Plant Physiology and Biochemistry 229 (2025) 110771

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Gene coexpression network analysis of galactomannan biosynthesis and endosperm maturation in species of the genus *Coffea*

Stéphane Dussert[a], Anna K. Stavrinides[a], Julien Serret[a], Virginie Vaissayre[a], Marie-Christine Combes[a], Fabienne Morcillo[a,b], Eveline Lefort[a,b], Stéphanie Rialle[c], Hervé Etienne[a,b], Philippe Lashermes[a], Thierry Joët[a,*]

[a] DIADE, Univ Montpellier, IRD, CIRAD, Montpellier, France
[b] CIRAD, DIADE, F-34398, Montpellier, France
[c] MGX-Montpellier GenomiX, Univ. Montpellier, CNRS, INSERM, Montpellier, France

ARTICLE INFO

Keywords:
Coffea
Cellular endosperm
Seed maturation
Cell wall storage polysaccharides
Mannans
Storage compounds

ABSTRACT

In a few important plant families and genera, including Arecaceae, Fabaceae and the genus *Coffea*, the main seed storage polysaccharide is not starch but cell wall galactomannans. Such seeds are albuminous with a persistent copious living endosperm that accumulates galactomannans. However, our understanding of the regulation of endosperm maturation, cell wall formation and galactomannan biosynthesis in albuminous seeds remains very limited. To gain insights into these processes, a large RNA-seq dataset was produced (14 coffee species × 5 endosperm developmental stages) and scrutinized using gene coexpression network analysis. The network revealed tight transcriptional coordination of the core galactomannan biosynthetic machinery with nucleotide sugar synthesis and transport, arabinogalactan protein and cellulose synthesis, and regulation of the trans-Golgi network. The orchestration of galactomannan and oil accumulation during endosperm maturation appeared to be exerted by the transcription factors FUS3, WRINKLED1, SHINE2 and DREB2D. The latter was the only coexpression partner of galactomannan biosynthetic genes. Numerous key genes of galactomannan accumulation were significantly upregulated in coffee somatic embryos overexpressing *DREB2D*, which showed increased production of UDP-galactose and diversion towards raffinose family oligosaccharides. Further, most genes of the galactomannan coexpression module were identified as DREB2D target genes by DAP-seq analysis.

1. Introduction

Plants store polysaccharides, oil and proteins in seeds to support the emergent young plant after germination until photosynthesis is fully established. Starch is by far the most common polysaccharide stored in seeds, where it is deposited intracellularly in specialised organelles called the amyloplasts (Aguirre et al., 2018). However, in species of certain genera and families, the main seed polysaccharide reserves are hemicelluloses that accumulate in cell walls, mostly mannans and xyloglucans (Delmer et al., 2024). The linear backbone of mannans consists of β-1,4-linked mannose residues with variable degrees of side substitutions. In seeds, the most common form of storage mannans is galactomannans, in which a varying proportion of mannosyl residues is substituted with an α-1,6-linked galactosyl residue (Scheller and Ulvskov, 2010). Seeds that store mannans in their cell walls are all albuminous seeds with a persistent cellularised living endosperm at the mature stage. The endosperm of albuminous seeds which store mannans may represent almost the entire seed mass and volume, as is the case in species of the monocot family Arecaceae (palms) and of the dicot genus *Coffea* (Joët et al., 2009; Dussert et al., 2013). The family Fabaceae is also known to contain numerous species which produce seeds with a persistent endosperm that accumulates mannans (Buckeridge, 2010). Among these species, seeds of the annual guar (*Cyamopsis tetragonoloba*) or the carob tree (*Ceratonia siliqua*) are exploited for extraction of galactomannans, which are widely used in the food, pharmaceutical, and cosmetic industries (Singh et al., 2018; Sharma et al., 2022).

Abbreviations: CWP, cell wall polysaccharides; DAP-Seq, DNA affinity purification sequencing; GMB, galactomannan biosynthesis.

* Corresponding author.

E-mail addresses: stephane.dussert@ird.fr (S. Dussert), anna_stavrinides@yahoo.com (A.K. Stavrinides), julien.serret@ird.fr (J. Serret), virginie.vaissayre@ird.fr (V. Vaissayre), marie-christine.combes@ird.fr (M.-C. Combes), fabienne.morcillo@cirad.fr (F. Morcillo), eveline.lefort@ird.fr (E. Lefort), stephanie.rialle@cnrs.fr (S. Rialle), herve.etienne@cirad.fr (H. Etienne), philippe.lashermes@ird.fr (P. Lashermes), thierry.joet@ird.fr (T. Joët).

https://doi.org/10.1016/j.plaphy.2025.110771
Received 9 July 2025; Received in revised form 25 October 2025; Accepted 12 November 2025

The mechanisms by which galactomannans are synthesized and stored in endosperm cell walls during maturation of albuminous seeds remain largely unresolved. However, three key galactomannan biosynthetic enzymes have been identified and characterised, using guar and two other Fabaceae species, fenugreek (*Trigonella foenum-graecum*) and senna (*Senna occidentalis*), as model systems. Two enzymes are Golgi-located, the mannan synthase (MANS), a CELLULOSE SYNTHASE-LIKE A (CSLA) enzyme that produces the β-1,4-linked mannan backbone (Dhugga et al., 2004), and the mannan α-galactosyltransferase (MAGT, previously known as galactomannan α-galactosyltransferase GMGT), which catalyses the transfer of galactosyl residues from UDP-galactose to the mannan backbone to assemble the galactosyl side chains (Edwards et al., 1999). The post-depositional modulation of the degree of galactose substitution is then operated by the cell wall-located α-galactosidase (AGAL) (Edwards et al., 1992; Joersbo et al., 2001). In addition to these three essential enzymes, another key factor may be needed for seed storage mannan synthesis, the MANNAN-SYNTHESIS RELATED (MSR) protein (Wang et al., 2013), which appeared to be an indispensable cofactor for coffee MANS1 mannan biosynthetic activity when heterologously expressed in *Pichia pastoris* (Voiniciuc et al., 2019). Additional insight comes from the progress in deciphering galactoglucomannan biosynthesis in Arabidopsis (Voiniciuc, 2022; Yoshimi et al., 2025). Among the nine CSLA proteins of Arabidopsis, CSLA2, CSLA3, CSLA7 and CSLA9 were demonstrated to elongate glucomannans with varying preference for Glc incorporation (Liepman et al., 2005; Goubet et al., 2009). By contrast to fenugreek MAGT, which accepts only homomannan oligosaccharides (Edwards et al., 1999), the Arabidopsis *MUCILAGE-RELATED10 (AtMUCI10)* encodes a strict glucomannan α-galactosyltransferase (Voiniciuc et al., 2015). Finally, in addition to carbohydrate substitutions, (gluco)mannans can be *O*-acetylated by specific members of the TRICHOME-BIREFRINGENCE-LIKE (TBL) family, namely mannan *O*-acetyltransfearases AtMOAT1-4 (Zhong et al., 2018). These findings represent the most recent advances in our understanding of mannan synthesis, highlighting that our knowledge of galactomannan accumulation in albuminous seeds remains scarce. In particular, the transcriptional regulation of this pathway remains largely unknown. More generally, our knowledge of the control of primary wall synthesis is extremely limited (Boerjan et al., 2024), in contrast to the considerable progress made in understanding how the synthesis of secondary wall components is regulated through MYB46 activity (Zhong et al., 2007).

This knowledge gap on storage mannan accumulation resembles that concerning the transcriptional orchestration of albuminous seed development and maturation, in comparison with our extended knowledge on that of dicot exalbuminous seeds (Fatihi et al., 2016; Alizadeh et al., 2021) and of cereal endosperm (Grimault et al., 2015; Zhang et al., 2016a; Song et al., 2024). LEC1-ABI3/FUS3/LEC2 (LAFL) proteins are well-characterised master regulators of seed maturation in exalbuminous seeds (Fatihi et al., 2016; Alizadeh et al., 2021). In particular, LEC2 and FUS3 activate the AP2/EREB transcription factor WRINKLED1 (WRI1) (Baud et al., 2007; Wang and Perry, 2013), which is thought to play a ubiquitous role in fatty acid synthesis in plants (Ma et al., 2013; Marchive et al., 2014). The role of the LAFL regulators and WRI1 in albuminous seeds is very poorly documented (Miray et al., 2021). However, in a previous work, we found that ABI3 and FUS3 transcription peaked at the filling stage in the developing *C. arabica* endosperm (Dussert et al., 2018). WRI1 expression coincides with that of fatty acid biosynthetic genes in the developing oil-rich endosperm of albuminous oil palm seeds (Dussert et al., 2013) and *Ricinus communis* (castor bean) seeds (Yang et al., 2019). Although other examples of transcriptome analyses of restricted developmental series have certainly been carried out, to our knowledge, a comprehensive analysis of the transcriptional orchestration of endosperm maturation in a dicot albuminous seed species has not yet been produced.

The endosperm of the coffee seed represents almost the entire seed mass and its main reserves are galactomannans and oil (Joët et al., 2009), making the coffee seed an interesting model to gain insights into galactomannan accumulation and endosperm maturation in dicot albuminous species through gene coexpression analysis. Several comprehensive reviews have detailed the advantages of this approach to investigate seed developmental processes (e.g. Righetti et al., 2015; Garg et al., 2017), infer gene function and prioritize candidate genes (e.g. Usadel et al., 2009; Li et al., 2015). It is worth noting that the two pioneer studies that used coexpression analysis in plant biology enabled the discovery of key players of cellulose and secondary wall biosynthesis in Arabidopsis (Brown et al., 2005; Persson et al., 2005). Since then, the successful use of gene coexpression network analysis to examine cell wall and oil biosynthesis in plants of major importance in both agriculture and forestry has been reported by numerous authors (e.g. Mochida et al., 2011; Voiniciuc et al., 2015; Guerin et al., 2016; Cui et al., 2021; Quan et al., 2021). When designing a coexpression study, the challenge is to obtain large but continuous variation in gene expression levels. In the present work, we took advantage of the conserved transcriptional programme of the developing endosperm in fourteen closely related *Coffea* species which differ in the galactomannan and oil contents of their seeds. Using a large RNA-seq dataset (69 transcriptomes), we built a gene coexpression network of the coffee endosperm, investigated whether LAFL regulators and WRI1 may orchestrate endosperm maturation, analysed coexpression modules to identify novel players and key processes associated with the deposition of cell wall storage polysaccharides and searched for the specific set of regulators for cell wall thickening and galactomannan biosynthesis. Coexpression network analysis identified the AP2/ERF transcription factor DREB2D as a candidate for galactomannan biosynthesis regulation. Its function was then investigated through overexpression in coffee somatic embryos and DNA affinity purification sequencing (DAP-seq) analysis.

2. Material and methods

2.1. Seed material

Seeds from fourteen coffee species conserved in the field genebank of the *Coffea* Biological Resources Center (BRC Coffea, maintained by IRD and CIRAD on Reunion Island) were used for transcriptome and chemical analyses. The fourteen *Coffea* species were *C. arabica* (ARA), *C. brevipes* (BRE), *C. canephora* (CAN), *C.* sp. 'Congo' (CON), *C. dewevrei* (DEW), *C. eugenioides* (EUG), *C. heterocalyx* (HET), *C. liberica* (LIB), *C. mauritiana* (MAUR), *C. pocsii* (POC), *C. pseudozanguebariae* (PSEU), *C. salvatrix* (SAL), *C. sessiliflora* (SES), *C. stenophylla* (STE). Seeds at five different maturation stages known to span the development of the endosperm (ST3 to ST7, Joët et al., 2009) were collected from three trees (of three distinct genotypes) per species. As the seed development stage varies considerably among the fourteen species studied, i.e. from ca. 3–12 months (Dussert et al., 2000), the developmental stages were based on marked anatomical and morphological seed and fruit traits that are shared across coffee species, as defined and described previously in *C. arabica* (Joët et al., 2009; Dussert et al., 2018). Briefly, stage 3 coincides with the rapid growth of the aqueous endosperm tissue which progressively replaces the perisperm in the locule (endosperm occupying half of the locule); at stage 4, the remaining perisperm resembles a thin green pellicle surrounding a soft milky endosperm; stage 5 is the peak of reserve deposition and corresponds to endosperm hardening due to massive deposition of galactomannans in the cell walls; stage 6 coincides with fruit veraison, while stage 7 corresponds to mature cherry fruits with red pericarp. After being cross sectioned, the seed was separated from the pericarp and immediately frozen in liquid nitrogen and stored at −80 °C. The endosperm was separated from the perisperm and the embryo while frozen. To minimise the genotypic effect, for each combination of species × development stage, endosperms from the three different accessions collected were evenly pooled prior to grinding and extraction of RNA, lipids, sugars and cell wall polysaccharides (CWP).

2.2. Production of transgenic coffee somatic embryos overexpressing the DREB2D gene

Coffee *DREB2D* cDNA (Genbank accession number XM_027236031), which was used as a template for PCR amplification, was cloned into the plant overexpression vector pMDC32, transgene expression being driven by a double sequence of the cauliflower mosaic virus 35S promoter. Coffee embryogenic calli (*C. arabica* cv. Caturra) were genetically transformed using recombinant *Agrobacterium tumefaciens* strain LBA1119 containing the recombinant plasmid, as previously described (Ribas et al., 2011). Somatic embryos regenerated from each hygromycin resistant callus (independent transgenic lines) were tested for transformation by PCR amplification of the selection gene *HPTII*.

2.3. Transcriptome analysis

For each of the 69 combinations of species × development stage (it was impossible to sample seeds at stage 3 in *C.* sp. Congo), endosperms were ground to a fine powder in an analytical grinder (IKA A10, Staufen, Germany) and total RNA was extracted from 70 mg of the powder using the Qiagen RNeasy Lipid Tissue Kit (Qiagen, Stanford, CA). Similar RNA extraction procedures were followed for transgenic somatic embryos. cDNA libraries were constructed using the TruSeq™ Stranded mRNA Sample Preparation Kit (Illumina, USA), then sequenced on an Illumina HiSeq 2500 (single reads, 100 nt) at the MGX platform (Montpellier Genomix, http://www.mgx.cnrs.fr/). After quality filtering using Cutadapt (quality score > Q30 and removal of reads shorter than 60 bp or longer than 140 bp), a minimum of 13 million reads per library were obtained (average of 27.4 million reads per library; Supplementary Table S1). The entire dataset has been deposited at the European Nucleotide Archive (ENA) under project numbers PRJEB32533 and PRJEB79959. Owing to the very low genetic divergence between the *Coffea* species (average of 1.3 % gene sequence variation (Cenci et al., 2012)), the trimmed reads of each library were mapped to the *C. canephora* coding transcriptome DNA reference sequence (25,574 CDS) (Denoeud et al., 2014) using BWA MEM (Li, 2013) with the default parameters, resulting in 60.9–76.3 % of mapped reads, independent of the species (Supplementary Table S1). The percentage of mapped reads did not differ significantly among the fourteen coffee species ($F_{1,13} = 1.49$; $P = 0.149$). Reads were counted using IDXstats in SAMtools (Li et al., 2009) and read counts were then normalized (RPKM). Principal component analysis of the whole transcriptome dataset showed grouping of samples at stages 3 and 4 in almost all species, grouping of samples of stages 6 and 7 in a second group, with stage 5 endosperms between the two groups, indicating a conserved transcriptional sequence of endosperm development in the genus *Coffea*, with major distinct switches governing endosperm growth, reserve deposition, and late maturation (Supplementary Fig. S1).

Genes with extremely low expression (<20 counts in total across the 69 tanscriptomes) were not used for subsequent analyses. The R-based DESeq2 software package (Love et al., 2014) was used to normalize RNAseq expression values for somatic embryos and to identify genes displaying significantly different expression levels between the controls and *DREB2D*-overexpressing lines, using four biological replications in both control and DREB2D+ somatic embryos (Supplementary Table S2), and a padj cutoff of 0.01. Mapman4 gene ontology annotation (Schwacke et al., 2019) was performed using Mercator (Lohse et al., 2014). The significance of functional term enrichment in each category of candidates was based on a FDR-adjusted hypergeometric distribution of Mapman BIN categories, using 23,797 genes with annotated BIN categories as a reference.

Transgenic *C. arabica* lines were also tested for overexpression of *DREB2D* and downstream putative target genes using qPCR analysis as described in Joët et al. (2014). The set of primers that enabled amplification of target genes is detailed in Supplementary Table S3. The level of expression of each gene was normalized to the geometric mean of expression levels of three validated coffee reference genes (Cc08g05690 Ubiquitin UBQ10, Cc00g15790 40S ribosomal protein S24, and Cc00g17460 14-3-3 protein; Cruz et al., 2009).

2.4. Coexpression analysis

Gene coexpression network analysis was performed using the procedure previously described in Guerin et al. (2016) after logarithmic transformation of read counts. The strategy used to build the coexpression network of galactomannan biosynthesis during coffee seed maturation combined both the guide-gene approach and the non-targeted approach, as defined by Aoki et al. (2007). The network was indeed constructed in the vicinity of a set of galactomannan biosynthetic guide genes, but their partners may belong to the whole coffee genome. The six guide genes used were *MANS1*, *MAGT1*, *VTC1* and *UGE1-3* genes, all described in a previous transcriptome analysis (Joët et al., 2014). These genes encode the enzymes required to produce the nucleotide sugar building blocks GDP-mannose and UDP-galactose (GDP-Man pyrophosphorylase, also known as VITAMIN C DEFECTIVE1, VTC1, and UDP-glucose 4-epimerase, UGE, respectively), to assemble the mannan backbone (mannan synthase, MANS), and to introduce the galactosyl side chains (galactosyltransferase, MAGT). During the first round of network construction, all positive connections between guide genes and their partners (the genes to which they are linked, hereafter termed P1) were computed. During the second round of network construction, partners of P1 were identified (P2) and then links between P2 were generated. Linear regressions were computed using in-house scripts. Two rounds of coexpression analysis using a $|R|$ threshold of 0.85 provided a tractable number of nodes (180) and edges (4878) for network construction. Gene interactions were visualized using the open source software Cytoscape (Shannon et al., 2003) and an organic layout. The Markov Cluster (MCL) algorithm (Enright et al., 2002) was used for module detection with an inflation of 3.5. Genes of the galactomannan biosynthesis (GMB) coexpression network were manually curated, taking into account recent relevant literature and Mapman annotation using Mercator software to retrieve Mapman functional categories (Lohse et al., 2014). Gene names were allocated based on the best Arabidopsis matches.

2.5. Determination of lipid and sugar contents and monosaccharide composition of cell wall material

Total lipids were extracted from 300-mg samples of freeze-dried endosperm powder using a modified Folch method (Laffargue et al., 2007). Sugars were extracted and measured by high-performance anion exchange chromatography coupled with pulsed amperometric detection (Dionex Chromatography Co., Sunnyvale, CA, USA) as detailed elsewhere (Dussert et al., 2006). The CWP content of the endosperm was measured using the defatted alcohol-insoluble residue method (Redgwell et al., 2003; Joët et al., 2014). Lipids, sugars and CWPs were all analysed in triplicate (from three different extractions) using a completely random experimental design (Supplementary Tables S4 and S5).

2.6. Recombinant DREB2D protein purification and electrophoretic mobility shift assay (EMSA)

The CDS for *DREB2D* was amplified from cDNA of *C. arabica* seeds using a primer pair containing restriction sites (5‘-ATCA**GGATCC**GAAGCTGACCGTAGTG-3’, and 5′-ATAT**CTCGAG**TCAGTTCCAGGGGTAGTTG-3′). The PCR product was cloned into the bacterial expression vector pET28(a) in frame with the N-terminal 6xHis tag by restriction digest. Recombinant proteins were produced in *Escherichia coli* DE3 (BL21) cells. Transformed cells were grown overnight at 37 °C, with 180 rpm shaking, in LB medium containing kanamycin (30 mg l^{-1}). Cultures were then diluted 1:100 in fresh 2 x YT media supplemented with kanamycin, and grown to an OD_{600} of 0.8, before chilling on ice. After induction by 0.1 mM isopropylthio-β-galactoside and overnight culture at 16 °C, bacteria were pelleted by centrifugation and lysed by sonication in Buffer A (50 mM phosphate buffer at pH 8, 50 mM glycine, 500 mM NaCl, 5 % glycerol, 10 mM imidazole) supplemented with a protease inhibitor cocktail (Promega) and 0.2 mg ml^{-1} lysozyme. Cellular debris was pelleted by centrifugation at 17000×*g* for 15 min at 4 °C, and the supernatant recovered and used to purify polyhistidine-tagged fusion proteins by incubation with nickel-nitrilotriacetic acid agarose resin (Qiagen) for 1 h at 4 °C. Soluble, tagged proteins were purified by washing the resin with Buffer B (the same as Buffer A, with 20 mM imidazole), and then eluted in Buffer C (the same as Buffer A, with 100 mM imidazole). Proteins were dialysed in Buffer D (50 mM Hepes buffer pH 7, 100 mM NaCl, 1 % glycerol, 1 mM DTT) with Amicon filtration units (10 KDa cutoff). The concentration of proteins was measured with Bradford reagent (Sigma-Aldrich) according to the manufacturer's instructions. Proteins were aliquoted, flash-frozen in liquid nitrogen and stored at −20 °C.

Oligonucleotide probes were designed based on the literature to obtain a core DREB recognition site (CCGAC), or with mutation to the DRE element (Feng et al., 2019). Probes were synthesized (Thermo Fisher) as single strands (Supplementary Table S6) and were hybridized by mixing the forward and reverse strand in equimolar amounts, heating to 95 °C and then slowly cooling to room temperature. The DRE forward single strand was tagged with AlexaFluor 488 on the 5' end to serve as labelled probe for the gel shift assays. Test solutions consisted of combinations of DRE fluorescent probe (at 200 or 100 ng per lane) and test probes with or without DREB2D. These solutions were mixed and incubated for 25 min before adding loading dye (TBE buffer, 80 % glycerol, 0.2 % bromophenol blue) and migrating in native gels (10 % polyacrylamide, 0.5 X TBE, ran at 100 V). The resulting gels were visualized using a Typhoon FLA 9500 imager (Amersham Biosciences) using the default settings for AlexaFluor488.

2.7. DNA affinity purification sequencing (DAP-seq) and analysis

Coffea canephora genomic DNA libraries were prepared for amp-DAP experiments and amplified using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, UK). The Halo-DREB2D fusion protein was obtained by coupled *in vitro* transcription/translation for 2 h at 27 °C using the TNT SP6 High-Yield Wheat Germ Protein Expression System (Promega, WI, USA). Halo-DREB2D synthesis was verified by Western blot using Anti-HaloTag® Monoclonal Antibodies (Promega). Halo-DREB2D purification and DNA affinity binding were performed following the detailed protocol described by Bartlett et al. (2017) and optimized by Hutin et al. (2023). Briefly, Halo-DREB2D proteins were immobilised on 20 μl of ChromoTek Halo-Trap Agarose Magnetic Beads for 1 h under agitation on a rotating wheel at 4 °C using 130 μl of cold DAP buffer (PBS supplemented with protease inhibitors (Roche), Nonidet P-40 (0.005 % v/v), and 1 mM TCEP (tris(2-carboxyéthyl) phosphine). Agarose beads complexed with Halo-DREB2D were then immobilised on a prechilled magnetic rack, washed 6 times with cold DAP buffer, resuspended in 100 μl of DAP buffer and incubated with 50 ng of gDNA librairies for 90 min at 4 °C on a rotating wheel. Beads were washed 10 times with 100 μl of DAP buffer, resuspended in 30 μl of elution buffer (Qiagen) and the bound DNA was eluted from the beads by heating at 98 °C for 10 min. A final DNA amplification step was carried out using the KAPA HiFi Real-Time PCR Library Amplification Kit (Kapa Biosystems, Boston, USA). Amp-DAP DNA libraries were purified with Agencourt AMPure XP beads (Beckman Coulter, Miami, FL) and sequenced on an Illumina NovaSeq 6000 on the MGX platform (Montpellier Genomix). The input gDNA libraries and Halo-DREB2D amp-DAP experiments were produced and sequenced in triplicate.

Read quality was analysed with FastQC (v0.11.9) and screening for contamination was performed using FastQ Screen (v0.15.1). The reads were then aligned with BWA (v0.7.17-r1188) with the -q 30 option on the *Coffea canephora* genome (Denoeud et al., 2014). The duplicate reads were removed using SAMtools (v1.17). After quality filtering, a minimum of 80 million paired-end reads per library was obtained. Peaks for each replicate were called for separately by comparing their alignment file with that of the control input DNA using MACS2 (v2.7.1) with option fe-cutoff 3 (Zhang et al., 2008). Peaks identified on individual replicates were compared using bedtools (v2.31.0) 'multiinter' function while the peak annotation was carried out using bedtools "closest" function to assign peaks to genes based on their proximity to the transcription start site (TSS), with a threshold of 2 kb or 4 kb upstream of the TSS. The MEME tool from MEME Suite (v5.5.5) was used to analyse the bound regions to identify enriched DNA sequence motifs, and Tomtom and RSAT tools were used to compare the detected motifs with databases (Bailey et al., 2015; Santana-Garcia et al., 2022).

3. Results

3.1. Endosperm lipid and cell wall polysaccharide contents vary considerably between coffee species

Based on a previous study showing that the seed lipid content varies greatly between *Coffea* species (Dussert et al., 2001), we hypothesized that the cell wall polysaccharide (CWP) content may also vary, assuming that oil and CWP were the two major storage reserves in all coffee species. The 14 coffee species studied here displayed considerable variability in the cell wall polysaccharide (CWP) content of their seeds (Fig. 1A), which ranged from 47.5 % dry matter (DM) in *C. salvatrix* to 81.1 % DM in *C. mauritiana*. The total lipid content of mature seeds also varied considerably from 10.8 % DM in *C. mauritiana* to 29.6 % DM in *C. pseudozanguebariae* (Fig. 1B). The cell wall polysaccharides (CWP) and lipid contents were highly negatively correlated ($R^2 = 0.79$, Fig. 1C), indicating that CWP and lipids collectively represent most of the mature endosperm reserves in all coffee species.

3.2. The galactomannan biosynthesis guided gene coexpression network reveals novel players and key processes associated with cell wall storage polysaccharide deposition

Two rounds of guided coexpression analysis using six key genes of galactomannan biosynthesis as guide genes (*MANS1*, *MAGT1*, *VTC1* and *UGE1-3*), the complete RNA-seq dataset (stages 3–7) and a $|R|$ threshold of 0.85, yielded two distinct networks of very contrasted sizes (Fig. 2A). The three *UGE* guide genes formed a very small network of six genes, while *MANS1*, *MAGT1* and *VTC1* formed a second large network of 174 genes, hereafter referred to as the galactomannan biosynthesis (GMB) network (Supplementary Table S7). The Markov clustering algorithm identified three modules in the GMB network. Modules 1, 2 and 3 contained 121, 43 and 11 genes, respectively. The three guide genes (*MANS1*, *VTC1*, and *MAGT1*) all belong to module 2, the core galactomannan biosynthesis module. Module 2 also includes *MSR1* (Cc00g16710), a key cofactor reported to be indispensable for coffee MANS1 mannan synthase activity (Voiniciuc et al., 2019) (Fig. 3).

Among the 174 genes of the GMB network, 124 genes (71 %) were assigned to functional groups and subgroups based on high amino acid sequence homology with proteins that have already been functionally characterised in model plants (Supplementary Table S7). It should be noted that no other members of the *MANS* and *MAGT* gene families were found in the GMB network, which was expected given that only *MANS1* and *MAGT1* were significantly expressed during endosperm development (Supplementary Table S8). The GMB network showed highly significant enrichment (FDR$<$0.0001) in the functional terms cell wall, cell wall precursor synthesis, cellulose synthesis, pectin synthesis and cell wall proteins (Fig. 2B). In all, the cell wall-related genes represented 29 % of the coexpression network annotated genes (36 genes). Among them, the coexpression network contains four genes that encode components of the cellulose synthase complex (CSC), i.e. CESA1, CESA2,

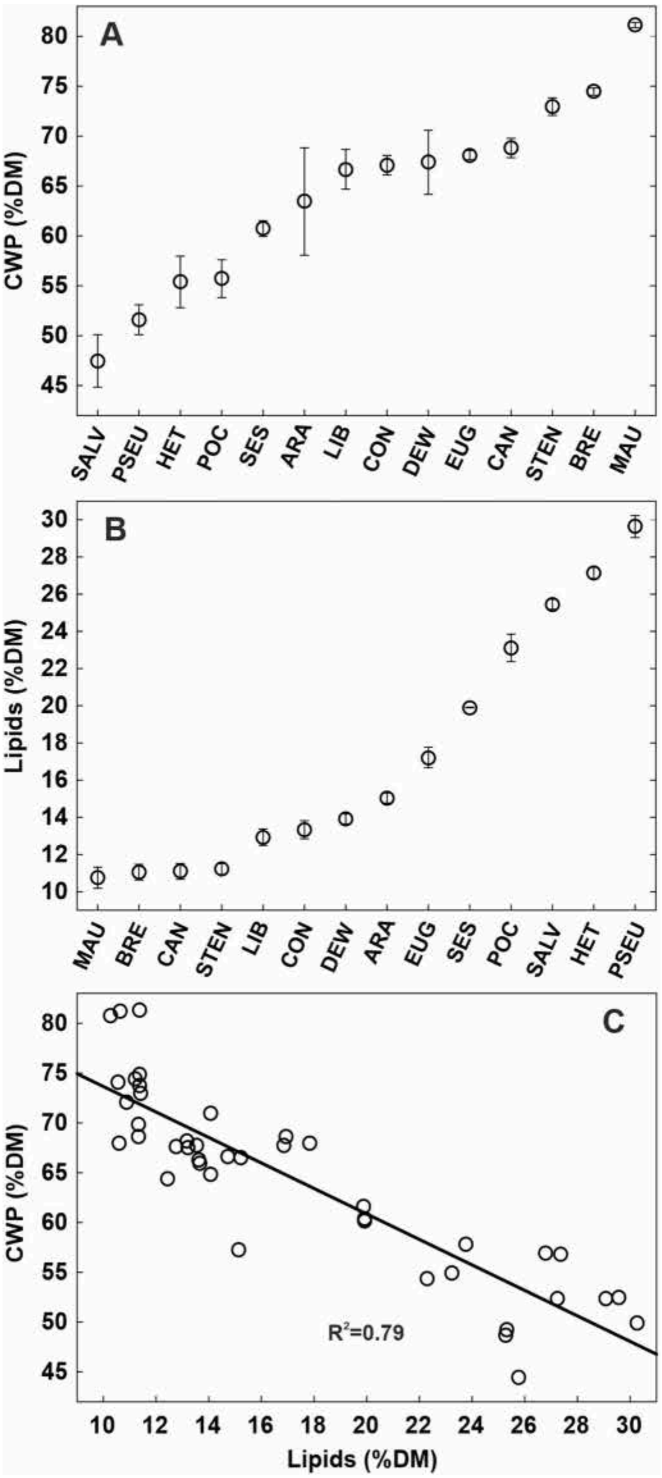

Fig. 1. Distribution of cell wall polysaccharides and lipid contents (% DM) in mature seeds of 14 *Coffea* species. Distribution of seed cell wall polysaccharide (A) and oil (B) contents in the 14 *Coffea* species studied. (C) Negative correlation between cell wall polysaccharide and oil contents. The 14 *Coffea* species are *C. arabica* (ARA), *C. brevipes* (BRE), *C. canephora* (CAN), *C.* sp. Congo (CON), *C. dewevrei* (DEW), *C. eugenioides* (EUG), *C. heterocalyx* (HET), *C. liberica* (LIB), *C. mauritiana* (MAUR), *C. pocsii* (POC), *C. pseudozanguebariae* (PSEU), *C. salvatrix* (SAL), *C. sessiliflora* (SES), *C. stenophylla* (STE).

CESA3, CESA6, and three key proteins of CSC secretory vesicles, KOR1, TRANVIA, and STELLO (Zhang et al., 2016b) (Fig. 3). TRANVIA (TVA) was recently described as a component of secretory compartments which promotes the delivery of cellulose synthase complexes to the plasma membrane through the trans-Golgi network (Vellosillo et al., 2021). The GMB network therefore provides a unique resource for identifying candidate genes associated with galactomannan accumulation, which functional role could be the subject of more in-depth studies in the future.

The network also includes COBRA-like protein 1 (*CBL1*) and Chitinase-like protein 2 (*CTL2*), both of which are involved in interactions between cellulose microfibrils and hemicellulose (Sánchez-Rodríguez et al., 2012; Aniento et al., 2022) (Fig. 3). Moreover, the presence of seven genes involved in the synthesis of cell wall proteins such as arabinogalactan proteins (AGPs), is also worth noting, as well as that of seven genes encoding well-characterised enzymes devoted to pectin synthesis and modification (CGR2, CGR3, PGLR, PME34, PME36, API, GAUT9) (Fig. 3). Finally, the GMB network also contains six genes dedicated to the biosynthesis and transport of nucleotide sugars (*GMD2*, *PPM*, *RGP1*, *UAFT*, *URGT6*, *UUAT1*), including key players of the GDP-Man pathway, such as the phosphomannomutase (PPM), which catalyses the interconversion of man-6P to man-1P and is essential for the biosynthesis of GDP-man, or URGT6, which transports UDP-gal into the Golgi lumen (Rautengarten et al., 2014) (Fig. 3). By contrast, KONJAC1 and 2, which regulate VTC1 activity for GDP-man biosynthesis (Sawake et al., 2015), were not retrieved in the GMB network.

Another functional class overrepresented in the GMB network is cellular transport (17 genes), with numerous genes involved in vesicle-mediated transport via the trans-Golgi-network, including KEA4, a Golgi-located cation/proton exchanger which suports Golgi function in cell-wall biosynthesis (Wang et al., 2019), SCAMP4, and several RAB-GTPases which play a major role in the regulation of membrane tethering and fusion in the trans-Golgi network and in the secretory pathway of cell wall polysaccharides (Lunn et al., 2013) (Fig. 3). Finally, key players in central carbon metabolism are also present in the GMB network, including SWEET3, CIF1, SUSY1, which are involved in sucrose import and cleavage, as well as the glycolytic enzymes HXK-5 and PGLYM-p (Fig. 3).

Together with AGAL1 and Trichome birefringence-like TBL38 (Cc02g14650), which is involved in O-acetylation of cell wall polymers during their deposition in Arabidopsis (Sun et al., 2020), the three *UGE* genes formed a small network of six genes that was not connected to the main GMB network (Fig. 2A). However, together with *TBL38* and *AGAL1*, *UGE* genes joined the GMB network when the ST3-ST5 RNA-seq dataset (42 transcriptomes) was used to build the network (Supplementary Fig. S2; Supplementary Table S9). Developmental stages ST3 to ST5 corresponded to the period of cell wall galactomannan deposition and oil biosynthesis (Joët et al., 2009, 2014). This ST3-ST5 network contains all the guide genes used for coexpression analysis, i. e. *MANS1*, *MAGT1*, *VTC1, UGE1, UGE2, UGE3*, key players of CWP synthesis, such as TVA and MSR1, other enzymes which catalyse mannan backbone modifications, such as the mannan endo-1, 4-beta-mannosidases MAN2 and MAN6 (Cc05g07220 and Cc06g00320, respectively), and the galactinol synthase GolS1b (Cc02g35340) involved in the conversion of UDP-galactose into raffinose family oligosaccharides (Fig. 3).

3.3. Gene coexpression analysis identifies FUS3 and WRI1 as pivotal regulators of coffee endosperm maturation

The GMB network includes seven transcription factors (TFs), namely, three TFs of the B3 superfamily (FUS3 and auxin response factors ARF4 and ARF8), three TFs of the AP2/ERF superfamily (including WRI1, SHINE2 and DREB2D) and one GATA TF (Fig. 2A; Supplementary Table S10). Since FUS3 and WRI1 are well-characterised master

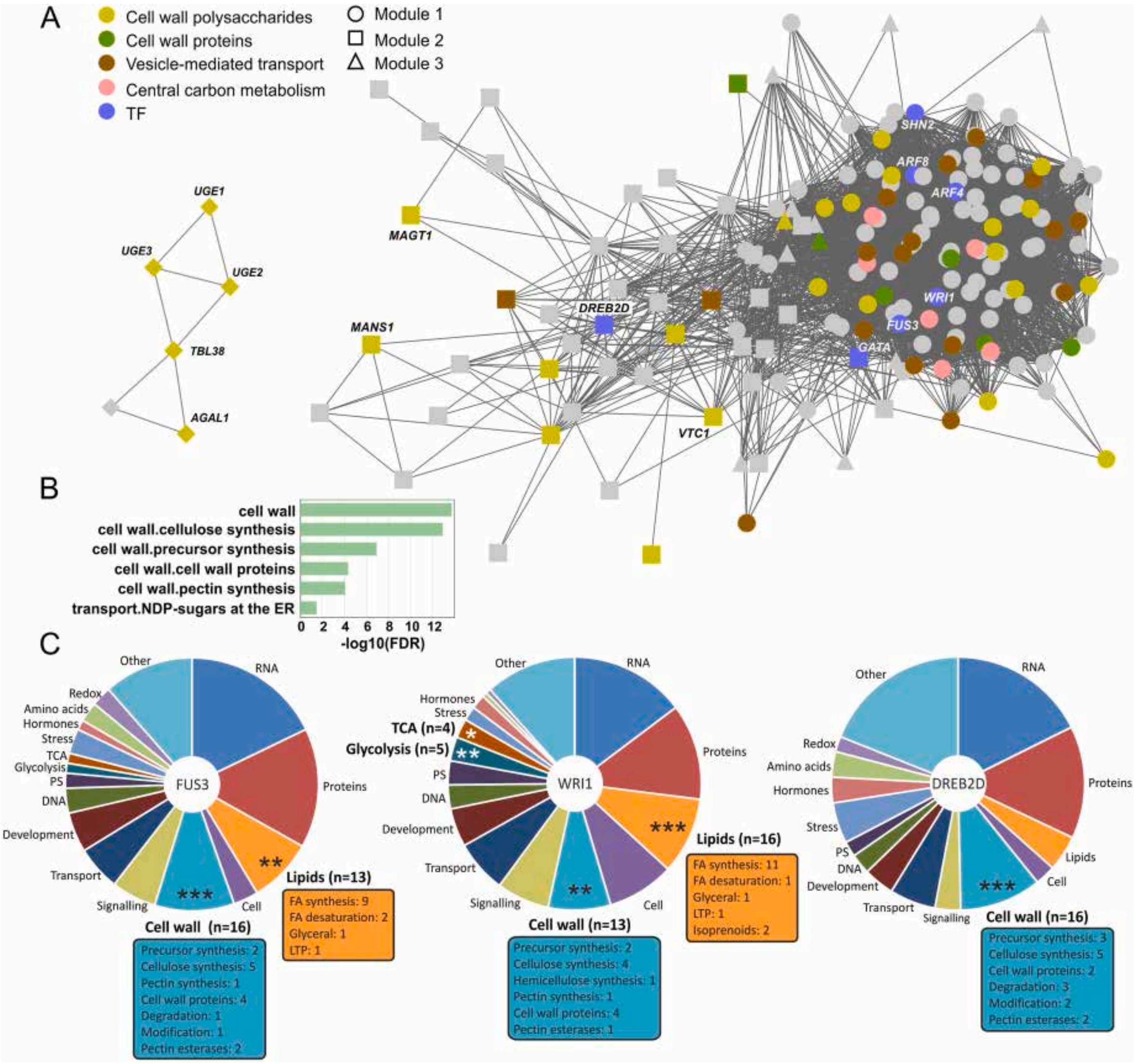

Fig. 2. Gene coexpression analysis of galactomannan biosynthesis. A. Gene coexpression analysis was performed using the whole transcriptome dataset, six galactomannan-related guide genes (*AGAL1*, *MAGT1*, *MANS1*, *VTC1*, *UGE1*, *UGE2*, *UGE3*), and an |R| threshold of 0.85. The three *UGE* guide genes formed a very small network of six genes (on the left), while *MANS1*, *MAGT1* and *VTC1* formed a large network comprising 174 genes, the galactomannan biosynthesis (GMB) network (on the right). The three modules detected in the GMB network using the Markov clustering algorithm are represented by different symbols (modules 1, 2 and 3 are represented by respectively, circles, squares and triangles). The most relevant functional groups of the GMB network genes are shown in colour: cell wall polysaccharide and protein synthesis, vesicle-mediated transport, central carbon metabolism, and transcription factors. **B.** False discovery rate (FDR) of the functional terms that showed significant enrichment in genes of the GMB network. **C.** Pie charts of functions assigned to the top 200 coexpression partners of FUS3, WRI1 and DREB2D. Significantly enriched functions are indicated by bold stars and the number of genes involved is given in brackets next to the name of the functional category. The number of genes in the different functional subclasses is given in the coloured boxes. FA, fatty acid biosynthetic genes; PS, photosynthesis-related genes; CWP, cell wall polysaccharides.

regulators of seed maturation and oil accumulation in model plants, we investigated the functions of their respective top 200 most positively connected genes according to the value of the Pearson coefficient in coexpression analysis. The top 200 coexpression partners of both *FUS3* and *WRI1* displayed a significant enrichment for lipid and cell wall related functions, including fatty acid synthesis, cellulose synthesis and cell wall proteins (Fig. 2C). Genes involved in glycolysis and tricarboxylic acid (TCA) cycle were also significantly overrepresented among the top 200 partners of *WRI1*. Although significantly expressed during endosperm development (Supplementary Fig. S3), the two seed master regulators LEL1 (LEC1-like) and ABI3 were not present in the GMB coexpression network. The expression profile of *LEL1* was very similar in all the coffee species, but with a transcription peak at ST3 prior to that of galactomannan biosynthesis genes (Supplementary Fig. S3). *ABI3* did not display a shared uniform transcriptional profile among the fourteen coffee species studied and therefore had no coexpression partners, even at an |R| threshold as low as 0.7 (Supplementary Fig. S3; Supplementary Table S10).

3.4. The AP2/ERF transcription factor DREB2D is a central node of the core galactomannan biosynthesis module

DREB2D (Cc10g02270) belongs to Module 2 of the GMB network and was the only TF that was a direct partner (first round of coexpression analysis) of the guide genes *MANS1*, *MAGT1* and *VTC1* (Fig. 2A). Phylogenetic analysis of DREB2D together with *Arabidopsis thaliana* members of DEHYDRATION-RESPONSIVE ELEMENT BINDING (DREB) subfamily A-2 of the APETALA2/ethylene-responsive factor (AP2/ERF) family identified DREB2D as a close relative to AtDREB2D (AT1G75490) and AtDREB2G (AT5G18450) (Supplementary Fig. S4). It is worth

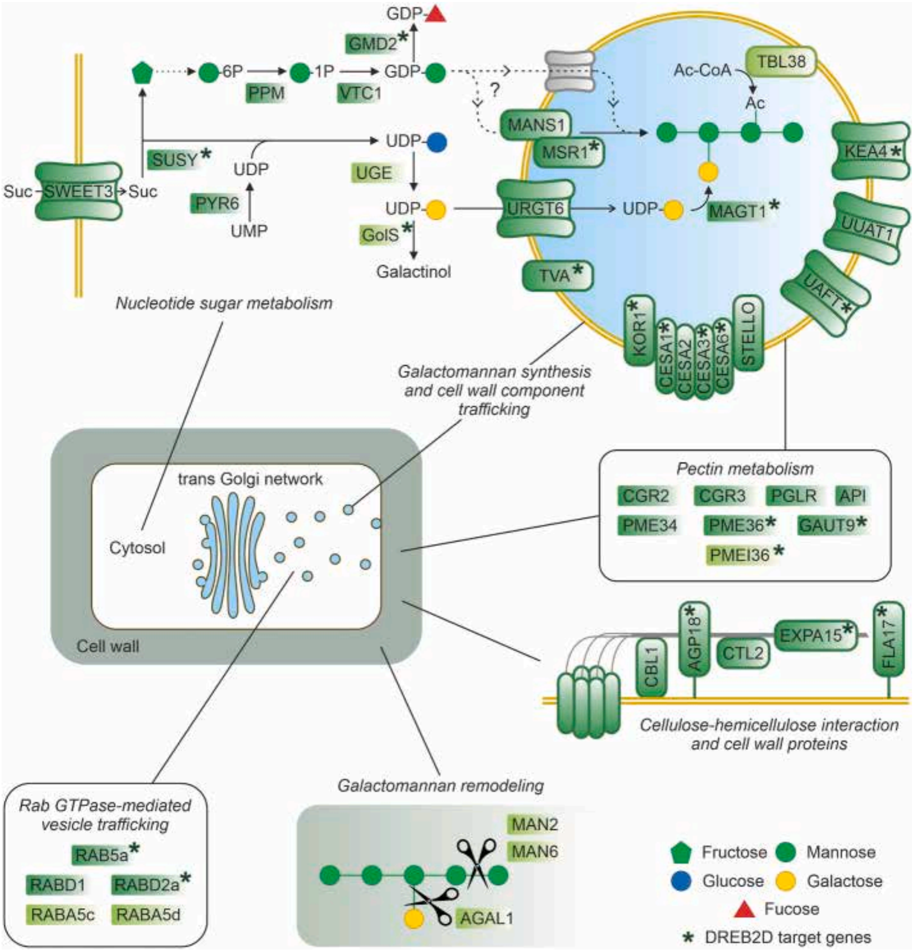

Fig. 3. Enzymes and proteins of the GMB coexpression network involved in the synthesis, deposition and modification of cell wall storage polysaccharides in the coffee endosperm. Scheme of the major biosynthetic pathways and metabolic processes occurring in different endosperm cell compartments. All enzyme and protein names are explained in Supplementary Tables S7 and S9. Enzymes and proteins highlighted in dark green are nodes of the GMB coexpression network. Enzymes and proteins followed by an asterisk are encoded by putative DREB2D target genes, as revealed by DAP-seq analysis. Enzymes and proteins highlighted in light green are nodes of the ST3-ST5 gene coexpression network. The scheme was adapted from Fig. 2 of the review by Voiniciuc (2022).

mentioning that *TVA* and *MSR1* were also direct coexpression partners of *DREB2D* in Module 2, with a Pearson coefficient of 0.90 and 0.93, respectively. Moreover, genes involved in the metabolism of cell wall polysaccharides and cellulose synthesis were significantly enriched in the top 200 coexpression partners of *DREB2D* (Fig. 2C). Finally, *DREB2D* was the only TF primarily connected to GMB guide genes, *MSR1*, *TBL38*, and *MAN6* in the coexpression network built using ST3-ST5 RNA-seq data (Supplementary Fig. S2). Gene coexpression network analysis using galactomannan biosynthetic genes as guide genes therefore provides evidence that DREB2D is a major candidate for direct transcriptional regulation of CWP biosynthesis and accumulation.

3.5. Overexpression of DREB2D in C. arabica somatic embryos alters soluble sugar composition

Transgenic *C. arabica* somatic embryos overexpressing the *DREB2D* cDNA sequence under the control of the constitutive 35S promoter of the Cauliflower mosaic virus were regenerated using *Agrobacterium*-mediated transformation. Real-time RT-PCR analysis showed a 6- to 250-fold increase in the expression of the *DREB2D* transgene in the eight independent transformed lines tested compared to control lines (Fig. 4A; Supplementary Table S11). The CWP content of DREB2D+ somatic embryos and controls did not differ significantly (Fig. 4B). No significant differences in the monosaccharide composition of CWPs was observed between control and DREB2D+ lines (Fig. 4C), either for mannose content or the major residues, namely arabinose, glucose, galactose and xylose, which together account for more than 80 % of total CWP monosaccharides. By contrast, soluble sugar profiling revealed that *DREB2D* overexpression altered the soluble sugar content, with a significant increase in raffinose and stachyose contents and a significant decrease in the inositol content (Fig. 4D). Raffinose increased dramatically, with accumulation levels more than 20 times higher in DREB2D+ lines (1.2 ± 0.5 %DM) than in controls (<0.1 %DM).

3.6. DREB2D modulates the expression of genes related to cell wall metabolism

An Illumina-based RNAseq experiment was then conducted using six DREB2D+ lines and four control lines (Supplementary Table S2). Deseq2 analysis revealed 881 differentially expressed genes (DEGs, Padj<0.01), including 532 up-regulated and 349 down-regulated genes in DREB2D+ somatic embryos (Supplementary Table S12). Among the significantly upregulated genes were several key genes of CWP biosynthesis and modification, including those encoding UGE1, UGE2, UGE3, AGAL1, MSR1, TRANVIA, PPM, TBL38, TBL42 and three mannan endo-1,4-beta-mannosidases (Fig. 5A; Supplementary Table S12). Of the 532 up-regulated genes, 110 genes showed a more than 4 times transcriptional induction in DREB2D+ transgenic lines compared to control embryos. Mapman enrichment analysis showed that this top 110 gene subset is significantly enriched in cell wall metabolism and sugar

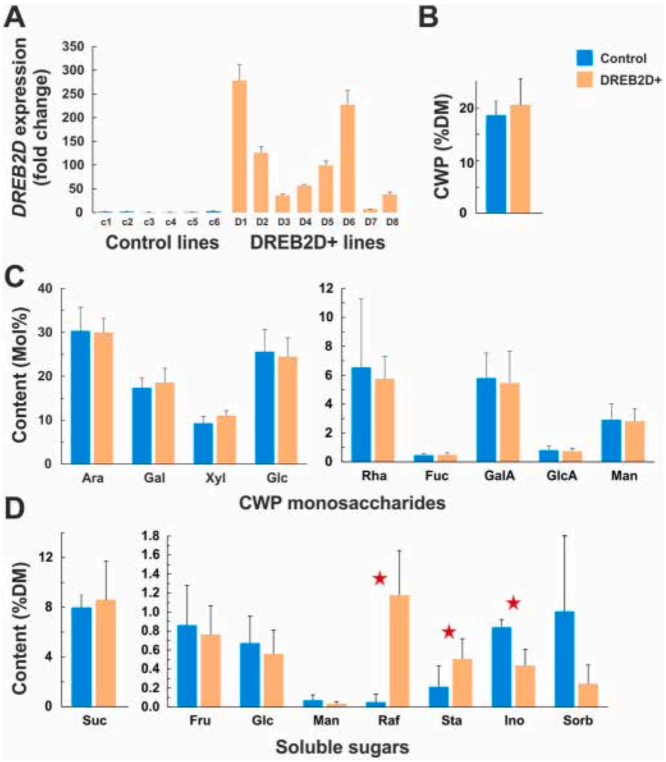

Fig. 4. Influence of DREB2D overexpression on the cell wall polysaccharide content and composition, and the soluble sugar contents of *C. arabica* somatic embryos. **A.** DREBD2 expression level as measured by Real-time RT-PCR in six control and eight DREBD2+ lines. **B.** Cell wall polysaccharide (CWP) content in control and DREBD2+ somatic embryos. **C.** Monosaccharide composition of CWP in control and DREBD2+ somatic embryos. **D.** Soluble sugar contents in control and DREBD2+ somatic embryos. Ara, arabinose; Fru, fructose; Fuc, fucose; Gal, galactose; GlcA, glucuronic acid; GalA, galacturonic acid; Glc, glucose; Ino, myo-inositol; Man, mannose; Raf, raffinose; Rha, rhamnose; Sorb, sorbitol; Sta, stachyose; Suc, sucrose; Xyl, xylose. The effect of DREB2D overexpression was tested using one-way ANOVA. Significant effects are indicated with an asterisk for P < 0.01.

metabolism functions (Fig. 5B). In addition, several late maturation functions were significantly enriched, including ABA metabolism, late embryogenesis abundant (LEA) proteins and abiotic stress response (Supplementary Table S13). In particular, nine genes encoding LEA proteins, as well as nine genes encoding small heat shock proteins (sHSP) were detected among up-regulated DEG in DREB2D+ somatic embryos (Supplementary Table S12).

Putative DREB2D target genes were also scrutinized using linear regression, based on the assumption of a quantitative relationship between the expression level of *DREB2D* and that of genes downstream of DREB2D. Correlations found between *DREB2D* and target gene expression levels revealed 38 candidate genes with a R^2 higher than 0.9 (FDR<0.01) (Supplementary Table S14). This subset of 38 candidate target genes contains key genes of CWP synthesis and modification, including *UGE1*, *UGE2*, *TBL38*, *AGAL1*, and the cellulose synthase-like *CSLC4* (Cc01g05870), whose orthologue is involved in xyloglucan synthesis in Arabidopsis (Kim et al., 2020) (Fig. 5C). Finally, the subset of genes quantitatively co-expressed with *DREB2D* in transformed somatic embryos include proteins typical of the seed maturation programme such as the LEA dehydrin Dh4 (Cc06g20920), the oil-body associated protein OBAP2B (Cc02g16040) and the lipid-droplet associated protein LDAP3 (Cc03g10150) (Fig. 5C).

Interpretation of functional enrichment of genes down-regulated in DREB2D+ lines is less straightforward because the cellular context of somatic embryos is very different from that of the endosperm, and these genes may not play any role during seed development. However, alpha-mannosidase genes (Cc04g05900, Cc09g07970), as well as several genes involved in pectin modifications such as polygalacturonases (Cc03g13740, Cc09g05370) and rhamnogalacturonate lyase (Cc10g08340), were significantly down-regulated in somatic embryos that overexpressed DREB2D. Furthermore, Mapman-based enrichment analysis significantly highlighted many processes that co-occur in plastids, including synthesis of thiamine, salicylic acid precursors, monoterpenes and carotenoids (Supplementary Table S13). In addition, many genes encoding components of the photosynthetic electron transfer chain were also detected among down-regulated genes. It is worth noting that visible differences in pigmentation were observed between DREB2D+ and control somatic embryos, most probably related to lower carotenoid and chlorophyll contents in DREB2D+ lines (Supplementary Fig. S5). These results suggest a role for DREB2D in the structural and functional changes in plastids associated with seed maturation. Another striking observation was the presence of all the key genes for caffeine biosynthesis (*XMT*, *MXMT* and *DXMT*) in the set of genes down-regulated in DREB2D+ lines. Since these genes are well-known markers of perisperm and early endosperm developmental stages (Denoeud et al., 2014), this result suggests that DREB2D may also act as a broad repressor of key processes of early endosperm development, thereby facilitating and synchronising entry into the maturation phase.

3.7. DREB2D binds the promoter of genes of the GMB coexpression network

DREB2D belongs to a class of TF proteins known to bind to C-repeat or dehydration response elements (CRT/DRE) in gene promoters (Jiang et al., 1996). An electrophoretic mobility shift assay (EMSA) was therefore performed to test CcDREB2D binding to artificial DRE elements *in vitro* (Fig. 6A). Binding of the DREB2D-His recombinant protein to the core DRE element (CCGAC) provided evidence for a retarded band in the gel, and for the absence of competition for binding using a test probe containing a mutated DRE element. DREB2D appeared able to bind both DRE (T**A**CCGAC) and CRT (T**G**CCGAC) elements. DNA affinity purification sequencing (DAP-seq) assays were then conducted to identify the target genes of DREB2D and to refine the putative motifs bound by DREB2D. Significant DREB2D-binding peaks (fold-enrichment >3 and FDR<0.05) common to the three DAP-seq replicates were found in the 2 kb promoter region of 3421 genes (Supplementary Table S15). This is illustrated in Fig. 6B with Cc05g12390 which encodes TRANVIA. GO-enrichment analysis performed on DREB2D target genes revealed significant enrichement in galactose metabolic processes, cellular glucose homeostasis and glucose binding (Supplementary Table S16). Moreover, many genes of the GMB coexpression network (32.2 %) were identified as DREB2D target genes by DAP-seq analysis (Fig. 3, Supplementary Table S7). These genes are involved in all the different processes of CWP biosynthesis, including sucrose cleavage (*SUSY*), nucleotide sugar metabolism and transport (*GMD2*, *UAFT*), galactomannan synthesis (*MSR*, *MAGT*), cellulose synthesis (*CESA1*, *CESA3*, *CESA6*, *KOR1*), pectin synthesis and modification (*GAUT9*, *PME36*, *PMEI36*), secretory vesicle trafficking (*TVA*, *SCAMP4*, *RAB5a*, *RABD2a*) and cell wall protein synthesis (*AGP18*, *FLA17*). MEME analysis of 2 kb promoter sequences of the DREB2D target genes resulting from DAP-seq analysis identified two highly significant (p = 6E-113 and 1E-15, respectively) C-repeat motifs potentially bound by DREB2D, hereafter referred as to DREB2D-BS1 and –BS2 (Fig. 6C). Neither motif contained a distinct CCGAC core motif of DRE *cis*-acting elements. However, the second motif is very similar to that bound by AtDREB2D, with significant Tomtom motif-motif similarity (FDR = 1.75E-06). The elements were then searched for in the proximal upstream region (2 kb) of eight selected genes of the GMB network representative of the different processes of CWP synthesis. All eight promoters contained either DREB2D-BS1 or DREB2D–BS2, or both motifs (Fig. 6C). For instance, DREB2D-BS1 was retrieved in the proximal upstream region of

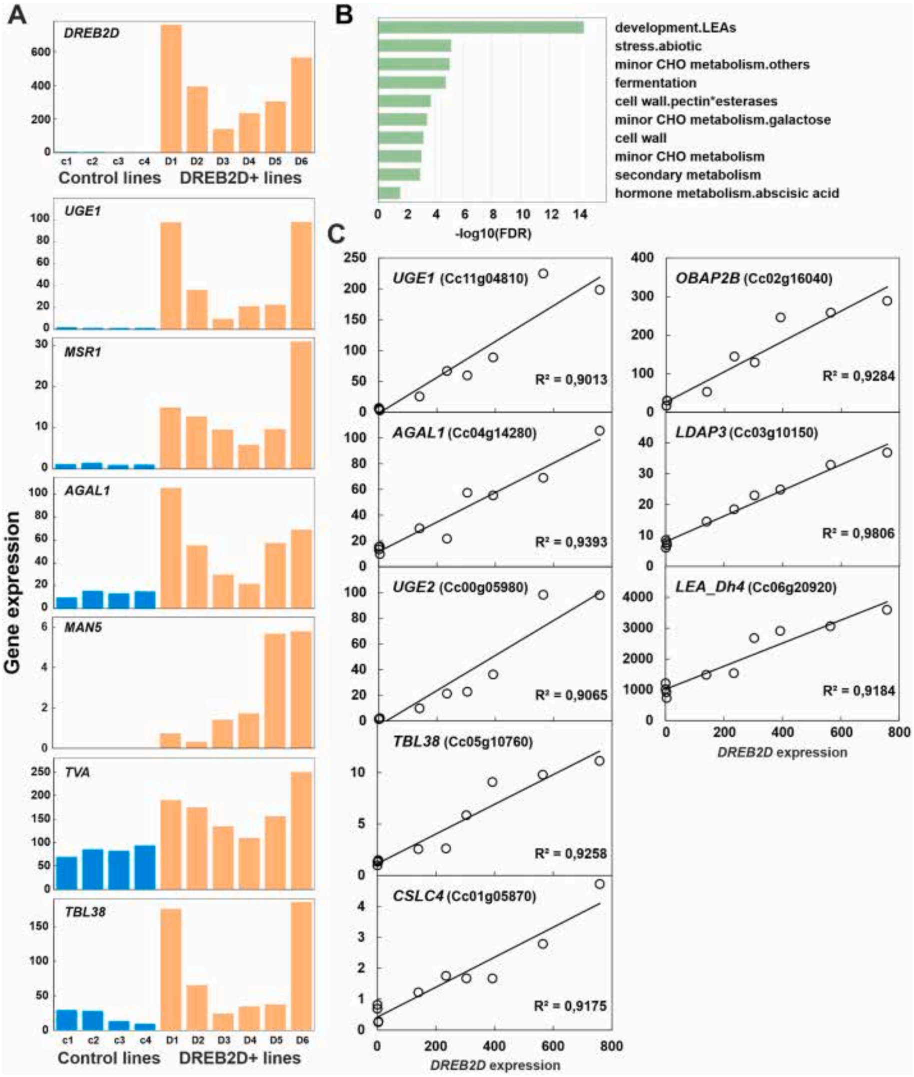

Fig. 5. Functional validation of DREB2D in coffee somatic embryos. **A**. Transcript amounts measured by RNA-seq of *DREB2D* and six putative target genes (*UGE1*, *MSR1*, *AGAL1*, *MAN5*, *TVA* and *TBL38*) in four control lines (blue bars) and six DREBD2+ lines (orange bars). **B.** False discovery rate (FDR) of the functional terms that show significant enrichment in the 110 genes which were massively (4 fold) upregulated in DREB2D+ transgenic somatic embryos compared to in control somatic embryos. Values are means (±SD) of four technical replicates for eight independent and six control lines transformed with empty PMDC vector. Real-time RT-PCR analysis of expression of key galactomannan biosynthetic genes in DREB2D+ transgenic somatic embryos. **C.** Highly significant positive correlations ($R > 0.9$) between transcript amounts (rpkm) of *DREB2D* and eight putative target genes (*UGE1, AGAL1, UGE2, TBL38*, *CSLC4*, *OBAP2D*, *LDAP3*, and *LEA_Dh4*). AGAL1, alpha-galactosidase; AGL, agamous-like; CSLC, cellulose-synthase-like-C; LDAP, lipid-droplet associated protein; LEA-dh, late embryogenesis abundant protein, dehydrin; OBAP2B, oil body associated protein 2B; TBL38, trichome-birefrengence-like, UGE, UDP-glucose 4-epimerase.

UGE1, but not DREB2D-BS2, *MAGT1* displays the opposite case, and both motifs were identified at the same locations in the promoter of *MANS1*.

4. Discussion

4.1. Transcriptional orchestration of synthesis, secretion and post-depositional modifications of galactomannans in the maturing endosperm

Thanks to the considerable variation in the endosperm galactomannan content observed between coffee species, we chose to use the interspecific genetic variability to build the gene coexpression network of endosperm maturation. The advantage of the mixed strategy (Guerin et al., 2016) used here to build a gene coexpression network, compared with a strict guide-gene approach (Joët et al., 2014), is being able to reveal novel key players of a biosynthetic pathway without prior information. Our gene coexpression network analysis first identified *MSR1* in the core GMB module that contains the three guide genes *MANS1*, *VTC1*, and *MAGT1*. Gene coexpression network analysis also provided evidence for concerted transcriptional regulation, along with galactomannan biosynthesis, of sucrose import and cleavage activities (*SWEET3*, *SUSY1*), as well as of glycolytic activity (*HXK-5*, *PGLYM-p*), nucleotide sugar metabolism for the synthesis of GDP-man and UDP-gal (*PPM*, *VTC1*, *UGE1*, *UGE2*, *UGE3*), and their import into Golgi (*URGT6*).

In addition to the upstream transcriptional control of central carbon metabolism and nucleotide sugar synthesis and transport which provide galactomannan building blocks, gene coexpression network analysis also revealed the tight orchestration of the multiple players involved in the functioning of Golgi and the trans-Golgi network, as well as the post-Golgi secretory pathway to the extracellular space. For instance, the GMB network included SCAMP4, one of the secretory carrier membrane proteins used as conventional markers for secretory vesicles, of which SCAMP2 is involved in the vesicular trafficking of pectin (Toyooka et al., 2009), plus key proteins STELLO and TRANVIA which play major roles in the assembly and delivery of CSCs. Similarly, RAB-GTPases, of which several members are present in the GMB network, play specialised roles

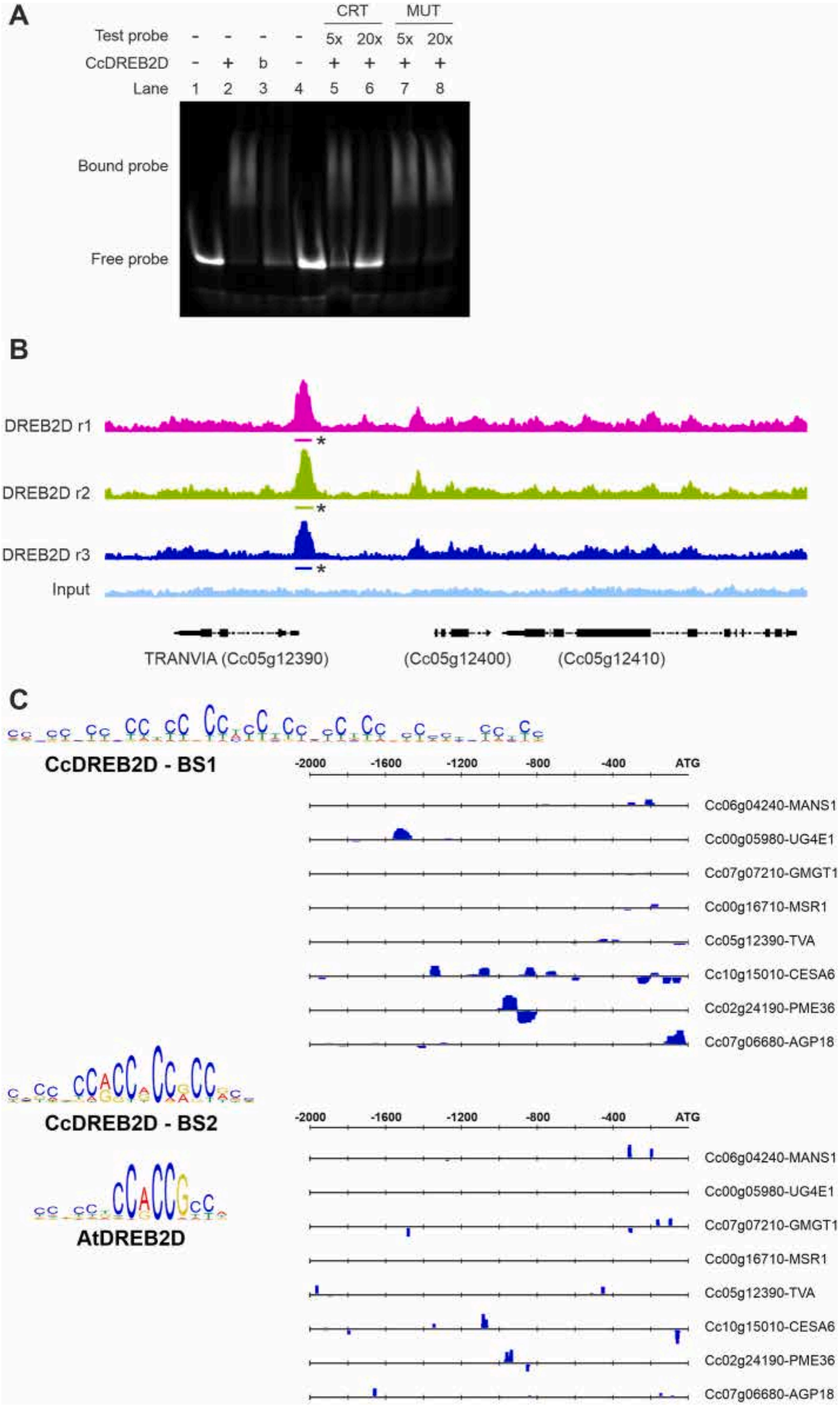

Fig. 6. DREB2D binds the promoter of genes of the GMB coexpression network. A. Electrophoretic Mobility Shift Assay (EMSA) of DREB2D and dsDNA probes. Imaging was done by detection of the labelled DRE probe. CcDREB2D protein was omitted in lanes indicated by '-', and included in lanes indicated by '+', in lanes indicated by 'b' it was boiled for 15 min at 98 °C. Labelled DRE probe was present in each lane (200 ng). Test probes (CRT, MUT) added to challenge the binding of the labelled DRE probe are indicated, (5x and 25x: five or 25 times excess compared to labelled DRE, '-' indicates no added test probe). **B.** DAP-seq DNA-binding profile of DREB2D (three replicates, r1-3) and the control (input). Horizontal bars below the plots represent the position of the significant peak regions in the promoter of *TRANVIA*. **C.** Logos of the two motifs (CcDREB2D-BS1 and -BS2) significantly enriched in the promoter sequences of the DREB2D target genes resulting from DAP-seq analysis. Number, position and strand of the two motifs in the proximal upstream region of eight selected genes of the GMB network.

in the transport of the different types of polysaccharides (cellulose, hemicellulose, pectin) from Golgi to the cell surface (Lunn et al., 2013). Furthermore, the GMB network includes the ECA4 component of the TPLATE adaptor complex, a major module for clathrin-mediated endocytosis (Gadeyne et al., 2014), for which evidence of a direct role in the recognition of cellulose synthase complexes for their internalization was recently put forward (Sánchez-Rodríguez et al., 2018). This is important since glycosyltransferases such as MANS and MAGT, require retrograde transport from secretory vesicles and late Golgi cisternae (trans) to earlier cisternae (cis) to maintain their steady-state spatial distribution across the Golgi stack (Zabotina et al., 2021). Finally, the GMB network revealed numerous receptor-like protein kinases which may help regulate CWP biosynthesis, since several receptor-like kinases have been shown to specifically control cellulose synthesis (Hématy et al., 2007) or to mitigate secondary cell wall thickening (Huang et al., 2018).

Coexpression analysis also highlighted the concerted regulation of galactomannan synthesis and post-depositional modifications, including the degree of galactose substitution (*AGAL1*), mannan remodelling (*MAN2* and *MAN6*) through transglycosylation (Prakash et al., 2012; Jacobson et al., 2025; Kikuchi et al., 2025) and acetylation (*TBL38*). Within the TBL family (DUF 231), TBL38 does not belong to subgroup II, whose members, including AXY4, are mannan O-acetyltransferases (MOATs) in Arabidopsis (Zhong et al., 2018). Rather, TBL38 belongs to subgroup IX, which includes PMR5, a functional acetyltransferase that mediates pectin acetylation (Chiniquy et al., 2019). The GMB network also includes three GDSL esterase genes close to the rice DARX1 protein, a Golgi-located enzyme that deacetylates the side chain of arabinoxylans and dynamically participates in acetylation/deacetylation of cell wall polymers during development (Zhang et al., 2017, 2019), acetylation of xylans being critical for interaction with cellulose fibrils in cell walls (Grantham et al., 2017).

4.2. Intertwining of the transcriptional control exerted on the specific biosynthetic machineries of the different cell wall polysaccharides

The pathway-guided gene coexpression network revealed tight transcriptional coordination of the galactomannan biosynthesis machinery with those of the other components of coffee endosperm CWP, *i. e.* arabinogalactan proteins, cellulose, and, to a lesser extent, pectic polysaccharides (Redgwell and Fischer, 2006; Li et al., 2021). Two different sets of CESA complexes are mobilised for primary cell wall synthesis during cell expansion and secondary cell wall thickening, respectively (McFarlane et al., 2014). It is worth noting that the *CESA* genes which are connected to galactomannan biosynthetic genes, *CESA1*, *CESA3*, and *CESA6*-like genes (*CESA2* and *CESA6*), are those encoding the three subunits that constitute the CSC recruited for the synthesis of the primary cell wall (Desprez et al., 2007; Persson et al., 2007). Moreover, the GMB network contains several genes involved in the metabolism of pectin, which mainly accumulates in primary walls. Collectively, these results thus suggest that storage of polysaccharides in the coffee endosperm only takes place in primary walls and corroborate previous immunohistological observations of coffee endosperm cell walls showing that the β-1,4-mannan-specific monoclonal antibody BGM C6 labels cross the entire wall cross section (Sutherland et al., 2004).

While the biosynthetic machinery of the different cell wall polysaccharides is highly coordinated at the transcriptional level, how coffee endosperm cells control the synthesis of large amounts of galactomannans in comparison with cellulose and pectin remains to be determined. The recruitment of three distinct UGE genes (*UGE1*, *UGE2*, *UGE3*) may help shift partitioning of most of the pool of UDP-Glu resulting from SUSY activity towards the synthesis of the galactomannan building block UDP-gal rather than cellulose, particularly in a cellular context where fatty acid synthesis also requires large supplies of sugar. In Arabidopsis, different coexpressed *UGE* genes act synergistically for CWP synthesis (Rösti et al., 2007; Seifert et al., 2002). Moreover, transporters of specific nucleotide sugars in the Golgi apparatus usually work unidirectionnally, and are key rate-limiting control points towards different types of CWP (Kleczkowski and Igamberdiev, 2020), as shown with UDP-gal transporters (Rautengarten et al., 2014; Abedi et al., 2016).

4.3. DREB2D, a secondary TF that plays a critical role in rewiring nucleotide sugar metabolism to galactomannan precursor synthesis

The present study provides considerable experimental evidence that DREB2D is a key TF for cell wall synthesis in the coffee endosperm: (i) it is the only TF of module 2 of the GMB network that groups numerous genes of the core galactomannan biosynthetic machinery, (ii) DREB2D is the only TF that is a direct partner of the guide genes *MANS1*, *MAGT1*, and *VTC1*, (iii) somatic embryos overexpressing *DREB2D* display significant up-regulation of numerous key genes of CWP biosynthesis and modification, including those encoding UGE1, UGE2, UGE3, AGAL1, MSR1, TRANVIA, PPM, TBL38, TBL42 and mannan endo-1,4-beta-mannosidases, (iv) DAP-seq analysis of DREB2D target genes identified many genes of the GMB coexpression network involved in sucrose cleavage, nucleotide sugar metabolism and transport, galactomannan synthesis, cellulose synthesis, pectin synthesis and modification, secretory vesicle trafficking and cell wall protein synthesis. Therefore, although gene coexpression network analysis is not sufficient to establish regulatory relationships, the present work illustrates the potential of this approach is to identify novel candidate regulators. DREB2D belongs to a TF family that is usually involved in abiotic stress tolerance (Agarwal et al., 2017). In Arabidopsis, overexpression of *DREB2D* partially rescues the desiccation sensitivity of the *abi3-5* mutant (González-Morales et al., 2016). In coffee, *DREB2G* is significantly upregulated during seed desiccation tolerance acquisition at the late maturation stage and *DREB2G* overexpression in somatic embryos triggers heat-shock proteins accumulation (Dussert et al., 2018). However, a recent study showed that PrDREB2D regulates the linolenate accumulation in oily *Paeonia rockii* albuminous seeds, by directly activating the expression of *PDAT2*, which codes for a phospholipid:diacylglycerol acyltransferase (Yang et al., 2024). PDAT plays a key role in the enrichment of seed oil in polyunsaturated fatty acids (Dahlqvist et al., 2000). As for the coffee DREB2D, the Arabidposis closest relative of PrDREB2D is AtDREB2D, indicating that the function of members of DREB subfamily A-2 of the AP2/ERF family is not restricted to abiotic stress tolerance but may also be involved in storage reserve accumulation during albuminous seed maturation. This important recent finding in tree peony (Yang et al., 2024) prompted us to investigate whether coffee DREB2D could also play a role in the high polyunsaturated fatty acid content of coffee endosperm oil (Joët et al., 2009). Interestingly, Cc06g11860, which codes for PDAT, and Cc06g11620, Cc07g12310 and Cc03g00040, which encode diacylglycerol acyltransferases, were identified as DREB2D target genes by DAP-seq analysis (Supplementary Table S16), suggesting DREB2D could also be involved in the regulation of triacylglycerol assembly in the coffee endosperm.

Because the three key genes *VTC1*, *MANS1* and *MAGT1* were not upregulated in DREB2D+ somatic embryos, additional cellular/molecular factors may be required to fully activate the galactomannan biosynthetic machinery and trigger their accumulation. In this respect, it is worth mentioning that PrDREB2D exerts its regulation of linoleate accumulation in the tree peony endosperm by recruiting the cofactor ABA-response element binding factor PrABF2b (Yang et al., 2024). Similarly, the functional interaction between the TFs HaDREB2 and HaHSFA9 synergistically trans-activates target sHSP genes during sunflower seed development (Díaz-Martín et al., 2005). When overexpressed in tobacco, HaDREB2 alone does not up-regulate sHSP genes, whereas, when co-expressed with HaHSFA9, it increases sHSP accumulation and improves seed longevity (Almoguera et al., 2009). Interestingly, *HSFA9* was identified in our ST3-ST5 coexpression network and may be an interesting candidate for further investigation of galactomannan biosynthesis in coffee. Finally, although it takes several years to obtain coffee plants that produce seeds starting from transformed somatic embryos, characterizing the seed transcriptome and polysaccharide composition of plants recovered from DREB2D+ somatic embryos will certainly provide additional information on galactomannan accumulation in albuminous seeds.

In the absence of functional galactomannan machinery in somatic embryos overexpressing *DREB2D*, the dramatic up-regulation of *UGE* genes presumably led to increased production of UDP-galactose and its diversion to other pathways such as the biosynthesis of raffinose family oligosaccharides. Indeed, raffinose and stachyose contents were found to be significantly increased in DREB2D+ lines. These results corroborate our previous findings in *C. arabica* seeds in which raffinose and stachyose were identified as a transient storage form that buffers excess UDP-galactose before it is remobilised and supplies auxiliary sources of building blocks for galactomannan synthesis (Joët et al., 2014). Since GolS1, which catalyses the first reaction of the raffinose family oligosaccharide pathway, uses UDP-gal and *myo*-inositol as a substrate, this could be a simple explanation for the smaller amounts of *myo*-inositol detected in DREB2D+ lines. Our results therefore point to a major role for DREB2D in controlling UDP-gal homeostasis during galactomannan accumulation in the coffee endosperm.

4.4. Mobilisation of an original set of master regulators for coffee endosperm maturation

LEC1-ABI3/FUS3/LEC2 (LAFL) proteins are well-characterised key regulators of embryo maturation, including oil and protein accumulation, in exalbuminous seeds (Fatihi et al., 2016; Alizadeh et al., 2021). Among these four so-called master regulators of seed development, only *FUS3* was retrieved in the GMB network of the coffee endosperm and the functions of its coexpression partners in the network suggest it plays a central role in oil and CWP accumulation. It should be noted that there is no orthologue of *AtLEC2* in the coffee genome. The function of FUS3 and ABI3 relatives has been suggested to be conserved in the dead endosperm of cereal caryopsis, but not that of LEC2 (Grimault et al., 2015). For instance, FUS3 displayed a conserved functionality in the barley endosperm for the control of storage protein accumulation, as shown by complementation of the loss-of-function *fus3* mutant in Arabidopsis (Moreno-Risueno et al., 2008). *FUS3* is significantly expressed in the transient endosperm of Arabidopsis seeds, which also stores oil, though ten-fold less than that stored in the embryo (Barthole et al., 2014). In Arabidopsis, FUS3 controls the synthesis of storage lipids, either by directly activating fatty acid biosynthetic genes or indirectly through the activation of *WRI1* (Wang and Perry, 2013).

WRI1 appeared to be a key regulatory hub in the GMB network of the coffee endosperm. WRI1 directly activates genes involved in late glycolysis and fatty acid synthesis during Arabidopsis seed maturation (Baud et al., 2007). WRI1 is thought to play a ubiquitous role in oil synthesis in plants (Ma et al., 2013). In the present work, we observed that numerous coexpression partners of *WRI1* are involved in fatty acid synthesis, glycolysis and in the TCA cycle, suggesting it may play a similar role in the coffee endosperm. Transcript profiling showed it is expressed concomitantly with fatty acid biosynthetic genes in the oily endosperm of oil palm (Dussert et al., 2013), *Paeonia ostia* (Yuan et al., 2022) and castor bean (Yang et al., 2019). The ability of the oil palm endosperm-specific WRI1-2 to trigger fatty acid synthesis has already been validated using transient expression in tobacco leaves (Dussert et al., 2013) and castor *WRI1* shown to complement Arabidopsis *wri1* mutants (Yang et al., 2019).

Our results thus suggest a conserved role for FUS3 and WRI1 as master TFs for storage compound accumulation in the cellular endosperm of coffee seeds. It is worth noting that WRI1 not only functions downstream of LEC2 (Baud et al., 2007) but also of FUS3 in Arabidopsis (Yamamoto et al., 2010). In addition to the canonical role of FUS3 and WRI1, the massive orientation of the carbon flow towards galactomannans may have been assisted by the selection of a particular assemblage of TFs during plant evolution. The identification of *SHINE2* (*SHN2*) in the GMB network is particularly striking in this context. Indeed, members of the SHINE clade of AP2 domain transcription factors were first shown to activate wax biosynthesis and to alter cuticle properties in Arabidopsis (Aharoni et al., 2004). However, they were later also described as key regulators of cellulose and hemicellulose biosynthesis pathways in rice and tobacco (Ambavaram et al., 2011; Liu et al., 2017). When ectopically expressed in these species, Arabidopsis and poplar *SHN2* were shown to coordinate the up-regulation of cellulose and other primary cell wall compound biosynthesis, as well as the down-regulation of lignin biosynthesis and many secondary NAC and MYB TFs that trigger secondary wall biosynthesis (Zhong and Ye, 2015). Microscopy and chemical analyses of sclerenchyma cells of rice leaves overexpressing *AtSHN2* further revealed a marked increase in cell wall thickness (around 45 %) compared to the wild type, mainly due to the increased deposition of cellulose and hemicelluloses in place of lignin (Ambavaram et al., 2011). Given that the gene regulatory network associated with SHN2, with both activator and repressor functions, appears to be highly conserved in both monocot and dicot plants, and that the present coffee galactomannan-guided co-expression network mimics large parts of this transcriptional repertoire, it is tempting to attribute a similar role to SHN2 in triggering the peculiar primary cell wall thickening programme of coffee endosperm parenchyma cells, in particular through the recruitment of secondary TFs specialised in the activation of galactomannan biosynthetic genes.

CRediT authorship contribution statement

Stéphane Dussert: Writing – original draft, Visualization, Supervision, Resources, Formal analysis, Data curation, Conceptualization. **Anna K. Stavrinides:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Julien Serret:** Writing – review & editing, Methodology, Investigation. **Virginie Vaissayre:** Writing – review & editing, Methodology, Investigation. **Marie-Christine Combes:** Writing – review & editing, Methodology, Investigation. **Fabienne Morcillo:** Writing – review & editing, Formal analysis, Data curation. **Eveline Lefort:** Writing – review & editing, Investigation. **Stéphanie Rialle:** Writing – review & editing, Methodology, Formal analysis. **Hervé Etienne:** Writing – review & editing, Methodology. **Philippe Lashermes:** Writing – review & editing, Methodology, Formal analysis, Data curation. **Thierry Joët:** Writing – original draft, Visualization, Supervision, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization.

Funding

Anna K. Stavrinides was financially supported by the European Molecular Biology Organisation (EMBO) under a long-term fellowship grant [ALTF 478-2017]. MGX acknowledges financial support from France Génomique National infrastructure, funded as part of “Investissement d’Avenir” program managed by Agence Nationale pour la Recherche [ANR-10-INBS-09].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors thank the *Coffea* Biological Resources Center (BRC Coffea, maintained by IRD and CIRAD in Reunion Island) for providing the plant material and the Plant Protection Platform (3P, IBiSA) for hosting visiting researchers. The authors are also grateful to Dr. Sylvain

Santoni and to the staff of the AGAP Institut platform for preparing the cDNA libraries. The authors acknowledge the ISO 9001 certified IRD i-Trop HPC (member of the South Green Platform) at IRD Montpellier for providing HPC resources that have contributed to the research results reported within this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plaphy.2025.110771.

Data availability

The transcriptome dataset has been deposited at the European Nucleotide Archive (ENA) under project numbers PRJEB32533 and PRJEB79959.

References

Abedi, T., Khalil, M.F.M., Asai, T., Ishihara, N., Kitamura, K., Ishida, N., Tanaka, N., 2016. UDP-galactose transporter gene hUGT1 expression in tobacco plants leads to hyper-galactosylated cell wall components. J. Biosci. Bioeng. 121, 573–583.

Agarwal, P.K., Gupta, K., Lopato, S., Agarwal, P., 2017. Dehydration responsive element binding transcription factors and their applications for the engineering of stress tolerance. J. Exp. Bot. 68, 2135–2148.

Aguirre, M., Kiegle, E., Leo, G., Ezquer, I., 2018. Carbohydrate reserves and seed development: an overview. Plant Reprod. 31, 263–290.

Aharoni, A., Dixit, S., Jetter, R., Thoenes, E., Arkel, G. Van, Pereira, A., 2004. The SHINE clade of AP2 domain transcription factors activates wax biosynthesis, alters cuticle properties, and confers drought tolerance when overexpressed in arabidopsis W inside a box sign. Plant Cell 16, 2463–2480.

Alizadeh, M., Hoy, R., Lu, B., Song, L., 2021. Team effort: combinatorial control of seed maturation by transcription factors. Curr. Opin. Plant Biol. 63, 102091.

Almoguera, C., Prieto-Dapena, P., Díaz-Martín, J., Espinosa, J.M., Carranco, R., Jordano, J., 2009. The HaDREB2 transcription factor enhances basal thermotolerance and longevity of seeds through functional interaction with HaHSFA9. BMC Plant Biol. 9, 75.

Ambavaram, M.M.R., Krishnan, A., Trijatmiko, K.R., Pereira, A., 2011. Coordinated activation of cellulose and repression of lignin biosynthesis pathways in rice. Plant Physiol. 155, 916–931.

Aniento, F., Medina Hernández, V.S. de, Dagdas, Y., Rojas-Pierce, M., Russinova, E., 2022. Molecular mechanisms of endomembrane trafficking in plants. Plant Cell 34, 146–173.

Aoki, K., Ogata, Y., Shibata, D., 2007. Approaches for extracting practical information from gene co-expression networks in plant biology. Plant Cell Physiol. 48, 381–390.

Bailey, T.L., Johnson, J., Grant, C.E., Noble, W.S., 2015. The MEME suite. Nucleic Acids Res. 43, W39–W49.

Barthole, G., To, A., Marchive, C., Brunaud, V., Soubigou-Taconnat, L., Berger, N., Dubreucq, B., Lepiniec, L., Baud, S., 2014. MYB118 represses endosperm maturation in seeds of Arabidopsis. Plant Cell 26, 3519–3537.

Bartlett, A., O'Malley, R.C., Huang, S.C., Galli, M., Nery, J.R., Gallavotti, A., Ecker, J.R., 2017. Mapping genome-wide transcription-factor binding sites using DAP-seq. Nat. Protoc. 12, 1659–1672.

Baud, S., Mendoza, M.S., To, A., Harscoët, E., Lepiniec, L., Dubreucq, B., 2007. WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in Arabidopsis. Plant J. 50, 825–838.

Boerjan, W., Burlat, V., Cosgrove, D.J., et al., 2024. Top five unanswered questions in plant cell surface research. The Cell Surface 11, 100121.

Brown, D.M., Zeef, L.A.H., Ellis, J., Goodacre, R., Turner, S.R., 2005. Identification of novel genes in arabidopsis involved in secondary cell Wall formation using expression profiling and reverse genetics. Plant Cell 17, 2281–2295.

Buckeridge, M.S., 2010. Seed cell wall storage polysaccharides: models to understand cell wall biosynthesis and degradation. Plant Physiol. 154, 1017–1023.

Cenci, A., Combes, M.C., Lashermes, P., 2012. Genome evolution in diploid and tetraploid *Coffea* species as revealed by comparative analysis of orthologous genome segments. Plant Mol. Biol. 78, 135–145.

Chiniquy, D., Underwood, W., Corwin, J., et al., 2019. PMR5, an acetylation protein at the intersection of pectin biosynthesis and defense against fungal pathogens. Plant J. 100, 1022–1035.

Cui, Y., Zeng, X., Xiong, Q., et al., 2021. Combining quantitative trait locus and co-expression analysis allowed identification of new candidates for oil accumulation in rapeseed. J. Exp. Bot. 72, 1649–1660.

Dahlqvist, A., Stahl, U., Lenman, M., Banas, A., Lee, M., Sandager, L., Ronne, H., Stymne, H., 2000. Phospholipid:diacylglycerol acyltransferase: an enzyme that catalyzes the acyl-coa-independent formation of triacylglycerol in yeast and plants. Proc. Natl. Acad. Sci. USA 97, 6487–6492.

Delmer, D., Dixon, R.A., Keegstra, K., Mohnen, D., 2024. The plant cell wall—dynamic, strong, and adaptable—is a natural shapeshifter. Plant Cell 36, 1257–1311.

Denoeud, F., Carretero-Paulet, L., Dereeper, A., et al., 2014. The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. Science 345, 1181–1184.

Desprez, T., Juraniec, M., Crowell, E.F., Jouy, H., Pochylova, Z., Parcy, F., Höfte, H., Gonneau, M., Vernhettes, S., 2007. Organization of cellulose synthase complexes involved in primary cell wall synthesis in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. 104, 15572–15577.

Dhugga, K.S., Barreiro, R., Whitten, B., et al., 2004. Guar seed β-Mannan synthase is a member of the cellulose synthase super gene family. Science 303, 363–366.

Díaz-Martín, J., Almoguera, C., Prieto-Dapena, P., Espinosa, J.M., Jordano, J., 2005. Functional interaction between two transcription factors involved in the developmental regulation of a small heat stress protein gene promoter. Plant Physiol. 139, 1483–1494.

Dussert, S., Chabrillange, N., Engelmann, F., Anthony, F., Louarn, J., Hamon, S., 2000. Relationship between seed desiccation sensitivity, seed water content at maturity and climatic characteristics of native environments of nine *Coffea* L. species. Seed Sci. Res. 10, 293–300.

Dussert, S., Chabrillange, N., Rocquelin, G., Engelmann, F., Lopez, M., Hamon, S., 2001. Tolerance of coffee (*Coffea* spp.) seeds to ultra-low temperature exposure in relation to calorimetric properties of tissue water, lipid composition, and cooling procedure. Physiol. Plantarum 112, 495–504.

Dussert, S., Davey, M.W., Laffargue, A., Doulbeau, S., Swennen, R., Etienne, H., 2006. Oxidative stress, phospholipid loss and lipid hydrolysis during drying and storage of intermediate seeds. Physiol. Plantarum 127, 192–204.

Dussert, S., Serret, J., Bastos-Siqueira, A., Morcillo, F., Déchamp, E., Rofidal, V., Lashermes, P., Etienne, H., Joët, T., 2018. Integrative analysis of the late maturation programme and desiccation tolerance mechanisms in intermediate coffee seeds. J. Exp. Bot. 69, 1583–1597.

Dussert, S., Guerin, C., Andersson, M., Joët, T., Tranbarger, T.J., Pizot, M., Sarah, G., Omore, A., Durand-Gasselin, T., Morcillo, F., 2013. Comparative transcriptome analysis of three oil palm fruit and seed tissues that differ in oil content and fatty acid composition. Plant Physiol. 162, 1337–1358.

Edwards, M.E., Scott, C., Gidley, M.J., Reid, J.S.G., 1992. Control of mannose/galactose ratio during galactomannan formation in developing legume seeds. Planta 187, 67–74.

Edwards, M.E., Dickson, C.A., Chengappa, S., Sidebottom, C., Gidley, M.J., Reid, J.S.G., 1999. Molecular characterisation of a membrane-bound galactosyltransferase of plant cell wall matrix polysaccharide biosynthesis. Plant J. 19, 691–697.

Enright, A.J., Dongen, S. Van, Ouzounis, C.A., 2002. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res. 30, 1575–1584.

Fatihi, A., Boulard, C., Bouyer, D., Baud, S., Dubreucq, B., Lepiniec, L., 2016. Deciphering and modifying LAFL transcriptional regulatory network in seed for improving yield and quality of storage compounds. Plant Sci. 250, 198–204.

Feng, W., Li, J., Long, S., Wei, S., 2019. A DREB1 gene from zoysiagrass enhances Arabidopsis tolerance to temperature stresses without growth inhibition. Plant Sci. 278, 20–31.

Gadeyne, A., Sánchez-Rodríguez, C., Vanneste, S., et al., 2014. The TPLATE adaptor complex drives clathrin-mediated endocytosis in plants. Cell 156, 691–704.

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. Plant J. 91, 1088–1107.

González-Morales, S.I., Chávez-Montes, R.A., Hayano-Kanashiro, C., Alejo-Jacuinde, G., Rico-Cambron, T.Y., de Folter, S., andHerrera-Estrella, L., 2016. Regulatory network analysis reveals novel regulators of seed desiccation tolerance in *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences USA 113, E5232–E5241.

Goubet, F., Barton, C.J., Mortimer, J.C., Yu, X., Zhang, Z., Miles, G.P., Richens, J., Liepman, A.H., Seffen, K., Dupree, P., 2009. Cell wall glucomannan in Arabidopsis is synthesised by CSLA glycosyltransferases, and influences the progression of embryogenesis. Plant J. 60, 527–538.

Grantham, N.J., Wurman-Rodrich, J., Terrett, O.M., et al., 2017. An even pattern of xylan substitution is critical for interaction with cellulose in plant cell walls. Nat. Plants 3, 859–865.

Grimault, A., Gendrot, G., Chaignon, S., Gilard, F., Tcherkez, G., Thévenin, J., Dubreucq, B., Depège-Fargeix, N., Rogowsky, P.M., 2015. Role of B3 domain transcription factors of the AFL family in maize kernel filling. Plant Sci. 236, 116–125.

Guerin, Chloé, Joët, T., Serret, J., et al., 2016. Gene coexpression network analysis of oil biosynthesis in an interspecific backcross of oil palm. Plant J. 87, 423–441.

Hématy, K., Sado, P.E., Tuinen, A. Van, Rochange, S., Desnos, T., Balzergue, S., Pelletier, S., Renou, J.P., Höfte, H., 2007. A receptor-like kinase mediates the response of arabidopsis cells to the inhibition of cellulose synthesis. Curr. Biol. 17, 922–931.

Huang, C., Zhang, R., Gui, J., Zhong, Y., Li, L., 2018. The receptor-like kinase AtVRLK1 regulates secondary cell wall thickening. Plant Physiol. 177, 671–683.

Hutin, S., Blanc-Mathieu, R., Rieu, P., Parcy, F., Lai, X., Zubieta, C., 2023. Identification of plant transcription factor DNA-Binding sites using seq-DAP-seq. In: Kaufmann, K., Vandepoele, K. (Eds.), Plant Gene Regulatory Networks, vol. 2698. Methods in Molecular Biology, Humana, New York, NY, pp. 119–145.

Jacobson, T., Edwards, M., Qiande, M., Robert, M., Moncrieff, J., Voiniciuc, C., 2025. Intracellular mannanases sustain matrix polysaccharide biosynthesis. bioRxiv 2025–02. https://doi.org/10.1101/2025.02.25.640093.

Jiang, Chao, Iu, B., Singh, J., 1996. Requirement of a CCGAC cis-acting element for cold induction of the BN115 gene from winter Brassica napus. Plant Mol. Biol. 30, 679–684.

Joersbo, M., Marcussen, J., Brunstedt, J., 2001. In vivo modification of the cell wall polysaccharide galactomannan of guar transformed with an alpha-galactosidase gene cloned from senna. Mol. Breed. 7, 211–219.

Joët, T., Laffargue, A., Salmona, J., Doulbeau, S., Descroix, F., Bertrand, B., Kochko, A. De, Dussert, S., 2009. Metabolic pathways in tropical dicotyledonous albuminous seeds: *Coffea arabica* as a case study. New Phytol. 182, 146–162.

Joët, T., Laffargue, A., Salmona, J., Doulbeau, S., Descroix, F., Bertrand, B., Lashermes, P., Dussert, S., 2014. Regulation of galactomannan biosynthesis in coffee seeds. J. Exp. Bot. 65, 323–337.

Kikuchi, A., Sato, E., Yoshimi, Y., et al., 2025. Atypical endo-β-1, 4-mannannases are necessary for normal glucomannan synthesis in Arabidopsis. bioRxiv 2025–02. https://doi.org/10.1101/2025.02.21.639209.

Kim, S.J., Chandrasekar, B., Rea, A.C., Danhof, L., Zemelis-Durfee, S., Thrower, N., Shepard, Z.S., Pauly, M., Brandizzi, F., Keegstra, K., 2020. The synthesis of xyloglucan, an abundant plant cell wall polysaccharide, requires CSLC function. Proc. Natl. Acad. Sci. USA 117, 20316–20324.

Kleczkowski, L.A., Igamberdiev, A.U., 2020. Optimization of nucleotide sugar supply for polysaccharide formation via thermodynamic buffering. Biochem. J. 477, 341–356.

Laffargue, A., de Kochko, A., Dussert, S., 2007. Development of solid-phase extraction and methylation procedures to analyse free fatty acids in lipid-rich seeds. Plant Physiol. Biochem. 45, 250–257.

Liepman, A.H., Wilkerson, C.G., Keegstra, K., 2005. Expression of cellulose synthase-like (Csl) genes in insect cells reveals that CslA family members encode mannan synthases. Proc. Natl. Acad. Sci. USA 102, 2221–2226.

Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/10.48550/arXiv.1303.3997.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The Sequence Alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Liu, Y., Wei, M., Hou, C., Lu, T., Liu, L., Wei, H., Cheng, Y., Wei, Z., 2017. Functional characterization of populus PsnSHN2 in coordinated regulation of secondary wall components in tobacco. Sci. Rep. 7, 42.

Li, Y., Pearl, S.A., Jackson, S.A., 2015. Gene networks in plant biology: approaches in reconstruction and analysis. Trends Plant Sci. 20, 664–675.

Li, Z., Zhang, C., Zhang, Y., Zeng, W., Igor, C., 2021. Coffee cell walls-composition, influence on cup quality and opportunities for coffee improvements. Food Quality and Safety 5, fyab012.

Lohse, M., Nagel, A., Herter, T., et al., 2014. Mercator: a fast and simple web server for genome scale functional annotation of plant sequence data. Plant Cell Environ. 37, 1250–1258.

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol. 15, 550.

Lunn, D., Gaddipati, S.R., Tucker, G.A., Lycett, G.W., 2013. Null mutants of individual RABA genes impact the proportion of different cell wall components in stem tissue of *Arabidopsis thaliana*. PLoS One 8, e75724.

Ma, W., Kong, Q., Arondel, V., Kilaru, A., Bates, P.D., Thrower, N.A., Benning, C., Ohlrogge, J.B., 2013. WRINKLED1, A ubiquitous regulator in oil accumulating tissues from arabidopsis embryos to oil palm mesocarp. PLoS One 8, e68887.

McFarlane, H.E., Döring, A., Persson, S., 2014. The cell biology of cellulose synthesis. Annu. Rev. Plant Biol. 65, 69–94.

Marchive, C., Nikovics, K., To, A., Lepiniec, L., Baud, S., 2014. Transcriptional regulation of fatty acid production in higher plants: molecular bases and biotechnological outcomes. Eur. J. Lipid Sci. Technol. 116, 1332–1343.

Miray, R., Kazaz, S., To, A., Baud, S., 2021. Molecular control of oil metabolism in the endosperm of seeds. Int. J. Mol. Sci. 22, 1621.

Mochida, K., Uehara-Yamaguchi, Y., Yoshida, T., Sakurai, T., Shinozaki, K., 2011. Global landscape of a Co-Expressed gene network in barley and its application to gene discovery in triticeae crops. Plant Cell Physiol. 52, 785–803.

Moreno-Risueno, M.Á., González, N., Díaz, I., Parcy, F., Carbonero, P., Vicente-Carbajosa, J., 2008. FUSCA3 from barley unveils a common transcriptional regulation of seed-specific genes between cereals and Arabidopsis. Plant J. 53, 882–894.

Persson, S., Wei, H., Milne, J., Page, G.P., Somerville, C.R., 2005. Identification of genes required for cellulose synthesis by regression analysis of public microarray data sets. Proc. Natl. Acad. Sci. U. S. A. 102, 8633–8638.

Persson, S., Paredez, A., Carroll, A., Palsdottir, H., Doblin, M., Poindexter, P., Khitrov, N., Auer, M., Somerville, C.R., 2007. Genetic evidence for three unique components in primary cell-wall cellulose synthase complexes in Arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 104, 15566–15571.

Prakash, R., Johnston, S.L., Boldingh, H.L., Redgwell, R.J., Atkinson, R.G., Melton, L.D., Brummell, D.A., Schröder, R., 2012. Mannans in tomato fruit are not depolymerized during ripening despite the presence of endo-β-mannanase. J. Plant Physiol. 169, 1125–1133.

Quan, M., Liu, X., Du, Q., Xiao, L., Lu, W., Fang, Y., Li, P., Ji, L., Zhang, D., 2021. Genome-wide association studies reveal the coordinated regulatory networks underlying photosynthesis and wood formation in Populus. J. Exp. Bot. 72, 5372–5389.

Rautengarten, C., Ebert, B., Moreno, I., et al., 2014. The Golgi localized bifunctional UDP-rhamnose/UDP-galactose transporter family of arabidopsis. Proc. Natl. Acad. Sci. U. S. A. 111, 11563–11568.

Redgwell, R.J., Curti, D., Rogers, J., Nicolas, P., Fischer, M., 2003. Changes to the galactose/mannose ratio in galactomannans during coffee bean (*Coffea arabica* L.) development: implications for in vivo modification of galactomannan synthesis. Planta 217, 316–326.

Redgwell, R., Fischer, M., 2006. Coffee carbohydrates. Braz. J. Plant Physiol. 18, 165–174.

Ribas, A.F., Dechamp, E., Champion, A., Bertrand, B., Combes, M.C., Verdeil, J.L., Lapeyre, F., Lashermes, P., Etienne, H., 2011. Agrobacterium-mediated genetic transformation of *Coffea arabica* (L.) is greatly enhanced by using established embryogenic callus cultures. BMC Plant Biol. 11, 92.

Righetti, K., Vu, J.L., Pelletier, S., et al., 2015. Inference of longevity-related genes from a robust coexpression network of seed maturation identifies regulators linking seed storability to biotic defense-related pathways. Plant Cell 27, 2692–2708.

Rösti, J., Barton, C.J., Albrecht, S., Dupree, P., Pauly, M., Findlay, K., Roberts, K., Seifert, G.J., 2007. UDP-glucose 4-epimerase isoforms UGE2 and UGE4 cooperate in providing UDP-galactose for cell wall biosynthesis and growth of *Arabidopsis thaliana*. Plant Cell 19, 1565–1579.

Sánchez-Rodríguez, C., Bauer, S., Hématy, K., et al., 2012. CHITINASE-LIKE1/POM-POM1 and its homolog CTL2 are glucan-interacting proteins important for cellulose biosynthesis in Arabidopsis. Plant Cell 24, 589–607.

Sánchez-Rodríguez, C., Shi, Y., Kesten, C., et al., 2018. The cellulose synthases are cargo of the TPLATE adaptor complex. Mol. Plant 11, 346–349.

Santana-Garcia, W., Castro-Mondragon, J.A., Padilla-Gálvez, M., et al., 2022. Rsat 2022: regulatory sequence analysis tools. Nucleic Acids Res. 50, W670–W676.

Sawake, S., Tajima, N., Mortimer, J.C., et al., 2015. KONJAC1 and 2 are key factors for GDP-Mannose generation and affect l-Ascorbic acid and glucomannan biosynthesis in Arabidopsis. Plant Cell 27, 3397–3409.

Scheller, H.V., Ulvskov, P., 2010. Hemicelluloses. Annu. Rev. Plant Biol. 61, 263–289.

Schwacke, R., Ponce-Soto, G.Y., Krause, K., et al., 2019. MapMan4: a refined protein classification and annotation framework applicable to multi-omics data analysis. Mol. Plant 12, 879–892.

Seifert, G.J., Barber, C., Wells, B., Dolan, L., Roberts, K., 2002. Galactose biosynthesis in Arabidopsis: genetic evidence for substrate channeling from UDP-D-galactose into cell wall polymers. Curr. Biol. 12, 1840–1845.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., Ideker, T., 2003. Cytoscape: a software Environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504.

Sharma, P., Sharma, S., Ramakrishna, G., Srivastava, H., Gaikwad, K., 2022. A comprehensive review on leguminous galactomannans: structural analysis, functional properties, biosynthesis process and industrial applications. Crit. Rev. Food Sci. Nutr. 62, 443–465.

Singh, S., Singh, G., Arya, S.K., 2018. Mannans: an overview of properties and application in food products. Int. J. Biol. Macromol. 119, 79–95.

Song, T., Huo, Q., Li, C., Wang, Q., Cheng, L., Qi, W., Ma, Z., Song, R., 2024. The biosynthesis of storage reserves and auxin is coordinated by a hierarchical regulatory network in maize endosperm. New Phytol. 243, 1855–1869.

Sun, A., Yu, B., Zhang, Q., et al., 2020. MYC2-activated trichome Birefringence-like37 acetylates cell walls and enhances herbivore resistance. Plant Physiol. 184, 1083–1096.

Sutherland, P.W., Hallett, I.C., MacRae, E., Fischer, M., Redgwell, R.J., 2004. Cytochemistry and immunolocalisation of polysaccharides and proteoglycans in the endosperm of green Arabica coffee beans. Protoplasma 223, 203–211.

Toyooka, K., Goto, Y., Asatsuma, S., Koizumi, M., Mitsui, T., Matsuoka, K., 2009. A mobile secretory vesicle cluster involved in mass transport from the golgi to the plant cell exterior. Plant Cell 21, 1212–1229.

Usadel, B., Obayashi, T., Mutwil, M., et al., 2009. Co-expression tools for plant biology: opportunities for hypothesis generation and caveats. Plant Cell Environ. 32, 1633–1651.

Vellosillo, T., Dinneny, J.R., Somerville, C.R., Ehrhardt, D.W., 2021. TRANVIA (TVA) facilitates cellulose synthase trafficking and delivery to the plasma membrane. Proc. Natl. Acad. Sci. U. S. A. 118, e2021790118.

Voiniciuc, C., Schmidt, M.H., Berger, A., Yang, B., Ebert, B., Scheller, H.V., North, H.M., Usadel, B., Günl, M., 2015. MUCILAGE-RELATED10 produces galactoglucomannan that maintains pectin and cellulose architecture in Arabidopsis seed mucilage. Plant Physiol. 169, 403–420.

Voiniciuc, C., Dama, M., Gawenda, N., Stritt, F., Pauly, M., 2019. Mechanistic insights from plant heteromannan synthesis in yeast. Proc. Natl. Acad. Sci. U. S. A. 116, 522–527.

Voiniciuc, C., 2022. Modern mannan: a hemicellulose's journey. New Phytol. 234, 1175–1184.

Wang, F., Perry, S.E., 2013. Identification of direct targets of FUSCA3, a key regulator of Arabidopsis seed development. Plant Physiol. 161, 1251–1264.

Wang, Y., Mortimer, J.C., Davis, J., Dupree, P., Keegstra, K., 2013. Identification of an additional protein involved in mannan biosynthesis. Plant J. 73, 105–117.

Wang, Y., Tang, R.J., Yang, X., Zheng, X., Shao, Q., Tang, Q.L., Fu, A., Luan, S., 2019. Golgi-localized cation/proton exchangers regulate ionic homeostasis and skotomorphogenesis in Arabidopsis. Plant Cell Environ. 42, 673–687.

Yamamoto, A., Kagaya, Y., Usui, H., Hobo, T., Takeda, S., Hattori, T., 2010. Diverse roles and mechanisms of gene regulation by the arabidopsis seed maturation master regulator FUS3 revealed by microarray analysis. Plant Cell Physiol. 51, 2031–2046.

Yang, W., Xin, Z., Zhang, Q., Zhang, Y., Niu, L., 2024. The tree peony DREB transcription factor PrDREB2D regulates seed α-linolenic acid accumulation. Plant Physiol. 195, 745–761.

Yang, Z., Liu, X., Li, N., Du, C., Wang, K., Zhao, C., Wang, Z., Hu, Y., Zhang, M., 2019. WRINKLED1 homologs highly and functionally express in oil-rich endosperms of oat and Castor. Plant Sci. 287, 110193.

Yoshimi, Y., Yu, L., Cresswell, R., Guo, X., Echevarría-Poza, A., Lyczakowski, J.J., Dupree, R., Kotake, T., Dupree, P., 2025. Glucomannan engineering highlights roles of galactosyl modification in fine-tuning cellulose-glucomannan interaction in Arabidopsis cell walls. Nat. Commun. 16, 1235.

Yuan, Y., Sun, H., Wang, X., Gao, L., Liu, C., Gai, S., Zhang, Y., 2022. Transcriptome of *Paeonia ostii* developing seeds revealed transcriptional factors, mRNAs and LncRNAs related to fatty acid synthesis and accumulation. Ind. Crop. Prod. 188, 115670.

Zabotina, O.A., Zang, N., Weerts, R., 2021. Polysaccharide biosynthesis: glycosyltransferases and their complexes. Front. Plant Sci. 12, 625307.

Zhang, B., Zhang, L., Li, F., Zhang, D., Liu, X., Wang, H., Xu, Z., Chu, C., Zhou, Y., 2017. Control of secondary cell wall patterning involves xylan deacetylation by a GDSL esterase. Nat. Plants 3, 17017.

Zhang, L., Gao, C., Mentink-Vigier, F., et al., 2019. Arabinosyl deacetylase modulates the arabinoxylan acetylation profile and secondary wall formation. Plant Cell 31, 1113–1126.

Zhang, Y., Liu, T., Meyer, C.A., et al., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.

Zhang, Y., Nikolovski, N., Sorieul, M., et al., 2016b. Golgi-localized STELLO proteins regulate the assembly and trafficking of cellulose synthase complexes in Arabidopsis. Nat. Commun. 7, 11656.

Zhang, Z., Zheng, X., Yang, J., Messing, J., Wu, Y., 2016a. Maize endosperm-specific transcription factors O2 and PBF network the regulation of protein and starch synthesis. Proc. Natl. Acad. Sci. USA 113, 10842–10857.

Zhong, R., Richardson, E.A., Ye, Z.-H., 2007. The MYB46 transcription factor is a direct target of SND1 and regulates secondary Wall biosynthesis in arabidopsis. Plant Cell 19, 2776–2792.

Zhong, R., Ye, Z.H., 2015. Secondary cell walls: biosynthesis, patterned deposition and transcriptional regulation. Plant Cell Physiol. 56, 195–214.

Zhong, R., Cui, D., Ye, Z.H., 2018. Members of the DUF231 family are O-Acetyltransferases catalyzing 2-O- and 3-O-Acetylation of Mannan. Plant Cell Physiol. 59, 2339–2349.