


RESOURCE

Domestication has altered gene expression and secondary metabolites in pea seed coat

Barbora Klčová^{1,†}, Jana Balarynová^{1,†}, Oldřich Trněný², Petra Krejčí³, Monika Zajacová Cechová³, Tatiana Leonova⁴, Daria Gorbach⁴, Nadezhda Frolova⁵, Elana Kysil⁴, Anastasia Orlova⁵, Christian Ihling⁶, Andrej Frolov⁵, Petr Bednár³ and Petr Smýkal^{1,*} 

¹Department of Botany, Faculty of Sciences, Palacky University, Šlechtitelů 27, Olomouc 773 71, Czech Republic,

²Agricultural Research Ltd., Zemědělská 1, Troubsko 664 41, Czech Republic,

³Department of Analytical Chemistry, Faculty of Sciences, Palacky University, 17. listopadu 1192/12, Olomouc 771 46, Czech Republic,

⁴Department of Bioorganic Chemistry, Leibniz-Institut für Pflanzenbiochemie, Weinberg 3, Halle (Saale) 06120, Germany,

⁵Laboratory of Analytical Biochemistry, Timiryazev Institute of Plant Physiology, Botanicheskaja 36, Moscow 127276, Russia, and

⁶Department of Pharmaceutical Chemistry and Bioanalytics, Institute of Pharmacy, Martin-Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3, Halle (Saale) 06120, Germany

Received 24 November 2023; accepted 9 March 2024; published online 5 April 2024.

*For correspondence (e-mail petr.smykal@upol.cz), <https://www.pf.upol.cz/bot/>.

[†]These authors contributed equally.

SUMMARY

The mature seed in legumes consists of an embryo and seed coat. In contrast to knowledge about the embryo, we know relatively little about the seed coat. We analyzed the gene expression during seed development using a panel of cultivated and wild pea genotypes. Gene co-expression analysis identified gene modules related to seed development, dormancy, and domestication. Oxidoreductase genes were found to be important components of developmental and domestication processes. Proteomic and metabolomic analysis revealed that domestication favored proteins involved in photosynthesis and protein metabolism at the expense of seed defense. Seed coats of wild peas were rich in cell wall-bound metabolites and the protective compounds predominated in their seed coats. Altogether, we have shown that domestication altered pea seed development and modified (mostly reduced) the transcripts along with the protein and metabolite composition of the seed coat, especially the content of the compounds involved in defense. We investigated dynamic profiles of selected identified phenolic and flavonoid metabolites across seed development. These compounds usually deteriorated the palatability and processing of the seeds. Our findings further provide resources to study secondary metabolism and strategies for improving the quality of legume seeds which comprise an important part of the human protein diet.

Keywords: domestication, gene expression, metabolomics, phenylpropanoid pathway, pea, seed proteomics, seed coat.

INTRODUCTION

The mature seed of an angiosperm consists of an embryo (E), a seed coat (SC), and in many cases, an endosperm (ES) (Baroux & Grossniklaus, 2019). Seed development involves the coordinated activities of these three genetically distinct entities. While E and ES are products of fertilization and combine paternal and maternal genotypes, SC is derived solely from maternal tissue. The E represents

the next generation, the ES is nourishing tissue, and SC has protective and dispersal functions. Although the morphology of seeds varies among different plant families and is used taxonomically, the fundamental components of the E (cotyledons, hypocotyl, and radicle) are highly conserved. Much of the information about the genetic regulation of seed development has been gained from the study of Arabidopsis mutants (Bradford & Nonogaki, 2009) and

related *Brassicacae* (Gao et al., 2022; MacGregor et al., 2015), tomato (Bizouerne et al., 2021), *Medicago truncatula* (Buitink et al., 2006; Fu et al., 2017; Gallardo et al., 2007; Verdier, Dessaint, et al., 2013; Verdier, Lalanne, et al., 2013), and other economically important legumes such as lentil (Yu et al., 2023), chickpea (Garg et al., 2017), and soybean (Gao et al., 2018) and pea (Balarynová et al., 2022, 2023; Cechová et al., 2017; Hradilová et al., 2017; Janská et al., 2019; Krejčí et al., 2022; Zablazková et al., 2021). The development of SC is important for the establishment of viable seeds; perturbation results in a decrease in seed viability and germination, essential for species survival in nature as well as in crop establishment.

Legumes represent one of the largest and most diverse families of flowering plants, with approximately 20 000 classified species (Lewis et al., 2005). Legume seed biology has been studied for more than 150 years using descriptive, physiological, biochemical, molecular, and genetic approaches (Le et al., 2007). Despite the importance of legumes in human nutrition, however, there is limited information on their seed development and particularly SC development in relation to domestication (except for economically important species such as soybeans). Most species of the legume family, including pea, have non-endospermic seeds with cotyledons serving as storage organs. In the mature legume seed, the E is enclosed solely by the SC as the only seed-covering layer. Seed coat development begins after fertilization events inducing the integuments of the ovule to differentiate. Young pea SC serves as a transient storage organ accumulating storage starch and proteins before the E starts its activity and SC serves as a buffering compartment during seed development (Rochat et al., 1995; Weber et al., 2005). Fully developed mature legume seed is composed of the E and the SC, where SC is formed by dead cells, yet metabolically active tissue (Grafí, 2020). The SC can further be divided into sub-regions based on morphological and anatomical features (Fu et al., 2017; Hradilová et al., 2017; Smýkal et al., 2014; Zablazková et al., 2021). During seed filling, SC supports storage compound synthesis by transmitting nutrients from the phloem, mainly sugars and amino acids, through the apoplast (Weber et al., 2005). However, the main role of the SC is to protect the mature seeds from adverse effects of the environment. Moreover, the maternally derived SC protects the E, mediates nutrients and signals from the maternal environment, limits desiccation during dormancy and germination, and facilitates seed dispersal (Haughn & Chaudhury, 2005; Smýkal et al., 2014). Besides a mechanical function, such protection is provided by the rich spectra of secondary metabolites such as flavonoids, impacting also the physiology and in the case of crops agronomic quality of the seed (Corso et al., 2020; Debeaujon et al., 2000; Lepiniec et al., 2006; MacGregor et al., 2015; Winkel-Shirley, 2001). In contrast to knowledge

about the E and ES, we have comparably less knowledge of the SC, especially at gene expression and molecular levels. Although some genes involved in SC metabolism have been identified (Balarynová et al., 2022; Du et al., 2017; Hradilová et al., 2017; Khan et al., 2014), the developmental pathways controlling SC development are not completely elucidated, and global genetic program associated with SC development was described mainly in Arabidopsis model (Verma et al., 2022) or closely related brassicas (Gao et al., 2022), while comparably less is known in legumes (Fu et al., 2017; Hradilová et al., 2017; Verdier, Dessaint, et al., 2013; Verdier, Lalanne, et al., 2013; Yu et al., 2023). Mutant analysis, also used in seed development, is extremely challenging in the case of subtle phenotypes, such as those with minor changes in metabolic processes. Comparative transcriptomics has greatly enhanced our ability to identify differential gene expression both in temporal as well as spatial resolution (Gao et al., 2022; Khan et al., 2014) and to identify gene regulatory networks, including *Medicago* seed dormancy and longevity (Verdier, Dessaint, et al., 2013; Verdier, Lalanne, et al., 2013).

Seeds accumulate a wide range of secondary metabolites, especially phenolics (phenylpropanoids and alkaloids) as a part of the desiccation tolerance (Sano et al., 2016) and chemical defense against abiotic (drought, temperature, toxic metals, and UV irradiation) and biotic (herbivores, biotic pathogens, and plant competitors) stress (Corso et al., 2020; Francoz et al., 2018). The accumulation is spatiotemporally regulated at transcriptional and post-transcriptional levels (Corso et al., 2020; Lepiniec et al., 2006). Flavonoids and lignins have received particular attention for being altered during plant domestication (Alseikh et al., 2021; Ku et al., 2020; Paauw et al., 2019) as the result of human selection. Tannin-rich seeds have often astringent taste and affect food digestibility (as a result of their defensive role to herbivory) and bioavailability. Especially outer integument has a higher amount of these substances, and phenolic oxidation and polymerization contribute to the physical protection of the seeds (Huss & Gierlinger, 2021; Steinbrecher & Leubner-Metzger, 2018). Chemical modifications (i.e., hydroxylation, glycosylation, methylation, and acetylation) contribute to metabolite's diversity and affect their physicochemical properties hence their bioactivity and transport (Corso et al., 2020). These are often environment and developmentally regulated, as in the case of seeds transmitting this information from the mother plant to the progeny (Boutet et al., 2022; Geshnizjani et al., 2019; MacGregor et al., 2015).

Advances in genomics provide an opportunity to assess the effect of domestication at the genome level by the comparison of progenitors and respective crops. In addition, the application of transcriptomics along with associated proteome and metabolome analysis provides the

opportunity to explore another dimension of domestication (Bellucci et al., 2014; Lu et al., 2016; Swanson-Wagner et al., 2012) as shown on soybean seed (Du et al., 2017; Gao et al., 2018; Lu et al., 2016), wheat glume (Zou et al., 2015), and our work on pea seed and pod development (Balarynová et al., 2022, 2023; Hradilová et al., 2017). There is evidence that spatiotemporal regulation of gene expression played an important role in the domestication process (Doebley et al., 2006; Du et al., 2017; Lu et al., 2016). Comparative analysis of wild progenitor and cultivated crop counterparts provides an attractive means to identify underlying genetic networks leading to profound changes occurring in a short evolutionary time (Smýkal et al., 2018).

In this study, we have analyzed the gene expression at the proteome, metabolome, and seed coat composition level during development using a set of cultivated and wild pea genotypes. We report the comprehensive gene expression changes related both to development as well as domestication status. Analysis of seed developmental stages revealed extensive modification of gene expression between wild pea progenitor and cultivated pea crop, resulting in altered proteomic and metabolite profiles of the mature seed coat.

RESULTS

Embryo and seed coat gene expression differ between wild and cultivated pea

To obtain spatial and temporal representation, we performed RNA sequencing analysis of isolated seed tissues (SC and E) at five developmental stages (D1–5). There were three genotypes representing wild (*P. elatius*, JI64, JI261, JI1794), two domesticated (*P. sativum*, JI92 landrace and cv. Cameor) and Ethiopian *P. abyssinicum* (PI358617, PA). Reads from each sample were mapped to the pea reference genome (Cameor V1.0) providing 65 246 gene models, hereafter referred to as genes. A total of 44 751 (68%) of these genes corresponded to genes from reference genome and 22 450 (32%) were newly assembled genes. In total, 147 854 transcripts were observed since 22 179 were multi-transcript genes with approximately 2.3 transcripts per gene. Although the reference genome is based on Cameor, we identified 4231 new genes in Cameor SC sample, 5478 in JI92, 6661 in JI64, and 5430 in JI1794. These newly identified genes can be either assembler noise but surely some of them represent new genes, especially from JI64, JI1794, and JI92 genotypes. The largest part of the Cameor new genes (2278 and 4231, respectively) is common for E and SC tissue, while 1291 and 662 are specific for SC or E, respectively. A total of 503 genes were determined as contamination (largely microbial) based on mapping to the non-plant database. A comparison of SC and E samples at the respective stages has shown a rather constant number of shared genes expressed in both tissues (13 000–15 000 with the highest

number in stage 1). In E, there is expressed comparably lower number (1000–2000) of tissue-specific genes than in SC (3000–4000). To see the dynamics of transcription during wild and cultivated pea seed development, a principal component analysis (PCA) was conducted on E and SC together (Figure 1a) and separately for each tissue (Figure 1b,c). Transcriptional dynamics of SC and E samples were characterized by two principal components (PC), where PC1 (24.62%) divided samples according to a cultivation status and PC2 (14.12%) described a direction of development. The biological replicates for each sample were clustered together as well as samples from the same tissue and the same developmental stage. Transcripts of SC of domesticated Cameor were set aside from other SC samples, while landrace JI92 was closer to wild pea. A significant difference in gene expression dynamics appeared between early and late developmental stages D1, D2, and D3, D4 and D5 in JI64, where the expression was increased 3- to 5-fold and 5- to 10-fold, respectively (Figure 1d). The hierarchical dendrogram showed clustering of E samples, where samples of D1 and D2, and D3–D5 stages of Cameor clustered together. Similarly in PA, stages D1–D4 clustered together and D5 was separated. In the case of other four genotypes (landrace JI92 and wild JI64, JI1794, and JI261) only replicates clustered together (Figure 1e). Samples of PA, stages D2–D5, made one shared cluster. Similarly, there was a cluster of wild genotype, JI1794, where stages D2–D4 clustered together. Other genotypes (landrace JI92, and wild: JI64, JI261) clustered separately (Figure 1e). At D1 stage, SC samples of all genotypes clustered together, while during the development, the clustering was more distinct. Thus, at early stages D1 and D2, wild and domesticated genotypes shared over 500 genes (Figure S1a). In later stages D4 and D5, there were about 150 shared genes. Surprisingly, in stage D3, which corresponds to the point of physiological maturation of pea seed, only 19 genes shared between all genotypes were found. Comparing genes within stages revealed much more shared genes among all three cultivated genotypes (Cameor, JI92, and PA) than among three wild genotypes (JI64, JI1794, and JI261) (Figure S1b). These results point to a change in the functions of SC through seed development, mainly during the time of physiological maturity establishment. As the main aim of the study related to domestication, we focused on the analysis of SC, as this tissue is crucial for the establishment of seed dormancy (Smýkal et al., 2014).

Gene co-expression network reveals differences related to development, dormancy, and domestication

To obtain a view of the transcriptional regulation across the SC development as well as domestication status, we performed an unbiased weighted gene co-expression network analysis (WGCNA) on a set of 26 951 SC-expressed genes (Figure 2a). This resulted in the identification of 20 gene

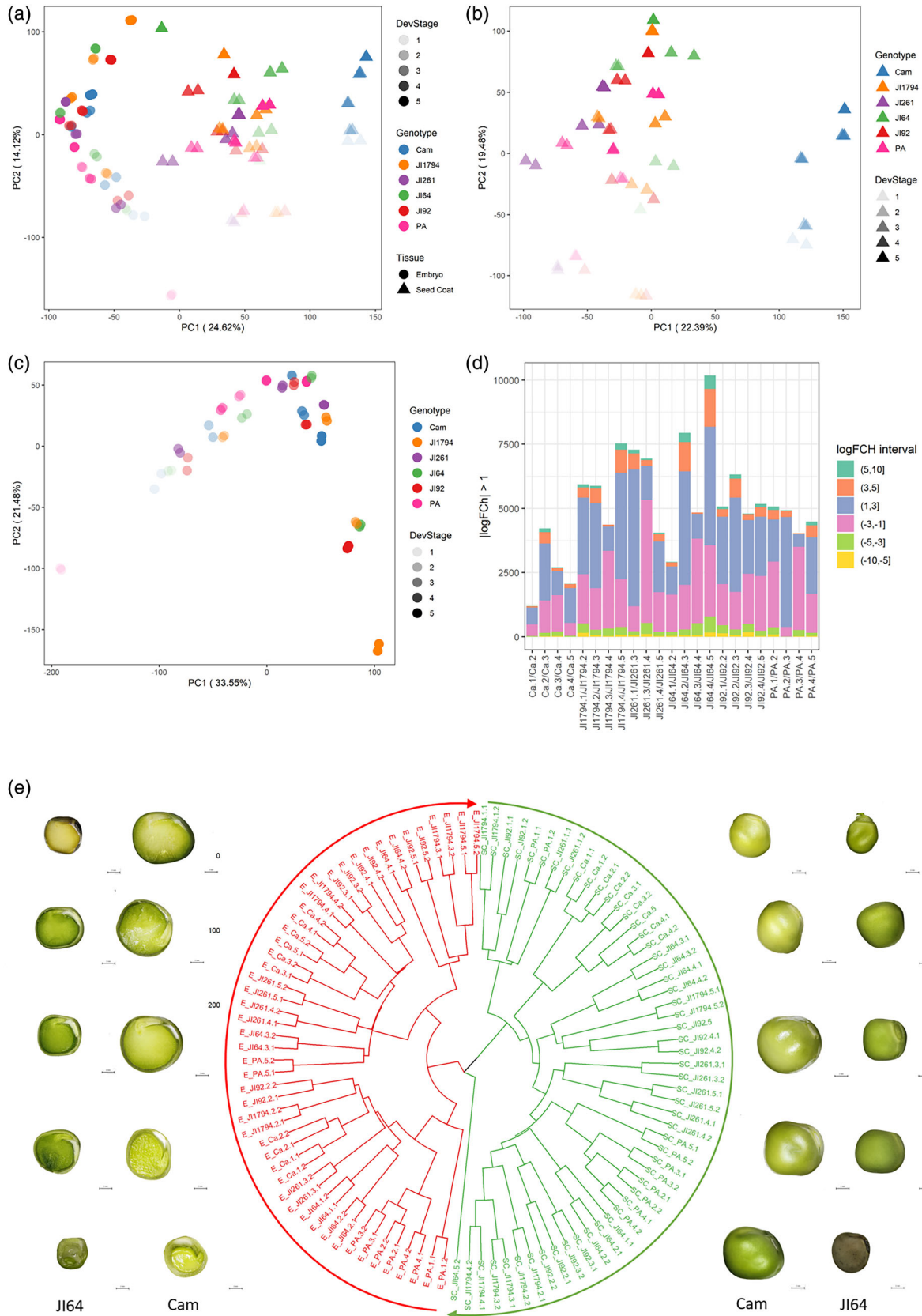


Figure 1. Principal component analysis (PCA) of transcriptome data of seed coat (a), embryo (E) samples (b), and SC and E samples together (c) showing the relationship of three domesticated (Cameor, JI92, and *Pisum abyssinicum* PI358617 = PA) and three wild (JI64, JI1794, and JI261) peas over five seed developmental stages (13, 17, 20, 23, and 28 DAP, labeled as D1-5). Gene expression dynamics of SC samples (d). Hierarchical cluster dendrogram illustrating the relationship between SC and E of studied developmental stages (e).

expression modules (Figure 2b), where each module was correlated with a module eigengene (ME, idealized representative gene to represent the overall expression trend of the module), the number of genes in each module varied from 48 (ME18) to 7442 (ME19) (Figure 2c). Every module had either positive or negative relation to seed traits (dormancy and developmental stages) at a particular level. While most of the modules did not show different expression patterns between wild and cultivated genotypes (Figure 2d) we focused further on four modules, which contained genes differing in the expression level between wild and domesticated genotypes. Of these, ME1-Blue and ME19-Turquoise showed a pattern of developmentally regulated genes up and down during the development, respectively, while ME15-Midnightblue was positively associated with seed dormancy, and ME6-Yellow, was positively associated with domestication. In addition, there was ME11-Black module with genes specific for Ethiopian *P. abyssinicum*. We analyzed the top four overrepresented GO terms in the ME1-Blue module (5377 genes) for the molecular function, which indicated: “oxidoreductase activity,” “transporter activity,” “kinase activity,” and “RNA binding activity” (Table S1a), while in ME19-Turquoise (7442 genes) “oxidoreductase activity,” “RNA binding activity,” “kinase activity,” and “hydrolase activity, acting on acid anhydrides” (Table S1b). GO terms “oxidoreductase activity,” “kinase activity,” “transporter activity,” and “vitamin binding” were also enriched in ME15-Midnightblue module (196 genes) (Table S1c). ME6-Yellow module (1584 genes) was characterized especially by the GO terms “oxidoreductase activity,” “transporter activity,” “RNA binding activity,” and “kinase activity” (Table S1d). The module with genes specific to Ethiopian pea (ME-Black, 802 genes) involved particularly the GO terms “RNA binding,” “lyase activity,” “hydrolase activity, acting on glycosyl bonds,” and “transferase activity, transferring glycosyl groups” (Table S1e).

Within 20 identified gene modules, 692 genes encoding transcription factors (TF) were found, with the most 12 abundant classes, including MYB (158), AP2/ERF (79), C2H2 (52), bZIP (42), WRKY (42), GRAS (36), bHLH (30), WD40 (29), B3 (27), Homeobox (25), HSF (19), and AUX/IAA (13) (Figure S2). Most TFs were identified within ME19-Turquoise and ME1-Blue modules supporting their role in the regulation of seed development.

Flavonoid biosynthesis genes are downregulated in domesticated pea

We analyzed differentially expressed genes (DEGs) between set of cultivated Cameor and wild pea genotypes

(JI64, JI1794, and JI261) labeled as WILD. The 2651 upregulated and 1809 downregulated genes were identified in WILD set at stages D2, D3, and D4. Most DEGs belong to GO category with oxidoreductase activity (257 genes), followed by kinase (183 genes) and RNA-binding activity (176 genes, Table S2). Among these, the genes belonging to ME19-Turquoise (230 genes, module related to development) and ME1-Blue (151 genes, module related to development) expression modules are the most abundant. In the ME19-Turquoise module, the DEG genes encoding leucoanthocyanidin reductase, flavonoid-3'-hydroxylase (TT7), chalcone-flavanone isomerase, or glutathione dehydrogenase were among the most highly expressed genes. Similarly, in the ME19-Turquoise module, DEGs encoded polyphenol oxidase (PPO), lipocalin, pectinesterase, and calmodulin-binding lipid transferase (Table S3a) were the most abundant. Moreover, to filter genetic variability in cultivated genotypes, we compared all wild genotypes against all cultivated ones. Notably, there were no DEGs identified between Cameor and JI92, and Cameor and PA. In this comparison, labeled as DOM, we found 154 and 170 up- and downregulated genes, respectively, characterized especially by GO terms oxidoreductase activity (26 genes), kinase activity (20 genes), and transporter activity (16 genes) (Table S3b). Among the most differentially expressed genes were polyphenol oxidase (PPO), a cytochrome P450 enzyme, or 12-oxophytodienoate reductase (involved in the biosynthesis of jasmonates).

We found chalcone synthase (CHS) genes encoding the key enzyme of flavonoid/isoflavonoid biosynthetic pathway together with chalcone flavanone isomerase (CHI), flavonoid -3'-hydroxylase (F3'H), dihydroflavanol/isoflavone reductase (DFR), flavanone/naringenin-3-dioxygenase (F3H), leucoanthocyanidin reductase (LAR), and anthocyanidin reductase (ANR) to be downregulated in Cameor (Table S4). Moreover, three genes encoding MYB111 and MYB12 transcription factors regulating early steps of flavonoid biosynthesis were also downregulated in Cameor SC samples. MYB12 was downregulated not only in Cameor but also in SC of wild JI64. Furthermore, a homolog of R2R3-MYB TT8 Glabrous 3 gene (*Psat6g060480*) was not expressed in Cameor (Table S4). Peroxidase gene *Psat7g051800*, which is expected to polymerize monolignols into lignin, showed higher expression level in wild pea genotypes. Similarly, PPO gene, already described in pea SC (Balarynová et al., 2022), was significantly more expressed in wild genotypes as well. Finally, PA were compared with domesticated Cameor and wild samples to get PA-specific DEGs. These included especially

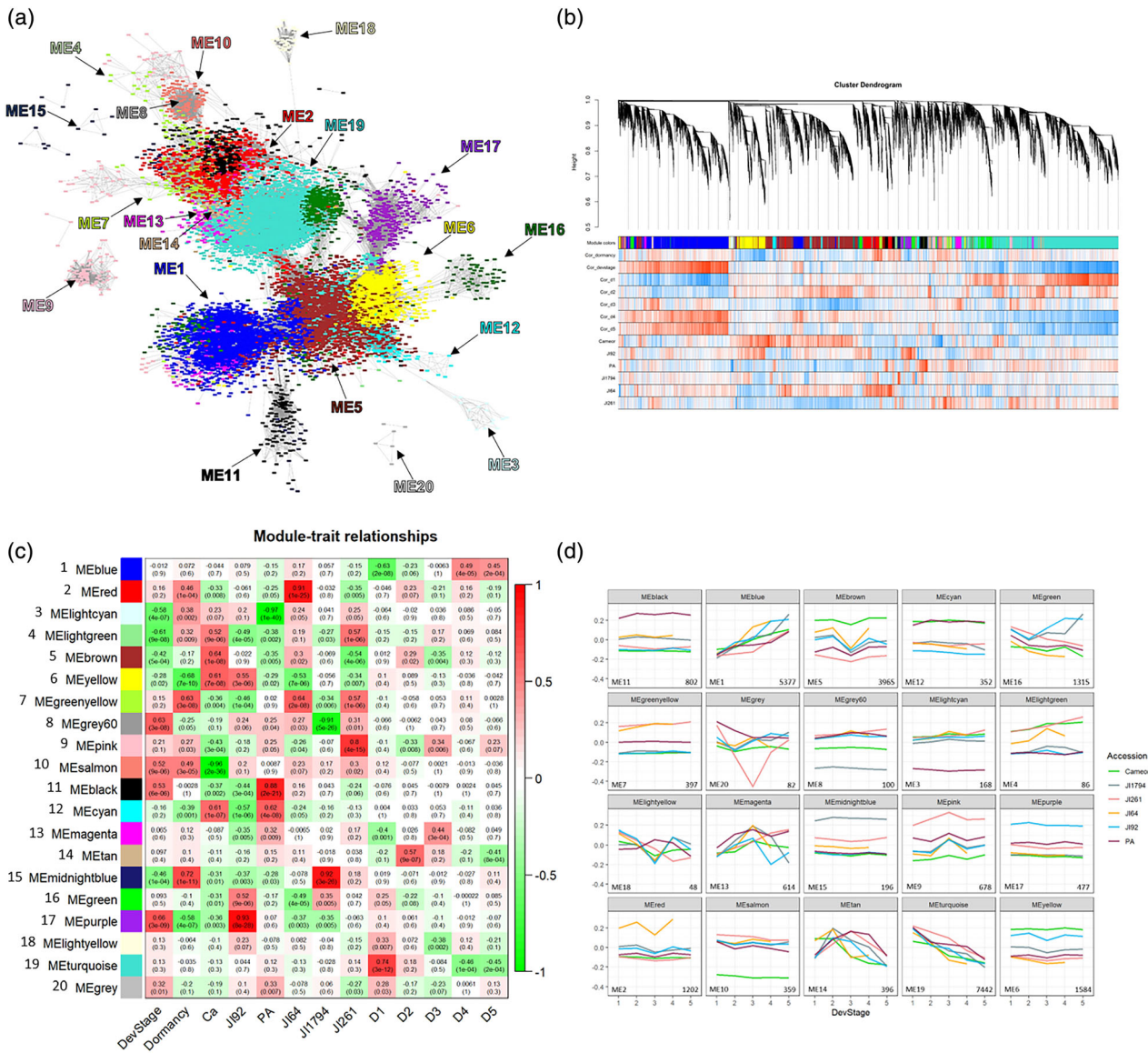


Figure 2. Weighted gene co-expression gene network of seed development of wild (J164, J11794, and J1261) and domesticated (Cameor, J192, and *Pisum abyssinicum*, PA) peas. Spatial resolution of gene modules (a) according to their relationship to particular traits, ME1 – blue, ME2 – red, ME3 – light cyan, ME4 – light green, ME5 – brown, ME6 – yellow, ME7 – green yellow, ME8 – gray 60, ME9 – pink, ME10 – salmon, ME11 – black, ME12 – cyan, ME13 – magenta, ME14 – tan, ME15 – midnight blue, ME16 – green, ME17 – purple, ME18 – light yellow, ME19 – turquoise, ME20 – gray. Module-trait relationship depiction (b). Cluster dendrogram showing hierarchical relationship between samples of SC and E (c). Expression of genes of different modules during pea development (d). The number at the right bottom indicates the number of genes contained in each module.

genes with oxidoreductase, kinase, and RNA-binding activity (Table S3c).

Domestication favored proteins involved in photosynthesis and protein metabolism at the expense of seed defense

As the representatives of domesticated and wild genotypes, we have compared domesticated landrace J192 and wild pea J164 SC samples at three developmental stages (D1, D2, and D6). The total SC protein fraction was

extracted (Table S5) and the correctness of protein determination was cross-validated by SDS-PAGE (Figure S3). Measurement of the total line densities revealed $155\ 825 \pm 11\ 280$ ($1.55 \times 10^5 \pm 0.11 \times 10^5$) arbitrary units (a.u.) and showed low relative standard deviation (7.24%) required for the precision of the protein determination. In total, 1570 and 1444 non-redundant proteins were identified in J192 and J164, respectively (Figures S4 and S5). Some proteins could not be identified in all groups, due to the phenomenon of under-sampling, that is, the numbers of

unique proteins could be expected to be lower than appeared as the result of the DDA experiments. To overcome this well-known limitation of the DDA experiments, label-free quantification with feature alignment across the whole dataset was accomplished. This analysis revealed 393 differentially expressed proteins detected among D1, D2, and D6 stages and two genotypes (J164 and J192). The principal component analysis (PCA) showed a clear separation between the developmental stages and genotypes with the first principal component (PC1) explaining 43.5% of the variability (Figure S6a). The highest difference was observed between the proteomes of the mature (D6) and early (D1 and D2) stages. Heatmap data visualization (Figure S6b) showed that SC sampled at the D6 stage clustered separately (indicating the most pronounced changes), whereby the D1 and D2 clustered separately but with a lower distance. This result makes the time-based paired comparisons more logical than the analysis of the genotype-related clusters. To address the time-based inter-genotype differences in more detail, we made the corresponding paired comparisons.

Functional annotation of differentially expressed proteins in SC from domesticated landrace J192 at D1 stage (Figure S7) showed higher abundance (in comparison to J164) of 73 proteins mostly involved in sugar (with

the most strongly regulated representative—sucrose-phosphatase 1, 5-fold) and protein metabolism (most regulated—isoaspartyl peptidase/L-asparaginase, 10-fold) and photosynthesis. Interestingly, despite that both J192 and wild J164 have SC pigmentation, proteins involved in epigallocatechin synthesis (anthocyanidin reductase), synthesis of cyclic metabolites derived from the C18 fatty acid α -linolenic acid (12-oxophytodienoic acid-reductase, 9-fold) were more abundant in the domesticated J192 (Table S6a). Furthermore, proteins involved in the detoxification of reactive oxygen species (ROS), such as glutathione S-transferase, thioredoxin, peroxiredoxin, and caleosin, were also more abundant in wild J192. On the other hand, 78 proteins involved in the protein and amino acid metabolism were less abundant in SC of domesticated J192 in comparison to the wild pea. Later in the development (D2 stage), the patterns of differential expression became less diverse and were represented with only 33 proteins (Figure S8, Table S6b). Proteins more abundant in J192 were mostly involved in secondary metabolism, lipid, amino acid, and protein metabolism and metabolism of hormones, while 12 proteins more abundant in wild J164 were involved in amino acid, nucleotide, and protein metabolism. Functional annotation of differentially expressed proteins in mature seeds (Figure 3) revealed

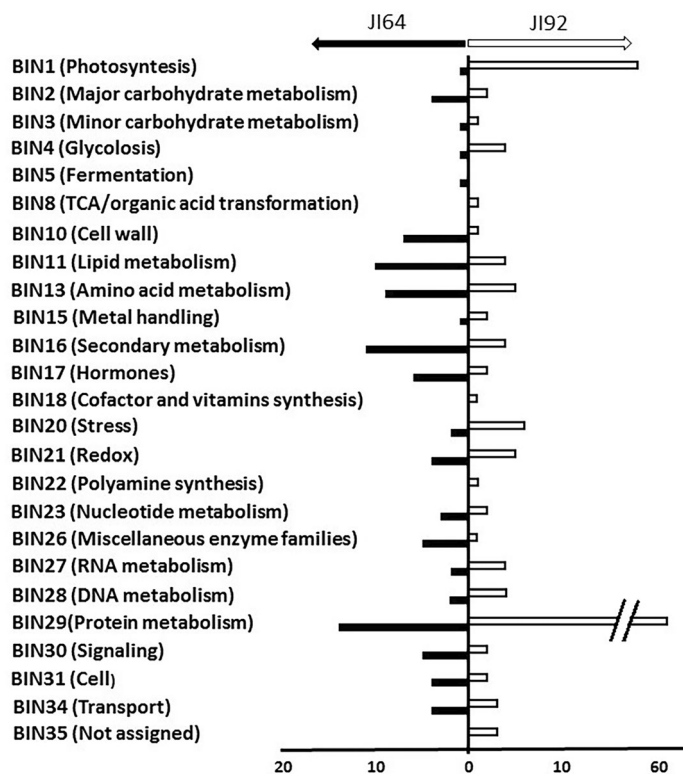


Figure 3. Functional annotation of the pea seed coat proteins (accomplished with the Mercator MapMan v3.6 tool), characterized with significantly higher expression in the mature seeds of J192 (white columns) and J164 (black columns).

127 polypeptides significantly more abundant in JI92, and these were involved in photosynthesis (e.g., photosystem II 22 kDa protein, 22-fold, chlorophyll a-b binding protein, 7-fold, chlorophyll a-b binding protein, 7-fold, ferredoxin-NADP reductase, 4-fold) and sugar metabolism (glyceraldehyde-3-phosphate dehydrogenase, fructose-bisphosphate aldolase, alpha-1,4 glucan phosphorylase, threonine synthase, aspartate aminotransferase, aspartate carbamoyltransferase 3). Furthermore, 58 proteins encoding ribosomal proteins, two Ca-binding proteins (caleosin and calcium ion-binding protein, 8-fold), and several seed storage proteins (late embryogenesis abundant protein-like, 53-fold, Group 3 LEA protein) were more represented in mature SCs of domesticated JI92 in comparison to wild JI64 (Table S6c).

Among the 86 proteins more abundant in D1 stage of the wild JI64, the polypeptides involved in secondary metabolism, cell wall formation, lipid, amino acid, protein metabolism, RNA processing (NOP56-like pre-RNA processing ribonucleoprotein, 122-fold), and DNA repair (putative DNA repair protein Rad4, 1173-fold). Notably, a protein with N-linked glycosylation activity (dolichodiphosphooligosaccharide-protein glycosyltransferase subunit 2) was highly abundant (4816-fold) in JI64 (Table S6d). Furthermore, polyphenol oxidase (PPO), an enzyme involved in the oxidation of diphenols, was more abundant in JI64 (34-fold). Later in the development (D2), proteins involved in RNA processing (NOP56-like pre-RNA processing ribonucleoprotein, 73-fold, RNA helicase, 7-fold) or DNA-repair (putative DNA repair protein Rad4, 1000-fold) were again more abundant in JI64 (Table S6e). The secondary metabolism functional class was represented by three highly upregulated enzymes of phenylpropanoid and polyphenol metabolism, with 151-fold upregulation of PPO in JI64 wild-type seeds (Balarynová et al., 2022), caffeic acid 3-*O*-methyltransferase (COMT, 2-fold), and dihydro flavanol 4-reductase (6-fold). In addition, proteins involved in sugar (sucrose synthase, 4-fold) and protein biosynthesis (carboxypeptidase, 104-fold, adenosylhomocysteinase, 11-fold, glutamine amidotransferase, 45-fold) (Figure S8) and secondary metabolism, lipid, amino acid, and metabolism of hormones were observed. Notably, there was the differential expression of lipoxygenase protein (434-fold) known to differ between JI64 and JI92 (North et al., 1989).

In the mature (D6) stage, the proteins of cell wall metabolism group more abundant in JI64 were represented by enzymes involved in the decarboxylation of sugar acids and pectin modification (pectinesterase, pectin acetylerase, each 2-4-fold), enzymes of fatty acid biosynthesis (Table S6f), with protease activity (class I glutamine amidotransferase, 214-fold), auxin (indole-3-acetyl-amido synthetase), GA (gibberellin 2-beta-dioxygenase, 4-fold) metabolism, and short-chain alcohol dehydrogenase (33-fold upregulated) (Figure 3). Dihydro flavanol 4-reductase (2-fold), cinnamic acid 4-

hydroxylase, flavanone-3-hydroxylase, cinnamyl alcohol dehydrogenase, and chalcone-flavanone isomerase family protein (1- to 3-fold) connected to monolignol and flavonoid biosynthesis were also more abundant in JI64. Among them, caffeic acid 3-*O*-methyltransferase (COMT, 10.7-fold) and polyphenol oxidase (133-fold) were more abundant in wild JI64 compared to JI92 landrace. By cellular localization, the SC proteins were predicted to be in cytoplasm, nucleus, and chloroplast compartments (34, 25, and 23%, in JI92 and 28, 27, and 12% in wild JI64, respectively) and this distribution remained unchanged throughout the development (Figure S9).

The seed coat of wild peas has a larger amount of soluble protective metabolites

The properties of the seed coat (especially its permeability for water and solutes) might depend both on its soluble constituents (i.e., those extractable with organic solvents and/or aq. alcoholic mixtures) and cell wall-bound components (i.e., extractable only upon alkali or acid treatment). Analysis of the extractable metabolome revealed 170 features, which were identified in the aqueous methanolic extracts (Table S7). As some metabolites appeared as several isomers or derivatives, the total number of identified primary metabolites was 109, represented by 20 amino acids, five amines, three fatty acids and their esters, 31 organic acids, 52 sugars (monosaccharides and their phosphorylated derivatives); di- and oligosaccharides, two phenolic compounds, three phosphoric acids, six representatives of other classes; and 34 could not be identified (Table S7). To address the changes in the primary metabolome associated with seed dormancy, we compared the abundances of individual metabolites isolated from the SC of domesticated (Cameor and JI92) and wild pea (JI1794, JI64, and JI261) genotypes, as shown in principal component analysis (PCA) score plot (Figure S10a). The comparison of Cameor and JI261 samples (t-test with Volcano plot representation) revealed 23 and 18 metabolites significantly ($P \leq 0.05$, $FC \geq 2$) down- and upregulated in the seeds of the domesticated genotype, respectively. In agreement with this, hierarchical clustering analysis revealed (Figure S10b) two contrasting clusters (Figure S11b). The metabolites, downregulated in the Cameor SC in comparison to wild JI261, were represented by mono- and oligosaccharides, and organic acids (including amino acids). The most pronounced differences, associated with domestication, for example, loss of dormancy, were observed for nicotinic acid (297-fold abundance decrease), mannitol (65-fold), 4-aminobutyric acid (GABA, 16-fold), and the unknown trisaccharide RI3295 (up to 29-fold) (Table S8). On the other hand, upregulated metabolites were represented by organic acids with the most significant changes observed in the levels of caffeic and dehydroascorbic acids (up to 21-fold).

Comparison of mature (D6) stage SC of non-dormant landrace JI92 and dormant wild JI261 revealed complete separation both in PCA score plot (with 88% of the explained difference in the PC1, Figure S12a) and hierarchical clustering (Figure S12b) with two pronounced clusters of 30 most confident features (Figure S12c). The t-test analysis with Volcano plot showed 15 and 25 metabolites (Figure S12d) significantly down- and upregulated, respectively (Table S9). Mannitol (20-fold) and unknown trisaccharide RI3295 (12-fold) were the most strongly downregulated, while upregulated metabolites were represented by organic acids, dehydroascorbic (85-fold) and ascorbic (14-fold) acid, as well as epigallocatechin (19-fold) and unknown metabolite RI 1833 (54-fold). Finally, a paired comparison of wild genotypes, that is, JI1794 versus JI261 (Figure S13) and JI64 versus JI261 (Figure S14) showed 18 and 9 downregulated metabolites, indicating genotypic specificities. Among them, sugar alcohols—mannitol (32-fold), galactinol (14-fold), and unknown trisaccharides (RI3295 and RI3206, 14-fold) (Figure S14) were found. These differences were accompanied by 2- to 3-fold higher relative contents of organic acids (glucaric, succinic, malonic, ribonic, malic) in wild JI64 along with a 2-fold difference in the levels of alanine (2.5-fold) (Tables S10 and S11).

Seed coats of wild peas are rich in cell wall-bound metabolites

To address the patterns of the metabolites covalently bound to the cell walls, analysis of SC and sequentially subjected to alkali and acid hydrolysis was performed to access ester- and ether-bound metabolome, respectively. The first analysis was expected to yield information on the periphery lignin and lignin-associated non-phenolic molecules, whereas the second one provided information on the composition of lignin itself (so-called core lignin). As, based on our previous experience, both these fractions were expected to be represented mostly with phenolics (Frolov et al., 2013), we used QqTOF-MS analysis in negative ion mode (Table S12). Importantly, due to the formation of different adduct ions and in-source fragmentation (which are both known to be universal limitations of electrospray ionization), each metabolite might be represented by multiple features. This increase in the potential analyte numbers essentially extends the analysis space and makes correction for multiple comparisons stricter. To compensate for this effect, we used Benjamini–Hochberg false discovery rate (FDR) correction at $P \leq 0.1$ (instead of the conventionally used threshold of $P \leq 0.05$).

The UHPLC-QqTOF-MS analysis of the alkali cell wall hydrolysates in negative ion mode yielded over 7500 signals of quasi-molecular negative ions (mostly the $[M-H]^-$ ones). These signals could be annotated with their exact m/z values (with a mass tolerance of 10 ppm) and retention

times (t_R). The resulting patterns showed clear differences between wild and domesticated SCs in the composition of the cell wall-bound metabolites (Figure S15a,b) and revealed 1791 and 18 metabolites significantly down- and upregulated in the non-dormant domesticated cultivar Cameor (Figure S15a,b). The application of a stricter (FC ≥ 10) filter reduced the list of downregulated features to 47 (Figure S16). Manual evaluation of the mass spectra and extracted ion chromatograms (XICs) allowed confirmation of 26 signals for reliable determination of elemental composition and MS/MS analysis (Table S13), whereas the other 21 either could not be confirmed or showed too low intensity (below 200 counts in the XICs). About half of the differentially abundant species could not be identified by the available MS/MS spectra (although even additional targeted MS/MS experiments were performed). The other 13 species could be tentatively annotated to certain structures of at least chemical classes based on their MS/MS fragmentation patterns. Remarkably, the second non-dormant domesticated genotype (JI92) showed an order of magnitude less abundant pattern of downregulated signals in comparison to the dormant JI261 (Figure S17), whereas the two wild pea genotypes (JI1794 and JI64) demonstrated less pronounced differences (Figures S18 and S19). Considering the differences in the patterns of the cell wall-bound metabolites, it can be seen that the loss of dormancy in the domesticated Cameor was associated with strongly pronounced (up to 3.9-fold decrease in IgFC, Table S13) downregulation of 13 compounds cleavable by alkali solutions. Among the seven tentatively identified species among those downregulated in cultured non-dormant genotype, four represented low-molecular phenolics/phenylpropanoids or conjugates (one of which, *O*-acetyl-2-phosphono-3-(4-hydroxyphenyloxy)-succinic acid, appeared to be the most depleted in the non-dormant SCs), one polyphenol and two fatty acid conjugates (which were more than 200-fold downregulated, Table S13). Surprisingly, the differences between the dormant genotype JI261 and the non-dormant genotype JI92 were much less pronounced both in the numbers and relative abundances of regulated metabolites. Thus, only five metabolites (two of which—unknowns) were annotated as downregulated in the non-dormant genotype. These three metabolites also represented the classes of low-molecular-weight phenolics and fatty acids. Remarkably, this downregulation was much less pronounced ($- \text{IgFC}$ did not exceed 2.0 in this case). A comparison of two dormant JI261 and JI64 yielded slightly different patterns of differentially abundant metabolites. These patterns were featured with phenolic-fatty acid conjugates (compounds 22–25). Interestingly, two of these conjugates (24, 25) were more abundant in the SCs from JI64, whereas two (22, 23) showed higher abundance in JI261. Finally, a comparison of JI261 with JI1794 did not reveal any phenolics and fatty acids among the only three

metabolites, which were annotated as differentially abundant. The MS/MS spectra of annotated cell wall-bound metabolites are shown in Figure S20.

Phenolic compounds are formed, oxidized, and accumulated during pea seed development

Untargeted metabolomics based on flow injection–electrospray ionization–high-resolution tandem mass spectrometry (FIA-ESI-HRTMS) was used to study metabolite dynamics during seed development. The score plot from the principal component analysis (PCA) shows the distribution of samples through the development and

domestication status. Cultivated and landrace genotypes were separated from the wild ones through development, with clear differences at D4–5–6 stages (Figure 4a), while at early D1–2 stages, the metabolic profiles were more similar both in negative (Figure 4a) and positive (Figure 4b) ionization mode. To obtain metabolic markers for early and late developmental stages, we compared early stages (D1 and D2) against late ones (D5 and D6). Identified compounds from the most significantly differential signals (MSDS) were taken from S-plot obtained by orthogonal projection to latent structures (OPLS-DA; datasets contained signals of 1888 markers in negative ionization mode and 635

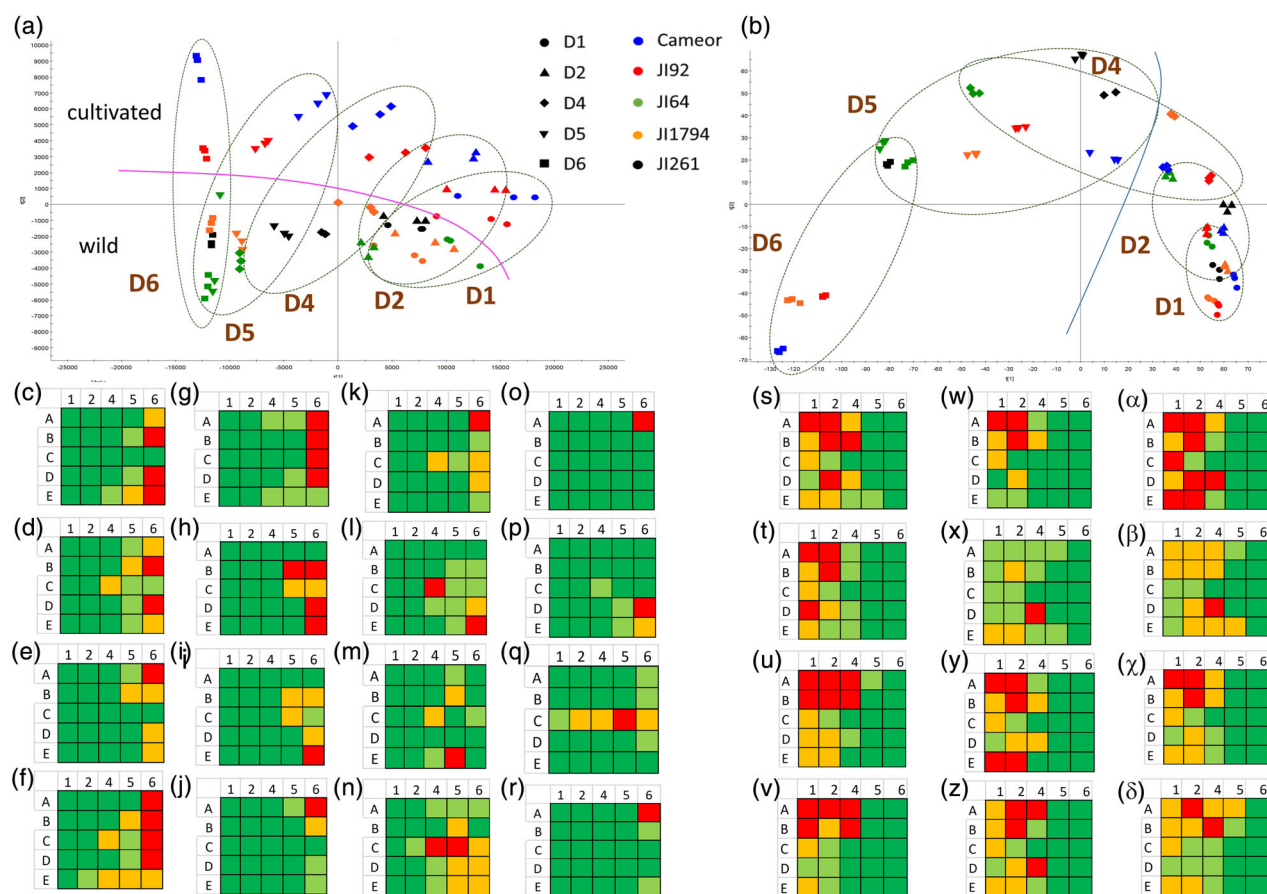


Figure 4. Principal component analysis of FIA-ESI-MS measurements of metabolite signal changes in negative (a) and positive (b) ESI modes. The seed coats of two domesticated (Cameor and JI92) and three wild (JI64, JI1794, and JI261) peas were studied at five developmental stages (13, 17, 23, and 28 DAP and mature seed, labeled as D1, D2, D4, D5, and D6, respectively). Heatmaps show profiles of most significantly differential metabolites in A–Cameor, B–JI92, C–JI64, D–JI1794, and E–JI261 (vertically); at developmental stages D1, D2, D4, D5, and D6 (horizontally from left to right); (c) coniferol dimers (m/z 357.1236); (d) hydroxybenzoyl glucoside (m/z 299.0843); (e) 7-O-methylaidzein (isofomononetin; m/z 267.0841); (f) sarmantosin (m/z 274.1028); (g) ophthalmic acid (m/z 288.1233); (h) 3',4',5'-trihydroxy-5,7,8-trimethoxy-2,3-dihydroflavone dihexoside (m/z 701.2209); (i) stachyose/cellotetraose (m/z 665.2463); (j) gibberellin A3/A6 (m/z 345.1344); (k) feruoylglucoside (m/z 355.0974); (l) gallic catechin trimer, type A (m/z 911.2072); (m) coniferol (caffeoyl)glucoside (m/z 539.1605); (n) hydroxybenzoic acid (m/z 137.0297); (o) nitidanin (tentatively; m/z 403.1433); (p) vaniloylconiferol (m/z 329.1020); (q) gallic acid (m/z 169.0217); (r) genistein (malonyl)glucoside (m/z 517.1082); (s) dehydrated hexose trimer with methyl group (m/z 273.0776, $([M + 2Na-2H]^{2+}/2)$); (t) hexose octamer with acetyl and methyl group (m/z 707.2167, $([M + 2Na]^{2+}/2)$); (u) hexose tetramer with carboxyl group and three methyl groups (m/z 723.1985, $([M + Na]^{+})$); (v) hexose hexamer with carboxyl group and three methyl groups (m/z 1065.3158, $([M + Na]^{+})$); (w) dehydrated hexose tetramer with methyl group (m/z 354.1065, $([M + 2Na]^{2+}/2)$); (x) hexose trimer with methyl group (m/z 281.5763, $([M + 2Na]^{2+}/2)$); (y) dehydrated hexose hexamer with methyl group (m/z 525.1646, $([M + 2Na]^{2+}/2)$); (z) hexose hexamer with methyl and carboxyl group (m/z 533.1569, $([M + 2Na-2H]^{2+}/2)$); (α) hexose octamer with three carboxyl groups (m/z 704.1968, $([M + 2Na]^{2+}/2)$); (β) hexose tetramer with methyl group (m/z 362.1019, $([M + 2Na-2H]^{2+}/2)$); (γ) dehydrated hexose octamer with methyl group (m/z 696.2092, $([M + 2Na]^{2+}/2)$); (δ) hexose dimer with carboxyl group (m/z 381.1023, $([M + Na]^{+})$).

markers in positive ionization mode, respectively, ordered according to $p(\text{corr})[1]$ coordinate) (Table S14). Identified MSDS more abundant in mature (D5 and D6) developmental stages were: coniferol dimers (m/z 357.1236); hydroxybenzoyl glucoside (m/z 299.0843); 7-O-methylidaidzein (isoformononetin; m/z 267.0841); sarmentosin (m/z 274.1028); ophthalmic acid (m/z 288.1233); 3',4',5'-trihydroxy-5,7,8-trimethoxy-2,3-dihydroflavone dihexoside (m/z 701.2209); stachyose/cellotetraose (m/z 665.2463); gibberellin A3/A6 (m/z 345.1344); feruoylglucoside (m/z 355.0974); gallic acid (m/z 169.0217); genistein (malonyl)glucoside (m/z 517.1082) (negative ion mode, Figure 4c–r; Table S15). Identification of metabolites was based on the exact mass measurement and fragmentation pattern. The main diagnostic fragmentation processes observed in MS/MS spectra of phenolic compounds included loss of dehydrated hexoses in the case of glycosylated forms, fission of polyphenolic skeletons (loss of water, CO, and retro-Diels–Alder fragmentation), loss of carboxyl group in the form of CO₂ and cleavage of methyl radical from methoxy groups. In the collision spectra of gibberellin(s), this included a loss of carboxyl group, water, and splitting of the middle five-membered cycle. Fragmentation of sarmentosin included loss of water and cyano-group. The selectivity of CID experiments in tandem mass spectra of hydroxybenzoyl glucoside and 3',4',5'-trihydroxy-5,7,8-trimethoxy-2,3-dihydroflavone dihexoside was increased using ion mobility separation (Figures S21 and S22).

In early developmental stages, modified sugars (dimers–octamers) containing methyl, acetyl, and carboxyl groups were found among MSDS significantly more abundant compared to late ones. In related MS/MS spectra, losses of dehydrated hexose(s) and sugar chains containing acetyl, methyl, and carboxyl groups were observed confirming the proposed structures. The decrease of modified sugar oligomers abundance during SC development can be explained by their polymerization and formation of insoluble long sugar chains including pectins (Figure 4s–t; Table S16).

Wild pea seed coats accumulate flavonol glycosides and gallic acid oligomers

We compared metabolites in SC samples from late D5 and D6 stages between two domesticated (Cameor and JI92) and three wild (JI64, JI1794, and JI261) pea genotypes. Score plot from PCA analysis showed the separation of domesticated samples from wild ones (Figure 5a). Among the most differential metabolites present at significantly higher amounts in wild SC compared to domesticated ones,

altogether nine metabolites were identified (Figure 5b–j; Table S17) including flavonol glycosides such as myricetin-galloyl glucoside, myricetin-3-caffeoyl glucoside, and phloridzin, a glycosylated dihydrochalcone phloretin. The remaining identified metabolites increased during SC development were gallic acid oligomers and glycosides (gallic acid di-/trimer glycosides, gallic acid trimer, gallic acid gallate glycoside), dehydrated gallic acid-trihexoside, and hydroxybenzoic acid. The main diagnostic fragmentation processes observed in MS/MS spectra included loss of dehydrated hexoses, acylhexoses and acyls (galloyl, caffeoyl), gallic acid units and gallic acid hexosides, fission of polyphenolic skeletons (loss of water, retro-Diels–Alder fragmentation), and loss of carboxyl group (-CO₂) (Table S17). Figures S23–S26 describe the MS/MS spectra where the ion mobility separation was applied (i.e., spectra of gallic acid gallate hexoside, myricetin 3-galloyl hexoside, myricetin 3-caffeoyl hexoside, and dehydrated gallic acid-trihexoside). LC/HRTMS analysis confirmed the presence of both *p*-hydroxybenzoic and salicylic acids (Figure S27) in comparison of dormant JI64 and domesticated landraces JI92.

Next, we focused on the phenylpropanoid pathway with a monolignol branch to check if there is a change in metabolic flow, as indicated by proteomic analysis. Signals of selected metabolites of the monolignol pathway were compared across the five developmental stages of five genotypes (Figure 6). Cinnamate and *p*-coumaraldehyde were detected as dominant metabolites during D1 and D2 stages, except for dormant JI261, where the expression was low in all stages. Metabolites of the monolignol pathway strongly correlated with gene expression of the ME1-Blue and ME19-Turquoise gene modules (Figure S28). Positive correlation with genes from the ME19-Turquoise module (decreasing expression during development) provided metabolites whose intensities decreased with development, such as phenylalanine, cinnamic acid, *p*-coumaric acid, *p*-coumaraldehyde (except JI261), and vanillin. Metabolites positively correlated with genes from the ME1-Blue module (increasing expression during development) were *p*-coumaroyl quinate, coniferyl-/5-hydroxyconiferyl-/sinapaldehyde (with two exceptions: 5-hydroxyconiferyl-/sinapaldehyde in JI261), *p*-coumaroyl-/coniferyl-/caffeoyl-/5-hydroxyconiferyl-, sinapyl-alcohol, and syringin (in JI261). During SC development, caffeic acid content increases in non-dormant genotypes, while changes only slightly in JI64. In the remaining two dormant genotypes, it first decreases then followed by an increase. SC of genotype JI261, the dormant one, showed a low abundance of mostly measured metabolites in comparison with other genotypes. Pigmented genotypes (JI92, JI64, JI1794, and JI261) have a higher expression of enzymes 4CL (4-coumarate:CoA ligase), HCT (hydroxycinnamoyl CoA:shikimate

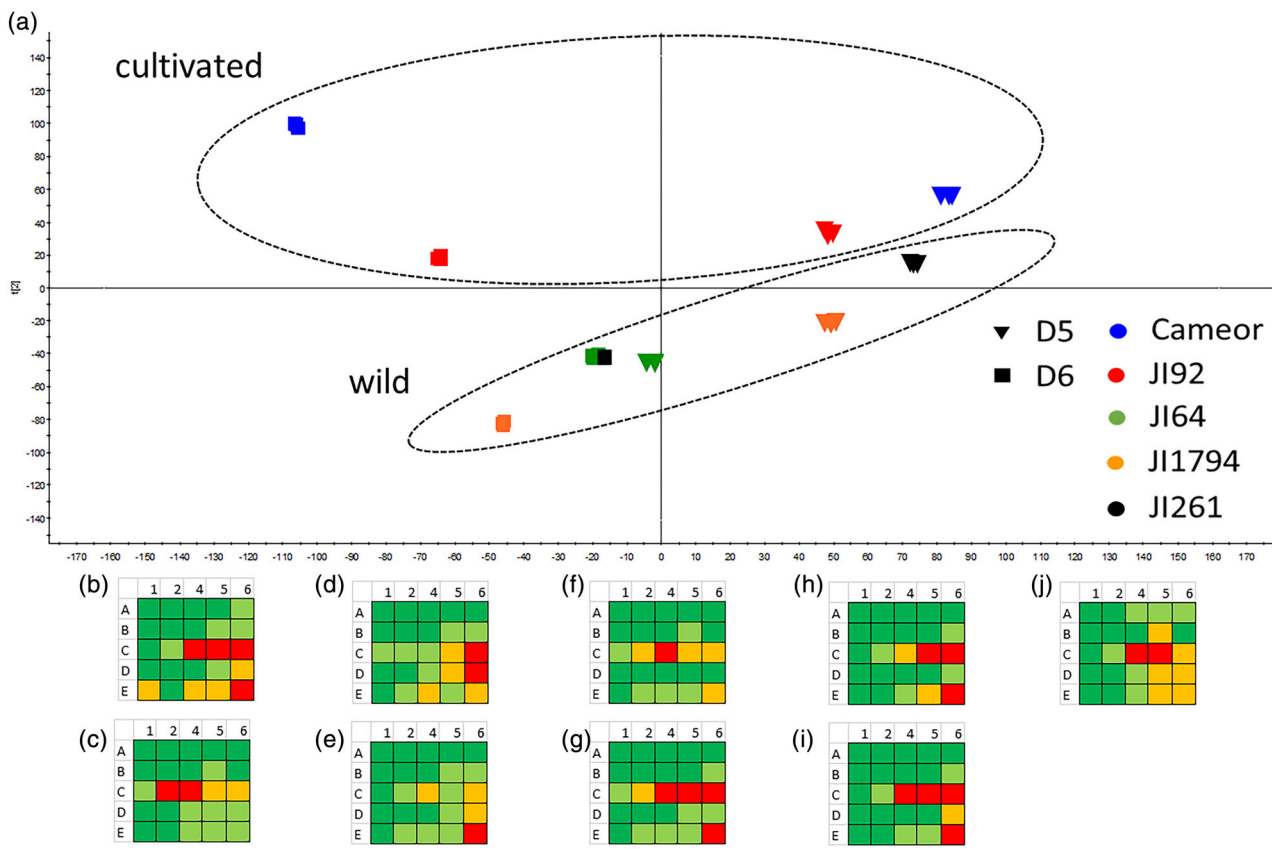


Figure 5. Principal component analysis of FIA-ESI-MS measurements of metabolites in negative ESI mode (a) describing differences among wild and cultivated genotypes at late developmental stages (D5-6). Heatmaps represent the profiles of identified the most significantly differential metabolites taken from S-plot provided by orthogonal partial least squares discriminant analysis. A-Cameor, B-JI92, C-JI64, D-JI1794, E-JI261 (vertically); D1, D2, D4, D5, and D6 describe the developmental stages (horizontally from left to right); (b) myricetin 3-galloyl glucoside (m/z 631.1017); (c) phlorizin (m/z 435.131); (d) myricetin 3-caffeoyl glucoside (m/z 641.1139); (e) gallic acid trimer glucoside (m/z 1073.1982); (f) gallic acid dimer (m/z 609.1249); (g) gallic acid trimer (m/z 913.183); (h) dehydrated gallic acid trihexoside (tentatively; m/z 771.1346); (i) gallic acid gallate glucoside (m/z 619.1041); (j) hydroxybenzoic acid (m/z 137.0247).

hydroxycinnamoyltransferase), and CAD (cinnamyl alcohol dehydrogenase) in SC. Finally, COMT (caffeic acid 3-O-methyltransferase) and CCoAMT (caffeoyl-CoA O-methyltransferase) are more expressed in wild pea genotypes (Table S18). Altogether, wild peas synthesize fewer monolignols, probably because of channeling a part of precursors to the flavonoid biosynthetic pathway. Identification of metabolites from the phenylpropanoid pathway was based on exact mass measurement and fragmentation pattern. The main diagnostic fragmentation processes observed in MS/MS spectra of phenolic compounds included loss of water, losses of CO_2 and acetic acid in the case of phenolic acids, losses of CO and acetaldehyde in the case of aldehydes, losses of CH_3OH and $\text{C}_2\text{H}_5\text{OH}$ in the case of alcohols and losses of shikimate/quinic acid and their combined losses with CO_2 , and acetic acid in the case of phenolic acids shikimate/quinic acid derivatives. In the collision spectra of glycosylated forms, the losses of hexoses were observed. Fragmentation of phenylalanine included loss of amino group (Table S19).

DISCUSSION

This work presents the first comprehensive gene expression, proteomic, and metabolomic profiling of pea seed coat development in relation to seed dormancy and domestication. It complements previous studies of gene expression, anatomy, and metabolite profiles of seed coats between wild and cultivated peas (Balarynová et al., 2022, 2023; Hradilová et al., 2017; Janská et al., 2019; Zablazková et al., 2021) (Figure 7). Although seed development involves the coordination of molecular, cellular, and biochemical processes among the embryo, endosperm, and seed coat, this study focused on the seed coat, as its permeability changed during domestication is important for dormancy in many species (Smýkal et al., 2014). Relatively advanced stages of seed development, from 13 DAP to the mature dry stage, were analyzed to capture changes related to differences in seed coat properties when dormancy is established (Balarynová et al., 2023; Janská et al., 2019; Smýkal et al., 2014). Moreover, this study compared

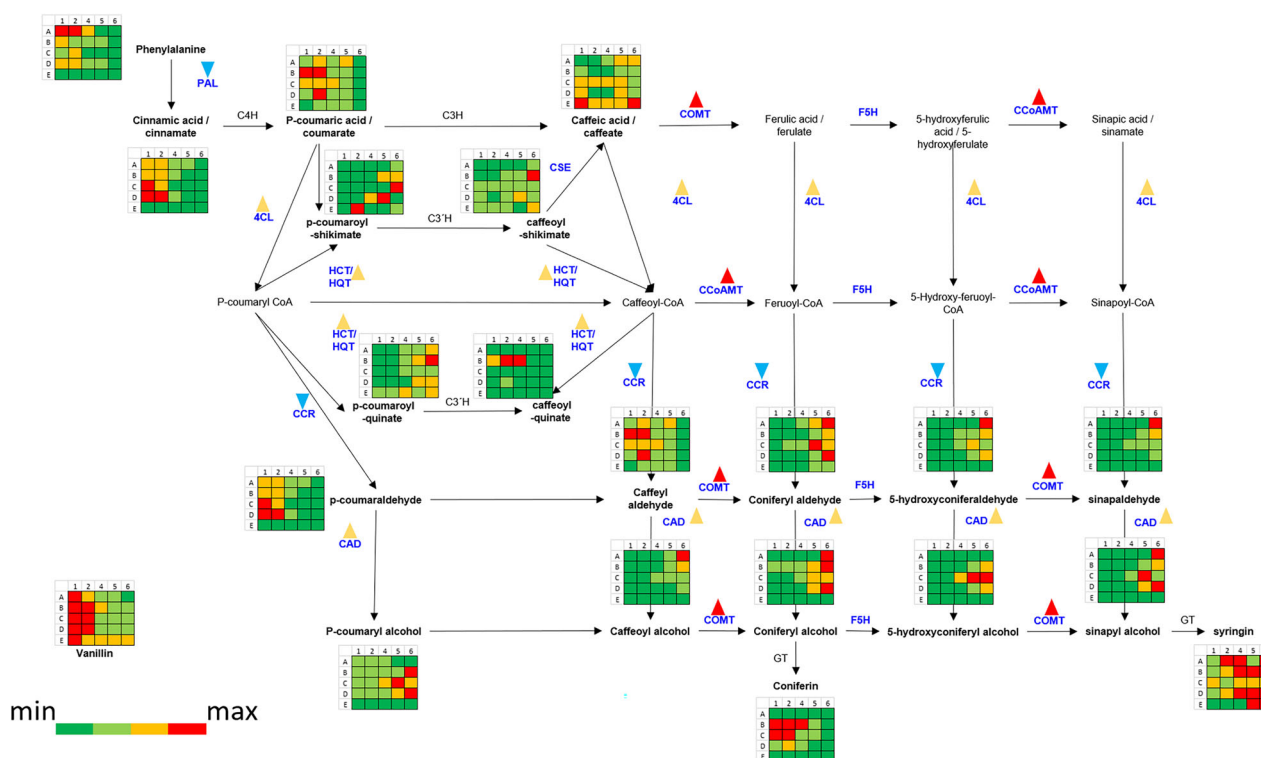


Figure 6. Scheme of monolignol biosynthetic pathway with metabolites identified in cultivated (Cameor and JI92) and wild (JI64, JI1794, and JI261) pea seed coats during development. Annotated enzymes (Figure S18) are written in blue letters. Genotypes and developmental stages are labeled as A-Cameor, B-JI92, C-JI64, D-JI1794, E-JI261 (vertically) and D1, D2, D4, D5, and D6 (horizontally from left to right), respectively. PAL: phenylalanine ammonia-lyase, C4H: cinnamate-4-hydroxylase, C3H: 4-coumarate-3-hydroxylase, COMT: caffeic acid O-methyltransferase, 4CL: 4-coumaroyl: CoA ligase, CSE: caffeoyl shikimate esterase, HCT/HQT: hydroxycinnamoyl CoA:shikimate hydroxycinnamoyltransferase, CCR: cinnamoyl CoA reductase, CCoAMT: caffeoyl CoA-3-methyltransferase, F5H: ferulate-5-hydroxylase, CAD: cinnamyl alcohol dehydrogenase.

wild and domesticated genotypes, in contrast to typical single-genotype studies of cultivated crop. This experimental setup of several genotypes and developmental stages, however, brought also some difficulties related to the multidimensionality of the analysis and genotypic specificities. The latter were visible in gene expression, protein, and metabolome profiles. Cameor, as representative of modern pea varieties set apart, while landrace (JI92) was closer to wild peas. Wild pea genotypes also displayed specificities. Interestingly, wild JI1794 had gene expression closer to cultivated pea. This might reflect its genetic relationship, as representative of wild progenitor closer to pea crop (Hellwig et al., 2022; Smýkal et al., 2017; Trněný et al., 2018). RNA-seq analysis revealed more tissue-specific genes expressed in the seed coat than in the embryo, supporting the important role of SC in the regulation of pea seed development (Smýkal et al., 2014) and is in agreement with results on lentil seeds (Yu et al., 2023). Transcriptomic data showed the largest change in gene expression between stages D2 and D3 corresponding to 17 and 20 DAP (days after pollination), respectively (Figure 1d). This relates to reaching the physiological maturity of the seeds (Balarynová et al., 2023), involving

accumulation of osmoprotectants and LEA proteins in connection to desiccation, antioxidant accumulation, and ABA metabolism. Conclusively, this developmental part is the point of the transmission from the early stage to the maturation phase characterized by changes in seed transcription and metabolism as shown by developmentally regulated genes (Figure 2b,d).

There are some comparable studies of the legume crop seed development. The issue of seed coat permeability was studied in soybean (Ranathunge et al., 2010) leading to the identification of genes involved in cutin and suberin formation and modulation of unsaturated fatty acid levels (regulated by fatty acid desaturases). These genes have not been detected in pea seed coat, except for lipoxygenases and thioredoxins, likely due to physiological differences between soybean (rich in oil) and pea seeds. When the black and brown soybean seeds were compared (Kovinich et al., 2011), genes of the phenylpropanoid/flavonoid pathways such as UDP-glycosyltransferase, anthocyanin-3'-O-methyltransferase, and caffeoyl-CoA-O-methyltransferase were identified as differentially expressed genes. This largely agrees with the

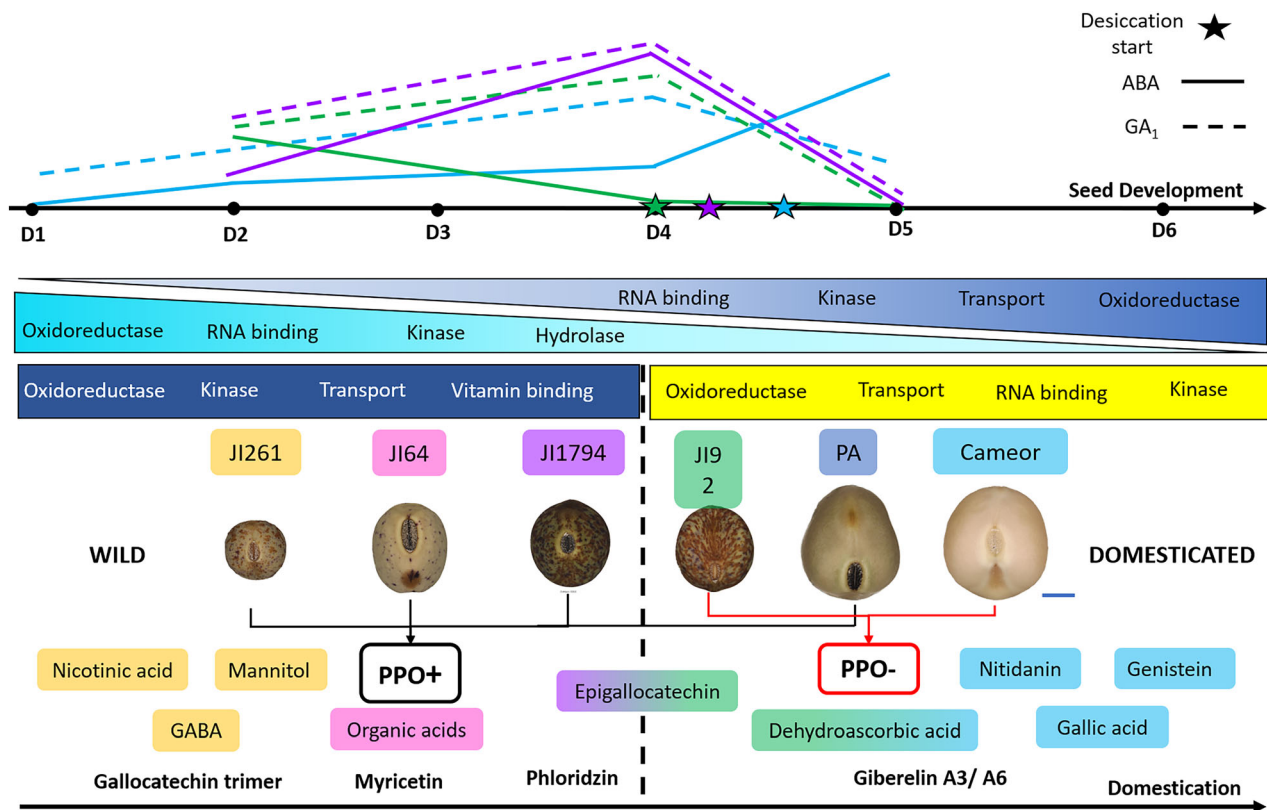


Figure 7. The scheme summarizing the principal results supplemented with several recent findings involving hormonal levels, point of seed desiccation, and functionality of the polyphenol oxidase (PPO) gene. The upper part shows abscisic acid (ABA) and gibberellin (GA) contents during pea seed development. The black arrow presents the direction of development, D1–D6 are studied developmental stages (13, 17, 20, 23, and 28 days after pollination and dry seed, respectively). The solid and dashed lines represent the levels of ABA and GA₁, respectively, while the line color corresponds to a specific pea genotype. The triangles mark the point at which the seed desiccates. This part of the scheme was adapted from Balarynová et al. (2023). The dark blue and turquoise triangles represent ME1-Blue (genes whose expression increases with development) and ME19-Turquoise (genes whose expression decreases with development) gene expression modules, respectively, described using the most abundant GO categories. The bottom part of the scheme is divided by the black dashed line into the right and left parts representing domesticated and wild pea genotypes, respectively. The rectangles represent ME15-Midnightblue and ME6-Yellow gene expression modules describing the most abundant GO categories of genes associated with dormancy and domestication, respectively. Balarynová et al. (2022) described functional and non-functional alleles of the PPO gene, represented by black and red arrows, respectively. This enzyme catalyzes oxidation and polymerization of the gallicocatechin in the pea seed coat (SC). SC metabolites specific for the particular pea genotype are highlighted by the same color as the corresponding genotype. Finally, the main metabolites identified in mature wild or domesticated pea genotypes are summarized.

findings in pea. Similarly, a comparison of variously colored pea seeds of domesticated genotypes identified phenylpropanoid pathway genes and metabolite differences (Ferraro et al., 2014) and this is reflected in polyphenolic profiles (Jha et al., 2019). Pigmented domesticated landraces (J192) apart from modern-type pea (Cameor) were deliberately included to capture differences related to domestication and dormancy. However, numerous genes, proteins, and metabolites of the phenylpropanoid pathway were found to be significantly upregulated in wild pea progenitor, in agreement with previous analyses (Balarynová et al., 2022; Cechová et al., 2017, 2019; Hradilová et al., 2017; Krejčí et al., 2022). Until now, the only genetically proven genes controlling seed coat permeability in legumes have been shown in soybean, encoding calcineurin-like protein and 1,4-β-glucanase (Jang et al., 2015; Sun et al., 2015) and those identified by mutagenesis in *Medicago*

truncatula, encoding KNOX4 transcription factor (Chai et al., 2016) altering beta-ketoacyl-CoA synthase expression (Chai et al., 2021) and ultimately affecting hydroxylated fatty acid profiles (Chai et al., 2016). Analysis of closely related alfalfa (*M. sativa*) showed a higher amount of fatty acid conjugates in hard (dormant) seeds (Wang et al., 2023). Despite the targeted analysis of respective homologous genes, studies have not found these to be associated with pea seed dormancy, suggesting that different mechanisms and pathways were altered during pea domestication (Smýkal et al., 2018). The convergence or parallelism of the domestication process is one of the main questions of man-directed evolution (Lenser & Theißen, 2013) and awaits the identification of respective genes in various phylogenetically and geographically distant species.

Recently, transcriptomic and metabolomic analysis was done on lentil seed development, although only at a

single cultivated lentil genotype, leading to the identification of isoflavonoid pathway genes expressed in later stages (Yu et al., 2023), in agreement with a previous (Hradilová et al., 2017) and this pea study. Since lentil and pea are phylogenetically closely related (Smýkal et al., 2015), these two studies are the most comparable, although lentil study did not include a wild progenitor. Phylogenetically more distant, yet topically relevant, is the study of seed coat development about hard seededness (permeability) in pomegranate (Qin et al., 2020) identifying genes and metabolites of the lignin and monolignol biosynthesis pathways, including coniferyl alcohol and sinapyl alcohol, accumulated in the pomegranate seed coat. This parallels of results on pea seed coat.

Although *Arabidopsis* seeds display physiological dormancy (Baskin & Baskin, 2014), a series of testa mutants (Appelhaagen et al., 2014) showed alteration in seed coat permeability and its importance for seed germination, related to phenylpropanoid metabolism (MacGregor et al., 2015). Moreover, it impacts also seed longevity (Sano et al., 2016), as flavonoids and proanthocyanidins are strong antioxidants. Legume seeds accumulate various secondary metabolites as protection against abiotic and biotic stress factors, for signaling and communication with the surrounding environment (Weston & Mathesius, 2013) and these are often species specific (Wink, 2013). These pathways are regulated by transcription factors (Zhao et al., 2013) and many of them were found within the ME1-Blue and ME19-Turquoise gene modules together with genes coding enzymes of phenylpropanoid and flavonoid biosynthetic pathway, similar to *Medicago truncatula* (Fu et al., 2017). Two TFs, namely, MYB5 and MYB14, were identified in *M. truncatula* seeds regulating the last steps of proanthocyanidins biosynthesis (Liu et al., 2014). In agreement with this, we have found MYB5 homolog in ME15-Midnightblue module positively correlated with dormancy (Figure 2b).

Domestication and seed coat modification

The impact of the domestication process on plant traits is well known (reviewed in Lenser & Theißen, 2013; Purugganan & Fuller, 2009; Smýkal et al., 2018). Mostly visible morphological changes were of primary focus and inspired Charles Darwin. Less is known about metabolic changes, except for visible seed and fruit pigmentation (Alseekh et al., 2021; Lyu et al., 2023; Maeda & Fernie, 2021; Paauw et al., 2019; Smýkal & Parker, 2023). However, it can be expected that besides morphology, human selection altered the nutritional value and composition of storage organs (fruits, seeds, and tubers). Examples include millennia-long selection and centuries of breeding for attractive color; reduced bitterness; and altered acidity, sweetness, and starchiness, as well as fragrances. Loss of

pigmentation in the course of domestication process is common both in plants and animals (Smýkal et al., 2018; Smýkal & Parker, 2023) and the underlying molecular mechanism has been identified in several legumes (Hellens et al., 2010; Pal et al., 2023). On the other hand, domesticated pea, lentil, or chickpea with non-dormant seeds can have both pigmented and non-pigmented seed coats (Balarynová et al., 2022; Hradilová et al., 2017; Pal et al., 2023; Sedláková et al., 2021, 2023) indicating that pigments (in broad sense) are not directly responsible for seed dormancy (reviewed in Smýkal et al., 2014). As seed metabolites are associated not only with nutritional value but also with physiological properties such as seed maturation, desiccation, and germination (Rao et al., 2014), it is not surprising to see differences related to domestication status in pea. Three classes of phenylpropanoid-derived compounds are particularly abundant in seeds: flavonoids, lignins, and lignans (Corso et al., 2020; Francoz et al., 2018) contributing also to seed color variation. These metabolites were found to play a role in seed dormancy in several plants such as pomegranate (Qin et al., 2020), *Rubus* (Wada et al., 2011), chickpea (Sedláková et al., 2023), alfalfa (Wang et al., 2023), and pea (Cechová et al., 2017, 2019; Hradilová et al., 2017; Janská et al., 2019; Krejčí et al., 2022) influencing likely seed coat permeability (Smýkal et al., 2014). Comparative analysis of wild progenitors, landraces, and modern varieties has shown a reduction in the content of these (reviewed in Ku et al., 2020). The question is if these have been selected directly, such as visual appearance or palatability (Cosson et al., 2022), or because of relation to other seed compounds, namely, storage proteins and sugars (Smýkal et al., 2018; Smýkal & Parker, 2023). The importance of phenylpropanoid pathway and alteration of respective metabolites in relation to pea domestication was shown recently. The activity of PPO differed between wild and domesticated pea seeds (Figure 7) as a result of loss-of-function mutation (Balarynová et al., 2022), and the same was found in faba bean seeds (Jayakodi et al., 2023). More enzymes involved in phenolic synthesis were altered during domestication. Transcriptomic and proteomic analyses showed a higher amount of caffeic acid O-methyltransferase (COMT) in wild pea seed coats. COMT enzyme controls important agronomic traits by regulating the level of lignin (Ma & Xu, 2008; Oraby & Ramadan, 2015) and melatonin (Chang et al., 2021) in plant tissue. Overexpression of COMT gene resulted in higher melatonin content increasing accumulation of osmolytes regulatory substances and stress tolerance in seeds (Bai et al., 2020; Chang et al., 2021; Lv et al., 2021). Suppressing of COMT caused a decrease in total lignin content of *Brassica napus* seeds (Oraby & Ramadan, 2015) and melatonin in *Arabidopsis* seeds resulting in a higher germination (Lv et al., 2021).

Wild pea seed coat is rich in defensive metabolites

Numerous metabolites accumulate in the seeds to provide physical and biochemical protection (Corso et al., 2020; Francoz et al., 2018). This is not surprising, considering the soil environment and herbivory. Consequently, a large number of SC genes (ME1-Blue module, Figure 2b) are related to primary and secondary metabolic processes such as cell wall biosynthesis, phenylpropanoid, flavonoid biosynthetic processes and annotated as having oxidoreductase activities, and this agrees with previous studies on *Medicago*, soybean or lentil (Fu et al., 2017; Kovinich et al., 2011; Yu et al., 2023). Notably, several peroxidases, glutathione peroxidase and primary amine oxidases were found to be associated with seed dormancy in *Medicago* (Renzi et al., 2020) and *Arabidopsis* (Renard et al., 2020) by the GWAS approach, and identified in *Medicago* SC development (Fu et al., 2017).

Legumes in particular are known to be rich in isoflavones. These are found in seeds and their content increase during germination (in chickpea; Wu et al., 2012). Their biological function in seeds is seed–microbe communication, facilitating mycorrhization (Dos Santos et al., 2020; Liu & Murray, 2016) and rhizobial colonization (Liu & Murray, 2016). The development of SC includes biosynthesis of phenolic secondary metabolites (phenolic acids, lignans, flavones, flavanols, and isoflavonoids, either glycosylated or present as aglycons) performing antioxidant, protective, and growth control functions. The production of galocatechin-based metabolites helps to protect plant tissues from fungal infection (Hammerbacher et al., 2018). Among identified metabolites present in SC of mature dry seed, there are several ones involved likely in seed defense, such as coniferol dimers, p-hydroxybenzoylconiferylalcohol, formononetin, and sarmentosine. Recently, these metabolites were found in hard-seeded alfalfa seeds (Wang et al., 2023). Formononetin is a precursor of pterocarpans, maackiain, and pisatin playing an important role in the interaction between fungi and the pea plant (DiCenzo & VanEtten, 2006; Kaimoyo & VanEtten, 2008). Pinoresinol is one of the structurally simplest lignan, a dimer of coniferyl alcohol, found in the wood and acts as seed coat protective neolignan (Yonekura-Sakakibara et al., 2020). Beta- and gamma-hydroxynitrile glucosides (such as sarmentosine) along with cyanogenic glucosides provide defense against pathogens and herbivores, communication with other plants, and adaptation to stressful environment. However, little is known about their biosynthesis and biological significance. Interestingly, arthropods which contain sarmentosin obtained from their host plants were reported to be strong deterrents against their predators (Bjarnholt et al., 2012).

Detection of salicylic acid (SA) and isomeric p-hydroxybenzoic acid (reviewed in Shah, 2003) particularly in SC of wild peas further supports the importance of

defense as SA is known to be an important signaling molecule in plant defense responses and it is involved in the modification of the redox balance (Wildermuth et al., 2001). Exogenous application of SA increased the level of unsaturated triglycerides, polyamines, and phenolics (such as quinic acid) in pea seeds (Berková et al., 2023). The relationship between SA and lignin levels was proposed and the association between phloridzin (dihydrochalcone glycoside related to flavonoids and produced on the side branch of phenylpropanoid pathway) SA and lignin was shown in apple (Chen et al., 2009; Zhou et al., 2019). Importantly, phloretin, an aglycon of phloridzin, is a substrate for polyphenol oxidase as shown in tomato (Kampatsikas et al., 2019) and pea PPO displays high homology to this (Balarynová et al., 2022). Phloridzin is particularly abundant in apples, including seeds, where its content is lowered during cold stratification required for seed germination (Lewak, 2011); however, knowledge about the physiological relevance of phloridzin *in planta* is limited (reviewed in Gosch et al., 2010) although its role in defense against fungal pathogens was described (Zhou et al., 2019).

Altogether, accumulated antioxidants, such as tocopherols, glutathione, ascorbate, polyols, quinones, and secondary metabolites such as flavonoids in seed coat are known to affect seed longevity (Sano et al., 2016). It is, thus, not surprising to find sugar alcohols (mannitol and galactinol) in seed coats, which are linked to seed longevity (de Souza Vidigal et al., 2016). Since ophthalmic acid (γ -glutamyl-L-2-aminobutyryl-glycine) is a marker of oxidative stress in plants and animals (Servillo et al., 2018), it fits into the antioxidative system and regulation of the redox state during seed maturation. Ascorbic acid (ASC) promotes seed germination when applied exogenously (Akram et al., 2017) while its endogenous level is involved in stress perception and redox homeostasis (De Tullio et al., 2004). Interestingly, ASC content increases during the early stages of development, followed by a decrease during the desiccation stage (De Tullio et al., 2004). This contrasts with the detection of a significant amount of both ACS and DHA in SC of mature dry pea seeds. Moreover, these have been differentially abundant between wild and cultivated pea SC samples (80- and 50-fold higher in cultivated genotypes; Tables S8 and S9).

Non-reducing carbohydrates (sucrose, raffinose, stachyose) are responsible for the induction of structural stability to membranes and proteins by replacing water molecules as a strategy to survive seed desiccation (Salvi et al., 2022). Indeed, galactinol and mannitol were differentially present, with higher amounts in wild pea SC. Seed desiccation of cultivated peas is slower than that of land-race and wild peas, and seed coat displays lower ABA and higher bioactive GA levels (Figure 7) than corresponding embryos (Balarynová et al., 2023). On the other hand, ABA and GA metabolic genes are more expressed in SC

compared to E (Balarynová et al., 2023). There is an intriguing relationship between Ca-binding proteins, lipid oxidation, and caleosin/peroxygenases detected among differentially expressed proteins. During seed germination, lipoxygenases (LOX) may function as storage proteins and aid the lipid bodies to degrade in the seed. Recently, these proteins were shown to have diverse functions (Hanano et al., 2023), including the final steps of hydroxyl-epoxy fatty acids formation and cuticle and suberin deposition. Notably, it has previously been identified as significantly more abundant long-chain hydroxylated fatty acids in SC of wild pea JI64 compared to landrace JI92 (Hradilová et al., 2017; Krejčí et al., 2022). Remarkably, the LOX activity is inhibited by some phenolic compounds, such as caffeic acid and quercetin (Szymanowska et al., 2009) suggesting feedback regulation. Since the LOX negatively influences the taste and smell of certain foods, as in the case of pea via *n*-hexanal, an off-flavor compound derived from lipoxygenase action (Bi et al., 2022), there has been a target of selection in breeding and possibly already during early domestication.

CONCLUSION

This work provides the first comprehensive gene expression, proteomic, and metabolomic profiling of pea seed coat development. Moreover, the comparison of wild and cultivated pea genotypes allowed analysis in relation to seed development, dormancy, and also domestication. The description of gene expression dynamics resulting in specific metabolic profiles provides new insight into pea domestication. This study has shown that domestication altered pea seed development and modified (mostly reduced) the transcript, protein, and metabolite composition of the seed coat, especially the compounds involved in seed defense and protection. It is tempting to speculate that wild pea seeds have higher longevity and stress tolerance. This remains to be tested experimentally. Since the seeds of wild peas must more precisely monitor their surroundings and respond to various biotic and abiotic signals, this is achieved by a higher diversity of proteins and metabolites present in their protective barrier (seed coat). These findings can be applied in the breeding of more resilient pea varieties.

EXPERIMENTAL PROCEDURES

Plant material

Cultivated pea cv. Cameor (*Pisum sativum* L., reference for pea genome sequence; Kreplak et al., 2019), primitive domesticated landraces JI92 and wild pea (*P. elatius* M.Bieb.) JI64, JI1794, and JI261 were used as in previous studies (Balarynová et al., 2022, 2023; Cechová et al., 2017; Hradilová et al., 2017; Janská et al., 2019). In addition, there was Ethiopian-origin *P. abyssinicum* A.Br. (PI358617) representing likely the second independent

domestication event (Hellwig et al., 2022; Trněný et al., 2018). In proteomic analysis, genotypes JI92 and JI64 were used as representative samples for the identification of contrasting protein groups with respect to dormancy status. Metabolomic analysis was performed with the set of five genotypes (Cameor, JI92, JI64, JI1794, and JI261, respectively). Plants were cultivated and seed developmental stages were labeled as described in Zablazková et al. (2021) and Balarynová et al. (2022, 2023). Randomly selected flowers were tagged on the day of opening (considered as day 0). Seed coats and embryos were separately sampled at 13, 17, 20, 23, and 28 DAP (days after pollination) and mature dry seeds, subsequently labeled as D1–D6. For proteomic analysis, seed coat samples at D1, D2, and D6 were used. The seeds were extracted from the harvested pods on ice. For transcriptomic and proteomic analysis, seed coats and embryos at D1–D5 were separated and rapidly placed in liquid nitrogen, and stored at -70°C for analyses. At the youngest (D1, 13 DAP) stage, the sample consisted of both SC and E, which was too small to be properly separated, while at later stages, SC and E were sampled separately.

RNA isolation and RNAseq analysis

Frozen seed coats were ground to a fine powder with liquid nitrogen using a sterile mortar and pestle. Total RNA was isolated using PureLink™ Plant RNA Reagent (Invitrogen). Residual DNA was removed by Baseline-ZERO DNase (Fisher Scientific, Czech Republic) treatment followed by phenol/chloroform extraction, and transcriptome profiles were generated using the Illumina NovaSeq platform performed by Novogene Ltd. (Cambridge, UK). Yield/quantity and purity were determined by using NanoDrop 2000 spectrophotometer (Fisher Scientific). The integrity of the RNA samples was examined with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Bioinformatic pipeline was prepared according to Pertea et al. (2016) with modifications. Reference pea genome *Pisum sativum* v1a version was used (Kreplak et al., 2019). The annotation file was converted from gff3 to gtf using *gffread* (Pertea & Pertea, 2020). Reference sequence index for mapping was prepared using HISAT2 v2.1.0 (Kim et al., 2015) on computational grid services of MetaCentrum using 400 GB RAM memory and 16 CPUs. Raw sequence reads were cleaned with adapter quality trimming process using BBDuk from BBTools v 38.73 (<https://jgi.doe.gov/data-and-tools/bbttools/bb-tools-user-guide/>) with parameters `minlen = 25 qtrim = r1 trimq = 10 ktrim = r k = 23 mink = 11 hdist = 1 tpe tbo`. Cleaned sequences were mapped to reference sequence transcripts using HISAT2 v2.1.0. Sequence Alignment Map files were converted to binary version by Samtools 1.9 (<http://www.htslib.org/>). Reference sequence-guided assembly was performed to reconstruct all isoforms of expressed genes using StringTie v2.1.0 (Pertea et al., 2015). Newly assembled transcripts from each genotype were compared with transcripts from reference sequences using *gffcompare* (<http://github.com/gpertea/gffcompare>). Merging all transcripts into an annotation file was performed using StringTie v2.1.0. Quantification of genes and transcripts was done by StringTie v2.1.0. The gene count matrix and transcripts count matrix were extracted using *prepDE.py* python script. Count matrices were used as input to identify differentially expressed genes using edgeR v3.30.3 (McCarthy et al., 2012) or DESeq2 v1.28.1 (Love et al., 2014) R packages. Normalization of the genes count matrix to FPKM values was performed using Ballgown v2.20.0 R package (Fu et al., 2020). The transcripts were joined to transcriptomic PeaAtlas (Alves-Carvalho et al., 2015) using blastn 2.5.0+ (Camacho et al., 2009). Transcripts were annotated using a customized EnTAP 0.10.6 annotation pipeline (Hart et al., 2020) against protein

databases eggnoG proteins, uniprot_sprot, plant_nr_database, uniprot_trembl. GO terms were searched in EggNOG database v 5.0.

PCA analysis was performed on FPKM normalized transcription levels with filtered out genes with variance less than 0.5 using *prcomp* function in R-4.2.2. (R Core Team, 2022). Data were scaled to have unit variance and zero centered before the analysis. PCA plot was made using *ggplot2* R library (Wickham, 2016). We used hierarchical clustering of the same FPKM values to observe directed relationship between samples. For this UPGMA incorporated in *hclust* and Euclidean distance in *dist* R functions was used. The circular tree was depicted using *factoextra* R library (Kassambara & Mundt, 2020). Common and distinct sets between samples and between developmental stages were shown by *upset* R library (Conway et al., 2017) with only more than 0.5 FPKM counts used.

Weighted gene co-expression network analysis (WGCNA)

Gene counts were subjected to a variance stabilizing transformation using the DESeq2 R package (Love et al., 2014), which also included data normalization. To prepare for the WGCNA analysis (Langfelder & Horvath, 2008), genes with counts transformed to 10 or greater that occurred in less than 20 samples were filtered out. Out of the initial 74 142 genes, 26 951 genes met these criteria and were utilized for the WGCNA analysis. The adjacency matrix was derived from a topological overlap matrix of “signed” type. Modules were identified as a single block using the following parameters: soft threshold power = 6, network type = “signed hybrid,” minModuleSize = 20, mergeCutHeight = 0.45, and deepSplit = 3. A gene co-expression network was visualized using Cytoscape 3.9.0, with edges included based on an adjacency threshold of 0.15 (Shannon et al., 2003) and arranged using a pre-fuse force-directed layout. GO enrichment analysis was performed using the BiNGO 3.0.3 app within Cytoscape (Maere et al., 2005), employing the hypergeometric test. Multiple testing correction was conducted using the Benjamini–Hochberg FDR method (Benjamini & Hochberg, 1995). Enriched GO terms were defined as those with false discovery rate (FDR) *P*-values less than 0.05. The functional content of GO enrichment terms in correlated gene groups was summarized by clustering the GO terms of genes within modules using GOMCL, a toolkit for clustering, evaluating, and extracting non-redundant associations of GO-based functions (Wang et al., 2020), with default parameters.

Protein isolation and processing

The total protein fraction was isolated by the phenol extraction (Mamontova et al., 2018, 2019). Dried SC samples were reconstituted in 10% (*w/v*) aq. SDS. The protein content was determined by 2D Quant Kit and cross-validated with SDS-PAGE (Leonova et al., 2022). Tryptic digestion relied on the filter-assisted sample preparation (FASP) protocol of Leonova et al. (2022). The resulting filtrates were desalted by solid-phase extraction (Mamontova et al., 2018, 2019). The completeness of tryptic digestion was verified by SDS-PAGE (Greifenhagen et al., 2016). Finally, the pre-cleaned eluates were freeze-dried overnight and stored at -20°C before analysis.

For the nano-LC-MS analysis, the sample set was randomized and standardized to quality controls (QCs, injected after each of six samples), and aliquots of a pool were obtained by mixing 10 μl of each tryptic digest. Individual digests (500 ng, 10 μl) dissolved in 3% (*v/v*) acetonitrile in 0.1% (*v/v*) aq. TFA were loaded onto an Acclaim PepMap 100 C₁₈ trap column (300 μm \times 5 mm, 3 μm particle size) for 15 min at the flow rate of 30 $\mu\text{l min}^{-1}$. The proteolytic peptides were separated at the flow rate of

300 nl min^{-1} on an Acclaim PepMap 100 C₁₈ column (75 μm \times 250 mm, particle size 2 μm) using an Ultimate 3000 RSLC nano-HPLC system coupled online to a hybrid LTQ Orbitrap XL mass spectrometer via a nano-ESI source equipped with a 30 μm i.d., 40-mm-long steel emitter (all Thermo Fisher Scientific, Bremen, Germany). The eluents A and B were 0.1% (*v/v*) aq. FA and 0.08% (*v/v*) FA in acetonitrile, respectively. The peptides were eluted with linear gradients ramping from 1 to 35% B over 90 min followed by 35–85% eluent B over 5 min. The column was washed for 5 min and re-equilibrated at 1% eluent B for 10 min. The nano-LC-MS analysis relied on data-dependent acquisition (DDA) experiments performed in the positive ion mode, comprising a survey Orbitrap-MS scan and MS/MS scans for the most abundant signals in the following 5 sec (at certain t_{R}) with charge states ranging from 2 to 6. The mass spectrometer settings and DDA parameters are summarized in the (Table S20). Identification of peptides and annotation of proteins relied on the search with the SEQUEST engine (run under Proteome Discoverer 2.2 software, Thermo Fisher Scientific, Bremen, Germany) against the pea genome database (Kreplak et al., 2019). Data processing (label-free quantification) and post-processing (statistical interpretation) were accomplished as described in (Table S21).

Analysis of primary metabolites

Analysis of the primary metabolites relied on the water–methanol extraction of samples (100 \pm 5 mg) as described by Chantseva et al. (2019) with minor changes. In detail, aliquots (50 μl) prepared without removal of lipophilic metabolites with hexane, were freeze-dried overnight. These were adjusted in a series of preliminary optimization experiments with 6.25, 12.5, 25, 50, 100, and 200 μl of a pooled extract. The residues were sequentially derivatized with methoxyamine hydrochloride (MOA) and *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA) according to Chantseva et al. (2019). The samples (1 μl) were analyzed by gas chromatography–electron ionization–quadrupole–mass spectrometry (GC-El-Q-MS) using a GCMS-QP2010 Ultra system under the control of GCMS Real Time Analysis software (Shimadzu Deutschland GmbH, Duisburg, Germany) and equipped with a CTC GC PAL Liquid Injector (Shimadzu Deutschland GmbH, Duisburg, Germany) and ZB-5MS capillary column (30 m \times 0.25 mm ID, 0.25 μm film thickness, Phenomenex, Aschaffenburg, Germany) under the instrumental settings summarized in Table S22.

Analysis of the cell wall-bound phenolics and related compounds

Isolation of cell walls was done according to Frolov et al. (2013) with some changes. In detail, 100 mg of SC samples was extracted three times with 1.8 ml of 80% (*v/v*) aq. MeOH by vortexing for 3 min followed by sonication for 10 min (Sonorex Super RK 510, Bandelin electronic, Berlin, Germany). The residues were sequentially extracted with 3 \times 1.8 ml of 70% (*v/v*) MeOH, water, 0.5% (*w/v*) SDS, 1 mol L⁻¹ aq. NaCl, water, MeOH, acetone, and *n*-hexane. After each extraction, the suspensions were centrifuged for 15 min at 3172 *g* (Eppendorf Centrifuge 5417R, Eppendorf AG, Hamburg, Germany), the supernatants were discarded and the pellets were re-suspended. After the last extraction, the residues were dried under air at room temperature (RT). The resulting material was defined as “cell wall preparation” and was supposed to be depleted of intracellular components (Frolov et al., 2013). For cleavage of ester bonds, 10 mg of each cell wall preparation was placed in amber glass vials, 1.5 ml of 1 mol L⁻¹ NaOH was added, the vials were sealed and nitrogen was purged for 10 min

before incubation in darkness for 20 h at 80°C. Afterward, the hydrolyzed samples (both aq. phase and solid residues) were quantitatively transferred to 2-ml polypropylene tubes by washing vials twice with 200 μl of 1 mol L⁻¹ aq. NaOH and centrifuged at 16 000 *g* at RT. The supernatants were transferred to 2-ml tubes, pH was adjusted to 3 with hydrochloric acid and samples were extracted three times with 6 ml of ethyl acetate supplemented with 2.2 μmol L⁻¹ 5-methylresorcinol (internal standard). The pooled organic fractions were dried under vacuum in rotary evaporator, reconstituted in 200 μl of 80% MeOH (aq.), and stored at -20°C. The pellets of the non-hydrolyzed residues were washed twice with MeOH, lyophilized overnight, transferred to glass vials, weighted, and supplemented with 1.5 ml of the dioxane – aq. 2 mol L⁻¹ HCl (9:1) mixture for hydrolysis of ether bonds. After purging with nitrogen for 10 min, the incubation at 90°C for 3 h was accomplished. The hydrolyzed samples were transferred to 2-ml polypropylene tubes and centrifuged as described above. The pH of the supernatants was adjusted to 3–4 with 1 mol L⁻¹ aq. sodium acetate and samples were extracted three times with 6 ml of dichloromethane. The pooled organic fractions were dried under vacuum in rotary evaporator, reconstituted in 200 μl of MeOH, and stored at -20°C before the analysis.

The composition of alkali and acid hydrolysates was analyzed by reversed-phase ultra-high-performance chromatography coupled online to quadrupole–time-of-flight mass spectrometry (RP-UHPLC-QqTOF-MS). Thereby, the chromatographic separations were performed with ACQUITY I-Class UPLC system (Waters, Eschborn, Germany) on Nucleoshell C1-8 column (150 \times 2 mm, 2.7 μm particle size, Macherey-Nagel, Düren, Germany) under the conditions specified in Table S23. The column effluents were transferred online in a hybrid QqTOF mass spectrometer TripleTOF 6600 (AB Sciex, Darmstadt, Germany) equipped with DuoSpray™ Ion Source, controlled by Analyst TF 1.7.1 software and operated in negative ion mode with sequential window acquisition of all theoretical fragment ion spectra (SWATH). Additional MS/MS experiments were performed as product ion scans when no clearly interpretable SWATH MS/MS spectra could be acquired. The MS settings in all experiments are summarized in Table S12. The primary data interpretation relied on the untargeted approach. Thereby, peak peaking, alignment, filtering, and relative quantification used MSDial software (<http://prime.psc.riken.jp/compms/msdial/main.html>), whereas exploration of the spectral data, calculation of elemental composition and interpretation of MS/MS spectra were accomplished in the PeakView™ (version 2.2) software (AB Sciex). Verification of peak integration was done by MultiQuant™ (version 3.0.2) software (AB Sciex). The data post-processing and statistical interpretation relied on the MetaboAnalyst 4.0 online software tool (free available via www.metaboanalyst.ca).

Untargeted metabolite analysis of the seed coat

Seed coats were ground to powder using a ball mill (Pulverisette 23, Fritsch, Germany). A mixture of 10 mg of powder and 1000 μl of extraction solution (acetone:water, 70:30) was sonicated for 1 h and shaken for 2 h. Extracts were centrifuged (18 400 *g*, 5 min) and supernatants were dried under nitrogen. Dried residues were dissolved in water:acetonitrile mixture (50:50) with the addition of 0.1% formic acid (v/v/v). High-resolution tandem mass spectrometer with ESI ion source (Z-spray) and QqTOF analyzer equipped with cyclic ion mobility cell (Select Series Cyclic IMS, Waters, Milford, MA, USA) were used in both positive and negative ionization modes. Masslynx software (version 4.2, Waters) was applied for system control and data treatment. Analyses were done in flow-injection analysis (FIA). Confirmation of the differential markers

identity was performed by direct infusion of SC extracts into ion source and by liquid chromatography/mass spectrometry analysis (LC/MS). Mass range 50–1200 Da, scan time 1 sec, capillary voltage 2.0 or 2.5 V, respectively, desolvation gas flow 800 L h⁻¹, nebulizer gas pressure 6 bar were used for all experimental arrangements. Instrument calibration was done using sodium formate solution (0.5 mmol L⁻¹, dissolved in 90:10 2-propanol/water). Leucine-enkephalin (50 μg μl^{-1} , dissolved in 50:50 acetonitrile/water +0.1% formic acid, v/v/v) was used for lock mass correction using a reference electrospray probe. FIA-MS analysis was performed using an ultra-performance liquid chromatography system ACQUITY I-Class (UPLC, Waters) interfaced to the Cyclic IMS mass spectrometer. Ten microliters of sample was injected using autosampler into the flow of carrier liquid (water: acetonitrile, 30:70, v/v with 0.1% formic acid, v/v) and delivered to the electrospray source at a flow rate of 0.1 ml min⁻¹. Run time was set at 5 min. MS parameters were set as follows: source temperature 100°C, desolvation temperature 250°C, trap and transfer collision energy was set at 6 and 4 eV, respectively. MS spectra were acquired in the range of 50–1200 Da with TOF settings for V-mode. Time of analysis was set at 3 min. Confirmatory collision experiments providing high number of MS scans in particular analyses were performed in MS/MS and MS/cIMS/MS modes (cIMS denotes the ion mobility separation of precursor ions in cyclic ion mobility cell) by direct infusion of extracts into ion source (10 ml min⁻¹). Collision spectra were collected for 1.5 min. ESI source parameters were set as follows: source temperature at 120°C, desolvation temperature at 220°C. For MS/MS experiments (cIMS cell not active), trap collision energy (TrapCE) was tuned in the range 10–35 eV, to achieve a sufficient yield of fragments and transfer collision energy (TransferCE) was set at 4 eV. LM resolution was set at 10 (arbitrary units). Lock mass spectra were collected at the end of each analysis for 15 sec. In the case of several metabolites (i.e., signals at *m/z* 619.1014; 631.1017; 641.1139; 771.1346; 299.0841; 701.1907), ion mobility separation was applied to increase the selectivity of MS experiments. MS/cIMS/MS was performed in single/multipass mode and cIMS parameters were set as follows: pushes per bin: 1, number of bins: 200, cyclic TW velocity: 375 m sec⁻¹, array TW velocity: 375 m sec⁻¹, TW ramp start height: 15 V, TW ramp end height: 35 V, TW ramping rate: 2.5 V msec⁻¹ and TW static height was selected depending on studied structures. Cyclic IMS sequence was performed in advanced mode and duration times of sequence parts were set as follows: injection time: 10 msec, separation time was edited to achieve required ion mobility separation with sufficient signal intensity in the same time, eject to pre-store time: 0.5 msec, eject time: 0.5 msec, reinject from pre-store time: 0.5 msec and eject and acquire time: 13.2 msec. Cyclic ion mobility separation was, thus, realized by tuning of two aforementioned parameters: TW static height and separation time in cyclic sequence. TrapCE was set at 6 eV and TransferCE was selected depending on studied structures. Tuned parameters are given in Table S24. All chemicals were purchased from Sigma-Aldrich (St. Louis, MO, USA). Ultrapure water was obtained from Milli-Q apparatus (Merck, Kenilworth, NJ, USA).

For LC/MS analysis, the hyphenation of ACQUITY UPLC I-class LC system with cyclic IMS mass spectrometer was used. The chromatographic separation was performed using Raptor ARC-18 column, 100 mm \times 2.1 mm, *dp* = 2.7 μm , Restek (reversed-phase mode). The volume injected into the column was 10 μl . Mobile phases, A: water with 0.1% formic acid (v/v) and B: acetonitrile with 0.1% formic acid (v/v), were used and the following gradient was used: 0 min – 95% A, 14 min – 0% A, 18 min – 0% A, 20 min – 95% A. Flow rate 0.25 ml min⁻¹ was applied.

Collected raw FIA-MS data were processed by MassLynx software (Waters) which comprised data extraction, normalization, and alignment (creation of data matrix). MarkerLynx parameters were set as follows: marker intensity threshold at 20 000 counts, peak separation at 0.05 Da, mass range was 50–1200 Da. A combined scan range type of analysis was selected. Background signal subtraction was applied as follows: scans to combine: 130–230, scans to subtract: 60–110 and 250–290. Prepared datasets were analyzed by multivariate statistics software EZinfo (version 3.0, Umetrics, Malmö, Sweden). Principal component analysis (PCA) and orthogonal projections to latent structures discriminant analysis (OPLS-DA) were used to reveal significantly differential signals and construct S-plots. Signals with the highest and lowest values of y-coordinates in S-plots (i.e., $p(\text{corr})[1]$, Correlation) were defined as markers, and differences in their intensities among particular genotypes were studied using Trend View tool in MarkerLynx. Normalized intensities of signals at particular m/z values were used for heatmap creation. The normalized values were calculated by dividing the intensity of selected signal of metabolite by the sum of intensities of all signals in the studied MS spectrum. The maximal and minimal normalized intensities were determined for each metabolite in the whole dataset (including all genotypes and all developmental stadia) and their differences were divided by the average value of all normalized intensities. The range of its intensities was divided into quarters and each quarter was then expressed by different color in the order of increasing intensity—green, light green, yellow, and red. These colors were then ascribed to each metabolite in particular genotype and developmental stadium according to its normalized intensity.

AUTHOR CONTRIBUTIONS

PS conceived and coordinated the study. PB and AF coordinated metabolomic and proteomic analysis. JB, BK, LK, PK, MZC, OT, TL, and CI performed the experiments. DG, AO, and OT contributed to bioinformatic data analysis. JB, BK, PK, OT, PB, AF, and PS contributed to manuscript preparation. All authors read and approved the manuscript. BK and JB contributed equally to this work.

ACKNOWLEDGEMENT

This work was supported by Grant Agency of Czech Republic 19-07155S, 24-10730S projects and from the project TowArds Next GENeration Crops, reg. no. CZ.02.01.01/00/22_008/0004581 of the ERDF Programme Johannes Amos Comenius. The work of A. Frolov group was infrastructurally supported by the Ministry of Science and Higher Education of the Russian Federation (theme # 122042700043-9). Open access publishing facilitated by Univerzita Palackeho v Olomouci, as part of the Wiley - CzechELib agreement.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Excel- GO_terms_MF_selected_WGCNA_modules.

Table S2. Excel- GO_terms_MF_DEGs_UP_and_DOWN.

Table S3. Excel- GO_terms_MF_DEGs_summary.

Table S4. Excel- List of DEGs involved in flavonoid pathway found in WILD gene set.

Table S5. Protein recoveries calculated for individual pea protein samples. Numbers 1, 2, 3 denote treatment groups corresponding to seed developmental stages (D1, D2 and mature seeds, respectively). Letters a–d denote biological replicates within the treatment groups.

Table S6. Excel- Annotation of proteins differentially expressed in wild and domesticated pea seed coat samples.

Table S7. Primary metabolites identified by spectral similarity library search and/or co-elution with authentic standards in pea seed coats aqua methanolic extracts. Metabolite analysis relied on GC-ESI-Q-MS analysis after derivatization of the lyophilized extracts with methoxamine hydrochloride (MOA) and *N*-methyl-*N*-(trimethylsilyl)trifluoroacetamide (MSTFA).

Table S8. Primary metabolites detected in the aq. methanolic extracts of mature Cameor seed coats demonstrating statistically significant up- and down-regulation in comparison to those of wild JI261.

Table S9. Primary metabolites of mature JI92 seed coats demonstrating statistically significant up- and down-regulation in comparison with those of wild JI261.

Table S10. Primary metabolites of mature JI1794 seed coats demonstrating statistically significant up- and down-regulation in comparison with those of JI261.

Table S11. Primary metabolites of mature JI64 seed coats demonstrating statistically significant up- and down-regulation in comparison with those of JI261.

Table S12. Mass analyzer settings applied for QqTOF-MS experiments in analysis of seed coat (cell wall) hydrolyzates and reference authentic standards.

Table S13. Cell wall-bound metabolites extracted from the seed coats of wild (JI64, JI1794, JI261) and domesticated (Cameor, JI92) peas upon alkali hydrolysis of corresponding isolated and purified cell wall material.

Table S14. Excel- Coordinates of markers in S-plot obtained from OPLS-DA analysis (FIA-ESI-HRTMS, negative ionization, lock mass uncorrected).

Table S15. List of identified significantly differential metabolites rising during seed coat development.

Table S16. List of identified significantly differential metabolites decreasing during seed coat development (positive ionization mode).

Table S17. List of identified metabolites with significantly higher content in wild compared cultivated genotypes in older developmental stages (D5-6).

Table S18. Excel- Expression of genes encoding enzymes of monolignol pathway in seed coats (SC) and embryos (E) of domesticated (Cameor, JI92 and *Pisum abyssinicum* PI358617) and wild (JI64, JI1794, JI261) peas over five seed developmental stages (13, 17, 20, 23, 28 DAP, labelled as 1-5). PAL: phenylalanine ammonia-lyase, C4H: cinnamate-4-hydroxylase, 4CL: 4-coumaroyl-CoA ligase, HCT: hydroxycinnamoyl CoA:shikimate hydroxycinnamoyltransferase, COMT: caffeic acid O-methyltransferase, CSE: caffeoyl shikimate esterase, CAD: cinnamyl alcohol dehydrogenase, CCR: cinnamoyl CoA reductase, CCoAMT: caffeoyl CoA-3-methyltransferase, F5H: ferulate-5-hydroxylase

Table S19. Studied metabolites of phenylpropanoid pathway.

Table S20. Instrument settings used in the proteomics LIT-Orbitrap-MS and -MS/MS experiments.

Table S21. Procedures and specific settings for data processing and post-processing of the proteomics data.

Table S22. Gas chromatographic (GC) separation conditions and electron ionization-quadrupole-mass spectrometry (EI-Q-MS)

settings for GC-ESI-Q-MS analysis of the primary metabolites in pea seed coats.

Table S23. Chromatographic conditions used for UHPLC separation of seed coat (cell wall) hydrolyzates and reference authentic standards.

Table S24. Variable parameters of MS/cIMS/MS measurements.

Figure S1. The dynamics of gene expression between studied developmental stages within all genotypes (a) or among genotypes in particular developmental stages (b).

Figure S2. Twelve representative groups of transcription factors described within 20 gene modules of pea SC. Visualized by Cytoscape 3.9.0.

Figure S3. SDS-PAGE electropherograms of the total protein fractions isolated from the seed coats of JI92 (a, c, e) and JI64 (b, d, f) seeds before and after tryptic hydrolysis. Numbers 1, 2, 3 denote seed developmental stages: DS1, DS2 and mature seeds, respectively. Letters a-d denote biological replicates. The aliquots (10 µg) of samples before hydrolysis (a, b), the incompletely digested aliquots left on filter unit after peptide elution (c, d) and aliquots of tryptic hydrolysates (corresponding to 5 µg of protein), (e, f) were loaded on gels. Inter-gel normalization relied on the total density of the Protein Ladder (PageRuler™ Prestained Protein Ladder #26616, 10–180 kDa) lane (St); the ND (non-digested) sample represents a reference protein not subjected to hydrolysis.

Figure S4. The numbers of tryptic peptides (a), possible proteins (b), and non-redundant proteins (protein groups) (c) identified in domesticated JI92 seed coats at developmental stages D1, D2 and D6. The tryptic digests ($n = 3$) obtained from seed coats were analyzed by nano-high performance liquid chromatography-electrospray ionization linear ion trap-orbital trap mass spectrometry (nanoHPLC-ESI-LIT-Orbitrap-MS) operated in positive DDA mode.

Figure S5. The numbers of tryptic peptides (a), possible proteins (b), and non-redundant proteins (protein groups, c) identified in wild pea JI64 seed coats at D1, D2 and D6 stages. The tryptic digests ($n = 3$), obtained from pea seedlings, were analyzed by nano-high performance liquid chromatography-electrospray ionization linear ion trap-orbital trap mass spectrometry (nanoHPLC-ESI-LIT-Orbitrap-MS) operated in positive DDA mode.

Figure S6. Principal component analysis (PCA) with score plot representation (a) accomplished for seed coat proteins differentially expressed at developmental stages D1 and D2 and in the mature state (D6) and hierarchical clustering with a heatmap representation (b).

Figure S7. Functional annotation (accomplished with the Mercator MapMan v3.6 tool) of the pea seed coat proteins isolated in stage D1. White and black columns denote the functional groups of the proteins, which were more expressed in the developing seeds of domesticated JI92 and wild JI64, respectively.

Figure S8. Functional annotation (accomplished with the Mercator MapMan v3.6 tool) of the pea seed coat proteins isolated in stage D2. White and black columns denote the functional groups of the proteins, which were more expressed in the developing seeds of the domesticated JI92 and wild JI64, respectively.

Figure S9. Prediction of sub-cellular localization of the proteins more expressed in the developing seeds of JI92 and JI64 with the BUSCA prediction tool (a, b – stage D1; c, d – stage D2; e, f – stage D6, respectively); compartments abbreviated as: APO – apoplast, CW – cell wall, CHL – chloroplast, CHL_oM – chloroplast outer membrane, CHL_tM – chloroplast thylakoid membrane, CHL_tl – chloroplast thylakoid lumen, CYT – cytosole, EX – extracellular space, GA – Golgi apparatus, ER – endoplasmic reticulum, MIT –

mitochondrion, NUC – nucleus, oM – organelle membrane PL – plasma membrane.

Figure S10. Evaluation of the differences in the metabolic profiles of the mature seeds obtained from the wild JI261 and domesticated Cameor by principal component analysis (PCA) with score plot representation (a) and hierarchical clustering with heatmap representation (b).

Figure S11. Representation of the differences in the metabolic profiles of the mature seed coats obtained from the wild JI261 and domesticated Cameor by the t-test with Volcano plot representation (a) and the top 30 differentially abundant metabolites demonstrating the most pronounced differences of corresponding GC-MS signals associated with seed dormancy (b). The p values in the Volcano plot were corrected for multiple comparisons, i.e. false discovery rates (FDR) were determined by Benjamini-Hochberg criterion (Benjamini-Hochberg method). The colored dots represent metabolites showing statistically significant differences with a threshold level of fold change (FC) ≥ 2 at $P \leq 0.05$ compared to controls. The blue and red dots represent the metabolites which were down- (\downarrow) and up-regulated (\uparrow) in the seed coats of the mature Cameor seeds in comparison to those obtained from the wild JI261. The metabolites marked with grey dots showed no statistically significant differences.

Figure S12. Evaluation of the differences in the metabolic profiles of the mature seeds obtained from the wild JI261 and domesticated JI92 by principal component analysis (PCA) with score plot representation (a) and hierarchical clustering with heatmap representation (b). The top 30 differentially abundant metabolites demonstrating the most pronounced differences of corresponding GC-MS signals associated with seed dormancy (c) and representation of the differences in the metabolic profiles of the mature seeds obtained from the wild JI261 and domesticated JI92 by the Volcano plot (d). The P values in the Volcano plot were corrected for multiple comparisons, i.e. false discovery rates (FDR) were determined by Benjamini-Hochberg criterion (Benjamini-Hochberg method). The colored dots represent metabolites showing statistically significant differences with a threshold level of fold change (FC) ≥ 2 at $P \leq 0.05$ compared to controls. The blue and red dots represent the metabolites which were down- (\downarrow) and up-regulated (\uparrow) in the mature JI92 seeds in comparison to those obtained from the wild JI261. The metabolites marked with gray dots showed no statistically significant differences.

Figure S13. Principal component analysis (PCA) illustrating distribution of metabolic profiles of mature seed coats of two wild pea genotypes, JI1794 and JI261 (a). Heatmap shows all (b) and the 30 top abundant metabolites (c). Volcano diagrams (d) characterize the differences between the levels of seed coat metabolites of JI1794 compared with those of JI261; x axis indicates multiple changes in the content of metabolites (FC - fold change), y -axis shows statistical significance of multiple changes (negative decimal logarithm of P -value). P -values were corrected by multiple comparison correction FDR (False Discovery Rate) using Benjamini-Hochberg criterion (Benjamini-Hochberg method). Colored dots represent metabolites showing statistically significant differences with a threshold level of FC ≥ 2 at P -value ≤ 0.05 compared to controls, namely, blue dots represent metabolites with lower levels compared to controls; red dots represent metabolites with higher levels compared to controls. Metabolites marked with gray dots have no statistically significant differences. The legend shows the number of metabolites whose levels increase (\uparrow) or decrease (\downarrow) compared to controls.

Figure S14. Principal component analysis (PCA) demonstrates the distribution of mature seed coat metabolic profiles of two wild

pea genotypes, JI64 and JI261(control) (a). Heatmap represents all (b) and the 30 top abundant metabolites (c). Volcano diagrams (d) characterize the differences between the levels of seed coat metabolites of JI64 compared with those of JI261, x axis - multiple changes in the content of metabolites (FC - fold change), y-axis - statistical significance of multiple changes (negative decimal logarithm of P -value). P -values were corrected by multiple comparison correction FDR (False Discovery Rate) using Benjamini-Hochberg criterion (Benjamini-Hochberg method). Colored dots represent metabolites showing statistically significant differences with a threshold level of $FC \geq 2$ at P -value ≤ 0.05 compared to controls, namely, blue dots represent metabolites with lower levels compared to controls; red dots represent metabolites with higher levels compared to controls. Metabolites marked with gray dots have no statistically significant differences. The legend shows the number of metabolites whose levels increase (\uparrow) or decrease (\downarrow) compared to controls.

Figure S15. Evaluation of the differences in the patterns of the cell wall-bound metabolites obtained from mature seed coats of wild JI261 and domesticated Cameor: principal component analysis (PCA) with score plot representation (a), hierarchical clustering with heatmap representation (b) and t -test analysis with the Volcano-plot representation (c).

Figure S16. Statistical analysis (t -test with Volcano plot representation) characterizing the differences between the levels of mature seed coat cell wall-bound metabolites of Cameor compared with those of wild JI261. x axis represents the multiple changes in the contents of metabolites (FC - fold change), whereas the y-axis shows the statistical significance of the corresponding individual changes (negative decimal logarithm of P). The resulting P -values were adjusted by the multiple comparison correction FDR (False Discovery Rate) procedure using the Benjamini-Hochberg criterion (Benjamini-Hochberg method). Coloured dots represent metabolites showing statistically significant differences with a threshold level of $FC \geq 10$ at $P \leq 0.1$ compared to the JI261. Namely, blue dots correspond to the metabolites demonstrating the lower levels compared to JI261. The legend shows the number of metabolites with the contents in the seed coats of the Cameor seeds higher (\uparrow) or lower (\downarrow) in comparison to the JI261.

Figure S17. Principal component analysis (PCA) illustrates the distribution of mature seed coat metabolic profiles of domesticated JI92 and wild JI261 (a). Heatmap shows all metabolites (b), Volcano diagrams (c, d) characterize the differences between the levels of seed coat metabolites of JI92 compared with those of JI261, x axis - multiple changes in the content of metabolites (FC - fold change), y-axis - statistical significance of multiple changes (negative decimal logarithm of P -value). P -values were corrected by multiple comparison correction FDR (False Discovery Rate) using Benjamini-Hochberg criterion (Benjamini-Hochberg method). Colored dots represent metabolites showing statistically significant differences with a threshold level of $FC \geq 2$ (c) and $FC \geq 4$ (d) at P -value ≤ 0.1 compared to controls, namely, blue dots represent metabolites with lower levels compared to controls; red dots represent metabolites with higher levels compared to controls. Metabolites marked with gray dots have no statistically significant differences. The legend shows the number of metabolites whose levels increase (\uparrow) or decrease (\downarrow) compared to controls.

Figure S18. Principal component analysis (PCA) shows the distribution of metabolic profiles of mature seed coats of two wild pea genotypes, JI1794 and JI261, control, (a). Heatmap shows a representation of all metabolites (b), Volcano diagram (c) characterizing the differences between the levels of seed coat metabolites of mature JI1794 seeds compared with those of JI261, x axis -

multiple changes in the content of metabolites (FC - fold change), y-axis - statistical significance of multiple changes (negative decimal logarithm of P -value). P -values were corrected by multiple comparison correction FDR (False Discovery Rate) using Benjamini-Hochberg criterion (Benjamini-Hochberg method). Colored dots represent metabolites showing statistically significant differences with a threshold level of $FC \geq 2$ at P -value ≤ 0.1 compared to controls, namely, blue dots represent metabolites with lower levels compared to controls; red dots represent metabolites with higher levels compared to controls. Metabolites marked with gray dots have no statistically significant differences. The legend shows the number of metabolites whose levels increase (\uparrow) or decrease (\downarrow) compared to controls.

Figure S19. Principal component analysis (PCA) demonstrates the distribution of mature seed coat metabolic profiles of two wild genotypes, JI64 and control JI261 (a). Heatmap representation of all metabolites (b). Volcano diagram (c) characterizing the differences between the levels of mature seed coat metabolites of JI64 compared with those of JI261, x axis - multiple changes in the content of metabolites (FC - fold change), y-axis - statistical significance of multiple changes (negative decimal logarithm of P -value). P -values were corrected by multiple comparison correction FDR (False Discovery Rate) using Benjamini-Hochberg criterion (Benjamini-Hochberg method). Colored dots represent metabolites showing statistically significant differences with a threshold level of $FC \geq 2$ at P -value ≤ 0.1 compared to controls, namely, blue dots represent metabolites with lower levels compared to controls; red dots represent metabolites with higher levels compared to controls. Metabolites marked with gray dots have no statistically significant differences. The legend shows the number of metabolites whose levels increase (\uparrow) or decrease (\downarrow) compared to controls.

Figure S20. Annotated cell wall-bound metabolites extracted from the seed coats of the dormant wild pea genotype JI261 and the seed coats from several pea genotypes varying in their dormancy (Cameor, JI92, JI64, and JI1794) upon alkali hydrolysis of corresponding isolated and purified cell wall material. The analysis relied on RP-UHPLC-QqTOF-MS accomplished with a Waters ACQUITY I-Class UPLC System (Waters GmbH, Eschborn, Germany) coupled on-line to a Triple-TOF6600 hybrid mass spectrometer (Sciex, Darmstadt, Germany) in the negative ion mode. The MS/MS spectra of individual metabolites are listed as Figure S20a–o. The number in the column behind a metabolite refers to the order in Table S13.

Figure S21. Ion mobility separation of m/z 299.0841. Mobilogram (a) of compounds with nominal mass 299. Peak with arrival time at 1.32 msec belongs to the studied structure of hydroxybenzoylglucoside with signal at m/z 299.0841. View to the signal at m/z 299.0841 in the zoomed spectrum without cyclic ion mobility separation (b). View to the signal at m/z 299.0841 in the zoomed spectrum with cyclic ion mobility separation (c). Spectrum was combined in the mobilogram range from 1.20 to 1.80 msec. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 299.0841 (d). Spectrum covers the full mobility range. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 299.0841 (e). Spectrum was combined in the mobilogram range from 1.20 to 1.80 msec.

Figure S22. Ion mobility separation of m/z 701.1907. Mobilogram (a) of compounds with nominal mass 701. Peak with arrival time at 9.50 msec belongs to the studied structure of 3',4',5'-trihydroxy-5,7,8-trimethoxy-2,3-dihydroflavone dihexoside with signal at m/z 701.1907. View to the signal at m/z 701.1907 in the

zoomed spectrum without cyclic ion mobility separation (b). View to the signal at m/z 701.1907 in the zoomed spectrum with cyclic ion mobility separation (c). Spectrum was combined in the mobilogram range from 9.20 to 10.00 msec. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 701.1907 (d). Spectrum covers the full mobility range. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation (e) in Transfer collision cell of m/z 701.1907. Spectrum was combined in the mobilogram range from 9.20 to 10.00 msec.

Figure S23. Ion mobility separation of m/z 619.1041. Mobilogram (a) of compounds with nominal mass 619. Peak with arrival time at 6.90 msec belongs to the studied structure of galocatechin galate hexoside with signal at m/z 619.1041. View to the signal at m/z 619.1041 in the zoomed spectrum without cyclic ion mobility separation (b). View to the signal at m/z 619.1041 in the zoomed spectrum with cyclic ion mobility separation (c). Spectrum was combined in the mobilogram range from 5.50 to 8.00 msec. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 619.1041 (d). Spectrum covers the full mobility range. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 619.1041 (e). Spectrum was combined in the mobilogram range from 5.50 to 8.00 msec.

Figure S24. Ion mobility separation of m/z 631.1017. Mobilogram (a) of compounds with nominal mass 631. Peak with arrival time at 1.65 msec belongs to the studied structure of myricetin 3-galloyl hexoside with signal at m/z 631.1017. View to the signal at m/z 631.1017 in the zoomed spectrum without cyclic ion mobility separation (b). View to the signal at m/z 631.1017 in the zoomed spectrum with cyclic ion mobility separation (c). Spectrum was combined in the mobilogram range from 1.75 to 2.00 msec. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 631.1017 (d). Spectrum covers the full mobility range. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 631.1017 (e). Spectrum was combined in the mobilogram range from 1.75 to 2.00 msec.

Figure S25. Ion mobility separation of m/z 641.1139. Mobilogram (a) of compounds with nominal mass 641. Peak with arrival time at 1.78 msec belongs to the studied structure of myricetin 3-cafeoyl hexoside with signal at m/z 641.1139. View to the signal at m/z 641.1139 in the zoomed spectrum without cyclic ion mobility separation (b). View to the signal at m/z 641.1139 in the zoomed spectrum with cyclic ion mobility separation (c). Spectrum was combined in the mobilogram range from 1.40 to 2.00 msec. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 641.1139 (d). Spectrum covers the full mobility range. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 641.1139 (e). Spectrum was combined in the mobilogram range from 1.40 to 2.00 msec.

Figure S26. Ion mobility separation of m/z 771.1346. Mobilogram (a) of compounds with nominal mass 771. Peak with arrival time at 2.90 msec belongs to the studied structure of dehydrated galocatechin-trihexoside with signal at m/z 771.1346. View to the signal at m/z 771.1346 in the zoomed spectrum without cyclic ion mobility separation (b). View to the signal at m/z 771.1346 in the zoomed spectrum with cyclic ion mobility separation (c). Spectrum was combined in the mobilogram range from 2.20 to 3.50 msec. MSMS spectrum with ion mobility separation and

fragmentation after collision induced activation in Transfer collision cell of m/z 771.1346 (d). Spectrum covers the full mobility range. MSMS spectrum with ion mobility separation and fragmentation after collision induced activation in Transfer collision cell of m/z 771.1346 (e). Spectrum was combined in the mobilogram range from 2.20 to 3.50 msec.

Figure S27. Reconstructed chromatograms of p-hydroxybenzoic and salicylic acids in DS5 of dormant JI64 and domesticated landraces JI92 (LC/HRTMS, negative ionization mode).

Figure S28. Module-trait relationship depiction showing the correlation between expression of the gene modules and the abundance of identified metabolites of the monolignol pathway.

OPEN RESEARCH BADGES



This article has earned an Open Data badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at NAsq data were deposited to Gene Expression Omnibus of the National Center for Biotechnology Information (NCBI) <https://www.ncbi.nlm.nih.gov/geo/under> the accession number GSE244961. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (<https://www.proteomexchange.org>) with the dataset identifier PXD047224 and 10.6019/PXD047224.

DATA AVAILABILITY STATEMENT

All RNA-seq data generated from this study have been deposited into the Gene Expression Omnibus of the National Center for Biotechnology Information (NCBI) under the accession number GSE244961. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE (Deutsch et al., 2023) partner repository with the dataset identifier PXD047224 and 10.6019/PXD047224.

REFERENCES

- Akram, N.A., Shafiq, F. & Ashraf, M. (2017) Ascorbic acid—a potential oxidant scavenger and its role in plant development and abiotic stress tolerance. *Frontiers in Plant Science*, **8**, 613.
- Alosekh, S., Scossa, F., Wen, W., Luo, J., Yan, J., Beleggia, R. et al. (2021) Domestication of crop metabolomes: desired and unintended consequences. *Trends in Plant Science*, **26**, 650–661. Available from: <https://doi.org/10.1016/j.tplants.2021.02.005>
- Alves-Carvalho, S., Aubert, G., Carrère, S., Cruaud, C., Brochot, A.L., Jacquinet, F. et al. (2015) Full-length de novo assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *The Plant Journal*, **84**, 1–19.
- Appelham, I., Thiedig, K., Nordholt, N., Schmidt, N., Huep, G., Sagasser, M. et al. (2014) Update on transparent testa mutants from *Arabidopsis thaliana*: characterisation of new alleles from an isogenic collection. *Planta*, **240**, 955–970.
- Bai, Y., Xiao, S., Zhang, Z., Zhang, Y., Sun, H., Zhang, K. et al. (2020) Melatonin improves the germination rate of cotton seeds under drought stress by opening pores in the seed coat. *PeerJ*, **8**, e9450.
- Balárynová, J., Klčová, B., Sekaninová, J., Kobrolová, L., Cechová, M.Z., Krejčí, P. et al. (2022) The loss of polyphenol oxidase function is associated with hilum pigmentation and has been selected during pea domestication. *New Phytologist*, **235**, 1807–1821.
- Balárynová, J., Klčová, B., Tarkovská, D., Turečková, V., Trnéný, O., Špundová, M. et al. (2023) Domestication has altered the ABA and gibberellin profiles in developing pea seeds. *Planta*, **258**, 25.
- Baroux, C. & Grossniklaus, U. (2019) Seeds—an evolutionary innovation underlying reproductive success in flowering plants. *Current Topics in Developmental Biology*, **131**, 605–642.

- Baskin, C.C. & Baskin, J.M. (2014) *Seeds: ecology, biogeography and evolution of dormancy and germination*, 2nd edition. New York: Academic Press.
- Bellucci, E., Bitocchi, E., Ferrarini, A., Benazzo, A., Biagetti, E., Klie, S. *et al.* (2014) Decreased nucleotide and expression diversity and modified co-expression patterns characterize domestication in the common bean. *Plant Cell*, **26**, 1901–1912.
- Benjamini, Y. & Hochberg, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, **57**, 289–300.
- Berková, V., Berka, M., Kameniarová, M., Kopecká, R., Kuzmenko, M., Shejbalová, Š. *et al.* (2023) Salicylic acid treatment and its effect on seed yield and seed molecular composition of *Pisum sativum* under abiotic stress. *International Journal of Molecular Sciences*, **24**, 5454.
- Bi, S., Lao, F., Pan, X., Shen, Q., Liu, Y. & Wu, J. (2022) Flavor formation and regulation of peas (*Pisum sativum* L.) seed milk via enzyme activity inhibition and off-flavor compounds control release. *Food Chemistry*, **380**, 132203.
- Bizouerne, E., Buitink, J., Vu, B.L., Vu, J.L., Esteban, E., Pasha, A. *et al.* (2021) Gene co-expression analysis of tomato seed maturation reveals tissue-specific regulatory networks and hubs associated with the acquisition of desiccation tolerance and seed vigour. *BMC Plant Biology*, **21**, 124.
- Bjarnholt, N., Nakonieczny, M., Kędzior, A., Debinski, D.M., Matter, S.F., Olsen, C.E. *et al.* (2012) Occurrence of sarmenosin and other hydroxynitrile glucosides in *Parnassius* (Papilionidae) butterflies and their food plants. *Journal of Chemical Ecology*, **38**, 525–537.
- Boutet, S., Barreda, L., Perreau, F., Totozafy, J.C., Mauve, C., Gakière, B. *et al.* (2022) Untargeted metabolomic analyses reveal the diversity and plasticity of the specialized metabolome in seeds of different *Camelina sativa* genotypes. *The Plant Journal*, **110**, 147–165.
- Bradford, K. & Nonogaki, H. (2009) *Seed development, dormancy and germination*. *Annual Plant Reviews*, Vol. 27. Oxford: Blackwell.
- Buitink, J., Leger, J.J., Guisle, I., Vu, B.L., Wuilleme, S., Lamirault, G. *et al.* (2006) Transcriptome profiling uncovers metabolic and regulatory processes occurring during the transition from desiccation-sensitive to desiccation-tolerant stages in *Medicago truncatula* seeds. *The Plant Journal*, **47**, 735–750.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.
- Cechová, M., Hradilová, I., Smýkal, P., Barták, P. & Bednár, P. (2019) Utilization of atmospheric solids analysis probe mass spectrometry for analysis of fatty acids on seed surface. *Analytical and Bioanalytical Chemistry*, **411**, 1169–1180. Available from: <https://doi.org/10.1007/s00216-018-1551-3>
- Cechová, M., Válková, M., Hradilová, I., Janská, A., Soukup, A., Smýkal, P. *et al.* (2017) Towards better understanding of pea seed dormancy using laser desorption/ionization mass spectrometry. *International Journal of Molecular Sciences*, **18**, 2196.
- Chai, M., Queralta Castillo, I., Sonntag, A., Wang, S., Zhao, Z., Liu, W. *et al.* (2021) A seed coat-specific β -ketoacyl-CoA synthase, KCS12, is critical for preserving seed physical dormancy. *Plant Physiology*, **186**, 1606–1615.
- Chai, M., Zhou, C., Molina, I., Fu, C., Nakashima, J., Li, G. *et al.* (2016) A class II KNOX gene, *KNOX4*, controls seed physical dormancy. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 6997–7002.
- Chang, T., Zhao, Y., He, H., Xi, Q., Fu, J. & Zhao, Y. (2021) Exogenous melatonin improves growth in hullless barley seedlings under cold stress by influencing the expression rhythms of circadian clock genes. *PeerJ*, **9**, e10740. Available from: <https://doi.org/10.7717/peerj.10740>
- Chantseva, V., Bilova, T., Smolikova, G., Frolov, A. & Medvedev, S. (2019) 3D-clinorotation induces specific alterations in metabolite profiles of germinating *Brassica napus* L. seeds. *Biological Communications*, **64**, 55–74.
- Chen, Z., Zheng, Z., Huang, J., Lai, Z. & Fan, B. (2009) Biosynthesis of salicylic acid in plants. *Plant Signaling & Behavior*, **4**, 493–496.
- Conway, J.R., Lex, A. & Gehlenborg, N. (2017) UpSetR: an R package for the visualization of intersecting sets and their properties. *Bioinformatics*, **33**, 2938–2940.
- Corso, M., Perreau, F., Mouille, G. & Lepiniec, L. (2020) Specialized phenolic compounds in seeds: structures, functions, and regulations. *Plant Science*, **296**, 110471.
- Cosson, A., Meudec, E., Ginies, C., Danel, A., Lieben, P., Descamps, N. *et al.* (2022) Identification and quantification of key phytochemicals in peas – linking compounds with sensory attributes. *Food Chemistry*, **385**, 132615.
- de Souza Vidigal, D., Willems, L., van Arkel, J., Dekkers, B.J.W., Hilhorst, H.W.M. & Bentsink, L. (2016) Galactinol as marker for seed longevity. *Plant Science*, **246**, 112–118.
- De Tullio, M.C., Liso, R. & Arrigoni, O. (2004) Ascorbic acid oxidase: an enzyme in search of a role. *Biologia Plantarum*, **48**, 161–166.
- Debeaujon, I., Léon-Kloosterziel, K.M. & Koornneef, M. (2000) Influence of the testa on seed dormancy, germination, and longevity in *Arabidopsis*. *Plant Physiology*, **122**, 403–414.
- Deutsch, E.W., Bandeira, N., Perez-Riverol, Y., Sharma, V., Carver, J., Mendonza, L. *et al.* (2023) The ProteomeXchange Consortium at 10 years: 2023 update. *Nucleic Acids Research*, **51**(D1), D1539–D1548.
- DiCenzo, G.L. & VanEtten, H.D. (2006) Studies on the late steps of (+) pisatin biosynthesis: evidence for (–) enantiomeric intermediates. *Phytochemistry*, **67**, 675–683.
- Doebley, J.F., Gaut, B.S. & Smith, B.D. (2006) The molecular genetics of crop domestication. *Cell*, **127**, 1309–1321.
- Dos Santos, J.V., Ribeiro, P.R.A., Carneiro, M.A.B., Soares, I.C., Fiorini, I.V.A., Cancellier, L.L. *et al.* (2020) Formononetin accelerates mycorrhization and increases maize production at low phosphorus application rates. *Anais da Academia Brasileira de Ciências*, **92**(Suppl 1), e20181371.
- Du, J., Wang, S., He, C., Zhou, B., Ruan, Y.-L. & Shou, H. (2017) Identification of regulatory networks and hub genes controlling soybean seed set and size using RNA sequencing analysis. *Journal of Experimental Botany*, **68**, 1955–1972.
- Ferraro, K., Jin, A.L., Nguyen, T.-D., Reinecke, D.M., Ozga, J.A. & Ro, D.-K. (2014) Characterization of proanthocyanidin metabolism in pea (*Pisum sativum*) seeds. *BMC Plant Biology*, **14**, 238.
- Francoz, E., Lepiniec, L. & North, H.M. (2018) Seed coats as an alternative molecular factory: thinking outside the box. *Plant Reproduction*, **31**, 327–342.
- Frolov, A., Henning, A., Böttcher, C., Tissier, A. & Strack, D. (2013) An UPLC-MS/MS method for the simultaneous identification and quantitation of cell wall phenolics in *Brassica napus* seeds. *Journal of Agricultural and Food Chemistry*, **13**, 1219–1227.
- Fu, F., Zhang, W., Li, Y.-Y. & Wang, H.L. (2017) Establishment of the model system between phytochemicals and gene expression profiles in macrosclereid cells of *Medicago truncatula*. *Scientific Reports*, **7**, 2580.
- Fu, J., Frazee, A.C., Collado-Torres, L., Jaffe, A.E. & Leek, J.T. (2020) *ballgown*: flexible, isoform-level differential expression analysis. R package version 2.20.0.
- Gallardo, K., Firnhaber, C., Zuber, H., Hélicher, D., Belghazi, M., Henry, C. *et al.* (2007) A combined proteome and transcriptome analysis of developing *Medicago truncatula* seeds: evidence for metabolic specialization of maternal and filial tissues. *Molecular & Cellular Proteomics*, **6**, 2165–2179.
- Gao, H., Wang, Y., Li, W., Gu, Y., Lai, Y., Bi, Y. *et al.* (2018) Transcriptomic comparison reveals genetic variation potentially underlying seed developmental evolution of soybeans. *Journal of Experimental Botany*, **69**, 5089–5104.
- Gao, P., Quilichini, T.D., Yang, H., Li, Q., Nilsen, K.T., Qin, L. *et al.* (2022) Evolutionary divergence in embryo and seed coat development of U's Triangle *Brassica* species illustrated by a spatiotemporal transcriptome atlas. *New Phytologist*, **233**, 30–51.
- Garg, R., Singh, V.K., Rajkumar, M.S., Kumar, V. & Jain, M. (2017) Global transcriptome and coexpression network analyses reveal cultivar-specific molecular signatures associated with seed development and seed size/weight determination in chickpea. *The Plant Journal*, **91**, 1088–1107.
- Geshnizjani, N., Sarikhani Khorami, S., Willems, L.A.J., Snoek, B.L., Hilhorst, H.W.M. & Ligterink, W. (2019) The interaction between genotype and maternal nutritional environments affects tomato seed and seedling quality. *Journal of Experimental Botany*, **70**, 2905–2918.
- Gosch, C., Halbwirth, H. & Stich, K. (2010) Phloridzin: biosynthesis, distribution and physiological relevance in plants. *Phytochemistry*, **71**, 838–843.
- Grafi, G. (2020) Dead but not dead end: multifunctional role of dead organs enclosing embryos in seed biology. *International Journal of Molecular Sciences*, **21**, 8024.
- Greiffenhagen, U., Frolov, A., Blüher, M. & Hoffmann, R. (2016) Site-specific analysis of advanced glycation end products in plasma proteins of type 2

- diabetes mellitus patients. *Analytical and Bioanalytical Chemistry*, **408**, 5557–5566.
- Hammerbacher, A., Raguschke, B., Wright, L.P. & Gershenzon, J.** (2018) Galocatechin biosynthesis via a flavonoid 3',5'-hydroxylase is a defence response in Norway spruce against infection by the bark beetle-associated sap-staining fungus *Endoconidiophora polonica*. *Phytochemistry*, **148**, 78–86.
- Hanano, A., Blée, E. & Murphy, D.J.** (2023) Caleosin/Peroxygenases: multifunctional proteins in plants. *Annals of Botany*, **131**, 387–409. Available from: <https://doi.org/10.1093/aob/mcad001>
- Hart, A.J., Ginzburg, S., Xu, M., Fisher, C.R., Rahmatpour, N., Mitton, J.B. et al.** (2020) EnTAP: bringing faster and smarter functional annotation to non-model eukaryotic transcriptomes. *Molecular Ecology Resources*, **20**, 591–604.
- Haughn, G. & Chaudhury, A.** (2005) Genetic analysis of seed coat development in Arabidopsis. *Trends in Plant Science*, **10**, 472–477.
- Hellens, R.P., Moreau, C., Lin-Wang, K., Schwinn, K.E., Thomson, S.J., Fiers, M.W.E.J. et al.** (2010) Identification of Mendel's white flower character. *PLoS One*, **5**, e13230.
- Hellwig, T., Abbo, S. & Ophir, R.** (2022) Phylogeny and disparate selection signatures suggest two genetically independent domestication events in pea (*Pisum L.*). *The Plant Journal*, **110**, 419–439.
- Hradilová, I., Trněný, O., Váľková, M., Cechová, M., Janská, A., Prokešová, L. et al.** (2017) A combined comparative transcriptomic, metabolomic, and anatomical analyses of two key domestication traits: Pod dehiscence and seed dormancy in Pea (*Pisum sp.*). *Frontiers in Plant Science*, **8**, 542.
- Huss, J.C. & Gierlinger, N.** (2021) Functional packaging of seeds. *New Phytologist*, **230**, 2154–2163.
- Jang, S.-J., Sato, M., Sato, K., Jitsuyama, Y., Fujino, K., Mori, H. et al.** (2015) A single-nucleotide polymorphism in an Endo-1,4-β-Glucanase gene controls seed coat permeability in soybean. *PLoS One*, **10**, e0128527.
- Janská, A., Pecková, E., Szczepaniak, B., Smýkal, P. & Soukup, A.** (2019) The role of the testa during the establishment of physical dormancy in the pea seed. *Annals of Botany*, **123**(5), 815–829. Available from: <https://doi.org/10.1093/aob/mcy213>
- Jayakodi, M., Golicz, A.A., Kreplak, J., Fechete, L.I., Angra, D., Bednár, P. et al.** (2023) The giant diploid faba genome unlocks variation in a global protein crop. *Nature*, **615**, 652–659.
- Jha, A.B., Purves, R.W., Elesaway, F.M., Zhang, H., Vandenberg, A. & Warrentin, T.D.** (2019) Polyphenolic profile of seed components of white and purple flower pea lines. *Crop Science*, **59**, 2711–2719.
- Kaimoyo, E. & VanEtten, H.D.** (2008) Inactivation of pea genes by RNAi supports the involvement of two similar O-methyltransferases in the biosynthesis of (+)-piscatin and of chiral intermediates with a configuration opposite that found in (+)-piscatin. *Phytochemistry*, **69**, 76–87.
- Kampatsikas, I., Bijelic, A. & Rompel, A.** (2019) Biochemical and structural characterization of tomato polyphenol oxidases provide novel insights into their substrate specificity. *Scientific Reports*, **9**, 4022.
- Kassambara, A. & Mundt, F.** (2020) *factoextra*: extract and visualize the results of multivariate data analyses. R package version 1.0.7. Available from: <https://CRAN.R-project.org/package=factoextra> [Accessed 10th October 2023].
- Khan, D., Millar, J.L., Girard, I.J. & Belmonte, M.F.** (2014) Transcriptional circuitry underlying seed coat development in *Arabidopsis*. *Plant Science*, **219–220**, 51–60.
- Kim, D., Langmead, B. & Salzberg, S.L.** (2015) HISAT: a fast spliced aligner with low memory requirements. *Nature Methods*, **12**, 357–360.
- Kovnich, N., Saleem, A., Arnason, J.T. & Miki, B.** (2011) Combined analysis of transcriptome and metabolite data reveals extensive differences between black and brown nearly-isogenic soybean (*Glycine max*) seed coats enabling the identification of pigment isogenes. *BMC Genomics*, **12**, 381.
- Krejčí, P., Cechová, M.Z., Nádvořníková, J., Barták, P., Kobřilová, L., Balarynová, J. et al.** (2022) Combination of electronically driven micro-manipulation with laser desorption ionization mass spectrometry – the unique tool for analysis of seed coat layers and revealing the mystery of seed dormancy. *Talanta*, **242**, 123303.
- Kreplak, J., Madoui, M.-A., Čápal, P., Novák, P., Labadie, K., Aubert, G. et al.** (2019) A reference genome for pea provides insight into legume genome evolution. *Nature Genetics*, **51**, 1411–1422.
- Ku, Y.-S., Contador, C.A., Ng, M.-S., Yu, J., Chung, G. & Lam, H.-M.** (2020) The effects of domestication on secondary metabolite composition in legumes. *Frontiers in Genetics*, **11**, 581357.
- Langfelder, P. & Horvath, S.** (2008) WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, **9**, 559. Available from: <https://doi.org/10.1186/1471-2105-9-559>
- Le, B.H., Wagmaister, J.A., Kawashima, T., Bui, A.Q., Harada, J.J. & Goldberg, R.B.** (2007) Using genomics to study legume seed development. *Plant Physiology*, **144**, 562–574.
- Lenser, T. & Theißen, G.** (2013) Molecular mechanisms involved in convergent crop domestication. *Trends in Plant Science*, **18**, 704–714. Available from: <https://doi.org/10.1016/j.tplants.2013.08.007>
- Leonova, T., Ihling, C., Saoud, M., Frolova, N., Rennert, R., Wessjohann, L.A. et al.** (2022) Does filter aided sample preparation (FASP) provide sufficient method linearity for quantitative plant shotgun proteomics? *Frontiers in Plant Science*, **13**, 874761.
- Lepiniec, L., Debeaujon, I., Routaboul, J.-M., Baudry, A., Pourcel, L., Nesi, N. et al.** (2006) Genetics and biochemistry of seed flavonoids. *Annual Review of Plant Biology*, **57**, 405–430.
- Lewak, S.** (2011) Metabolic control of embryonic dormancy in apple seed: seven decades of research. *Acta Physiologiae Plantarum*, **33**, 1–24.
- Lewis, G., Schrire, B., Mackinder, B. & Lock, M.** (2005) *Legumes of the world*. Kew, Surrey, UK: Royal Botanic Gardens.
- Liu, C., Jun, J.H. & Dixon, R.A.** (2014) MYB5 and MYB14 play pivotal roles in seed coat polymer biosynthesis in *Medicago truncatula*. *Plant Physiology*, **165**(4), 1424–1439. Available from: <https://doi.org/10.1104/pp.114.241877>
- Liu, C.-W. & Murray, J.D.** (2016) The role of flavonoids in nodulation host-range specificity: an update. *Plants*, **5**, 33.
- Love, M.I., Huber, W. & Anders, S.** (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, **15**, 550.
- Lu, X., Li, Q.-T., Xiong, Q., Li, W., Bi, Y.D., Lai, Y.C. et al.** (2016) The transcriptomic signature of developing soybean seeds reveals the genetic basis of seed trait adaptation during domestication. *The Plant Journal*, **86**, 530–544.
- Lv, Y., Pan, J., Wang, H., Reiter, R.J., Li, X., Mou, Z. et al.** (2021) Melatonin inhibits seed germination by crosstalk with abscisic acid, gibberellin, and auxin in Arabidopsis. *Journal of Pineal Research*, **70**, e12736.
- Lyu, X., Li, Y.H., Li, Y., Li, D., Han, C., Hong, H. et al.** (2023) The domestication-associated L1 gene encodes a eucomic acid synthase pleiotropically modulating pod pigmentation and shattering in soybean. *Molecular Plant*, **16**(7), 1178–1191. Available from: <https://doi.org/10.1016/j.molp.2023.06.003>
- Ma, Q.H. & Xu, Y.** (2008) Characterization of a caffeic acid 3-O-methyltransferase from wheat and its function in lignin biosynthesis. *Biochimie*, **90**, 515–524.
- MacGregor, D.R., Kendall, S.L., Florance, H., Fedi, F., Moore, K., Paszkiewicz, K. et al.** (2015) Seed production temperature regulation of primary dormancy occurs through control of seed coat phenylpropanoid metabolism. *New Phytologist*, **205**, 642–652.
- Maeda, H.A. & Fernie, A.R.** (2021) Evolutionary history of plant metabolism. *Annual Review of Plant Biology*, **72**, 185–216.
- Maere, S., Heymans, K. & Kuiper, M.** (2005) BiNGO: a cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics*, **21**, 3448–3449.
- Mamontova, T., Afonin, A.M., Ihling, C., Soboleva, A., Lukasheva, E., Sulima, A.S. et al.** (2019) Profiling of seed proteome in pea (*Pisum sativum L.*) lines characterized with high and low responsivity to combined inoculation with nodule bacteria and arbuscular *Mycorrhizal fungi*. *Molecules*, **24**, E1603.
- Mamontova, T., Lukasheva, E., Mavropolo-Stolyarenko, G., Proksch, C., Bilova, T., Kim, A. et al.** (2018) Proteome map of pea (*Pisum sativum L.*) embryos containing different amounts of residual chlorophylls. *International Journal of Molecular Sciences*, **19**, E4066.
- McCarthy, D.J., Chen, Y. & Smyth, G.K.** (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research*, **40**(10), 4288–4297. Available from: <https://doi.org/10.1093/nar/gks042>
- North, H., Casey, R. & Domoney, C.** (1989) Inheritance and mapping of seed lipoyxygenase polypeptides in *Pisum*. *Theoretical and Applied Genetics*, **77**, 805–808.
- Oraby, H.F. & Ramadan, M.F.** (2015) Impact of suppressing the caffeic acid O-methyltransferase (*COMT*) gene on lignin, fiber, and seed oil

- composition in *Brassica napus* transgenic plants. *European Food Research and Technology*, **240**, 931–938. Available from: <https://doi.org/10.1007/s00217-014-2397-3>
- Paauw, M., Koes, R. & Quattrocchio, F.M. (2019) Alteration of flavonoid pigmentation patterns during domestication of food crops. *Journal of Experimental Botany*, **70**, 3719–3735.
- Pal, L., Dwivedi, V., Gupta, S.K., Saxena, S., Pandey, A. & Chattopadhyay, D. (2023) Biochemical analysis of anthocyanin and proanthocyanidin and their regulation in determining chickpea flower and seed coat colour. *Journal of Experimental Botany*, **74**, 130–148.
- Perteau, G. & Perteau, M. (2020) GFF utilities: GffRead and GffCompare. *F1000Research*, **9**, 304.
- Perteau, M., Kim, D., Perteau, G.M., Lee, J.T. & Salzberg, S.L. (2016) Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, **11**, 1650–1667.
- Perteau, M., Perteau, G.M., Antonescu, C.M., Chang, T.C., Mendell, T. & Salzberg, S.L. (2015) StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, **33**, 290–295.
- Purugganan, M.D. & Fuller, D.Q. (2009) The nature of selection during plant domestication. *Nature*, **457**, 843–848. Available from: <https://doi.org/10.1038/nature07895>
- Qin, G., Liu, C., Li, J., Qi, Y., Gao, Z., Zhang, X. *et al.* (2020) Diversity of metabolite accumulation patterns in inner and outer seed coats of pomegranate: exploring their relationship with genetic mechanisms of seed coat development. *Horticulture Research*, **7**, 1–14.
- R Core Team. (2022) *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/> [Accessed 22nd June 2023].
- Ranathunge, K., Shao, S., Outob, D., Gijzen, M., Peterson, C.A. & Bernards, M.A. (2010) Properties of the soybean seed coat cuticle change during development. *Planta*, **231**, 1171–1188.
- Rao, J., Cheng, F., Hu, C., Quan, S., Lin, H., Wang, J. *et al.* (2014) Metabolic map of mature maize kernels. *Metabolomics*, **10**, 775–787. Available from: <https://doi.org/10.1007/s11306-014-0624-3>
- Renard, J., Niñoles, R., Martínez-Almonacid, I., Gayubas, B., Mateos-Fernández, R., Bissoli, G. *et al.* (2020) Identification of novel seed longevity genes related to oxidative stress and seed coat by genome-wide association studies and reverse genetics. *Plant, Cell & Environment*, **43**, 2523–2539.
- Renzi, J.P., Duchoslav, M., Brus, J., Hradilová, I., Pechanec, V., Václavěk, T. *et al.* (2020) Physical dormancy release in *Medicago truncatula* seeds is related to environmental variations. *Plants (Basel)*, **9**, 503.
- Rochat, C., Wuilleme, S., Boutin, J.-P. & Hedley, C.L. (1995) A mutation at the *rb* gene, lowering ADPGPPase activity, affects storage product metabolism of pea seed coats. *Journal of Experimental Botany*, **46**, 415–421. Available from: <https://doi.org/10.1093/jxb/46.4.415>
- Salvi, P., Varshney, V. & Majee, M. (2022) Raffinose family oligosaccharides (RFOs): role in seed vigor and longevity. *Bioscience Reports*, **42**, BSR20220198.
- Sano, N., Rajjou, L., North, H.M., Debeaujon, I., Marion-Poll, A. & Seo, M. (2016) Staying alive: molecular aspects of seed longevity. *Plant Cell Physiology*, **57**, 660–674.
- Sedláková, V., Hanáček, P., Grulichová, M., Zablazková, L. & Smýkal, P. (2021) Evaluation of seed dormancy, one of the key domestication traits in chickpea. *Agronomy*, **11**, 2292.
- Sedláková, V., Zeljković, S.Č., Štefelová, N., Smýkal, P. & Hanáček, P. (2023) Phenylpropanoid content of chickpea seed coats in relation to seed dormancy. *Plants (Basel)*, **12**, 2687.
- Servillo, L., Castaldo, D., Giovane, A., Rosario, C., D'Onofrio, N., Cautela, D. *et al.* (2018) Ophthalmic acid is a marker of oxidative stress in plants as in animals. *Biochimica et Biophysica Acta*, **1862**, 991–998.
- Shah, J. (2003) The salicylic acid loop in plant defence. *Current Opinion in Plant Biology*, **6**, 365–371.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D. *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, **13**, 2498–2504.
- Smýkal, P., Coyne, C.J., Ambrose, M.J., Maxted, N., Schaefer, H., Blair, M.W. *et al.* (2015) Legume crops phylogeny and genetic diversity for science and breeding. *Critical Reviews in Plant Sciences*, **34**, 43–104.
- Smýkal, P., Hradilová, I., Trněný, O., Brus, J., Rathore, A., Bariotakis, M. *et al.* (2017) Genomic diversity and macroecology of the crop wild relatives of domesticated pea. *Scientific Reports*, **7**, 17384.
- Smýkal, P., Nelson, M.N., Berger, J.D. & Von Wettberg, E.J.B. (2018) The impact of genetic changes during crop domestication on healthy food development. *Agronomy*, **8**, 26.
- Smýkal, P. & Parker, T. (2023) Domestication-related changes in seed dispersal and pigmentation: visual selection and functional trait? *Molecular Plant*, **16**, 1240–1242.
- Smýkal, P., Vernoud, V., Blair, M.W., Soukup, A. & Thompson, R.D. (2014) The role of the testa during development and in establishment of dormancy of the legume seed. *Frontiers in Plant Science*, **5**, 351.
- Steinbrecher, T. & Leubner-Metzger, G. (2018) Tissue and cellular mechanics of seeds. *Current Opinion in Genetics & Development*, **51**, 1–10.
- Sun, L., Miao, Z., Cai, C., Zhang, D., Zhao, M., Wu, Y. *et al.* (2015) GmHs1-1, encoding a calcineurin-like protein, controls hard-seededness in soybean. *Nature Genetics*, **47**, 939–943.
- Swanson-Wagner, R., Briskine, R., Schaefer, R., Hufford, M.B., Ross-Ibarra, J., Myers, C.L. *et al.* (2012) Reshaping of the maize transcriptome by domestication. *Proceedings of the National Academy of Sciences of the United States of America*, **109**(29), 11878–11883.
- Szymanowska, U., Jakubczyk, A., Baraniak, B. & Kur, A. (2009) Characterisation of lipoxygenase from pea seeds (*Pisum sativum* var. Telephone L.). *Food Chemistry*, **116**, 906–910.
- Trněný, O., Brus, J., Hradilová, I., Rathore, A., das, R., Kopecký, P. *et al.* (2018) Molecular evidence for two domestication events in the pea crop. *Genes*, **9**, 535.
- Verdier, J., Dessaint, F., Schneider, C. & Abirached-Darmency, M. (2013) A combined histology and transcriptome analysis unravels novel questions on *Medicago truncatula* seed coat. *Journal of Experimental Botany*, **64**, 459–470.
- Verdier, J., Lalanne, D., Pelletier, S., Torres-Jerez, I., Righetti, K., Bandyopadhyay, K. *et al.* (2013) A regulatory network-based approach dissects late maturation processes related to the acquisition of desiccation tolerance and longevity of *Medicago truncatula* seeds. *Plant Physiology*, **163**, 757–774.
- Verma, S., Attuluri, V.P.S. & Robert, H.S. (2022) Transcriptional control of Arabidopsis seed development. *Planta*, **255**, 90.
- Wada, S., Kennedy, J.A. & Reed, B.M. (2011) Seed-coat anatomy and proanthocyanidins contribute to the dormancy of *Rubus* seed. *Scientia Horticulturae*, **130**, 762–768.
- Wang, G., Oh, D.-H. & Dassanayake, M. (2020) GOMCL: a toolkit to cluster, evaluate, and extract non-redundant associations of Gene Ontology-based functions. *BMC Bioinformatics*, **21**, 139.
- Wang, X., Zhang, H., Song, R., Sun, M., Liu, P., Tian, P. *et al.* (2023) Multiple omics datasets reveal significant physical and physiological dormancy in alfalfa hard seeds identified by multispectral imaging analysis. *The Crop Journal*, **11**, 1458–1468.
- Weber, H., Borisjuk, L. & Wobus, U. (2005) Molecular physiology of legume seed development. *Annual Review of Plant Biology*, **56**, 253–279.
- Weston, L.A. & Mathesius, U. (2013) Flavonoids: their structure, biosynthesis and role in the rhizosphere, including allelopathy. *Journal of Chemical Ecology*, **39**, 283–297.
- Wickham, H. (2016) *ggplot2: elegant graphics for data analysis*. New York: Springer-Verlag. ISBN 978-3-319-24277-4, <https://ggplot2.tidyverse.org> [Accessed 22nd June 2023].
- Wildermuth, M.C., Dewdney, J., Wu, G. & Ausubel, F.M. (2001) Isochorismate synthase is required to synthesize salicylic acid for plant defence. *Nature*, **414**(6863), 562–565. Available from: <https://doi.org/10.1038/35107108>
- Wink, M. (2013) Evolution of secondary metabolites in legumes (Fabaceae). *South African Journal of Botany*, **89**, 164–175.
- Winkel-Shirley, B. (2001) Flavonoid biosynthesis. A colorful model for genetics, biochemistry, cell biology, and biotechnology. *Plant Physiology*, **126**, 485–493.
- Wu, Z., Song, L., Feng, S., Liu, Y., He, G., Yioe, Y. *et al.* (2012) Germination dramatically increases isoflavonoid content and diversity in chickpea (*Cicer arietinum* L.) seeds. *Journal of Agricultural and Food Chemistry*, **60**, 8606–8615.
- Yonekura-Sakakibara, K., Yamamura, M., Matsuda, F., Ono, E., Nakabayashi, R., Sugawara, S. *et al.* (2020) Seed-coat protective neolignans are

- produced by the dirigent protein AtDP1 and the laccase AtLAC5 in Arabidopsis. *The Plant Cell*, **33**, 129–152.
- Yu, B., Gao, P., Song, J., Yang, H., Qin, L., Yu, X. *et al.* (2023) Spatiotemporal transcriptomics and metabolic profiling provide insights into gene regulatory networks during lentil seed development. *The Plant Journal*, **115**, 253–274.
- Zablatzká, L., Balarynová, J., Klčová, B., Kopecký, P. & Smýkal, P. (2021) Anatomy and histochemistry of seed coat development of wild *Pisum sativum* subsp. *elatius* (M. Bieb.) Asch. et Graebn. and domesticated pea (*Pisum sativum* subsp. *sativum* L.). *International Journal of Molecular Sciences*, **22**, 4602. Available from: <https://doi.org/10.3390/ijms22094602>
- Zhao, L., Gao, L., Wang, H., Chen, X., Wang, Y., Yang, H. *et al.* (2013) The R2R3-MYB, bHLH, WD40, and related transcription factors in flavonoid biosynthesis. *Functional and Integrative Genomics*, **13**, 75–98.
- Zhou, K., Hu, L., Li, Y., Chen, X., Zhang, Z., Liu, B. *et al.* (2019) MdUGT88F1-mediated phloridzin biosynthesis regulates apple development and Valsa canker resistance. *Plant Physiology*, **180**(4), 2290–2305. Available from: <https://doi.org/10.1104/pp.19.00494>
- Zou, H., Tzarfati, R., Hübner, S., Krugman, T., Fahima, T., Abbo, S. *et al.* (2015) Transcriptome profiling of wheat glumes in wild emmer, hulled landraces and modern cultivars. *BMC Genomics*, **16**, 777. Available from: <https://doi.org/10.1186/s12864-015-1996>