

Integrative transcriptomic and HPTL-ESI-MS analysis reveals a strong coordination between genes and pathways contributing to high erucic triacylglycerol biosynthesis during Pennycress (*Thlaspi arvense* L.) seed maturation

Ana Claver

EEAD - CSIC Zaragoza: Estacion Experimental de Aula Dei

María Ángeles Luján

EEAD - CSIC Zaragoza: Estacion Experimental de Aula Dei

José Manuel Escuín

Instituto de Carboquímica: Instituto de Carboquímica

María Savirón

University of Zaragoza: Universidad de Zaragoza

María Victoria López

EEAD - CSIC Zaragoza: Estacion Experimental de Aula Dei

Rafael Picorel

EEAD - CSIC Zaragoza: Estacion Experimental de Aula Dei

Carmen Jarne

Universidad de Zaragoza

Vicente Cebolla

ICB: Instituto de Carboquímica

Miguel Alfonso (✉ alfonso@eead.csic.es)

Estacion Experimental de Aula Dei <https://orcid.org/0000-0002-8190-5156>

Research Article

Keywords: *Thlaspi arvense*, seed, oil, TAG, DGAT, PDAT, VLCFAs, erucic acid.

Posted Date: August 7th, 2023

DOI: <https://doi.org/10.21203/rs.3.rs-3221172/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Integrative transcriptomic and HPTL-ESI-MS analysis reveals a strong coordination between genes and pathways contributing to high erucic triacylglycerol biosynthesis during Pennycress (*Thlaspi arvense* L.) seed maturation

Ana Claver¹, María Ángeles Luján¹, José Manuel Escuín², María Savirón³, M^a Victoria López⁴, Rafael Picorel¹, Carmen Jarne², Vicente L. Cebolla² and Miguel Alfonso¹

¹ Department of Plant Biology, EEAD-CSIC, Avda. Montañana 1005, 50059 Zaragoza (Spain)

² Instituto de Carboquímica, ICB-CSIC, C/ Miguel Luesma Castán 4, 50018 Zaragoza (Spain)

³CEQMA-CSIC-Universidad de Zaragoza, Facultad de Ciencias, c/ Pedro Cerbuna 12, 50009, Zaragoza (Spain).

⁴Department of Soil and Water Conservation, EEAD-CSIC, Avda. Montañana 1005, 50059 Zaragoza (Spain)

*** Correspondence:**

M. Alfonso, Department of Plant Biology, Estación Experimental de Aula Dei-Consejo Superior de Investigaciones Científicas (CSIC), Avda Montañana 1005, 50059 Zaragoza, SPAIN. Tel: +34 976 71 60 59; Fax: +34 976 716045; E-mail address: alfonso@eead.csic.es

Key message Transcriptomics and HPTLC-ESI-MS analysis identified the complex coordination between acyl-lipid biosynthetic pathways for high erucic TAG accumulation during Pennycress seed maturation

Abstract

Thlaspi arvense (Pennycress) is an emerging feedstock for biofuel production because of its high seed-oil content enriched in erucic acid (35-40% of total fatty acids). A transcriptomic analysis together with an HPTLC-ESI-MS and MSⁿ positional analysis characterization of the TAG fractions in five different seed maturation stages was performed to study the high erucic seed-oil biosynthesis in Pennycress. The RNA-Seq data showed a temporal pattern of regulation in which genes involved in the synthesis of carbon precursors and fatty acids were highly abundant at the earlier stages of maturation, then decreasing in the mature ones. Genes involved in TAG biosynthesis showed a complex temporal regulation in which *DGAT2* and *PDAT1* showed higher expression at the earlier stages of maturation, coincident with the rapid incorporation of 22:1 to TAG. Genes from the acyl-editing pathway showed higher expression at the earlier stages of seed maturation, suggesting a major role of this pathway at the initial stages of seed-oil biosynthesis. HPTLC-ESI-MS analysis showed an increase of TAG species containing VLCFAs (22:1 and 24:1) with 18:2 at *sn*-2 position with seed maturation. This increase was concomitant with an increase of *DGAT1* and *PDAT2* mRNA levels, suggesting a higher Kennedy pathway activity. However, the concerted regulation between DGAT and PDAT enzymes suggested a coordination between pathways, isoforms and substrate pools for the synthesis of high erucic TAG in Pennycress.

Keywords: *Thlaspi arvense*, seed, oil, TAG, DGAT, PDAT, VLCFAs, erucic acid.

INTRODUCTION

Field Pennycress (*Thlaspi arvense* L.) is a winter annual species that belongs to the Brassicaceae family. Pennycress has attracted the attention of researchers as a promising alternative oilseed feedstock for biodiesel production because of its high seed-oil content and fatty acid composition. Pennycress is a prolific seed producer (Fan et al., 2013; Claver et al., 2017). Pennycress seeds contain around 29-40% oil (w/w), depending on the varieties, which is twice the amount present in other oil commodities like soybean or sunflower, and very similar to that found in camelina (Moser, 2012; Claver et al., 2017; Altendorf et al., 2019; López et al., 2021). Because of its high seed-oil content and fatty acid profile, with high amount of erucic acid (22:1^{Δ13}; 30-35% of total fatty acids), Pennycress oil can be used for biodiesel and biojet production with excellent properties like high cetane number, good low temperature behaviour and low oxidation properties when compared to other oil commodities (Moser et al., 2009; Moser, 2012; Fan et al., 2013). Many research efforts are being held at the agronomical level, directed towards a future crop improvement focusing in some important agronomic traits like dormancy, vernalization, or seed dehiscence (Sedbrook et al., 2014; Cubins et al., 2019; López et al., 2021). At the molecular level, complete genomic sequencing (Chopra et al., 2018; Gang et al., 2021; García Navarrete et al., 2022; Nunn et al., 2022), and transcriptome assembly of Pennycress genes has been reported, providing tools for its breeding (Dorn et al., 2013; 2015). Other studies have focused in the metabolite profiling of Pennycress seed embryos (Tsogtbaatar et al., 2015; Johnston et al., 2022) or lipidomics (Romsdahl et al., 2023) providing information of lipid species accumulating in the Pennycress seed. However, the enzymes and pathways responsible of the high seed-oil and erucic acid content of the Pennycress seed are still poorly understood.

As a member of the Brassicaceae family, Pennycress is closely genetically related with the model plant *Arabidopsis thaliana*, *Camelina sativa*, or *Brassica napus*. In Brassicaceae, Very Long Chain Fatty Acids (VLCFAs), like eicosanoic acid (20:1^{Δ11}) or erucic acid (22:1^{Δ11}), are present in their seed-oils, although their content and distribution is highly variable among plant species. In *Arabidopsis*, 22:1 levels in seed lipids are very low (<2.5%), being 20:1 the major VLCFAs species representing a 15-20% of the total fatty acids in seeds (Sun et al., 2013; Li-Beisson et al., 2013; Claver et al., 2020). In other species, such as *Brassica napus*, *Crambe abyssinica* or *Thlaspi arvense*, the total erucic acid content in the seed can range from 39 to 60% (Sun et al., 2013; Claver et al., 2020).

As an example, while *T. arvense* shows a high 22:1 content, another *Thlaspidae* like *T. caerulescens* shows low 22:1 levels, similar to that from *Arabidopsis* (Claver et al., 2020). The reasons of this heterogeneity remain unclear. Our group recently performed a functional characterization of the Pennycress *TaFAE1* elongase, responsible of the biosynthesis of erucic acid in the endoplasmic reticulum (ER) through the sequential elongation of C18 acyl-CoA substrates to produce 20:1-CoA and 22:1-CoA (Ghanevati and Jaworski, 2001; Katavic et al., 2001; Bates et al., 2013). The complementation of a series of *Arabidopsis* mutant lines with the Pennycress *TaFAE1* gene indicated that the elongase from Pennycress showed higher affinity to 20:1-CoA substrates than the *Arabidopsis* one, indicating that different enzyme affinities might explain the different erucic acid content in their seed-oil (Claver et al., 2020). In fact, differences in substrate affinity have been reported for many enzymes of the seed-oil biosynthetic pathway in different plant species (Lager et al., 2013; Aznar-Moreno et al., 2015; Demski et al., 2019; Claver et al., 2020). These results illustrate the need to investigate the genes and pathways, as well as the different isoforms in each pathway, that contribute to TAG biosynthesis in Pennycress.

Triacylglycerol (TAG) was the major reservoir of erucic acid in Pennycress seeds, increasing during seed maturation (Claver et al., 2017; Romsdahl et al., 2023). TAG is the major fraction in plant seed-oils, representing an 80-90% of total seed lipids (Bates et al., 2013). Two pathways contribute to TAG biosynthesis in the seed. On one hand, TAG biosynthesis is performed by a series of enzymes (glycerol 3-phosphate acyltransferase, GPAT; lyso-phosphatidyl-acyltransferase LPAT; phosphatidic acid phosphatase, PAP; and acyl-CoA:diacylglycerol acyltransferase (DGAT) that perform the sequential acylation of the *sn-1*, *sn-2* and *sn-3* position of the glycerol backbone through the Kennedy pathway (Ohlrogge and Browse, 1995). DGAT enzymes are in charge of the final acylation to produce TAG, which activity has been shown to determine the carbon flow into TAG (Weselake et al., 2009; Li et al., 2010; Bates et al., 2013). On the other hand, acyl-editing reactions use an acyl group from the membrane phospholipid phosphatidyl-choline (PC) to its incorporation to DAG for TAG biosynthesis through the action of the phosphatidyl choline acyl-transferase (PDAT), (Bates and Browse, 2012; Wang et al., 2012; Bates et al., 2013). LysoPC, generated by the extraction of the acyl group, is converted again into PC through the action of lyso-phosphatidyl choline acyl-transferase (LPCAT). This PC/lysoPC pool acts as a reservoir for PC/DAG

interconversion by phosphatidylcholine-diacylglycerol-choline phosphotransferase (PDCT), being another pathway of diacylglycerol (DAG) supply for TAG synthesis (Bates and Browse, 2012; Wang et al., 2012; Bates et al., 2013). Not only differences in elongase activity, but also in the incorporation of VLCFA to TAG through both pathways have been pointed out as possible bottlenecks responsible of the different erucic acid content in plant seed-oils (Guan et al., 2013). In *Arabidopsis*, 40% of the PUFAs found in TAG are believed to be incorporated to TAG through the acyl-editing pathway (Lu et al., 2009). On the contrary, in *Crambe abyssinica*, a high erucic acid containing species, *sn-1* and *sn-3* positions of TAG used acyl groups incorporated outside the acyl-editing pathway, suggesting a major role of DGAT enzymes (Guan et al., 2013). Moreover, differences in substrate affinity among specific *Bn*DGAT2 isoforms to incorporate erucyl-CoA to TAG in Brassica cultivars with different erucic acid content, have been suggested to play a relevant role in the fatty acid composition of plant seed-oils (Demski et al., 2019).

In this work, we have studied seed-oil content and its fatty acid composition during Pennycress seed maturation both at the molecular and lipidomic levels in an attempt to identify the enzymes and pathways responsible of the high seed-oil content and erucic acid incorporation into the Pennycress seed-oil. To that end, particular effort was made in an HPTLC-ESI-MS study and MSⁿ positional analysis of the TAG fraction of the Pennycress seed-oil lipids during seed maturation, identifying the different acyl groups associated to TAG. Our data confirmed the rapid incorporation of 22:1 to TAG at the initial stages of seed maturation and the increase of TAG species enriched in VLCFAs like 20:1, 22:1 and even 24:1, particularly during the oil-filling stage of the Pennycress seed. Positional analysis showed that 18:2 was detected at *sn-2* position in the most abundant TAG species, independently of the maturation stage. Expression of genes involved in oil and fatty acid biosynthesis during the maturation of the Pennycress seed showed a complex pattern of regulation in which enzymes belonging to both the acyl-editing or Kennedy pathways were expressed in a concerted manner with changes between the earlier and the latter stages of seed maturation, suggesting a strong coordination between pathways, isoforms and substrate pools for seed-oil biosynthesis and the incorporation of VLCFAs to TAG in Pennycress.

MATERIALS AND METHODS

Plant materials

Pennycress seeds (*Thlaspi arvense* L.) from the SPRING32 germline were used for gene expression and lipidomic analysis. These seeds were obtained from the Nottingham Arabidopsis Stock Centre (NASC), UK. Seeds were germinated in plates on wet Whatman paper without addition of any other supplement. For germination, seeds were vernalized for 3 days at 4 °C and then moved to a growth chamber for additional 10-14 days. No vernalization treatment was required for fully development of flowers and seeds in this germline. Once germinated, seeds were transferred to pots containing a 75:25 mixture of substrate (peat moss, Kekkila White 420W: vermiculite) and grown in a bioclimatic chamber under a light intensity of 120-150 $\mu\text{mol m}^{-2} \text{s}^{-1}$, with a 16h/8h light/dark photoperiod at 22 °C and a relative humidity of 45%. For seed maturation studies, seeds from five developmental stages corresponding to GREEN (G), GREENYELLOW (GY), YELLOWGREEN (YG), YELLOW (Y) and MATURE (M) were chosen for analysis, similarly as described in Claver et al., (2017). Seeds separated from the pods, corresponding to these different maturation stages, were harvested, frozen in liquid nitrogen, and stored at -80 °C for further analysis unless indicated otherwise.

RNA isolation, cDNA synthesis and qPCR expression analysis

Total RNA was isolated from 0.1 g of *Thlaspi arvense* seeds from the five maturation stages analyzed (GREEN, GREENYELLOW, YELLOWGREEN, YELLOW and MATURE) using the CTAB-LiCl extraction method of Gasic et al. (2004). RNA concentration and integrity were measured in a Nanodrop 2000 UV-Vis Spectrophotometer (Thermo Scientific). cDNAs were synthesized from 3 μg of total RNA using ThermoScript Reverse Transcriptase (Life Technologies) and oligo dT primer, according to the manufacturer's instructions. Quantitative PCR (qRT-PCR) of target genes was performed using a 7500 Real Time PCR System (Applied Biosystems), SYBR Green Master Mix (Applied Biosystems), and specific primers (Supplementary Table 1). The Ct values were calculated relative to *ACT2* and *EF1 α* reference genes using $2^{-\Delta\Delta\text{Ct}}$ method (Livak and Schmittgen, 2001). Data were obtained from the analysis of at least three biological samples with three independent technical repeats for each sample.

RNA-Seq analysis.

RNA-Seq libraries were prepared and sequenced on an Illumina NovaSeq6000 at Novogene Ltd (www.novogene.uk). Ten libraries were obtained from two biological replicates of each of the five developmental stages that were studied in this work. Messenger RNA was purified from total RNA using poly-T oligo-attached magnetic beads. The first strand cDNA was synthesized using random hexamer primers, followed by the second strand cDNA synthesis using either dUTP for directional library or dTTP for non-directional library. The library was checked with Qubit and real-time PCR for quantification and bioanalyzer for size distribution detection. Quantified libraries were pooled and sequenced. Original image data file from high-throughput sequencing was transformed to sequenced reads by CASAVA. Raw data were stored in FASTQ(fq) files, containing sequences of reads and corresponding base quality. Once raw reads were cleaned, alignments were performed with HISAT2 (Mortazavi et al., 2008). Mapped regions were classified as exons, introns, or intergenic regions, and annotated with respect to the Pennycress reference genome (www.ncbi.nlm.gov/assembly/GCA_91186555.2), Nunn et al. (2022).

Gene expression level was estimated by FPKM values (short for the expected number of Fragments Per Kilobase of transcript sequence per Millions base pairs sequenced), (Mortazavi et al., 2008). Correlation of the gene expression levels between samples was estimated by Pearson coefficient greater than 0.92 and the R^2 greater than 0.8. Up-regulated and down-regulated genes were identified for each seed maturation stage comparison and co-expression Venn diagrams were generated. The screening criteria used for differential expressed genes was $\log_2(\text{FoldChange}) \geq 1$ and $\text{padj} \leq 0.05$. Similar expression patterns were clustered together using the fpkm values of genes and homogenized the row (Z-score). The overall results of FPKM cluster analysis, clustered using the $\log_2(\text{FPKM}+1)$ value, were generated. When required, heatmaps were generated using the online tool http://bar.utoronto.ca/ntools_heatmapper_plus.cgi using the $\log_2$ ratio of (FPKM + 1) in each sample pair.

We used the clusterProfiler (Yu G, 2012) software for enrichment analysis, including GO Enrichment, DO Enrichment, KEGG and Reactome database Enrichment. GO terms with $\text{padj} < 0.05$ were regarded as significant enrichment. In the results of the GO enrichment analysis, the most significant 30 Terms were selected for display. The

different colors represent the three GO subclasses of BP, CC, MF. KEGG pathways with $p\text{-adj} < 0.05$ were regarded as significant enrichment. In the KEGG enrichment results, the most significant 20 KEGG pathways were selected for display.

Lipid and fatty acid composition analysis

Total lipids were extracted from Pennycress seeds (0.1g) with chloroform:methanol (2:1, v:v) as described by Bligh and Dyer (1959). For total seed-oil quantification, we followed the method from Li et al., (2006) through direct whole seed transmethylation using triheptadecanoin (30-35 μg) as internal standard. Fatty acid methyl esters of total lipids were analysed by GC-FID as described in Claver et al., (2017).

HPTLC-MS analysis

Standards and chemicals

1-Oleoyl-*rac*-glycerol (C18:1, $\geq 99\%$, 111-03-5 CAS); 1,2-dioleoyl-*rac*-glycerol (C18:1/C18:1, $\geq 97\%$, 2442-61-7 CAS); glyceryl trioleate (C18:1/C18:1/C18:1, $>99\%$, 22-32-7 CAS); oleic acid (C18:1, $\geq 99\%$, 112-80-1 CAS); linoleic acid (C18:2, $\geq 99\%$, 60-33-3 CAS); erucic acid (C22:1, $\geq 99\%$, 112-86-7 CAS); methyl oleate (99%, 112-62-9 CAS); cholesteryl oleate ($\geq 98\%$, 303-43-5 CAS); cholesteryl stearate (96%, 35602-69-8 CAS), were purchased from Sigma-Aldrich (Madrid, Spain). Notation adopted for the identity of a molecular species was (x:y) where x is the carbon number of the fatty acid-chains of the molecule; and y corresponds to the total number of double bonds. HPLC-grade *n*-heptane (C7, 99.9%), tert-butyl methyl ether (MTBE, 99.9%), glacial acetic acid (AcH, 99.8%), methanol (MeOH, 99.9%) dichloromethane (DCM, 99.9%) and tetrahydrofuran (THF, 99.9%, without stabilizer) were purchased from PanReac (Barcelona, Spain). LC/MS- grade MeOH was purchased from Fisher (UK). HPTLC silica gel 60 plates (20 $\times$ 10 cm) without fluorescence indicator were employed, from Merck (Darmstadt, Germany).

High-Performance Thin-Layer chromatography-densitometry-tandem mass spectrometry

Instruments for sample application, chromatographic development, densitometry and HPTLC-MS coupling were from CAMAG (Muttens, Switzerland).

Plate pre-conditioning

Before being used, plates were immersed in THF for cleaning by diffusion. Subsequently, they were dried at 70° C and vacuum (50 mbar) for 15 min. Clean plates provided stable baselines, monitored by UV at 190 nm. To avoid possible impurities from the solvent itself being deposited uniformly on the plate, an additional pre-development with the chosen mobile phase was carried out, in the absence of sample, up to 90 mm migration distance (m.d.).

Standards and sample application

Solutions of each above individual standards were also applied in triplicate on the same plate (concentration: 0.33-2 mg/ml per standard in DCM:MeOH, (1:1 v:v); application of effective mass: 3 µg/band). In order to optimize the applied sample volume and thus save sample, the ATS4 filling quality method was used. In a given plate, minimal distance between tracks was 6 mm; distances from the lateral and lower plate edges were 10 mm. One or more tracks were left empty, as blanks. Five seed maturation stages (G, GY, YG, Y and M) were studied: Three samples (lipid extracts) per maturation stage corresponding to three different batches were analyzed by HPTLC-densitometry-tandem MS. Each sample was applied on three different plates (9 measurements per sample). Samples were dissolved (3-4 mg ml⁻¹) in DCM:MeOH (1:1 v:v). 4 µl/band were applied on the corresponding HPTLC silica gel plate, at least in triplicate, as 4-mm bands, by using the Automatic TLC Sampler (ATS4) system.

Chromatographic development and densitometric detection

Isocratic chromatographic development up to 70 mm-migration distance was performed in a horizontal developing chamber (20 x 10 cm) using a mixture of C7, MTBE and AcH (70:30:1, v:v:v). Three plates per lipid extract sample of each maturation stage were developed on different days. Detection was carried out using a TLC Scanner 3 densitometer in mode UV at 190 nm. Baseline of chromatograms was corrected manually. WinCATS software (v 1.4.3.6336) was used to control and process data from sample application, chromatography and densitometry.

Coupling with tandem mass spectrometry (MS)

TLC-MS Interface 2 was used from an extraction of each TG peak directly from the plate. A detail of elution-based interface description and operation can be found elsewhere

(Sancho-Albero et al., 2022). It was equipped with an oval, 4 x 2-mm extraction head that was positioned on the corresponding TG-band maximum, whose the x,y coordinates were provided by WinCats software, using a laser crosshair. Then the interface head was lowered. MeOH was delivered for band extraction at 0.2 mL/min by using a PU-2080 HPLC pump (Jasco, Tokyo, Japan). The eluate is directed through a 2- μ m stainless steel frit to remove silica gel and then sent to the mass spectrometer. Electrospray ESI-MS in positive mode (ESI⁺) was selected, and mass spectra were registered on an Ion trap Amazon Speed Spectrometer (Brüker Daltonics, Bremen, Germany). ESI⁺-MS was conducted with capillary and endplate offset voltages of -4500 and -500 V, 36 psi as pressure of the nebulizer gas (N₂), 6.0 L/min as flow rate of the drying gas (N₂) and 120 °C as drying gas temperature. Spectra were acquired in the *m/z* 70–1500 range at the ultra scan mode. Bruker Daltonics Trap Control software packages v 8.0 and Data Analysis v 5.2 were used to control the mass spectrometer and process data. For each TG peak, several HPTLC- ESI⁺-MS experiments were performed from replicate bands, and confirmation of identity was carried out by MS². These experiments were performed from different plates. The HPTLC-ESI⁺-MS operating conditions are specified for each case in the Results and Discussion section or in the Supplementary Information.

RESULTS

Pennycress seeds incorporate 22:1^{Δ13} to seed-oil at the early stages of seed maturation

Pennycress seeds are characterized by a high level of erucic acid (22:1^{Δ13}), which is the major fatty acid present in the Pennycress seed-oil, reaching values between 29-40% depending on the varieties (Claver et al., 2017; Altendorf et al., 2019). We already monitored seed-oil and fatty acid accumulation in two Pennycress germplines, NASC and FRENCH of European origin (Claver et al., 2017). In this work, the SPRING32 germline was used (Chopra et al., 2018; McGinn et al., 2019). Seeds from five developmental stages from the youngest to the final mature seeds were used and designated as GREEN (G, 12 DAF); GREENYELLOW (GY, 19 DAF); YELLOWGREEN (YG, 26 DAF); YELLOW (Y, 33DAF); and MATURE seeds (M, 45 DAF), (Fig. 1a). Total lipid content was analysed in each developmental stage on a fresh weight basis following the method by Li et al., (2006). Total lipid content from G to GY-YG almost doubled from average values of 48 µg of total lipids to 92 µg. Seed oil content still increased at the Y stage, reaching higher values (185 µg of total lipids) at the M stage (Fig. 1b). These results are similar to those reported previously in Pennycress for the NASC and FRENCH lines (Claver et al., 2017) and from *Crambe abyssinica*, another species with seed-oil content enriched in erucic acid (Furmanek et al., 2014) or in *Camelina sativa* (Marmon et al., 2017). Temporal changes in fatty acid composition were also analysed in total lipid fractions from each developmental stage. As expected, erucic acid (22:1^{Δ13}) was the major fatty acid species detected in all seed maturation stages. Interestingly, erucic acid levels were already high at the initial stage G (Fig. 1c), reaching average values of 28% of total fatty acids. Erucic acid levels increased during Pennycress seed maturation, peaking at 34-35% of total fatty acids at the GY and YG stages (Fig. 1c). Then erucic acid levels slightly decreased at the later stages of seed maturation reaching average values of 32-33% of total fatty acids (Fig. 1c). Other VLCFAs like 20:1 and 24:1 were also detected in all seed maturation stages, increasing their levels during the whole seed maturation process (Fig. 1c).

Differential gene expression at the different stages of Pennycress seed maturation.

To investigate at the molecular level the oil-biosynthetic pathway occurring during the whole seed maturation process in Pennycress, we followed two complementary strategies. On one hand, we performed an RNA-Seq analysis covering the complete transcriptome changes during maturation at the five stages described above and, secondly, we performed a qPCR gene expression analysis on selected genes involved in seed-oil and fatty acid biosynthesis, among other functions. For the transcriptome analysis, ten libraries corresponding to two biological replicates of the five different developmental stages were constructed. For each library, raw reads, clean reads, quality parameters as Q20 (%), Q30 (%) and QC (%), as well as the mapped percentage were first monitored. The results are available in Table S2. A 79.97 % of the clean reads were detected in exonic regions, while 2.71% and 17.30% were detected in intronic and intergenic regions, respectively. In general, and taking into account the latest annotation of the Pennycress genome available (García-Navarrete et al., 2022), which estimated 28,034 genes from which 27,213 corresponded to protein coding genes, our RNA-Seq analysis identified 20,015 protein coding genes that covers a 73.54 % of the total Pennycress genome. The quality of the data was tested through a Pearson correlation analysis, showing that all libraries from the biological replicates were highly related and, therefore, good for the gene expression analysis. When pairwise comparisons were analysed between seed maturation stages, the results indicated that 3,443 differentially expressed genes (DEGs) were identified in the GY vs G comparison, 6,406 were detected in the YG vs G comparison, 9,214 were obtained in the Y vs G comparison, and 10,994 were observed in the M vs G comparison (Fig. 2). Interestingly, the number of identified DEGs increased during seed maturation and were not clustered to the initial maturation stages (Fig. 2), indicating specific gene expression dynamics all-through the seed maturation process. The ratio of up-regulated to down-regulated genes also changed during seed maturation. While at the initial stages (GREEN or GREENYELLOW) the up-regulated genes were slightly higher than the down-regulated ones, at the late maturation stages (YELLOW or MATURE), the number of down-regulated genes was higher than that of the up-regulated genes (Fig. 2).

GO category and hierarchical clustering analysis of DEGs

Gene ontology (GO) category analysis was performed with the identified DEGs for each maturation stage comparison. We split the results into down-regulated and up-regulated genes to facilitate its interpretation. With respect to biological processes (BP), the GO analysis of down-regulated DEGs in the GY vs G comparison showed a high proportion

of genes associated with DNA: “DNA replication”, “DNA metabolic process” or “DNA conformation changes” and also with “photosynthesis”, among other categories (Fig. 3). This association of DEGs in the BP categories was also confirmed in the cellular component (CC) one. Again, DEGs associated to “photosynthesis”, “thylakoid”, “chromosome”, “nucleosome” or “DNA packaging” were found among the most represented associations. GO analysis of the up-regulated DEGs detected in the GY vs G comparison also showed that “defence response”, “carbohydrate biosynthesis” or “transport” and “membrane transport” categories were detected as the most relevant associations in the BP category (Fig. 3). When the CC analysis was performed, two associations, “lipid droplet” and “monolayer-surrounded lipid storage body”, were detected as the most represented ones. This could be consistent with the beginning of the seed-oil filling process as shown in Fig. 1b. Finally, it is particularly relevant the observation of two groups of up-regulated DEGs detected in the GY vs G comparison for the molecular function (MF) category, related with “transferase activity, transferring acyl groups”, (Fig. 3).

In the YG vs G comparison, the BP category analysis still showed DEGs groups related with “DNA replication” or “photosynthesis” among the most represented ones (Fig. 3b). However, other categories related with “peptide biosynthesis” or “peptide metabolism” were also detected. In the CC category analysis, DEGs related to “photosynthesis” or “photosynthetic structures” were still present. However, DEGs related with “non-membrane bound organelles” or “ribosomes” were the most represented genes. The analysis of the up-regulated DEGs showed some new groups of genes in the BP category, such as “fatty acid biosynthesis” or “fatty acid metabolism” as well as those related with “carbohydrate biosynthesis, or metabolism”, (Fig. 3). Again, “lipid droplet” and “monolayer surrounded lipid storage body”, genes related with the formation of lipid droplets and oil bodies, were the most represented groups (Fig. 3). The MF category for this comparison showed that DEG genes related with “acyl transferase activity” were still present at this stage. It is also interesting the identification of a group of DEGs related with “oxidoreductase activity”, “phosphatidyl-inositol kinase activity” and “transcription regulation activity” (Fig. 3).

The Y vs G stage comparison was also analysed. No great differences were found in the GO analysis of the down-regulated DEG genes of this maturation stage with respect to the YG vs G, which might not be surprising (Fig. 4). When the up-regulated DEGs were

analysed, “oligosaccharide metabolism and biosynthesis” genes were highly represented as well as “lipid metabolic process” ones (Fig. 4). “Lipid droplet” and “monolayer-surrounded lipid storage bodies” were still highly represented in the Y vs G comparison, suggesting that oil droplet formation was still active. As occurred in the YG vs G comparison, groups of genes with “acyl-transferase activity” were still detected (Fig. 4). Some relevant changes were observed when the GO analysis was performed on the M vs G comparison. Genes related with “lipid biosynthetic process” or “fatty acid biosynthetic process”, which were found in the up-regulated DEG list in the YG vs G or Y vs G comparisons, were found now at the down-regulated BP category in the M vs G comparison (Fig. 4). This might be consistent with the fact that oil-biosynthesis and oil filling might be reducing or even stopping at this seed maturation stage. Consistent with this, no genes related with “lipid droplet” or “oil bodies” formation were found in the GO analysis of the up-regulated DEGs (Fig. 4). The same was true for DEGs with acyl-group transferase activity in the MF category with respect to Y vs G or YG vs G (Fig. 4). On the contrary, many DEGs related to RNA processing or protein ubiquitination were detected, consistent with the end of the seed maturation process.

We performed a gene clustering analysis of the DEGs using the log2 of FPKM values. The results showed that the gene expression patterns could be adjusted to minimally 4 main clusters that could be in some cases sub-divided into new sub-clusters to a total of 11 different clusters. The results are shown in Fig. 5. Cluster 1 grouped all the genes that decreased their expression from the G to the M maturation stages. This Cluster 1 could be sub-divided in four sub-clusters in which sub-cluster 1a contained 390 genes showing a strong decrease from G to M stages (Fig. 5). Genes encoding proteins involved in photosynthesis, encoding lipid transfer proteins (LTPs) and acyl-carrier proteins (ACPs) were highly represented in this sub-cluster. Sub-clusters 1b and 1c grouped genes that decreased their expression during seed maturation with a completely different pattern; while those grouped in 1b (249 genes) decreased rapidly from G to GY or YG and then maintained their expression, those grouped in 1c (428 genes) showed high expression levels during the three initial stages of seed maturation, then decreasing at the Y and M final stages. Sub-cluster 1d (6581 genes), grouped most of the down-regulated genes in all stages of seed maturation, showing higher expression at the early stages of seed maturation, then decreasing in the most mature ones (Fig. 5). Genes encoding lipid transfer proteins (LTPs), Long Acyl CoA Synthetases (*LACS1*, *LACS4*, and *LACS9*),

Glycerol phosphate acyl-transferases like *GPAT1*, *GPAT6*, or *GPAT7*; Acyl-carrier proteins (*ACP1*, *ACP2*, *ACP4* or *ACP5*) or genes encoding fatty acid desaturases like the ER omega-3 desaturase *FAD3* or the plastidial desaturases *FAD4* and *FAD6* were present in this sub-cluster (Fig. 5).

Cluster 2 grouped all genes that increased from G to GY or YG stages and then decreased to Y and M final stages (Fig.5). Two sub-clusters, sub-cluster 2a (221 genes) and 2b (266 genes) could be identified by differences in the expression rates (Fig. 5). Many seed storage proteins were detected in this cluster as well as other genes like the *FAEI* elongase, responsible of the synthesis of 22:1, some lipid transfer proteins (*LTP5*, *LTP20*) and also glycerol 3-phosphate acyltransferases like *GPAT4*. Cluster 3 grouped those genes which increased their expression from the initial stages of seed maturation and maintained their expression level to the end of the maturation of the seed (Fig. 5). As occurred with Cluster 1, 4 sub-clusters could be differentiated: Sub-cluster 3a contained 4.365 genes. Interestingly, most of the genes involved in oil biosynthesis were detected in this cluster. This included *DGAT1*, *PDAT2*, *LPAT1*, two *GPATs* (*GPAT5* and *GPAT9*) and also interestingly genes encoding oleosins (*OLE1*, *OLE2*), OBAPs (*OBAP2B*) or Seipins (*SEI2*). The presence of these genes in this cluster was consistent with TAG biosynthesis and seed-oil filling as showed in Fig. 1. Sub-cluster 3b (226 genes), 3c (166 genes) and 3d (274 genes) was also differentiated. In sub-cluster 3c genes like *OBAP1A*, *OBAP2A*, *SEIPIN1*, all encoding proteins involved in lipid droplet accumulation were observed (Table S3). Finally, Cluster 4 (46 genes) grouped all the genes that increased dramatically their expression levels all through the seed maturation process, being higher at the mature stage (Fig. 5). Many late embryogenesis abundant proteins like *LEA1* were detected in this cluster. In general, the clustering analysis was consistent with the GO analysis, showing a decrease of the expression of genes encoding photosynthetic proteins, consistent with the loss of chlorophyll or plastids during seed maturation. On the other hand, the data suggested that fatty acid biosynthesis or modification was higher at the early stages of seed maturation (GREEN, GREENYELLOW) and then decreased. In addition, genes involved in TAG biosynthesis and oil accumulation increased their expression between GREEN to YELLOWGREEN stages and then decreased their expression with more or less intensity, consistent with seed maturation.

Expression dynamics of genes involved in carbohydrate metabolism and fatty acid biosynthesis and modification

Fatty acid biosynthesis is a plastidial process in which many different enzymes, including the two condensing enzymes, *KASC1* and *KASC2*, the two fatty acid ACP thioesterases, *FATA* and *FATB*, responsible of hydrolysing 16:0-ACP and 18:0-ACP substrates for export to the ER, acyl-carrier proteins (ACPs) or Long Acyl Chain Synthetases (LACS) are involved. Genes encoding most if not all these plastid associated enzymes were down-regulated during Pennycress seed maturation, with higher values at the G stage then decreasing to the M stage (Fig. 6). Similarly, several genes encoding 3-ketoacyl-CoA synthase family members, involved in the biosynthesis of VLCFA (Very Long Chain Fatty Acids), like *KCS18*, *KCS8* or *KCS16* were also down-regulated during Pennycress seed maturation showing an expression pattern similar to that of the FA biosynthetic genes (Fig. 6). Expression of these genes, more concretely that of *KCS18* (*FAE1*), was consistent with 20:1 and 22:1 fatty acid accumulation during Pennycress seed maturation (Fig. 1c). Several enzymes involved in fatty acid modification like the $\Delta 9$ acyl-lipid desaturases *ADS1* and *ADS2*, involved in the desaturation of VLCFAs; the ER desaturases *FAD2* and *FAD3*, responsible of the biosynthesis of 18:2 and 18:3 fatty acids, or the plastidial *FAD4* desaturase, responsible of 16:1 synthesis in the plastid also showed a similar down-regulation pattern along with Pennycress seed maturation (Fig. 6). Interestingly, *GPAT* genes, encoding the glycerol-3-phosphate acyl-transferases (GPAT), the first step in TAG biosynthesis through the Kennedy pathway, showed a complex pattern of regulation in which some isoforms (*GPAT1*, *GPAT2* or *GPAT4*) decreased during seed maturation (Fig. 6) while others like *GPAT5*, *GPAT9* or even *GPAT8* (increased from G to the YG stage), increased their expression upon seed maturation.

It is also worth mentioning that not only genes involved in fatty acid biosynthesis and modification showed specific expression patterns. Other genes involved in lipid transport, like *LTP4*, *LTP5*, *LTP6* and *LTP12* also showed a similar downregulated expression pattern, being higher at the earlier stages of seed maturation and then decreasing in the later ones (Fig. 6). Particularly, *LTP4* and *LTP6* showed an important modification of their expression values suggesting a relevant role in the transport of acyl-lipids in the seed.

Glycerol-3-phosphate and acetyl-CoA are the carbon sources necessary for FA and TAG biosynthesis in the seed. Several genes involved in carbon assimilation like glycerol-3-

phosphate dehydrogenases (G3PDHs), Pyruvate dehydrogenases (PyrDHs), phosphoenolpyruvate carboxylases (PEPC) or phosphoenolpyruvate carboxylases:kinases (PEPCK) were monitored. As shown in Fig. 6, all of them showed a similar down-regulation pattern with Pennycress seed maturation, showing higher expression values at the earlier stages of seed maturation (G, GY) to then decrease in the later ones (Y, M), consistent with those involved in FA biosynthesis.

Expression of genes encoding transcription factors (TFs)

390 genes encoding different types of TFs were identified in our RNA-Seq data (Supporting information 1). Among these data, the MYB-family was the most represented one, with 77 DEGs encoding MYB TFs (Fig. S1). Other families like bHLH or AP2-ERF were also highly represented (Fig. S1). *ABI5* and *bHLH109* were found as the most up-regulated TFs during Pennycress seed maturation; their expression was very low at the G stage but increased at the Y and M seed stages (Fig. S1). *ABI5* encodes a transcription factor involved in seed germination and signalling (Skubacz et al., 2016) while *bHLH109* encodes a protein involved in embryo formation (Nowak and Gaj, 2016). Conversely, *MYB60* and *bZIP2* genes, were found as the most downregulated ones (Fig. S1). *MYB60* controls stomatal opening while *bZIP2* is involved in DNA-dependent transcription (Fusconi et al., 2013). *MYB60* expression was high in the G stage where photosynthesis in the seed pod is still active. Similarly, higher transcription is expected in younger tissues than in mature ones. *WRINKLED1* is considered the master regulator of oil biosynthesis in plants (Cernac and Benning, 2004). Other TFs like *FUSCA3* and *LEC2* are upstream transcription factors acting on *WRI1* (Baud et al., 2007). The expression patterns of *WRI1*, *LEC2* or *FUS3*, were similar, showing higher expression at the earlier stages of seed development and decreasing at the later ones (Fig. S1). This pattern is consistent with the activation of genes involved in fatty acid and lipid biosynthesis during seed maturation.

Expression dynamics of genes involved in lipid biosynthesis during Pennycress seed maturation

At this point, we focused our analysis on those genes involved in lipid biosynthesis in the seed as well as the biosynthesis or modification of fatty acids incorporated to TAG during seed maturation, particularly VLCFAs like erucic acid. On one hand, we used the RNA-Seq data (normalized FPKM values) to monitor the expression of several selected genes. On the other hand, we performed a qPCR analysis on samples from each maturation stage of the same selected genes, comparing both expression data and also validating the RNA-Seq results. *TaDGAT1* and *TaDGAT2* genes, encode two diacyl-glycerol acyl transferases with high homology, 90,63 and 81.3 %, respectively with respect to the Arabidopsis *AtDGAT1* and *AtDGAT2* genes (Routaboul et al., 1999; Shockey et al., 2006). *AtDGAT1* is essential for seed-oil production through the Kennedy pathway (Routaboul et al., 1999) while *AtDGAT2* seemed to have high specificity towards unusual fatty acids like ricinoleic, vernolic, among others through this pathway (Shockey et al., 2006; Brugal et al., 2008; Li et al., 2010; Demski et al., 2019). Similarly, the lyso-phosphatidic acid-acyl transferase (*TaLPAT*) that participates in DAG biosynthesis in the Kennedy pathway was also included. In higher plants, LPATs are a multigenic family. In Arabidopsis, *AtLPAT1* has been involved in the synthesis of PA in the plastid (Kim and Huang, 2004) while *AtLPAT2* is involved in *de novo* synthesis of PA in the ER (Kim et al., 2005). Both LPAT genes were included in the analysis. The acyl-editing pathway was also monitored through the expression of several genes. *TaPDAT* encodes the phospholipid-diacylglycerol acyl transferase, that is the enzyme responsible of the biosynthesis of TAG from DAG produced through the acyl-editing pathway (Zhang et al., 2009). As occurred with *DGAT*, two *PDAT* genes were identified in the Pennycress genome. *TaPDAT1* and *TaPDAT2*, with high homology with their Arabidopsis orthologs. We also monitored the expression of two genes involved in the acyl-group exchange between PC and DAG: *TaPDCT*, encoding the PC-diacylglycerol-acyltransferase and the *TaLPCAT*, lyso-phosphatidyl-choline acyl transferase (Lager et al., 2013). Monitoring the expression of these genes covers the activity of the acyl-editing pathway at the molecular level. Other genes like the fatty acid elongase *TaFAE1*, responsible of 20:1-CoA and 22:1-CoA elongation from 18:1-CoA, *TaFAD2*, the omega-6 desaturase responsible of the biosynthesis of 18:2 at the RE and *WRINKLED1*, the TF involved in the control of many genes of the lipid biosynthetic pathway in the seed (Cernac and Benning, 2004; Baud et al., 2007) were also included in the analysis. The results are shown in Fig. 7.

Pennycress accumulates high levels of 22:1 during seed maturation (Claver et al., 2017; 2020). Expression of the *TaFAE1* gene, encoding the elongase responsible of 22:1 production, increased 3 to 5 fold between the initial G stage to the GY and YG maturation stages declining thereafter either in the RNA-Seq data and in the qPCR analysis (Fig. 7). This induction of *TaFAE1* at the initial stages of Pennycress seed maturation could be consistent with the rapid availability of 22:1-CoA in the acyl-CoA pool for its early incorporation to TAG as shown in Fig. 1c, or reported previously by our group in Pennycress accessions of European origin (Claver et al., 2017). A very similar gene expression profile was obtained for the *FAD2* desaturase, with a two-fold increase of mRNA levels obtained both at the RNA-Seq data and qPCR analysis at the GY stage to then decrease its mRNA levels upon Pennycress seed maturation (Fig. 7). This again was consistent with the high 18:2 levels in the Pennycress seed-oil at the early stages of seed maturation (Fig. 1c).

TaDGAT1 expression levels increased gradually during Pennycress seed maturation, particularly between the YG and Y stages, showing a 2,5-3 fold maximum increase at the Y stage (Fig. 7). This expression profile was again consistent with the oil-filling profile of the Pennycress seed (Fig. 1b). Interestingly, *TaDGAT2* expression results from both RNA-Seq and qPCR data showed higher mRNA levels at the initial maturation stages, G and GY, further declining upon seed maturation in the YG, Y and M stages (Fig. 7). It is worth mentioning that the FPKM values of both *DGAT1* and *DGAT2* genes indicated that *DGAT2* mRNA was more abundant than that of *DGAT1* at the G stage, suggesting a specific role of *TaDGAT2* in TAG biosynthesis, particularly at the initial stages of Pennycress seed maturation. Expression of *TaLPAT1* and *TaLPAT2* was also monitored. *TaLPAT1* followed a similar profile to that of *TaDGAT1*, with an mRNA increase from the YG stage to the Y and M stages between 2-3.5 fold, both in the RNA-Seq and qPCR data (Fig. 7). On the contrary, *TaLPAT2* showed very subtle changes in its expression, if any, during Pennycress seed maturation (Fig. 7).

The expression of genes encoding enzymes of the acyl-editing pathway was also monitored. As occurred with the two *DGAT* genes, the two *TaPDAT* genes analysed showed completely different expression patterns. It is worth mentioning that the *TaPDAT1* gene was not detected in the RNA-Seq data, and only the *TaPDAT2* gene was found. qPCR data indicated that *TaPDAT1* mRNA levels were high at the initial stages of seed development to then rapidly decline upon Pennycress seed maturation (Fig. 7).

On the contrary, the *TaPDAT2* gene showed a continuous increase in mRNA levels upon Pennycress seed maturation reaching maximum expression levels at the Y stage both in the qPCR and in RNA-Seq data. These results might suggest different roles for both PDAT enzymes at the early (*PDAT1*) or late (*PDAT2*) stages of seed maturation. Interestingly, the expression of the *TaPDCT* gene showed a similar gene expression pattern to that of *PDAT1*, with high expression values at the early stages of seed maturation (Fig. 7). The other enzyme of the acyl editing pathway, *TaLPCAT*, responsible of the re-incorporation of an acyl group to PC, maintained its expression levels between the G to the YG stage to then slowly decrease at the Y and M stages, suggesting that *LPCAT* expression and/or activity might not be limiting for seed-oil accumulation

We also monitored the expression of genes involved in lipid droplet and oil body formation. These oil droplets increase their number and size upon seed maturation (Farese and Walther, 2009). Genes encoding two oleosins, *OLE1* and *OLE2* showed a similar increase in mRNA levels both in the RNA-Seq data and in the qPCR analysis, with maximum values between the YG and Y stages, in the moment with the highest oil-filling occurring in the seed (Fig. S2). Interestingly, *OBAP1a* showed a similar increase in mRNA levels during seed maturation although its expression levels were much higher than OLEOSINs (15-30 fold at the Y stage; Fig. S2). Nevertheless, the expression of these genes was consistent with oil-filling and the expression of lipid biosynthetic genes during Pennycress seed maturation.

HPTLC-ESI-MS characterization of the TAG pool composition during Pennycress seed maturation.

To further monitor how the erucic acid incorporated to TAG, we used a lipidomics approach based in the application of HPTLC - UV densitometry - MS to analyse the chemical composition of the different TAG species at each seed developmental stage. This analytical platform has proven to be useful for lipidomic analysis in complex lipid mixtures (Jarne et al., 2018; Jarne et al., 2021; Sancho-Albero et al., 2022). Our analysis focused on TAGs as it constitutes the major fraction of the total seed lipids in Pennycress (Claver et al., 2017; Romsdahl et al., 2022). Two different lipid extract samples per stage (G, GY, YG, Y and M), which corresponded to two extraction batches, were analyzed by HPTLC-densitometry-MS. Percentages of TAGs in samples, as well as intra- and inter-

plate HPTLC repeatability results (expressed in Area counts) for the separated TAG peaks are presented in Table 1, including the mean, the relative standard deviation (RSD%), and the coefficient of variation for a confidence interval of 95% (CV).

The selected isocratic development in acidic medium (C7-TMBE-AcH, 70:30:1, v/v/v) allowed to separate at baseline most of neutral lipid families in samples, e.g. mono-(MG), di- (DG), tri-acylglycerides (TG), fatty acids (FA), fatty acyl-(FAE) and cholesteryl esters (ChOE), and phosphatidylcholine (PC, at the application point), over a total m.d. of 70 mm migration distance (m.d.). HPTLC separation was first tested on silica gel plates using the standards mentioned in Materials and Methods. Chromatograms related to standards are shown in Fig. S3. HPTLC chromatograms corresponding to samples at the different maturation stages, detected at UV 190 nm. Development conditions were selected to clearly separate TAG (55 mm, m.d.) from the other neutral lipid families (Fig. S4).

Identification of TAG species by HPTLC-Ion Trap MS

In addition to comparing the migration distances of the corresponding TAG standards and those of the separated bands in the samples, the attribution of identity to TAG class and the structural identification of their molecular species was performed by collecting the HPTLC-ESI⁺-MS and MS² spectra directly from the chromatographic plate using the described interface. Fig. 8 shows the HPTLC-ESI⁺-MS spectra of TAG zones corresponding to each Pennycress seed maturation stage. For all samples, mass spectra were recorded at the same ionization time in order to compare relative ion intensities of different TAG species in each maturation stage. In general, eight major TAG species were identified in all seed maturation stages with maxima intensities at m/z 853.8; 879.8; 907.8; 935.9; 959.9; 989.9; 1017.9 and 1046.0. TAGs are weakly basic esters and, under the HPTLC and ESI⁺ conditions readily lead to the formation of [TAG+Na]⁺, which in our case can be fragmented to yield structural information (Jarne et al., 2018; Jarne et al., 2021; Sancho-Albero et al., 2022). Thus, the fragmentation pattern of sodium adducts of TAG species usually contains the following fragment ions: [M+Na-R_xCOOH]⁺, corresponding to the losses of fatty acyl substituents as fatty acids, and [M+Na-R_xCOONa]⁺, corresponding to the losses of fatty acyl substituents as their sodium salts (with x= 1, 2 and 3), (Han, 2016). Ions m/z corresponding to TAG species

are reported in Table 2 along with their fragmentation patterns (MS²). Essentially, the TAG species identified were similar, with some slight variation, to those reported using a reversed-phase HPLC and a triple quadrupole MS technology, (Romsdahl et al., 2022). These results validate the use of HPTLC-ESI-MS technology for the analysis of TAG species in complex lipid mixtures like Pennycress seed lipid fractions.

Relative abundance of TAG species with the maturation stages

For all samples, mass spectra were recorded at the same ionization time in order to compare relative ion intensities of different TAG species in each maturation stage. Although a quantitative analysis and a direct comparison of intensities in different plates is excluded, there is an ion intensity-concentration relationship for each sample that is supported by several reasons. In general, ESI-MS response factor of lipids mostly depends on the head class group and minimally on its aliphatic chain composition during a full-scan mass. Therefore, the influence of fatty acyl chain length is far less significant (Han, 2016). This means that we can hypothesize that the TAG species that are included in the chromatographic corresponding to TAG family from a given sample have a similar response per unit mass. Moreover, ESI ionization in HPTLC-densitometry-MS was carried out under strictly the same conditions for each TAG species. For all samples, mass spectra were recorded at the same ionization time in order to compare relative ion intensities of different TAG species in each maturation stage. The HPTLC mobile phase is removed before densitometry and subsequent automated band extraction was performed using the same solvent (MeOH) in all cases. No solvent gradients were employed for ESI, avoiding in this manner the gradient effect on nebulization and lipid response. Finally, the surface extracted on the HPTLC-separated TAG band by the interface in each run is the same and is representative of the separated band for comparative purposes. Therefore, ion intensities from ESI⁺-MS spectra of each sample corresponding to the maturation stages can be compared. Fig. 9 shows normalized intensities of ions corresponding to the identified major TAG species. Their relative abundance varied significantly during seed maturation. Thus, at the early G and GY stages, TAG species like 58:5, 60:4 and 62:4 containing 20:1 and 22:1 fatty acids, were the most abundant species detected in the MS analysis (Fig. 9a and 9b). The presence of 22:1 in the most abundant TAG species confirmed the rapid incorporation of erucic acid to TAG at the early stages of seed maturation (G, GY), (Claver et al., 2017). Upon seed

maturation, the TAG composition of YG and Y stages showed a change with respect to the G and GY initial stage: an increase of TAG 62:4 corresponding to the specie containing 22:1/18:2/22:1 can be observed with regard to the other species (Fig. 9c and 9d). Finally, at mature stage (M), intensities of TAG 58:5, 60:4 and 62:4 showed similar high intensities (Fig. 9e). These results indicated that upon maturation, an increase in TAG species containing VLCFAs like 22:1 or 20:1 occurred. The presence of (C24:1)FA in the TAG 64:4 (refer Table 2, ion at 1046.0 m/z) was consistent with the fatty acid composition of total lipids from mature seeds and our previous TLC-GC TAG analysis in NASC and FRENCH lines (Claver et al., 2017). These data clearly indicated that VLCFAs are the major fatty acid species esterified to TAG at the end of the Pennycress seed maturation showing that erucic as well as 20:1 or 24:1 incorporation to TAG were boosted upon oil-filling stages.

***sn*-positional analysis**

At this point, we decided to perform a positional analysis of the acyl moieties esterified to TAG in Pennycress. This analysis might provide useful information about the selectivity and/or specificity of the enzymes involved in TAG biosynthesis and, therefore, about the pathway responsible of its synthesis. According to Ramaley et al., (2015), $[M+Na]^+$ adducts from ion trap are the most suitable for TAG regioisomer analysis as they produced the most consistent level of positional sensitivity for the fragmentation. Methods for TAG isomers identification were described utilizing differential fragmentation efficiency of FAs from different *sn*-positions. The preferential loss of the fatty acid at positions *sn*-1/3 seems to be general for TAG molecules regardless of energy and instrumentation employed (Hsu et al., 2010). This is consistent with the notion that the α -hydrogen of the fatty acid substituent at *sn*-2, is more labile than that located at fatty acid substituent at *sn*-1 and *sn*-3, which leads to the formation of two ions of similar abundance corresponding to the losses of the fatty acids substituents at *sn*-1 and at *sn*-3 and those are significantly more abundant than the ion corresponding to the loss of the fatty acid substituent at *sn*-2 (Hsu et al., 2010). Thus, the intensities of the resulting fragment ions reflect the FA distribution in the glycerol backbone.

TAG regioisomers can be identified with MSⁿ methods, but analysis of TAG enantiomers is not possible because the fragmentation methods cannot distinguish

between *sn*-1 and *sn*-3 FAs due to the identical fragmentation efficiencies of FAs from these positions. In natural products, many isobaric TAG species, including isomers may produce shared isobaric fragment ions, making the analysis of TAG regioisomers even more challenging. However, on several occasions where the most abundant ion comes from a single triad of FA composition, we have been able to identify the FA in the *sn*-2 position. Therefore, ion at m/z 1017.9 corresponds to the sodium adduct of TAG(C62:4) $[C_{65}H_{118}O_6Na]^+$. As the stability of $[M+Na]^+$ was high, a consecutive fragmentation was achieved in the ion-trap MS to have verification of identity. Hence, the respective HPTLC-ESI⁺-MS/MS spectrum of the precursor ion at m/z 1017.9 showed two ion products corresponding to losses of fatty acyl substituents as fatty acids: at m/z 679.6 (most intense) which corresponds to $[M+Na-R_{1,3}COOH]^+$, $R_{1,3}= C(22:1)FA$, and at m/z 737.62 which corresponds to $[M+Na-R_2COOH]^+$, $R_2= C(18:2)FA$; and two much less abundant ions products corresponding to losses of fatty acyl substituents as their sodium salts: 657.6 m/z which corresponds to $[M+Na-R_{1,3}COONa]^+$, $R_{1,3} = C(22:1)FA$, and 715.7 m/z which corresponds to $[M+Na-R_2COONa]^+$, $R_2= C(18:2)FA$. ESI MS/MS spectra were carried out using He as the collision gas, an optimal amplitude voltage of 0.6 V and an isolation width for the precursor ion of 1 m/z units (Fig. 10). Results are consistent with a TAG structure of 22:1/18:2/22:1 with linoleic acid at the *sn*-2 position. This TAG is already present at the youngest state, in all stages, and becoming the most important in GY, YG, Y and M. It is also possible to identify *sn*-2 position in the ion at m/z 989.8 and in the ion at m/z 1046.0. In the others, it is more difficult to identify the *sn*-2 position since several TAG species can contribute to the same ion, as observed in the ESI mass spectra (Fig. S5). This analysis should not be used as the basis for excluding the presence of other isomers.

DISCUSSION

Manipulation of plant seed-oils has attracted the interest of breeders and biotechnologists because their multiple applications for food sector, industrial uses and more recently for biofuel production. However, alterations of plant seed-oil composition are diffculted by the complexity of the biosynthetic pathways that involve transport of acyl groups between different cellular organelles, enzymes acting on different pathways or isoforms or enzymes performing similar activities acting in a same pathway. The knowledge of these biosynthetic pathways is particularly relevant in plant species that accumulate unusual

fatty acids. This is the case of Pennycress, an emerging feedstock for biofuel production. Pennycress is characterized by a high seed-oil content enriched in erucic acid (Claver et al., 2017; Altendorf et al., 2019), with a completely different composition to that from other Brassicaceae like *Arabidopsis* or *Camelina* to which Pennycress is phylogenetically related. The different enzymes acting in the Pennycress seed-oil biosynthetic pathways that incorporate high amounts of erucic acid to TAG as well as their concerted regulation are still far from being understood. In this work, we have addressed this question through an integrative transcriptomic (RNA-Seq) and qPCR expression analysis together with an HPTLC-ESI-MS approach to analyse the different TAG species accumulating during seed maturation in order to identify, first, which genes of which seed-oil biosynthetic pathway are acting during the different stages of seed maturation and, secondly, correlate these results with the incorporation of erucic acid and other acyl groups to TAG during the whole seed maturation process.

In this work we performed a whole genome RNA-Seq analysis of gene expression in five different maturation stages, GREEN, GREENYELLOW, YELLOWGREEN, YELLOW and MATURE, that cover the whole seed maturation process. A recent transcriptome analysis was reported in Pennycress in natural variants with differences in seed-oil content (Arias et al., 2023). In their study, two early developmental stages in that might correspond to our initial GREEN stage and an even earlier stage were used (Arias et al., 2023). The analysis of the five different developmental stages in this work has allowed us to perform a complete study of the temporal pattern of gene expression during the whole process of seed maturation, observing the fluctuations between pathways and between enzymes of the same pathway from the earlier (GREEN, GREENYELLOW) to the late (YELLOW, MATURE) stages of seed maturation. This is illustrated in the data from Fig. 2a, where the number of DEGs increased from 3,443 at the GREENYELLOW vs GREEN comparison to 10,994 DEGs in the MATURE vs GREEN comparison. These fluctuations were particularly relevant in those genes encoding enzymes and isoforms from the two main seed-oil biosynthetic pathways, providing new information of the TAG biosynthetic pathway in Pennycress.

In general, our RNA-Seq data showed that genes encoding enzymes involved in carbon assimilation like glycerol-3-phosphate dehydrogenases, pyruvate dehydrogenases, or phosphoenol pyruvate decarboxylases were highly expressed at the early stages of seed maturation and then decreased at the latter stages of seed maturation (Figs. 3, 4, 5 and 6).

The same occurred with those involved in fatty acid biosynthesis like FATA, FATB, KAS1, KAS2 or LACS, or those encoding acyl-ACP carrier proteins involved in the transfer of acyl groups from the plastid to the ER. This pattern is consistent with the generation of carbon precursors for fatty acid biosynthesis in the plastid and the synthesis of fatty acids prior to their incorporation to acyl-lipids. Other genes that showed high expression values at the GREEN stage were those encoding photosynthetic proteins from PSII or PSI as well as other photosynthetic membrane formation or processes (Figs. 3 and 4). Expression of these genes also decreased during Pennycress seed maturation consistent with the loss of chlorophyll and plastids as seen in Fig. 1a. This high expression of photosynthetic genes at the earlier stages of seed maturation is consistent with previous data from Arias et al., (2023) and might suggest a role of carbon assimilation by photosynthesis for fatty acid biosynthesis at the earlier stages of seed maturation. Several genes encoding lipid transfer proteins (LTPs), like *LTP4*, *LTP5* and *LTP6* showed also higher expression at the earlier stages of seed development to decrease in the mature stage. The most interesting one was *LTP4*, where the differences in expression were very high (Fig. 6). It is worth mentioning that the expression of these genes from lipid transfer proteins was different to that from those involved in oil bodies and lipid droplet formation, like *OLE1*, *OLE2* and *OBAP1A*, that peaked at the intermediate stages of seed maturation, GREENYELLOW, YELLOWGREEN or even YELLOW, coincident with the oil-filling stage in the seed (Figs. 3, 4 and Fig. S2), indicating a temporal separation between the transfer of acyl-groups and the synthesis of lipid bodies.

Most if not all the genes encoding enzymes participating in either the Kennedy or the acyl-editing pathways were covered by our RNA-Seq data and qPCR analysis validation. This included many isoforms of genes participating in each of the two pathways like DGATs, PDATs or LPATs. We correlated these expression data with the HPTLC-ESI-MS results analysing the incorporation of 22:1 and other acyl groups to TAG. Our data showed that 22:1 was rapidly incorporated into TAG at the early stages of seed development as seen in the fatty acid total lipid content (Fig. 1c) or the characterization of the different TAG species during Pennycress seed maturation by HPTLC-MS analysis (Figs. 8 and 9; Table 2). 58:5 (18:1/18:2/22:1) was the major TAG species at this stage (Figs. 8 and 9). Other species like 60:4 (20:1/18:2/22:1) or 62:4 (22:1/18:2/22:1) were also present. These results confirmed our previous TLC-GC data (Claver et al., 2017) or those recently obtained using MS quadrupole analysis (Romsdahl et al., 2023). VLCFA-

containing species increased during Pennycress seed maturation; at the GREENYELLOW or YELLOW stages, 62:4 (22:1/18:2/22:1) was now the major TAG species, with a decrease of 58:5 or even 60:4 (Figs, 8 and 9). This increase in 22:1 incorporation to TAG was consistent with the expression of the *TaFAE1* elongase, that increased from GREEN to YELLOWGREEN, up to 3-4 fold, coinciding with the oil filling of the seed (Fig. 7). Moreover, the expression of the FAD2 desaturase gene, encoding the omega-3 endoplasmic reticulum desaturase that synthesizes 18:2, also followed that from the FAE1 elongase (Fig. 7). These expression values were also consistent with the presence of 18:2 at the *sn*-2 position in TAG species in all seed maturation stages (Fig.10 and Fig. S5). It is worth mentioning that both the elongase and the FAD2 desaturase act on the same substrate but on different moieties, while the elongase acts on 18:1 esterified to ACP prior its incorporation to lipids, the desaturases act on fatty acids (18:1) already esterified to lipids (Ghanevati and Jaworski, 2001; Ohlrogge and Browse, 1995), suggesting that both enzymes could be using separate substrate pools.

DGAT1 and DGAT2 are the main acyltransferases acting on the Kennedy pathway for the last acylation of TAG (Weselake et al., 2009; Li et al., 2012; Bates et al., 2013). Analysis of Arabidopsis mutants indicated that DGAT1 was the major acyltransferase involved in TAG biosynthesis (Katavic et al., 1995; Regmi et al., 2020), while the role of DGAT2 is less understood although it has been related with the incorporation of unusual fatty acids to TAG (Demsky et al., 2019). Both RNA-Seq and qPCR analysis showed a complete opposite pattern of expression of both genes in Pennycress during seed maturation. *DGAT2* showed higher expression at the earlier stages, then decreasing in the later ones while *TaDGAT1* expression showed 3.5 to 4.5-fold increases at the YELLOW and MATURE stages, consistent with the oil-filling of the seed and the detection of 62:4 TAG species as the major ones (Fig. 7). A correlation of both observations, suggested a coordination between both DGAT enzymes in which *TaDGAT2* may help to the rapid incorporation to TAG of 22:1 at the earlier stages of seed development while *TaDGAT1* would be responsible of the major acyl-transferase activity of the Kennedy pathway during seed maturation and oil filling. This hypothesis was consistent with previous observations in another high erucic acid containing species like *Crambe abyssinica* (Furmanek et al., 2014). In that work, although significant PDAT activity (10% of the

total acyl transferase activity) was detected, particularly at the stages of rapid seed-oil accumulation, specific activity of DGAT enzymes on erucoyl-CoA was higher following seed maturation compared to the early developmental stages (Furmanek et al., 2014). Furthermore, lipidomic and mutational analysis of *DGAT* and *PDAT* genes in *Camelina sativa* (an 18:3 enriched TAG containing species) also suggested that TAG biosynthesis was dominated by DGAT1 in the cotyledonal tissues (Marmon et al., 2017). This role of *TaDGAT1* would be supported by the expression pattern of the *TaLPAT* genes, particularly *TaLPAT1* (Fig. 7) or some *GPAT* genes (*GPAT5*, *GPAT8* and *GPAT9*, Fig. 6). Interestingly, although pennycress is a VLCFA containing species in TAG, our positional analysis detected 18:2 at the *sn*-2 position in TAG (Figs 10 and S5). Species with uncommon acyl moieties at the *sn*-2 position, like eruyl-CoA, possess additional seed-specific LPATs (Laurant and Huang, 1992). Our results do not support this observation in Pennycress. Nevertheless, our expression data also support a relevant role of the acyl-editing pathway in the incorporation of acyl groups and 22:1 to TAG. As occurred with the two *DGAT* genes, expression of the two *PDAT* genes showed a complete opposite gene expression pattern. *TaPDAT1* showed high expression values at the GREEN initial stage of seed maturation, being then rapidly downregulated in the rest of the stages (Fig. 7). On the contrary, expression of *TaPDAT2* showed an increase in mRNA levels with maturation, showing a 3.5-fold increase at the YELLOW stage (Fig. 7). Interestingly, the expression of the *PDCT* gene showed a similar expression pattern to that of *PDAT1* (Fig. 7). Expression of the *LPCAT* gene did not show significant changes during Pennycress seed maturation (Fig. 7). These results altogether suggested that the acyl-editing pathway should be important at the early stages of seed development, consistent with a higher proportion of 18:2 in TAG (Figs. 8 and 9, Table 2). It is worth mentioning that in our previous TLC-GC analysis of seed lipid fractions during seed maturation, we detected DAG with 22:1 (or 20:1) in all seed maturation stages, although 18:2 was the major FA esterified to DAG in most stages (Claver et al., 2017). This DAG pool could be used either for DGAT or PDAT for TAG biosynthesis. In that sense, our expression data support a role of PDAT1 in the incorporation of 22:1 to TAG, particularly at the earlier stages of seed development. In *Arabidopsis* (mostly containing 18:1 and 18:2 fatty acids esterified to TAG), it has been estimated that 40% of fatty acids in TAG were originated from the acyl-editing pathway (Lu et al., 2009). It is difficult to determine to what extent the acyl-editing pathway participates in TAG biosynthesis in the Pennycress seed. Only metabolic flux analysis together with functional analysis of DGAT

and PDAT mutants will help to clarify this point. Nevertheless, the data suggest that both the Kennedy and the acyl-editing pathways contribute to TAG biosynthesis during Pennycress seed maturation and that their contribution may vary during the process. It could be hypothesized from the expression patterns of the genes from the different pathways and even from isoforms of the same pathway that different substrate pools are used by DGAT and PDAT enzymes for TAG biosynthesis in the seed. In that sense, it was recently reported that in Arabidopsis, *AtPDAT1* and *AtDGAT2* used a different larger bulk of PC-derived DAG pool than that used by *AtDGAT1* (Regmi et al., 2020). Our data showed a concerted regulation of *DGAT2* and *PDAT1* genes, suggesting that connexions between acyl-lipid fluxes might be occurring during maturation and seed-oil biosynthesis in the Pennycress seed.

DATA AVAILABILITY

All datasets generated for this study are included in the manuscript and in the supplementary data.

AUTHOR CONTRIBUTIONS

AC, ML grew the Pennycress plants, obtained the RNA and lipid extracts. AC and ML performed the qPCR determinations and prepared the samples for RNA-Seq. JME, CJ, MS and VC, performed all the experiments related with HPTLC and ESI-MS experiments and interpreted the MS data. MA, RP and MVL developed the idea of the project and coordinated the obtention of the data and the analysis. MA and MVL acquired the funding for the project. MA performed the analysis of the RNA-Seq data and prepared the manuscript. All authors discussed, commented and approved the final version of the manuscript.

FUNDING

This work was supported by the Spanish Ministry of Research and Innovation (MICINN) and FEDER (Grant PID2021-1265630-B100) and Gobierno de Aragón, (Grants A09-20R for Research Groups in Aragón and LMP194_21, Research in Strategic Lines).

ACKNOWLEDGEMENTS

We wish to thank Patricia Lorente and Marina de la Vega for their excellent technical assistance.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

- Altendorf K, Isbell T, Wyse DL, Anderson JA (2019) Significant variation of seed oil content, fatty acid profile, and seed weight in natural populations of field Pennycress (*Thlaspi arvense* L.). *Ind. Crop Prod.* 129: 261-268
- Arias, C.L., García Navarrete, L.T., Mukundi, E., Swanson, T., Yang F., Hernández, J., Grotewold E., Alonso A. P. (2023) metabolic and transcriptomic study of pennycress natural variation identifies targets for oil improvement. *Plant Biotechnol. J.*, <https://doi.org/10.1111/pbi.14101>.
- Aznar-Moreno, JA, Denolf, P, Van Audenhove, K, De Bodt, S, Engelen S, Fahy D, Wallis JG and Browse, J (2015) Type 1 diacylglycerol acyltransferases of *Brassica napus* preferentially incorporate oleic acid into triacylglycerol. *J. Exp. Bot.* 66: 6497-6506
- Bates PD, Browse J (2012) The significance of different diacylglycerol synthesis pathways on plant oil composition and bioengineering. *Front. Plant Sci.* 3: 147-150
- Bates PD, Stymne S, Ohlrogge J (2013) Biochemical pathways in seed oil synthesis. *Curr. Op. Plant Biol.* 16: 358-364
- Baud, S., Mendoza, M.S., To, A., Harscoët, E., Lepiniec, L., Dubreucq, B. (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. *Plant J.*, 50: 825-838
- Bligh EG, Dyer WJ (1959) A rapid method of total lipid extraction and purification. *Can. J. Biochem. Physiol.* 37: 911-917
- Bouaid, A., Martínez, M., Aracil, J., 2009. Production of biodiesel from bioethanol and Brassica carinata oil: oxidation stability study. *Biores. Tech.* 100, 2234-2239
- Brugal J, Shockey J, Lu CF, Dyer J, Larson T, Graham I, Browse J, (2008) Metabolic engineering of hydroxyl fatty acid production in plants: RcDGAT2 drives dramatic increases in ricinoleate levels in seed oil. *Plant Biotechnol. J.* 6: 819-831
- Cernac A, Benning C. (2004) WRINKLED1 encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in Arabidopsis. *Plant J.* 40(4):575-85. doi: 10.1111/j.1365-313X.2004.02235.x.

- Chopra, R., Johnson, E.B., Daniels, E., McGinn, M., Dorn, K.M., Esfahanian, M., Folstad, N., Amundson, K., Altendorf, K., Betts, K., Frels, K., Anderson, J.A., Wyse, D.L., Sedbrook, J.C., Marks, M.D. (2018) Translational Genomics Using Arabidopsis as a Model Enables the Characterization of Pennycress Genes Through Forward and Reverse Genetics. *Plant J.* 96: 1093-1105
- Claver A, Rey R, López MV, Picorel R, Alfonso M (2017) Identification of target genes and processes involved in erucic acid accumulation during seed development in the biodiesel feedstock Pennycress (*Thlaspi arvense* L.). *J. Plant Physiol.* 208: 7-16
- Claver A, de la Vega M, Rey-Giménez R, Luján MA, Picorel R, López MV, Alfonso M (2020) Functional analysis of β -ketoacyl-CoA synthase from biofuel feedstock *Thlaspi arvense* reveals differences in the triacylglycerol biosynthetic pathway among Brassicaceae. *Plant Mol. Biol.* 104: 283-296
- Cubins, J.A., Wells, M.S., Frels, K., Ott, M.A., Forcella, F., Johnsson, G.A., Walia, M.K., Becker, R.L., Gesch, R.W. (2019) Management of pennycress as a winter annual cash cover crop. A review. *Agron. Sustain. Dev.* 39, 1-11
- Demski K, Jeppson S, Lager I, Mistak A, Jasieniecka-Gazarkiewicz K, Waleron M, Szymne S, Banas A (2019) Isoforms of Acyl-CoA:diacylglycerol acyltransferase2 differ substantially in their specificities toward erucic acid. *Plant Physiol.* 181: 1468-1479
- Dorn KM, Fankhauser JD, Wyse DL, Marks MD (2013) De novo assembly of the pennycress (*Thlaspi arvense*) transcriptome provides tools for the development of a winter cropland biodiesel feedstock. *Plant J.* 75: 1028-1038
- Dorn KM, Fankhauser JD, Wyse DL, Marks MD (2015) A draft genome of field pennycress (*Thlaspi arvense*) provides tools for the domestication of a new winter biofuel crop. *DNA Res.* 22: 121-131
- Fan J, Shonnard DR, Kalnes TN, Johnsen PB, Rao S (2013) A life cycle assessment of pennycress (*Thlaspi arvense* L.) –derived jet fuel and diesel. *Biomass Bioener.* 55: 87-100
- Farese RV Jr, Walther TC (2009) Lipid droplets finally get a little R-E-S-P-E-C-T. *Cell* 139: 855-860

- Furmanek T, Demski K, Banas W, Haslam R, Napier J, Stymne S, Banas A (2014) The utilization of the Acyl-CoA and the involvement PDAT and DGAT in the biosynthesis of erucic acid-rich triacylglycerols in *Crambe* seed oil. *Lipids* 49: 327-333
- Fusconi, F., Simeoni, F., Francini, P., Cominelli, E., Conti, L., Riboni, M., Simonini L., Martin CR, Tonelli C, Galbiati, M. (2013) The *Arabidopsis thaliana* MYB60 promoter provide a tool for the spatio-temporal control of gene expression in stomatal guard cells. *J. Exp. Bot.* 64: 3361-3371.
- Gasic, K., Hernández, A., Korban, S., 2004. RNA extraction from different apple tissues rich in polyphenols and polysaccharides for cDNA library construction. *Plant Mol. Biol. Rep* 22, 437-438.
- García Navarrete, T., Arias, C., Mukundi, E., Alonso A.P., Grotewold E., (2022) natural variation and improved genome annotation of the emerging biofuel crop field pennycress (*Thlaspi arvense*). *G3* (Bethesda), 12, jkac084
- Ghanevati M, Jaworski JG (2001) Active-site residues of a plant membrane-bound fatty acid elongase beta-ketoacyl- CoA- synthase, FAE1 KCS. *Biochim. Biophys. Acta* 1530: 77-85
- Guan R, Lager I, Li X, Stymne S, Zhu L-H (2014) Bottlenecks in erucic acid accumulation in genetically engineered ultrahigh erucic acid *Crambe abyssinica*. *Plant Biotech. J.* 12: 193-203.
- Han, X. (2016) Comprehensive mass spectrometry of lipids. In: *Lipidomics: Comprehensive Mass Spectrometry of Lipids*, ch. 9, p.222. Wiley on line library, New Jersey DOI: [10.1002/9781119085263.ch9](https://doi.org/10.1002/9781119085263.ch9)
- Hsu F-F., Turk, J. (2010) Electrospray ionization multiple-stage linear ion-trap mass spectrometry for structural elucidation of triacylglycerols: Assignment of fatty acyl groups on the glycerol backbone and location of double bonds. *J. Am. Soc. Mass Spectrom.* 21: 657–669 <https://pubs.acs.org/doi/10.1016/j.jasms.2010.01.007>
- Jarne, C., Savirón, M., Lapieza, M. P., Membrado, L., Orduna, J., Galbán, J., Garriga, R., Morlock, G. E., Cebolla, V. L. (2018) High-Performance Thin-Layer Chromatography Coupled with Electrospray Ionization Tandem Mass Spectrometry for Identifying Neutral Lipids and Sphingolipids in Complex Samples. *J. AOAC Int.* 101: 1993–2000.

- Jarne, C., Membrado, L., Savirón, M., Vela, J., Orduna, J., Garriga, R., Galbán, J., Cebolla, V.L. (2021) Globotriaosylceramide-related biomarkers of fabry disease identified in plasma by high-performance thin-layer chromatography-densitometry-mass spectrometry *J. Chromatogr. A*. 1638: 461895.
- Johnston C, García Navarrete LT, Ortiz E, Romsdahl TB, Guzha A, Chapman KD, Grotewold E, Alonso AP (2022) Effective mechanisms for improving seed oil production in pennycress (*Thlaspi arvense* L.) highlighted by integration of comparative metabolomics and transcriptomics. *Frontiers in Plant Sci.* 13: 943585 doi: 10.3389/fpls.2022.943585
- Katavic V, Friesen W, Barton DL, Gossen KK, Giblin EM, Luciw T, An J, Zou J-T, Mackenzie SL, Keller WA, Males D, Taylor DC (2001) Improving erucic acid content in rapeseed through biotechnology: What can the *Arabidopsis FAE1* and the yeast *SLC1-1* genes contribute. *Crop Sci.* 41: 739-747
- Kim HU, Huang AH (2004) Plastid lysophosphatidyl acyltransferase is essential for embryo development in *Arabidopsis*. *Plant Physiol.* 134: 1206-1216
- Kim HU, Li, Y, Huang AH (2005) Ubiquitous and endoplasmic reticulum lysophosphatidyl acyltransferase, LPAT2, is essential for female, but not male gametophyte development in *Arabidopsis*. *The Plant Cell* 17: 1073-1089
- Lager, I, Lindberg Yilmaz J, Zhou XR, Jasieniecka K, Kazachkov M, wang P, Zou J, Weselake R, Smith MA, Bayon S, Dyer JM, Shockey JM, Heinz E, Green A, banas A, Stymne S (2013) Plant Acyl-CoA:lysophosphatidylcholine acyltransferases (LPCATs) have different specificities in their forward and reverse reactions. *J. Biol. Chem.* 288: 36902-36914
- Laurent P, Huang AH. Organ- and development-specific acyl coenzyme A lysophosphatidate acyltransferases in palm and meadowfoam. (1992) *Plant Physiol.* 99(4):1711-5. doi: 10.1104/pp.99.4.1711.
- Li, Y.H., beisson, F., Pollard, M., Ohlrogge, J. (2006) Oil content of *Arabidopsis* seeds: the influence of seed anatomy: light and plant to plant variation. *Phytochem.* 67: 904-915
- Li R, Yu K, Hildebrand DF (2010) *DGAT1*, *DGAT2* and *PDAT* expression in seeds and

- other tissues of epoxy and hidroxy fatty acid accumulating plants. *Lipids* 45: 145-157
- Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, Baud S, Bird D, Debono A, Durrett TP, *et al.* (2013) Acyl-lipid metabolism. *The Arabidopsis book* 11: 161
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time PCR and the 2(-delta Delta C(T)) method. *Methods* 25: 402-408
- López MV, de la Vega M, Gracia R, Claver A, Alfonso M (2021) Agronomic potential of two European pennycress accessions as a Winter crop under Mediterranean conditions. *Ind. Crop Prod.* 159: 113107 doi.org/10.1016/j.indcrop.2020.113107
- Lu, C., Xin, Z., Ren, Z., Miquel, M., 2009. An enzyme regulating triacylglycerol composition is encoded by the ROD1 gene of Arabidopsis. *Proc. Natl. Acad. Sci. USA* 106, 18837-18842
- Marmon, S., Sturvenant D., Herrfurth, C., Chapman, K., Stymne, S., Feussner, I. (2017) Two acyltransferases contribute differently to linolenic acid levels in seed oil. *Plant Physiol.* 173: 2081-2095
- McGinn M, Phippen WB, Chopra R, Bansal S, Jarvis BA, Phippen ME, Dorm KM, Esfahanian M, Nazarens TJ, Cahoon EB, Durrett TP, Marks MD, Sedbrook JC (2018) Molecular tools enabling pennycress (*Thlaspi arvense*) as a model plant and oilseed cash cover crop. *Plant Biotechnol. J.* 1-13 doi: 10.1111/pbi.13014
- Mortazavi A, Williams B A, McCue K, *et al.* (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* 5: 621-628
- Moser BR, Knothe G, Vaughn SF, Isbell TA (2009) Production and evaluation of biodiesel from field Pennycress (*Thlaspi arvense* L.) oil. *Energy Fuels* 23: 4149-4155
- Moser RB (2012) Biodiesel from alternative oilseed feedstocks: camelina and field pennycress. *Biofuels* 3: 193-209
- Nowak, K., Gag M.D. (2016) Stress related function of bHLH109 in somatic embryo induction in Arabidopsis. *J. Plant Physiol.* 193: 119-126
- Nunn, A., Rodríguez-Arévalo, I., Tandukar, Z., Frels, K., Contreras-Garrido, A.,

- Carbonell-Bejerano, P., Zhang, P., Ramos Cruz, D., Jandrasits, K., Lanz, C., Brusa, A., Mirouze, M., Dorn, K., Galbraith, D. W., Jarvis, B. A., Sedbrook, J. C., Wyse, D. L., Otto, C., Langenberger, D., Stadler, P. F., Weigel, D., Marks, M. D., Anderson, J. A., Becker, C. and Chopra, R. (2022) Chromosome-level *Thlaspi arvense* genome provides new tools for translational research and for a newly domesticated cash cover crop of the cooler climates. *Plant Biotechnol. J.*, <https://doi.org/10.1111/pbi.13775>
- Ohlroge J, Browse J (1995) Lipid biosynthesis. *Plant Cell* 7: 957-970
- Ramaley, L., Herrera, L.C. & Melanson, J.E. (2015) Quantitative Analysis of TAG in Oils Using Lithium Cationization and Direct-Infusion ESI Tandem Mass Spectrometry. *J Am Oil Chem Soc* **92**, 323–334. <https://doi.org/10.1007/s11746-015-2604-9>
- Regmi, A., Shockey, J., Kotapati, H.-K., bates P.D. (2020) Oil-producing metabolons containing DGAT1 use separate substrate pools from those containing DGAT2 or PDAT. *Plant Physiol.* 184: 720-737
- Romsdahl, T.B., Cocuron J-C., Pearson, M. J., Alonso, A.P., Chapman K.D. (2022) A lipidomics platform to analyze the fatty acid compositions of non-polar and polar lipid molecular species from plant tissues: Examples from developing seeds and seedlings of pennycress (*Thlaspi arvense*). *Front. Plant Sci.* 13: <https://www.frontiersin.org/articles/10.3389/fpls.2022.1038161/full>
- Routaboul JM, Benning C, Bechtold N, caboche M, lepiniec L (1999) The TAG1 locus of Arabidopsis encodes for a diacylglycerol acyltransferase. *Plant Physiol. Biochem.* 37: 831-840
- Sancho-Albero, M., Jarne, C., Savirón, M., Martín-Duque, P., Membrado, L., Cebolla, V.L., Santamaría, J.. (2022) High-Performance Thin-Layer Chromatography-Densitometry-Tandem ESI-MS to Evaluate Phospholipid Content in Exosomes of Cancer Cells. *Int. J. Mol. Sci.* 23: 1150-1164.
- Sedbrook, J.C., Phippen, W.B., Marks, M.D., (2014) New approaches to facilitate rapid domestication of a wild plant to an oilseed crop: example pennycress (*Thlaspi arvense* L.) *Plant Sci.* 227: 122-132
- Shockey, J.M., Gidda, S.K., Chapital, D.C., Kuan, J.-C., Dhanoa, P.K., Bland, J.M., Rothstein, S.J., Mullen, R.T., Dyer, J.M. (2006). Tung tree DGAT1 and DGAT2

- have nonredundant function in triacylglycerol biosynthesis and are localized to different subdomains of the endoplasmic reticulum. *The Plant Cell* 18, 2294-2313.
- Skubacz, A., Daskowska-Golec, A, Szarejko, I. (2016) The role and regulation of ABI5 (ABA-insensitive 5) in plant developmental, stress responses and phytohormone crosstalk. *Frontiers I Plant Sci.* 7, 1884.
- Stuart Grossert, J., Cubero Herrera, L., Ramaley, L., Melanson, J.E. (2014) Studying the Chemistry of Cationized Triacylglycerols Using Electrospray Ionization Mass Spectrometry and Density Functional Theory Computations. *J. Am. Soc. Mass Spectrom.* 25: 1421-1440. <https://pubs.acs.org/doi/10.1007/s13361-014-0917-9>
- Sun X, Pang H, Guo H, Yan Q, Hang Y (2013) Evolutionary pattern of the *FAEI* gene in *Brassicaceae* and its correlation with the erucic acid trait. *PLoS One* 8: 1-12
- Tsogtbaatar E, Cocuron JC, Corchado Sonera M, Alonso AP (2015) Metabolite fingerprinting of pennycress (*Thlaspi arvense* L.) embryos to assess active pathways during oil synthesis. *J. Exp. Bot.* 66: 4267-4277
- Zhang M, Fan J, Taylor DC, Ohlrogge JB (2009) DGAT1 and PDAT1 acyltransferases have overlapping functions in *Arabidopsis* triacylglycerol biosynthesis and are essential for normal pollen and seed development. *Plant Cell* 21: 3885-3901
- Wang L, Shen W, Kazachkov M, Chen G, Chen Q, Carlsson AS, Stymne S, Weselake RJ, Zou J (2012) Metabolic interactions between the Lands Cycle and the Kennedy pathway of glycerolipid synthesis in *Arabidopsis* developing seeds. *Plant Cell* 24: 4652-4669
- Weselake RJ, Taylor DC, Rahman MH, Shah S, Laroche A, McVetty PBE, Harwood JL (2009) Increasing the flow of carbon into seed oil. *Biotechnol. Adv.* 27: 866-878
- Yu G, Wang LG, Han Y, He QY (2012) ClusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* 16(5):284-7. doi: 10.1089/omi.2011.011.

Table 1. Percentages of TAG peak in UV (190 nm)-HPTLC chromatograms of oil samples, and intra- and inter-plate HPTLC repeatability results for the separated TAG peaks (expressed in Area counts).

Maturation stage/batch	TAG%	Intraplate Repeatability				Interplate Repeatability			
		Mean	SD	RSD%	CV	Mean	SD	RSD%	CV
G/1	91.1%	9442.71	389.06	4.12	359.84	9377.38	297.94	3.18	180.06
G/2	88.1%	12284.57	370.23	3.01	342.42	12600.08	500.08	3.97	302.22
GY/1	92.7%	15556.71	847.40	5.45	783.74	15968.54	774.08	4.85	467.81
GY/2	91.9%	14830.29	600.25	4.05	555.15	17465.08	510.93	2.93	308.78
YG/1	95.3%	14830.29	600.25	4.05	555.15	15787.08	1241.20	7.86	750.12
YG/2	95.0%	16867.14	385.73	2.29	356.75	17140.92	642.26	3.75	388.15
Y/1	95.8%	15531.14	403.41	2.60	373.10	16668.62	1408.25	8.45	851.07
Y/2	95.5%	10920.29	284.65	2.61	263.26	11932.92	1367.02	11.46	826.15
M/1	94.0%	15779.43	402.49	2.55	372.25	17312.85	2184.14	12.62	1319.98

Table 2 Identified molecular species by HPTLC-ESI⁺-MS and their fragmentation patterns (MS²) in separated TAG peaks of oil samples.

[TAG+Na] ⁺ <i>m/z</i>	[M+Na-RCOOH] ⁺		Molecular Formula	TAG
	<i>m/z</i>	(Cx;y)FA		
907.8	597.5 627.5 651.5	C20:1 C18:2 C16:0	C ₅₇ H ₁₀₄ NaO ₆	16:0/18:2/20:1
935.9	587.5 655.6 679.6	C22:1 C18:2 C16:0	C ₅₉ H ₁₀₈ NaO ₆	16:0/18:2/22:1
959.9	621.5 679.6	C22:1 C18:2	C ₆₁ H ₁₀₈ NaO ₆	18:2/18:2/22:1
989.9	651.5 679.6 709.6	C22:1 C20:1 C18:2 (<i>sn</i> -2)	C ₆₃ H ₁₁₄ NaO ₆	20:1/18:2/22:1
1017.9	679.6 737.7	C22:1 C18:2 (<i>sn</i> -2)	C ₆₅ H ₁₁₈ NaO ₆	22:1/18:2/22:1
1046.0	679.6 707.6 765.7	C24:1 C22:1 C18:2 (<i>sn</i> -2)	C ₆₇ H ₁₂₂ NaO ₆	22:1/18:2/24:1

FIGURE LEGENDS

Fig. 1. Total lipid and fatty acid composition content during Pennycress seed maturation. (a) Photograph showing each of the five maturation stages (GREEN, GREENYELLOW, YELLOWGREEN, YELLOW and MATURE) used in this study. Seeds were extracted from the pod, pooled for each stage, and used in the analysis (b); Total lipid per seed on a fresh weight basis; (c) Fatty acid composition from total lipids extracted from the different stages of seed maturation; G, GREEN seed, GY, GREENYELLOW seed, YG, YELLOWGREEN seed, Y, YELLOW seed, M, MATURE seed. Data were obtained from three independent biological replicates. Data represent means $\pm$ SD.

Fig. 2. Distribution of DEG genes identified in the RNA-Seq analysis. Total (grey bars), up-regulated (red bars) and down-regulated (green bars) genes in each seed maturation stage. DEGs were identified by a $\log_2\text{ratio} \geq 1$ and a $\text{padj} \leq 0.05$.

Fig. 3. GO categorization analysis of DEG genes identified in the GREENYELLOW vs GREEN (upper panels) and YELLOWGREEN vs GREEN (lower panels) pairwise comparisons. Downregulated genes are shown on the left and upregulated ones on the right. Orange, green and blue colors indicate biological process (BP), cellular component (CC) and molecular functions (MF) categories respectively. Number of genes in each category is indicated in the bars.

Fig. 4. GO categorization analysis of DEG genes identified in the YELLOW vs GREEN (upper panels) and MATURE vs GREEN (lower panels) pairwise comparisons. Downregulated genes are shown on the left and upregulated ones on the right. Orange, green and blue colors indicate biological process (BP), cellular component (CC) and molecular functions (MF) categories respectively. Number of genes in each category is indicated in the bars.

Fig. 5. HCL clustering of genes obtained for each seed maturation stage in the RNA-Seq analysis. ClusterProfiler was used for the analysis. The black line in each cluster represents the average estimated variation for each cluster. The number of genes in each cluster is indicated in each figure.

Fig. 6. Differential expression of genes related to carbon assimilation and fatty acid biosynthesis for each of the five seed maturation stages used in this work. Values represent average $\log_2(\text{fpkm} + 1)$ values from each of the biological repeats and were used

to generate heatmaps from http://bar.toronto.ca/ntools/cgi-bin/ntools_heatmapper_plus.cgi. The *Thlaspi arvense* annotated genome (Nunn et al., 2022) was used for identification of the gene ID. FATA and FATB, fatty acid acyl-ACP thioesterases; KASC, ketoacyl-ACP synthases; ACP, acyl-carrier proteins; LACS, long-chain acyl-CoA synthetases; ADS, Acyl desaturase; FAD, fatty acid desaturase, GPAT, glycerol 3-phosphate acyl-transferase; LTP, lipid transfer protein; G3PDH, Glycerol 3-phosphate dehydrogenase; PPC, phosphoenol pyruvate carboxylase; PyrDH, pyruvate dehydrogenase; PKP, phosphoenol pyruvate kinase; PDC, phosphoenol pyruvate decarboxylase.

Fig. 7 Expression profiling of individual genes and isoforms involved in VLCFA and TAG biosynthesis during seed maturation by RNA-Seq (light grey bars) and qPCR (dark grey bars). For RNA-Seq data, expression levels represented by FPKM values were normalized to GREEN stage. The genes analysed (*FAE1*, *FAD2*, *DGAT1*, *DGAT2*, *LPAT1*, *LPAT2*, *PDAT1*, *PDAT2*, *PDCT*, *LPCAT* and *WRI1*) are indicated in each figure. For qPCR analysis, data were obtained from three independent pools of seeds from five plants of each line. Data represent means of at least three biological replicates.

Fig. 8. HPTLC-ESI⁺-MS profiles of TAG fraction separated by HPTLC and extracted from the plate, using the interface, for each of the different Pennycress seed maturation stage stage: a) Green, b) Green-Yellow, c) Yellow-Green, d) Yellow, and e) Mature

Fig. 9. Normalized HPTLC-ESI⁺-MS ion intensities of TAG species found in the corresponding TAG peaks of oil samples at the different maturation stages: a) Green, b) Green-Yellow, c) Yellow-Green, d) Yellow, and e) Mature.

Fig. 10. HPTLC-ESI⁺-MS/MS of most abundant ion at m/z 1018.0.

Supplementary material

Fig. S1. Differential expression of genes encoding transcription factors. (a) Distribution of DEGs among different transcription factor families. (b) Differential expression of genes encoding transcription factors for each of the five seed maturation stages used in this work. Values represent average log₂(fpkm + 1) values from each of the biological repeats and were used to generate heatmaps from <http://bar.toronto.ca/ntools/cgi->

[bin/ntools_heatmapper_plus.cgi](#). The *Thlaspi arvense* annotated genome (Nunn et al., 2022) was used for identification of the gene ID.

Fig. S2. Expression profiling of individual genes and isoforms involved in oil bodies formation during seed maturation by RNA-Seq (light grey bars) and qPCR (dark grey bars). For RNA-Seq data, expression levels represented by FPKM values were normalized to GREEN stage. The genes analysed, *OLE1*, *OLE2*, and *OBA1a*, are indicated in each figure. For qPCR analysis, data were obtained from three independent pools of seeds from five plants of each line. Data represent means of at least three biological replicates.

Fig. S3. HPTLC chromatograms of standards at UV 190 nm: pic1: 1-Oleoyl-rac-glycerol, pic 2: 1,2-dioleoyl-rac-glycerol, pic 3: linoleic acid, pic 4: oleic acid, pic 5: erucic acid, pic 6: glyceryl trioleate, pic 7: methyl oleate, pic 8: cholesteryl stearate, and pic 9: cholesteryl oleate.

Fig. S4. HPTLC chromatograms corresponding to samples at the different maturation stages detected at UV 190 nm: a) GREEN, b) GREENYELLOW, c) YELLOWGREEN; d) YELLOW, and e) MATURE.

Fig. S5. HPTLC-ESI⁺-MS/MS of molecular species found in separated and extracted TAG peak (907.8, 935.9, 959.9, 989.9 and 1046.0)

Table S1 primers used for qPCR analysis in this study.

Table S2. Total reads, clean reads and quality parameters of the RNA-Seq data.

Supporting information 1: RNA-Seq data set containing all the results for each pairwise comparison used in this study.

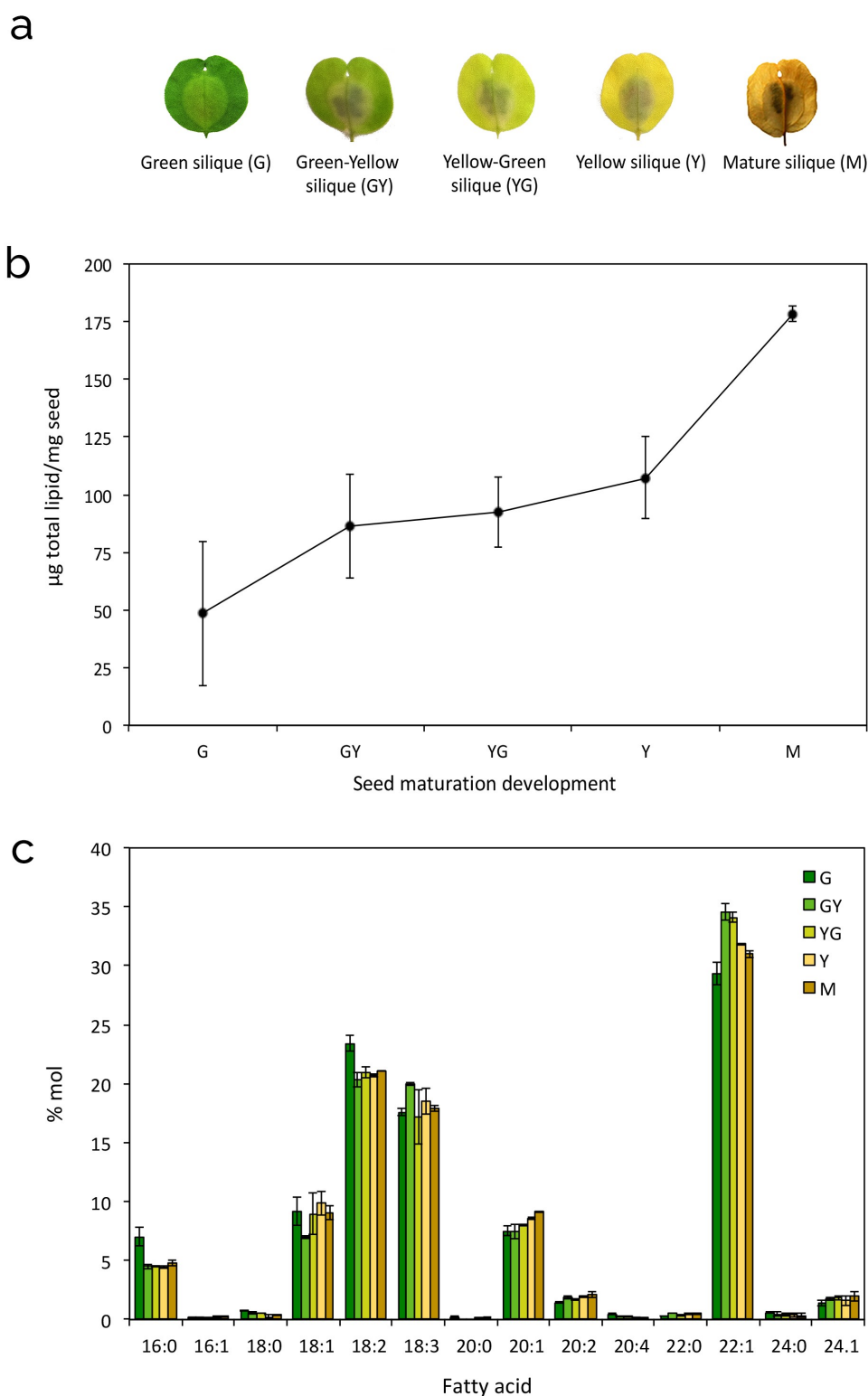


Fig.1. Total lipid and fatty acid composition content during Pennycress seed maturation. (a) photograph showing each of the five maturation stages used in this study, (b) Total lipid per seed on a fresh weight basis, (c) Fatty acid composition from total lipids extracted from the different stages of seed maturation; G, green seed, GY, green-yellow seed, YG, yellow-green seed, Y, yellow seed, M, mature seed. The photograph shows the different maturation stages of the Pennycress seed analysed in this study. Seeds were pooled for each stage and data were obtained from three independent biological replicates. Data represent means $\pm$ SD.

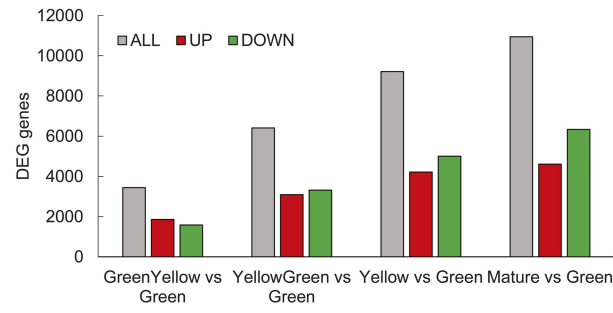
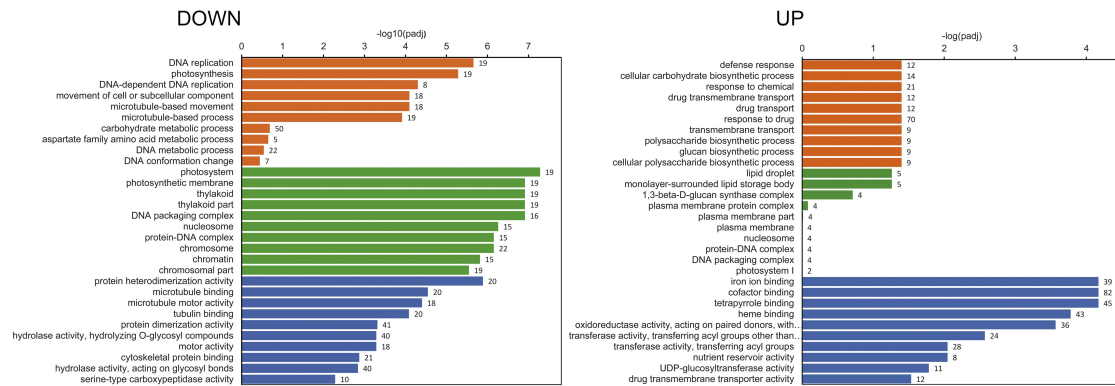


Fig.2. Distribution of DEG genes identified in the RNA-Seq analysis. (a) Total (grey bars), up-regulated (red bars) and down-regulated (green-bars) genes in each seed maturation stage. DEGs were identified by a $\log_2\text{ratio} \geq 1$ and a $\text{padj} \leq 0.05$.

GreenYellow vs Green



YellowGreen vs Green

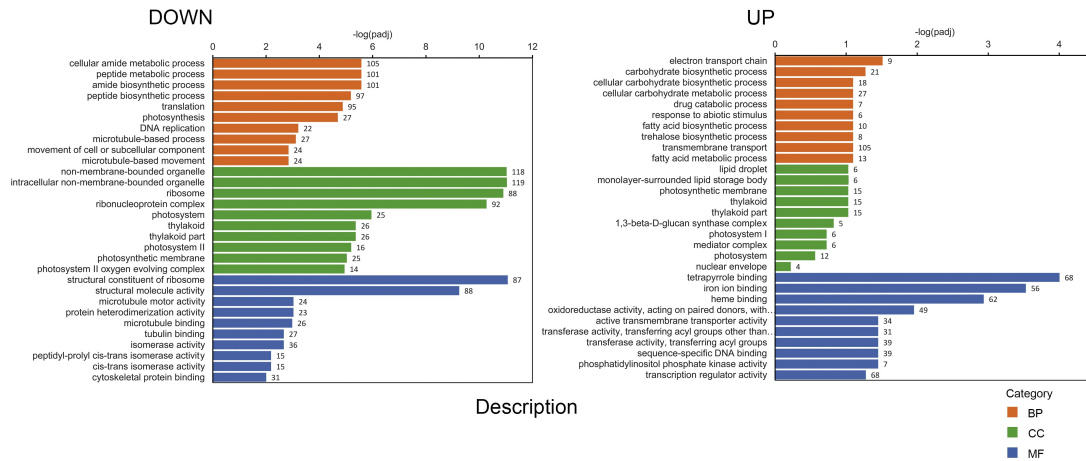
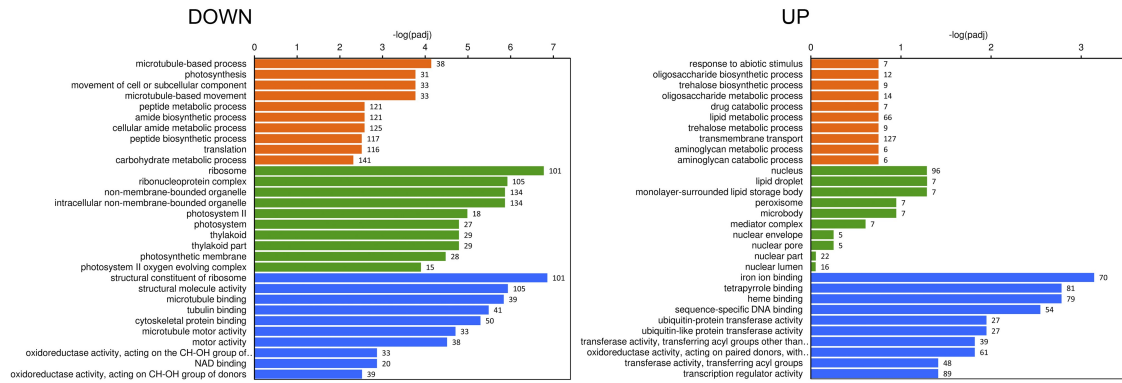


Fig.3. GO categorization analysis of DEG genes for each comparison at the GRRENYELLOW vs GREEN (upper panels) and YELLOWGREEN vs GREEN (lower panels) comparisons. Downregulated genes are shown on the left and upregulated ones on the right. Orange, green and blue colors indicate biological process, cellular component and molecular functions categories respectively. Number of genes in each category is indicated in the bars.

Yellow vs Green



Mature vs Green

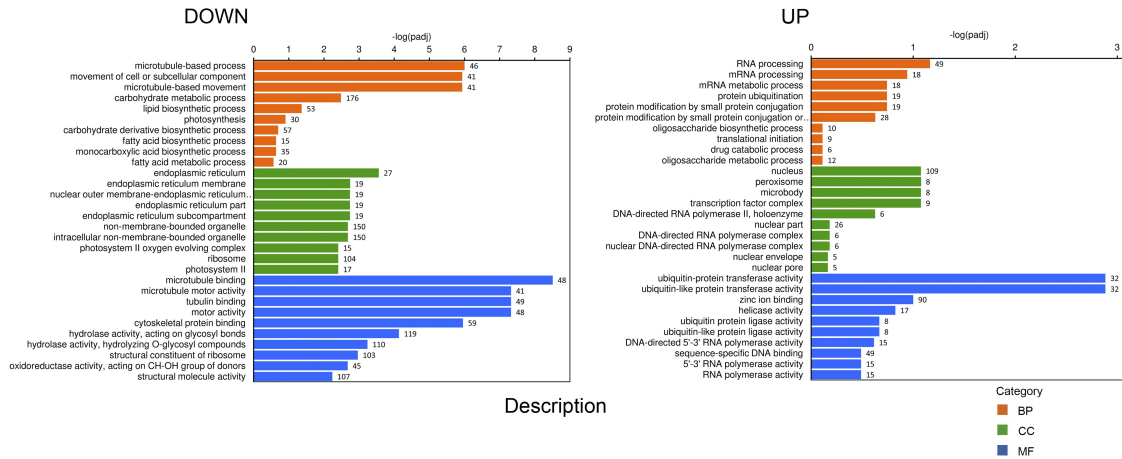


Fig. 4. GO categorization analysis of DEG genes for each comparison at the YELLOW vs GREEN (upper panels) and MATURE vs GREEN (lower panels) comparisons. Downregulated genes are shown on the left and upregulated ones on the right. Orange, green and blue colors indicate biological process, cellular component and molecular functions categories respectively. Number of genes in each category is indicated in the bars.

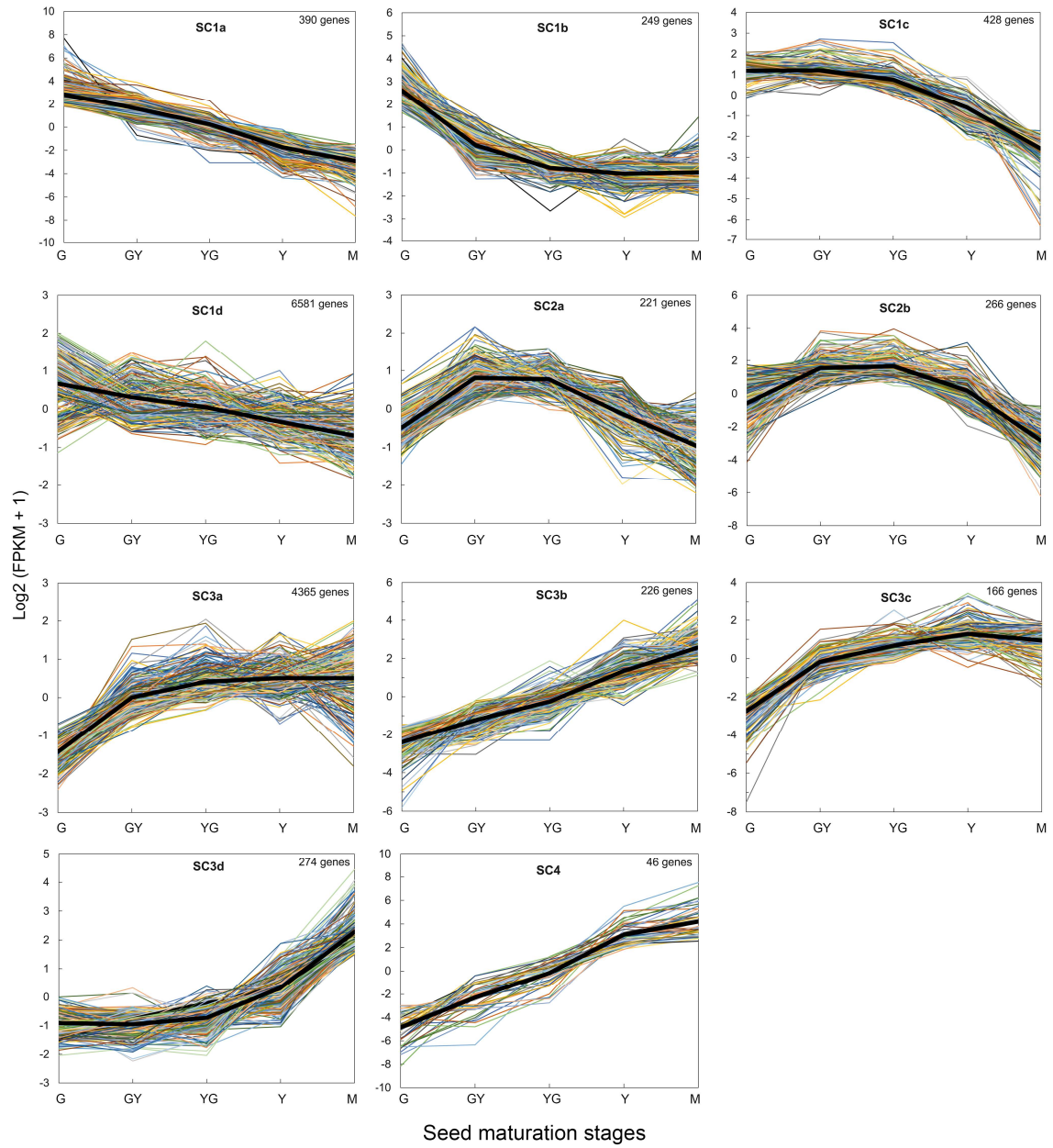


Fig. 5. HCL clustering of genes obtained for each seed maturation stage in theRNA-Seq analysis. The black line in each cluster represents the average estimated variation for each cluster. The number of genes in each cluster is indicated in each figure.

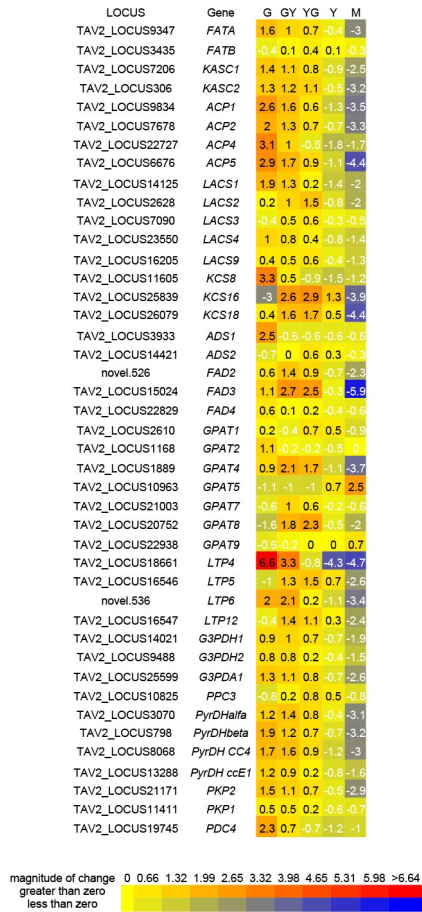


Fig. 6. Differential expression of genes involved in carbon precursors and fatty acid biosynthesis for each seed maturation stage, Values represent the $\log_2(\text{fpkm} + 1)$ values. Heatmaps were constructed using the online tool: <http://www.unitoronto>

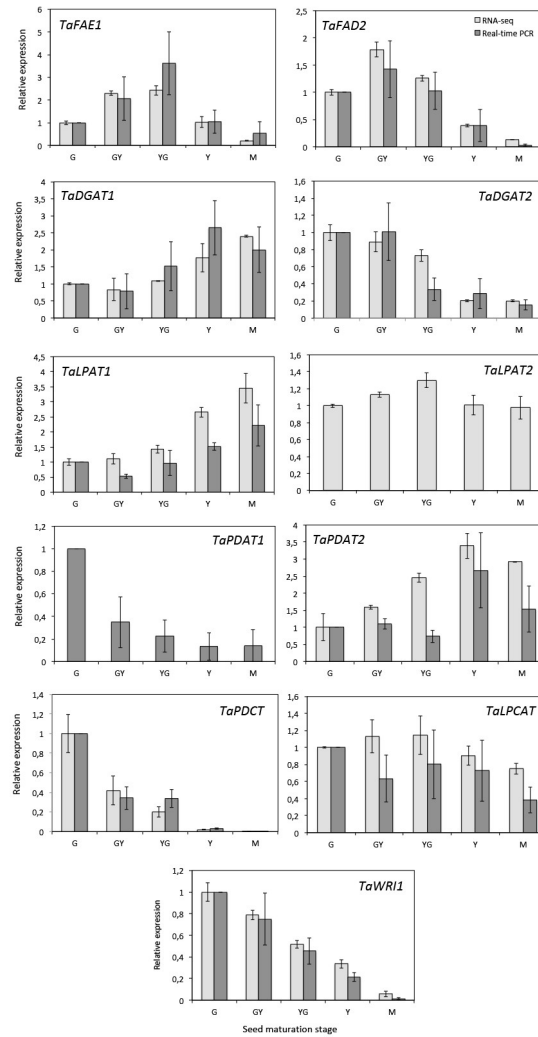


Fig. 7 RNA-Seq + qPCR expression analysis of selected genes involved in TAG biosynthesis and regulation. Both the RNA-seq data (normalized fpkm values) and qPCR expression data are represented for each gene. The genes analysed (*FAE1*, *FAD2*, *DGAT1*, *DGAT2*, *LPAT1*, *LPAT2*, *PDAT1*, *PDAT2*, *PDCT*, *LPCAT* and *WRI1*) are indicated in each figure. For qPCR analysis, data were obtained from three independent pools of seeds from five plants of each line. Data represent means of at least three biological replicates.

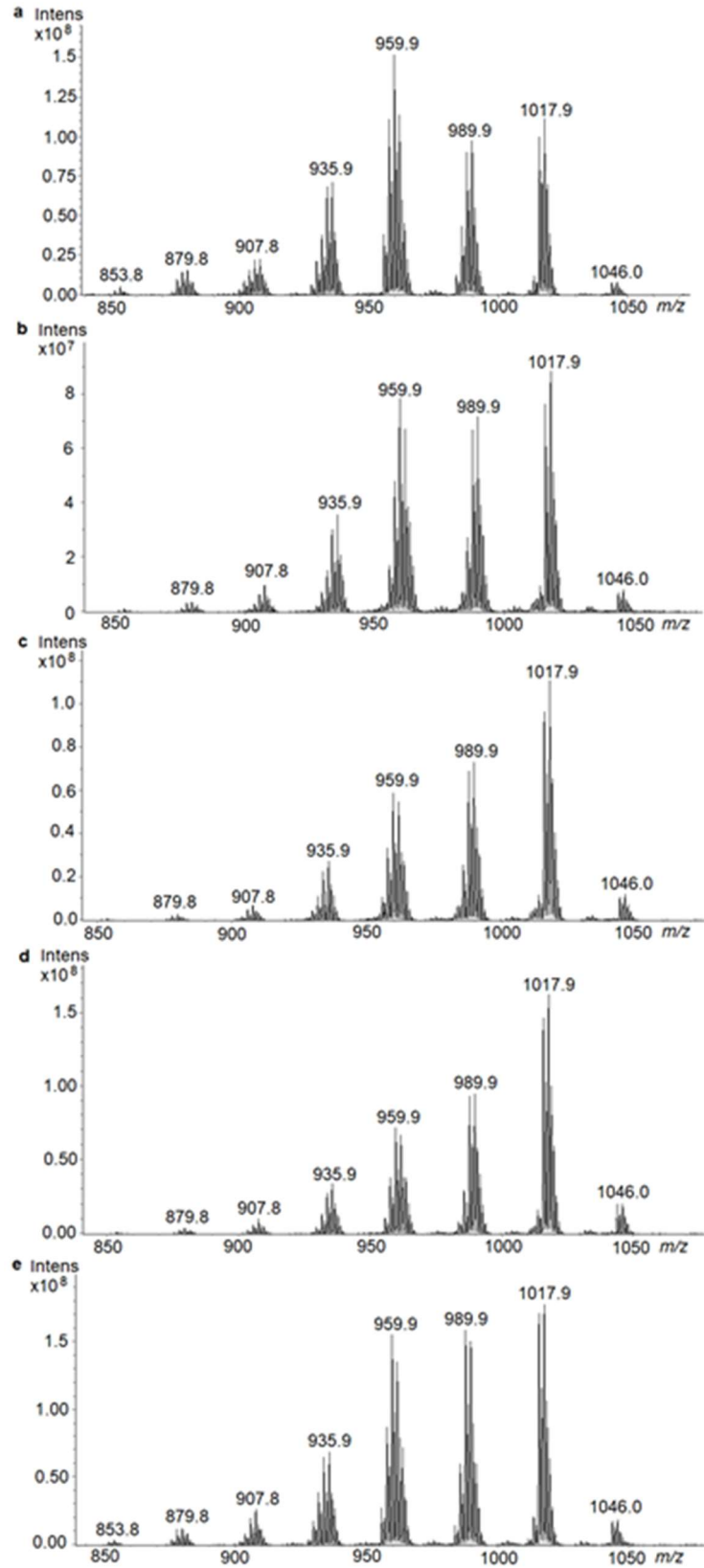


Fig 8. HPTLC-ESI⁺-MS profiles of TAG fraction separated by HPTLC and extracted from the plate, using the interface, for each of the different Pennycress seed maturation stage stage: a) Green, b) Green-Yellow, c) Yellow-Green, d) Yellow, and e) Mature

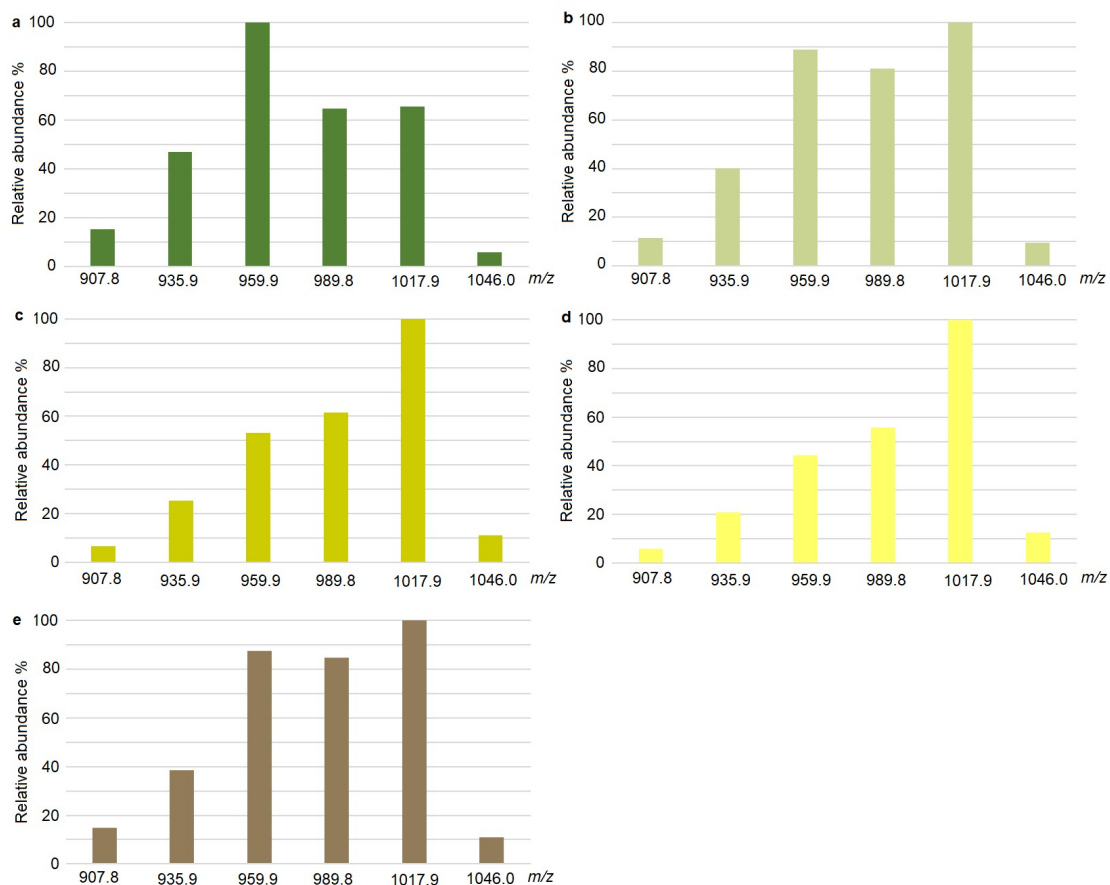


Fig 9. Normalized HPTLC-ESI⁺-MS ion intensities of TAG species found in the corresponding TAG peaks of oil samples at the different maturation stages: a) Green, b) Green-Yellow, c) Yellow-Green, d) Yellow, and e) Mature

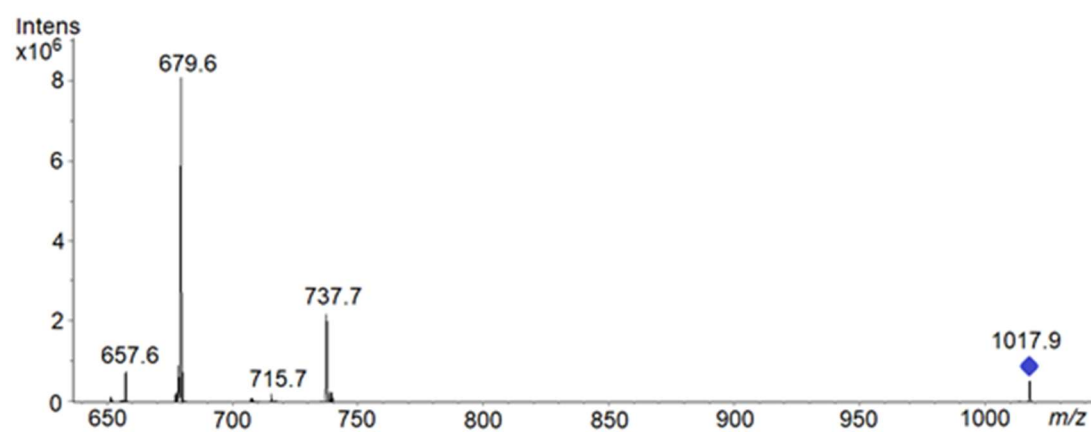


Fig 10. HPTLC-ESI⁺-MS/MS of most abundant ion at m/z 1018.0.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Fig.S1.docx](#)
- [Fig.S2.docx](#)
- [Fig.S3.docx](#)
- [Fig.S4.docx](#)
- [Fig.S5.docx](#)
- [tableS1primers.docx](#)
- [TableS2.QCdatastatistics.xlsx](#)
- [SeedmaturationstagesRNAseqdeg.xlsx](#)