



# Transcriptome dynamics of cork oak (*Quercus suber*) somatic embryogenesis reveals active gene players in transcription regulation and phytohormone homeostasis of embryo development

Tiago Capote<sup>1,2</sup> · Anabel Usié<sup>1,3</sup> · Pedro Barbosa<sup>1</sup> · Marcos Ramos<sup>1,3</sup> · Leonor Morais-Cecílio<sup>2</sup> · Sónia Gonçalves<sup>1,4</sup> 

Received: 7 August 2017 / Revised: 20 March 2019 / Accepted: 16 May 2019 / Published online: 11 June 2019

© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

Cork oak (*Quercus suber* L.) is one of the most important Mediterranean forest tree species. The last decades have been marked by a decline in this species. Implementation of breeding programs is fundamental to revert this trend. Somatic embryogenesis is the system of choice for clonal propagation, constituting a valuable tool for embryo production and improved genotype testing. In this study, the cork oak transcriptome during somatic embryogenesis was characterized in four stages of development to identify relevant genes in the process and to understand the molecular and biochemical events occurring in each specific stage. A total 66,693 candidate coding regions were predicted from the generated de novo transcriptome assembly. Differential gene expression analysis identified 11,507 genes distributed in 30 clusters with distinct gene expression patterns and enriched in various biological process GO terms. Results show 1159 differentially expressed genes coding for transcription regulators, namely transcription factors (76%) with important roles in embryogenesis, like orthologous of AINTEGUMENTA-like, PLETHORA, CYTOKININ RESPONSE FACTOR, GATA transcription factors, and AUXIN RESPONSE FACTORS genes. Results also show 250 differentially expressed phytohormone-related genes involved in important aspects of embryogenesis as tissue specification, differentiation, and embryogenesis competence. Finally, we identified a group of genes with functions in cellular protection and abiotic stress tolerance coding for LATE EMBRYOGENESIS ABUNDANT proteins. Cork oak embryogenesis transcriptome characterization represents a tool for future biotechnological applications. Our results provide a molecular insight into embryo development, establishing a basis for further research towards improvement of somatic embryogenesis in cork oak.

**Keywords** *Quercus suber* · Somatic embryogenesis · Transcriptomics · Transcription factor · Hormone · Plant biotechnology

Communicated by J. Wright

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11295-019-1353-6>) contains supplementary material, which is available to authorized users.

✉ Sónia Gonçalves  
sonia.goncalves@cebal.pt

<sup>1</sup> Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo (CEBAL), Instituto Politécnico de Beja (IPBeja), Rua Pedro Soares, 7801-908 Beja, Portugal

<sup>2</sup> Linking Landscape, Environment, Agriculture and Food, Instituto Superior de Agronomia, University of Lisbon, Tapada da Ajuda 1349-017, Lisbon, Portugal

<sup>3</sup> Instituto de Ciências Agrárias e Ambientais Mediterrânicas (ICAAM), Universidade de Évora, Évora, Portugal

<sup>4</sup> Present address: Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB101SA, UK

## Introduction

Cork oak (*Quercus suber* L.) is one of the most important Mediterranean forest tree species. Native to dry and semi-arid regions, it plays an important ecological role as environmental protector since it is one of the critical species of the *montado* ecosystem contributing to soil conservation as well as flora and fauna biodiversity. Socially, cork oak stands to underpin the basis of a sustainable activity, providing a source of income among rural populations across the Mediterranean basin due to the economic value of cork.

In the last two decades, climate change, exotic pathogens, and pest translocations from ecosystems separated in the past have caused extensive decline and mortality of several tree species worldwide (Allen et al. 2010). One of the main tree genera affected by this event is *Quercus* (Braisier 1996). In the

Iberian Peninsula, a severe decline affecting European oak forests has been well reported since the 1980s (Brasier 1992; Brasier 1996), in which cork oak is largely representative. The use of natural regeneration seeds for propagation does not allow the rescue of valuable phenotypes (Valladares et al. 2004), only identified in adult trees. Since the capacity for vegetative propagation of *Quercus* genera through traditional methods is very low (Vieitez et al. 2012), somatic embryogenesis emerges as a valuable tool for integration in breeding programs. Induction of *Q. suber* somatic embryogenesis has been successfully achieved by distinct research groups (Bueno et al. 2000; Pinto et al. 2002; Pintos et al. 2008; Vieitez et al. 2012). The possibility to modulate somatic embryogenesis and embryo development in cork oak allows the storage of cork oak embryos for several years, which is difficult owing to recalcitrance of the acorns. In addition, it opens the possibility to perform functional genomic studies such as genetic transformation for herbicide resistance (Álvarez et al. 2009), gene silencing (Zhang et al. 2012), and overexpression (Mallón et al. 2014) or even precise genome editing by new emerged techniques like Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/CRISPR-associated protein 9 (Cas9) system (Fan et al. 2015).

Somatic embryos follow similar developmental patterns as their zygotic counterparts. Embryo development and maturation requires the concerted action of several signaling pathways integrating genetic, epigenetic, and hormonal regulation (Gutierrez et al. 2007; Feng et al. 2010; Gao et al. 2012). Phytohormonal homeostasis modulates embryo development involving several types of hormones including abscisic acid (ABA), auxin (AUX), brassinosteroid (BR), cytokinin (CK), ethylene (ET), gibberellin (GA), jasmonic acid (JA), and salicylic acid (SA) modulating embryo development.

Embryo development proceeds through a series of spatially and temporally regulated gene expression networks from which many genes have already been characterized (Gao et al. 2012). In woody species, recent advances provided insight into mechanisms involved in transition from morphologically mature to physiologically mature somatic embryo namely in *Picea asperata* (Jing et al. 2017). In *Citrus sinensis*, small RNAs and degradome high-throughput sequencing revealed a group of miRNAs/siRNA as well as their targets differentially regulating specific biological process and transcription factor expression (Wu et al. 2015).

An effort to explain key regulatory processes involved in distinct developmental stages of *Q. suber* somatic embryogenesis has been made by a proteomic approach (Gomez-Garay et al. 2013). However, precise molecular mechanisms involved in somatic embryo induction, maturation, and germination in this species are still poorly known.

An assessment of the transcriptome of embryo development will contribute to the comprehension of genes involved in this process and may help to improve the process of in vitro

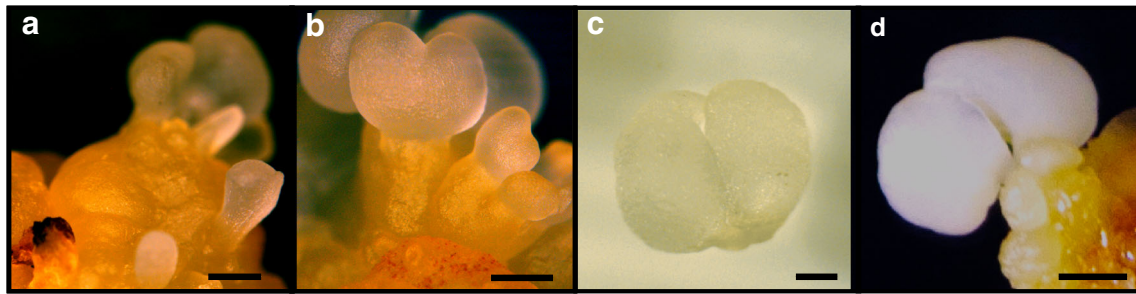
embryo production. Identification of gene expression patterns associated with specific stages of embryo development is critical to understand the molecular and biochemical basis characteristic of each specific stage of embryo development. Knowledge generated by such studies will contribute to and enhance the quality of mature embryos for use in forestation processes.

Next-generation sequencing applied to transcriptomics enables the unraveling of developmental transcriptome dynamics. RNA-Seq and the quantification of differentially expressed genes not only describes genes being expressed in a particular stage of development but also quantifies the expression levels of each transcript. In this context, we have analyzed the transcriptome of *Q. suber* somatic embryo in four distinct embryo developmental stages, from globular to mature cotyledonary. With this approach, we expected to identify potential important genes in *Q. suber* somatic embryo development and maturation, focusing specifically on transcripts coding for transcription factors, transcription regulators, and chromatin regulators, as well as transcripts related to hormone biosynthesis, metabolism, transport, and signaling.

## Materials and methods

### Sample preparation and RNA sequencing

Branches up to 5 cm in diameter were collected from one adult cork oak tree growing in Cercal do Alentejo, Portugal, in April 2011. Segmented branches without lateral branches and leaves, 20 cm in length and between 2 and 4 cm in diameter, were forced to sprout in a fitoclima S600PLH (Aralab) at  $25 \pm 1$  °C, with a 16-h photoperiod provided by cool-white fluorescent tubes ( $200 \mu\text{mol m}^{-2} \text{s}^{-1}$ ) and 90–100% relative humidity. Expanding leaves with 1–2 cm from the base to the apex were excised from epicormic shoots and used as initial explants for somatic embryogenesis induction, according to Álvarez et al. (2004). *Quercus suber* embryogenic clusters were grown in MSSH medium, containing MS micronutrients (Murashige and Skoog 1962) and SH macronutrients (Schenk and Hildebrandt 1972) and supplemented with MS vitamins (Murashige and Skoog 1962), on a climatic chamber at 25 °C under a light/dark cycle of 16/8 h. Somatic embryogenic clusters were subcultured every 30 days. After 4 years in culture, embryos were selected and pooled according to developmental stages. We selected four distinct *Q. suber* embryo developmental stages (Fig. 1a): (1) the globular stage (ST1), in which embryos start to emerge from the pro embryogenic masses acquiring a radial symmetrical morphology; (2) the heart/torpedo stage (ST2), where emergence of the first discernible organs occurs, the two cotyledons, and a change in embryo morphology from radial to bilateral symmetry is visible; (3)



**Fig. 1** Morphology of cork oak somatic embryo in the different developmental stages used for RNA-Seq analysis. Somatic embryo at the globular (ST1) (a), heart/torpedo-shaped (ST2) (b), early cotyledonary (ST3) (c), and mature cotyledonary (ST4) (d) stages. Bars = 0.5 mm in (a, b, c), 0.5 cm in (d)

the immature cotyledonary stage (ST3) where the cotyledons are well defined and most embryogenic tissue and organs are formed; (4) and the mature cotyledonary stage (ST4), in which embryos pass through several morphological and biochemical changes, namely the expanding of the storage organs, the cotyledons, repression of germination, and acquisition of desiccation tolerance (Von Arnold et al. 2002).

Embryos from each distinct stage (20–40 fresh embryos) were pooled together and ground with a mortar in the presence of liquid nitrogen for RNA extraction. Total RNA was extracted from the four embryo stages described above using the AMBION RNAqueous® Total RNA Isolation kit (Thermo Fisher Scientific, Waltham, MA, USA) including Plant RNA Isolation Aid solution (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's instructions. RNA samples were treated with AMBION TURBO DNA-free™ Kit (Thermo Fisher Scientific, Waltham, MA, USA) for DNA removal. RNA samples were analyzed by electrophoretic separation of nucleic acids in a Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) and used for cDNA synthesis and library construction for sequencing when RIN > 8.

RNA-seq library preparation and sequencing was performed by the Beijing Genomics Institute BGI (Hong-Kong, China). Paired-end libraries were prepared following the protocol of Illumina® TruSeq RNA Library Preparation kit and sequenced in one lane using the Illumina HiSeq™ 2000 platform, with a read length of 100 bp. Raw reads were cleaned by removing adaptor sequences, empty reads, and low-quality sequences.

### Transcriptome data assembly and annotation

Illumina paired-end reads were pre-processed according to quality (minimum of 20) and length (minimum of 80) using a custom Perl script combined with Sickle (Joshi and Fass 2011) in order to remove or trim from the dataset reads with low average quality and reads containing undetermined nucleotides (N's). Pre-processed reads were assembled with the Trinity platform (Grabherr et al. 2011). Assembly statistics

were generated by quast 2.3 (Gurevich et al. 2013). Protein sequence prediction was performed with Transdecoder (<http://transdecoder.github.io>) from the transcriptome assembly, which extracts the longest open reading frames (ORFs) and identifies ORFs with homology to known proteins using Blastp and hmmer against Uniprot and Pfam databases producing a set of candidate coding regions.

Pre-processed paired-end reads were aligned to the assembled contigs using BWA (Li and Durbin 2010). Mapped reads with a mapping quality (MAPQ) value  $\geq 10$ , a single best hit (X0:i:1), and no suboptimal or alternative hits (X1:i:0) were kept, selected with Samtools (Li et al. 2009), and considered as unique mapped reads. Digital differential expression analysis was performed using edgeR (Robinson et al. 2009), examining differential expression (DE) from a table of read counts generated by unique mapped reads through the featureCounts R function. A common dispersion value was used considering the biological coefficient of variation (BCV) to be 0.1, which is a typical value reported for RNA-Seq next-generation sequencing data studies involving genetically identical organisms arising from controlled experiments (McCarthy et al. 2012). Genes with very low counts across the dataset of predicted genes were discarded to avoid interferences with statistical analysis. After DE analysis, correction for multiple testing was performed by applying a cutoff  $\leq 0.01$  FDR (false discovery rate), to produce a final list of differentially expressed genes (DEGs) and digital gene expression values in counts per million (CPM).

The transcriptome assembly and the whole set of DEGs were functionally annotated using Blastx against the non-redundant (nr) protein plant sequence database from NCBI. Results were loaded into Blast2GO (Conesa and Gotz 2008) to perform a protein domain and gene ontology (GO) annotation using the InterProScan (Zdobnov and Apweiler 2001; Jones et al. 2014) function. ANNEX annotation (annotation argumentation) function was used to refine annotations (Conesa and Gotz 2008) and GOSlim (plants) was used to generate plant specific GO terms. Results were graphically represented with Cytoscape (Shannon et al. 2003).

Overrepresented GO terms of differentially expressed genes were identified with BiNGO (Maere et al. 2005) (Cytoscape plugin) using as a reference a custom GO annotation of the assembled sequences generated and annotated with Blast2GO. From those results, an enrichment analysis was performed. GO terms with  $p$  values  $< 0.05$  were considered to be significantly different and enriched in our sets of DEGs.

### Gene clustering analysis and differential expression profiling

Hierarchical clustering was performed using MultiExperiment Viewer (MeV v4.9.0) (Saeed et al. 2006) with the Euclidean distance and average linkage model. K-means clustering was done with the Euclidean distance to produce distinct expression profile clusters. A  $\log_2$  transformation was applied to all expression values before the clustering analysis.

### Differential gene expression validation by RT-qPCR

A set of 19 DEGs was analyzed by reverse transcription quantitative real-time PCR (RT-qPCR) in order to validate expression profiles obtained in bioinformatics analyses (Online Resource 1). Two micrograms of total RNA previously extracted for RNA-Seq were used for cDNA synthesis using the QuantiTect Reverse Transcription kit (Qiagen, Valencia, CA, USA), which includes an additional genomic DNA elimination step and uses a mix of oligo (dT) and random hexamer primers. Specific primers were designed by Primer3Plus (Untergasser et al. 2007). RT-qPCR experiments were carried out in an iCycler iQ5 Instrument (Bio-Rad Laboratories, Hercules, CA, USA) using the Sso Advanced Universal SYBR Green Master mix (Bio-Rad Laboratories, Hercules, CA, USA) in 96-well plates. Three replicates were performed in reaction mixtures of 20  $\mu\text{L}$  containing 10  $\mu\text{L}$  of Master mix, 400 nM of each specific primer pair (forward/reverse) and 1  $\mu\text{L}$  of cDNA with a dilution of 1:100 as template. All selected genes were amplified with the following PCR program with a single fluorescent reading taken at the end of each cycle: 95  $^{\circ}\text{C}$  at 10 min, 45 cycles of 10 s at 95  $^{\circ}\text{C}$ , 15 s at 60  $^{\circ}\text{C}$ , 20 s at 60  $^{\circ}\text{C}$ , and 15 s at 72  $^{\circ}\text{C}$ , except for *AINTEGUMENTA (ANT)*, *NUCLEAR CAP-BINDING PROTEIN SUBUNIT 1 (ABH1)*, and *TRANSLATIONALLY CONTROLLED TUMOR PROTEIN 1 (TCPCI)* for which the annealing temperature was 61  $^{\circ}\text{C}$ . To distinguish specific from nonspecific products and primer dimers, a melting curve was obtained immediately after amplification. Normalization was performed using two reference genes: *ACTIN (ACT)* and *CLATHRIN ADAPTOR COMPLEXES (CACs)*. Normalized relative quantities (NRQ)

were calculated by  $NRQ = \frac{E_{goi}^{\Delta Ct, goi}}{\sqrt[f]{\prod_o E_{ref_o}^{\Delta Ct, ref_o}}}$  where  $E$  is the amplification efficiency for each primer pair,  $f$  the number of

reference genes used to normalize data,  $goi$  the gene of interest,  $ref$  the reference gene and  $\Delta Ct$  is the  $Ct$  of the sample with higher  $Ct$  across samples minus the  $Ct$  value of the sample in test (Hellemans et al. 2007). Expression values from RNA-Seq experiments were calculated by the edgeR package. Both RNA-Seq and RT-qPCR gene expression values were submitted to  $\log_2$  transformation and compared by Pearson correlation.

### Gene annotation

To identify genes encoding for transcription factors (TFs), transcription regulators (TRs), and chromatin regulators (CRs), all DEGs were analyzed with the PlantTFcat tool (Dai et al. 2013). For analysis and classification of homeostasis-related phytohormones and genes involved in germination, a Blastx against the Arabidopsis Hormone Database (AHD2.0, <http://ahd.cbi.pku.edu.cn/>) (Jiang et al. 2011) was performed with all DEGs ( $E$  value  $\leq 10^{-10}$ , % identity  $> 60$ , query coverage  $> 70$ ) producing a list of genes. Other embryogenesis-related genes were manually identified searching for genes annotated with “embryo” and “embryogenesis” terms.

### Data availability

Raw reads of transcriptomic data have been deposited to the NCBI Short Read Archive (SRA) under accessions numbers SAMN05898489, SAMN05898490, SAMN05898491, and SAMN05898492. The four records correspond to *Quercus suber* somatic embryo development Biosamples stage 1 (SAMN05898489), stage 2 (SAMN05898490), stage 3 (SAMN05898491), and stage 4 (SAMN05898492).

## Results

### RNA-Seq analysis and global functional annotation

A de novo transcriptome sequence assembly was generated from a genotype embryo cell line to study *Q. suber* somatic embryo development and maturation. Sequencing of cDNA libraries of four different embryo developmental stages generated 160,366,471 paired-end reads, of which 156,088,387 were kept after pre-processing. The transcriptome assembly resulted in 143,960 contigs ranging from 500 to 16,793 bp and a N50 length of 1734 bases (Table 1). A total of 66,693 candidate coding regions within the assembled transcriptome were identified, from which 22,109 were co-expressed among the four embryo developmental stages (Fig. 2a). In addition, 671, 811, 1014, and 1480 genes were found exclusively present in ST1, ST2, ST3, and ST4, respectively. Also, there were 373 genes uniquely present in ST1 and ST2, 786 genes unique

**Table 1** Statistics of *Quercus suber* somatic embryo assembled transcriptome

Nucleotide length	Contigs
0–500 bp	37,514
500–1000 bp	92,395
> 1000 bp	51,565
Total	181,474
Largest contig	16,793
N50	1734
N75	1071
L50	24,276
L75	48,030
GC (%)	40.6

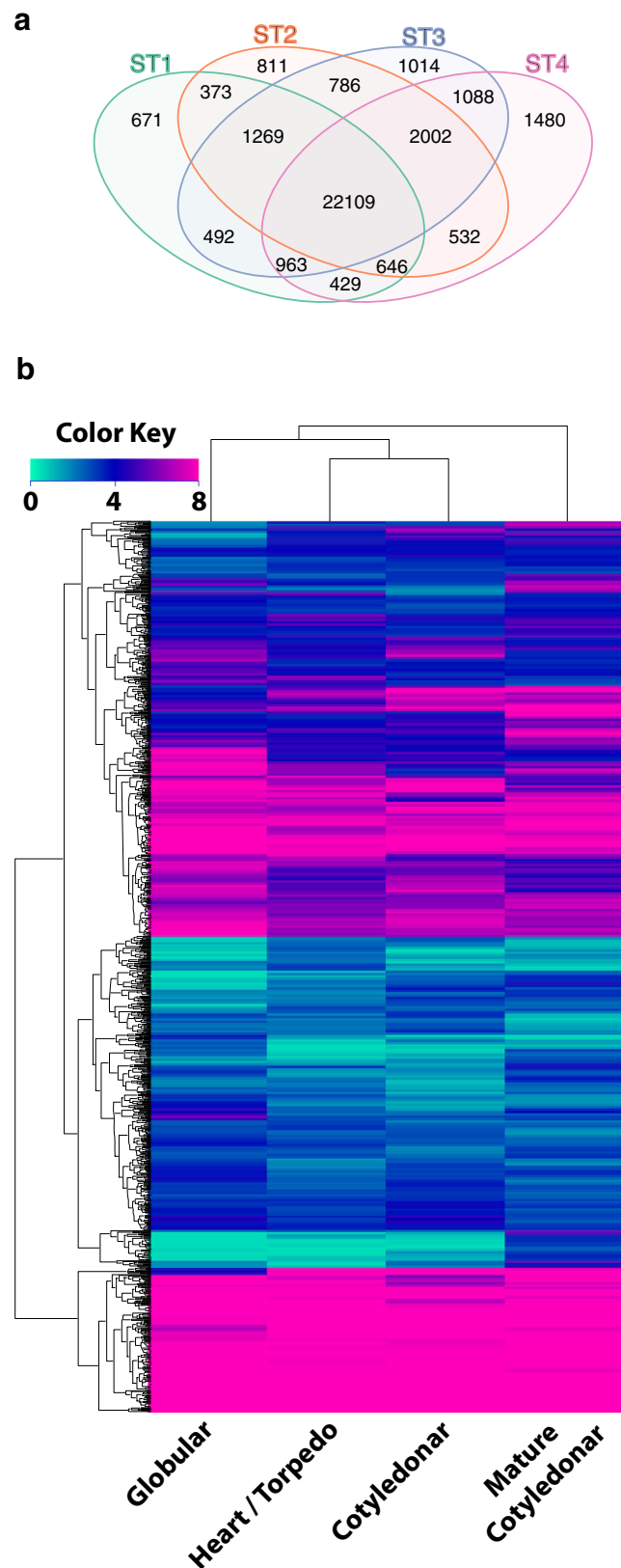
to ST2 and ST3 stages and 1088 genes that were expressed only in ST3 and ST4.

After normalization and  $\log_2$  transformation, a heatmap was generated to show a global overview of gene expression, during somatic embryo development (Fig. 2b). Hierarchical clustering showed that ST2 and ST3 grouped closer to ST1 while ST4 clustered in a different group (Fig. 2b). Moreover, multiple gene groups were clustered together showing similar expression patterns across developmental stages.

Functional categories were assigned to all genes in terms of GO (Fig. 3). A total of 29,725 genes (44.57%) were assigned to one or more plant GOSlim categories. From those, 18,360 (61.77%) were involved in biological process, 24,712 (83.14%) with molecular functions, and 14,088 (47.40%) with cellular components. Also, a total of 8558 (28.80%) genes were annotated to three main GO categories. Regarding biological processes, the main represented GO categories at level 2 are cellular process (31.73%) and metabolic process (28.28%). Other well-represented biological processes such as single-organism process (8.77%) and response to stimulus (5.68%) were also identified (Fig. 3). In terms of molecular function categories, GO terms at level 3 associated with the annotated genes reveal that heterocyclic compound binding (23.08%), organic cyclic compound binding (23.08%), transferase activity (17.11%), small molecule binding (16.34%), and hydrolase activity (13.46%) are the most represented GO terms (Fig. 3). Cellular component-associated GOs at level 4 reveal that most of the genes belong to intracellular membrane-bounded organelle (28.75%), cytoplasm (28.28%), and cytoplasmic part (23.73%) components (Fig. 3).

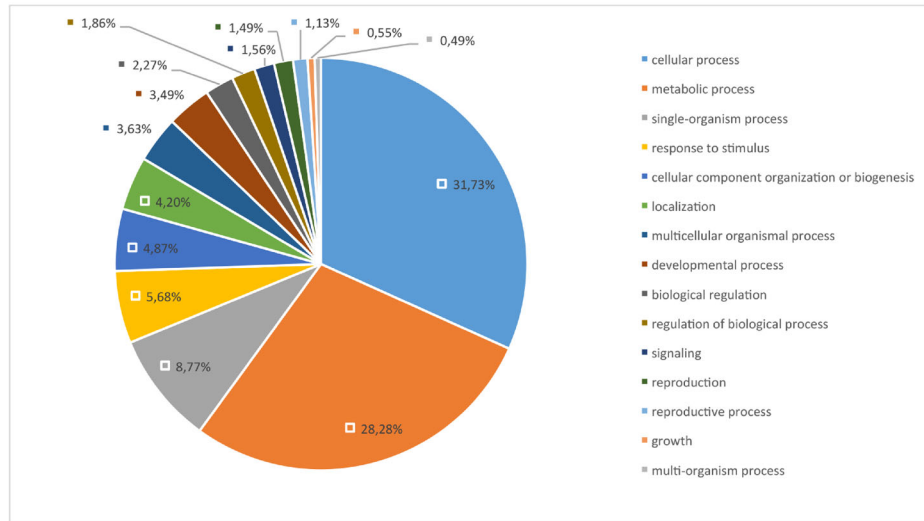
### Gene expression and differential gene expression analysis

Unique mapped reads of each cDNA library were used for gene expression analysis by the featureCounts R function. The analysis identified 11,507 DEGs (FDR < 0.01) between all developmental stages. The genes were differentially

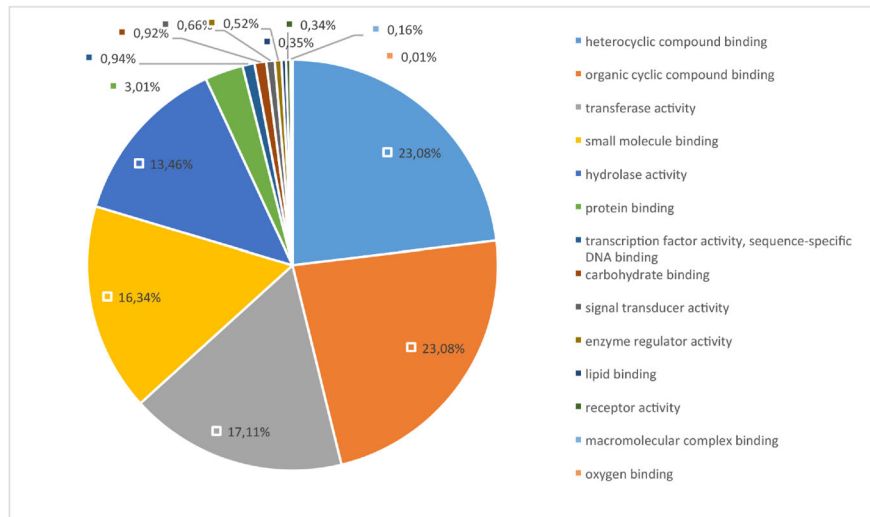


**Fig. 2** Venn diagram of all expressed genes between the four developmental stages (top). Hierarchical clustering of DE genes among libraries (bottom). Green indicates downregulated and red indicates upregulated genes

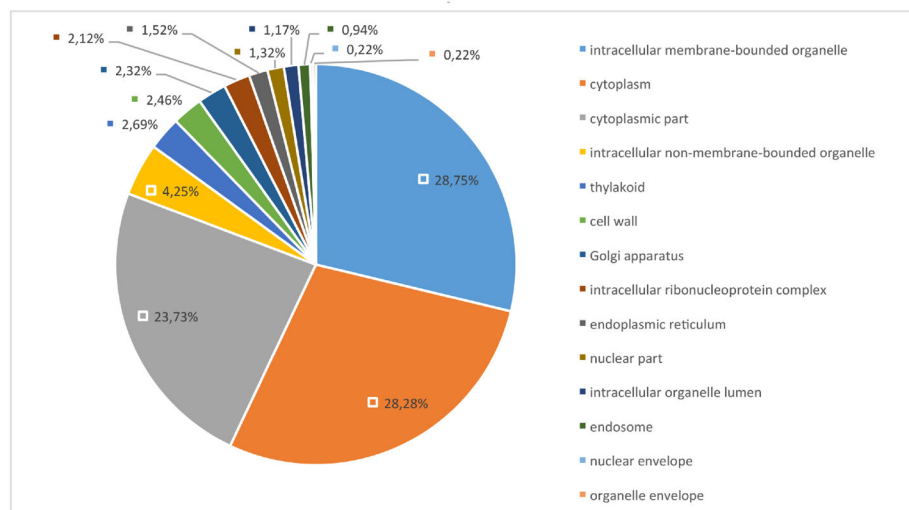
### Biological Process (Level 2)



### Molecular Function (Level 3)



### Cellular Component (Level 4)



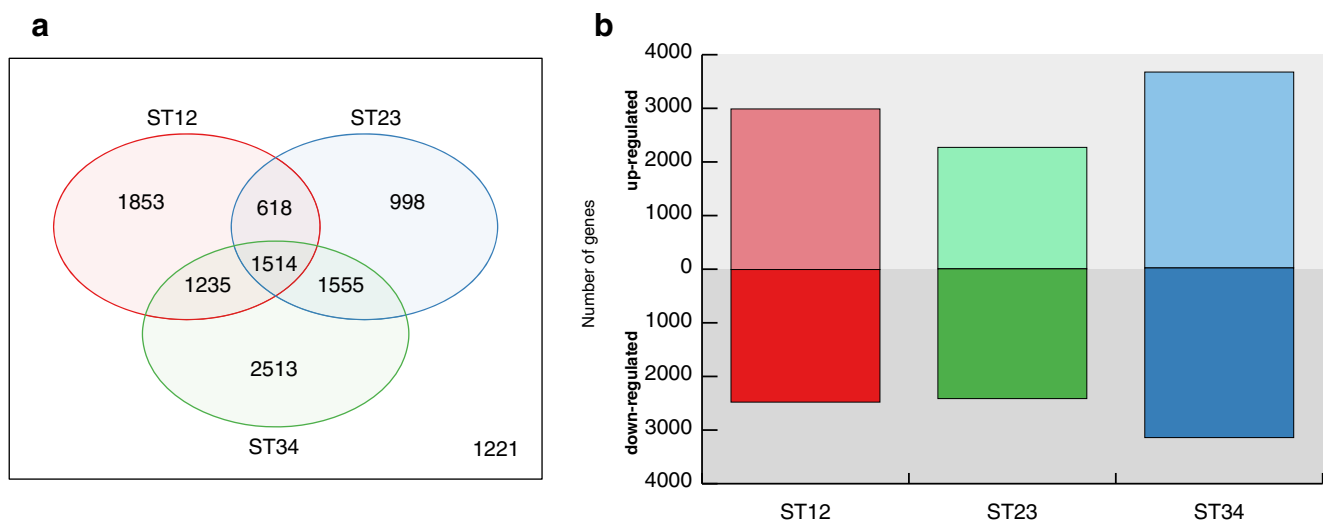
**Fig. 3** Gene Ontology functional classification of all expressed genes on biological process at level 2, molecular function at level 3 and cellular localization at level 4

expressed when compared: ST1 vs ST2 (ST12), ST2 vs ST3 (ST23), ST3 vs ST4 (ST34), and non-sequential developmental stages (Fig. 4a). Smear plots of log (fold change) vs average log (count per million) and respective volcano plots show the distribution of DEGs among sequential developmental stages (Online Resource\_2). In ST12, the analysis identified 2995 upregulated and 2473 downregulated genes (Fig. 4b). In ST23, 2265 and 2420 up- and downregulated genes were found, respectively (Fig. 4b). Finally, in ST34, the total number of DEGs is higher when compared with that of ST12 and ST23, accounting 3653 genes upregulated and 3164 downregulated (Fig. 4b).

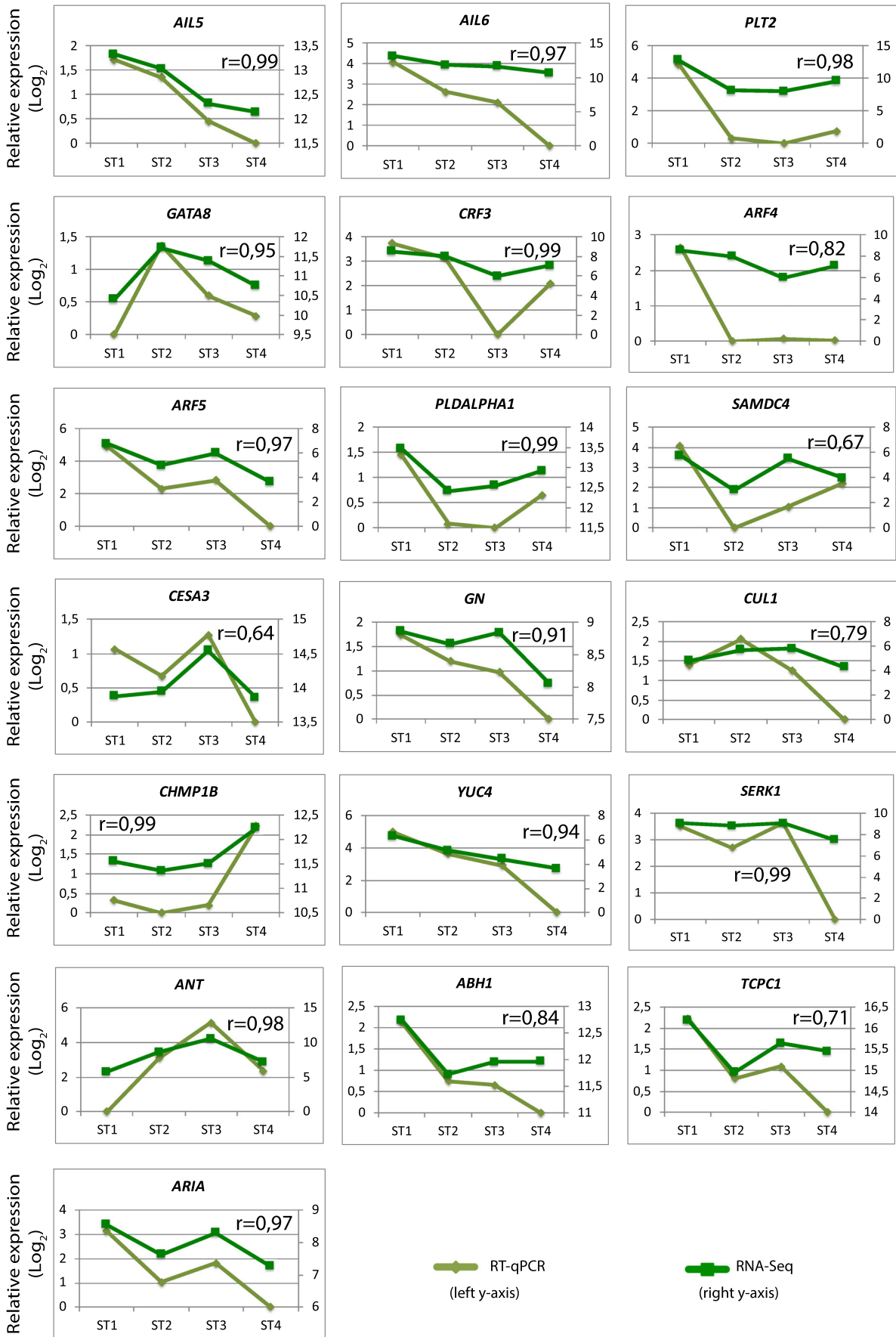
In order to understand patterns of gene expression during somatic embryo development, a hierarchical clustering analysis with the K-means method using Euclidean distance after  $\log_2$  expression values transformation was performed. K-means clustering revealed 30 clusters with different expression patterns containing between 121 and 743 significantly regulated genes for the four developmental stages (Online Resource\_3).

The confirmation of the expression levels obtained by the bioinformatics analyses was made by RT-qPCR of 19 genes related with embryogenesis and with distinct profiles across somatic embryo developmental stages (Fig. 5). Correlation of gene expression levels obtained by RT-qPCR and Illumina sequencing was demonstrated by Pearson's correlation in which the majority of the genes showed a high correlation level:  $r > 0.9$ . A moderate correlation was also observed ( $0.6 > r > 0.9$ ) for six of the selected genes.

The GO annotation for the 11,507 DEGs revealed that cellular and metabolic processes are the main categories represented in biological processes, each accounting for 31.51% and 28.31%, respectively. Genes involved in other important biological processes such as single-organism processes (8.95%), response to stimulus (6.60%), cellular component organization (4.57%), localization (3.99%), multicellular organism processes (3.38%), and developmental processes (3.33%) were also identified (Online Resource\_4). Regarding molecular function categories, GO terms for the DEGs are mainly related with binding and catalytic activities. In the catalytic subset, the main groups represented were heterocyclic compound binding (21.71%), organic cyclic compound binding (21.71%), small molecule binding (15.05%), protein binding (4.09%), and carbohydrate binding (1.34%). In the subset of catalytic activities, two main groups were represented: transferase activity (18.63%) and hydrolase activity (13.67%). Transcription factor activity (1.30%) and signal transducer activity (0.80%) were also represented (Online Resource\_4). In terms of cellular components, identified genes are mainly related to intracellular membrane-bounded organelle (29.07%), cytoplasm (27.41%), cytoplasmic part (23.27%), intracellular non-membrane-bounded organelle (4.12%), thylakoid (3.76%), and cell wall (2.53%) (Online Resource\_4). Data from all GOSlim-annotated DE genes are listed in Online Resource\_5. In order to understand the enrichment occurrence of overrepresented biological process in DEGs clusters defined in K-means hierarchical analysis, we compared gene clusters proportion with its transcriptome assembly occurrence. This analysis showed that significantly enriched GO ontologies were generation of precursor metabolites and energy (clusters 5, 8, and 25); secondary metabolic processes (cluster 8 and 11); cell cycle (cluster 15 and 22); photosynthesis (cluster 5 and 8); DNA metabolic processes



**Fig. 4** Downregulated and upregulated genes (a) and Venn diagram showing the differentially expressed genes among three comparisons of developmental stages pairs (b)





◀ **Fig. 5** Validation of RNA-Seq transcript expression profiles. The comparison of RNA-Seq and RT-qPCR data assays was calculated with the Pearson correlation and expressed by the  $r$  value. In the  $y$ -axis is represented the  $\log_2$  of the relative expression level values from RNA-Seq data (right  $y$ -axis scale) and RT-qPCR (left  $y$ -axis scale) in each developmental stage

(cluster 15); translation (cluster 16); cellular component organization; DNA metabolic processes, embryo development, and epigenetic regulation of gene expression (all at cluster 22); cellular protein modification process and carbohydrate metabolic process (cluster 28); and response to biotic and external stimulus (cluster 29) (Fig. 6). These results give a broad view of overrepresented biological processes involved in some gene clusters, allowing monitoring of gene expression patterns with complementary association to biological processes, molecular function, and cellular components.

### Transcription factors, transcription regulators, and chromatin regulator genes involved in cork oak somatic embryogenesis

In order to identify and categorize TF, TR, and CR genes related with somatic embryo developmental stages, a list of plant TF, TR, and CR genes was generated from libraries ST1, ST2, ST3, and ST4. The analysis identified 2674 TF, TR, and CR genes from which 1159 were differentially expressed among developmental stages (Online Resource\_6). The most predominant family type of regulators is TFs (76%), followed by transcription factor iterators (8%) and CRs (4%). Among TFs, the most abundant belongs to C2H2, MYB-HB-like, WD40-like, AP2-EREBP, bHLH, WRKY, NAM, and bZIP families of TFs, to the CCHC (Zn) family of TRs and to the PHD family of CRs. GO annotation enrichment analysis was performed to identify overrepresented biological processes related with the identified differentially expressed regulators. The most relevant GO terms were associated with response to endogenous stimulus, post-embryonic development, signal transduction, response to abiotic stimulus, cell communication, flower development, cell differentiation, and anatomical structure morphogenesis (Fig. 7).

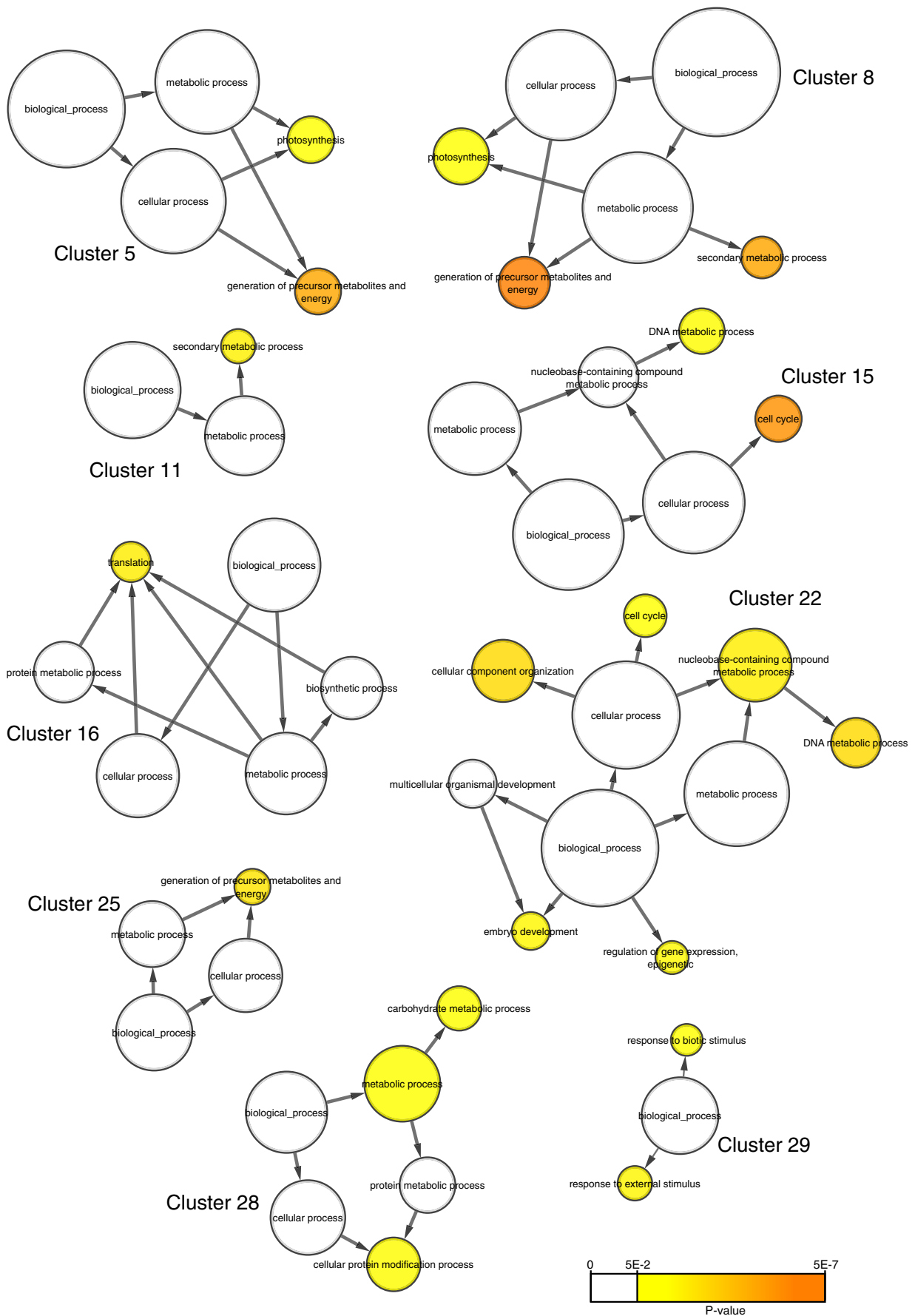
A search for genes involved in embryo development and maturation was made taking into account blast similarity and gene function in *Arabidopsis thaliana*. AP2-EREBP TFs with high homology for *AINTEGUMENTA*-like (*AIL*) gene were identified, namely homologs of *AINTEGUMENTA-LIKE 5* (*AIL5*) and *AINTEGUMENTA-LIKE 6* (*AIL6*). Interestingly, a homolog of *PLETHORA 2* (*PLT2*) which is also an AP2-EREBP transcription factor is upregulated in ST1, being 25–27 times more expressed than that in ST2 and ST3, and a homolog of the *AINTEGUMENTA* (*ANT*) gene was found to be upregulated in ST3. With an opposite expression pattern a *CYTOKININ RESPONSE FACTOR 3* (*CRF3*) homolog was

identified as downregulated in the ST3 stage. We also found a homolog of *GATA transcription factor 8* (*GATA8*) expressed in all developmental stages, with a particularly higher expression in ST2 and ST3.

AUXIN RESPONSE FACTORS (ARFs) are TFs, which control auxin-regulated gene transcription. Two homologs of this type of TFs, *AUXIN RESPONSE FACTOR 4* (*ARF4*) and *ARF5*, were also identified, both of which were upregulated in ST1.

### Expression profile of homeostasis-related phytohormone genes

In order to explore hormone-mediated transcriptional regulation in gene expression during embryo development, we mapped the DEGs to eight categorical hormone-related pathways including abscisic acid, auxin, brassinosteroid, cytokinin, ethylene, gibberellin, jasmonic acid, and salicylic acid. The analysis revealed that 250 genes associated with hormone biosynthesis, response, signaling, receptors, and metabolism showed significant differential expression during somatic embryo development (Online Resource\_7). The main group showing significant DE is related with abscisic acid (27.6%), followed by auxin (23.2%), ethylene (20.0%), jasmonic acid and salicylic acid (7.6%), gibberellin (5.6%), brassinosteroid (5.2%), and cytokinin (3.2%). In terms of biological functions, we found relevant *Arabidopsis thaliana* homolog genes within the ABA group, namely the *PHOSPHOLIPASE D ALPHA 1* (*PLDALPHA1*) and the *NUCLEAR CAP-BINDING PROTEIN SUBUNIT 1* (*ABH1*), which were highly expressed in all developmental stages. Within the AUX group, we found homologs of *AUXIN EFFLUX CARRIER COMPONENT 2* (*PIN2*), which was highly upregulated in ST1; homologs of *CULLIN-1* (*CUL1*) and *TOPLESS* (*TPL*) were found to be upregulated in both ST2 and ST3. A homolog of *S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME 4* (*SAMDC4*) was also identified as upregulated in ST1 and ST3. Related with AUX biosynthesis, a homolog of *INDOLE-3-PYRUVATE MONOOXYGENASE YUCCA4* (*YUC4*) was identified as upregulated in ST1 and the *GUANINE-NUCLEOTIDE EXCHANGE FACTOR GNOM* (*GN*) was downregulated at ST4. Also related with AUX homeostasis, we identified a putative gene encoding *ESCRT-RELATED PROTEIN CHMP1B* (*CHMP1B*), which was highly expressed in all embryo developmental stages, especially in ST4. Downregulated in ST4 was a putative coding sequence related with the BR signaling pathway, a homolog of the *SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1* (*SERK1*) gene. The approach used to identify the hormone-mediated transcriptional regulated genes, also allowed the identification of an ET and JA responsive gene highly expressed in all developmental stages, especially in ST3, the *CELLULOSE SYNTHASE A CATALYTIC*



**Fig. 6** Biological process network of GO term enrichment for DE genes in K-means clusters with over-represented GO terms. The node size represents the number of genes associated to a given GO term and node fill color reflects the adjusted  $P$  value

*SUBUNIT 3 (CESA3)* homolog. A group of hormone-related genes not expressed in at least one developmental stage were also identified, namely *THERMOSPERMINE SYNTHASE ACAULIS5 (ACL5)* *XYLOGLUCAN ENDOTRANSGLUCOSYLASE/HYDROLASE PROTEIN 24 (XTH24)* and three *HISTIDINE KINASE* isoforms (*HK1*, *HK3*, and *HK4*).

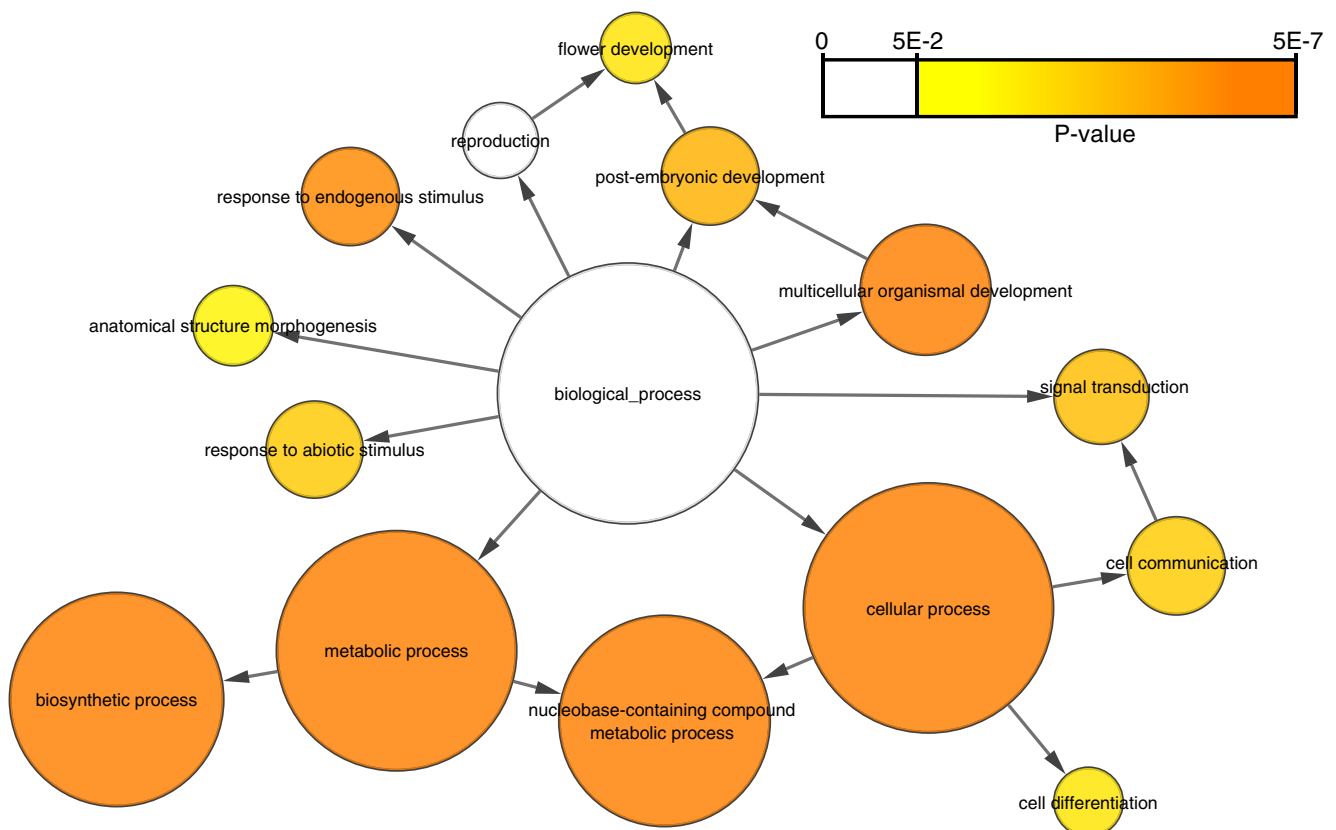
### Expression profile of genes involved in germination

Plant hormones can affect several biological processes including seed dormancy and germination (Graeber et al. 2012). Although seeds do not occur in the *in vitro* somatic embryo culture system used in this study, genes involved in seed germination may have analog function in the embryo maturation process. In fact, genes described as involved in seed germination and related with hormone homeostasis were found to be differentially expressed, namely transcripts with homology for ABA- and GA-related genes, which are the major class of hormones responsible for seed germination (Miransari and

Smith 2014). Within these, we highlight ABA-related genes like homologs of the *ABC TRANSPORTER C FAMILY MEMBER 5 (ABCC5)*, *ARM REPEAT PROTEIN INTERACTING WITH ABF2 (ARIA)*, and *BETA-D-XYLOSE 6 (BXL6)* which were upregulated in ST2 and ST3. Related with CK homeostasis, two *HISTIDINE KINASE (HK)* were identified, an *HK3* homolog upregulated in ST4 and an *HK4* homolog downregulated in ST2. Related with GA signaling pathway, one homolog of the *DELLA PROTEIN (RGL1)* upregulated in ST3 was found. Related with JA biosynthesis, a homolog of *3-KETOACYL-COA THIOLASE 2 (PED1)* which showed an increasing expression profile through ST1 to ST4 and a homolog of the *TWO PORE CALCIUM CHANNEL PROTEIN 1 (TPC1)* downregulated at ST4 were reported.

### Expression profiles of other embryogenesis-related genes

Embryogenesis-related genes may be involved in a broad range of biological processes and molecular functions. Besides TFs, TRs, CRs, and phytohormones that regulate diverse developmental changes in embryogenesis, which were categorized above, it is reasonable to consider that there are more genes involved in embryo development than genes which occur in the described categories. In order to overcome



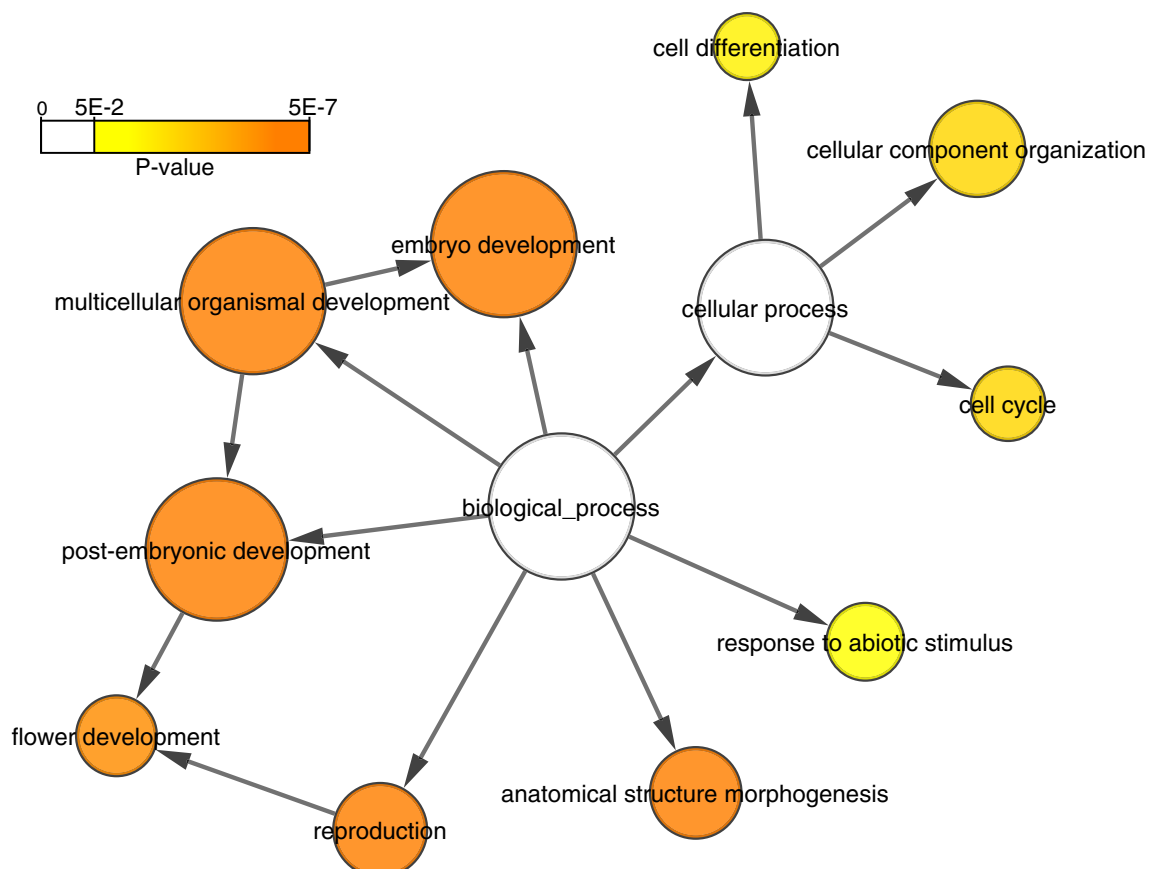
**Fig. 7** Biological process network of GO term enrichment of DE transcription factors. The node size represents the number of genes associated to a given GO term and node fill color reflects the adjusted  $P$  value

this, we manually searched the DE gene dataset for blast descriptions and GO terms related with embryo and embryogenesis. Furthermore, we excluded genes previously described in this study such as TFs, TRs, CRs, and homeostasis-related phytohormone genes. A list of 67 genes was compiled and a GO enrichment map, using the DEGs as reference set, was generated (Fig. 8). Gene ontologies related with embryo development and post-embryonic development were the most significantly enriched categories in this subset of genes. Besides of the relevant role in embryogenesis, these genes are also involved in anatomical structure morphogenesis, response to reproduction and flower development, abiotic stimulus and cellular processes (cell cycle, cellular component organization, and cell differentiation). A list of these genes, as well as their expression levels, is summarized in Online Resource\_8. The most relevant gene groups that are important to highlight contain embryo defective, embryo abundant, and late embryogenesis abundant protein (LEA) homologs with different expression profiles during the four developmental stages and three pentatricopeptide repeat protein (PPR) homologs downregulated in ST2, namely the *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN*

*At3g06430*, *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN At3g49240*, and *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN At3g53700*. A homolog of *SUMO-ACTIVATING ENZYME SUBUNIT 2 (SAE2)* and a homolog of *SERINE/THREONINE-PROTEIN KINASE-LIKE PROTEIN ACR4 (ACR4)* which were both upregulated in ST1 and a homolog of the *ABC TRANSPORTER I FAMILY MEMBER 6 (ABC16)* upregulated in ST4 were also identified. Interestingly, a homolog of *TRANSLATIONALLY CONTROLLED TUMOR PROTEIN 1 (TCPC1)* with high expression values during all developmental stages, particularly in ST1, was identified.

## Discussion

*Quercus suber* somatic embryos are an important biological resource for future breeding programs in this species. We provide the first overview of the transcriptome dynamics of somatic embryo development in *Q. suber* which will increase our understanding of the complex mechanism underlying the process. A total of 66,693 genes were identified with distinct



**Fig. 8** Biological process network of GO term enrichment of DE genes related with embryogenesis GO terms not included in the transcription factors and hormone-related genes sets. The node size represents the

number of genes associated to a given GO term and node fill color reflects the adjusted *P* value

roles in various biological processes, where 11,507 were differentially expressed between embryo stages. Several genes grouped in different expression clusters (Fig. 5) associated with biological processes namely generation of precursor metabolites and energy, secondary metabolic processes, cell cycle, cellular component organization, embryo development, and response to external stimulus (Fig. 6). We centered the analyses on transcripts with potential functions in embryogenesis, namely TF, TR, and CR, transcripts related with phytohormone homeostasis and germination (Table 2).

### Transcription factors, transcription regulators, and chromatin regulators involved in cork oak somatic embryogenesis and somatic embryo development

Recent genetic studies have identified genes involved in embryogenesis initiation and progression as well as in embryo development (Tzafrir et al. 2004; Abid et al. 2010; Radoeva and Weijers 2014). Many of these genes encode TFs able to control various plant biological processes regulating gene expression such as embryonic and post-embryonic development or embryo germination (Radoeva and Weijers 2014). In this study, 1159 differentially expressed TFs with gene ontologies deeply related with response to stimulus, signal transduction, and structure development were identified (Fig. 7). Particularly, we found a group of AP2-ERE BP TFs, including *AIL5*, *AIL6*, *PLT2*, and *ANT* differentially expressed across the four stages of cork somatic embryo development. AP2-ERE BP transcription factor family play an essential role in cell proliferation and embryogenesis (El Ouakfaoui et al. 2010). *BABY BOOM*, a member of the family, is preferentially expressed in developing embryos, and when overexpressed, it induces the formation of somatic embryos from leaves, cotyledons, and shoot apical meristem (Boutilier et al. 2002). A close homolog of *BBM* is *AIL5*, also known as *EMBRYOMAKER* (*EMK*) or *PLT5* (Tsuwamoto et al. 2010; Prasad et al. 2011). Our data revealed an ortholog of *AIL5* in cork oak. Its expression pattern shows that the gene is highly expressed in the four developmental stages, which suggests that *QsAIL5* may have an active function during cork oak somatic embryogenesis and similarly to its ortholog in *Arabidopsis* may have an important role in embryonic identity maintenance (Tsuwamoto et al. 2010). In addition, *AIL6* together with *PLT1*, *PLT2*, and *BBM* have redundant functions in root meristem and embryo differentiation and in combination with *AIL5* control shoot phyllotaxis (Galinha et al. 2007; Prasad et al. 2011). *PLT2* is also described as responsible for controlling the patterning within the root stem cell niche and is required for root stem cell activity during embryogenesis (Aida et al. 2004). Therefore, we hypothesize that *QsAIL6* and *QsPLT2* may play a role in cork oak hypocotyl and root

development. Transcription of *PLT* genes is stimulated by auxin and is dependent on ARF transcription factors (Benjamins and Scheres 2008; Mahonen et al. 2014). In this context, it is particularly interesting to observe that both genes show identical decreasing expression from ST1 to ST4, which may partially explain the difficulty of cork oak somatic embryo root formation without the use of plant growth regulators. Specifically, *ARF5* is critical for embryonic root formation (Schlereth et al. 2010), embryo axis formation, and vascular development (Hardtke and Berleth 1998). This study found an ortholog of *ARF5*, which is upregulated in ST1. The relative high expression of *QsARF5*, *QsPLT2*, *QsAIL6* in ST1 suggests that these TFs may act together in early stages of cork oak somatic embryogenesis to control cell identity and embryo differentiation similar to what is observed in *Glycine max* (El Ouakfaoui et al. 2010). However, detection of other members of the ARF family in the same stage points to a redundant function of genes of this family. We found an ortholog of *AtARF4* also upregulated in ST1. In *Q. suber*, *QsARF4* expression is specifically associated with the initial phase of acorn development (Miguel et al. 2015). In *Arabidopsis*, *AtARF4* seems to be redundant in the establishment of tissue patterning and floral meristem determinacy (Liu et al. 2014). This points to the importance and occurrence of tissue organization and patterning in early stages of cork oak development; however, further studies have been performed in order to understand the function of the ARF genes in cork oak embryogenesis. CYTOKININ RESPONSE FACTORS (CRFs) also belong to the AP2-ERE BP TF family. Analysis of loss-of-function mutations revealed that CRFs function redundantly in order to regulate embryo, cotyledon, and leaf development (Rashotte et al. 2006). Our data shows an ortholog of *CRF3* with differential expression in ST1. As a transcription factor, *QsCRF3* may have a key role in cork oak embryo development linking the complex transcriptional regulatory network involving TFs and phytohormones action. Within the differentially expressed TFs identified we found also a member of the GATA TF family. Its orthology with *AtGATA8*, which is a positive regulator of seed germination (Liu et al. 2005), lead us to hypothesize a putative role of *QsGATA8* in somatic embryo germination since we observed a reduction of gene expression levels at ST3 and ST4, which may also be related with the difficulty for cork oak somatic embryos to convert into emblings, documented in several studies (García-Martin et al. 2001; Pinto et al. 2002; Hernández et al. 2003).

### Homeostasis-related phytohormone genes involved in cork oak somatic embryogenesis and somatic embryo development

Hormones are involved in modulating gene expression by controlling the abundance of two types of gene regulatory

**Table 2** Expression values in counts per million (CPM) of relevant genes identified in cork oak somatic embryo development

Gene ID	Gene name	Description	Relevance/function	Developmental stage gene expression (CPM)			
				ST1	ST2	ST3	ST4
Transcription Factors, transcription regulators, and chromatin regulators							
TR81709	AIL5	AP2-like ethylene-responsive transcription factor AIL5	Embryonic identity maintenance (Tsuwamoto et al. 2010), shoot phyllotaxis (Galinha et al. 2007; Prasad et al. 2011)	10,226	8338	5097	4510
TR47049	AIL6	AP2-like ethylene-responsive transcription factor AIL6	Root meristem, embryo differentiation, shoot phyllotaxis (Galinha et al. 2007; Prasad et al. 2011)	8806	3598	3103	1568
TR72814	PLT2	AP2-like ethylene-responsive transcription factor PLT2	Root meristem and embryo differentiation (Aida et al. 2004)	7091	276	256	752
TR22775	ANT	AP2-like ethylene-responsive transcription factor ANT	Cell proliferation and embryogenesis (El Ouakfaoui et al. 2010)	54	387	1508	145
TR67174	ARF4	Auxin response factor 4	Acorn development (Miguel et al. 2015), tissue patterning and meristem determinacy (Liu et al. 2014)	291	40	81	128
TR42927	ARF5	Auxin response factor 5	Embryonic root and embryo axis formation (Schlereth et al. 2010); vascular development (Hardtke and Berleth 1998)	109	32	64	12
TR22093	CRF3	Ethylene-responsive transcription factor CRF3	Component of CK signaling pathway involved in embryo development (Rashotte et al. 2006)	372	255	63	137
TR17944	GATA8	GATA transcription factor 8	Seed germination (Liu et al. 2005)	1353	3386	2678	1716
Phytohormone homeostasis-related genes							
TR46556	CUL1	Cullin 1	Hormone response and signaling, morphogenesis and control of circadian clock (Shen et al. 2002; Willems et al. 2004; Harmon et al. 2008; Mockaitis and Estelle 2008)	28	53	56	20
TR58693	SAMDC4	S-Adenosylmethionine decarboxylase proenzyme 4	Embryogenesis and plant growth (Ge et al. 2006)	53	8	46	16
TR24883	YUC4	Probable indole-3-pyruvate monooxygenase YUCCA 4	Embryogenesis and organs formation (Cheng et al. 2007)	79	35	21	12
TR2935	CHMP1B	ESCRT-related protein CHMP1B	Embryo development and cellular differentiation (Spitzer et al. 2009)	3021	2628	2918	4840
TR63436	SERK1	Somatic embryogenesis receptor kinase 1	Embryogenic competence regulation (Hu et al. 2005)	535	452	532	182
TR14543	ACL5	Thermospermine synthase ACAULIS 5	Vascular differentiation (Baima et al. 2014)	0	5	11	67
TR13159	ATHB8	Homeobox-leucine zipper protein ATHB8	Vascular differentiation (Milhinhos et al. 2013)	929	752	2098	488
Germination-related genes							
TR52039	ABCC5	ABC transporter C family member 5		648	1065	1027	354
TR31857	ABCC5	ABC transporter C family member 5	Primary root elongation, lateral root formation and germination (Gaedeke et al. 2001; Martinoia et al. 2002; Klein et al. 2003; Lee et al. 2004)	0	41	211	44
TR70220	ABCC5	ABC transporter C family member 5		2	2	0	52
TR41255	ABCC5	ABC transporter C family member 5		78	24	7	98
TR52039	ABCC5	ABC transporter C family member 5		648	1065	1027	354
TR15229	ARIA	Arm repeat protein interacting with ABF2	Seed germination and young seedling growth (Kim et al. 2004);	375	197	315	154
TR55482	TPC1	Two pore calcium channel protein 1	Germination involving ABA (Peiter et al. 2005) and JA homeostasis (Bonaventure et al. 2007)	42	22	40	11
TR62491	RGL1	DELLA protein RGL1	Negative regulator of GA-dependent germination (Wen and Chang 2002)	1479	1103	3154	1970
TR37074	HK3	Histidine kinase isoform 3	Mediator of seed germination via CK action (Riefler et al. 2006)	3606	4893	3761	5938
TR8919	HK4	Histidine kinase isoform 4		856	344	803	651

**Table 2** (continued)

Gene ID	Gene name	Description	Relevance/function	Developmental stage gene expression (CPM)			
				ST1	ST2	ST3	ST4
<i>Other embryogenesis-related genes</i>							
TR72405	PPR At3g06430	Pentatricopeptide repeat-containing protein At3g06430	Essential to embryogenesis (Lu et al. 2011)	591	236	585	666
TR37357	PPR At3g49240	Pentatricopeptide repeat-containing protein At3g49240	Germination and embryo development (Cushing et al. 2005)	1880	809	1383	902
TR45089	PPR At3g53700	Pentatricopeptide repeat-containing protein At3g53700	Involved cell death, signal transduction and transcriptional regulation in embryogenesis (Pagnussat et al. 2005)	476	183	308	481
TR11611	ABC16	ABC transporter I family member 6	Chloroplast differentiation during embryogenesis lethal (Xu and Møller 2004)	2	4	14	51
TR41251	SAE2	SUMO-activating enzyme subunit 2	Embryo development (Saracco et al. 2007)	6624	3344	3291	3058
TR80906	TCPC1	Translationally controlled tumor protein 1	Positive regulator of mitotic growth (Brioudes et al. 2010)	74,652	31,669	51,111	44,733
TR2413	ACR4	Serine threonine-protein kinase-like protein ACR4	Cell differentiation in embryogenesis (Tanaka et al. 2002)	2413	513	1041	254

proteins, TFs and transcriptional repressors integrating and coordinating plant developmental response. In this study, we found 250 DEGs associated with hormone biosynthesis, response, signaling, receptors, and metabolism. In early formed cork oak embryos, endogenous high levels of AUX associated with high methylation levels were reported after embryogenesis induction (Rodríguez-Sanz et al. 2014). We identified at least 3 AUX-related genes involved in embryogenesis and post-embryonic development: *CUL1*, *SAMDC4*, and *YUC4*. *CUL1* targets various proteins involved in hormone response and signaling, morphogenesis, and control of circadian clock (Shen et al. 2002; Willems et al. 2004; Harmon et al. 2008; Mockaitis and Estelle 2008). Several studies showed a reduced AUX response in *cull1* loss-of-function mutant development (Shen et al. 2002; Hellmann et al. 2003; Moon et al. 2007) and arrest in early embryogenesis stages in *Arabidopsis* (Shen et al. 2002). In this study, *CUL1* is expressed in all developmental stages and may play a key role in AUX homeostasis in cork oak embryogenesis. An ortholog of *SAMDC4*, an important decarboxylase in plant polyamine biosynthetic pathway, was also identified. In *Arabidopsis*, *SAMDC4* is essential for plant embryogenesis, normal growth, and development (Ge et al. 2006). In the present study, *QsSAMDC4* is expressed during all developmental stages and is downregulated in ST2 and ST4. *YUC4* is a AUX responsive gene and is related with hormone biosynthesis and plant development. YUC genes are mainly expressed in meristems, young primordia, vascular tissues, and reproductive organs. In *Arabidopsis*, Cheng and co-authors (Cheng et al. 2007) showed that AUX synthesized by YUC

genes is essential for establishment of the basal body region during embryogenesis and formation of embryonic and post-embryonic organs. Our data shows the presence of an ortholog of *YUC4* upregulated in ST1. This result may suggest a potential important involvement of *QsYUC4* in cork oak embryogenesis initiation and maintenance. The *CHMP1B* gene was also identified, which is related with multi vesicular body formation and sorting of *PIN* AUX carriers and essential for embryo development (Spitzer et al. 2009). In *Arabidopsis*, *CHMP1B* is involved in cellular differentiation and embryo symmetry establishment (Spitzer et al. 2009). We found *QsCHMP1B* to be expressed in all developmental stages, with high relative expression values, especially in ST4. Therefore, *QsCHMP1B* may be relevant in embryo symmetry establishment and cellular differentiation in cork oak embryo development, particularly in ST4. An ortholog of *SERK1*, a BR receptor, was identified as downregulated in ST4. *SERK1* is described as a key factor in embryogenic competence regulation, as it regulates cell death and increases stress resistance (Hu et al. 2005). Moreover, *Arabidopsis* seedlings overexpressing *AtSERK1* exhibited an increased efficiency for initiation of somatic embryogenesis, which led to consider that an increase in *AtSERK1* expression of 3-to 4-fold is sufficient to confer embryogenic competence in culture (Hecht et al. 2001). Our results show that secondary embryogenesis was observed in embryo clusters in ST1 to ST3, but this capacity was lost in ST4, where embryos follow a maturation pathway expressing *QsSERK1* from ST1 to ST3 with similar mRNA expression levels, however in ST4 *QsSERK1* decreases about 2.5-to 3-fold

resembling that *QsSERK1* may have the same ability as its ortholog in *Arabidopsis*.

Our data also revealed several differential expressed genes not expressed in at least one developmental stage. *ACL5* is not expressed in ST1 but its expression increases from ST2 to ST4. In *Arabidopsis*, *ACL5* is required for xylem specification by regulating and preventing premature death of xylem vessel elements (Hanzawa et al. 2000; Kakehi et al. 2008; Muñiz et al. 2008). Authors demonstrated that thermospermine produced by *ACL5* is one of the factors downstream to AUX synthesis contributing to the regulation of vascular differentiation (Baima et al. 2014). Recently, Milhinhos and co-workers (Milhinhos et al. 2013) described a feedback regulatory path in poplar secondary xylem in which thermospermine levels are controlled by an AUX-dependent feedback loop mechanism involving the poplar ortholog of *Arabidopsis thaliana* *HOMEBOX 8 (ATHB8)*. Specifically, *ATHB8* is expressed in procambial cells and is directly regulated by *ARF5* (Baima et al. 2014). Our data reveals the presence of two *QsATHB8* orthologs, expressed in all developmental stages and with similar expression patterns. Our report describes the presence of well-known key players in AUX-mediated process including hypocotyl elongation and differentiation of xylem precursor cells. Taking into account the recalcitrance of cork oak somatic embryos to elongate and form leaves in vitro, identification of these players may constitute important findings to design modulation strategies for cork oak plantlets conversion.

### Genes involved in cork oak somatic embryo germination

The proper maturation and germination of somatic embryos constitutes one of the most important factors in conversion into plants (Vieitez et al. 2012). Pérez et al. 2015 reported a decrease in ABA content and in 5-methyl-deoxycytidine during cork oak somatic embryo germination. Although epigenetic control and dynamic ABA levels appear to play an important role for the correct maturation and subsequent germination of embryos (Pérez et al. 2015), identification of other elements is needed to fully understand the process.

The current study also identifies several expressed genes related with germination, namely *ABCC5* homologs, upregulated in stage ST2 and ST3. *ABCC5* is involved in primary root elongation and lateral root formation and also plays a role in ABA-mediated germination inhibition (Gaedeke et al. 2001; Martinoia et al. 2002; Klein et al. 2003; Lee et al. 2004). These transcripts may have important roles in cork oak somatic embryos germination, opening the gate for future studies about this topic. In addition, results revealed the presence of *ARIA* orthologs, which acts as positive regulator of ABA response (Kim et al. 2004). In our data, the putative *QsARIA* is downregulated in ST4 which may reflect a

decreasing effect of ABA germination inhibition, however the gene is still expressed which may be insufficient for embryo germination to take place in the studied embryogenic cell line. Also involved in regulation of germination is *QsTPC1* downregulated in ST4. *TPC1* is reported to inhibit ABA-dependent germination (Peiter et al. 2005) and to be involved in JA homeostasis (Bonaventure et al. 2007). Taking together the expression patterns of *ARIA* and *TPC1*, the results led us to hypothesis the presence of a genetic driving force inhibiting cork oak embryo germination, specifically at ST4. In contrast to ABA regulation of germination, GA promotes seed germination in many plant species (Seo et al. 2009). The regulation of GA homeostasis is in part related with *DELTA* protein action (Cao et al. 2005). In our data, we found an ortholog of *RGL1*, a member of the *DELTA* family, which is a well-known negative regulator of GA response (Wen and Chang 2002), upregulated in ST3 and ST4. *RGL1* function is related with negative regulation of the GA signaling pathway, therefore the high expression levels observed particularly in ST3 and ST4 may explain in part the low cork oak embryo germination rates by impairing GA action in germination. Another plant hormone which plays positive and negative regulatory roles in many aspects of plant growth and development is cytokinin. Riefler and co-workers demonstrate that different receptors were able to mediate CK control of seed germination, particularly *AtHK4* and *AtHK3*, showing that CK is a negative regulator of seed germination, and distinct pathways are being controlled by different CK-hormone receptors (Riefler et al. 2006). Expression levels of *HK3* and *HK4* *Q. suber* orthologs are opposed along somatic embryogenesis development, meaning when expression of one gene decreases the expression of the other increases. This evidence may be related with redundant function of histidine kinase receptors. These findings may reveal players acting in cork oak germination inhibition however further studies are required to clarify the effect of the genes in cork oak somatic embryo conversion. It will be interesting to study how the genes respond to different hormone treatments, namely to ABA, GA and CK and how the repression and overexpression of these genes will affect conversion. These findings may have a deep impact on future cork oak reforestation and breeding programs since the low percentage of somatic embryo germination is a crucial factor in the process.

### Other relevant genes in cork oak somatic embryogenesis

Somatic embryogenesis is marked by high mitotic rate and accompanied by morphogenetic events affected by many factors, such as stress signals and primary compound synthesis. A proteomic approach of *Q. suber* somatic embryogenesis already revealed important players in the process namely proteins associated with glycolysis, hormone biosynthesis, compound storage,



cell division, reactive oxygen species detoxification, and stress response (Gomez-Garay et al. 2013). An important group of proteins involved in response to biotic and abiotic stress is the LEA protein family (Hincha and Thalhammer 2012). We identified several LEA transcripts, namely hydroxyproline-rich glycoproteins and three PPR proteins differentially expressed and closely related to embryo development aspects. The first PPR gene identified was an ortholog of the *PPR At3g06430*. *PPR At3g06430* mutants are related with delayed embryogenesis and germination abnormalities (Lu et al. 2011). The second *PPR* gene identified is an ortholog of *PPR At3g49240* also known as *EMB1796*. The *emb1796* mutants resulted in developmental failure beyond globular stage, exhibiting consistent and severe developmental arrest (Cushing et al. 2005). Finally, the third PPR gene identified is an ortholog of *PPR At3g53700* which is involved in protein degradation, cell death, signal transduction, and transcriptional regulation required for early embryogenesis (Pagnussat et al. 2005). Furthermore, we identified orthologs of *ABCI6* and *SAE2*, other two genes essential for embryogenesis. *AtABCI6* plays an important role in plastid Fe-S cluster maintenance and repair during embryogenesis and as *AtSAE2* their mutants are embryonic lethal (Xu and Møller 2004; Saracco et al. 2007). The presence of these genes in the data may constitute evidence for a group of essential genes for cork oak embryo development. In fact, the GO term “embryo development” is enriched in the “other relevant genes in cork oak somatic embryogenesis” subset (Fig. 8). Plant development is a process involving precise coordination and regulation of cell proliferation and differentiation. *TCPC1* is a positive regulator of mitotic growth controlling the duration of the cell cycle both in animals and in plants and therefore acting in a complex but poorly understood regulation network (Brioudes et al. 2010). *ACR4* is another important player in cell division, particularly in meristems, root tips, and lateral initiation zones of the pericycle (Watanabe et al. 2004; De Smet et al. 2008; Stahl et al. 2009). In addition, it is required during embryogenesis and embryo development (Tanaka et al. 2002). The expression of *QsTCPC1* increases from ST1 to ST4, probably promoting cell division and embryo growth observed through developmental stages. Also, *ACR4* is upregulated in ST1 and ST3 which may indicate a high meristematic activity at these stages of development. Identification of these genes represents an important advance in the knowledge of cork oak somatic embryogenesis. Nevertheless, functional studies will be needed in the near future in order to determine the exact role of each referred gene.

This work allowed for the characterization of cork oak somatic embryogenesis and for the identification of several genes in four distinct stages of embryo development, from a unique tree genotype. Genes related with transcriptional regulation were specially addressed in this study since they were the main class of regulators identified throughout the embryo developmental process and were involved in over-represented specific processes such as response to cell differentiation, endogenous stimulus, post-embryonic development, and signal

transduction (Figs. 6, 7, and 8). Hormone-related genes were also revealed by the presence of genes involved in root formation and development as well as genes involved in germination which showed active expression dynamics in cork oak somatic embryogenesis. Functional studies of the genes identified in this work will be of great importance to better understand their role in somatic embryo development. However, it is fundamental to take into account that this study used a single tree genotype and that the results may not reflect what happens in other genotypes. However, the knowledge gathered in this study and in upcoming functional studies will be valuable for the design of successful regeneration protocols that can be integrated into future breeding programs. Additionally, the dataset generated in this work contributes to the enlargement of cork oak genomics resources.

**Acknowledgments** The authors acknowledge Fundação para a Ciência e a Tecnologia (FCT) for awarding a PhD grant to Tiago Capote (SFRH/BD/69785/2010) and for funding António Marcos Ramos, Anabel Usié, and Pedro Barbosa through the Project Investigador FCT IF/01015/2013/CP1209/CT0001 - Genomics and bioinformatics applied to Portuguese plant and animal genetic resources, the project PTDC/AGR-FOR/3356/2014, and the research unit LEAF Unit UID/AGR/04129/2013. We also thank the Program Alentejo 2020 funded through the European Fund for Regional Development under the scope of LENTIDEV – A molecular approach to cork porosity (REF: ALT20-03-0145-FEDER-000020).

**Funding** Fundação para a Ciência e a Tecnologia (FCT) funded the PhD grant for Tiago Capote (SFRH/BD/69785/2010). António Marcos Ramos, Anabel Usié, and Pedro Barbosa were funded by the Project Investigador FCT IF/01015/2013/CP1209/CT0001 - Genomics and bioinformatics applied to Portuguese plant and animal genetic resources. This work was funded by Program Alentejo 2020, through the European Fund for Regional Development under the scope of LENTIDEV – A molecular approach to cork porosity (REF: ALT20-03-0145-FEDER-000020).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

- Abid G, Jacquemin J, Sassi K, Muhovski Y (2010) Gene expression and genetic analysis during higher plants embryogenesis. *Biotechnol Agron Soc Environ* 14:667–680
- Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh YS, Amasino R, Scheres B (2004) The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. *Cell* 119: 119–120. <https://doi.org/10.1016/j.cell.2004.09.018>
- Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Vennetier M, Kitzberger T, Rigling A, Breshears DD, Hogg EH(T), Gonzalez P, Fensham R, Zhang Z, Castro J, Demidova N, Lim JH, Allard G, Running SW, Semerci A, Cobb N (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For Ecol Manag* 259:660–684. <https://doi.org/10.1016/j.foreco.2009.09.001>

- Álvarez R, Alonso P, Cortizo M et al (2004) Genetic transformation of selected mature cork oak (*Quercus suber* L.) trees. *Plant Cell Rep* 23:218–223. <https://doi.org/10.1007/s00299-004-0810-2>
- Álvarez R, Álvarez JM, Humara JM, Revilla Á, Ordás RJ (2009) Genetic transformation of cork oak (*Quercus suber* L.) for herbicide resistance. *Biotechnol Lett* 31:1477–1483. <https://doi.org/10.1007/s10529-009-0033-2>
- Baima S, Forte V, Possenti M, Peñalosa A, Leoni G, Salvi S, Felici B, Ruberti I, Morelli G (2014) Negative feedback regulation of auxin signaling by ATHB8/ACL5-BUD2 transcription module. *Mol Plant* 7:1006–1025. <https://doi.org/10.1093/mp/ssu051>
- Benjamins R, Scheres B (2008) Auxin: the looping star in plant development. *Annu Rev Plant Biol* 59:443–465. <https://doi.org/10.1146/annurev.arplant.58.032806.103805>
- Bonaventure G, Gfeller A, Proebsting WM, Hörtensteiner S, Chételat A, Martinoia E, Farmer EE (2007) A gain-of-function allele of TPC1 activates oxylipin biogenesis after leaf wounding in *Arabidopsis*. *Plant J* 49:889–898. <https://doi.org/10.1111/j.1365-313X.2006.03002.x>
- Boutillier K, Offringa R, Sharma VK et al (2002) Ectopic expression of BABY BOOM triggers a conversion from vegetative to embryonic growth. *Plant Cell* 14:1737–1749. <https://doi.org/10.1105/tpc.001941>
- Braisier CM (1996) Phytophthora cinnamomi and oak decline in southern Europe. Environmental constraints including climate change. *Ann Sci For* 53:347–358. <https://doi.org/10.1051/forest:19960217>
- Brasier CM (1992) Oak tree mortality in Iberia. *Nature* 360:539. <https://doi.org/10.1038/360539a0>
- Brioudes F, Thierry A-M, Chambrier P, Mollereau B, Bendahmane M (2010) Translationally controlled tumor protein is a conserved mitotic growth integrator in animals and plants. *Proc Natl Acad Sci U S A* 107:16384–16389. <https://doi.org/10.1073/pnas.1007926107>
- Bueno MA, Gómez A, Manzanera JA (2000) Somatic and Gametic Embryogenesis in *Quercus Suber* L. In: Jain SM, Gupta PK, Newton RJ (eds) *Somatic embryogenesis in woody plants: volume 6*. Springer Netherlands, Dordrecht, pp 479–508
- Cao D, Hussain A, Cheng H, Peng J (2005) Loss of function of four DELLA genes leads to light- and gibberellin-independent seed germination in *Arabidopsis*. *Planta* 223:105–113. <https://doi.org/10.1007/s00425-005-0057-3>
- Cheng Y, Dai X, Zhao Y (2007) Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in *Arabidopsis*. *Plant Cell* 19:2430–2439. <https://doi.org/10.1105/tpc.107.053009>
- Conesa A, Gotz S (2008) Blast2GO: a comprehensive suite for functional analysis in plant genomics. *Int J Plant Genomics* 2008:619832:1–12. <https://doi.org/10.1155/2008/619832>
- Cushing DA, Forsthoefel NR, Gestaut DR, Vernon DM (2005) *Arabidopsis* emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. *Planta* 221:424–436. <https://doi.org/10.1007/s00425-004-1452-x>
- Dai X, Sinharoy S, Udvardi M, Zhao PX (2013) PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. *BMC Bioinformatics* 14:321. <https://doi.org/10.1186/1471-2105-14-321>
- De Smet I, Vassileva V, De Rybel B et al (2008) Receptor-like kinase ACR4 restricts formative cell divisions in the *Arabidopsis* root. *Science* 322:594–597. <https://doi.org/10.1126/science.1160158>
- Fan D, Liu T, Li C, Jiao B, Li S, Hou Y, Luo K (2015) Efficient CRISPR/Cas9-mediated targeted mutagenesis in *Populus* in the first generation. *Sci Rep* 5:12217. <https://doi.org/10.1038/srep12217>
- Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. *Science* 330(6004):622–627. <https://doi.org/10.1126/science.1190614>
- Gaedeke N, Klein M, Kolkusaoglu U, Forestier C, Müller A, Ansoerge M, Becker D, Mamnun Y, Kuchler K, Schulz B, Mueller-Roeber B, Martinoia E (2001) The *Arabidopsis thaliana* ABC transporter AtMRP5 controls root development and stomata movement. *EMBO J* 20:1875–1887. <https://doi.org/10.1093/emboj/20.8.1875>
- Galinha C, Hofhuis H, Luijten M, Willemsen V, Bilou I, Heidstra R, Scheres B (2007) PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* 449:1053–1057. <https://doi.org/10.1038/nature06206>
- Gao M, Gropp G, Wei S (2012) Combinatorial networks regulating seed development and seed filling. In: Ken-Ichi Sato (ed) *Embryogenesis*. InTech, pp 189–228
- García-Martín G, González-Benito ME, Manzanera JA (2001) *Quercus suber* L. somatic embryo germination and plant conversion: pretreatments and germination conditions. *In Vitro Cellular & Developmental Biology-Plant* 37(2):190–198. <https://doi.org/10.1007/s11627-001-0033-y>
- Ge C, Cui X, Wang Y, Hu Y, Fu Z, Zhang D, Cheng Z, Li J (2006) BUD2, encoding an S-adenosylmethionine decarboxylase, is required for *Arabidopsis* growth and development. *Cell Res* 16:446–456. <https://doi.org/10.1038/sj.cr.7310056>
- Gomez-Garay A, Lopez JA, Camafeita E, Bueno MA, Pintos B (2013) Proteomic perspective of *Quercus suber* somatic embryogenesis. *J Proteome* 93:314–325. <https://doi.org/10.1016/j.jprot.2013.06.006>
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>
- Graeber K, Nakabayashi K, Miatton E et al (2012) Molecular mechanisms of seed dormancy. *Plant Cell Environ* 35:1769–1786. <https://doi.org/10.1111/j.1365-3040.2012.02542.x>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUASt: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C (2007) Combined networks regulating seed maturation. *Trends Plant Sci* 12:294–300. <https://doi.org/10.1016/j.tplants.2007.06.003>
- Hanzawa Y, Takahashi T, Michael AJ, Burtin D, Pineiro M, Coupland G, Komeda Y (2000) ACAULIS5, an *Arabidopsis* gene required for stem elongation, encodes a spermine synthase. *EMBO J* 19:4248–4256. <https://doi.org/10.1093/emboj/19.16.4248>
- Hardtke CS, Berleth T (1998) The *Arabidopsis* gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. *EMBO J* 17:1405–1411. <https://doi.org/10.1093/emboj/17.5.1405>
- Harmon F, Imaizumi T, Gray WM (2008) CUL1 regulates TOC1 protein stability in the *Arabidopsis* circadian clock. *Plant J* 55:568–579. <https://doi.org/10.1111/j.1365-313X.2008.03527.x>
- Hecht V, Vielle-Calzada JP, Hartog MV, Schmidt EDL, Boutilier K, Grossniklaus U, de Vries SC (2001) The *Arabidopsis* SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. *Plant Physiol* 127:803–816. <https://doi.org/10.1104/pp.010324>
- Hellemans J, Mortier G, De Paepe A et al (2007) qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* 8:R19. <https://doi.org/10.1186/gb-2007-8-2-r19>
- Hellmann H, Hobbie L, Chapman A, Dharmasiri S, Dharmasiri N, del Pozo C, Reinhardt D, Estelle M (2003) *Arabidopsis* AXR6 encodes CUL1 implicating SCF E3 ligases in auxin regulation of embryogenesis. *EMBO J* 22:3314–3325. <https://doi.org/10.1093/emboj/cdg335>

- Hernández I, Celestino C, Alegre J, Toribio M (2003) Vegetative propagation of *Quercus suber* L. by somatic embryogenesis. II. Plant regeneration from selected cork oak trees. *Plant Cell Rep* 21:765–770. <https://doi.org/10.1007/s00299-003-0604-y>
- Hincha DK, Thalhammer A (2012) LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem Soc Trans* 40:1000–1003. <https://doi.org/10.1042/BST20120109>
- Hu H, Xiong L, Yang Y (2005) Rice SERK1 gene positively regulates somatic embryogenesis of cultured cell and host defense response against fungal infection. *Planta* 222:107–117. <https://doi.org/10.1007/s00425-005-1534-4>
- Jiang Z, Liu X, Peng Z, Wan Y, Ji Y, He W, Wan W, Luo J, Guo H (2011) AHD2.0: an update version of arabidopsis hormone database for plant systematic studies. *Nucleic Acids Res* 39:1–7. <https://doi.org/10.1093/nar/gkq1066>
- Jing D, Zhang J, Xia Y, Kong L, OuYang F, Zhang S, Zhang H, Wang J (2017) Proteomic analysis of stress-related proteins and metabolic pathways in *Picea asperata* somatic embryos during partial desiccation. *Plant Biotechnol J* 15:27–38. <https://doi.org/10.1111/pbi.12588>
- Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S (2014) InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>
- Joshi N, Fass J (2011) Sickle: a sliding-window, adaptive, quality-based trimming tool for FastQ files (version 1.33) [software]. Available at <https://github.com/najoshi/sickle>
- Kakehi JJ, Kuwashiro Y, Niitsu M, Takahashi T (2008) Thermospermine is required for stem elongation in *Arabidopsis thaliana*. *Plant Cell Physiol* 49:1342–1349. <https://doi.org/10.1093/pcp/pen109>
- Kim S, Choi H, Ryu H-J, Park JH, Kim MD, Kim SY (2004) ARIA, an *Arabidopsis* arm repeat protein interacting with a transcriptional regulator of abscisic acid-responsive gene expression, is a novel abscisic acid signaling component. *Plant Physiol* 136:3639–3648. <https://doi.org/10.1104/pp.104.049189>
- Klein M, Perfus-Barbeoch L, Frelet A, Gaedeke N, Reinhardt D, Mueller-Roeber B, Martinoia E, Forestier C (2003) The plant multidrug resistance ABC transporter AtMRP5 is involved in guard cell hormonal signalling and water use. *Plant J* 33:119–129
- Lee EK, Kwon M, Ko J-H, Yi H, Hwang MG, Chang S, Cho MH (2004) Binding of sulfonylurea by AtMRP5, an *Arabidopsis* multidrug resistance-related protein that functions in salt tolerance. *Plant Physiol* 134:528–538. <https://doi.org/10.1104/pp.103.027045>
- Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Liu PP, Koizuka N, Martin RC, Nonogaki H (2005) The BME3 (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of *Arabidopsis* seed germination. *Plant J* 44:960–971. <https://doi.org/10.1111/j.1365-313X.2005.02588.x>
- Liu X, Dinh TT, Li D, Shi B, Li Y, Cao X, Guo L, Pan Y, Jiao Y, Chen X (2014) AUXIN RESPONSE FACTOR 3 integrates the functions of AGAMOUS and APETALA2 in floral meristem determinacy. *Plant J* 80:629–641. <https://doi.org/10.1111/tbj.12658>
- Liu Y, Li C, Wang H, Chen H, Berg H, Xia Y (2011) AtPPR2, an *Arabidopsis* pentatricopeptide repeat protein, binds to plastid 23S rRNA and plays an important role in the first mitotic division during gametogenesis and in cell proliferation during embryogenesis. *Plant J* 67:13–25. <https://doi.org/10.1111/j.1365-313X.2011.04569.x>
- Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21:3448–3449. <https://doi.org/10.1093/bioinformatics/bti551>
- Mahonen AP, ten TK, Siligato R et al (2014) PLETHORA gradient formation mechanism separates auxin responses. *Nature* 515:125–129. <https://doi.org/10.1038/nature13663>
- Mallón R, Valladares S, Corredoira E, Vieitez AM, Vidal N (2014) Overexpression of the chestnut CsTL1 gene coding for a thaumatin-like protein in somatic embryos of *Quercus robur*. *Plant Cell Tissue Organ Cult* 116:141–151. <https://doi.org/10.1007/s11240-013-0390-3>
- Martinoia E, Klein M, Geisler M, Bovet L, Forestier C, Kolukisaoglu Ü, Müller-Röber B, Schulz B (2002) Multifunctionality of plant ABC transporters—more than just detoxifiers. *Planta* 214:345–355. <https://doi.org/10.1007/s004250100661>
- McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res* 40:4288–4297. <https://doi.org/10.1093/nar/gks042>
- Miguel A, de Vega-Bartol J, Marum L, Chaves I, Santo T, Leitão J, Varela MC, Miguel CM (2015) Characterization of the cork oak transcriptome dynamics during acorn development. *BMC Plant Biol* 15:158. <https://doi.org/10.1186/s12870-015-0534-1>
- Milinhos A, Prestele J, Bollhöner B, Matos A, Vera-Sirera F, Rambla JL, Ljung K, Carbonell J, Blázquez MA, Tuominen H, Miguel CM (2013) Thermospermine levels are controlled by an auxin-dependent feedback loop mechanism in *Populus* xylem. *Plant J* 75:685–698. <https://doi.org/10.1111/tbj.12231>
- Miransari M, Smith DL (2014) Plant hormones and seed germination. *Environ Exp Bot* 99:110–121. <https://doi.org/10.1016/j.envexpbot.2013.11.005>
- Mockaitis K, Estelle M (2008) Auxin receptors and plant development: a new signaling paradigm. *Annu Rev Cell Dev Biol* 24:55–80. <https://doi.org/10.1146/annurev.cellbio.23.090506.123214>
- Moon J, Zhao Y, Dai X, Zhang W, Gray WM, Huq E, Estelle M (2007) A new CULLIN 1 mutant has altered responses to hormones and light in *Arabidopsis*. *Plant Physiol* 143:684–696. <https://doi.org/10.1104/pp.106.091439>
- Muñiz L, Minguet EG, Singh SK et al (2008) ACAULIS5 controls *Arabidopsis* xylem specification through the prevention of premature cell death. *Development* 135:2573–2582. <https://doi.org/10.1242/dev.019349>
- Murashige T, Skoog F (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiol Plant* 15:473–497. <https://doi.org/10.1111/j.1399-3054.1962.tb08052.x>
- El Ouakfaoui S, Schnell J, Abdeen A et al (2010) Control of somatic embryogenesis and embryo development by AP2 transcription factors. *Plant Mol Biol* 74:313–326. <https://doi.org/10.1007/s11103-010-9674-8>
- Pagnussat GC, Yu H-J, Ngo QA et al (2005) Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. *Development* 132:603–614. <https://doi.org/10.1242/dev.01595>
- Peiter E, Maathuis FJM, Mills LN, Knight H, Pelloux J, Hetherington AM, Sanders D (2005) The vacuolar Ca<sup>2+</sup>-activated channel TPC1 regulates germination and stomatal movement. *Nature* 434:404–408. <https://doi.org/10.1038/nature03381>
- Pérez M, Viejo M, LaCuesta M, Toorop P, Cañal MJ (2015) Epigenetic and hormonal profile during maturation of *Quercus suber* L. somatic embryos. *J Plant Physiol* 173:51–61. <https://doi.org/10.1016/j.jplph.2014.07.028>
- Pinto G, Valentim H, Costa A, Castro S, Santos C (2002) Somatic embryogenesis in leaf callus from a mature *Quercus suber* L. tree. *In Vitro Cellular & Developmental Biology-Plant* 38(6):569–572. <https://doi.org/10.1079/IVP2002352>

- Pintos B, Bueno MA, Cuenca B, Manzanera JA (2008) Synthetic seed production from encapsulated somatic embryos of cork oak (*Quercus suber* L.) and automated growth monitoring. *Plant Cell Tissue Organ Cult* 95:217–225. <https://doi.org/10.1007/s11240-008-9435-4>
- Prasad K, Grigg SP, Barkoulas M, Yadav RK, Sanchez-Perez GF, Pinon V, Blilou I, Hofhuis H, Dhonukshe P, Galinha C, Mähönen AP, Muller WH, Raman S, Verkleij AJ, Snel B, Reddy GV, Tsiantis M, Scheres B (2011) Arabidopsis PLETHORA transcription factors control phyllotaxis. *Curr Biol* 21:1123–1128. <https://doi.org/10.1016/j.cub.2011.05.009>
- Radoeva T, Weijers D (2014) A roadmap to embryo identity in plants. *Trends Plant Sci* 19:709–716. <https://doi.org/10.1016/j.tplants.2014.06.009>
- Rashotte AM, Mason MG, Hutchison CE, Ferreira FJ, Schaller GE, Kieber JJ (2006) A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. *Proc Natl Acad Sci U S A* 103:11081–11085. <https://doi.org/10.1073/pnas.0602038103>
- Riefler M, Novak O, Strnad M, Schmu T (2006) Arabidopsis cytokinin receptor mutants reveal functions in shoot growth, leaf senescence, seed size, germination, root development, and cytokinin. *Metabolism* 18:40–54. <https://doi.org/10.1105/tpc.105.037796.1>
- Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140. <https://doi.org/10.1093/bioinformatics/btp616>
- Rodríguez-Sanz H, Manzanera JA, Solís MT, Gómez-Garay A, Pintos B, Risueño MC, Testillano PS (2014) Early markers are present in both embryogenesis pathways from microspores and immature zygotic embryos in cork oak, *Quercus suber* L. *BMC Plant Biol* 14:1–18. <https://doi.org/10.1186/s12870-014-0224-4>
- Saeed AI, Bhagabati NK, Braisted JC et al (2006) TM4 microarray software suite. *Methods Enzymol* 411:134–193. [https://doi.org/10.1016/S0076-6879\(06\)11009-5](https://doi.org/10.1016/S0076-6879(06)11009-5)
- Saracco S a, Miller MJ, Kurepa J, Vierstra RD (2007) Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant Physiol* 145:119–134. <https://doi.org/10.1104/pp.107.102285>
- Schenk RU, Hildebrandt AC (1972) Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. *Can J Bot* 50:199–204. <https://doi.org/10.1139/b72-026>
- Schlereth A, Möller B, Liu W, Kientz M, Flipse J, Rademacher EH, Schmid M, Jürgens G, Weijers D (2010) MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. *Nature* 464:913–916. <https://doi.org/10.1038/nature08836>
- Seo M, Nambara E, Choi G, Yamaguchi S (2009) Interaction of light and hormone signals in germinating seeds. *Plant Mol Biol* 69:463–472. <https://doi.org/10.1007/s11103-008-9429-y>
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, 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. <https://doi.org/10.1101/gr.1239303>
- Shen W-H, Parmentier Y, Hellmann H, Lechner E, Dong A, Masson J, Granier F, Lepiniec L, Estelle M, Genschik P (2002) Null mutation of AtCUL1 causes arrest in early embryogenesis in Arabidopsis. *Mol Biol Cell* 13:1916–1928. <https://doi.org/10.1091/mbc.E02-02-0077>
- Spitzer C, Reyes FC, Buono R, Sliwinski MK, Haas TJ, Otegui MS (2009) The ESCRT-related CHMP1A and B proteins mediate multivesicular body sorting of auxin carriers in Arabidopsis and are required for plant development. *Plant Cell* 21:749–766. <https://doi.org/10.1105/tpc.108.064865>
- Stahl Y, Wink RH, Ingram GC, Simon R (2009) A signaling module controlling the stem cell niche in Arabidopsis root meristems. *Curr Biol* 19:909–914. <https://doi.org/10.1016/j.cub.2009.03.060>
- Tanaka H, Watanabe M, Watanabe D, Tanaka T, Machida C, Machida Y (2002) ACR4, a putative receptor kinase gene of Arabidopsis thaliana, that is expressed in the outer cell layers of embryos and plants, is involved in proper embryogenesis. *Plant Cell Physiol* 43:419–428. <https://doi.org/10.1093/pcp/pcf052>
- Tsuwamoto R, Yokoi S, Takahata Y (2010) Arabidopsis EMBRYOMAKER encoding an AP2 domain transcription factor plays a key role in developmental change from vegetative to embryonic phase. *Plant Mol Biol* 73:481–492. <https://doi.org/10.1007/s11103-010-9634-3>
- Tzafirir I, Pena-muralla R, Dickerman A et al (2004) Identification of genes required for embryo development in Arabidopsis. *Plant Physiol* 135:1206–1220. <https://doi.org/10.1104/pp.104.045179> **published**
- Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res* 35:W71–W74. <https://doi.org/10.1093/nar/gkm306>
- Valladares S, Toribio M, Celestino C, Vieitez AM (2004) Cryopreservation of embryogenic cultures from mature *Quercus suber* trees using vitrification. *Cryo-Letters* 25:177–186
- Vieitez AM, Corredoira E, Martínez MT, San-José MC, Sánchez C, Valladares S, Vidal N, Ballester A (2012) Application of biotechnological tools to *Quercus* improvement. *Eur J For Res* 131:519–539. <https://doi.org/10.1007/s10342-011-0526-0>
- Von Arnold S, Sabala I, Bozhkov P et al (2002) Developmental pathways of somatic embryogenesis. *Plant Cell Tissue Organ Cult* 69:233–249. <https://doi.org/10.1023/A:1015673200621>
- Watanabe M, Tanaka H, Watanabe D, Machida C, Machida Y (2004) The ACR4 receptor-like kinase is required for surface formation of epidermis-related tissues in Arabidopsis thaliana. *Plant J* 39:298–308. <https://doi.org/10.1111/j.1365-313X.2004.02132.x>
- Wen C-K, Chang C (2002) Arabidopsis RGL1 encodes a negative regulator of gibberellin responses. *Plant Cell* 14:87–100. <https://doi.org/10.1105/tpc.010325>
- Willems AR, Schwab M, Tyers M (2004) A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. *Biochim Biophys Acta, Mol Cell Res* 1695:133–170. <https://doi.org/10.1016/j.bbamcr.2004.09.027>
- Wu XM, Kou SJ, Liu YL, Fang YN, Xu Q, Guo WW (2015) Genomewide analysis of small RNAs in nonembryogenic and embryogenic tissues of citrus: microRNA- and siRNA-mediated transcript cleavage involved in somatic embryogenesis. *Plant Biotechnol J* 13:383–394. <https://doi.org/10.1111/pbi.12317>
- Xu XM, Møller SG (2004) AtNAP7 is a plastidic SufC-like ATP-binding cassette/ATPase essential for Arabidopsis embryogenesis. *Proc Natl Acad Sci U S A* 101:9143–9148. <https://doi.org/10.1073/pnas.0400799101>
- Zdobnov EM, Apweiler R (2001) InterProScan—an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* 17:847–848. <https://doi.org/10.1093/bioinformatics/17.9.847>
- Zhang Y, Wang Y, Wang C (2012) Gene overexpression and gene silencing in birch using an agrobacterium-mediated transient expression system. *Mol Biol Rep* 39:5537–5541. <https://doi.org/10.1007/s11033-011-1357-2>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.