Multi-omic analyses unveil temporal and spatial distribution of specialized metabolites in seeds of *Camelina sativa*

Léa Barreda[1±], Céline Brosse[1±], Stéphanie Boutet[1], Nicolas Klewko[1], Delphine De Vos[1], Tracy Francois[1], Boris Collet[1], Damaris Grain[1], Céline Boulard[1], Benoît Bernay[2], François Perreau[1], Loïc Lepiniec[1], Loïc Rajjou[1], Massimiliano Corso[1*]

[1] Université Paris-Saclay, INRAE, AgroParisTech, Institut Jean-Pierre Bourgin for Plant Sciences (IJPB), 78000, Versailles, France

[2] Plateforme Proteogen, SFR ICORE 4206, Université de Caen Basse-Normandie, Esplanade de la paix, 14032 Caen cedex, France

[±] Equal contribution

[*] Correspondence to: Massimiliano Corso massimiliano.corso@inrae.fr

E-mails:
LB lea.barreda@inrae.fr
CB celine.brosse@inrae.fr
SB stephanie.boutet@inrae.fr
NK nicolas.klewko57@gmail.com
DDV delphine.de-vos@inrae.fr
TF tracy.francois@inrae.fr
BC boris.collet@inrae.fr
DG damaris.grain@inrae.fr
CB celine.boulard@inrae.fr
DC delphine.charif@inrae.fr
BB benoit.bernay@unicaen.fr
FP francois.perreau@inrae.fr
LL loic.lepiniec@inrae.fr
LR loic.rajjou@inrae.fr
MC massimiliano.corso@inrae.fr

Running head: Specialized metabolites in camelina seed tissues

One Sentence Summary: Metabolomic, proteomic and transcriptomic analyses unveil specialized metabolites signature and distribution in Camelina seed coat-endosperm and embryo

Keywords: *Camelina sativa*, metabolic signature, glucosinolates, seed development, seed germination, untargeted metabolomics, transcriptomics, proteomics

Abstract

Seeds of Brassicaceae species produce a large diversity of specialized metabolites (SMs) that strongly influence their quality, with beneficial or toxic effects on human and animal nutrition, and provide resistance to biotic or abiotic stresses. While the distribution of these compounds has been described in leaves and roots tissues, very limited information is available about their spatio-temporal accumulation in seeds of model or crop plants.

Camelina sativa (camelina) is an oilseed Brassicaceae cultivated for human and animal nutrition, and for industrial uses. While we previously explored in detail SM diversity and plasticity, no information is available about SM distribution and expression of SM-related proteins and genes in camelina seeds.

In this study we used untargeted metabolomics (LC-MS/MS), proteomics (DIA) and transcriptomics (RNA-Seq) to analyse synthesis, transport, modifications and degradations of SMs that are accumulated in the different seed tissues (i.e. seed coat, endosperm, and embryo) at 6 developmental and 2 germination stages. Our results showed specific patterns for many SMs, and related proteins or genes, during seed coat and embryo development. We also showed that, differently from *Arabidopsis thaliana* seeds, the defence and antinutritional glucosinolates compounds were accumulated in both the seed coat and endosperm, and the corresponding degradation products isothiocyanates were present at high level in the embryos of dry seeds, in *C. sativa*.

Characterizing the spatial dynamics of seed SMs will contribute to the development of crops with an optimized distribution of beneficial and toxic metabolites for seeds quality and animal nutrition.

Introduction

Plants produce a myriad of specialized metabolites (SMs) with tremendous interest for agriculture, nutrition and health (Arimura and Maffei, 2016). The large diversity of chemical structures and, hence, of functions observed among specialized metabolic classes arises from different functional groups (decorations or modifications) commonly added to SM core structures (Barreda et al., 2024a). SM structures can indeed be modified by the addition or removal of functional groups, hydroxylation, methylation, glycosylation and acylation being the most frequent modification reactions.

Various SMs are widely accumulated in seeds of Brassicaceae model and crop species, which thus constitute an interesting model to study SMs diversity and functions (Corso et al., 2020; Corso et al., 2021; Baud et al., 2023). Major metabolic classes that are accumulated in seeds of Brassicaceae species, including the model species *Arabidopsis thaliana* and the crops *Camelina sativa* and *Brassica napus* (rapeseed), are the nitrogen-containing compounds glucosinolates, and many phenylpropanoids, including flavonols, proanthocyanidins (PAs, or condensed tannins), cinnamic acids and monolignols (Halkier and Gershenzon, 2006; Quéro et al., 2016; Barreda et al., 2024a).

While some SM classes can be directly synthetized in seeds, others are synthetized in different organs or tissues of the mother plant and then transported to seeds. In *A. thaliana*, the flavonoids proanthocyanidins (PAs) and flavonols are synthetized in seeds (Lepiniec et al., 2006; Corso et al., 2020). In particular, the aglycone and decorated quercetin and kaempferol flavonols are accumulated at high concentration in the seed coat, endosperm and embryo. The localization of flavonols in *A. thaliana* seeds depends on their decorations, particularly on the glycosylation pattern, the monoglycosylated flavonols being accumulated in the seed coat and/or endosperm, and di and tri –glycosylated flavonoids accumulated in the embryo of Arabidopsis seeds (Routaboul et al., 2006). PAs are specifically synthetized in the seed coat, where they are accumulated at high concentrations at early Arabidopsis seed developmental stages, starting from 1-2 days after pollination (Lepiniec et al., 2006). Differently from flavonoids, glucosinolates (GSLs) are synthetized in maternal tissues outside of the seeds and then transported to seeds in *A. thaliana* (Nour-Eldin and Halkier, 2009; Sanden et al., 2023; Xu et al., 2023). Recent results nicely showed that, in *A. thaliana* seeds, GSL glucoiberin and glucoraphanin are synthetized in the funiculus and then transported to the seed embryo, where they accumulate and are subject to modifications (Sanden et al., 2023; Xu et al., 2023). In particular, the side chain of those two methionine-derived aliphatic GSL is hydroxylated by the 2-oxoglutarate-dependent dioxygenase AOP3 to form hydroxylated/hydroxylalkyl GSL (OH-GSL) (Kliebenstein et al., 2001). Previous reports suggested that GSL sinapoylation and benzoylation occurring on the -OH from the side chain structure are catalysed by SCPL17, a serine carboxypeptidase-like, while the synthesis of benzoyl compounds is catalysed by the benzoate-CoA ligase BZO1 (Reichelt et al., 2002; Kliebenstein et al., 2007; Lee et al., 2012; Sanden et al., 2023). Many abundant cinnamic acids in Brassicaceae seeds derive from sinapic acid core structure and decorated metabolites, mainly sinapate esters (COO-R) (Baumert et al., 2005; Corso et al., 2020). In *A. thaliana*, the synthesis of sinapate esters starts with the glycosylation of sinapic

acid to form sinapoyl-glucose (sinapoyl-G), that constitutes the source of sinapate for the production of other sinapate-esters. These include the conversion of sinapoyl-G to sinapoyl-malate by the sinapoylglucose: malate sinapoyltransferase (SMT), the synthesis of 1,2-di-sinapoyl-G from two molecules of sinapoyl-G catalyzed by the sinapoyl-glucose:sinapoyl-glucose sinapoyltransferase or the synthesis of sinapoylcholine (sinapine) by 1-O-sinapoyl-beta-glucose:choline sinapoyltransferase (SCT) (Fraser et al., 2007; Stehle et al., 2008b; Stehle et al., 2008a; Li et al., 2010; Barreda et al., 2024a). While many of the above cited SMs are essential for Brassicaceae seeds to cope with abiotic and biotic stresses (Corso et al., 2020), many efforts have been done by breeders to drastically reduce their content in seeds since some of them are considered antinutritional compounds for human and animal consumption (Di Gioia et al., 2019). Among them sinapine (an alkaloidal amine) is one of the most abundant metabolites synthetized in Brassicaceae model and crop species and can constitute up to 2% of seed weight in these species (Menard et al., 2022). Sinapine reduces protein quality and digestibility and, because of its bitter taste, makes the meal obtained from Brassicaceae seeds less suitable and palatable for animal consumption (Bell, 1993; Huang et al., 2008; Menard et al., 2022). Similarly to other SMs accumulated in seeds, GSLs are part of the arsenal of compounds that plant species use to defend themselves against insects, vertebrate herbivores and pathogens (Dalling et al., 2020). GSLs, such as glucoraphanin, glucobrassicin and sinigrin, are considered antinutritional because of their bitter taste or their potential toxicity due to the accumulation of degradation products such as isothiocyanates, thiocyanates, nitriles and epithionitriles. Nevertheless, many reports highlighted a wide range of beneficial effects on human health, including but not exclusively due to their anticancerogenic, antibacterial, antidiabetic, and antioxidant activities (Di Gioia et al., 2019). In addition, despite GSLs were considerably reduced in Brassicaceae seed by breeding programs, limited information is available about the activity and/or toxicity of a large set of aglycone and decorated (e.g. glycosylated, sinapoylated and benzoylated) GSLs, together with their related degradation products.

Seeds from Brassicaceae crop species are used for oil and protein extraction, and for valorisation of co-products, including many SMs, but less information is available about SM diversity in these seeds and, hence, modifications, tissue-distribution and accumulation during seed development, unlike the extensive studies conducted in *A. thaliana*. Among these species *Camelina sativa* (camelina) is an oilseed crop cultivated for human and animal nutrition, and for industrial uses, increasingly used as model for research on Brassicaceae species. The characteristics of camelina make it a valuable crop for diversifying and intensifying production in agroecological cropping systems (Zubr, 1997; Neupane et al., 2022). Camelina seeds oil is rich in omega-3, polyunsaturated fatty acids (PUFA) and vitamin E (tocopherol), and is used in food, cosmetics and as a biofuel (Faure and Tepfer, 2016; Zanetti et al., 2017). Besides oil, camelina seeds accumulate high levels of SMs, including many phenylpropanoids and glucosinolates with a wide range of biological functions (Quéro et al., 2016; Alberghini et al., 2022; Boutet et al., 2022). *Camelina sativa* is hexaploid (2n=6x=40), with three sets of chromosomes and therefore three genomes (Kagale et al., 2014). *C. sativa* genome (n = 20, N^6N^7H) originated from a

hybridization between the parental genomes *C. neglecta*–like (n = 13, N^6N^7) and *C. hispida* (n = 7, H) (Mandáková et al., 2020). In particular, every chromosome or chromosomal region of *A. thaliana*, was represented in three independent chromosomes in the genome of *C. sativa*. We recently analyzed the seed SM landscape in camelina seeds, highlighting the accumulation of a large set of SM classes, including GSLs, flavonoids, cinnamic acids, monolignols, terpenoids and amino acids and derivatives (Boutet et al., 2022). This work documented major effects of the environment on the stimulation of seed SMs in *C. sativa* seeds and showed that SMs, including GSLs and phenylpropanoids, showed a higher environmentally-induced plasticity with respect to most of the primary metabolites, including some sugars, fatty acids, proteins and lipids (Alberghini et al., 2022; Boutet et al., 2022). While we explored in detail SM diversity and plasticity, few information is available about SM distribution and expression of SM-related proteins and/or genes in developing and germinating seed tissues, including coat, endosperm and embryo, of camelina and other oilseed crop species (Fang et al., 2012). Knowing the spatiotemporal synthesis, transport and degradation of SMs in a crop species such as camelina would be indeed of main importance to develop new tools to finely modulate the ratio between beneficial and antinutritional compounds in seeds.

In this study we used untargeted metabolomics, proteomics and transcriptomics to analyze synthesis, transport, modifications and degradations of SMs that are accumulated in the different seed tissues of *C. sativa* (i.e. seed coat, endosperm, and embryo), during seed development and germination. Our results showed specific signatures for many SMs during seed development. In addition, we highlighted major differences in the accumulation patterns of important SM categories in the seed coat and embryo compared to what has been observed in *A. thaliana* seeds. In particular, while the defense and antinutritional GSL compounds are mainly accumulated in the embryo in *A. thaliana*, in *C. sativa* seeds GSLs were accumulated in the seed coat and endosperm, and the corresponding degradation products (isothiocyanates) were present at high level in the embryos of dry seeds.

Results

Camelina sativa seed coat and embryo development

Camelina sativa seeds reach maturity in about 50 days from the onset of flowering. Plants were cultivated in a greenhouse (see materials and methods for more detailed information). Seeds were harvested at 13, 21, 28, 35, 42 Days After Flowering (DAF), at dry seeds, and at two germination stages (6 hours and 24 hours from seed imbibition) (Figure 1). Seed coat and endosperm (SCE), and seed embryo (SE) were separated and used for multi-omic analyses. As for the ratio between the two seed tissues, at 13 DAF 77% of the seed weight is constituted by the SCE, at 21 DAF the two tissues represent together 50% of the seed weight, while from 28 DAF the SE represents around 70% of the seed weight (Figure 1).

Specialized metabolite diversity in camelina seed coats and embryos

Untargeted SM profiling revealed a large diversity in camelina seeds (Figure 2; Table S1 and S2A). Untargeted metabolomic analyses (positive and negative ElectroSpray Ionization, ESI+ and ESI-) allowed the detection of 4061 peaks of putative metabolites (i.e. metabolic features) (Table S2B). Molecular network analyses enabled the clustering of SM compounds according to their MS/MS spectra (Figure 2A, Table S2A) and was used to improve the annotation of the unknown metabolic features (Olivon et al., 2018; Boutet et al., 2022). Annotated SMs in camelina seed coats and embryos belong to several classes, among which amino acids (AA) and derivatives (208 SMs), cinnamic acids (176 SMs) and flavonols (64 SMs) were the most numerous, followed by nucleosides and derivatives, terpenoids, flavan-3-ols and proanthocyanidins (PAs), fatty amides, carbohydrates, carboxylic acids, glucosinolates (GSLs), flavonoids, sulfoxides, lignans, glutathione and derivatives, isothiocyanates (ITCs) and amines (Figure 2B and Table S2A).

The average intensity of SMs in seed coats and embryos allowed the identification of the top-10 metabolites with highest relative intensity (Table S1). Sinapine (cinnamic acids/alkaloids) was the SM showing the highest intensity for both seed coat/endosperm (SCE) and embryos (SE). In addition, two ITCs (9-Methylsulfinylnonyl ITC and 10-Methylsulfinyldecyl ITC) deriving from the degradation of two major GSL compounds, rutin (flavonol) and adenosine (nucleosides and derivatives) were highly accumulated metabolites common to seed coats and embryos. Among the SMs with high intensity and specific to the SCE, there were one ITC (11-Methylsulfinylundecyl ITC), the GSL glucoarabin (9-methylsulfinylnonyl GSL) and a GSL derivative (p2553; proposed formula: $C_{17}H_{32}NO_7S_2^+$), together with the epicatechin (flavan-3-ol) and a cinnamic acid with unknown formula (p8). Among the seed-embryo–specific SMs with the highest intensity there were a triglycosylated flavonol, putatively annotated as quercetin-(xyloside-rhamnoside-glucoside), a sulfoxide (p29, proposed formula $C_{13}H_{25}NOS$) and three AA and derivatives (gamma-Glu-Leu, p13 with proposed formula $C_{11}H_{20}N_2O_5S$ and (2S)-2-amino-5-[[(1S)-1-carboxy-2-methylsulfanylethyl]amino]-5-oxopentanoic acid).

Camelina seed specialized metabolome is shaped by the tissue and the developmental stage

The untargeted metabolomic data obtained from seed coats and embryos harvested at six developmental and two germination stages were used to perform a principal component analysis (PCA). 24,3 % and 19,6 % of the variability observed among the samples can be explained respectively by the axis PC1, which is associated to the type of tissue, and the axis PC2, which is associated to the developmental stage (Figure 2C). The PCA performed allowed to clearly separate the samples according to the tissue, confirming that seed coat/endosperm and embryo present contrasting specialized metabolome signatures. Hence, Differentially Accumulated Metabolites (DAMs) according to the “developmental stage” (13 days after flowering (DAF) to dry seeds, and germinating seeds) factor were identified (ANOVA, FDR<0.05) separately for the two seed tissues considered (SCE and SE). The statistical analysis highlighted that the accumulation of 62% (2523) and 73% (2973) of SMs was impacted by the

seed developmental stage in the SCE and SE, respectively (Figure 2D).

A core-set of specialized metabolites is regulated during seed embryo and coat-endosperm development and maturation

In order to identify stage-specific DAMs, a hierarchical clustering analysis was conducted for SCE and SE tissues using DAMs intensities. The analysis allowed the separation of the DAMs into five clusters for seed coat-endosperm (I to V; Figure 3A) and four clusters for seed embryo (VI to IX; Figure 3B). The clusters allowed the identification of DAMs accumulated from early (13 DAF) to late (42 DAF and dry seeds) seed developmental stages, and of germination-specific DAMs. Afterwards, using the DAMs belonging to each cluster, an enrichment analysis was performed to identify the metabolic categories that are characteristics of each tissue and developmental stage. The enrichment analysis was performed using the metabolic categories identified with the molecular network analysis, both the 'annotated' and 'not-annotated' (or 'unknown') ones (Table S2A, column "Metabolite class"). In particular, not-annotated metabolites that clustered together in the molecular network analysis (Figure 2A) and therefore have similar chemical structures, were regarded as a distinct metabolic category. Besides the SMs assigned to known metabolic categories, several not-annotated SMs showing similar MS/MS profiles were enriched in different clusters displaying specificities to some seed tissue and/or developmental stage (Figure 3A and B).

Regarding the SCE specialized metabolome (Figure 3A), carboxylic acids and flavan-3-ols and PAs were found to be enriched and specific to cluster I, characterized by DAMs with higher accumulations at 21 DAF and 28 DAF. In addition, glucosinolates, and nucleosides and derivatives SMs were found to be enriched in cluster II, which included DAMs with high intensity at 13 DAF. Cinnamic acids, AA and derivatives, flavonols and lignans categories were enriched in cluster III, which was the biggest for SCE tissues and showed SMs with increasing accumulation from 28 to 42 DAF. Finally, AA and derivatives, sulfoxides, amines and carbohydrates were enriched in clusters IV and V that are specific to seeds during germination stages.

Concerning the seed embryo (SE) SMs (Figure 3B), no enriched metabolic categories were found for SMs from cluster VI, which is characterized by DAMs more accumulated during germination 24h after imbibition. Cluster VII, which includes DAMs with the highest accumulation mainly at the early seed developmental stage (13 DAF), is enriched for fatty amides and terpenoids. Cluster VIII is the largest one and contains SMs with increasing intensity from 35 DAF to dry seeds, and showed an enrichment for several metabolic classes, including cinnamic acids, AA and derivatives, flavonols, carboxylic acids and flavan-3-ols and PAs. From work on Arabidopsis we know that flavan-3-ols and derivatives are not produced in the embryo (Lepiniec et al., 2006) and that these metabolites could originate from degraded cells of the endothelium and, hence, will not be further discussed in this article. Finally, AA and derivatives and isothiocyanates (GSLs degradation products) were found to be enriched in cluster IX, which is characterized by DAMs with higher accumulation predominantly at 21 and 28 DAF.

According to the results highlighted by the enrichment analysis and the broad interest for seed biology and quality, we will focus on phenylpropanoid (flavonoids, cinnamic acids and monolignols) and GSL metabolic categories, and on genes or enzyme categories involved in their modifications (i.e. glycosylation, methylation, hydroxylation and acylation) (Barreda et al., 2024b).

Proteomic analyses allowed the identification of SM-related proteins with specific expression pattern during seed development

To get a better understanding of the regulation and accumulation dynamics of SM-related pathways in camelina SCE and SE, data-independent acquisition (DIA) proteomic combined with ion mobility spectrometry analyses were performed on the same samples used for untargeted metabolomics. These analyses allowed to measure the abundance of 11462 proteins (Table S3). The PCA performed with proteomic data separated the samples according to seed developmental and germination stages (Figure 4A). As for metabolomic data, Differentially Accumulated Proteins (DAPs) impacted by the "developmental stage" factor were identified (ANOVA, FDR<0.05) separately for SCE and SE (Figure S1). The statistical analysis highlighted that the accumulation of 14% (995) and 33% (2418) DAPs was specifically affected by the seed developmental stage in SCE and SE, respectively, while 53% (3912) of detected proteins were differentially expressed and common to SCE and SE (Figure 4B and S1). Next, the proteins involved in GSLs, flavonoids, cinnamic acids and monolignols biosynthesis/pathways, and the enzymes putatively involved in SM modifications that were differentially expressed in SCE (333 DAPs) and SE (374 DAPs) during development, were identified. Among these DAPs, 273 were in common between SCE and SE (Table S3 and Figure 5). A hierarchical clustering analysis was conducted for SCE and SE seed tissues using GSLs, flavonoids, cinnamic acids and monolignols –related DAPs (Figure 5A and 5B). DAPs were separated into six clusters for SCE (Figure 5A) and seven clusters for seed embryo (Figure 5B), according to their accumulation patterns. In SCE, most DAPs belonging to selected SM categories were accumulated at earlier seed developmental stages, either at 13 and/or 21 DAF (clusters IV and VI). Similar results were observed for SE, except for proteins involved in SM glycosylation (belonging to glycosyltransferase GT1 group) that were accumulated at several seed developmental and maturation stages, particularly during seed germination (cluster VII).

Multiomic data highlighted seed coat-endosperm and embryo -specific accumulation patterns of proteins and metabolites related to glucosinolates, flavonoids and cinnamic acids

The accumulation of metabolites and proteins belonging to GSLs (Figures 6 and 7), flavonoids (Figures 8 and 9), and cinnamic acids and monolignols (Figures S2 and S3) categories was further characterized in the SE and SCE. Given that *C. sativa* has a hexaploid genome, three homeolog genes could code for the proteins involved in a specific enzymatic activity. Proteomic results showed in this study allowed the identification of the proteins copies specifically accumulated in SE and SCE and involved in SM biosynthesis, modifications and degradation (Table S3; Figure 5). For GSLs, cinnamic acids and

monolignols one copy was expressed for most proteins coding for the same enzymatic activity for both SCE and SE, while flavonoid proteins were mostly expressed at 1 copy for SCE and two copies for SE (Figure S4).

The accumulation and composition of GSLs and their degradation products that were differentially accumulated in the SCE and/or the SE were remarkably different between the two seed tissues considered in this study (Figure 7). Most GSLs that were differentially accumulated during seed development, including the seed-specific aliphatic GSL 8-methyl-sulfinyl-octyl-GSL, glucoarabin (9-methyl-sulfinyl-nonyl-GSL; GSL9), glucocamelinin (10-methyl-sulfinyl-decyl-GSL; GSL10) and gluconesliapaniculatin (11-methyl-sulfinyl-undecyl-GSL; GSL11), were heavily accumulated in the SCEfrom early seed developmental stage (13 DAF) to later stages (35 DAF) (Figure 7A). The indolic GSLs glucobrassicin and neoglucobrassicin showed the highest intensity at 13 DAF. Differently from GSL metabolites, most of their degradation products (ITCs) were differentially accumulated only in the embryo, where they showed the highest accumulation level at dry mature seed stage in SE (Figure 7A, 7B). The contrasting GSL distribution between SCE and SE is further highlighted when comparing the relative accumulation levels of the main annotated and differentially accumulated GSL and ITC degradation products between these tissues (Figure 7B). Glucoarabin (GSL9) was the GSL with the highest relative abundance, and showed its peak at 35 DAF in SCE. Consistently, the main GSL degradation product was the 9-Methylsulfinylnonnyl ITC that showed the maximum relative intensity at 13 DAF for SCE and in dry seeds embryo. Gluconesliapaniculatin (GSL11) was the second most abundant GSL and peaked at 35 DAF, followed by the glucocamelinin (GSL10), which was characterized by two peaks at 13 and 35 DAF (Figure 7B). 10-Methylsulfinyldecyl and 11-Methylsulfinylundecyl ITCs, which are the corresponding degradation products of GSL10 and GSL11 were differentially and highly accumulated in the SE in dry mature seeds (Figure 7B). Finally, 8-methylsulfinyloctylglucosinolate was also specific to the SCE, but showed lower relative intensity compared to GSL9, GSL10 and GSL11. In most cases, the accumulation of proteins involved in GSL side-chain elongation and core-structure formation was higher at earlier seed developmental stages (13 and 21 DAF) in both SCE and SE (Figure 6). Among the proteins related to GSL modifications we identified two Flavin-Monooxygenases, namely FMOGS-OX 5 and 7, putatively implicated in GSL hydroxylation. These proteins showed the highest expression at early (FMOGS-OX5, 13-21 DAF) or late (FMOGS-OX7, 35 DAF to dry seeds, specific to SE) seed developmental stages. FMOGS-OX5 was also found in the SE during germination. Proteins playing a role in GSL degradation were expressed during the whole seed development and germination in both SCE and SE, with some specificities and exceptions. For exemple PYK10 myrosinase accumulation was highest at the germination stage, while TGG1 myrosinase was expressed throughout seed development in SCE and SE, and also during germination specifically in SCE tissues.

As observed for GSLs, the accumulation of different flavonoid classes showed marked differences according to the seed tissue and the developmental stage considered (Figure 9). In particular, in SCE

tissue flavan-3-ols and derivatives were accumulated from 13 to 28 DAF in their aglycone form (epicatechin, catechin and procyanidin B1 and B2), while glycosylated flavan-3-ols (epicatechin and catechin glycosides) showed the highest accumulation level at 35 and 42 DAF. While most flavonol metabolites, corresponding to quercetin and kaempferol mono-, di- and tri- glycosides, were accumulated at later seed developmental stages in SCE (35-42 DAF) and SE (dry mature seeds), isorhamnetin (i.e. methylated quercetin) -glycoside was accumulated at earlier developmental stages in both tissues (Figure 9). Interestingly, when the isorhamnetin presented a second -glycosyl group (isorhamnetin-(Rha-Glc), it was mainly accumulated at later seed developmental stages in both tissues (Figure 9). Enzymes involved in the biosynthesis of the core flavonoid pathway and in flavonol specific pathway were expressed at 13-21 DAF in both SCE and SE, with a shift between the two tissues for some of them (Figure 7). This is the case of FLS1 which showed the highest expression at 21 DAF in the SCE and at 13 DAF in the SE. On the contrary FLS3, showed an earlier expression in SCE compared to the SE. ANR, involved in flavan-3-ols synthesis showed its highest expression in the SCE at early developmental stages. As concerns proteins involved in flavonoid glycosylation, it is worth mentioning that the UGT glycosyltransferase F7GlcT was specific to the embryo and showed the highest expression during the germination stage.

While for both GSLs and flavonoids the metabolite landscape and composition were markedly different between the SCE and SE, fewer differences between these two tissues were observed for cinnamic acid composition (Figure S3). Sinapine was the most abundant metabolite belonging to this category in both tissues (Table S1), and showed the highest intensity at 35-42 DAF for SCE and in dry seeds for SE. Similarly, other cinnamic acid derivatives, e.g. sinapoyl, feruloyl and coumaroyl glycosides, were accumulated at 35-42 DAF in the SCE and at dry seeds in the embryo. In addition, guaiacyl-ferulic acid-glucose was more abundant during germination in SCE and SE. Some differentially accumulated proteins involved in cinnamic acids synthesis or decoration showed specific expression in seed embryo (Figure S2). This is the case of the sinapic acid:UDP-glucose glucosyltransferase (SGT1), involved in sinapic acid glycosylation, of the Sinapoylglucose:malate sinapoyltransferase (SMT), which produced sinapoyl-malate, and of 1-O-sinapoyl-beta-glucose:choline sinapoyltransferase, catalysing the synthesis of sinapine.

Transcriptomic analyses in seed embryo.

Transcriptomics were performed on a subset of samples compared to those used for metabolomic and proteomic analyses. In particular, seed embryo at 13, 21, 28, 35, 42 DAF and at 6 hours after germination stage were subjected to RNA-Seq analyses (Figure 10, Figure S5 and Table S4). As observed for proteomic and untargeted metabolomic analyses, large differences in gene expression (FDR < 0.05) were observed between seed developmental stages, and between seed development at 42 DAF and at 6 hours of germination. The PCA confirms these results and the quality of the data obtained since it allowed a clear separation of embryos from different developmental stages and showed that the three

biological replicates were clustered together (Figure 10). Next, we extracted Differentially Expressed Genes (DEGs) involved in GSL, flavonoid, and cinnamic acid synthesis, transport, decorations (or modifications), degradation and regulation (Figure 10 and Figure S6). Many DEGs related to GSLs were expressed at 13 DAF and showed generally an earlier expression peak compared to what has been observed for corresponding proteomic data. As for flavonoids, all DEGs involved in their biosynthesis and most DEGs involved in their decoration showed the highest expression level at 13 DAF. *MYB11* and *MYB111* transcription factors (TFs) involved in the regulation of flavonol metabolites, were expressed at 13DAF, while *TTG1* TF showed its expression peak at germination stage. DEGs related to cinnamic acids were characterized by larger differences concerning their expression pattern. In particular, genes coding for Serine-carboxypeptidase-like enzymes involved in sinapic acid ester biosynthesis showed the highest expression level at 13 DAF (*SCPL12-FTP1*), 21 DAF (*SCPL9-SNG2-SCT*), 28 DAF (*SCPL7*), 42 DAF (*SGT1*) and during germination for *SST* (*SCPL9*).

Discussion

The accumulations of SMs strongly influence the quality of plant-derived products. While the spatial distribution of these compounds in plant leaves and roots tissues has been described for some model and crop species (Brown et al., 2003; Czerniawski et al., 2021), very limited information and mainly restricted to primary metabolites is available about their spatial accumulation in seed tissues (Zhang et al., 2021). Given their major impact on seed quality, elucidating SM distribution in seed coat, endosperm and embryo during development and germination is of main interest in Brassicaceae species, including the oilseed crop *C. sativa* (Kagale et al., 2014; Faure and Tepfer, 2016; Zanetti et al., 2021).

In the present study we combined untargeted metabolomic, proteomic and transcriptomic analyses to unveil seed SM accumulation (synthesis, transport and/or modifications/decorations) in developing seed coat and endosperm, and seed embryo. Our results emphasized a large diversity of SM in camelina seeds, as shown in a previous study (Boutet et al., 2022).

A progressive but significant evolution of the specialized metabolome can be observed during embryo development. The sample replicates of the earliest stage of seed development studied (13 DAF) and the samples replicates of the latest seed development stage (dry seeds) are the most distant; and intermediary developmental stage samples are distributed between these two samples, following a chronological order. Interestingly, the embryo specialized metabolome of 6h and 24h germinating seeds appear to be similar to those of seed embryo in the early and middle seed developmental stages, respectively (Figure 2C). Hence, these data suggest that there is a considerable specialized metabolome change/shift occurring in the embryo from dry seed to germinating seed and that the metabolome is afterwards following an evolution equivalent to the one that takes place during seed development. Similar results have been observed for the SCE, for which seeds harvested at 13 DAF and germinating seeds at 24h after imbibition showed a similar specialized metabolome.

RNA-Seq transcriptomic analysis revealed that the expression of a majority of genes involved in GSLs, flavonoids, cinnamic acids, and monolignols biosynthetic pathways peaks at early seed developmental stage (13 DAF) in the embryo (Figure 10 and Figure S6). Consistently, a large part of the enzymes involved in GSL, flavonoid, cinnamic acid and monolignol synthesis were accumulated in the embryo between 13 DAF to 21 DAF. (Figure 5B), consistent with a delay between gene expression and the accumulation of corresponding proteins. Interestingly, enzymes from SCE were also mostly expressed at early stage of seed development (Figure 5B). However, SMs displayed contrasting accumulation patterns between both seed tissues (Figure 3). Indeed, whereas both SE and SCE had some groups of SMs highly accumulated at 13 DAF, a large part of SCE SMs were mainly accumulated at 35 and 42 DAF stages, whereas the majority of embryo SMs were highly accumulated at dry seed stage (Figure 3). Both of those SCE and SE SM clusters were enriched in AA and derivatives, cinnamic acids and flavonols, indicating that those metabolic categories are biosynthesised and/or stored in SCE during seed development and transported to SE between the maturation (42 DAF) and dry seed stage (Figure 3). Hence, those results emphasize the importance of SCE tissues, at the interface between the mother plant

and the embryo, in the support of embryonic growth (Yan et al., 2014). PCA analyses on untargeted metabolomic data highlighted that, while both factors are impacting SMs diversity, metabolite signature seems more similar within the same tissue during seed development than in different tissues at a specific developmental stage (Figure 2C). For example, proanthocyanidins (PAs) are specifically accumulated in *Arabidopsis thaliana* seed coat from the early seed developmental stages (1-2 DAF) and reach the maximum peak at 6-7 DAF for Col-0 ecotype and 12 DAF for Cvi-0 ecotype (Routaboul et al., 2012; Corso et al., 2020). In addition, PA accumulation is paralleled by the expression of the *BANYULS* gene, which is involved in their synthesis (Debeaujon et al., 2003). This trend is confirmed in camelina seeds, where the anthocyanidin reductase (ANR) protein, encoded by a *BANYULS* homolog (Xie et al., 2003), showed the highest expression peak at 13 DAF in SCE samples (Figure 7) and, consistently, the flavan-3-ols derivatives (catechin, epicatechin and procyanidin) are highly accumulated at 13, 21 and 28 DAF (Figure 8). Differently from the aglycone forms, catechin and epicatechin glucosides (as most glycosylated and acylated flavonoids) are accumulated at later stages of seed coat development in camelina, as shown for *Vitis vinifera* seeds (Zerbib et al., 2018) (Figure 8), consistent with an increase stability of these glycosylated metabolites, as previously shown by Raab et al. (2010).

Besides flavonoids, the defensive and antinutritional GSLs and related ITC compounds, deriving from GSL degradation, showed contrasting accumulation pattern when comparing SCE to embryo tissues, in camelina seeds (Figure 6). While some GSLs have a negative impact on seed oil and protein quality and digestibility, and are in some cases toxic for animals and humans, others are powerful antioxidants with nutritional and health benefits (Traka, 2016; Di Gioia et al., 2019). Hence, they have been a target of domestication for some Brassicaceae (e.g. Brassica and Camelina species) for which many efforts are still being done to modulate the content or nature of GSLs in seeds (Nour-Eldin et al., 2017; Hölzl et al., 2023; Mann et al., 2023). Nevertheless, GSLs are antioxidants accumulated in response to abiotic stresses (Rao et al., 2021; Mácová et al., 2022) and are essential for plant resistance to pathogens, so much so that their systematic elimination could prove counterproductive. In this work we highlighted that camelina GSLs and Arabidopsis seeds are characterized by contrasting GSL accumulation patterns. In particular, while in camelina seed GSLs are mostly accumulated in the SCE and ITCs (GSL-degradation products) are accumulated in the embryo, in *A. thaliana* they are exclusively accumulated in the embryo (Sanden et al., 2023), highlighting two contrasting patterns of synthesis and transport of these compounds. As shown in this and other studies (Berhow et al., 2013; Amyot et al., 2019), camelina seeds mainly accumulate long carbon(C)-chain GSLs, notably C9-C11 methylsulfinylalkyl glucosinolates, while Arabidopsis seeds accumulate GSLs with shorter lateral carbon chain. In this study we also showed the presence of the C8 8-methyl-sulfinyl-octyl-GSL in SCE (Figure 7B) that is not usually detected in camelina seeds. The contrasting GSL accumulation pattern between these two Brassicaceae species could be partially due to the absence of benzoylated and sinapoylated GSLs in camelina seeds, whereas they accumulate at high level in Arabidopsis embryos (Kliebenstein et al., 2007; Lee et al., 2012; Bussell et al., 2014; Sanden et al., 2023). In particular, sinapoyl- and benzoyl-

GSLs could be less toxic than non-decorated GSLs or could be transported into the vacuole, thus facilitating their accumulation in the embryo. On the contrary, non-decorated GSLs accumulate in seed coat and/or endosperm avoiding the accumulation of these toxic compounds in the embryo. Nevertheless, no clear information is available regarding the biological activity and role of acylated-GSLs in seeds, which deserves further investigation.

Interestingly, fewer SMs showed peaks of accumulation at germination stages (6h Germ and 24h Germ) compared to the seed developmental stages for both tissues (Figure 3). This suggests that many SMs are degraded rather than synthetized during germination. Similarly, there were fewer proteins strongly accumulated at germination stages compared to seed developmental stages for both tissues (Figure 5). However, it is worth mentioning that glycosyltransferases were found to be highly expressed in SE at both 6h and 24h germination stages (Figure 5). This could be linked to starch and sugar remobilization occurring during the germination and establishment of the seedling. Likewise, carbohydrates metabolic category was found to be specifically enriched upon seed germination (Figure 3). Moreover, 9-Methylsulfinylnonyl, 10-Methylsufnyldecyl and 11-Methylsulfinylundecyl ITCs were found to decrease progressively from dry seed stage to 24h Germination stage (Figure 7). This decrease of ITCs was concomitant with an increase in expression of several enzymes involved in GSL degradation, and thus potentially ITC production (Figure 10), suggesting that ITCs may have a role during camelina seed germination by providing for instance an important source of AA, sulfur atoms or other molecules.

To conclude, GSLs, flavonoids, cinnamic acids, and monolignols are largely synthetized and accumulated in camelina seeds and, therefore, determine their nutritional quality and play a key role in seed interaction with biotic and abiotic environments. Characterizing the spatial dynamics of seed SMs will contribute to the development of crops with an optimized distribution of beneficial and toxic metabolites for seeds quality and animal nutrition or to the development of new strategies (e.g. fractionation or germination) to reduce the level of undesired compounds after harvest.

Materials and Methods

Plant culture and seed harvest

Camelina sativa (L.) "Céline" genotype was used for the plant culture and analyses. Seeds of the three biological replicates (10 plants/replicate) were sown with a gap of two weeks between each replicate and were cultivated in different part of the greenhouse. Plants were cultivated in 16 cm diameter pots (from TEKU) in a greenhouse: 16h light (23-25 °C) / 8h dark (15°C) (Osram Powerstar E40 HQI-BT 400W lights). Relative humidity room was set on 45%. A nutrient solution containing 140 milligrams of nitrogen per liter (140 mg N L^{-1}) was used to irrigate the plants.

Flowers were tagged just before opening to count opened flowers and harvest seeds at defined stages (Sakai et al., 2018). Seed embryos and coats/endosperms were separated and harvested at 13, 21, 28, 35, 42 DAF, at dry mature seed stage, and at 6 hours and 24 hours, after imbibition, during germination.

For seed pictures acquisition, siliques from each time-point were carefully dissected to harvest the seeds. Chloral hydrate solution (chloral hydrate:H_2O_2:glycerol (8:2:1, w/v/v) was used to bleach seed coats and fix the samples that were placed between slide and slip (Boisson et al., 2001). Seeds were then stored for 24h to 72h at 4°C before imaging. Images were acquired on Axioplan 2 imaging epifluorescence microscope after several weeks of seed coat clearing (Zeiss, Göttingen, Germany).

Specialized metabolites extraction and analysis

Metabolite extraction and injection. Untargeted metabolomic data analyses and date were described in (Boutet et al., 2022). Briefly, metabolites were extracted from 20 mg of seed coats and embryos using a MeOH: Methyl-tert-butyl: H2O (1:3:1) buffer, while polar and semipolar metabolites were separated from the oil and protein fractions using a MeOH: H2O (1:3) buffer. Polar and semipolar metabolite fraction was used for metabolomic analyses. Untargeted metabolomic data were acquired using a UHPLC system (Ultimate 3000 Thermo) coupled to quadrupole time of flight mass spectrometer (Q-Tof Impact II Bruker Daltonics, Bremen, Germany). For chromatographic separation a Nucleoshell RP 18 plus reversed-phase column was used. Samples were injected in positive and negative ionisation mode (ESI+ and ESI-).

Data processing and ion annotation. ESI+ and ESI- were processed using MZmine 2.52 software (http://mzmine.github.io/).

Metabolite annotation was performed in four steps: i) LC-MS/MS data were compared with a home-made library (IJPB chemistry and metabolomic platform) containing more than 400 standards or experimental common features; ii) LC-MS/MS data were also searched against the available MS2 spectral libraries (Massbank NA, GNPS Public Spectral Library, NIST14 Tandem, NIH Natural Product and MS-Dial); iii) a molecular network analysis was used to assign not-annotated metabolites to a chemical family (see Boutet *et al.*, 2022 and Olivon *et al.*, 2018 for further information); iv) Sirius software (https://bio.informatik.uni-jena.de/software/sirius/) was used to assign a putative annotation to

metabolic features that were not annotated during the previous steps. Raw data were normalised on the internal standard (Apigenin) and weight of seeds used for the extraction.

Statistical analyses of metabolomic data. Statistical analyses were performed using Metaboanalyst 5.0 software (Pang et al., 2021; Pang et al., 2022). In particular, an ANOVA has been conducted to identify differentially accumulated metabolites among seed developmental stages, separately for SCE and SE (adjusted p value < 0.05).

Transcriptomic analyses (RNA-Seq)

RNA extraction. RNAs were extracted with the PicoPure™ RNA Isolation Kit (ThermoFisher) following the corresponding protocol with slight modifications. Briefly, 2 to 5 mg of seeds were grinded with some polyvinylpolypyrrolidone (PVPP; phenolic compound absorbing agent) in liquid nitrogen. 100 µL of extraction buffer was added to each sample. Following, samples were vortexed, incubated at 42°C for 30 min (500 rpm) using a ThermoMixer™ C (Eppendorf), and centrifuged (2 min, 3000 g). Supernatants were collected, without picking-up the pelleted material, and placed in the pre-conditioned RNA purification columns to be centrifuged (1 min, 16 000 g). The cell extracts obtained were mixed with 100 µL of 70 % ethanol by pipetting up and down. The mixtures were placed into the pre-conditioned purification columns which were centrifuged 2 min at 100 g, to allow RNA binding to the columns, immediately followed by a centrifugation to remove flow-through (30 sec, 16 000 g). The columns were washed by adding 100 µL of wash buffer 1 and centrifugation (1 min, 8 000 g). DNase treatment was made by adding 40 µL of DNase solution mix (10µl DNase I Stock solution + 30µl Buffer RDD; RNase-Free DNase Set, Qiagen) to each column, followed by 15 min of incubation (ambient temperature) and washing of the column (addition of 40 µL of wash buffer 1 and centrifugation 30 sec at 8 000 g). Columns were washed twice with 100 µL of wash buffer 2 (2 min at 8 000 g) and a last centrifugation was made after flow-through waste discard to remove all traces of wash buffer (1 min, 16 000 g). 12 µL of elution buffer were added to each column, which were incubated 1 min at ambient temperature before RNA elution centrifugations (1 min at 1 000 g followed by 1 min at 16 000 g). Obtained RNAs were stored at -80 °C until sequencing.

RNA sequencing. Sequencing of messenger RNAs (mRNAs) was performed by BGI Genomics (Hong Kong, https://www.bgi.com). Prior to cDNA library construction (Eukaryotic Strand-specific Transcriptome), quality and concentrations of mRNAs were checked again with the bioanalyzer from BGI platform. The sequencing (paired-end reads of 150bp, 70M clean reads per sample) was performed with DNA Nanoballs Technology (DNBSEQ™). After sequencing, raw reads were filtered by the BGI platform with the filtering software SOAPnuke (version 1.5.2, https://github.com/BGI-flexlab/SOAPnuke) (Cock et al., 2009; Chen et al., 2018) (Parameters used: l 15 -q 0.2 -n 0.05) to reach a Phred+33 fastq quality score.

Annotation and statistical analysis of transcriptomic data. High quality raw reads (Céline strain) were aligned against the *Camelina sativa* genome (strain: DH55, genome assembly:

GCA_000633955.1, annotation: https://www.camelinadb.ca/downloads/: gff and fasta to be made available if no more available in this site) using GSNAP (gmap-2021-12-17) with a set of options allowing to find novel splicing and to be tolerant to SNP between DH55 and Céline (VCF files, to be made available). htseq-count (HTSeq-0.12.4-python2) was used with the following options (-m intersection-nonempty --secondary-alignments=ignore -a 20 -s no) to count the number of reads by gene. Orthologous genes of *Arabidopsis thaliana* were searched in *Camelina sativa* using Plant.ensembl website (using Biomart) and CamRegBase (synthelogs data).

Statistical analyses were performed to identify Differentially Expressed Genes using the Edge R package (Robinson et al., 2010). Statistical analyses were performed to identify genes that were differentially expressed among developmental and germination stages in the SCE and SE separately (FDR < 0.05).

Proteomic analyses (DIA)

Sample preparation and analysis. Five µg of each protein extract were prepared using a modified Gel-aided Sample Preparation protocol (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4409837/). Samples were digested with trypsin/Lys-C overnight at 37°C. For nano-LC fragmentation, protein or peptide samples were first desalted and concentrated onto a µC18 Omix (Agilent) before analysis.

The chromatography step was performed on a NanoElute (Bruker Daltonics) ultra-high-pressure nano flow chromatography system. Approximatively 200ng of each peptide sample were concentrated onto a C18 pepmap 100 (5mm x 300µm i.d.) precolumn (Thermo Scientific) and separated at 50°C onto a reversed phase Reprosil column (25cm x 75µm i.d.) packed with 1.6µm C18 coated porous silica beads (Ionopticks). Mobile phases consisted of 0.1% formic acid, 99.9% water (v/v) (A) and 0.1% formic acid in 99.9% ACN (v/v) (B). The nanoflow rate was set at 250 nl/min, and the gradient profile was as follows: from 2 to 30% B within 70 min, followed by an increase to 37% B within 5 min and further to 85% within 5 min and reequilibration.

MS experiments were carried out using a TIMS-TOF pro mass spectrometer (Bruker Daltonics) with a modified nano electrospray ion source (CaptiveSpray, Bruker Daltonics). A 1400 spray voltage with a capillary temperature of 180°C was typically employed for ionizing. MS spectra were acquired in the positive mode in the mass range from 100 to 1700 m/z and 0.60 to 1.60 1/k0 window. In the experiments described here, the mass spectrometer was operated in PASEF DIA mode with exclusion of single charged peptides. The DIA acquisition scheme consisted of 16 variable windows ranging from 400 to 1200 m/z.

Protein identification. Database searching and LFQ quantification (using XIC) was performed using DIA-NN (version 1.6.1) (Demichev et al., 2020). The UniProt *Camelina sativa* database was used for library-free search / library generation. For RT prediction and extraction mass accuracy, we used the default parameter 0.0, which means DIA-NN performed automatic mass and RT correction. Top six fragments (ranked by their library intensities) were used for peptide identification and quantification. The FDR was set to 1% at the peptide precursor level. The variable modifications allowed were as

follows: Nterm-acetylation and Oxidation (M). In addition, C-Propionoamide was set as fix modification. “Trypsin/P” was selected. Data were filtering according to a FDR of 1%. Cross-run normalisation was performed using RT-dependent.

Identification of differentially expressed proteins. To quantify the relative levels of protein abundance between different groups, datas from DIA-NN were then analysed using DEP package from R. Briefly, proteins that are identified in 2 out of 3 replicates of at least one condition were filtered, missing data were imputed using random draws from a manually defined left-shifted Gaussian distribution and differential enrichment analysis was based on linear models and empherical Bayes statistic. A 1.2-fold increase in relative abundance and a 0.01 pvalue were used to determine enriched proteins. ANOVA were performed from PERSEUS to determine those enriched proteins. A 1.2-fold increase in relative abundance and a 0.05 FDR were used.

Clustering analyses

The R package ‘PheatMap” was used to perform hierarchical clustering analysis (hclust) and create heatmaps using omics data.

ACKNOWLEDGEMENTS

This work was supported by AgroParisTech (grant "Appel à projets scientifiques 2020" to M.C. and L. R.), the "Département de Biologie et Amélioration des Plantes" (BAP) of INRAE (grant "Appel à projets scientifiques BAP 2020" to M.C.), by the project 'Seed MetSpe' of the Labex Saclay Plant Sciences-SPS (ANR-10-LABX-0040-SPS to L.L. and M.C.) and by a project financed by Plant2Pro Carnot institute ('BrassiMet' to M.C.).

The IJPB benefits from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007). This work has benefited from the support of IJPB's Plant Observatory technological platforms. We are especially thankful to Patrick Grillot (IJPB, INRAE-Versailles) for his help during the plant culture in the greenhouse.

The authors would like to thank as well Dr Mark Tepfer (retired, former IJPB, INRAE) for the constructive discussion on *Camelina sativa* species.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

DATA AVAILABILITY

The metabolomic data and metadata have been deposited at the MassiVE data repository portal with the identifier MSV000094688 (doi:10.25345/C5HH6CH59).

The transcriptomic data have been deposited at the National Center for Biotechnology Information (NCBI) Transcriptome Shotgun Assembly Sequence Database (TSA) with BioProject identification PRJNA835304.

The proteomic data have been deposited at the iProX (integrated Proteome resources) data repository portal with the identifier IPX0008673001

AUTHOR CONTRIBUTIONS

MC directed the research. MC and LR designed the research. MC, DDV, FT, DG and CB grew the plants and harvested seed coat and embryos at defined developmental stages. SB generated the LC-MS/MS data and performed the initial quality analyses on the metabolomic data. SB, FP, LB, NK and CB performed the metabolites annotation with MetGem, SIRIUS and using MS/MS spectra. MC, LB, CB, SB, FP, NK and LL analysed and interpret the untargeted metabolomic data. BC prepared the samples for proteomic analyses. BB performed the proteomic analyses. MC analysed transcriptomic data. MC, LB and CB performed the statistical analyses on multiomic data. MC, LB and CB generated all Fig.s and supplemental material. MC and LB wrote the paper, which was edited by CB, LR and LL. All the authors commented on and approved the manuscript.

FIGURE LEGEND

Figure 1. *Camelina sativa* seeds used in this study.

Seed coat and embryo samples used in this study were harvested at 13, 21, 28, 35, 42 days after flowering. Dry mature and germinating seeds (6h and 24h) were also used.

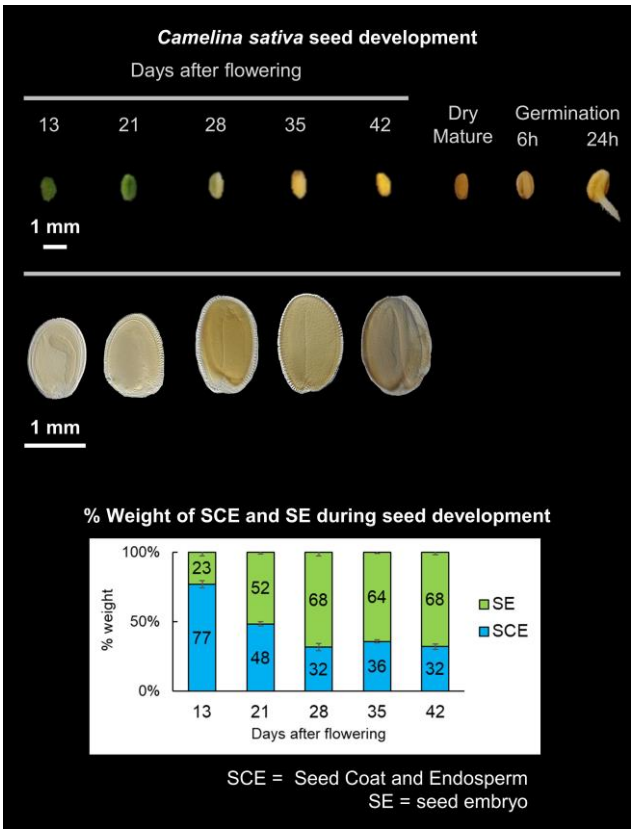
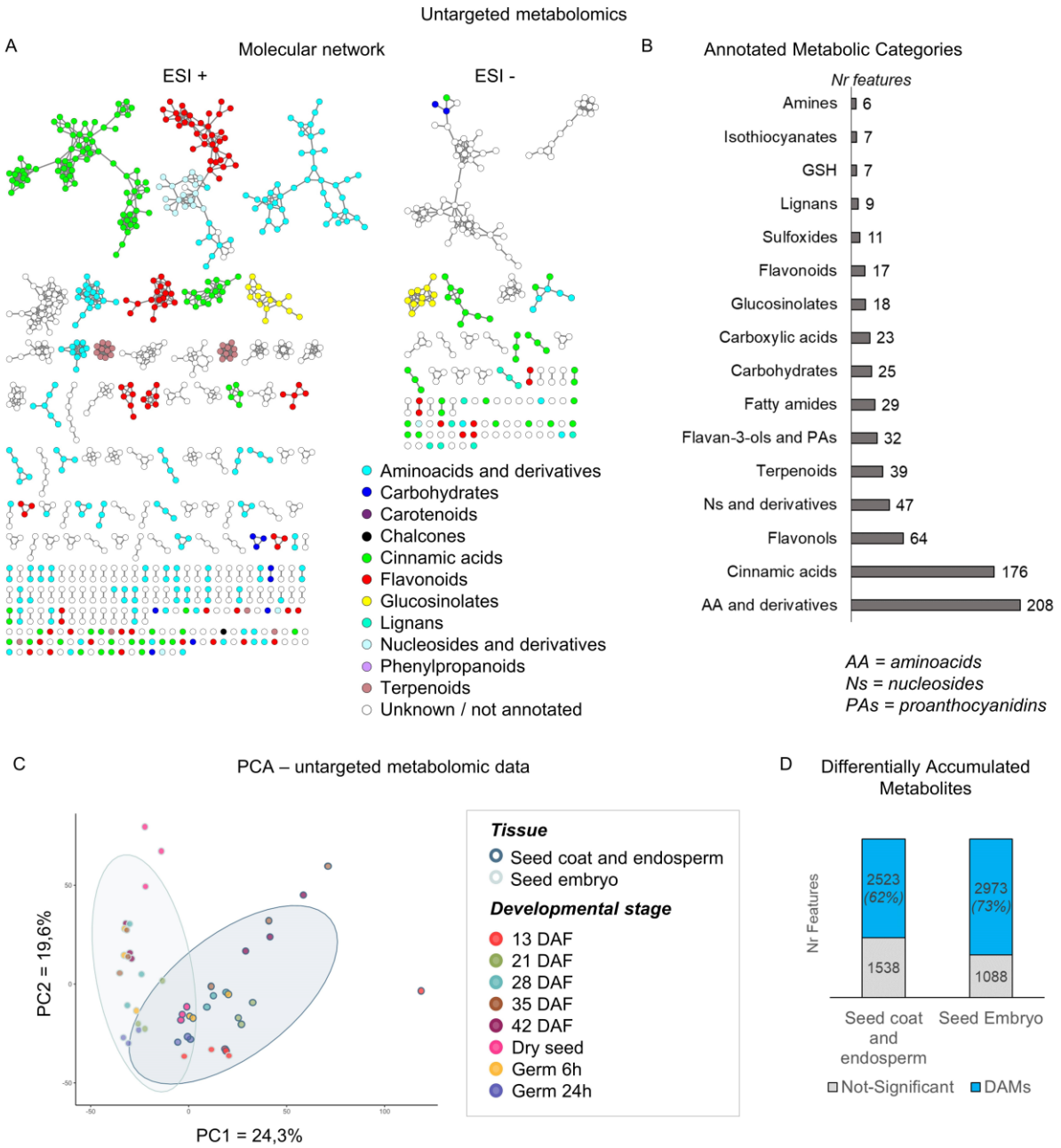


Figure 2. Untargeted metabolomics in developing and germinating SE and SCE

A) The molecular networks are shown for metabolite analyses carried out in positive (ESI+) and negative (ESI-) ionization modes. Different colours correspond to different metabolic classes. Metabolites are grouped based on their chemical structures. Cosine similarity scores of 0.85 and 0.65 were used for ESI+ and ESI-, respectively. **B)** Histogram showing the annotated/known metabolic categories. The number of metabolites that belong to each category is indicated in the figure. Categories with less than two metabolite features are not shown. **C)** Principal component analyses using the untargeted metabolomic data of SCE and SE harvested at several developmental stages and during germination. DAF = Days After Flowering. **D)** Venn diagram indicating the number of Differentially Accumulated Metabolite features (DAMf) according to the tissue and developmental stage factors.



645

Figure 3. Specialized metabolites signatures and stage-specific enrichment of metabolic categories during SE and SCE development and germination

Hierarchical clustering and enrichment analyses of Differentially Accumulated Metabolites (DAMs) specific to SCE **(A)** and SE **(B).** Annotated and unknown enriched (hypergeometric test, $p < 0.05$) metabolic categories of DAMf clusters specific to seed coats and embryos are showed.

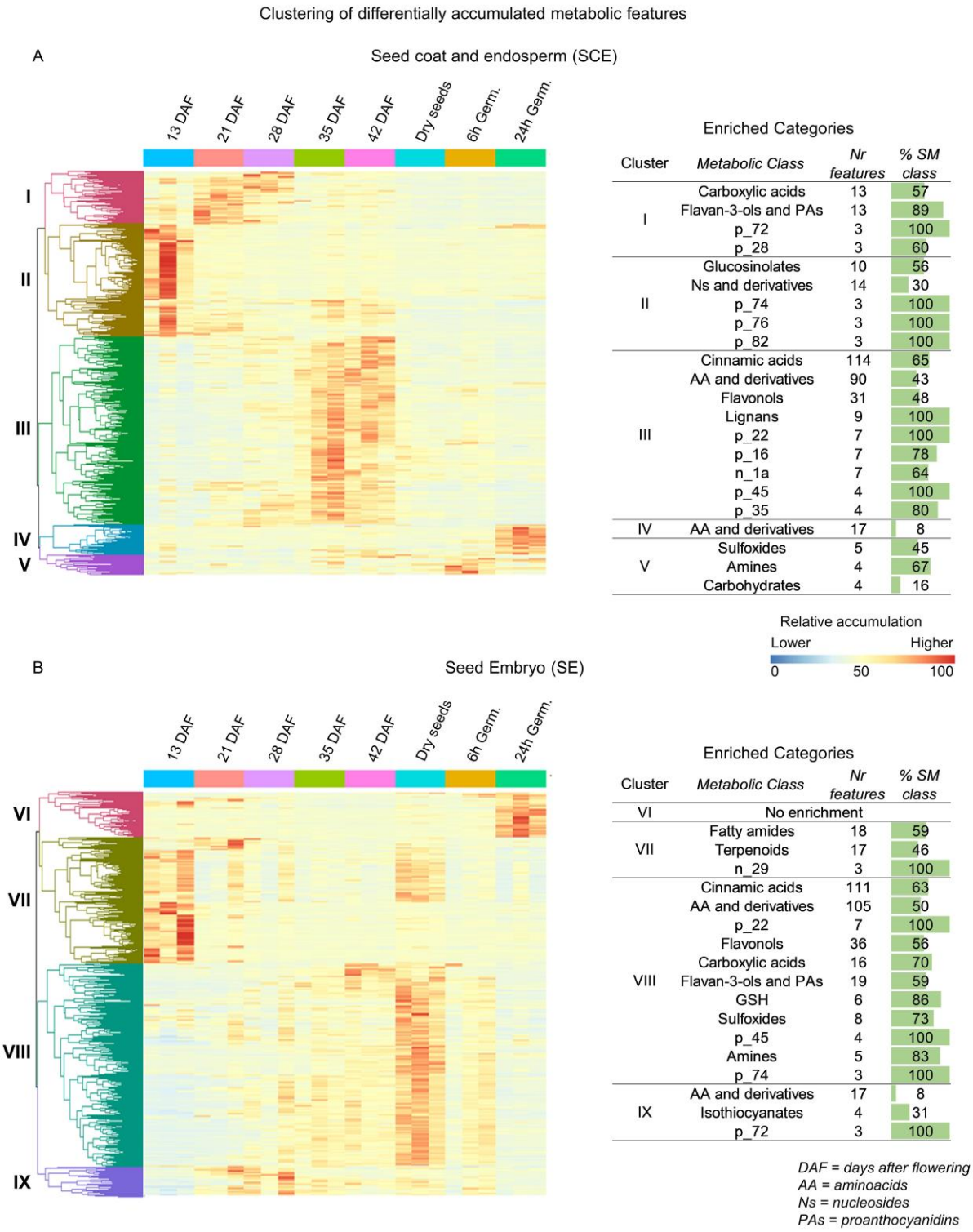


Figure 4. Proteomic analyses in developing SCE and SE

A) Principal component analyses using the proteomic data of SCE and SE harvested at several developmental stages and during germination. DAF = Days After Flowering. **B)** Histogram indicating the number of Differentially accumulated proteins (DAPs) according to the developmental stage in seed coat and embryo.

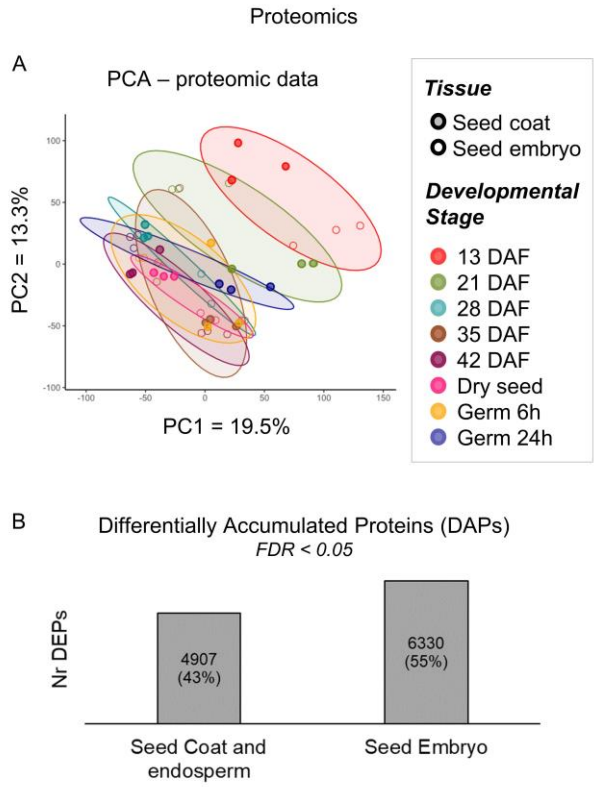


Figure 5. Accumulation pattern of specialized metabolites -related proteins

A) Hierarchical clustering of SM-related DAPs in SCE and SE. **B)** Expression behaviour of proteins involved in the biosynthesis, transport and degradation of the major specialized metabolite pathways, and of the main SM-decoration/modification categories. GT1 = glycosyltransferases, group 1.

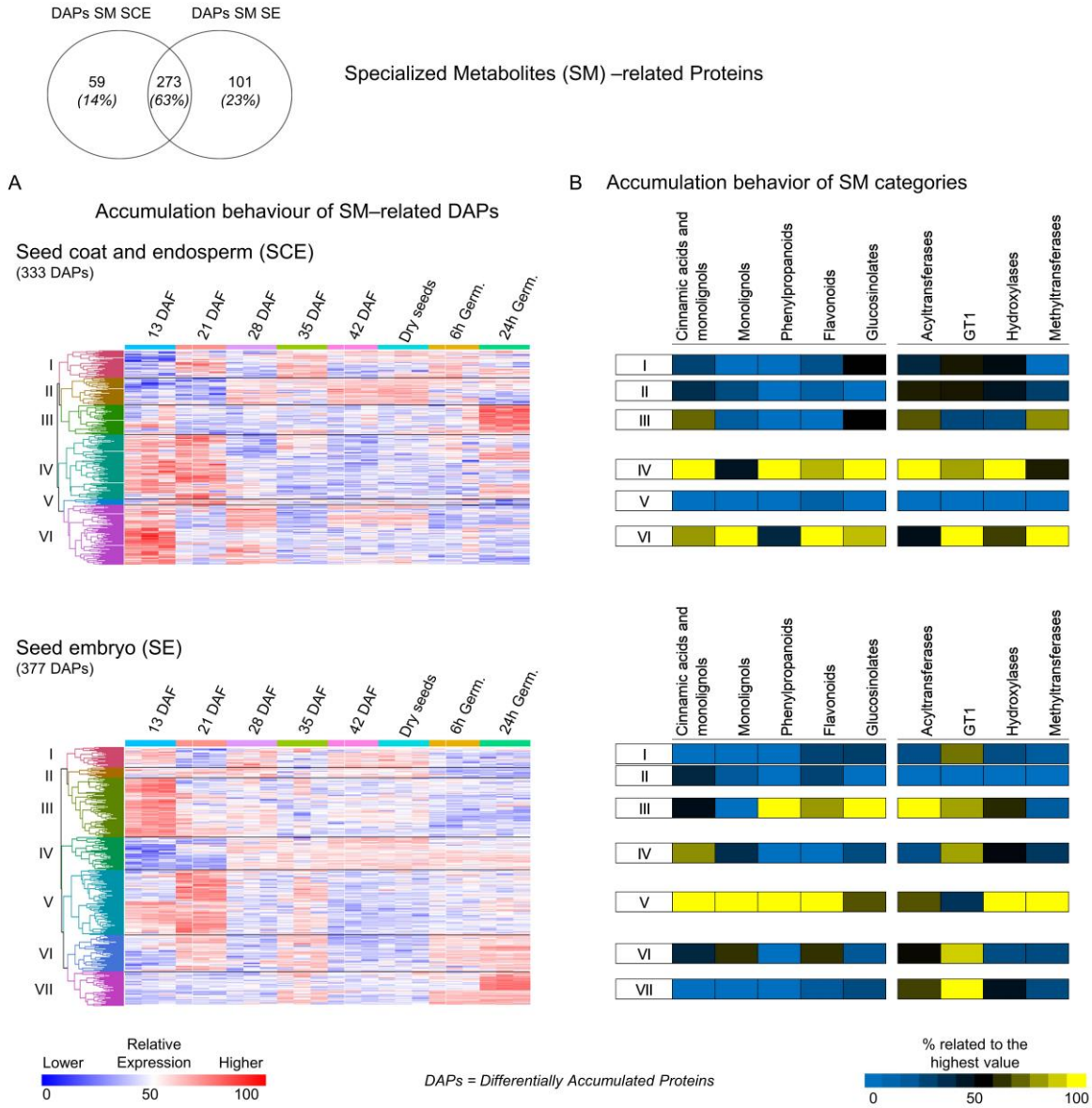


Figure 6. Expression of proteins involved in glucosinolate pathways

Expression of proteins involved in GSL biosynthesis, side-chain elongation and modifications, and degradation in camelina SCE and SE. Only differentially accumulated proteins (DAPs, ANOVA, FDR < 0.05) are showed in the figure.

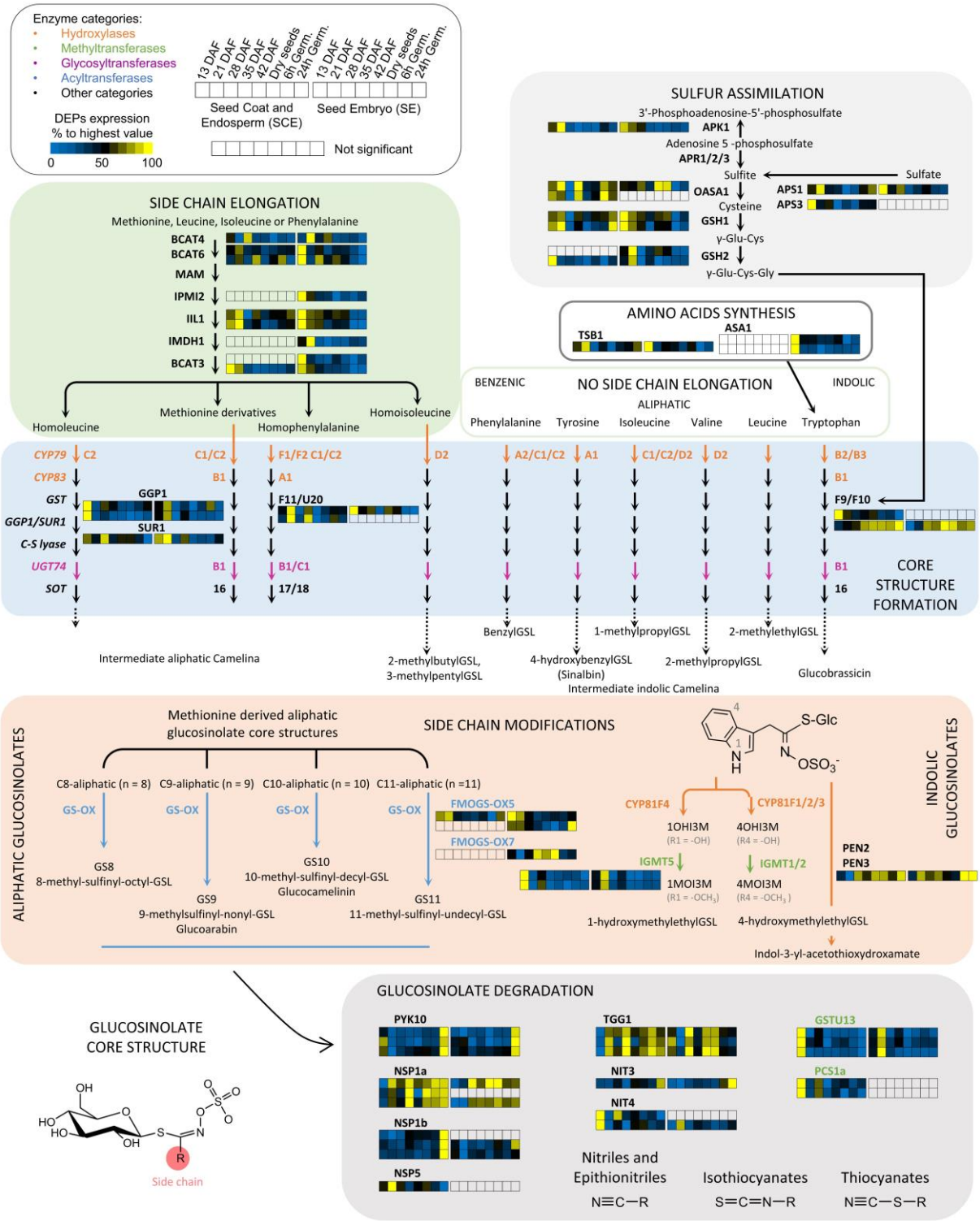


Figure 7. Accumulation profiles of glucosinolates and isothiocyanates metabolites in SCE and SE **A)** Heatmap showing the glucosinolates and isothiocyanates degradation products that are differentially accumulated during seed development in SCE and SE tissues. **B)** Histograms showing the major glucosinolates and isothiocyanates specialized metabolites accumulated in SCE and/or SE.

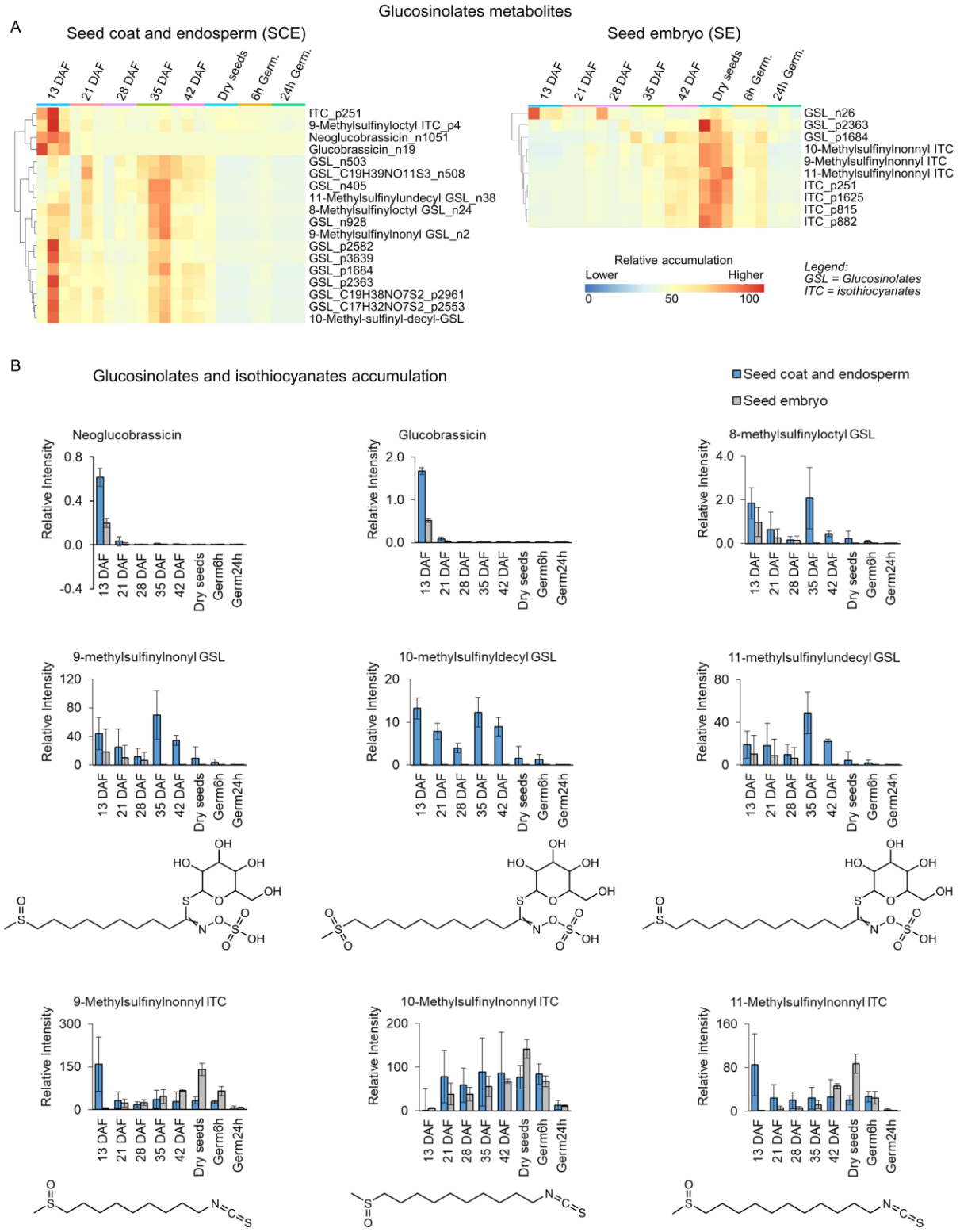


Figure 8. Expression of proteins involved in flavonoid pathways

Expression of proteins involved in flavonoid biosynthesis, degradation, transport and glycosylation in camelina SCE and SE. Only differentially accumulated proteins (DAPs, ANOVA, FDR < 0.05) are showed in the figure.

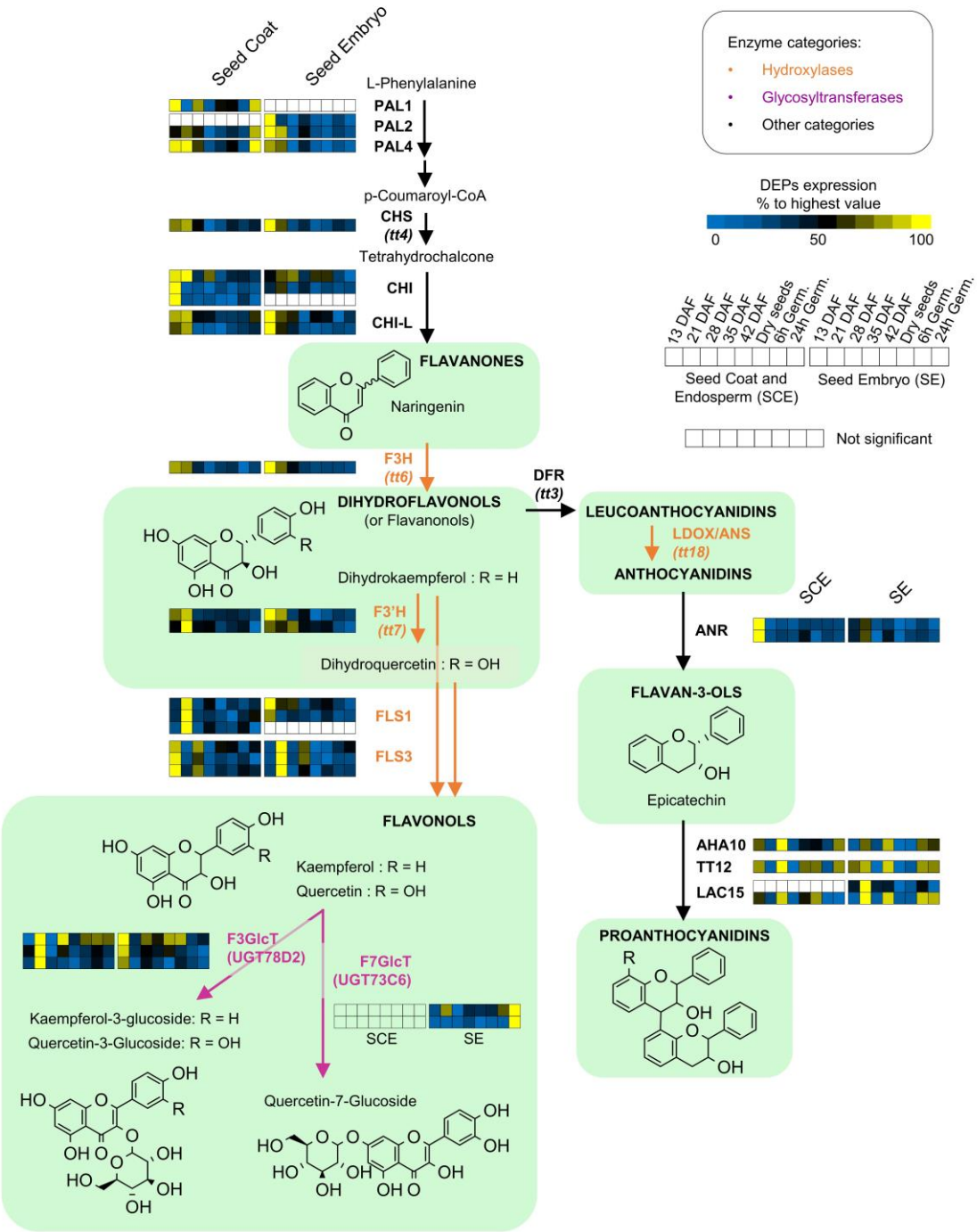


Figure 9. Accumulation profiles of flavonoids metabolites in SCE and SE

Heatmap showing flavonoid metabolites that are differentially accumulated during seed development in SCE and SE tissues.

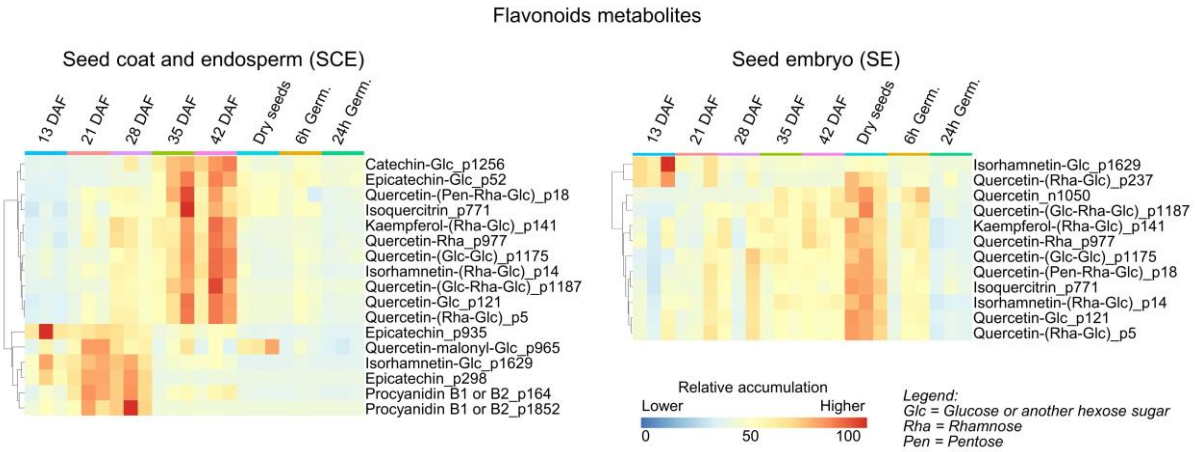
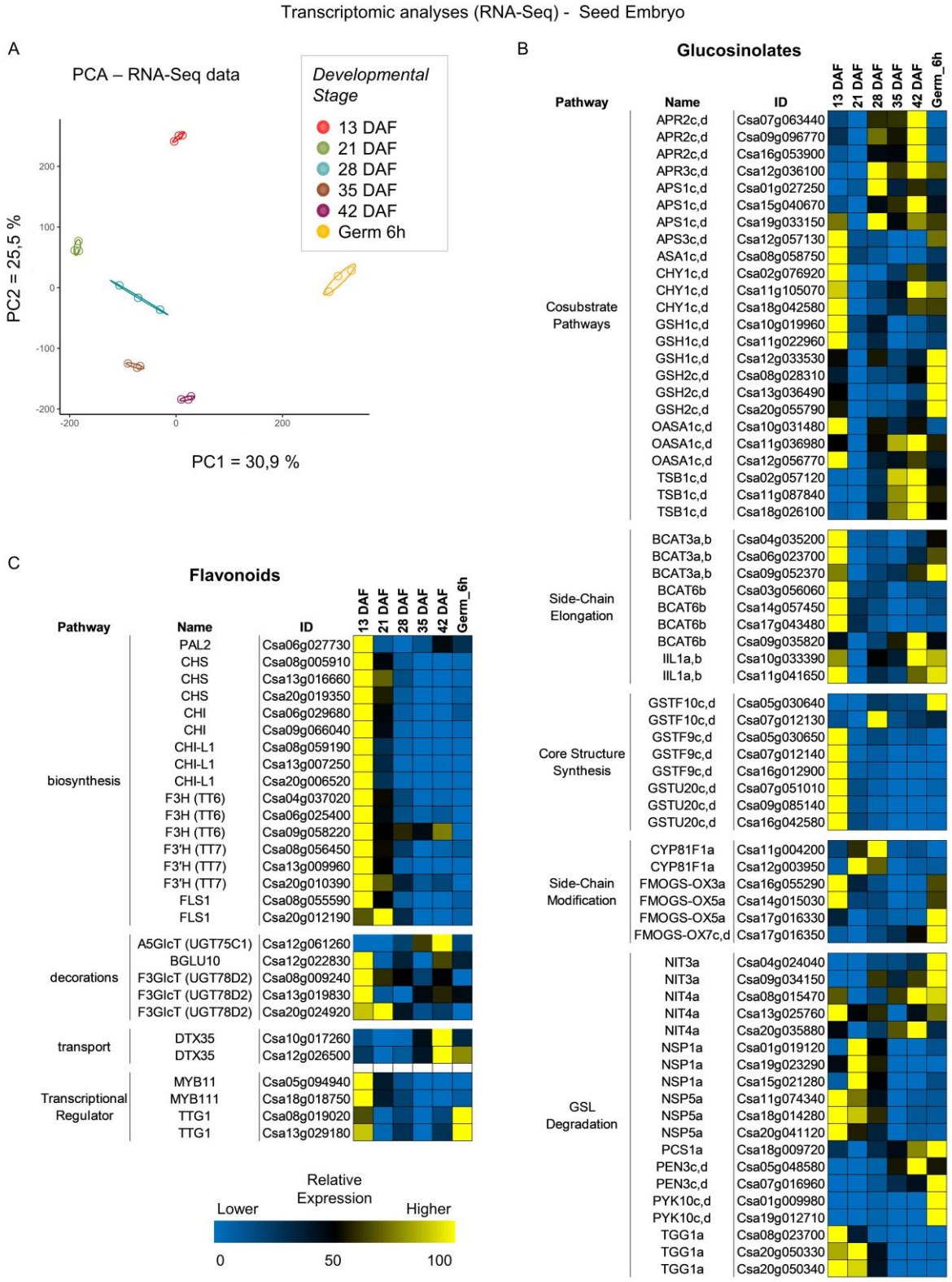


Figure 10. RNA-Seq transcriptomic analyses and expression of flavonoid and GSL-related genes **A)** Principal component analyses using the RNA-Seq transcriptomic data of SCE and SE harvested at several developmental stages and during germination. DAF = Days After Flowering. **B)** Heatmap showing the expression behaviour of GSL and flavonoid –related genes during seed development in SCE and SE tissues.



695

SUPPORTING INFORMATION

Table S1. Top-10 most accumulated specialized metabolites in *C. sativa* seed coats and endosperms, and in the embryos. Average intensity of each metabolic feature is expressed as relative intensity according to the internal standard Apigenin.

Table S2. *Camelina sativa* seed specialized metabolome (LC-MS/MS)

(a) Annotation and information on Camelina sativa seed coat/endosperm and embryo specialized metabolites. Identification (ID), ionization mode (mode), m/z, retention time (RT), Molecular Network clusters, Annotation (based on homemade library, standards and public databases), metabolite class (based on molecular network clusters and annotation), and source of the annotation are indicated for each metabolite. n = metabolite or cluster from negative ionization mode, p = metabolite or cluster from positive ionization mode.

(b) Relative accumulation of Camelina sativa seed coat and embryo specialized metabolites (LC-MS/MS). Data are showed for two tissues, eight developmental stages, and three biological replicates considered in this study. The QCs (quality control) were obtained by mixing all the samples considered in the analyses. SCE = seed coat and endosperm, SE = seed embryo. DAF = days after flowering.

Table S3. *Camelina sativa* seed proteome (DIA)

Relative accumulation of *Camelina sativa* seed coat/endosperm and embryo proteins (DIA, TIMS-TOF). Data are showed for two tissues, eight developmental stages, and three biological replicates considered in this study. Corresponding *A. thaliana* (At ID) and protein description are showed when available. SCE = seed coat and endosperm, SE = seed embryo. DAF = days after flowering

Table S4. *Camelina sativa* seed transcriptome (RNA-Seq)

Expression of Camelina sativa seed coat/endosperm and embryo genes (RNA-Seq). Data are showed for embryo tissue, six developmental stages, and three biological replicates considered in this study. Corresponding *A. thaliana* (At ID) and gene description are showed when available. SE = seed embryo. DAF = days after flowering.

Figure S1. Accumulation profiles of Differentially accumulated proteins (DAPs)

Heatmaps showing DAPs that are differentially accumulated during seed development in SCE and SE tissues.

Figure S2. Expression of proteins involved in cinnamic acids and monolignols pathways

Expression of proteins involved in cinnamic acids and monolignols biosynthesis and decorations in camelina SCE and SE. Only differentially accumulated proteins (DAPs, ANOVA, FDR < 0.05) are showed in the figure.

Figure S3. Accumulation profiles of cinnamic acids and monolignols metabolites in SCE and SE

Heatmap showing cinnamic acids and monolignols metabolites that are differentially accumulated during seed development in SCE and SE tissues.

Figure S4. Expression of SM-related homeologs proteins

The number of copies expressed for proteins differentially expressed (FDR < 0.05) in SE and SCE and involved in GSLs, flavonoids, and cinnamic acids and monolignols pathways.

Figure S5. Differentially Expressed Genes (DEGs) in the embryo (SE) during seed development

Figure S6. Expression of cinnamic acids and monolignols -related genes

Heatmap showing the expression behaviour of cinnamic acids and monolignols –related genes during seed development in SCE and SE tissues.

REFERENCES

Alberghini B, Zanetti F, Corso M, Boutet S, Lepiniec L, Vecchi A, Monti A (2022) Camelina [Camelina sativa (L .) Crantz] seeds as a multi-purpose feedstock for bio-based applications. Ind Crop Prod **182**: 114944

Amyot L, McDowell T, Martin SL, Renaud J, Gruber MY, Hannoufa A (2019) Assessment of Antinutritional Compounds and Chemotaxonomic Relationships between Camelina sativa and Its Wild Relatives. J Agric Food Chem **67**: 796–806

Arimura G, Maffei M (2016) Plant Specialized Metabolism: Genomics, Biochemistry, and Biological Functions, Taylor and. Taylor and Francis, New York, NY

Barreda L, Brosse C, Boutet S, Perreau F, Rajjou L, Lepiniec L, Corso M (2024a) Specialized metabolite modifications in Brassicaceae seeds and plants: diversity, functions and related enzymes. Nat Prod Rep **41**: 834–859

Barreda L, Brosse C, Boutet S, Perreau F, Rajjou L, Lepiniec L, Corso M (2024b) Specialized metabolite modifications in Brassicaceae seeds and plants: diversity, functions and related enzymes. Nat Prod Rep DOI: 10.1039/d3np00043e

Baud S, Corso M, Debeaujon I, Dubreucq B, Job D, Marion-Poll A, Miquel M, North H, Rajjou L, Lepiniec L (2023) Recent progress in molecular genetics and omics-driven research in seed biology. Comptes Rendus - Biol **345**: 61–110

Baumert A, Milkowski C, Schmidt J, Nimtz M, Wray V, Strack D (2005) Formation of a complex pattern of sinapate esters in Brassica napus seeds, catalyzed by enzymes of a serine carboxypeptidase-like acyltransferase family? Phytochemistry **66**: 1334–1345

Bell JM (1993) Factors affecting the nutritional value of canola meal: A review. Can J Anim Sci **73**: 689–697

Berhow MA, Polat U, Glinski JA, Glensk M, Vaughn SF, Isbell T, Ayala-Diaz I, Marek L, Gardner C (2013) Optimized analysis and quantification of glucosinolates from Camelina sativa seeds by reverse-phase liquid chromatography. Ind Crops Prod **43**: 119– 125

Boisson M, Gomord V, Audran C, Berger N, Dubreucq B, Granier F, Lerouge P, Faye L, Caboche M, Lepiniec L (2001) Arabidopsis glucosidase I mutants reveal a critical role of N-glycan trimming in seed development. EMBO J **20**: 1010–1019

Boutet S, Barreda L, Perreau F, Totozafy J-C, Mauve C, Gakière B, Delannoy E, Martin-Magniette M-L, Monti A, Lepiniec L, et al (2022) Untargeted metabolomic analyses reveal the diversity and plasticity of the specialized metabolome in seeds of different Camelina sativa genotypes. Plant J **110**: 147–165

Brown PD, Tokuhisa JG, Reichelt M, Gershenzon J (2003) Variation of glucosinolate accumulation among different organs and developmental stages of Arabidopsis thaliana. Phytochemistry **62**: 471–481

Bussell JD, Reichelt M, Wiszniewski AAG, Gershenzon J, Smith SM (2014) Peroxisomal ATP-binding cassette transporter COMATOSE and the multifunctional protein ABNORMAL INFLORESCENCE MERISTEM are required for the production of benzoylated metabolites in arabidopsis seeds. Plant Physiol **164**: 48–54

Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, et al (2018) SOAPnuke: A MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience **7**: 1–6

Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM (2009) The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Res **38**: 1767–1771

Corso M, Perreau F, Mouille G, Lepiniec L (2020) Specialized Phenolic compounds in seeds: structures, functions, and regulations. Plant Sci **296**: 110471

Corso M, Perreau F, Rajjou L, Ben Malek R, Lepiniec L, Mouille G (2021) Specialized metabolites in seeds. Adv. Bot. Res. pp 35–70

Czerniawski P, Piasecka A, Bednarek P (2021) Evolutionary changes in the glucosinolate biosynthetic capacity in species representing Capsella, Camelina and Neslia genera. Phytochemistry **181**: 112571

Dalling JW, Davis AS, Arnold AE, Sarmiento C, Zalamea PC (2020) Extending Plant Defense Theory to Seeds. Annu Rev Ecol Evol Syst **51**: 123–141

Debeaujon I, Nesi N, Perez P, Devic M, Grandjean O, Caboche M, Lepiniec L (2003) Proanthocyanidin-Accumulating Cells in Arabidopsis Testa: Regulation of Differentiation and Role in Seed Development. Plant Cell **15**: 2514–2531

Demichev V, Messner CB, Vernardis SI, Lilley KS, Ralser M (2020) DIA-NN: neural networks and interference correction enable deep proteome coverage in high throughput. Nat Methods **17**: 41–44

Fang J, Reichelt M, Hidalgo W, Agnolet S, Schneider B (2012) Tissue-Specific Distribution of Secondary Metabolites in

Rapeseed (Brassica napus L.). PLoS One. doi: 10.1371/journal.pone.0048006

Faure J-D, Tepfer M (2016) Camelina, a Swiss knife for plant lipid biotechnology. OCL **23**: D503

Fraser CM, Thompson MG, Shirley AM, Ralph J, Schoenherr JA, Sinlapadech T, Hall MC, Chapple C (2007) Related arabidopsis serine carboxypeptidase-like sinapoylglucose acyltransferases display distinct but overlapping substrate specificities. Plant Physiol **144**: 1986–1999

Di Gioia F, Pinela J, Bailón ADH, Fereira ICFR, Petropoulos SA (2019) The dilemma of "good" and "bad" glucosinolates and the potential to regulate their content. Glucosinolates Prop Recover Appl. doi: 10.1016/B978-0-12-816493-8.00001-9

Halkier BA, Gershenzon J (2006) Biology and biochemistry of glucosinolates. Annu Rev Plant Biol **57**: 303–333

Hölzl G, Rezaeva BR, Kumlehn J, Dörmann P (2023) Ablation of glucosinolate accumulation in the oil crop Camelina sativa by targeted mutagenesis of genes encoding the transporters GTR1 and GTR2 and regulators of biosynthesis MYB28 and MYB29. Plant Biotechnol J **21**: 189–201

Huang J, Rozwadowski K, Bhinu VS, Schäfer U, Hannoufa A (2008) Manipulation of sinapine, choline and betaine accumulation in Arabidopsis seed: Towards improving the nutritional value of the meal and enhancing the seedling performance under environmental stresses in oilseed crops. Plant Physiol Biochem **46**: 647–654

Kagale S, Koh C, Nixon J, Bollina V, Clarke WE, Tuteja R, Spillane C, Robinson SJ, Links MG, Clarke C, et al (2014) The emerging biofuel crop Camelina sativa retains a highly undifferentiated hexaploid genome structure. Nat Commun **5**: 3706

Kliebenstein DJ, D'Auria JC, Behere AS, Kim JH, Gunderson KL, Breen JN, Lee G, Gershenzon J, Last RL, Jander G (2007) Characterization of seed-specific benzoyloxyglucosinolate mutations in Arabidopsis thaliana. Plant J **51**: 1062–1076

Kliebenstein DJ, Lambrix VM, Reichelt M, Gershenzon J, Mitchell-Olds T (2001) Gene duplication in the diversification of secondary metabolism: Tandem 2-oxoglutarate-dependent dioxygenases control glucosinolate biosynthesis in arabidopsis. Plant Cell **13**: 681–693

Lee S, Kaminaga Y, Cooper B, Pichersky E, Dudareva N, Chapple C (2012) Benzoylation and sinapoylation of glucosinolate R-groups in Arabidopsis. Plant J **72**: 411–422

Lepiniec L, Debeaujon I, Routaboul J-M, Baudry A, Pourcel L, Nesi N, Caboche M (2006) Genetics and Biochemistry of Seed Flavonoids. Annu Rev Plant Biol **57**: 405–430

Li X, Bergelson J, Chapple C (2010) The ARABIDOPSIS accession pna-10 is a naturally occurring sng1 deletion mutant. Mol Plant **3**: 91–100

Mácová K, Prabhullachandran U, Štefková M, Spyroglou I, Pěnčík A, Endlová L, Novák O, Robert HS (2022) Long-Term High-Temperature Stress Impacts on Embryo and Seed Development in Brassica napus. Front Plant Sci **13**: 844292

Mandáková T, Pouch M, Brock JR, Al-Shehbaz IA, Lysak MA (2020) Origin and Evolution of Diploid and Allopolyploid Camelina Genomes Were Accompanied by Chromosome Shattering. Plant Cell **31**: 2596–2612

Mann A, Kumari J, Kumar R, Kumar P, Pradhan AK, Pental D, Bisht NC (2023) Targeted editing of multiple homologues of GTR1 and GTR2 genes provides the ideal low-seed, high-leaf glucosinolate oilseed mustard with uncompromised defence and yield. Plant Biotechnol J 1–14

Menard GN, Langdon M, Bhunia RK, Shankhapal AR, Noleto-Dias C, Lomax C, Ward JL, Kurup S, Eastmond PJ (2022) Diverting phenylpropanoid pathway flux from sinapine to produce industrially useful 4-vinyl derivatives of hydroxycinnamic acids in Brassicaceous oilseeds. Metab Eng **70**: 196–205

Neupane D, Lohaus RH, Solomon JKQ (2022) Realizing the Potential of Camelina sativa as a Bioenergy Crop for a Changing Global Climate. Plants **11**: 772

Nour-Eldin HH, Halkier BA (2009) Piecing together the transport pathway of aliphatic glucosinolates. Phytochem Rev **8**: 53–67

Nour-Eldin HH, Madsen SR, Engelen S, Jørgensen ME, Olsen CE, Andersen JS, Seynnaeve D, Verhoye T, Fulawka R, Denolf P, et al (2017) Reduction of antinutritional glucosinolates in *Brassica* oilseeds by mutation of genes encoding transporters. Nat Biotechnol **35**: 377–382

Olivon F, Elie N, Grelier G, Roussi F, Litaudon M, Touboul D (2018) MetGem Software for the Generation of Molecular Networks Based on the t-SNE Algorithm. Anal Chem **90**: 13900–13908

Pang Z, Chong J, Zhou G, Anderson De Lima Morais D, Chang L, Barrette M, Gauthier C, Jacques P-´ E, Li S, Xia J

(2021) MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. Nucleic Acids Res. doi: 10.1093/nar/gkab382

Pang Z, Zhou G, Ewald J, Chang L, Hacariz O, Basu N, Xia J (2022) Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data. Nat Protoc **17**: 1735–1761

Quéro A, Molinié R, Mathiron D, Thiombiano B, Fontaine JX, Brancourt D, Van Wuytswinkel O, Petit E, Demailly H, Mongelard G, et al (2016) Metabolite profiling of developing Camelina sativa seeds. Metabolomics **12**: 1–14

Raab T, Barron D, Arce Vera F, Crespy V, Oliveira M, Williamson G (2010) Catechin Glucosides: Occurrence, synthesis, and stability. J Agric Food Chem **58**: 2138–2149

Rao SQ, Chen XQ, Wang KH, Zhu ZJ, Yang J, Zhu B (2021) Effect of short-term high temperature on the accumulation of glucosinolates in Brassica rapa. Plant Physiol Biochem **161**: 222–233

Reichelt M, Brown PD, Schneider B, Oldham NJ, Stauber E, Tokuhisa J, Kliebenstein DJ, Mitchell-Olds T, Gershenzon J (2002) Benzoic acid glucosinolate esters and other glucosinolates from Arabidopsis thaliana. Phytochemistry **59**: 663–671

Robinson MD, Mccarthy DJ, Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma Appl NOTE **26**: 139–140

Routaboul J-M, Dubos C, Beck G, Marquis C, Bidzinski P, Loudet O, Lepiniec L (2012) Metabolite profiling and quantitative genetics of natural variation for flavonoids in Arabidopsis. J Exp Bot **63**: 3749–3764

Routaboul JM, Kerhoas L, Debeaujon I, Pourcel L, Caboche M, Einhorn J, Lepiniec L (2006) Flavonoid diversity and biosynthesis in seed of *Arabidopsis thaliana*. Planta **224**: 96–107

Sakai K, Taconnat L, Borrega N, Yansouni J, Brunaud V, Paysant-Le Roux C, Delannoy E, Martin Magniette M-L, Lepiniec L, Denis Faure J, et al (2018) Combining laser-assisted microdissection (LAM) and RNA-seq allows to perform a comprehensive transcriptomic analysis of epidermal cells of Arabidopsis embryo. Plant Methods **14**: 10

Sanden NCH, Kanstrup C, Crocoll C, Schulz A, Nour-Eldin HH, Halkier BA, Xu D (2023) An UMAMIT-GTR transporter cascade controls glucosinolate seed loading in Arabidopsis. Nat Plants doi: 10.1038/s41477-023-01598-4

Stehle F, Brandt W, Schmidt J, Milkowski C, Strack D (2008a) Activities of Arabidopsis sinapoylglucose:malate sinapoyltransferase shed light on functional diversification of serine carboxypeptidase-like acyltransferases. Phytochemistry **69**: 1826–1831

Stehle F, Stubbs MT, Strack D, Milkowski C (2008b) Heterologous expression of a serine carboxypeptidase-like acyltransferase and characterization of the kinetic mechanism. FEBS J **275**: 775–787

Traka MH (2016) Health Benefits of Glucosinolates. Adv Bot Res **80**: 247–279

Xie D-Y, Sharma SB, Paiva NL, Ferreira D, Dixon RA (2003) Role of anthocyanidin reductase, encoded by BANYULS in plant flavonoid biosynthesis. Science (80-) **299**: 396–399

Xu D, Christian N, Sanden H, Hansen LL, Belew ZM, Madsen SR, Meyer L, Jørgensen ME, Hunziker P, Veres D, et al (2023) Export of defensive glucosinolates is key for their accumulation in seeds. Nature **617**: 132–138

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

Zanetti F, Alberghini B, Marjanovi A, Grahovac N, Rajkovi D, Kiprovski B, Monti A (2021) Camelina, an ancient oilseed crop actively contributing to the rural renaissance in Europe. A review. Agron Sustain Dev **41**: 2

Zanetti F, Eynck C, Christou M, Krzyżaniak M, Righini D, Alexopoulou E, Stolarski MJ, Van Loo EN, Puttick D, Monti A (2017) Agronomic performance and seed quality attributes of Camelina (Camelina sativa L. crantz) in multi-environment trials across Europe and Canada. Ind Crops Prod **107**: 602–608

Zerbib M, Cazals G, Enjalbal C, Saucier C (2018) Identification and Quantification of Flavanol Glycosides in Vitis vinifera Grape Seeds and Skins during Ripening. Molecules. doi: 10.3390/molecules23112745

Zhang S, Ghatak A, Bazargani MM, Bajaj P, Varshney RK, Chaturvedi P, Jiang D, Weckwerth W (2021) Spatial distribution of proteins and metabolites in developing wheat grain and their differential regulatory response during the grain filling process. Plant J **107**: 669–687

Zubr J (1997) Oil-seed crop: Camelina sativa. Ind Crops Prod **6**: 113–119