizan la pectina, como la poligalacturonasa, pectina metil esterasa y la pectato liasa (Rogers et al. 2001). En cuanto a la pectina metil esterasa, está codificada por siete genes de *Arabidopsis*: uno de ellos regula la viabilidad del polen y el crecimiento del tubo polínico, uno se localiza en la punta del tubo polínico e inactiva a *PPME1*, otro es necesario para el crecimiento del tubo polínico, el cuarto está implicado en la remodelación de las pectinas en el polen, otro es un gen restringido al grano de polen, y los dos últimos están relacionados con el crecimiento del tubo polínico.

Las NADPH oxidasas (NOX), conocidas en las plantas como Rbohs (homólogos de la oxidasa respiratoria), son enzimas que generan el radical superóxido  $O_2^-$  en el apoplasto durante su acción enzimática, lo que conduce a un incremento de las especies reactivas de oxígeno (ROS) extracelulares. En *Arabidopsis* se codifican diez Rbohs, con dos formas específicas del polen cuya actividad es esencial para el crecimiento del tubo polínico (Jiménez-Quesada et al. 2015). Aunque la composición y la construcción del tubo polínico se conocen bien, los mecanismos funcionales y reguladores no están claros, sin embargo, todos los modelos descritos implican al  $Ca^{2+}$  como molécula de señalización esencial. Se sabe que un gradiente de  $Ca^{2+}$  centrado en la punta es esencial para la germinación del polen y el crecimiento del tubo polínico. El incremento del gradiente de  $Ca^{2+}$  se correlaciona con el crecimiento del tubo. La señal de  $Ca^{2+}$  puede ser decodificada y transmitida por las proteína quinasas dependientes del calcio (CDPKs) que pueden ser activadas directamente por el  $Ca^{2+}$  y fosforilar los efectores posteriores para regular innumerables procesos biológicos, entre los que se encuentra el crecimiento del tubo polínico. Las CDPKs han sido identificadas en todo el reino vegetal y constituyen una gran familia multigénica en varias especies de plantas (Yang et al. 2021).

Se sugiere que la ruta de señalización  $Ca^{2+}$ /CPKs puede regular el crecimiento del tubo polínico manteniendo las concentraciones de iones intracelulares adecuadas en el ápice a través de canales iónicos diversificados y bien ajustados (Yang et al. 2021). Se ha visto que la interrupción del gradiente de calcio en la punta provoca la inhibición del crecimiento del tubo (Golovkin and Reddy 2003). Además, esta ruta puede interactuar con otras moléculas de señalización como las ROS, que son generadas por las Rboh y que también participan en el crecimiento del tubo polínico (Yang et al. 2021).

La calmodulina (CaM), un sensor de calcio multifuncional ubicuo en todos los eucariotas, media la acción del calcio mediante la regulación de la actividad de muchas proteínas no relacionadas. (Golovkin and Reddy 2003) aislaron tres proteínas de unión a CaM estre-

chamente relacionadas (NPG1, NPGR1 y NPGR2) de Arabidopsis. La primera se expresa solo en el polen, mientras que las otras dos se expresan en el polen y en otros tejidos. NPG1 es esencial para la germinación del polen (Golovkin and Reddy 2003).

En la tabla 2 recogemos los genes imprescindibles en el desarrollo del tubo polínico que se han descrito en Arabidopsis

**Tabla 2.** Genes imprescindibles en el desarrollo del tubo polínico descritos en Arabidopsis.

<b>Gen</b>	<b>UniProt</b>	<b>Función</b>
<i>BET11</i>	Q9M2J9	Necesario para el transporte de vesículas desde el retículo endoplasmático hasta el complejo de Golgi
<i>BET12</i>	Q94CG2	Necesario para el transporte de vesículas desde el retículo endoplasmático hasta el complejo de Golgi
<i>AtCPK2</i>	Q38870	Decodifica señales de Ca <sup>2+</sup>
<i>AtCPK14</i>	P93759	Decodifica señales de Ca <sup>2+</sup>
<i>AtCPK16</i>	Q7XJR9	Decodifica señales de Ca <sup>2+</sup>
<i>AtCPK17</i>	Q9FMP5	Decodifica señales de Ca <sup>2+</sup>
<i>AtCPK20</i>	Q9ZV15	Decodifica señales de Ca <sup>2+</sup>
<i>AtCPK24</i>	Q9SIQ7	Decodifica señales de Ca <sup>2+</sup>
<i>AtCPK26</i>	Q9SZM3	Decodifica señales de Ca <sup>2+</sup>
<i>AtCPK34</i>	Q3E9C0	Decodifica señales de Ca <sup>2+</sup>
<i>COG6</i>	Q6NMI3	Esencial en el crecimiento del tubo polínico
<i>AtTMEM18</i>	F4HUY1	Importante para el crecimiento del tubo polínico
<i>SEC1A</i>	Q9C5P7	Importante en el tráfico de vesículas durante el crecimiento de la punta del tubo polínico
<i>ROP1 GTPase</i>	P92978	Implicado en el control de la polaridad de la célula durante el crecimiento de la punta del tubo polínico
<i>AtGEF12</i>	Q9CA89	Mantiene la actividad de Rop en el tubo polínico y controla el crecimiento polarizado del tubo polínico
<i>RabA4d</i>	Q9LH50	Tráfico de vesículas intracelulares y transporte de proteínas durante el crecimiento de la punta del tubo polínico
<i>ACA9</i>	Q9LU41	Catalizar la hidrólisis del ATP
<i>CNGC18</i>	Q9LEQ3	Necesario para el crecimiento direccional del tubo polínico
<i>VANGUARD</i> pectinesterase	Q5MFV6	Actúa en la modificación de la pared celular
<i>AtMGD2</i>	O82730	Implicado en la construcción de la membrana plasmática del tubo polínico
<i>AtMGD3</i>	Q9SI93	Implicado en la construcción de la membrana plasmática del tubo polínico
<i>NPG1</i>	Q8GZN1	Proteínas de unión a CaM esencial para la germinación del polen
<i>NPGR1</i>	Q9CB03	Proteínas de unión a CaM esencial para la germinación del polen

Gen	UniProt	Función
<i>NPGR2</i>	Q66GN3	Proteínas de unión a CaM esencial para la germinación del polen

Así, el pistilo tiene una doble función: la producción de los gametos femeninos en el ovario y la discriminación de los granos de polen que se posan en el estigma. Durante la fase progámica, que abarca desde el momento en el que el polen llega al estigma hasta la fecundación, se producen interacciones entre el polen y el pistilo, donde el éxito de la fecundación depende del reconocimiento del polen y de que se den las condiciones apropiadas para su hidratación, la germinación del tubo polínico y el crecimiento direccional hacia los óvulos, procesos que tienen lugar en los tejidos del estigma/estilo (Quiapim et al. 2009). En consecuencia, merecerá la pena averiguar si todas estas actividades están también presentes durante el desarrollo del tubo polínico del olivo.

### 1.6. Alérgenos del olivo

El polen del olivo es una importante fuente de alergia en los países mediterráneos (Liccardi et al. 1996) y es de los más alergénicos que se producen en nuestras latitudes. Los alérgenos son proteínas que se expresan significativamente en el polen y muchos de ellos tienen un papel fundamental en su fisiología (Jiménez-Quesada et al. 2015). En el caso del olivo, se han identificado y caracterizado 15 alérgenos hasta la fecha (tabla 3).

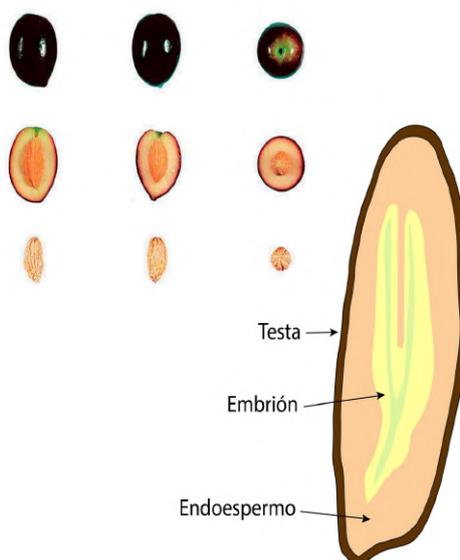
**Tabla 3.** Alérgenos del olivo. Todos ellos han sido descritos en polen, menos el marcado con \* que es el único presente en el fruto. Hemos usado las secuencias de nucleótidos, menos para los alérgenos marcados con \*\* que se ha empleado su secuencia de proteínas. *Ole e 1-Ole e 3*, *Ole e 5* y *Ole e 6* son secuencias de la variedad Picual, *Ole e 8* es una secuencia de acebuche y el resto son secuencias de variedad no especificada. (Modificada de (Carmona 2020)).

Alérgeno	Familia	NCBI
<i>Ole e 1</i>	Ole e 1-like	AY159880, AF532760, AF532754, AF532753
<i>Ole e 2</i>	Profilina	DQ317582, DQ317581, DQ640904, DQ663558, DQ663557, DQ663556, DQ663555, DQ663554, DQ663553
<i>Ole e 3</i>	Polcalcina	AY172113
<i>Ole e 4**</i>	Producto de degradación de Ole e 9	P80741
<i>Ole e 5</i>	Cu/Zn superóxido dismutasa	EU250797, EU250796, EU250795, EU247090, EU247089, EU247088
<i>Ole e 6**</i>	ND	AY221958, AAO34670.1
<i>Ole e 7**</i>	Proteína de transferencia de lípidos (LTP)	P81430.2

Alérgeno	Familia	NCBI
<i>Ole e 8</i>	Proteína de unión a calcio	XM_023029144
<i>Ole e 9</i>	1.3- $\beta$ -glucanasa	AF249675
<i>Ole e 10</i>	Proteína de unión a carbohidratos (CBM-43)	AY082335
<i>Ole e 11</i>	Pectina metilesterasa	DQ026502, GU084173
<i>Ole e 12</i>	Isoflavona reductasa-like	EU927297
<i>Ole e 13*</i>	Taumatina-like (TLP)	GU084174
<i>Ole e 14</i>	Poligalacturonasa	KC920918
<i>Ole e 15</i>	Ciclofilina	MG742363

### 1.7. La semilla

La semilla del olivo consta de tres partes: la testa, que es la capa que rodea la semilla; el endospermo, que protege al embrión y le suministra los nutrientes necesarios para su crecimiento (Yan et al. 2014); y el embrión (figura 2). Estos tres tejidos están dispuestos uno dentro del otro y se desarrollan de forma sincronizada tras la fecundación para generar una semilla viable (Ingram and Gutierrez-Marcos 2015).



**Figura 2.** Fotografía del fruto del olivo, variedad Picual (Universidad de Córdoba) y representación de los tejidos de la semilla del olivo (Modificada de (Zafra et al. 2018)).

En la semilla, (Maestri et al. 2019) determinaron que el porcentaje de aminoácidos es-

enciales individuales es de un 46 %. Este parámetro es un indicador de la calidad de las proteínas y supera a los de la harina de trigo integral ( $\approx 27$  %). También es una fuente rica en fibra dietética total ( $\approx 47$  % del peso seco), lípidos ( $\approx 30$  %) y proteínas ( $\approx 17$  %) y cuenta con cantidades significativas de macroelementos esenciales (K, Ca, Mg, Na y P) y microelementos (Zn, Mn y Cu). Los compuestos fenólicos están presentes en cantidades relativamente elevadas ( $\approx 2,8$  mg/g del peso seco), principalmente los del grupo de los compuestos secoiridoides (derivados del ácido elenólico).

Basándonos en el perfil nutricional y los componentes nutraceuticos, la semilla del olivo tiene un alto potencial como fuente de aceite comestible, proteínas o harinas (Maestri et al. 2019). Actualmente, la extracción del aceite de oliva virgen extra se realiza triturando el fruto al completo, incluido el hueso, para después prensar y decantar dicho aceite separándolo del alpechín y los elementos sólidos (Consejo Oleícola Internacional 1996). Así, el aceite contiene componentes de la semilla que contribuyen al aroma y a otras propiedades como la actividad peroxidasa (Zafra et al. 2018). Sin embargo, hay pruebas de que el aceite extraído de aceitunas deshuesadas es de mejor calidad que el que se obtiene de la molienda de la aceituna entera. Por esto se prevé que la producción de AOVE a partir de los frutos deshuesados aumente, con el consecuente incremento de huesos de aceituna como desecho. Actualmente, estos huesos se aprovechan para producir energía (Maestri et al. 2019) mediante la quema, una práctica muy contaminante.

Las proteínas de almacenamiento de la semilla (SSP, *seed storage proteins*) se generan durante el desarrollo de la semilla y se almacenan en las vacuolas de almacenamiento de proteínas (PSV) del embrión y del endospermo (Zafra et al. 2018). Estas proteínas se almacenan constantemente y se degradan cuando la semilla germina. Las proteínas más abundantes en la semilla madura de olivo son las prolaminas (aproximadamente un 70 % del total de las proteínas de la semilla), que pertenecen a la familia de las 11S (Alché, et al. 2006). Otra familia de SSP importantes son las globulinas 7S que incluyen a las  $\beta$ -conglutinas y las vicilinas (Zafra et al. 2018). Ambas se han descrito en la semilla del olivo.

Las SSP son las proteínas vegetales más consumidas por el ser humano y tienen una gran calidad nutricional. Sin embargo, las presentes en las semillas del olivo, que es uno de los principales cultivos de los países mediterráneos, no se destinan al consumo humano ya que en la producción de aceite se emplean los frutos enteros (Alché, et al. 2006). Por tanto, el estudio de los procesos biológicos de la semilla que estarían implicados en la acumulación de estos interesantes productos podría contribuir a aumentar el interés nutraceutico y económico de las semillas de olivo.

## 2. Objetivos

En la introducción hemos visto que, a pesar del gran interés socioeconómico del cultivo del olivo, poco se sabe de muchos aspectos de su funcionamiento. En este trabajo se han planteado pues los siguientes objetivos:

1. Caracterizar funcionalmente el polen de Picual y determinar si hay nuevos genes relacionados con el crecimiento del tubo polínico.
2. Identificar los genes de Picual que codifican los alérgenos conocidos para ver sus perfiles de expresión durante el desarrollo del tubo polínico.
3. Caracterizar el transcriptoma de la semilla de Picual.
4. Crear un atlas de expresión de Picual con la información de polen y semilla.

### 3. Material y Métodos

#### 3.1. Polen de Picual

Se recogió polen de la variedad Picual de olivos situados en la Estación Experimental del Zaidín (CSIC, Granada, España) durante la antesis y se almacenó a -80 °C hasta su posterior uso. La germinación del polen se hizo *in vitro*. El primer paso consistió en la prehidratación en una cámara húmeda en la oscuridad a temperatura ambiente durante 1 hora. A continuación, el polen se transfirió al medio de germinación (10 % [p/v] de sacarosa, 0,03 % [p/v] de Ca[NO<sub>3</sub>]<sub>2</sub>, 0,01 % [p/v] de KNO<sub>3</sub>, 0,02 % [p/v] de MgSO<sub>4</sub>, y 0,03 % [p/v] de ácido bórico), como se describe en (Zienkiewicz et al. 2010). El polen se mantuvo a temperatura ambiente en oscuridad y bajo agitación continua, y se tomaron muestras antes de su hidratación (dos réplicas) y en diferentes momentos de la germinación (dos réplicas a 0 h y tres réplicas a 1, 3 y 6 h). El ARN se extrajo con el kit RNeasy Plant Total RNA (Quiagen, U.S.A.). Las muestras fueron secuenciadas en el NextSeq 550 del Servicio de Ultrasecuenciación de la Universidad de Málaga.

En función de las muestras que tenemos, hemos planteado cuatro comparaciones: 0 h frente a 1 h, 1 h frente a 3 h, 3 h frente a 6 h y 1 h frente a 6 h. Las tres primeras son importantes porque siguen la evolución temporal del desarrollo del tubo polínico. La comparación P1H-P6H también resulta de interés porque podemos ver los cambios que se producen desde que se activa el desarrollo del tubo polínico hasta que termina.

#### 3.2. Semilla de Picual

Se ha trabajado con frutos de olivo (aceitunas) de la variedad Picual en tres estadios de maduración: verde, envero (momento de maduración del fruto, en el que cambia su color gradualmente) y maduro que se recolectaron de olivos situados en la Estación Experimental del Zaidín. Los frutos se despulparon manualmente y, tras retirar la cáscara, las semillas se dividieron en dos partes: embrión y endospermo+testa. Una vez separadas, las muestras se congelaron a -80°C y se almacenaron hasta su uso.

Para la extracción de ARN, se trituraron entre 0,05 y 0,1 g de muestra en un mortero en frío con nitrógeno líquido, con la precaución de que haya siempre nitrógeno líquido para evitar la degradación del ARN y las proteínas. El ARN total se extrajo con el kit RNeasy Plant Total RNA (Quiagen, U.S.A.) y se mandó secuenciar al Servicio de Ultrasecuenciación de la UMA.

Tenemos siete muestras de semilla:

- Una réplica de la semilla completa madura (SEED).
- Tres réplicas de los tejidos de la testa y el endospermo (T+End).
- Tres réplicas del tejido del embrión (EMBR).

La única comparación que podemos hacer en este caso es la de T+End frente a EMBR y usar la semilla completa para verificar la presencia del transcrito, pero no su cuantificación, dado que la contribución de embrión y endospermo es muy diferente.

### 3.3. Extracción de proteínas

En este trabajo, una de las cuestiones que se plantean es si existen diferencias en la semilla, ya sea entre los tejidos o los días de maduración. Por eso, llevamos a cabo un experimento de extracción de proteínas. Dicho experimento también sirve para comprobar que las proteínas se han extraído correctamente y no ha habido degradación.

Para la extracción de proteínas, partimos del mismo polvo que usamos para la secuenciación y lo resuspendemos en 1,5 ml de tampón de extracción (Tris-HCl pH 6.8 125mM, dodecil sulfato sódico (SDS) al 2 %, ditioneitol (DTT) al 1 %, Cocktail Inhibidor de Proteasas (CIP) (Sigma) 1/5000, fluoruro de fenilmetilsulfonilo (PMSF) 1/1000). Agitamos las muestras durante 20 minutos a 4 °C y las centrifugamos a  $12.500 \times g$  durante 10 minutos a la misma temperatura. Todo el procedimiento se llevó a cabo en frío para no desnaturar ni degradar las proteínas. Los restos sólidos precipitan y sedimentan, quedando la fracción soluble en el sobrenadante, que recuperamos con la ayuda de una jeringuilla con cuidado de no coger la capa de lípidos que se forma en la superficie. De ese sobrenadante preparamos alícuotas de 200  $\mu$ l en tubos Eppendorf, que almacenamos a -20 °C hasta su utilización.

### 3.4. Separación electroforética

Una vez terminamos con la extracción de proteínas, las separamos con una electroforesis SDS-PAGE. En cada uno de los pocillos cargamos 150  $\mu$ g de proteína por muestra. Utilizamos geles de poli(acrilamida) al 12 % de 15 pocillos con una capacidad de 15  $\mu$ l cada uno (Criterion™ TGX Stain-Free™ precast gels, 4-20 % BioRad). El paso previo a cargar las muestras fue la desnaturación de las proteínas. Para esta tarea, utilizamos un bloque térmico a 99 °C durante 1 min con la precaución de mantener los tubos Eppendorf abiertos mientras se calientan, ya que, debido a las altas temperaturas que se manejan y el incremento de presión dentro del tubo, estos podrían llegar a romperse. Una vez desna-

turalizadas las proteínas se procede a cargar el gel y a separar las muestras con un voltaje de 120 V durante 1 h y 30 min. Una vez terminada la electroforesis, escaneamos los geles con el sistema Gel Doc<sup>TM</sup> Imager (Bio-Rad) con el *software* Image Lab con activación mediante luz UV durante 1 min. De esta forma, obtuvimos el perfil de expresión de proteínas y comprobamos que todas las calles de electroforesis tenían la misma cantidad de proteína (normalización). Es necesario teñir las proteínas con azul de Coomassie antes de escanear los geles (Brunelle and Green 2014).

### **3.5. Análisis bioinformáticos**

Los datos que hemos utilizado de Picual mapeados sobre el transcriptoma de Picual y el de acebuche fueron limpiados, mapeados y cuantificados por transcrito por Rocío Bautista (Unidad de Bioinformática – SCBI). Durante el desarrollo de este proyecto he aprendido a utilizar las herramientas bioinformáticas y he limpiado y mapeado los datos sobre el genoma, que son los que usaremos en el atlas de olivo.

#### **3.5.1. Análisis de calidad**

FastQC (Andrews 2010) es una herramienta rápida y ampliamente usada de control de calidad para datos de secuencias de alto rendimiento. Hemos usado la versión 0.11.9 para analizar los datos de RNA-Seq antes y después de la limpieza para comprobar que se ha realizado correctamente. Hemos usado la versión 1.11 de MultiQC (Ewels et al. 2016) para crear un único informe a partir de los generados por FastQC, lo que facilita la visualización de los resultados.

#### **3.5.2. Preprocesamiento de las lecturas**

Hemos usado Trimmomatic (Bolger et al. 2014) para recortar las lecturas de Illumina. Hemos empleado la versión 0.39 con los ajustes por defecto para limpiar las lecturas brutas, eliminar fragmentos con baja calidad y adaptadores.

#### **3.5.3. Mapeo y cuantificación**

Hemos usado HISAT2 (Kim et al. 2019) para mapear lecturas de NGS sobre un genoma de referencia (olivo de la variedad Picual la mayoría de las veces). Hemos usado la versión 2.1.0 y los ajustes por defecto. Para manejar los resultados de mapeo hemos usado Samtools (Danecek et al. 2021) versión 1.13.

La cuantificación de los genes la hemos hecho con StringTie (Pertea et al. 2015) a partir de los archivos generados por Samtools. La versión utilizada ha sido la 1.3.3b, con los

valores por defecto.

#### 3.5.4. Determinación de ortologías

Hemos determinado los ortólogos proteicos de Picual con *Arabidopsis*, con las variedades de olivo Farga y Arbequina y con el acebuche con el *software Full-LengtherNext* (Seoane 2016). La ortología con *Arabidopsis thaliana* se ha hecho por ser la planta modelo mejor anotada, que nos dará más información que las anotaciones de olivo. Los criterios utilizados son los que vienen por defecto y la versión empleada ha sido la 1.0.1.

Todos los programas mencionados hasta ahora han sido ejecutados en el supercomputador Picasso. De aquí en adelante se harán en mi ordenador personal.

#### 3.5.5. Expresión diferencial

Hemos utilizado la herramienta web iDEP 0.95 (Ge et al. 2018) para analizar datos de RNA-Seq para obtener los genes con expresión diferencial (DEG). Para ello, hemos llevado a cabo un filtrado (mínimo 1 CPM en 2 librerías) para eliminar los genes que no se expresan o se expresan poco y nos introducirían ruido (variabilidad). La normalización de los recuentos se hizo con EdgeR. Después, hacemos un análisis de componentes principales (PCA) para comprobar que las muestras se agrupan por condición experimental. Finalmente, la expresión diferencial se hizo con un cambio de expresión de 2 y una FDR < 0,05. El resto de los valores son los que vienen por defecto. Para poder trabajar con más fluidez, he hecho una instalación local de esta herramienta (<https://github.com/iDEP-SDSU/idep>) en mi ordenador personal.

#### 3.5.6. Análisis funcionales

Es recomendable hacer una representación funcional de las listas de genes para saber en qué procesos celulares están implicados a partir de los términos GO y las vías KEGG (Ashburner et al. 2000). Para ello hemos usado REVIGO (Supek et al. 2011), que resume listas de términos GO encontrando un subconjunto no redundante representativo.

Este análisis debe completarse con un análisis de enriquecimiento funcional en el que se detecten qué procesos biológicos se ven específicamente alterados por los DEG de nuestro análisis. Para ello hemos usado la herramienta web ShinyGO 0.76 (Ge et al. 2020) con los valores por defecto.

#### 3.5.7. Genomas y proteomas

Los genomas, transcriptomas y proteomas que hemos utilizado en este TFM se resu-

men en la tabla 4.

**Tabla 4.** Genomas, transcriptomas y proteomas que hemos utilizado en este trabajo.

Organismo	Tipo	Enlace de descarga
<i>Arabidopsis thaliana</i>	CDS	<a href="http://ftp.ensemblgenomes.org/pub/plants/release-53/fasta/arabidopsis_thaliana/cds/Arabidopsis_thaliana.TAIR10.cds.all.fa.gz">http://ftp.ensemblgenomes.org/pub/plants/release-53/fasta/arabidopsis_thaliana/cds/Arabidopsis_thaliana.TAIR10.cds.all.fa.gz</a>
<i>Arabidopsis thaliana</i>	Proteínas	<a href="http://ftp.ensemblgenomes.org/pub/plants/release-53/fasta/arabidopsis_thaliana/pep/Arabidopsis_thaliana.TAIR10.pep.all.fa.gz">http://ftp.ensemblgenomes.org/pub/plants/release-53/fasta/arabidopsis_thaliana/pep/Arabidopsis_thaliana.TAIR10.pep.all.fa.gz</a>
<i>Olea europaea</i> var. Picual	CDS	<a href="https://genomaolivar.dipujaen.es/downloads/Sequences/Olea_europaea_CDS_v061.fasta.gz">https://genomaolivar.dipujaen.es/downloads/Sequences/Olea_europaea_CDS_v061.fasta.gz</a>
<i>Olea europaea</i> var. Picual	Proteínas	<a href="https://genomaolivar.dipujaen.es/downloads/Sequences/Olea_europaea_proteins_v061.fasta.gz">https://genomaolivar.dipujaen.es/downloads/Sequences/Olea_europaea_proteins_v061.fasta.gz</a>
<i>Olea europaea</i> var. Picual	Genoma	<a href="https://genomaolivar.dipujaen.es/downloads/Annotations/Olea_europaea_gene_models_v061.gff.zip">https://genomaolivar.dipujaen.es/downloads/Annotations/Olea_europaea_gene_models_v061.gff.zip</a>
<i>Olea europaea</i> var. Arbequina	RNA	<a href="https://download.cncb.ac.cn/gwh/Plants/Olea_europaea_Olive_GWHAOPM00000000/GWHAOPM00000000.RNA.fasta.gz">https://download.cncb.ac.cn/gwh/Plants/Olea_europaea_Olive_GWHAOPM00000000/GWHAOPM00000000.RNA.fasta.gz</a>
<i>Olea europaea</i> var. Arbequina	Proteínas	<a href="https://download.cncb.ac.cn/gwh/Plants/Olea_europaea_Olive_GWHAOPM00000000/GWHAOPM00000000.Protein.faa.gz">https://download.cncb.ac.cn/gwh/Plants/Olea_europaea_Olive_GWHAOPM00000000/GWHAOPM00000000.Protein.faa.gz</a>
<i>Olea europaea</i> var. Farga	CDS	<a href="https://denovo.cnag.cat/filebrowser/download/939">https://denovo.cnag.cat/filebrowser/download/939</a>
<i>Olea europaea</i> var. Farga	Proteínas	<a href="https://denovo.cnag.cat/filebrowser/download/930">https://denovo.cnag.cat/filebrowser/download/930</a>

### 3.5.8. Agrupamiento de genes

El agrupamiento (*clustering*) es el método de aprendizaje computacional no supervisado más extendido en el análisis de datos. Es una técnica de segmentación de datos que los divide en varios grupos en función de un parámetro de similitud. El agrupamiento de todos los posibles DEGs (esto es todos los genes que aparezcan como DEG en al menos una de las comparaciones) es la primera etapa para proponer la correulación de los genes del mismo grupo. Hemos trabajado con los logaritmos de los recuentos para aplanar el comportamiento de los genes que no cambien mucho. Luego los hemos filtrado por un coeficiente de variación elevado ( $CV > 20\%$ ). Los datos logarítmicos de expresión de los DEG que pasan este filtro los escalamos a una distribución gaussiana con la función *scale()* del paquete *base* (v 3.6.2) de *R*. Con estos valores se calculan las distancias entre los perfiles de expresión con la función *dist()* del paquete *stats* (versión 3.6.2) con el méto-

do euclidiano. Si los datos no estuvieran escalados, las distancias dependerían el nivel de expresión.

Para determinar los límites de cada *cluster* se utilizan diferentes métodos de vinculación (*linkage*). El método a elegir dependerá del comportamiento de los datos. Hemos usado la función *agnes()* del paquete *cluster* (v 2.1.3) para seleccionar el método más apropiado para formar el dendrograma con la función *hclust()* del paquete *stats*, que con los datos de este trabajo siempre es el método de Ward. También determinamos el número óptimo de *clusters* con la función *NbClust()* del paquete *NbClust* (v 3.0.1) que se lo pasamos a la función *cutree()* del paquete *dendextend* (v 1.15.2) para asignar los genes que corresponden a cada *cluster*. Esta estrategia tiene el inconveniente de formar grupos muy pequeños, por lo que también utilizamos la función *cutreeDynamic()* del paquete *dynamicTreeCut* (v 1.63-1) para formar *clusters* con al menos 10 genes que tienen más sentido biológico.

Una vez determinados los *clusters*, representamos sus perfiles de expresión para lo que necesitamos calcular la mediana de cada *cluster* en cada muestra. Hemos usado la función *aggregate()* del paquete *stats* (versión 3.6.2).

Con los genes de cada *cluster* hacemos entonces un análisis de información funcional tal como hemos descrito más arriba para saber en qué están involucrados.

### 3.5.9. Construcción de redes

Una red es una construcción matemática de nodos y aristas. Gráficamente, las redes se representan como un conjunto de componentes (nodos) y sus relaciones representadas por enlaces (aristas) que conectan pares de nodos. Las relaciones pueden ser interacciones físicas, similitud de secuencia, correlación de la expresión, etc. En este trabajo usamos la correlación entre los perfiles de expresión obtenida con la función *cor()* del paquete *stats* para formar estas redes con las funciones del paquete *igraph* (v 1.3.1). Usamos como umbral de correlación con el método de Pearson una  $r = 0,75$  porque  $r^2 = 0,56$  indica que más de la mitad (en concreto el 56 %) de la varianza de ambos perfiles está compartida. Obtenemos así una matriz de adyacencia para representar la fuerza de la conexión de cada par de nodos (genes). Finalmente, la red formada de cada *cluster* se analiza con el índice de Kleinberg gracias a la función *hub\_score()* para determinar las comunidades que la forman.

### 3.5.10. Construcción del atlas de expresión del olivo (OliveAtlas)

El atlas de expresión del olivo contiene un conjunto de herramientas que nos permiten

visualizar los datos de expresión que se han analizado en este TFM. Este atlas se ha implementado usando EasyGDB (Fernandez-Pozo and Bombarely 2022) en un servidor virtual local que estará disponible en <https://oliveatlas.uma.es/> tras su publicación. Para hacer el atlas con las secuencias de semilla y polen, mapeamos las lecturas limpias, y obtuvimos la expresión de los genes tal como se cuenta más arriba. Para construir el atlas, hemos instalado los recursos de EasyGDB en un contenedor *Docker* y hemos colocado las tablas en formato CSV de los datos de expresión en el directorio *scr/expression\_data*.

### 3.5.11. Identificación de alérgenos en Picual

Hemos utilizado las secuencias codificantes disponibles en el NCBI para todos los alérgenos, menos para el *Ole e 4* y *Ole e 7* de los que solo conocemos la secuencia de aminoácidos (tabla 3) y las hemos comparado contra el transcriptoma de Picual (tabla 4) con BlastN (salvo para *Ole e 4* y *Ole e 7* que hemos utilizado BlastP contra el proteoma de Picual). Representamos los perfiles de expresión a lo largo del desarrollo polínico de los alérgenos determinados.

### 3.5.12. Genes de la autoincompatibilidad esporofítica

En el género *Brassica* se han descrito seis genes (tabla 1) que tienen un papel en la autoincompatibilidad esporofítica (Yamamoto and Nishio 2014). Hemos determinado sus ortólogos proteicos en Picual y hemos estudiado sus perfiles de expresión.

### 3.5.13. Genes imprescindibles para el desarrollo del tubo polínico

A partir de la lista de genes de *Arabidopsis* que son imprescindibles para el desarrollo del tubo polínico que hemos mencionado en el apartado 1.5 de la introducción, hemos determinado los ortólogos de Picual para cada uno de ellos comparándolos con BLASTP contra el proteoma de picual. En la tabla 2 están los genes, su descripción y el número de acceso que hemos usado para buscar el ortólogo.

## 4. Resultados

### 4.1. Mapeo de las lecturas

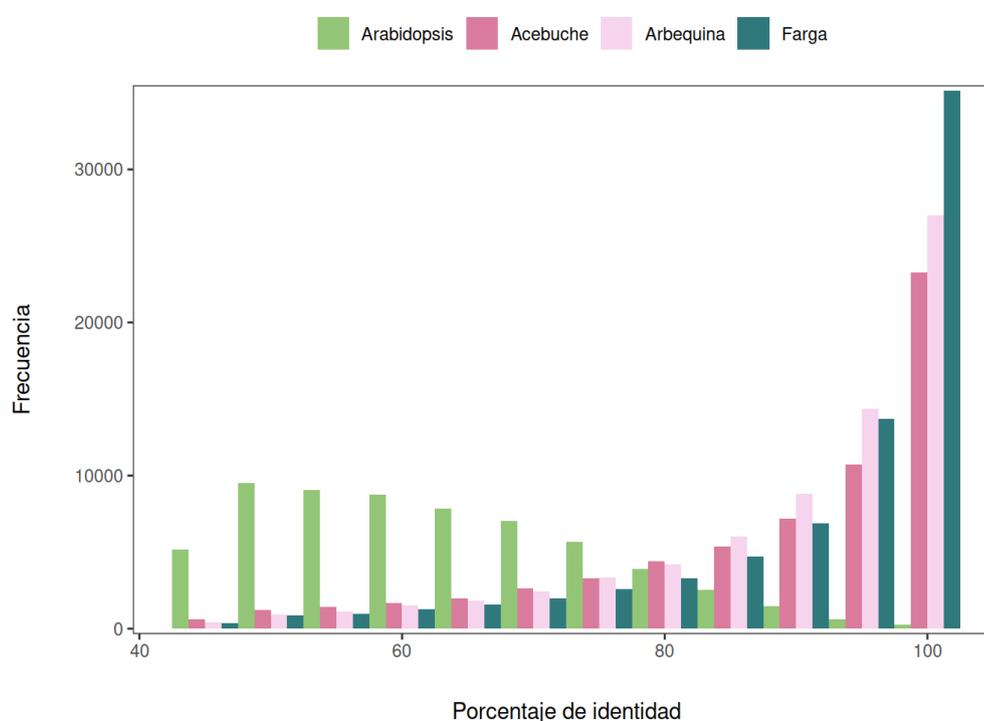
Pese a que el material secuenciado es de la variedad Picual, el acebuche es la especie silvestre de la que surgen el resto de variedades (Besnard et al. 2001, 2013; Diez et al. 2015) por lo que sería de esperar que la mayoría de los genes presentes en Picual lo estén también en el acebuche, con mayor o menor número de copias. Por eso nos planteamos la posibilidad de mapear las lecturas sobre el transcriptoma de acebuche. Sin embargo, la tasa de alineamiento global sobre el olivo silvestre es visiblemente peor que sobre Picual (~75 % frente a ~6 %, tabla 5). Estos datos nos indican que lo más adecuado es alinear las lecturas contra Picual.

**Tabla 5.** Número de lecturas brutas de cada muestra y tasa de alineamiento global sobre el transcriptoma cuando se usa Picual o acebuche como referencia para el mapeo.

Muestra	Número de lecturas	Tasa de alineamiento global en Picual	Tasa de alineamiento global en acebuche
SEED	84.524.144	74,36 %	5,52 %
T+End_1	92.008.804	78,13 %	5,73 %
T+End_2	77.082.748	75,94 %	5,63 %
T+End_3	91.655.172	74,47 %	5,58 %
EMBR_1	75.888.108	75,48 %	5,33 %
EMBR_2	85.464.854	74,32 %	5,65 %
EMBR_3	85.065.344	73,88 %	5,69 %
P0H_1	82.902.018	72,89 %	6,53 %
P0H_2	88.165.266	74,07 %	6,72 %
P1H_1	80.002.740	70,01 %	5,95 %
P1H_2	81.535.524	69,61 %	5,91 %
P1H_3	86.619.832	72,07 %	6,58 %
P3H_1	76.360.584	74,02 %	6,67 %
P3H_2	86.769.778	73,36 %	6,57 %
P3H_3	94.148.126	74,07 %	6,73 %
P6H_1	90.185.476	74,49 %	6,79 %
P6H_2	95.480.454	74,48 %	6,80 %
P6H_3	85.397.554	73,97 %	6,75 %

Estos datos de mapeo nos hacen preguntarnos si el acebuche que se ha utilizado para

este es realmente un olivo o si se trata de otra especie diferente, para lo que hemos determinado los ortólogos proteicos de Picual con otras variedades de olivo, como son Farga y Arbequina, y con *Arabidopsis* como control externo (y para su posterior uso en las herramientas de análisis funcional), con la herramienta *Full-LengthNext*. El histograma de los porcentajes de identidad entre estos proteomas y el de Picual (figura 3) indica que Farga es la más parecida a Picual, seguida por Arbequina y después por el acebuche. Como cabía esperar, el acebuche mantiene un porcentaje de identidad con Picual considerablemente mayor que el de *Arabidopsis*, por lo las diferencias observadas a nivel de nucleótidos entre Picual y acebuche se ven muy atenuadas a nivel de aminoácidos.



**Figura 3.** Histograma del porcentaje de identidad de los ortólogos proteicos de Picual con *Arabidopsis* y las variedades de olivo acebuche, Arbequina y Farga.

## 4.2. Análisis del desarrollo del tubo polínico

### 4.2.1. Filtrado de genes

Como contamos con muestras del desarrollo del tubo polínico de Picual a diferentes tiempos (0 h, 1 h, 3 h y 6 h) ya mapeadas sobre el transcriptoma, y dado que Picual contiene 81.484 modelos de gen debemos retirar todos los que no tengan una expresión clara antes de la correspondiente normalización. Podemos hacerlo de dos formas: (1) filtrar y

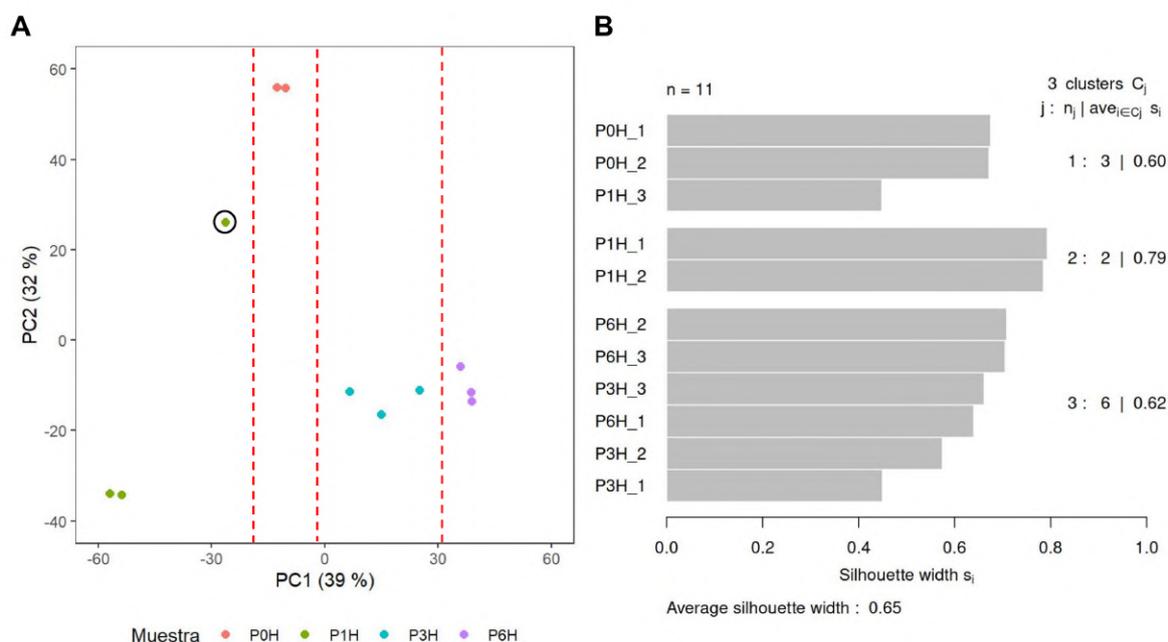
normalizar los recuentos de cada comparación por separado o (2) filtrar y normalizar los recuentos de todas las muestras a la vez. El primer enfoque resulta de interés cuando lo que queremos es estudiar los genes en sí (para determinar genes expresados diferencialmente o biomarcadores, por ejemplo). El segundo lo utilizamos para ver funciones. Hemos utilizado la herramienta web iDEP.95 y en el filtrado nos quedamos con aquellos genes que tienen al menos 1 CPM en cada 2 o 3 librerías, en función de si la muestra de P0H (con solo dos réplicas) interviene o no. De todos los genes, 19.828 se expresan a 0 h, 20.544 a 1 h 20.811 a 3 h y 21.024 a 6 h tras un filtrado a 1 CPM en 2 librerías.

#### 4.2.2. Congruencia de las réplicas

A continuación, hacemos un análisis preliminar de los datos mediante componentes principales (PCA) para determinar si las muestras secuenciadas se agrupan por condición experimental. En la figura 4A observamos que las muestras se separan bien por condición experimental en la primera dimensión, lo que indica que son bastante diferentes entre sí. Dicha separación no es tan evidente con las muestras de P3H y P6H, de lo que podemos deducir que estas dos condiciones se parecen más entre sí de lo que se parecen al resto. También merece la pena destacar que una de las réplicas de P1H se aleja del comportamiento de las otras dos réplicas y se sitúa más cerca de las de P0H. Como solo se dispone de tres réplicas de P1H, no parece coherente eliminarla. Además, hemos de tener en cuenta la propia variabilidad biológica de las muestras y que dos de las réplicas se tomaron en una campaña de muestreo y la tercera corresponde a otra distinta, lo que queda claro que influye.

Para comprobar si estas diferencias pueden afectar al resultado final, sometimos las muestras a una agrupación jerárquica con el método *ward.D2* que se valoró en un gráfico de silueta (figura 4B) para determinar si alguna observación (muestra) presenta una contribución negativa a su grupo. Cuando indicamos que las muestras deben agruparse en 4 *clusters*, uno por condición experimental, una de las réplicas de P3H no contribuye ni positiva ni negativamente, las otras dos forman un grupo con las tres réplicas de P6H (informe complementario 2), y una de las réplicas de P1H se agrupa mejor con las de P0H, como ya intuíamos en el PCA (figura 4A). Si tomamos esta información y hacemos 3 *clusters* vemos que no hay cambio con lo que respecta a las muestras de P0H y P1H, pero, en esta situación, la réplica de P3H que antes se "agrupaba" sola pasa a formar parte del *cluster* al que pertenecen sus otras réplicas y las de P6H, contribuyendo positivamente (figura 4B). Además, los otros dos *clusters* se agrupan mejor que antes (de un valor de 0,57

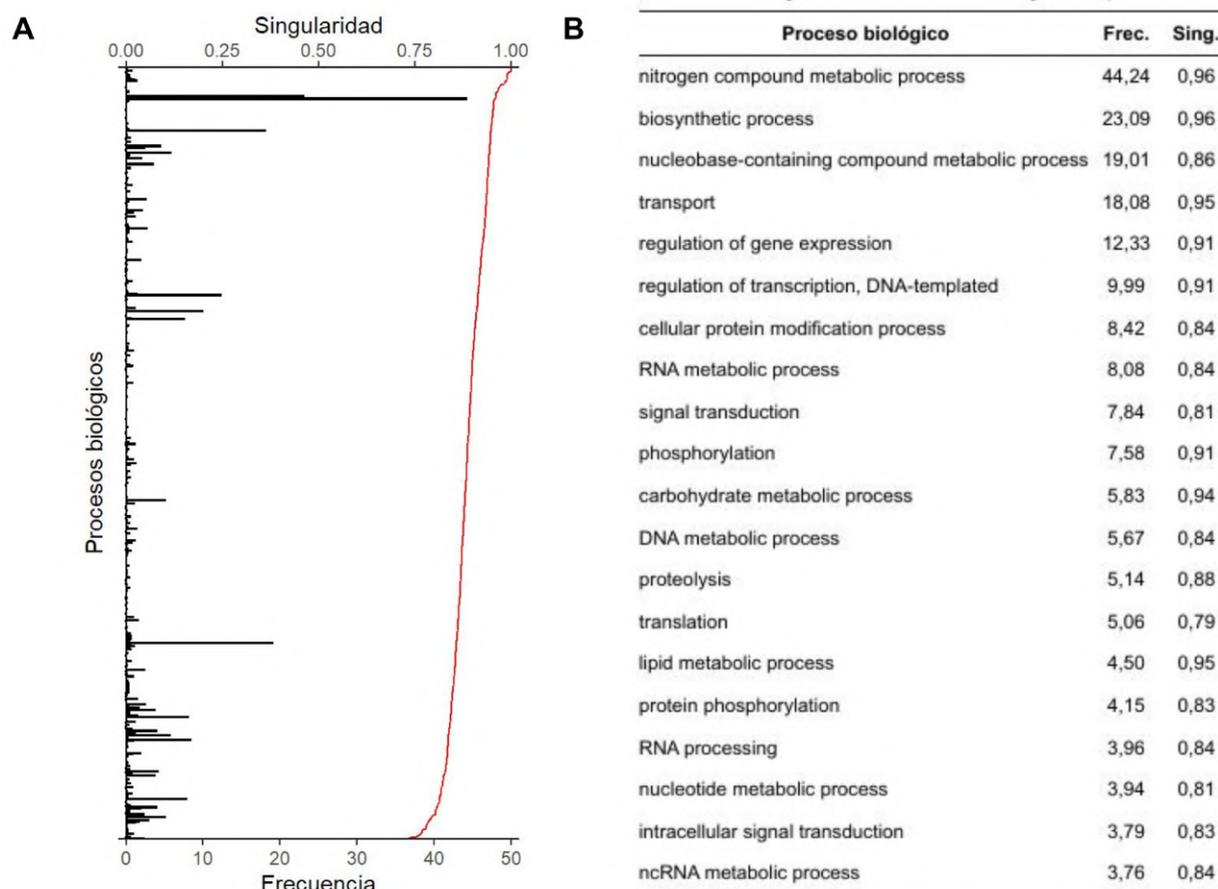
a 0,60 para el primer *cluster* y de 0,72 a 0,79 para el segundo). Esto nos indica que no debemos eliminar ninguna de nuestras muestras, ya que todas ellas contribuyen de forma positiva y no van a empeorar la calidad de los análisis, aunque algunas no sean unas réplicas exactas.



**Figura 4.** La congruencia de las réplicas de las muestras del desarrollo del tubo polínico se ilustra con un (A) PCA en el que se han usado las líneas rojas discontinuas para definir los agrupamientos, y se señala la muestra de P1H que se parece más a las P0H. (B) Gráfico de silueta de las muestras cuando se agrupan en tres *clusters*. Independientemente de cómo se agrupan, todas las muestras contribuyen positivamente al análisis.

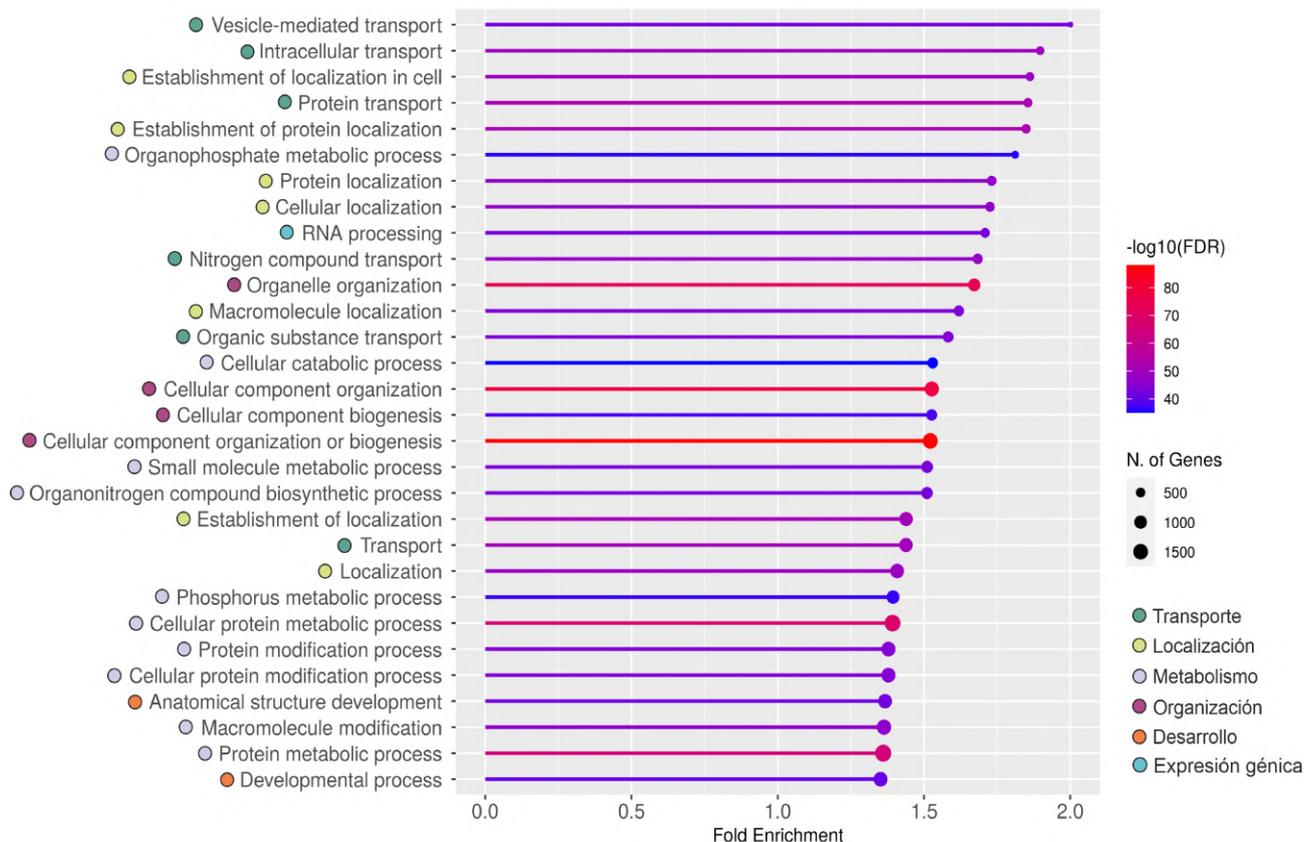
#### 4.2.3. Funciones celulares del polen

Para saber los procesos biológicos que están implicados en el desarrollo del tubo polínico llevamos a cabo un análisis funcional de los 22.937 genes con expresión fiable con la herramienta web REVIGO. Hemos usado los términos GO (Gene Ontology) de los ortólogos en Arabidopsis de los genes de olivo determinados más arriba. Aunque en este proceso perdemos información porque no todos los genes de olivo tienen un ortólogo claro en Arabidopsis, los resultados obtenidos son mejores que si se usa la anotación del borrador de Picual (resultados no mostrados). De 22.937 genes obtenemos 19.741 términos GOs únicos. Si representamos gráficamente los resultados obtenemos la figura 5A donde vemos que la mayoría de los procesos biológicos tienen una frecuencia baja y que esta no va ligada a su singularidad (relevancia funcional para el proceso analizado). En la figura 5B se recogen los 20 primeros más frecuentes y su índice de singularidad.



**Figura 5.** Procesos biológicos que se dan en los genes expresados en el polen representados (A) en función de su frecuencia (barras) y su singularidad (línea roja), o como tabla (B) con los 20 más frecuentes.

Puesto que son muchos los procesos biológicos obtenidos para el polen, es conveniente analizar aquellos que se enriquecen (figura 6). De estos, el que se enriquece más ( $n = 306$ ,  $FE = 2,00$ ) es el correspondiente al transporte mediado por vesículas, seguido por el transporte intracelular ( $n = 418$ ,  $FE = 1,90$ ). También se enriquecen muchos procesos relacionados con la localización y la organización, así como el metabolismo.

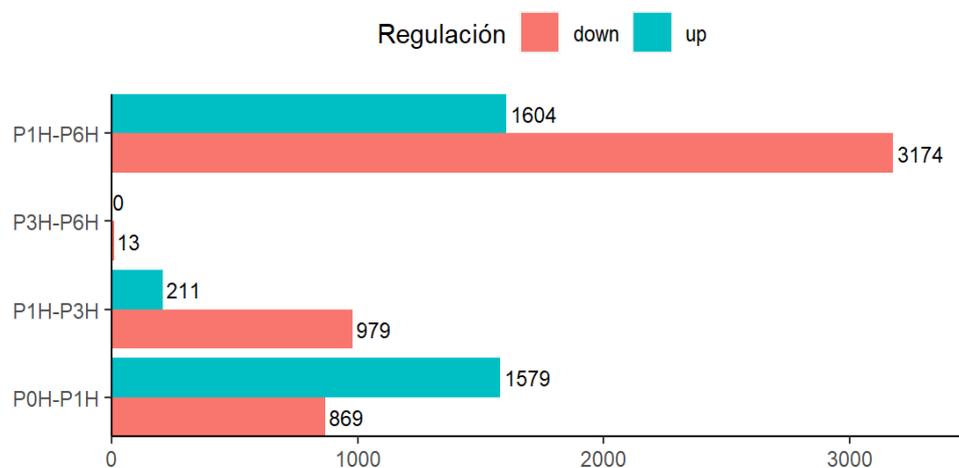


**Figura 6.** Enriquecimiento funcional de los genes que se expresan en el polen. En el eje de abscisas se representa las veces que se enriquece el término. En el eje de ordenadas tenemos los procesos biológicos que se enriquecen. El tamaño del punto viene determinado por el número de genes que pertenecen al proceso biológico en cuestión. Por último, el color indica la significación estadística (FDR) del enriquecimiento (más significativo cuanto más rojo), sabiendo que los procesos se seleccionaron por tener un FDR < 0,05. (Obtenida con ShinyGO 0.76).

#### 4.2.4. Expresión diferencial del desarrollo del tubo polínico

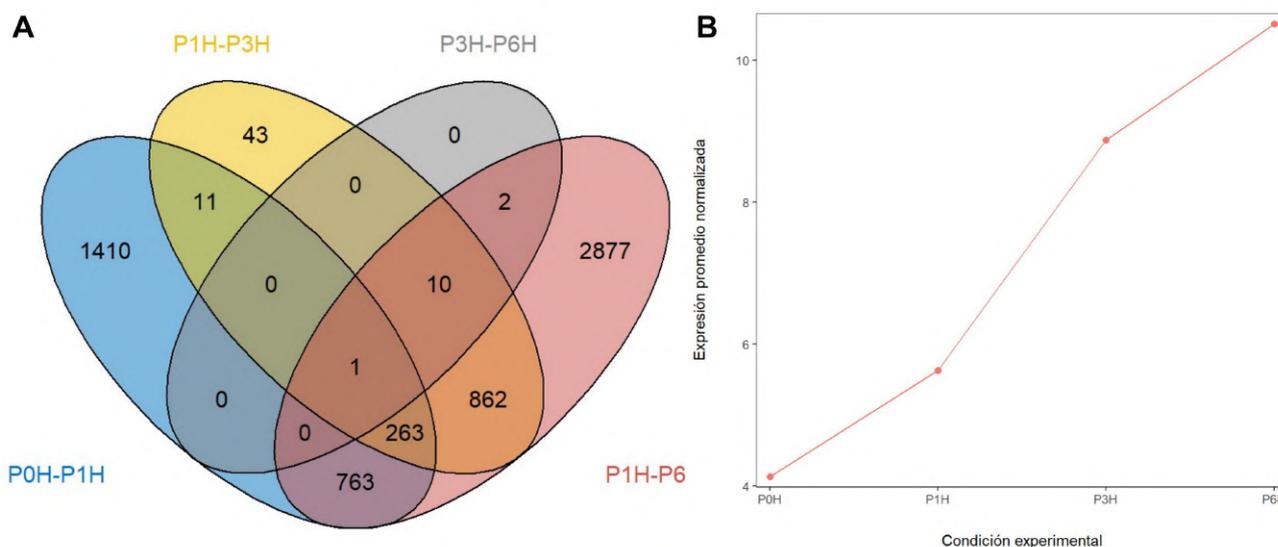
Analizamos ahora la expresión diferencial de los genes a lo largo del desarrollo del tubo polínico (figura 7). La primera comparación fue entre el polen seco (0 h) y a 1 h de la hidratación, con lo que evaluamos el impacto de la hidratación. Este análisis definió un número elevado de genes activados (1.579) y uno más reducido de genes reprimidos (869). Para definir los cambios en los niveles de transcripción que se producen después de la hidratación y durante el crecimiento del tubo polínico, comparamos los puntos temporales consecutivos P1H-P3H y P3H-P6H, así como P1H-P6H para saber qué cambia entre el comienzo del crecimiento del tubo hasta el momento en que se considera que ya está totalmente formado. No parece intuitivo que haya más genes reprimidos que activados, pero eso no es más que el reflejo de cómo se han hecho las comparaciones. Los genes "down" de la comparación P1H-P3H indican que se están activando a 3 h, y lo mismo con las

comparaciones P1H-P6H y P3H-P6H lo que sugiere que aunque el polen es transcripcionalmente activo en seco, hay una enorme cantidad de genes que se activan durante el desarrollo del tubo. El hecho de que en la comparación P3H-P6H obtengamos un número muy reducido de genes (13) activados a 6 h ("down") sugiere que al cabo de tres horas ya casi toda la maquinaria del tubo está en marcha. Curiosamente, cuando se consideraba que las muestras de 3 h y 6 h formaban una sola condición experimental, se obtuvieron menos genes con expresión diferencial que intervienen en procesos poco significativos (resultados no mostrados). Todos los DEGs se pueden consultar en el informe complementario 4.



**Figura 7.** Número de DEGs para cada una de las comparaciones. En azul aparecen los DEGs que se sobreexpresan para la primera condición de cada par de comparaciones. En rojo, los DEGs que se sobreexpresan para la segunda condición.

Podemos visualizar los DEGs de estas comparaciones en un diagrama de Venn (figura 8A), para determinar cuántos hay en común entre ellas y, además, cuántos DEGs son únicos en cada comparación. Los cambios de expresión más bruscos se dan en la comparación P1H-P6H (2.877 DEGs exclusivos). Además, hay un DEG común a las cuatro comparaciones, el gen *Oleur061Scf1025g00027.1*. Este gen es un DEG positivo en todas las comparaciones, por lo que su expresión va en aumento significativo a medida que se desarrolla el tubo polínico (figura 8B). Está anotado en OliveTreeDB como poseedor de un dedo de zinc (InterPro), un RING-type (TrEMBL) y que se parece significativamente a BOI-related E3 ubiquitin-protein ligase 1 (SwissProt).

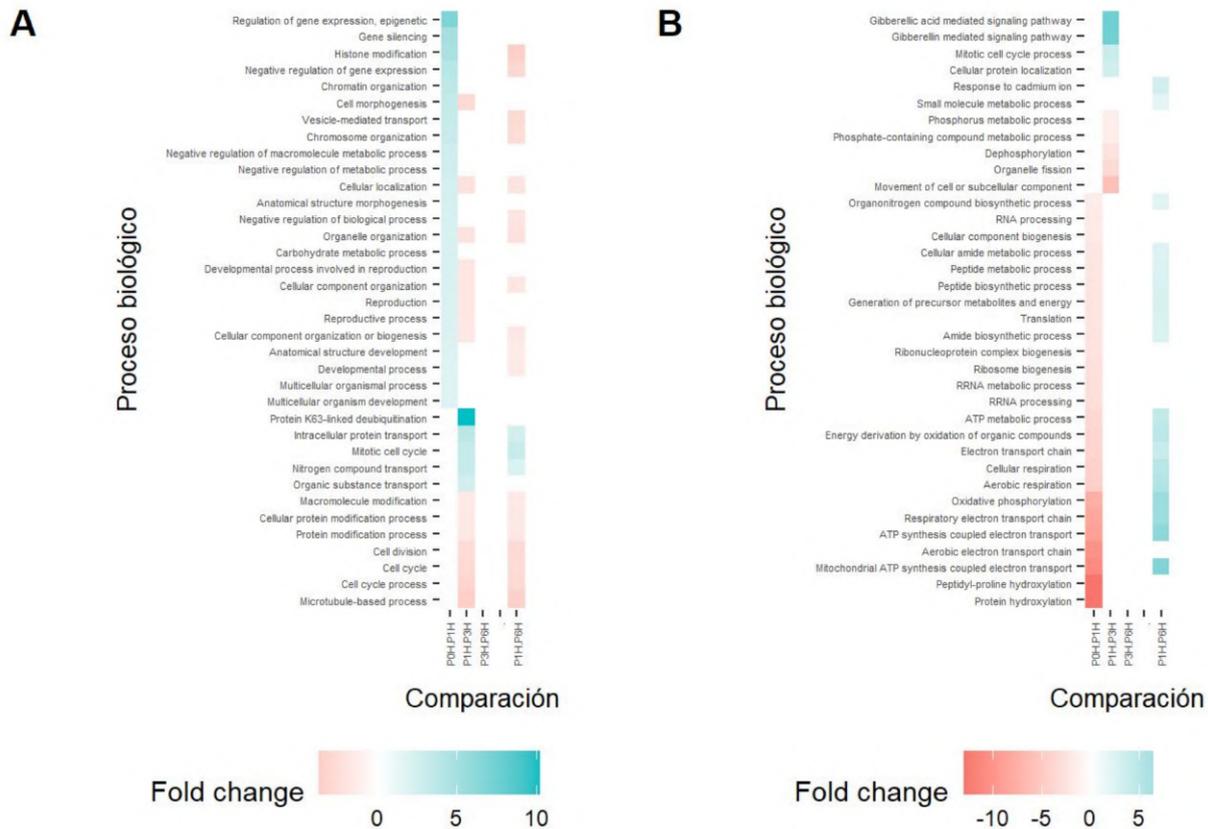


**Figura 8.** DEG del desarrollo del tubo polínico distribuidos (A) en un diagrama de Venn con cada comparación. (B) Perfil de expresión del DEG común a todas las comparaciones.

Para saber en qué funciones están significativamente implicados los DEG, llevamos a cabo un análisis de enriquecimiento funcional de los procesos biológicos de la Ontología Genética (GO). Para ello, utilizamos la herramienta web ShinyGO v0.76 con los ortólogos de Picual en *Arabidopsis*. El resultado, cuyas gráficas de enriquecimiento biológico se recogen en el informe complementario 1 y que se resumen en la figura 9, se puede resumir de la siguiente manera:

1. Con la hidratación del polen (P0H-P1H), a tiempo 0 h abundan los procesos relacionados con la regulación de la expresión génica, el desarrollo, el metabolismo, el transporte y la reproducción (figura S1). Los que tienen un FDR más significativo son los relacionados con la organización de los componentes celulares. A 1 h aumentan los procesos relacionados con la producción de energía (ATP) y la traducción (figura S2).
2. En el paso de 1 h a 3 h, los DEGs activados a 1 h están relacionados con procesos de transporte, localización intracelular, desubiquitinación y señalización mediada por el ácido giberélico (figura S3). En 3 h, los procesos que aparecen son los que intervienen en el movimiento celular, la mitosis, la división celular y los microtúbulos (figura S4). El más significativo es la organización de los componentes celulares. Esto indica con claridad que el tubo se está alargando y que por él empiezan a desplazarse los componentes celulares, e incluso que la célula espermática en alguno de ellos ha podido comenzar a dividirse.
3. Cuando estudiamos todo el cambio entre 1 h y 6 h, vemos que a 1 h siguen abun-

dando los procesos relacionados con la energía (figura S5) y el más significativo es la biosíntesis de compuestos organonitrogenados (como las amidas). A 6 h encontramos procesos relacionados con el ciclo celular, la localización y el transporte (figura S6) y el más significativo es la organización de orgánulos, todo en consonancia con lo observado en la etapa anterior.



**Figura 9.** Resumen de los procesos biológicos que se van enriqueciendo (en rojo) durante el desarrollo del tubo polínico y los que van desapareciendo (en azul).

De todos los procesos enriquecidos, son comunes en todos los casos (12) aquellos relacionados con la organización celular, la localización, el transporte y el metabolismo. La comparación de P0H-P1H es la que más procesos únicos tiene (18) a destacar los de silenciamiento génico, organización de la cromatina y metabolismo del ARN ribosómico. Además, solo hay dos procesos que se dan exclusivamente en la comparación P1H-P6H, que son el proceso metabólico de las moléculas pequeñas y la respuesta al cadmio.

#### 4.2.5. Coexpresión de genes durante el desarrollo del tubo polínico

Para conocer mejor lo que pasa durante el desarrollo no basta con los genes que se activan o reprimen significativamente, sino que también se deben conocer los que se co-expresan porque podrían revelar el proceso común que gobierna su coexpresión. Para

ello habrá que comparar los perfiles de expresión y buscar los que más correlacionados estén, un procedimiento que empieza por el análisis de *clustering* jerárquico de los genes calificados como DEG en al menos una comparación y que presentan un coeficiente de variación (CV) de más de un 20 %, porque los genes que varían poco no tiene sentido correlacionarlos (la explicación detallada de cómo se ha realizado se puede encontrar en el informe complementario 2 que es el informe del análisis con R). El *clustering* jerárquico devuelve una matriz de distancias con la que se calcula un dendograma que "podamos" de forma dinámica para que los grupos formados tengan al menos 10 genes (a diferencia del método tradicional en el que se pueden obtener grupos con nada más que uno o dos genes). Determinamos que el número de *clusters* dinámicos es 6, con los números de genes de la tabla 6.

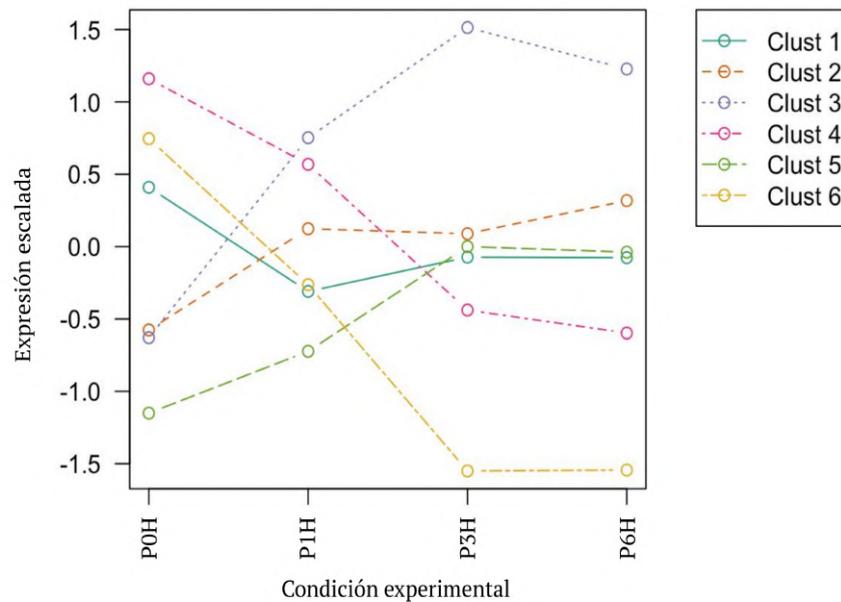
**Tabla 6.** *Clusters* del *clustering* jerárquico y número de genes que tiene cada uno.

<b>Cluster</b>	1	2	3	4	5	6
<b>Número de genes</b>	42	31	28	27	22	19

Representamos a continuación sus perfiles de expresión para ver qué comportamiento tienen (figura 10). Vemos que hay dos *clusters* (4 y 6) cuya expresión es elevada en el polen maduro (seco) y cae a lo largo del desarrollo del tubo polínico. También hay otros dos *clusters* (3 y 5) que tienen el comportamiento inverso. Aunque los datos se escalaron para evitar el sesgo por nivel de expresión, en ambas parejas de *clusters* se aprecia que uno tiene mayor expresión que el otro.

Analizamos funcionalmente los *clusters* para conocer en qué procesos se activan o reprimen los genes coordinadamente. Empezando por los genes que aumentan la expresión durante el proceso:

- El *cluster* 3 nos lleva a procesos de crecimiento de tubo polínico y otras respuestas a estímulos (calor, hipersensibilidad, salinidad y frío).
- El *cluster* 5 señala los procesos de citocinesis, metabolismo de lípidos, respuesta a heridas e hipersensibilidad, biosíntesis de almidón y transcripción.
- El *cluster* 2, que sube ligeramente la expresión con respecto al polen seco, participa en procesos de respuesta a heridas, transporte de nitrato y desfosforilación de proteínas.

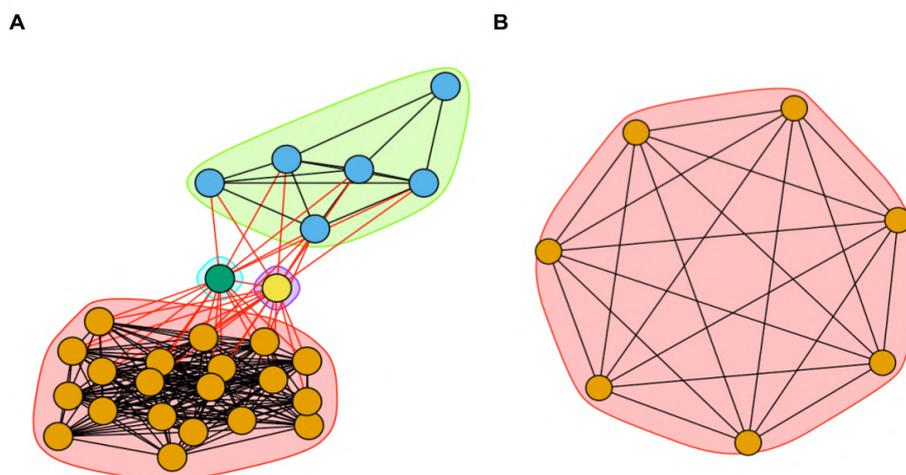


**Figura 10.** Perfiles de expresión de los seis *clusters* determinados por *clustering* jerárquico.

Los *clusters* que siguen el comportamiento opuesto son:

- El *cluster* 4 interviene en el plegamiento de las proteínas, transcripción y respuesta a estímulos (frío, hipoxia y estrés osmótico).
- El *cluster* 6 está relacionado con la biosíntesis del fosfatidilinositol, respuesta al estrés oxidativo y plegamiento de proteínas.
- Finalmente, el *cluster* 1, que solo baja ligeramente la expresión, interviene en procesos de organización del citoesqueleto, desarrollo de la raíz, señalización, plegamiento y fosforilación de proteínas, y respuestas a distintos estímulos (sal, luz UV-B y frío, principalmente).

En general se puede decir que el polen seco contiene toda una batería de actividades preparadas para responder a estímulos externos que le indiquen que ha de permanecer quiescente o germinar el tubo polínico, y que mantiene en reposo el genoma, y que cuando germinan se pone en marcha en primer lugar el mecanismo de producción de energía, el desenrollamiento del DNA para expresar los genes que sintetizan las proteínas necesarias de la pared celular y el tubo polínico, y la división de la célula espermática.



**Figura 11.** Representación de las comunidades que forman la red de co-expresión del *cluster 3* obtenido (A) por *clustering* jerárquico, en la que los nodos azules coinciden con el *cluster 9* (B) obtenido por k-medias, que está implicado en el desarrollo del tubo polínico.

Para afinar más el proceso, hemos determinado las posibles comunidades de genes dentro de cada *cluster*. De ellos, el *cluster 3* (figura 11A), que es el que interviene en procesos relacionados con el crecimiento del tubo polínico, cuenta con cuatro subgrafos. Uno está formado por 20 genes (rojo), relacionado con la respuesta a estreses y la modificación de histonas. Otro está formado por seis genes (verde) que interviene en los procesos relacionados con el crecimiento del tubo polínico, un grupo que también hemos identificado cuando se han hecho agrupaciones por k-medias (*cluster 9*, figura 11B e informe complementario 2). Las otras dos comunidades (azul y morado) están formadas por un solo gen que claramente sirven de coordinación para los otros dos grandes grupos. Por desgracia, se desconoce la naturaleza funcional de estos dos genes, aunque uno de ellos (Oleur061Scf0076g02032.1) es uno de los 55 genes que más se expresan en el polen (véase más adelante) que solo hemos visto que tiene un dominio de unión al DNA relacionado con la familia MYB según una proteína de *Salix suchowensis* (tabla 7). Esto sugiere fuertemente que este gen es uno de los principales reguladores de la elongación del tubo polínico, y que está sin caracterizar, al menos en el olivo. Del otro gen (Oleur061Scf5931-g00007.1) no tenemos más información en ninguna otra especie más allá de proteína predicha o desconocida.

Si comparamos los seis genes del subgrafo verde del *cluster 3* con los siete del *cluster 9*, apreciamos que hay cinco en común y tres diferentes (tabla 7).

**Tabla 7.** Genes de los *clusters* 3 y 9. Cinco son comunes entre ambos, el marcado con \* es único del *cluster* 3 y los marcados con \*\* solo aparecen en el *cluster* 9.

Gen	Término GO	Anotación
Oleur061Scf2388g00003.1	GO:0009860	Pollen tube growth
Oleur061Scf2178g06022.1	GO:0009860	Pollen tube growth
Oleur061Scf8557g00002.1		Zinc finger C <sub>2</sub> H <sub>2</sub> -type
Oleur061Scf0658g00034.1		Zinc finger, RING/FYVE/PHD-type
Oleur061Scf2783g01016.1		Zinc finger C <sub>2</sub> H <sub>2</sub> -type
*Oleur061Scf1025g00027.1		Zinc finger, RING/FYVE/PHD-type
**Oleur061Scf5931g00007.1		HTH Myb
**Oleur061Scf0076g02032.1		HTH Myb

La posible anotación de los genes marcados con \*\* se hizo buscando ortólogos más remotos con BlastP y obtuvimos que ambos genes se describen como miembro de la familia de reguladores transcripcionales Myb HTH en *Salix suchowensis* (p. ident. del 34 %).

#### 4.3. Autoincompatibilidad esporofítica en el olivo

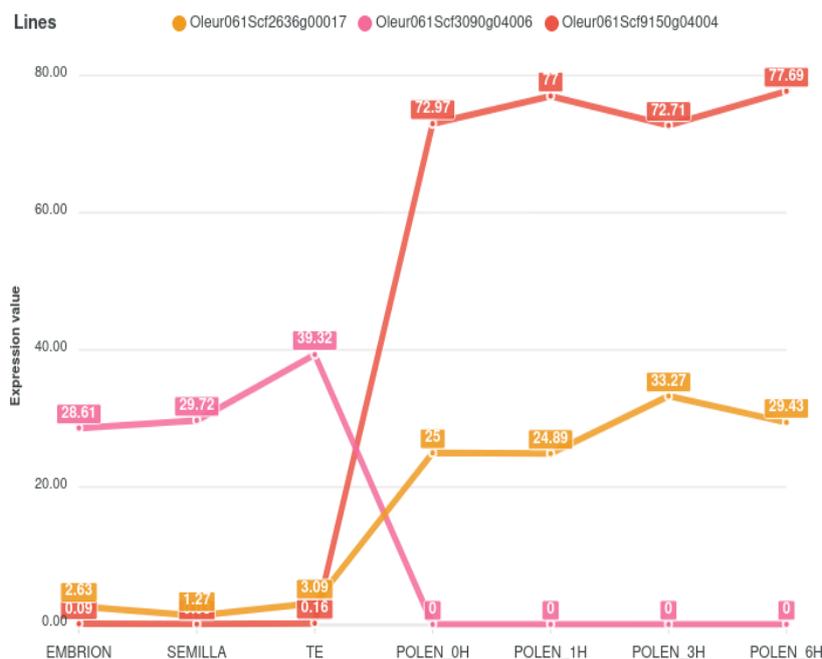
En el género *Brassica* se han descrito seis genes (tabla 1) que tienen un papel en la autoincompatibilidad esporofítica (Yamamoto and Nishio 2014). Para saber si, como proponen (Breton and Bervillé 2012; Breton et al. 2014; Koubouris et al. 2014), el olivo presenta este tipo de autoincompatibilidad, hemos determinado sus ortólogos proteicos en Picual (tabla 8) y hemos estudiado sus perfiles de expresión en el polen (figura 12), aunque todos menos *SCR* se detectan en el estigma.

**Tabla 8.** Genes de la SSI descritos en *Brassica*, sus ortólogos en Picual y la identidad.

Gen	Ortólogo en Picual	Identidad (%)
<i>SRK</i>	Oleur061Scf0072g00021.1	42,18 %
<i>SCR</i>	Oleur061Scf7007g03003.1	52,17 %
<i>MLPK</i>	Oleur061Scf3090g04006.1	67,32 %
<i>ARC1</i>	Oleur061Scf2636g00017.1	51,21 %
<i>THL-1</i>	Oleur061Scf9150g04004.1	55,86 %

De los seis genes descritos en la tabla 1, cinco tienen un ortólogo lejano en Picual. Curiosamente, dos de ellos se expresan en el polen (*ARC1* y *THL-1*) y uno en la semilla (*MLPK*) (figura 12), pero el transcrito de *SCR* no se detecta. Esto último confirma que el gen se debe estar expresando en la antera, pero que la proteína sí que se localiza en el

polen (Yamamoto and Nishio 2014), y que los genes de *ARC-1*, *THL-1* y *MLPK* realizan más funciones de señalización, además de la SSI.



**Figura 12.** Perfil de expresión de los ortólogos de olivo para los genes de *Brassica* responsables de la autoincompatibilidad gametofítica.

#### 4.4. Genes imprescindibles en el desarrollo del tubo polínico

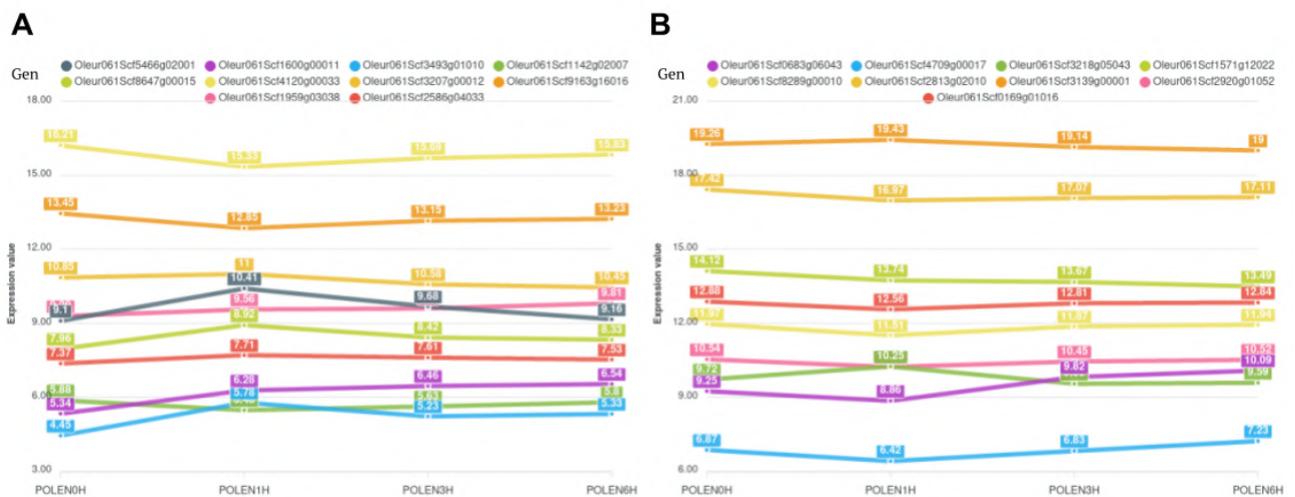
En la introducción se comentaron todas las funciones enzimáticas que se ha visto que son muy importantes para el tubo polínico en *Arabidopsis*. Hemos sacado de TAIR (<https://www.arabidopsis.org>) la lista de todos los que tienen actividad peroxidasa, esterasa, hidrolasa, lipasa, poligalacturonasa y pectina metil esterasa, así como las pectinas, y la callosa, y otros genes de *Arabidopsis* que se han descrito en la bibliografía que son necesarios para el desarrollo del tubo polínico (tabla 2) (Golovkin and Reddy 2003; Qin et al. 2009; Bolaños-Villegas et al. 2015; Dou et al. 2016; Simeunovic et al. 2016; Rui et al. 2020, 6; Beuder et al. 2022). Con todos ellos, hemos sacado de nuestra lista de ortólogos el gen de olivo que les corresponde (tabla 9) y hemos estudiado sus perfiles de expresión (figuras 13 y 14).

**Tabla 9.** Genes imprescindibles en el desarrollo del tubo polínico de *Arabidopsis* y sus ortólogos en Picual. Con objetivo de abreviar, hemos recogido solo aquellos que se expresan en nuestras muestras de polen.

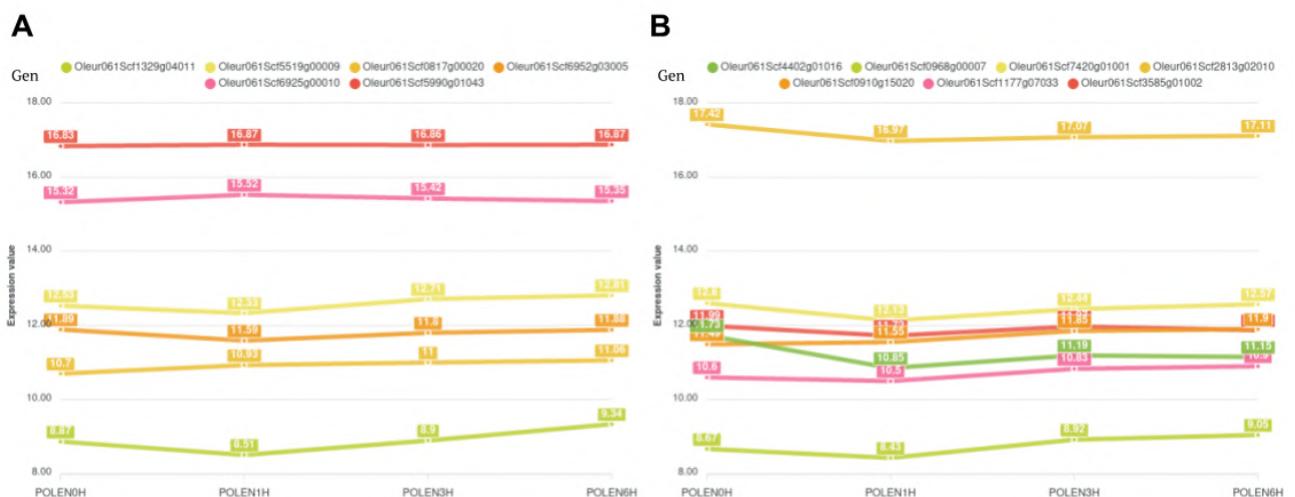
<b>Gen</b>	<b>Identificador en Arabidopsis</b>	<b>Identificador en Picual</b>	<b>Identidad (%)</b>
Peroxidasa	AT1G44970	Oleur061Scf3493g01010	76,49 %
Peroxidasa	AT4G16270	Oleur061Scf5466g02001	63,17 %
Esterasa	AT4G30140	Oleur061Scf2586g04033	41,24 %
Lipasa	AT4G29460	Oleur061Scf3207g00012	54,81 %
Lipasa	AT4G29470	Oleur061Scf3207g00012	47,62 %
Lipasa	AT4G30140	Oleur061Scf2586g04033	41,24 %
Hidrolasa	AT3G24350	Oleur061Scf1959g03038	63,25 %
Pectina	AT3G56960	Oleur061Scf9163g16016	65,82 %
Callosa	AT2G13680	Oleur061Scf4120g00033	79,43 %
Poligalacturonasa	AT4G20050	Oleur061Scf1142g02007	66,37 %
<i>BET11</i>	AT3G58170	Oleur061Scf8647g00015	77,31 %
<i>BET12</i>	AT4G14455	Oleur061Scf1600g00011	58,68 %
Pectin metil esterasa	AT1G69940	Oleur061Scf1571g12022	53,5 %
Pectin metil esterasa	AT2G47030	Oleur061Scf2813g02010	42,91 %
Pectin metil esterasa	AT2G47040	Oleur061Scf2813g02010	41,33 %
Pectin metil esterasa	AT2G47050	Oleur061Scf3139g00001	21,56 %
Pectin metil esterasa	AT5G07410	Oleur061Scf1571g12022	53,5 %
<i>COG6</i>	AT1G31780	Oleur061Scf2920g01052	72,84 %
<i>SEC1A</i>	AT1G02010	Oleur061Scf3218g05043	70,52 %
<i>NPG1</i>	AT2G43040	Oleur061Scf8289g00010	68,52 %
<i>NPGR1</i>	AT1G27460	Oleur061Scf4709g00017	53,76 %
<i>NPGR2</i>	AT4G28600	Oleur061Scf0683g06043	58,55 %
<i>AtCPK14</i>	AT2G41860	Oleur061Scf5519g00009	74,86 %
<i>AtCPK16</i>	AT2G17890	Oleur061Scf0817g00020	73,16 %
<i>AtCPK17</i>	AT5G12180	Oleur061Scf5990g01043	81,26 %
<i>AtCPK20</i>	AT2G38910	Oleur061Scf1329g04011	71,94 %
<i>AtCPK24</i>	AT2G31500	Oleur061Scf6925g00010	65,13 %
<i>AtCPK26</i>	AT4G38230	Oleur061Scf6952g03005	89,35 %
<i>AtCPK34</i>	AT5G19360	Oleur061Scf5990g01043	81,24 %
<i>ROP1</i> GTPase	AT3G51300	Oleur061Scf1177g07033	95,43 %
<i>AtGEF12</i>	AT1G79860	Oleur061Scf0910g15020	60,20 %
<i>RabA4d</i>	AT3G12160	Oleur061Scf0968g00007	82,14 %
<i>ACA9</i>	AT3G21180	Oleur061Scf4402g01016	69,39 %
<i>CNGC18</i>	AT5G14870	Oleur061Scf7420g01001	69,42 %

VANGUARD pectinesterase	AT3G62170	Oleur061Scf2813g02010	40,04 %
AtMGD2	AT5G20410	Oleur061Scf3585g01002	81,57 %
AtMGD3	AT2G11810	Oleur061Scf3585g01002	79,04 %

Ninguno de los genes imprescindibles parece cambiar significativamente su expresión durante el desarrollo del tubo polínico, aunque sí que es interesante destacar que para las lipasas obtenemos seis ortólogos en Picual de los que solo se expresan dos y uno de ellos (el gen Oleur061Scf2586g04033 es el mismo que para la esterasa).



**Figura 13.** Perfiles de expresión de (A) la callosa, la pectina, la lipasa, la hidrolasa, las peroxidasas, *BET11*, la esterasa, la poligalacturonasa y *BET12* (B) las cuatro pectin metil esterases, *NPG1*, *COG6*, *SEC1A*, *NPGR2* y *NPGR1*. (Obtenida con OliveAtlas).



**Figura 14.** Perfiles de expresión de (A) las CDPKs y (B) los genes críticos para el desarrollo del tubo polínico. (Obtenida con OliveAtlas).

## 4.5. Alérgenos

En otro de los objetivos de este trabajo queríamos saber cuáles eran los genes concreto de Picual que codifican los alérgenos tan conocidos en la especie, y ver si, además de estar presentes en el polen seco (que es de donde se caracterizaron), también están presentes cuando crece el tubo polínico.

### 4.5.1. Ortología

Para determinar los transcritos de Picual que codifican los alérgenos, hemos utilizado las secuencias codificantes disponibles en el NCBI (para todos los alérgenos menos para el *Ole e 4* y *Ole e 7* de los que solo disponemos de las secuencias de proteínas y *Ole e 6* ya que su secuencia de nucleótidos no nos devolvía similitud con ningún transcrito) y las hemos comparado con el transcriptoma de Picual como base de datos (*subject*) con BlastN (salvo para *Ole e 4*, *Ole e 6* y *Ole e 7* que hemos utilizado BlastP contra el proteoma de Picual). Los ortólogos resultantes se indican en la tabla 10. En el caso de *Ole e 5* y *Ole e 11*, tenemos varios transcritos que nos han devuelto un porcentaje de identidad muy elevado que nos impide determinar cuál es el correcto. Esto puede deberse a que haya varios parálogos en Picual, o bien a que el gen esté duplicado o sea una quimera dado el estado de borrador del genoma de esta variedad. También hemos colocado dos posibles ortólogos para *Ole e 1*, aunque probablemente el verdadero sea el que hemos denominado 1.1.

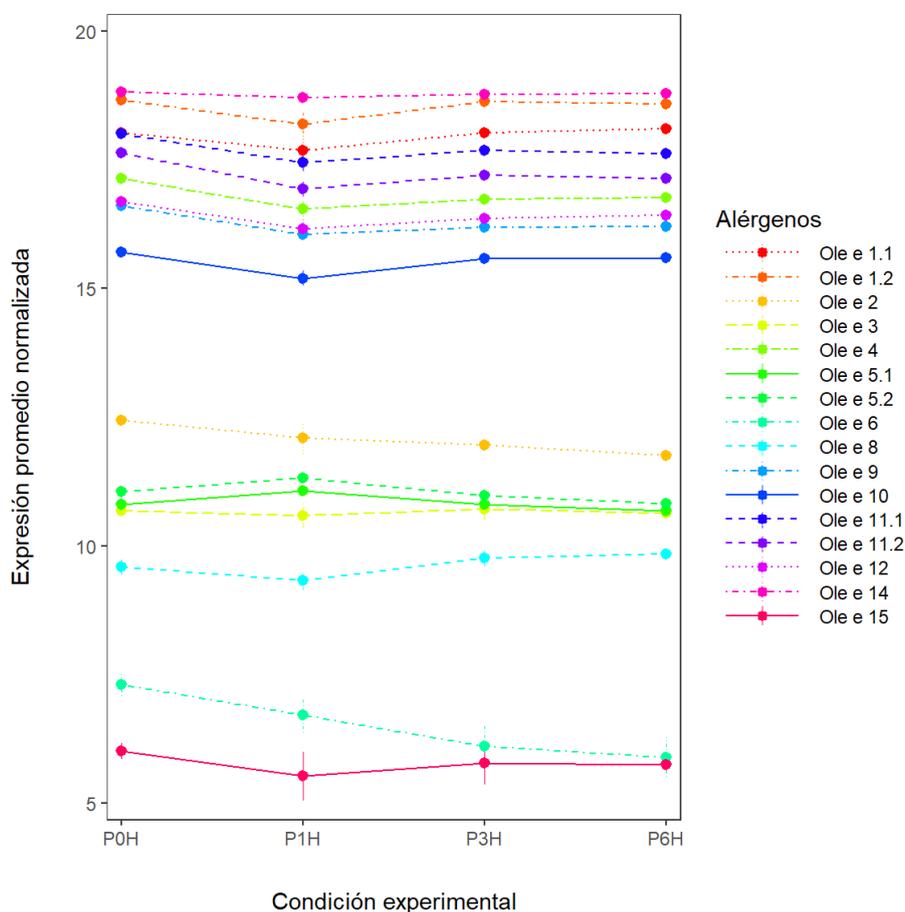
**Tabla 10.** Propuesta de ortología de los alérgenos del olivo en Picual.

Alérgeno	Transcrito de Picual	Ident. (%)	Alérgeno	Transcrito de Picual	Ident. (%)
<i>Ole e 1.1</i>	Oleur061Scf2365g01009.1	98 %	<i>Ole e 8</i>	Oleur061Scf1812g02029.1	100 %
<i>Ole e 1.2</i>	Oleur061Scf2365g01011.1	>96 %	<i>Ole e 9</i>	Oleur061Scf3514g03028.1	>98 %
<i>Ole e 2</i>	Oleur061Scf0405g02004.1	>97 %	<i>Ole e 10</i>	Oleur061Scf1517g08025.1	100 %
<i>Ole e 3</i>	Oleur061Scf0677g01016.1	97 %	<i>Ole e 11.1</i>	Oleur061Scf0658g00038.1	100 %
<i>Ole e 4*</i>	Oleur061Scf4268g00005.1	100 %	<i>Ole e 11.2</i>	Oleur061Scf1025g00024.1	99 %
<i>Ole e 5.1</i>	Oleur061Scf1858g00035.1	>99 %	<i>Ole e 12</i>	Oleur061Scf3497g00001.1	100 %
<i>Ole e 5.2</i>	Oleur061Scf7408g00008.1	>99 %	<i>Ole e 13</i>	Oleur061Scf2948g03005.1	99 %
<i>Ole e 6*</i>	Oleur061Scf5231g01021.1	64 %	<i>Ole e 14</i>	Oleur061Scf1515g00017.1	>98 %
<i>Ole e 7*</i>	Oleur061Scf3787g01003.1	90 %	<i>Ole e 15</i>	Oleur061Scf0157g06006.1	99 %

Todos los alérgenos, menos el *Ole e 7* y el *Ole e 13* (por ser de fruto) se expresan en nuestras muestras de polen. La inesperada ausencia de *Ole e 7* ya se debatió en la tesis de Rosario Carmona (Carmona 2020).

#### 4.5.2. Perfil de expresión

Para conocer el perfil y nivel de expresión de estos alérgenos, los representamos gráficamente en función de las muestras de polen (figura 15). El nivel de expresión de los alérgenos se mantiene prácticamente constante durante el desarrollo del tubo polínico. Además, los alérgenos con mayor expresión son el *Ole e 1* y el *Ole e 14*, mientras que *Ole e 15* es el que menos se expresa y que explica por qué ha sido el último en identificarse (San Segundo-Acosta et al. 2019).



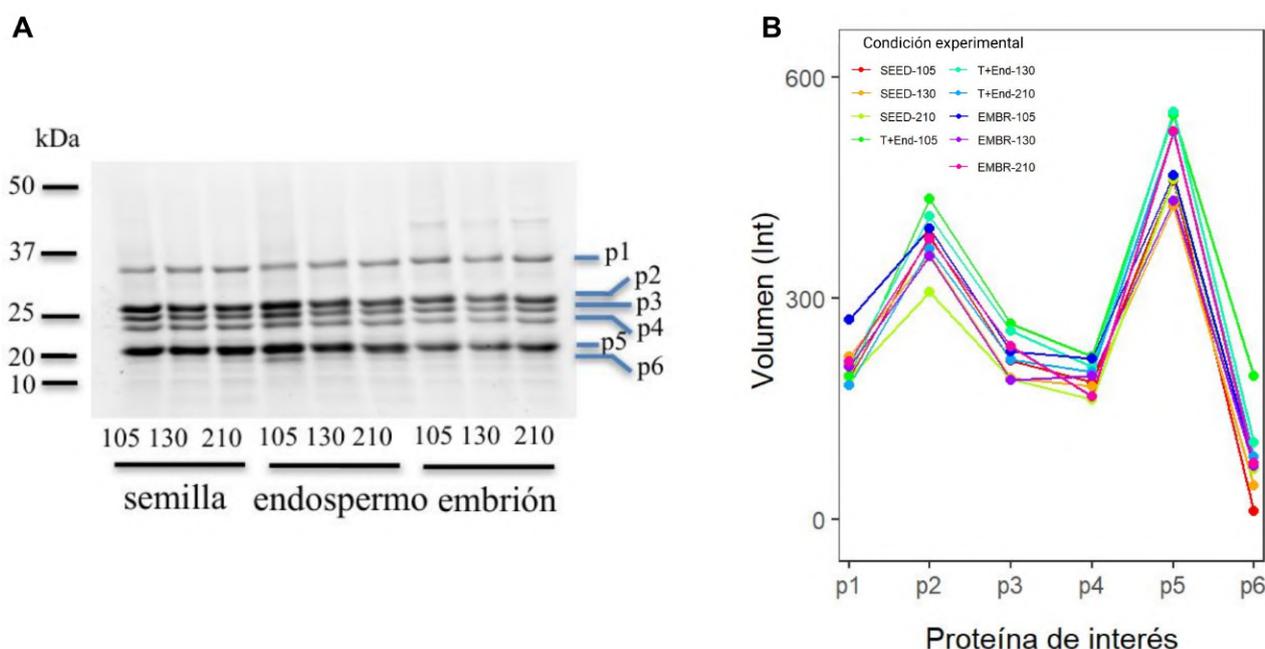
**Figura 15.** Perfil de expresión de los alérgenos propuestos.

## 4.6. Caracterización transcripcional de la semilla de olivo

### 4.6.1. Perfil proteico

Aunque el objetivo a largo plazo es el estudio del endospermo (junto con la testa) y el embrión por separado en distintos tiempos de maduración para conocer cómo se acumu-

lan los componentes nutracéuticos de interés, hemos empezado por conocer si ambos tejidos presentan perfiles proteicos fácilmente diferenciables en los genes de poliacrilamida teñidos con Coomassie. Para ello, hemos llevado a cabo un breve experimento de extracción de proteínas como se describe en los materiales y métodos. A continuación, se desnaturalizan y se separan en una electroforesis SDS-PAGE. Como resultado obtenemos que las proteínas mayoritarias son muy parecidas y no hay cambios significativos en las semillas de Picual (figura 16A), y tampoco en las de Arbequina (no se muestran por ser exactamente iguales a los de Picual). El gel de acrilamida teñido se escaneó con el Gel Doc™ para cuantificar y demostrar que los perfiles son prácticamente indistinguibles entre las condiciones (figura 16B). Como no se aprecia un cambio a simple vista en las proteínas minoritarias ni tan siquiera entre tejidos tan distintos como el endospermo y el embrión, las diferencias se conseguirán manifestar con más facilidad con un análisis transcripcional de las muestras.

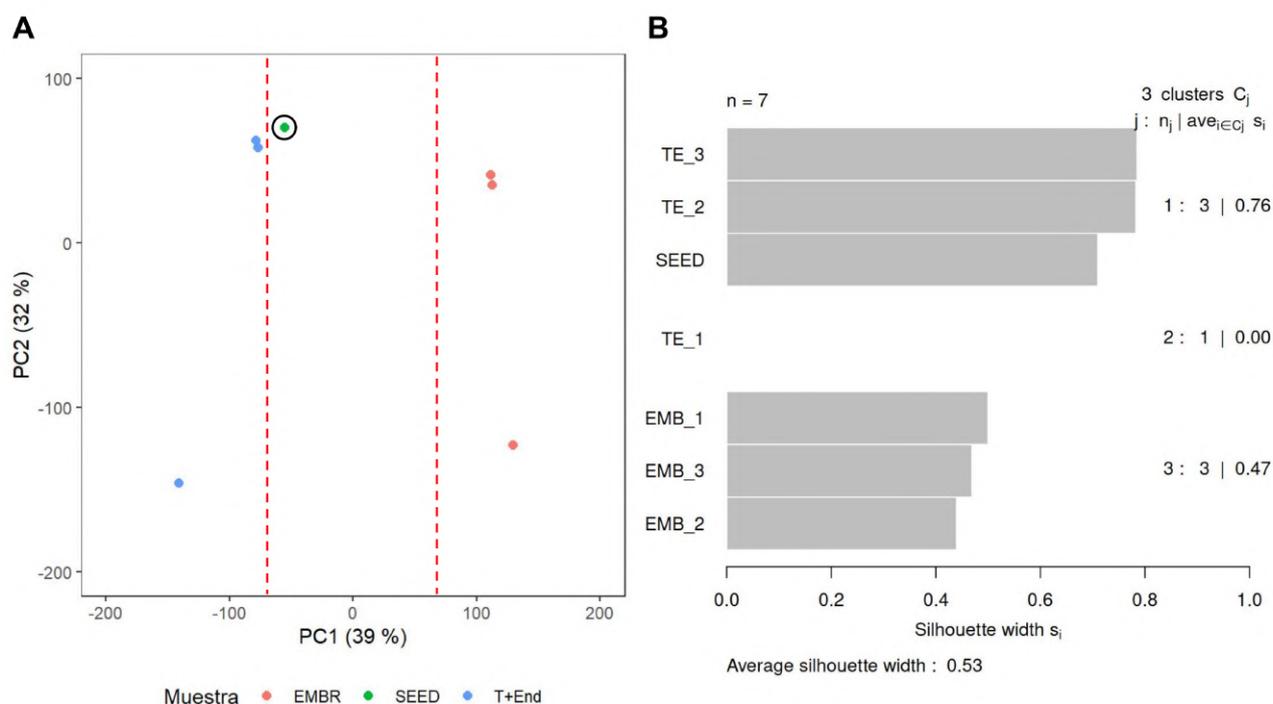


**Figura 16.** Análisis de las proteínas de semilla. (A) SDS PAGE del extracto de proteínas de Picual y (B) perfiles de cuantificación. SEED: semilla completa. T+End: testa y endospermo. EMBR: embrión. 105, 130 y 210 son los días después de la anthesis que caracterizan los tres estadios principales de maduración.

#### 4.6.2. Filtrado de los genes

Para este trabajo disponíamos solo de tres réplicas de embriones, otras tres de endospermo y testa y una muestra de semilla completa, todos en el momento inicial de la semilla (día 105), mapeadas sobre el transcriptoma de Picual. Al igual que con el polen, se fil-

traron los genes con expresión poco fiable antes de normalizarlos. Hemos utilizado también la herramienta web iDEP.95 y en el filtrado nos quedamos con los genes que tienen al menos 1 CPM en 3 librerías. En este caso, no se expresaban nunca 18.085 genes (22 %) y de los 44.096 (54 %) que se expresaban, solo 39.410 (48 %) superaron el filtrado. Lo primero que queda claro es que se están expresando más genes que en el polen.

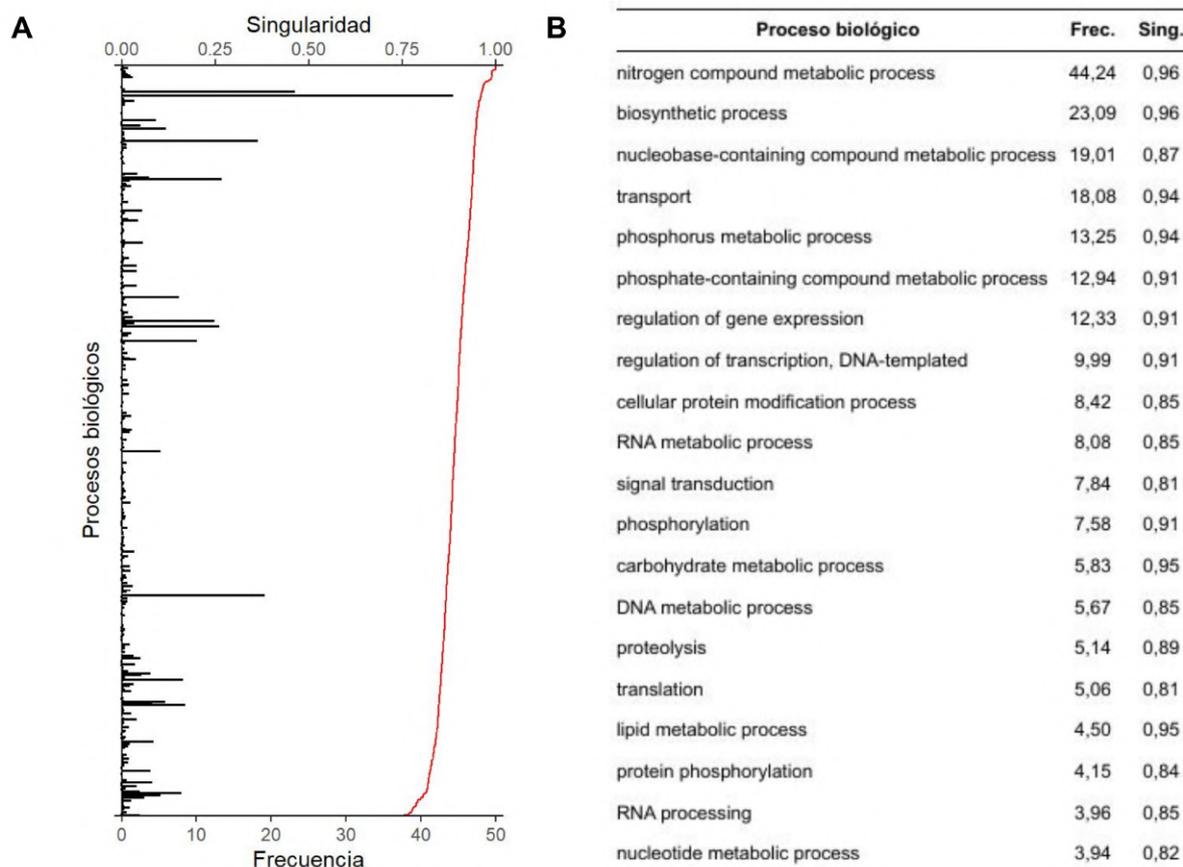


**Figura 17.** La congruencia de las réplicas de las muestras de los tejidos de la semilla se ilustra con un (A) PCA en el que se han usado las líneas rojas discontinuas para definir los agrupamientos, y se señala la muestra de semilla porque está más cerca de las T+End. (B) Gráfico de silueta de las muestras cuando se agrupan en tres *clusters*. Independientemente de cómo se agrupan, todas las muestras contribuyen positivamente al análisis.

#### 4.6.3. Congruencia de las réplicas

De nuevo queríamos ver que las muestras se agrupan por condición experimental. En el PCA de la semilla (figura 17A) podemos ver que las muestras de embrión (EMBR) y las de testa y endospermo (T+End) se separan bien en la primera dimensión, lo que nos indica que son diferentes entre sí. La única muestra de semilla completa está más cerca de las réplicas de T+End, pero es lo esperable porque el tejido que ocupa la mayor parte de la semilla es el endospermo. También llama la atención como, tanto para T+End como para EMBR, hay dos réplicas que están muy juntas y una tercera que aparece más separada. Al igual que en el polen, son datos que vienen de dos campañas de muestreo y secuenciación diferentes. Cuando hacemos un *clustering* jerárquico y lo valoramos con un

gráfico de silueta (figura 17B) con 3 *clusters*, uno por condición experimental, una de las réplicas de T+End no contribuye ni positiva ni negativamente. Además, las otras dos réplicas de T+End se agrupan bien con la de SEED y las tres de EMBR se agrupan entre ellas. Como ninguna muestra nos introduce ruido, trabajaremos con todas ellas. Cuando repetimos el *clustering* jerárquico sin la muestra de semilla e indicando que haga solo 2 *clusters*, entonces los grupos son los esperados (informe complementario 2). Todo ello indica que las muestras son aptas para el análisis.



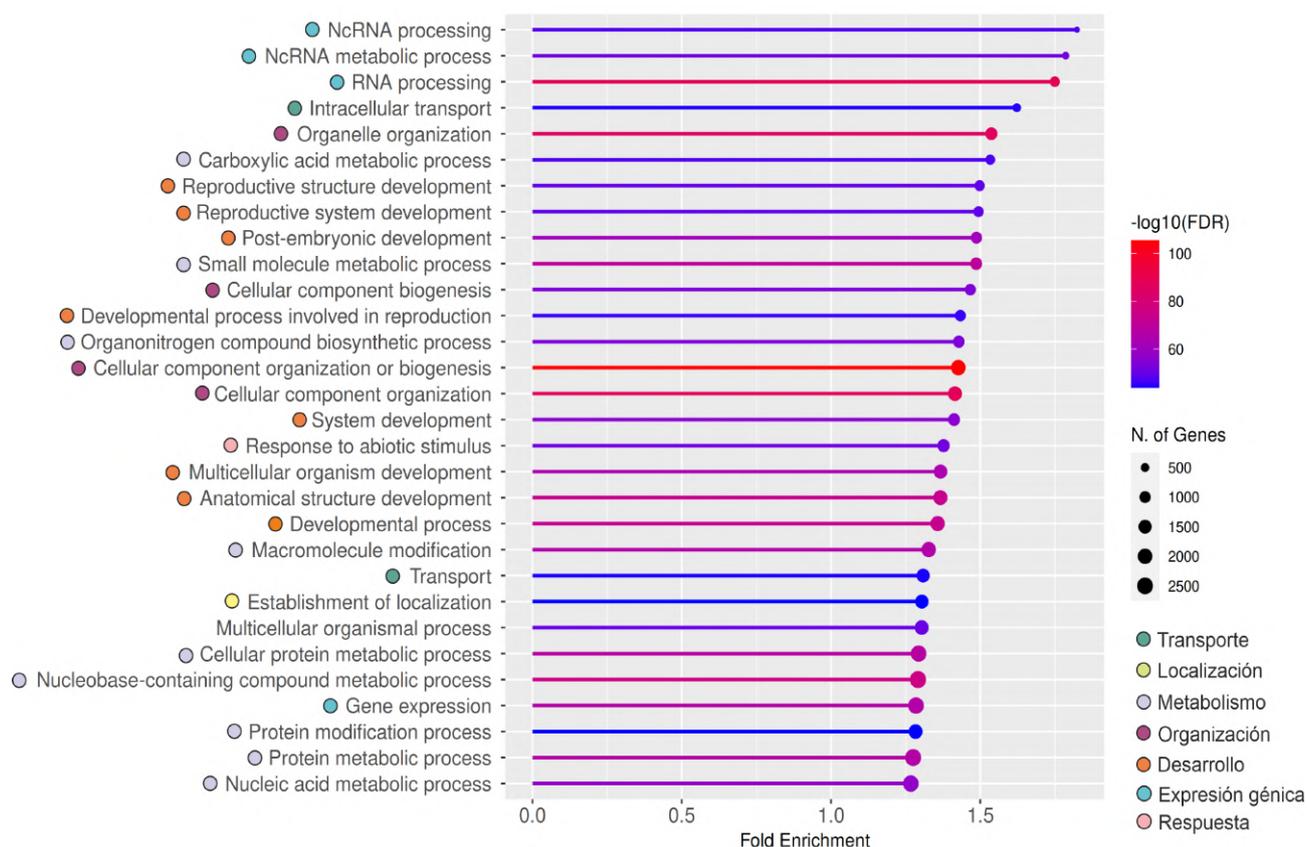
**Figura 18.** Procesos biológicos que se dan en los genes expresados en la semilla representados (A) en función de su frecuencia (barras) y su singularidad (línea roja), o como tabla (B) con los 20 más frecuentes.

#### 4.6.4. Funciones celulares de la semilla

Un análisis funcional con REVIGO de los genes que se expresan en la semilla (figura 18A) nos indica que la mayoría de los procesos biológicos tienen una frecuencia baja y esta tampoco va ligada a su singularidad. En la figura 18B tenemos una tabla con los 20 procesos más frecuentes. De nuevo, para este análisis hemos utilizado los ortólogos en *Arabidopsis* y sus términos GOs asociados, por lo que de 39.410 genes nos quedamos

con 33.445 términos GOs únicos.

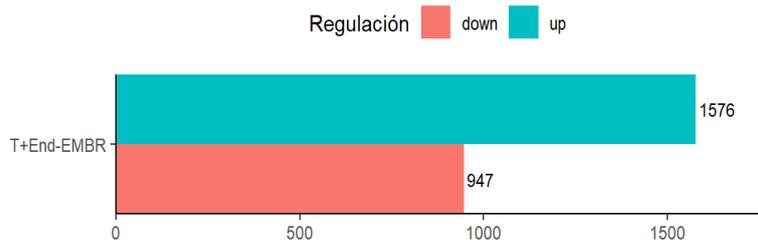
Ahora que ya sabemos los principales procesos biológicos que se dan en la semilla, es conveniente analizar los que se enriquecen (figura 19) que aparecen relacionados con el procesamiento ( $n = 343$ ,  $FE = 1,82$ ) y el metabolismo ( $n = 400$ ,  $FE = 1,79$ ) de ARN no codificante, en un claro ejemplo de la puesta en marcha de la maquinaria de regulación que necesita, sobre todo, el embrión. También se dan procesos de transporte y desarrollo.



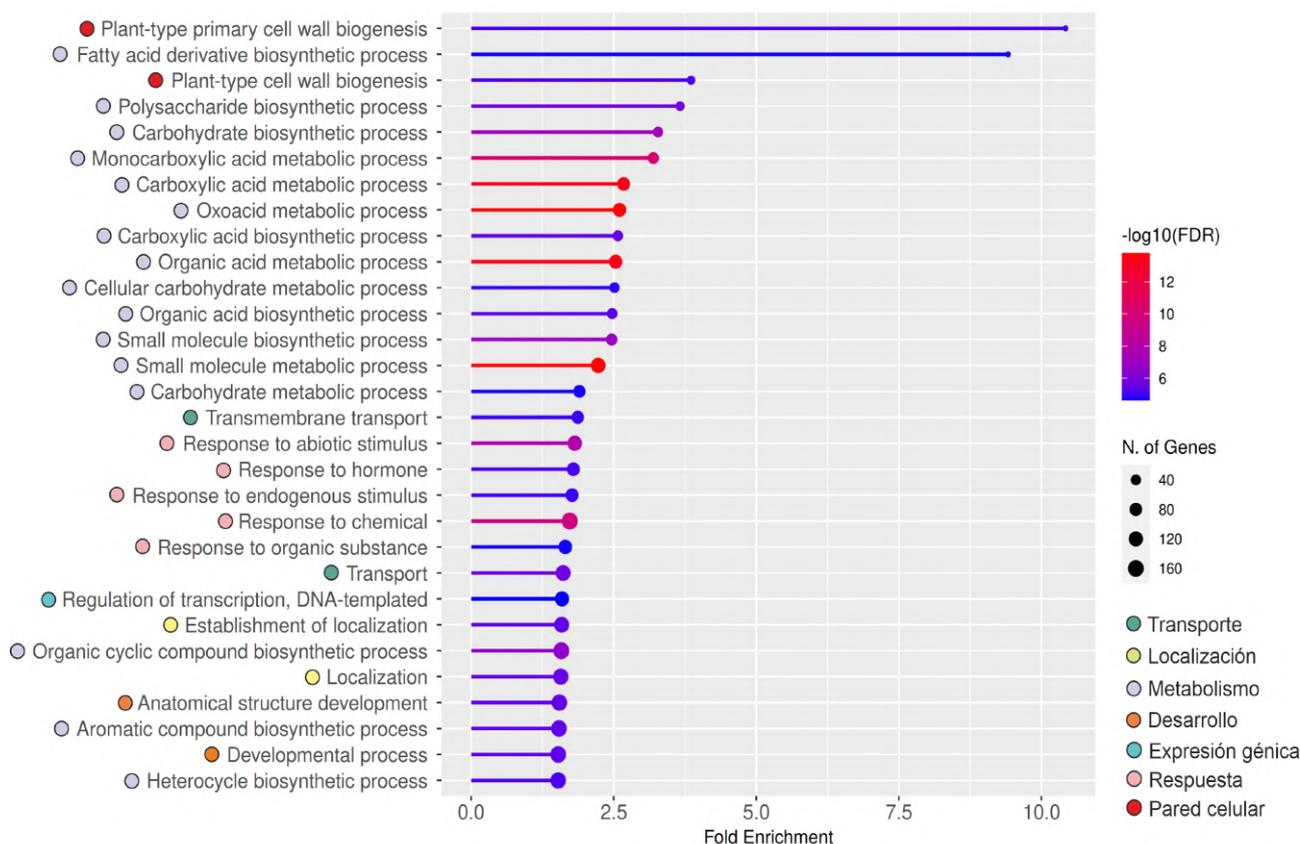
**Figura 19.** Enriquecimiento funcional de los genes que se expresan en la semilla. Los detalles son como en la figura 6. (Obtenida de ShinyGO 0.76)

#### 4.6.5. Expresión diferencial de los tejidos de la semilla

Analizamos los cambios de expresión diferencial que se producen en la semilla entre la testa+endospermo y el embrión (figura 20). Este análisis definió un número elevado de genes que se sobreexpresan en T+End (1.576) y un número más reducido de genes que lo hacen en el EMBR (947) en consonancia con la movilización de recursos del endospermo para alimentar al embrión durante el inicio de su desarrollo.



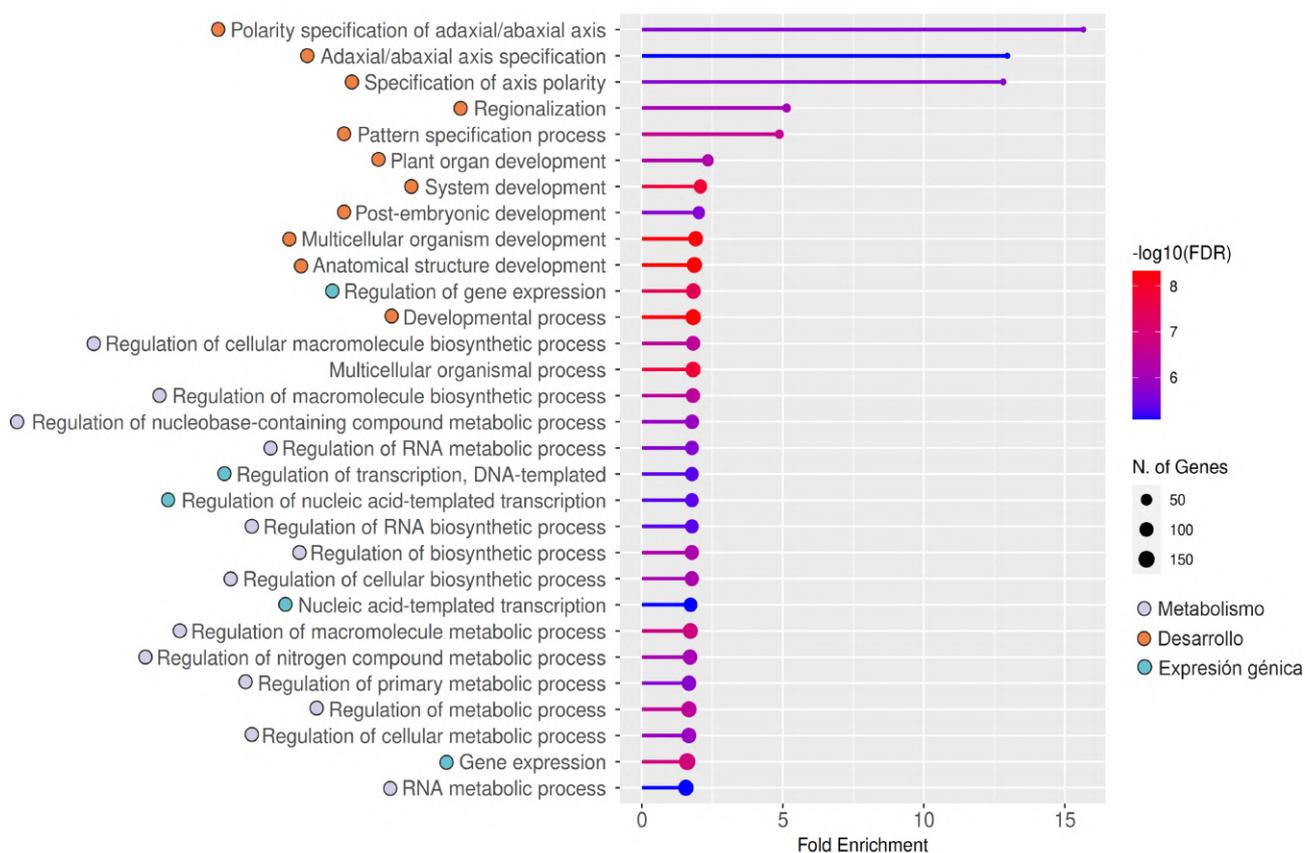
**Figura 20.** Número de DEGs para la comparación de T+End frente a EMBR.



**Figura 21.** Procesos biológicos que se enriquecen en T+End en los DEGs de la comparación T+End-EMBR. (Obtenida con ShinyGO 0.76).

Analizamos entonces por enriquecimiento funcional con ShinyGO v0.76 estos DEGs empleando el correspondiente ortólogo en Arabidopsis. Los procesos que se enriquecen en la T+End (figura 21) son los relacionados con la síntesis de glúcidos y grasas, el transporte transmembranario y procesos de respuesta, de nuevo en concordancia con la función conocida del endospermo. En cuanto a los procesos que se enriquecen en el EMBR (figura 22) son los que regulan la polaridad, la especialización de tejidos y órganos, la síntesis de macromoléculas y biosíntesis lo que es inequívoca señal de un organismo en de-

sarrollo.



**Figura 22.** Procesos biológicos que se enriquecen en EMBR en los DEGs de la comparación T+End-EMBR. (Obtenida con ShinyGO 0.76).

### Coexpresión de genes de los tejidos de la semilla

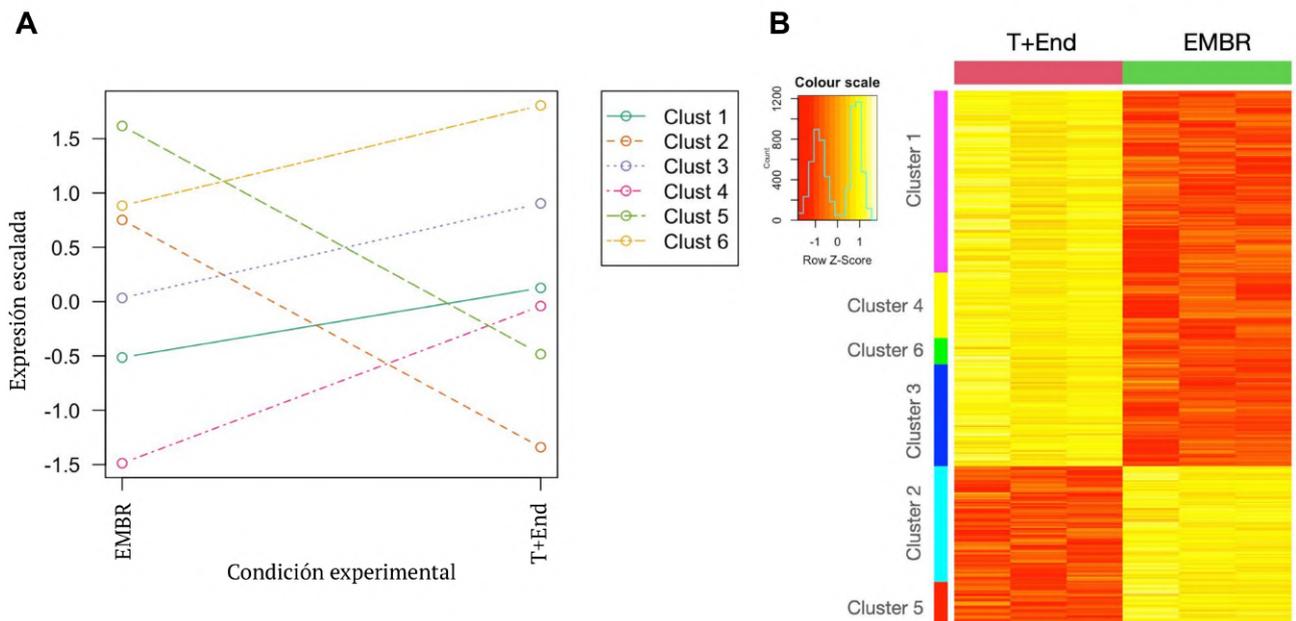
El análisis de *clustering* jerárquico con "poda" dinámica del dendograma nos devolvió seis *clusters* (figura 23A) con los números de genes recogidos en la tabla 11.

**Tabla 11.** *Clusters* del *clustering* jerárquico y número de genes que tiene cada uno.

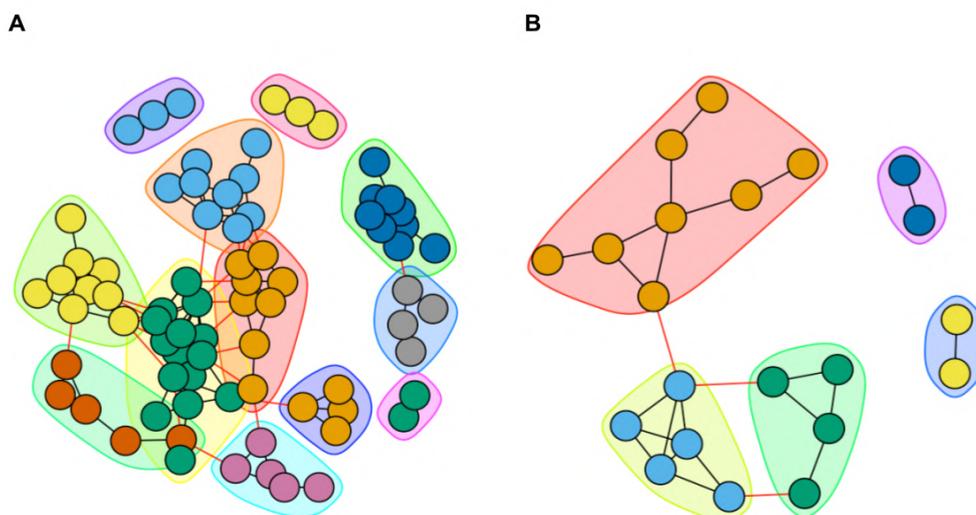
<b>Cluster</b>	1	2	3	4	5	6
<b>Número de genes</b>	368	234	205	132	87	54

Se dividen en dos perfiles de expresión principales: los que tienen una expresión más baja en EMBR y aumenta en T+End (1, 3, 4 y 6) y los que tienen una expresión elevada en EMBR y cae en T+End (2 y 5). El informe completo sobre este análisis se encuentra en la informe complementario 3. Puesto que las pendiente son similares, la única diferencia real entre los grupos parece estar en su nivel de expresión. El verdadero parecido entre

los distintos *clusters* se aprecia muy bien en la matriz cromática (*heatmap*) de la figura 23B.



**Figura 23.** Perfil de expresión de los *clusters* de las muestras de semilla (A) obtenidos por *clustering* jerárquico o como una (B) matriz cromática (*heatmap*).



**Figura 24.** *Clusters* número 5 (A) y número 6 (B) del *clustering* jerárquico, los únicos que presentan varias comunidades numerosas y *hubs*.

Hemos determinado las comunidades de co-expresión de cada uno de los seis *clusters*. En la mayoría, las comunidades están formadas solo por 2 a 5 genes, pero en los *clusters* 5 y 6 las redes contienen algunos *hub* significativos (nodos verdes en el *cluster* 5 y nodos

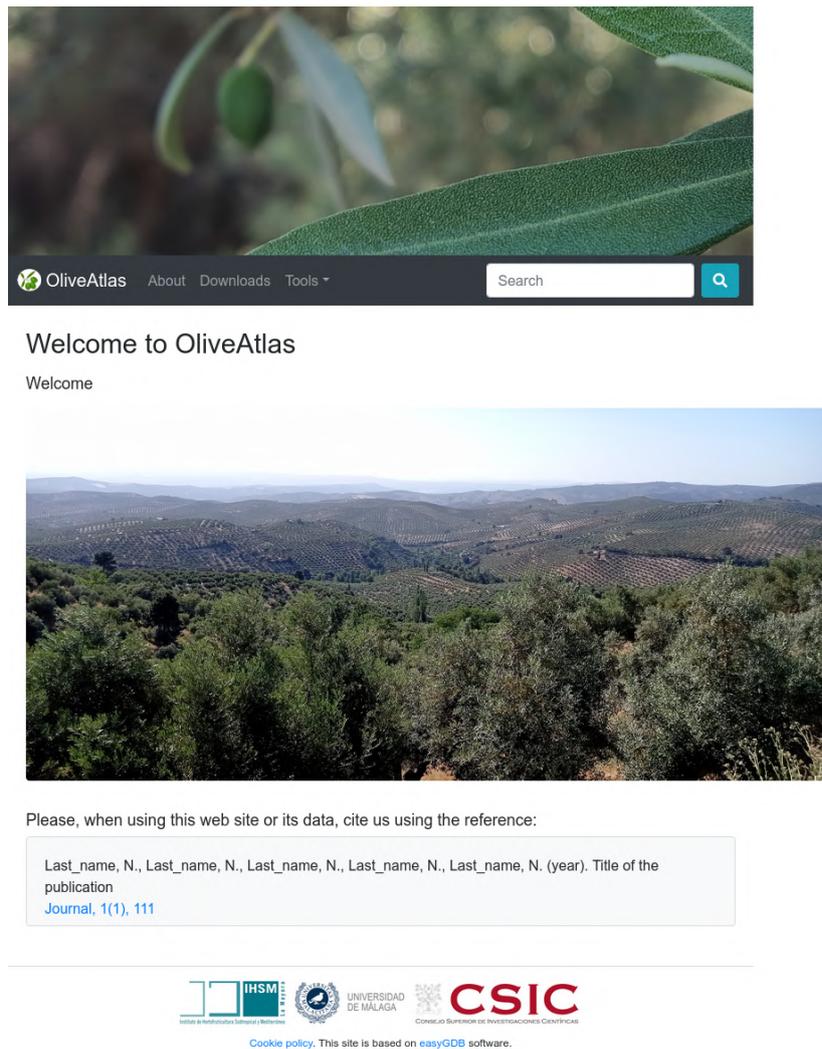
azul claro en el *cluster* 6). En el *cluster* 5 (figura 24A) se dan procesos de regulación y biosíntesis (figura S7), mientras que en el *cluster* 6 (figura 24A) tenemos procesos de metabolismo de la pared celular y metabolismo secundario.

#### 4.7. El atlas de expresión del olivo

Construir un atlas de expresión del olivo (figura 25) nos va a facilitar obtener información de los patrones de expresión de los genes. Cuando se ponga en funcionamiento estarán disponibles todos los datos de Picual analizados en este TFM (muestras de polen y semilla) y datos replicados encontrados en bases de datos públicas que se están localizando en nuestro laboratorio (muestras de fruto, flor, hoja, meristemos, raíz, tallo y polen, muestras de estreses abióticos como frío y heridas en la raíz, y muestras a estreses bióticos como infección por *Verticillium dahliae*). En el futuro contemplamos extender este atlas a todas las variedades para las que haya un genoma fiable.

El atlas de expresión incluye múltiples métodos de visualización para explorar y comparar los datos de expresión en una gran variedad de tejidos y condiciones experimentales. Su herramienta principal es la expresión de genes, que nos permite hacer una búsqueda en la base de datos de una lista de genes de interés y representar sus perfiles de expresión, así como una matriz cromática con los niveles de expresión medios en cada réplica, una gráfica los valores de expresión de cada réplica para cada gen de forma individual y una tabla resumen (figura 26). Estará conectado a OliveTreeDB (<https://genomaolivar.dipujaen.es>) para el uso de herramientas como búsqueda de genes por su ID o anotación, BLAST, JBrowse, extracción de secuencias y extracción de anotaciones. En el futuro se plantea añadir una herramienta de búsqueda de ortólogos entre algunas variedades de olivo.

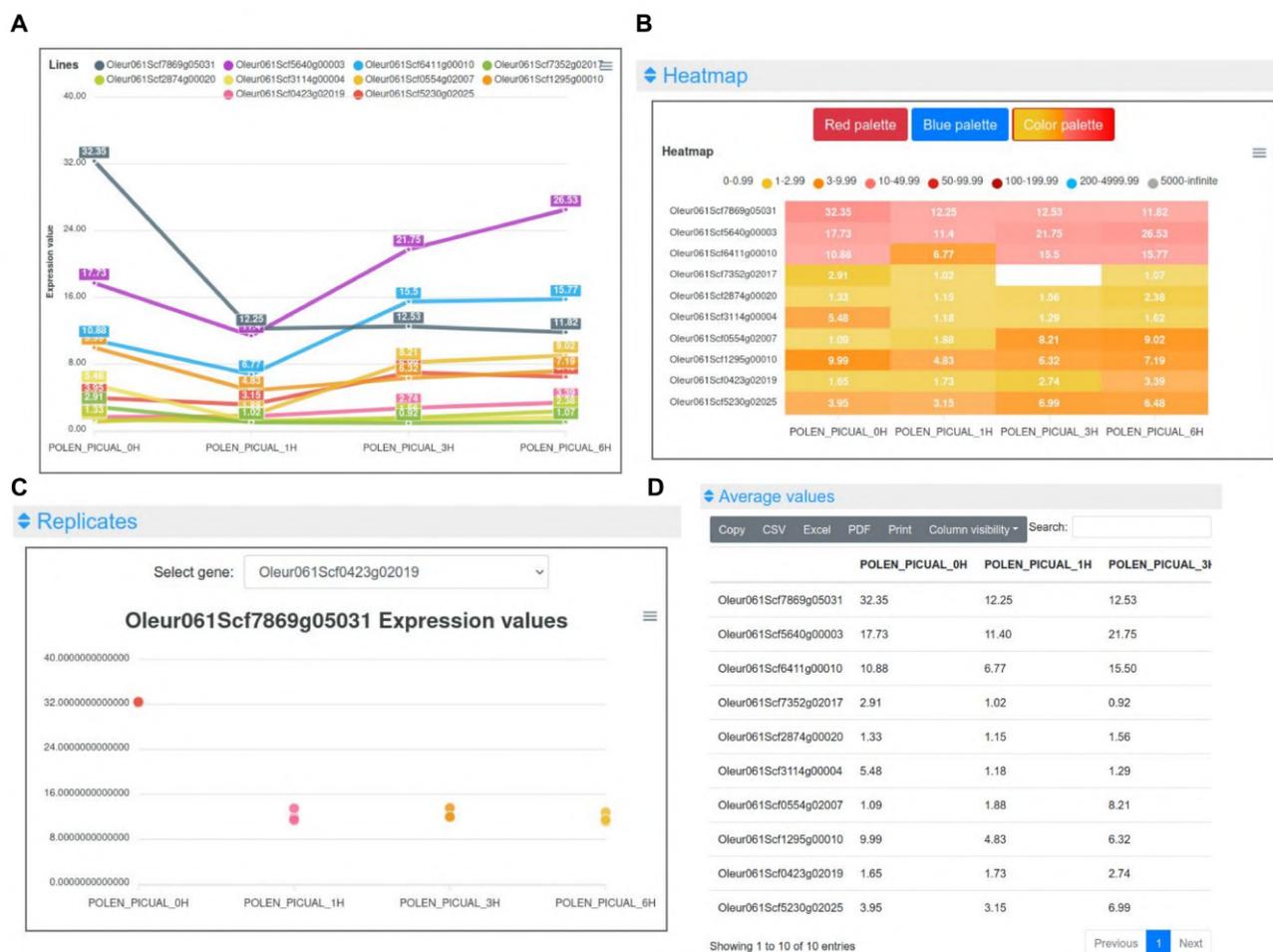
Actualmente es accesible de forma local, pero en los próximos meses se podrá consultar en un servidor web (<https://oliveatlas.uma.es/>).



**Figura 25.** Página de bienvenida a OliveAtlas.

#### 4.7.1. Construcción del atlas

Para construir el atlas, hemos instalado los recursos de EasyGDB (Fernandez-Pozo and Bombarely 2022) en un contenedor Docker, hemos colocado los datos de expresión en el directorio `scr/expression_data` y ya lo tenemos listo para funcionar localmente a través de la dirección `localhost:8000/easy_gdb/`.

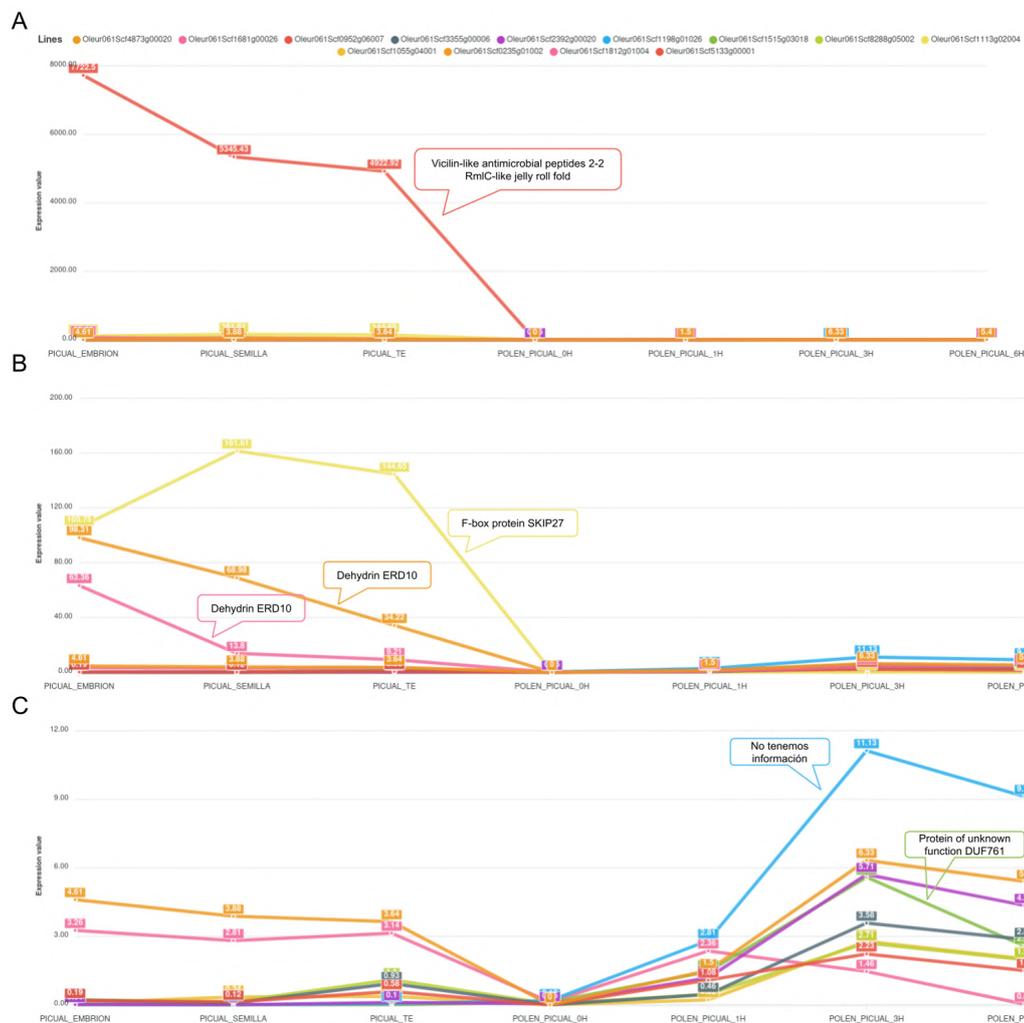


**Figura 26.** Herramientas disponibles en OliveAtlas. (A) Gráfica de los perfiles de expresión. (B) Matriz cromática con los niveles de expresión medios de cada réplica. (C) Gráfica de los valores de expresión cada réplica para cada gen. (D) Tabla resumen.

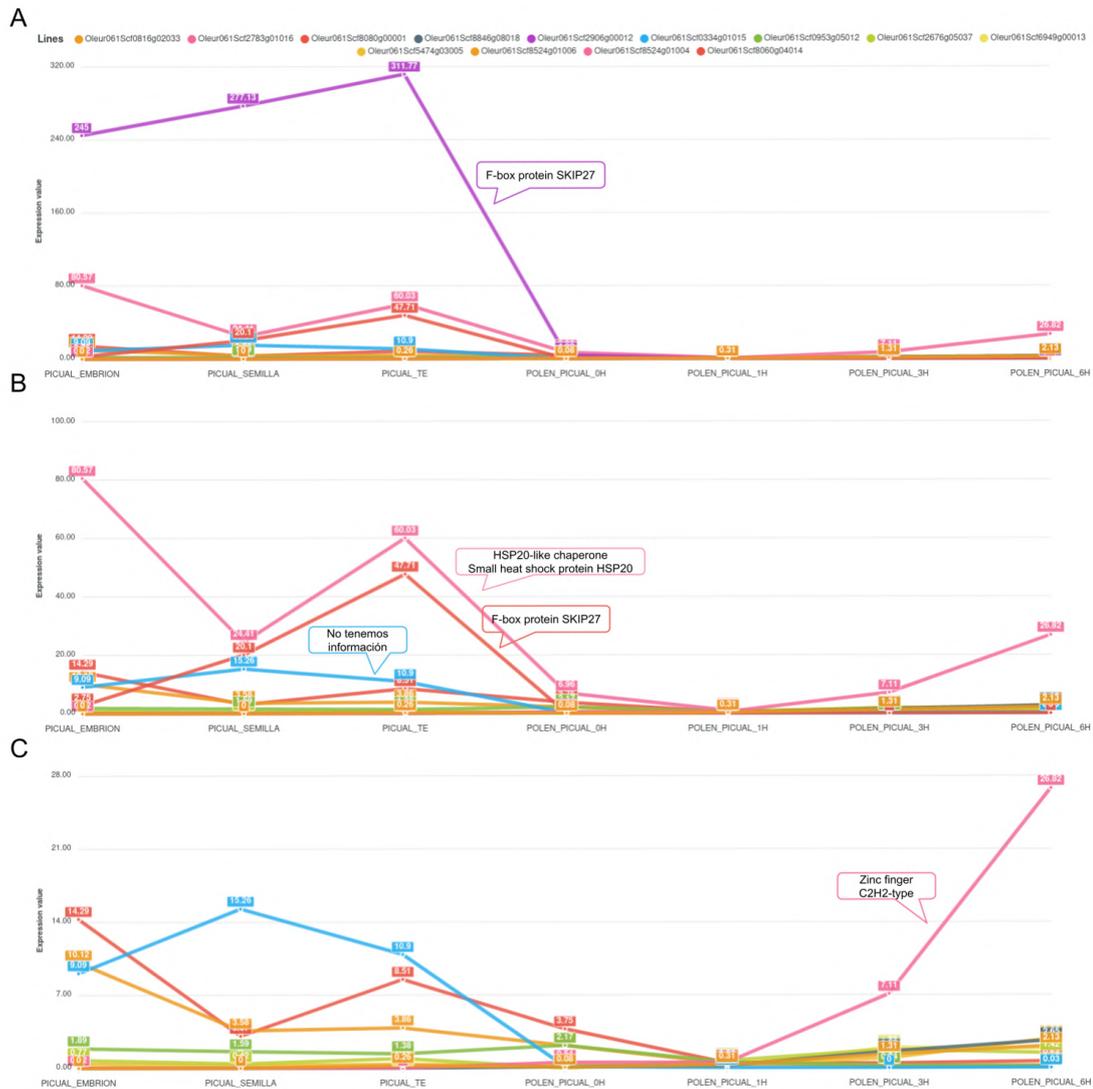
#### 4.7.2. Genes que más varían durante el desarrollo del tubo polínico

Para ilustrar la comodidad que es tener este atlas, analizamos de los perfiles de expresión de los 55 genes que más cambian en el desarrollo del tubo polínico cuando se calcula el coeficiente de variación del logaritmo de la expresión de cada gen a lo largo de todas las muestras. Los 55 son los que tienen un CV > 15 %, independientemente de si son DEG o no. Si nos centramos en los 15 primeros (por limitación de la interfaz), vemos que el gen *Oleur061Scf5133g00001*, pese a ser uno de los genes más variables del polen (CV ≈ 24 %), es una proteína similar a las vicilinas (típicas de semilla) que se expresa considerablemente más en la semilla (figura 27A). Su elevada expresión en la semilla nos impide ver el comportamiento de los otros genes. Si lo quitamos con un simple clic, observamos que tenemos un problema similar con los genes *Oleur061Scf1113g02004*

(CV ≈ 19 %), Oleur061Scf0235g01002 (CV ≈ 22 %) y Oleur061Scf1681g00026 (CV ≈ 21 %), que son una proteína F-box SKIP27 y dos dehidrinas ERD10, respectivamente (figura 27B), también típicas de semilla, con un nivel de expresión elevadísimo en esos tejidos. Al ocultar sus perfiles de expresión con sendos clic, nos aparecen dos genes con comportamientos interesantes: el gen Oleur061Scf1198g01026 (CV ≈ 22 %) y el gen Oleur061Scf1515g03018 (CV ≈ 20 %). Ninguno de ellos se expresa en la semilla, mientras que en el polen aumenta mucho su expresión entre la primera y la tercera hora de desarrollo y luego disminuye a las 6 h (figura 27C). Lamentablemente, ninguno de estos dos genes está anotado y no sabemos qué función pueden desempeñar. Por tanto, acabamos de descubrir dos genes muy expresados durante el desarrollo del tubo polínico para los que no tenemos ninguna información y que merecen un futuro estudio más profundo.



**Figura 27.** Perfil de expresión en las muestras de semilla y de polen de los 15 primeros genes más variables en el desarrollo del tubo polínico. (Modificada del Atlas del olivo, en desarrollo).



**Figura 28.** Perfil de expresión en las muestras de semilla y de polen de los 15 segundos genes más variables en el desarrollo del tubo polínico. (Modificada del Atlas del olivo, en desarrollo).

Si analizamos los siguientes 15 genes, Oleur061Scf2906g00012 (CV ≈ 17 %) es una proteína F-box SKIP27 que se expresa mucho más en la semilla que en el polen (figura 28A), como la mayoría de las F-box. Si ocultamos su expresión, tenemos los genes Oleur061Scf8524g01004 (CV ≈ 18 %), Oleur061Scf8080g00001 (CV ≈ 17 %) y Oleur061Scf0334g01015 (CV ≈ 18 %), que son una chaperona similar a la HSP20, otra proteína F-box SKIP27 y un gen cuya función desconocemos, tienen un comportamiento similar en la semilla (figura 28B) y, lo más interesante, la chaperona aumenta claramente su expresión a las 6 h de desarrollo del tubo polínico. Si ocultamos estos tres genes, podemos apreciar al gen Oleur061Scf2783g01016 (CV ≈ 18 %) que tiene un comportamiento muy interesante y es que su expresión se dispara casi cincuenta veces desde tiempo

0 h a 6 h (figura 28C). Este gen en un dedo de zinc tipo  $C_2H_2$  que podría ser responsable de la regulación génica de las últimas etapas de este proceso.

Llama la atención que encontremos tantos genes tan poco estudiados entre los que más varían durante el desarrollo del tubo polínico. De hecho, de los 55 genes hay 13 que no conocemos la función que desempeñan. Para intentar anotarlos, hemos hecho un BlastP de su secuencia de proteínas contra la base de datos de proteínas vegetales no redundantes. De estos transcritos, Oleur061Scf2371g06028.1 se describe como proteína de almacenamiento de semillas similar a la vicilina (p. ident. del 40 %), como proteína tipo Nipped-B (p. ident. del 57 %) y como proteína similar a la Histona Lisina N-Metiltransferasa, H3 lisina-4 específica (p. ident. del 45 %) en *Actinidia chinensis* var. *chinensis*. El transcrito Oleur061Scf2388g00003.1 aparece como factor  $\rho$  de intercambio de nucleótidos de guanina en *Perilla frutescens* y en *Salix suchowensis* (p. ident. del 44 %). Oleur061Scf6949g00013.1 está anotado como proteína similar a la chaperonina en *Actinidia chinensis* var. *chinensis* (60 % de indentidad); proteína de unión de la membrana plasmática en *Parasponia andersonii* (59 % de identidad), *Perilla frutescens* (sube al 65 %) y *Trema orientale* (59 %); y proteína transmembranaria en *Prunus mume* (57 % de identidad) y *Senna tora* (60 % de indentidad). Por último, Oleur061Scf8846g08018.1 se describe como proteína similar a la oxidorreductina 1 de retículo endoplasmático (60 % de identidad) y a una proteína transmembranaria (56 % de identidad) en *Pyrus ussuriensis* x *Pyrus communis* (60 % de identidad). El transcrito Oleur061Scf0076g02032.1 se describe como proteína HTH Myb en *Salix suchowensis* (34 % de identidad), que lo coloca como un inequívoco factor de transcripción.

## 5. Discusión

### **Las diferencias en el mapeo podrían deberse a la procedencia de las variedades**

Las diferencias observadas en el mapeo cuando se usa Picual o acebuche como referencia (tabla 5) y los resultados del porcentaje de identidad de los ortólogos proteicos de estas dos variedades (figura 3) nos indican que Picual y acebuche podrían parecerse menos de lo que cabría esperar. Sin embargo, en (Besnard et al. 2001) se compararon los perfiles RAPDs (amplificación aleatoria de ADN polimórfico o *Random Amplification of Polymorphic DNA*) de 121 cultivares de olivo con los de 20 poblaciones naturales de acebuche de las zonas oriental y occidental de la Cuenca del Mediterráneo y determinaron que los cultivares de Israel, Turquía, Siria, Grecia y Sicilia estaban cerca del grupo oriental de acebuches, mientras que los clones de Italia continental, Francia continental, Córcega, España y el Magreb estaban más cerca del grupo occidental. Dado que el acebuche secuenciado era un individuo localizado en la ciudad de Orhangazi de la provincia de Bursa (Turquía) (Unver et al. 2017), mientras que las muestras que hemos estudiado son de individuos de la variedad Picual cultivados en la Estación Experimental del Zaidín (Granada, España), la distinta procedencia geográfica podría explicar muy bien el bajísimo porcentaje de mapeo (~6 %) de las lecturas de Picual sobre la referencia de acebuche. En cambio, el genoma secuenciado de Farga se hizo a partir de muestras tomadas de un individuo de la región del Maestrazgo (Teruel, España) de unos mil años de antigüedad (Cruz et al. 2016). La variedad Arbequina procede de Palestina, pero se introdujo en España en la comarca de Arbeca en Lérida en el siglo XVII (<https://centrodeolivaryaceite.com/arbequina-una-variedad-en-expansion/>), aunque se secuenció un individuo de Mianning, en la provincia de Sichuan (China) (Rao et al. 2021) donde fue introducido en 1964 desde Albania (Wang et al. 2019), que pertenece a la Cuenca del Mediterráneo occidental. Dada la plasticidad fenotípica (y probablemente del genoma) del olivo (Mousavi et al. 2019; Badad et al. 2021), parece que Farga, Picual y Arbequina podrían considerarse occidentales, y diferentes al acebuche turco, y que esto justificaría la diferencia tan grande que se observa en los mapeos, ya que ambos proceden de dos grupos de olivos silvestres diferentes. En cambio, es más difícil explicar que dado que Arbequina es la variedad que llegó más recientemente a la península de las analizadas, se parezca tanto a Picual (figura 3), por lo que es posible que las diferencias de secuencia en los genomas se deban a otros factores que no hemos contemplado.

## **El número de genes expresados se incrementa a lo largo del desarrollo del tubo polínico**

El genoma de olivo cuenta con 81.484 transcritos tentativos derivados de otros tantos modelos de gen predichos en el genoma (Unver et al. 2017). Como de los 19.828 que se expresan a 0 h se pasa a 21.024 a 6 h y además hay 25.373 transcritos que nunca se expresan, planteamos la posibilidad de que dicho genoma está sobreestimado y será necesario muestrear otros tejidos y etapas de desarrollo del olivo para confirmarlo, algo que ya se está haciendo en nuestro laboratorio. De hecho, en (Carmona et al. 2015) ya se planteaba que el transcriptoma reproductivo representativo de olivo debía rondar como mucho los 28.972 transcritos tentativos, un valor muy cercano al número de genes expresados en el polen que hemos detectado (22.937).

En (Wang et al. 2008) se describe un estudio similar al realizado en este TFM: describieron los cambios en el transcriptoma de los granos de polen maduros desecados a los granos de polen hidratados y luego a los tubos polínicos de *Arabidopsis*. Encontraron que el número de genes expresados aumentó significativamente desde el polen maduro desecado hasta el polen hidratado y, de nuevo, hasta los tubos polínicos en crecimiento. En nuestro caso, hay un aumento de 716 genes entre el polen maduro desecado (muestras P0H) y el polen hidratado (P1H), y un aumento de 480 genes a lo largo del desarrollo del tubo polínico hasta las 6 h, unas cifras muy cercanas a los 692 genes y 255 que se describen para *Arabidopsis* en (Wang et al. 2008), respectivamente.

### **Un posible gen clave del desarrollo del tubo polínico**

En el análisis de expresión diferencial hallamos un DEG que resulta de especial interés porque su expresión va en aumento significativo a medida que se desarrolla el tubo polínico (figura 8B). La anotación de gen indica que tiene un dedo de zinc de tipo RING (*Really Interesting New Genes*) presente principalmente en las ubiquitina ligasas de tipo E3, que desempeñan funciones muy importantes en las plantas relacionadas con el crecimiento, desarrollo, resistencia al estrés (sequía, salinidad, temperatura, oxidación y metales dañinos) y transducción de señales (Han et al. 2022). Su función más reconocida es la de modificar y degradar las proteínas oportunas. Este resultado nos incita a pensar que la degradación de proteínas tiene una función relevante en el desarrollo del tubo polínico. Pero es que, además, se ha visto que estas proteínas se pueden acumular en el núcleo (Sun et al. 2019) y funcionar como factor de transcripción, y así regular la respuesta al estrés, ácido abscísico, etileno y otras (Han et al. 2022). Recientemente se ha descubierto que otra proteína RING regula el estrés hídrico gracias a su actividad en el ciclo de ubiquitinación

(Chen et al. 2022). ¿Podría también regular la expresión de algunos genes que intervienen en el desarrollo de este órgano? Esto podría ser indicio de que estamos ante una nueva proteína pluriempleada (*moonlight protein*) (Singh and Bhalla 2020), aunque habría que comprobar en el olivo en qué localización intracelular aparece.

### **El perfil funcional del tubo polínico de olivo es similar al de otras plantas**

Se sabe que el polen, antes de germinar, tiene una expresión baja en genes relacionados con las vías energéticas y la traducción, pero cuenta con una gran reserva de ribosomas y ARNt necesario para la traducción durante el crecimiento del tubo polínico (da Costa-Nunes and Grossniklaus 2004). Esto explicaría por qué vemos un enriquecimiento de estos dos procesos a 1 h de la germinación (figura S2) y por qué estos dos procesos aparecen reprimidos a 0 h. De hecho, en (Wang et al. 2008) se describe que la expresión de los genes relacionados con la transcripción, la transducción de señales y el transporte celular se modificaron significativamente durante la germinación del polen y el crecimiento del tubo, que son procesos que en nuestro análisis funcional aparecen entre los 20 más frecuentes (figura 5B). Con todo esto, parece que el polen maduro está controlando que esa germinación no se produzca antes de entrar en contacto con el estigma, ya que se enriquecen procesos relacionados con la regulación de la expresión génica (figura S1). A 1 h también se dan procesos relacionados con el ácido giberélico (figura S3), crucial en la germinación del polen y el desarrollo del tubo polínico en arroz (Chhun et al. 2007) y en *Arabidopsis* (Singh et al. 2002). A las 3 h se enriquecen procesos relacionados con la mitosis y la división celular, probablemente porque en ese momento se esté produciendo la mitosis polínica II (McCormick 2004). Más concretamente, se ha visto que la expresión génica en el gametofito masculino está sesgada hacia la señalización y el transporte de vesículas (Pina et al. 2005). En cuanto a la señalización, se enriquece la ruta de señalización mediada por el ácido giberélico a 1 h (figura S3), pero las CDPKs, que están involucradas en el crecimiento del tubo polínico (Yang et al. 2021) tienen un perfil de expresión constante durante todo el desarrollo del tubo polínico (figura 14A). El transporte de vesículas se enriquece en el polen maduro (figura S1) y en el polen tras 6 h de germinación (figura S6), por lo que comprobamos que en el olivo las vesículas también son relevantes justo antes de la germinación y en el último estadio del desarrollo del tubo.

En nuestras muestras de polen maduro se expresan muchos genes de respuesta a estrés abiótico y biótico para proteger el polen y la mayoría se inhiben en la primera hora de germinación (figura 10). Esto tiene sentido porque el proceso de fecundación puede ser interrumpido por cambios ambientales y el polen es muy sensible a estreses como el frío

o una humedad baja (Luria et al. 2019), lo que también podría explicar que en el *cluster 3* (figura 11) haya una comunidad que interviene en procesos del crecimiento del tubo polínico que se co-expresa de forma estrecha con una subcomunidad relacionada con la respuesta a estreses.

### **El polen contiene proteínas que no se transcriben en él y eso dificulta el estudio transcripcional de la autoincompatibilidad**

El estudio de los genes de la figura 12 no nos permite confirmar ni desmentir que en el olivo se dé la DSI, como se describe en (Saumitou-Laprade et al. 2017b) ni tampoco podemos descartar la SSI, ya que todos los genes, salvo *SCR*, se expresen en el estigma. Solo podemos confirmar que no es en el polen donde se expresa el transcrito de *SCR*, al igual que sugirió (Carmona 2020) para el alérgeno *Ole e 7*. Tanto con *SCR* como con *Ole e 7*, la proteína se detecta en el polen pero no así su transcrito.

También vemos que otros de los genes del proceso descritos en el estigma que también se expresan en el polen (*ARC1* y *THL-1*) y otro más en la semilla (*MLPK*), por lo que claramente intervienen en otras señalizaciones intracelulares fuera del estigma o, en su defecto, los ortólogos determinados son muy lejanos y nos llevan a un parálogo incorrecto en el olivo. Seguiremos buscando genes que se sepa que intervienen en la DSI por localización en el genoma de los microsatélites asociados a los locus de DSI descritos en (Besnard et al. 2020) y (Mariotti et al. 2021), y mejorando la ortología y anotación de los genes de Picual.

### **Los genes esenciales para el desarrollo del tubo polínico y los alérgenos se mantienen constantes**

Cuando estudiamos el perfil de expresión de los alérgenos (figura 15) y de los genes importantes para el desarrollo del tubo polínico (figuras 13 y 14) apreciamos que dicho perfil es bastante plano entre 0 h y 6 h. Esto indica que los genes esenciales ya están activados en el polen en reposo y su expresión continua siendo necesaria durante el desarrollo del tubo polínico, aunque no se observan cambios significativos de niveles de expresión.

Con respecto a los alérgenos, hemos determinado con bastante seguridad los ortólogos de Picual (tabla 10) para los 15 alérgenos de olivo que hay descritos hasta la fecha (tabla 3) con una identidad elevada en todos los casos (salvo en *Ole e 6*). Todos, menos el *Ole e 7* y el *Ole e 13* se expresan en nuestras muestras de polen de una forma constante durante el desarrollo del tubo polínico (figura 15). La ausencia de *Ole e 13* tiene sentido porque es un alérgeno presente en el fruto, pero la de *Ole e 7* es un misterio que ya

apuntaba Rosario Carmona en su tesis (Carmona 2020). Carmona propuso que el transcrito se sintetiza en otro tejido, donde se traducirá a proteína, y que solo la proteína sea transportada al grano de polen. Otro aspecto llamativo es que el transcrito de *Ole e 1*, el primer alérgeno que se identificó por ser el más abundante, es el segundo en abundancia en todos los estadios, detrás de *Ole e 14*. Lo notable de este hecho es que *Ole e 14* se ha descrito muy recientemente (Oeo-Santos et al. 2018) (de ahí la numeración), por lo que, a pesar de que el transcrito sea el más abundante, la proteína no se expresa de manera equivalente. Quizá este gen, una poligalacturonasa involucrada en elongación celular, abscisión de órganos, liberación de microsporas y crecimiento del tubo polínico, sea uno más de los que se encuentra en el polen seco para tenerlo preparado para responder rápidamente ante cualquier situación adversa o favorable del entorno.

### **En el desarrollo del tubo polínico intervienen más reguladores de los que se conocen**

Los análisis de coexpresión de genes sirven para detectar grupos de genes que respondan al mismo proceso o vía celular, o bien para localizar posibles reguladores comunes de todos ellos. Lo más destacable que se ha encontrado en este trabajo es el *cluster 3* del polen (figura 11A), que está implicado en procesos de respuesta a diferentes estreses, como son el calor, el frío y la salinidad, así como en el desarrollo del tubo polínico. Dicho *cluster* tiene el perfil de expresión que más aumenta con el tiempo (figura 10), aunque disminuye ligeramente a las 6 h, por lo que quizás el tubo ya se puede considerar formado a las 6 h. Adicionalmente, de los genes que pertenecen a la comunidad relacionada con el crecimiento del tubo polínico, dos de ellos tienen un dedo de zinc de tipo RING y otros dos son factores de transcripción Myb que según lo descrito en (Leydon et al. 2013) parecen necesarios para la diferenciación del tubo polínico.

Los *clusters 4* y *6* están involucrados en procesos de respuesta al frío, a la hipoxia y al estrés osmótico, así como al estrés oxidativo, respectivamente. Dichos *clusters* tienen una expresión elevada a 0 h, pero disminuyen a lo largo del desarrollo del tubo (hasta llegar a inhibirse, en el caso del *cluster 6*, figura 10), lo que nos confirma lo que habíamos intuido con los datos de expresión diferencial: hay respuestas al estrés que están activas en el polen maduro y desaparecen durante la germinación por no ser necesarias. Además, los genes que regulan el crecimiento del tubo polínico y la respuesta a estreses están conectados por dos genes de función desconocida de los que uno de ellos podría ser un factor de transcripción de la familia Myb.

Gracias al atlas de expresión de Picual hemos encontrado información pertinente de

genes que varían mucho su expresión en el polen, pero se expresan considerablemente más en semilla y regulan dos proteínas de interés que son las vicilinas y las proteínas F-box (Zafra et al. 2018; Jain et al. 2007). Esto nos hace pensar que quizá estas dos proteínas estudiadas en semilla no tengan una función única y sean relevantes en los procesos que se dan en otros tejidos. También encontramos otras proteínas reguladoras que son importantes y a las que no se les ha prestado atención.

### **El endospermo tiene muy activo el metabolismo para aportar nutrientes al embrión**

Hemos hallado que no se aprecian diferencias claras entre las proteínas de la testa y endospermo (T+End), o del embrión (EMBR), cuando se miran mediante un gel de proteínas, porque aparecen 6 proteínas muy mayoritarias que lo enmascaran todo (figura 13), por lo que es necesario un análisis transcripcional de las muestras para encontrar cambios más sutiles. Los análisis a tiempo 0 h indican que hay diferencias significativas entre T+End y EMBR (figuras 21 y 22), por lo que es necesario realizar un estudio a diferentes tiempos de maduración, que se encuentra en curso.

En líneas generales, de la comparación realizada deducimos que el endospermo tiene enriquecidos procesos de metabolismo como la biosíntesis de ácidos grasos y la síntesis de glúcidos para alimentar al embrión, así como el transporte transmembranario que es el proceso necesario para intercambiar esos nutrientes y procesos de respuesta que es la forma de que ambos tejidos se comuniquen (figura 21). En cuanto al embrión, se enriquecen procesos relacionados con la polaridad, ya que está en crecimiento y se va a definir la orientación de los diferentes tejidos, así como la especialización de dichos tejidos y órganos (figura 22). Estos resultados concuerdan con lo descrito en (Yan et al. 2014).

## 6. Limitaciones y líneas futuras

Sería interesante repetir los mismos análisis utilizando el genoma en lugar del transcriptoma. Por ello, estamos desarrollando un flujo de trabajo en el que partimos de las lecturas brutas y las alineamos contra un genoma de referencia de Picual, sin olvidar los pasos intermedios del control de la calidad de las lecturas brutas, el preprocesamiento, el control de calidad de las lecturas limpias, y la posterior cuantificación de la expresión génica. Una vez que este flujo esté puesto a punto, pretendemos implementarlo y volver a analizar las muestras estudiadas en este trabajo para poder analizar las posibles diferencias, así como muestras de Picual obtenidas de la base de datos del SRA con la intención de añadirlas a un atlas de expresión del olivo que tenemos en desarrollo (OliveAtlas).

También merece la pena mejorar la ortología con Arabidopsis para minimizar los casos en los que los ortólogos lejanos nos devuelvan información poco precisa y añadir una ortología con chopo como modelo de árbol. Sería conveniente mejorar las anotaciones de Picual para añadir un archivo GMT a la instalación local de iDEP como referencia para los análisis de enriquecimiento, lo que nos permitirá realizar los análisis funcionales directamente con Picual y no depender tanto de su ortología con Arabidopsis.

Creemos necesario estudiar en detalle todos esos genes clave que hemos determinado y que no se sabe lo que son en ninguna especie, ya que es probable que intervengan en procesos importantes en el desarrollo del tubo polínico.

## 7. Conclusiones

1. El genoma de las variedades de olivo europeas se parecen más entre sí que al genoma de acebuche, probablemente debido a la segregación geográfica y la plasticidad del olivo.

2. El perfil funcional del desarrollo del tubo polínico del polen de olivo indica que el polen maduro sin hidratar ya tiene activados una gran cantidad de genes en previsión de las situaciones favorables y adversas que se pueda encontrar. Una vez que el polen comienza a hidratarse, empieza a sintetizar energía y a desactivar los genes de respuesta a estrés que parecen innecesarios y que a partir de las 3 h, el tubo polínico ya está prácticamente formado, incluida la división del núcleo espermático.

3. Se han encontrado algunos transcritos de función desconocida con dominios de unión al ADN (seguramente sea reguladores), así como varias proteínas de la familia RING, además de dos genes que tienen mucho interés porque conectan la respuesta a estreses con el desarrollo del tubo polínico. Proponemos comenzar con la caracterización de la proteína con dedo de Zn de tipo RING Oleur061Scf1025g00027, los transcritos muy expresados en el polen de función desconocida Oleur061Scf1198g01026 y Oleur061Scf1515g03018 y los dos genes responsables de la relación entre el crecimiento del tubo polínico y la respuesta a estreses Oleur061Scf0076g02032.1 y Oleur061Scf5931-g00007.1.

4. Hemos caracterizado que el transcriptoma que se expresa en el polen está formado por 22.937 transcritos, que se podrá consultar en un atlas. En ellos hemos identificado los que codifican los 14 alérgenos de polen conocidos hasta la fecha. Tanto estos transcritos como los de muchas otras actividades importantes para el polen y el tubo polínico no presentan expresión diferencial, pero sí se expresan constantemente en el desarrollo del tubo polínico.

5. No hay diferencias prominentes entre las proteínas que se observan por separación electroforética en el endospermo y en el embrión de la semilla de olivo en reposo. Sin embargo, el perfil transcripcional de ambos tejidos es muy diferente, como atestiguan los 2.523 genes con expresión diferencial que demuestran que el endospermo tiene activados los procesos biológicos necesarios para aportar nutrientes a un embrión que comienza a diferenciarse y a establecer su polaridad.

## Material complementario

El material complementario se puede consultar en <https://doi.org/10.5281/zenodo.6777805> y consta de los siguientes archivos:

- inf\_comp\_01.pdf: informe PDF con las figuras complementarias (de la S1 a la S9) del enriquecimiento funcional de polen y semilla.
- inf\_comp\_02.html: informe HTML del script de R provisional que hemos usado para realizar correlaciones, agrupaciones (*clustering*) y redes con los datos de polen.
- inf\_comp\_03.html: informe HTML obtenido con el mismo script de R anterior para realizar correlaciones, agrupaciones (*clustering*) y redes con los datos de semilla.
- inf\_comp\_04.ods: informe ODS con cuatro pestañas que incluyen información sobre:
  - ST1 → ortología Arabidopsis-Picual.
    - id\_Arabidopsis: identificador en Arabidopsis.
    - id\_Picual: identificador en Picual.
  - ST2 → DEGs para la comparación P0H-P1H.
    - Regulacion: Down cuando ese gen se activa y Up cuando el gen se reprime.
    - id\_Picual: identificador en Picual
    - ExprMed: valor de expresión media.
    - Pvalue: valor de  $P$ .
    - Pvalueadj: valor de  $P$  ajustado.
  - ST3 → DEGs para la comparación P1H-P3H. Misma distribución que la anterior.
  - ST4 → DEGs para la comparación P3H-P6H. Misma distribución que la anterior.
  - ST5 → DEGs para la comparación P1H-P6H. Misma distribución que la anterior.
  - ST6 → DEGs para la comparación T+End-EMBR. Misma distribución que la anterior.

## 8. Bibliografía

- Alagna F, Caceres ME, Pandolfi S, Collani S, Mousavi S, Mariotti R, Cultrera NGM, Baldoni L, Barcaccia G. 2019. The paradox of self-fertile varieties in the context of self-incompatible genotypes in olive. *Front Plant Sci* **10**: 725.
- Alagna F, Cirilli M, Galla G, Carbone F, Daddiego L, Facella P, Lopez L, Colao C, Mariotti R, Cultrera N, et al. 2016. Transcript analysis and regulative events during flower development in olive (*Olea europaea* L.) ed. E. Albertini. *PLOS ONE* **11**: e0152943.
- Alché, J de D, Jiménez-López JC, Wang W, Castro-López AJ, Rodríguez-García MI. 2006. Biochemical characterization and cellular localization of 11 S type storage proteins in olive (*Olea europaea* L.) Seeds. *J Agric Food Chem* **54**: 5562–5570.
- Andrews S. 2010. *FastQC: a quality control tool for high throughput sequence data [Online]*. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**: 25–29.
- Badad O, Lakhssassi N, Zaid N, El Baze A, Zaid Y, Meksem J, Lightfoot DA, Tombuloglu H, Zaid EH, Unver T, et al. 2021. Genome wide MeDIP-Seq profiling of wild and cultivated olives trees suggests DNA methylation fingerprint on the sensory quality of olive oil. *Plants Basel Switz* **10**: 1405.
- Baldoni L, Tosti N, Ricciolini C, Belaj A, Arcioni S, Pannelli G, Germana MA, Mulas M, Porceddu A. 2006. Genetic structure of wild and cultivated olives in the central mediterranean basin. *Ann Bot* **98**: 935–942.
- Besnard G, Baradat P, Breton C, Khadari B, Bervillé A. 2001. Olive domestication from structure of oleasters and cultivars using nuclear RAPDs and mitochondrial RFLPs. *Genet Sel Evol* **33**: S251.
- Besnard G, Cheptou P, Debbaoui M, Lafont P, Hugueny B, Dupin J, Baali-Cherif D. 2020. Paternity tests support a diallelic self-incompatibility system in a wild olive (*Olea europaea* subsp. *laperrinei*, Oleaceae). *Ecol Evol* **10**: 1876–1888.
- Besnard G, Khadari B, Navascués M, Fernández-Mazuecos M, El Bakkali A, Arrigo N, Baali-Cherif D, Brunini-Bronzini de Caraffa V, Santoni S, Vargas P, et al. 2013. The complex history of the olive tree: from Late Quaternary diversification of Mediterranean lineages to primary domestication in the northern Levant. *Proc R Soc B Biol Sci* **280**: 20122833.
- Beuder S, Lara-Mondragón C, Dorchak A, MacAlister CA. 2022. SEC1A is a major Arabidopsis Sec1/Munc18 gene in vesicle trafficking during pollen tube tip growth. *Plant J* **110**: 1353–1369.
- Bolaños-Villegas P, Guo C-L, Jauh G-Y. 2015. Arabidopsis Qc-SNARE genes *BET11* and *BET12* are required for fertility and pollen tube elongation. *Bot Stud* **56**: 21.

- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120.
- Bredemeijer GMM. 1984. The role of peroxidases in pistil-pollen interactions. *Theor Appl Genet* **68**: 193–206.
- Breton C, Terral J-F, Pinatel C, Médail F, Bonhomme F, Bervillé A. 2009. The origins of the domestication of the olive tree. *C R Biol* **332**: 1059–1064.
- Breton CM, Bervillé A. 2012. New hypothesis elucidates self-incompatibility in the olive tree regarding S-alleles dominance relationships as in the sporophytic model. *C R Biol* **335**: 563–572.
- Breton CM, Farinelli D, Shafiq S, Heslop-Harrison JS, Sedgley M, Bervillé AJ. 2014. The self-incompatibility mating system of the olive (*Olea europaea* L.) functions with dominance between S-alleles. *Tree Genet Genomes* **10**: 1055–1067.
- Brunelle JL, Green R. 2014. Coomassie Blue Staining. In *Methods in Enzymology*, Vol. 541 of, pp. 161–167, Elsevier <https://linkinghub.elsevier.com/retrieve/pii/B9780124201194000136> (Accessed June 25, 2022).
- Carmona R. 2020. Estudio bioinformático del transcriptoma reproductivo del olivo (*Olea europaea* L.) y aplicaciones. Universidad de Málaga, Málaga.
- Carmona R, Zafra A, Seoane P, Castro AJ, Guerrero-Fernández D, Castillo-Castillo T, Medina-García A, Cánovas FM, Aldana-Montes JF, Navas-Delgado I, et al. 2015. ReprOlive: a database with linked data for the olive tree (*Olea europaea* L.) reproductive transcriptome. *Front Plant Sci* **6**. <https://www.frontiersin.org/article/10.3389/fpls.2015.00625> (Accessed June 20, 2022).
- Chen S, Xu K, Kong D, Wu L, Chen Q, Ma X, Ma S, Li T, Xie Q, Liu H, et al. 2022. Ubiquitin ligase OSRINGZF1 regulates drought resistance by controlling the turnover of OsPIP2 ;1. *Plant Biotechnol J* pbi.13857.
- Chhun T, Aya K, Asano K, Yamamoto E, Morinaka Y, Watanabe M, Kitano H, Ashikari M, Matsuoka M, Ueguchi-Tanaka M. 2007. Gibberellin regulates pollen viability and pollen tube growth in rice. *Plant Cell* **19**: 3876–3888.
- Consejo Oleícola Internacional, ed. 2000. *Catalogo mundial de variedades de olivo*. Consejo Oleícola Internacional, Madrid.
- Consejo Oleícola Internacional. 1996. *Enciclopedia mundial del olivo*. Consejo Oleícola Internacional, Barcelona.
- Cruz F, Julca I, Gómez-Garrido J, Loska D, Marcet-Houben M, Cano E, Galán B, Frias L, Ribeca P, Derdak S, et al. 2016. Genome sequence of the olive tree, *Olea europaea*. *GigaScience* **5**: s13742-016-0134–5.
- Cuevas González J. 2017. Biología reproductiva y cainismo en *Olea europaea* L. [https://www.ual.es/application/files/3715/3900/3646/Leccion\\_inaugural\\_curso\\_2018-2019\\_final.pdf](https://www.ual.es/application/files/3715/3900/3646/Leccion_inaugural_curso_2018-2019_final.pdf).

- da Costa-Nunes JA, Grossniklaus U. 2004. Unveiling the gene-expression profile of pollen. *Genome Biol* **5**: 205.
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008.
- Diez CM, Trujillo I, Martinez-Urdiroz N, Barranco D, Rallo L, Marfil P, Gaut BS. 2015. Olive domestication and diversification in the Mediterranean Basin. *New Phytol* **206**: 436–447.
- Dou X-Y, Yang K-Z, Ma Z-X, Chen L-Q, Zhang X-Q, Bai J-R, Ye D. 2016. AtTMEM18 plays important roles in pollen tube and vegetative growth in Arabidopsis. *J Integr Plant Biol* **58**: 679–692.
- Ewels P, Magnusson M, Lundin S, Käller M. 2016. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**: 3047–3048.
- Fernandez-Pozo N, Bombarely A. 2022. EasyGDB, a low-maintenance and highly customizable system to develop genomics portals ed. P. Robinson. *Bioinformatics* btac412.
- Ge SX, Jung D, Yao R. 2020. ShinyGO: a graphical gene-set enrichment tool for animals and plants. *Bioinformatics* **36**: 2628–2629.
- Ge SX, Son EW, Yao R. 2018. iDEP: an integrated web application for differential expression and pathway analysis of RNA-Seq data. *BMC Bioinformatics* **19**: 534.
- Golovkin M, Reddy ASN. 2003. A calmodulin-binding protein from Arabidopsis has an essential role in pollen germination. *Proc Natl Acad Sci* **100**: 10558–10563.
- Green PS. 2002. A revision of *Olea* L. (Oleaceae). *Kew Bull* **57**: 91.
- Han G, Qiao Z, Li Y, Yang Z, Wang C, Zhang Y, Liu L, Wang B. 2022. RING zinc finger proteins in plant abiotic stress tolerance. *Front Plant Sci* **13**: 877011.
- Ingram G, Gutierrez-Marcos J. 2015. Peptide signalling during angiosperm seed development. *J Exp Bot* **66**: 5151–5159.
- Irene S, Salvatore P, Adela O. 2010. Programmed-cell-death hallmarks in incompatible pollen and papillar stigma cells of *Olea europaea* L. under free pollination. *Plant Cell Rep* **29**: 561–572.
- Jain M, Nijhawan A, Arora R, Agarwal P, Ray S, Sharma P, Kapoor S, Tyagi AK, Khurana JP. 2007. F-Box proteins in rice. Genome-wide analysis, classification, temporal and spatial gene expression during panicle and seed development, and regulation by light and abiotic stress. *Plant Physiol* **143**: 1467–1483.
- Jiménez-Quesada MJ, Traverso JÁ, Zafra A, Jimenez-Lopez JC, Carmona R, Claros MG, de Dios Alché J. 2015. Identification and *in silico* analysis of NADPH oxidase homologues involved in allergy from an olive pollen transcriptome. In *Bioinformatics and Biomedical Engineering* (eds. F. Ortuño and I. Rojas), *Lecture Notes in Computer*

*Science*, pp. 450–459, Springer International Publishing, Cham.

- Jiménez-Ruiz J, Ramírez-Tejero JA, Fernández-Pozo N, Leyva-Pérez M de la O, Yan H, Rosa R de la, Belaj A, Montes E, Rodríguez-Ariza MO, Navarro F, et al. 2020. Transposon activation is a major driver in the genome evolution of cultivated olive trees (*Olea europaea* L.). *Plant Genome* **13**: e20010.
- Kessler SA, Lindner H, Jones DS, Grossniklaus U. 2015. Functional analysis of related CrRLK1L receptor-like kinases in pollen tube reception. *EMBO Rep* **16**: 107–115.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol* **37**: 907–915.
- Klaas M, Yang B, Bosch M, Thorogood D, Manzanares C, Armstead IP, Franklin FCH, Barth S. 2011. Progress towards elucidating the mechanisms of self-incompatibility in the grasses: further insights from studies in *Lolium*. *Ann Bot* **108**: 677–685.
- Koubouris GC, Breton CM, Metzidakis IT, Vasilakakis MD. 2014. Self-incompatibility and pollination relationships for four Greek olive cultivars. *Sci Hort* **176**: 91–96.
- Langgut D, Garfinkel Y. 2022. 7000-year-old evidence of fruit tree cultivation in the Jordan Valley, Israel. *Sci Rep* **12**: 7463.
- Leydon AR, Beale KM, Woroniecka K, Castner E, Chen J, Horgan C, Palanivelu R, Johnson MA. 2013. Three Myb transcription factors control pollen tube differentiation required for sperm release. *Curr Biol* **23**: 1209–1214.
- Liccardi G, D'Amato M, D'Amato G. 1996. Oleaceae pollinosis: a review. *Int Arch Allergy Immunol* **111**: 210–217.
- Luria G, Rutley N, Lazar I, Harper JF, Miller G. 2019. Direct analysis of pollen fitness by flow cytometry: implications for pollen response to stress. *Plant J* **98**: 942–952.
- Maestri D, Barrionuevo D, Bodoira R, Zafra A, Jiménez-López J, Alché J de D. 2019. Nutritional profile and nutraceutical components of olive (*Olea europaea* L.) seeds. *J Food Sci Technol* **56**: 4359–4370.
- Mariotti R, Pandolfi S, De Cauwer I, Saumitou-Laprade P, Vernet P, Rossi M, Baglivo F, Baldoni L, Mousavi S. 2021. Diallelic self-incompatibility is the main determinant of fertilization patterns in olive orchards. *Evol Appl* **14**: 983–995.
- McCormick S. 2004. Control of male gametophyte development. *Plant Cell* **16**: S142–S153.
- Mousavi S, de la Rosa R, Moukhli A, El Riachy M, Mariotti R, Torres M, Pierantozzi P, Stanzione V, Mastio V, Zaher H, et al. 2019. Plasticity of fruit and oil traits in olive among different environments. *Sci Rep* **9**: 16968.
- Muñoz-Sanz JV, Zuriaga E, Cruz-García F, McClure B, Romero C. 2020. Self-(In)compatibility systems: target traits for crop-production, plant breeding, and biotechnology. *Front Plant Sci* **11**. <https://www.frontiersin.org/article/10.3389/fpls.2020.00195> (Accessed June 1, 2022).

- Oeo-Santos C, Mas S, Quiralte J, Colás C, Blanca M, Fernández J, Feo Brito F, Villalba M, Barderas R. 2018. A hypoallergenic polygalacturonase isoform from olive pollen is implicated in pollen-pollen cross-reactivity. *Int Arch Allergy Immunol* **177**: 290–301.
- Perteau M, Perteau GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* **33**: 290–295.
- Pina C, Pinto F, Feijó JA, Becker JD. 2005. Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. *Plant Physiol* **138**: 744–756.
- Qin Y, Leydon AR, Manziello A, Pandey R, Mount D, Denic S, Vasic B, Johnson MA, Palanivelu R. 2009. Penetration of the stigma and style elicits a novel transcriptome in pollen tubes, pointing to genes critical for growth in a pistil. *PLOS Genet* **5**: e1000621.
- Quiapim AC, Brito MS, Bernardes LAS, daSilva I, Malavazi I, DePaoli HC, Molfetta-Machado JB, Giuliatti S, Goldman GH, Goldman MHS. 2009. Analysis of the *Nicotiana tabacum* stigma/style transcriptome reveals gene expression differences between wet and dry stigma species. *Plant Physiol* **149**: 1211–1230.
- Rallo L, Díez CM, Morales-Sillero A, Miho H, Priego-Capote F, Rallo P. 2018. Quality of olives: a focus on agricultural preharvest factors. *Sci Hort* **233**: 491–509.
- Rao G, Zhang J, Liu X, Lin C, Xin H, Xue L, Wang C. 2021. *De novo* assembly of a new *Olea europaea* genome accession using nanopore sequencing. *Hortic Res* **8**: 1–12.
- Rejón JD, Zienkiewicz A, Rodríguez-García MI, Castro AJ. 2012. Profiling and functional classification of esterases in olive (*Olea europaea*) pollen during germination. *Ann Bot* **110**: 1035–1045.
- Rogers HJ, Bate N, Combe J, Sullivan J, Sweetman J, Swan C, Lonsdale DM, Twell D. 2001. Functional analysis of cis-regulatory elements within the promoter of the tobacco late pollen gene g10. *Plant Mol Biol* **45**: 577–585.
- Rui Q, Wang J, Li Y, Tan X, Bao Y. 2020. Arabidopsis COG6 is essential for pollen tube growth and Golgi structure maintenance. *Biochem Biophys Res Commun* **528**: 447–452.
- Rutley N, Twell D. 2015. A decade of pollen transcriptomics. *Plant Reprod* **28**: 73–89.
- San Segundo-Acosta P, Oeo-Santos C, Benedé S, de los Ríos V, Navas A, Ruiz-Leon B, Moreno C, Pastor-Vargas C, Jurado A, Villalba M, et al. 2019. Delineation of the olive pollen proteome and its allergenome unmasks cyclophilin as a relevant cross-reactive allergen. *J Proteome Res* **18**: 3052–3066.
- Sanz-Cortés F, Martínez-Calvo J, Badenes ML, Bleiholder H, Hack H, Llacer G, Meier U. 2002. Phenological growth stages of olive trees (*Olea europaea*). *Ann Appl Biol* **140**: 151–157.
- Saumitou-Laprade P, Vernet P, Vekemans X, Billiard S, Gallina S, Essalouh L, Mhaïs A,

- Moukhli A, El Bakkali A, Barcaccia G, et al. 2017a. Elucidation of the genetic architecture of self-incompatibility in olive: evolutionary consequences and perspectives for orchard management. *Evol Appl* **10**: 867–880.
- Saumitou-Laprade P, Vernet P, Vekemans X, Castric V, Barcaccia G, Khadari B, Baldoni L. 2017b. Controlling for genetic identity of varieties, pollen contamination and stigma receptivity is essential to characterize the self-incompatibility system of *Olea europaea* L. *Evol Appl* **10**: 860–866.
- Seoane P. 2016. Creación y uso de los flujos de trabajo bioinformáticos. Universidad de Málaga, Málaga.
- Simeunovic A, Mair A, Wurzinger B, Teige M. 2016. Know where your clients are: subcellular localization and targets of calcium-dependent protein kinases. *J Exp Bot* **67**: 3855–3872.
- Singh DP, Jermakow AM, Swain SM. 2002. Gibberellins are required for seed development and pollen tube growth in *Arabidopsis*. *Plant Cell* **14**: 3133–3147.
- Singh N, Bhalla N. 2020. Moonlighting proteins. *Annu Rev Genet* **54**: 265–285.
- Sun J, Sun Y, Ahmed RI, Ren A, Xie M. 2019. Research progress on plant RING-finger proteins. *Genes* **10**: 973.
- Supek F, Bošnjak M, Škunca N, Šmuc T. 2011. REVIGO summarizes and visualizes long lists of Gene Ontology terms. *PLOS ONE* **6**: e21800.
- Talavera S, Buirra A, Quintanar A, García MA, Fernández Piedra P, Aedo C, eds. 2017. *Flora iberica. Vol. 16,2: Compositae (partim) / editores: S. Talavera (Sevilla), A. Buirra (Madrid), A. Quintanar (Madrid), M.Á. García (Madrid), M. Talavera (Sevilla), P. Fernández Piedra (Sevilla) & C. Aedo (Madrid)*. Real Jardín Botánico, CSIC, Madrid.
- Terral J-F, Alonso N, Capdevila RB i, Chatti N, Fabre L, Fiorentino G, Marival P, Jordá GP, Pradat B, Rovira N, et al. 2004. Historical biogeography of olive domestication (*Olea europaea* L.) as revealed by geometrical morphometry applied to biological and archaeological material. *J Biogeogr* **31**: 63–77.
- Unver T, Wu Z, Sterck L, Turktas M, Lohaus R, Li Z, Yang M, He L, Deng T, Escalante FJ, et al. 2017. Genome of wild olive and the evolution of oil biosynthesis. *Proc Natl Acad Sci* **114**: E9413–E9422.
- Wang J, Zhang D, Farooqi TJA, Ma L, Deng Y, Jia Z. 2019. The olive (*Olea europaea* L.) industry in China: its status, opportunities and challenges. *Agrofor Syst* **93**: 395–417.
- Wang L, Zhang J, Peng D, Tian Y, Zhao D, Ni W, Long J, Li J, Zeng Y, Wu Z, et al. 2022. High-Quality genome assembly of *Olea europaea* subsp. *cuspidata* provides insights into its resistance to fungal diseases in the summer rain belt in East Asia. *Front Plant Sci* **13**. <https://www.frontiersin.org/article/10.3389/fpls.2022.879822> (Accessed June 26, 2022).
- Wang Y, Zhang W-Z, Song L-F, Zou J-J, Su Z, Wu W-H. 2008. Transcriptome analyses

show changes in gene expression to accompany pollen germination and tube growth in *Arabidopsis*. *Plant Physiol* **148**: 1201–1211.

Yamamoto M, Nishio T. 2014. Commonalities and differences between Brassica and *Arabidopsis* self-incompatibility. *Hortic Res* **1**: 14054.

Yan D, Duermeyer L, Leoveanu C, Nambara E. 2014. The functions of the endosperm during seed germination. *Plant Cell Physiol* **55**: 1521–1533.

Yang H, You C, Yang S, Zhang Y, Yang F, Li X, Chen N, Luo Y, Hu X. 2021. The role of calcium/calcium-dependent protein kinases signal pathway in pollen tube growth. *Front Plant Sci* **12**. <https://www.frontiersin.org/article/10.3389/fpls.2021.633293> (Accessed June 3, 2022).

Zafra A, M'rani-Alaoui M, Lima E, Jimenez-Lopez JC, Alché J de D. 2018. Histological features of the olive seed and presence of 7S-type seed storage proteins as hallmarks of the olive fruit development. *Front Plant Sci* **9**. <https://www.frontiersin.org/article/10.3389/fpls.2018.01481> (Accessed June 22, 2022).

Zienkiewicz K, Castro AJ, Alché J de D, Zienkiewicz A, Suárez C, Rodríguez-García MI. 2010. Identification and localization of a caleosin in olive (*Olea europaea* L.) pollen during in vitro germination. *J Exp Bot* **61**: 1537–1546.