Physiol Mol Biol Plants (July–August 2018) 24(4):535–549
https://doi.org/10.1007/s12298-018-0526-3

RESEARCH ARTICLE

Characterization of the cork formation and production transcriptome in *Quercus cerris* × *suber* hybrids

Brígida Meireles[1] · Ana Usié[1,2] · Pedro Barbosa[1] · Ana Margarida Fortes[3] · André Folgado[1] · Inês Chaves[1] · Isabel Carrasquinho[4] · Rita Lourenço Costa[4,5] · Sónia Gonçalves[1,7] · Rita Teresa Teixeira[6] · António Marcos Ramos[1,2] · Filomena Nóbrega[4]

Received: 30 November 2017 / Revised: 13 March 2018 / Accepted: 20 March 2018 / Published online: 11 April 2018

Abstract Cork oak is the main cork-producing species worldwide, and plays a significant economic, ecological and social role in the Mediterranean countries, in particular in Portugal and Spain. The ability to produce cork is limited to a few species, hence it must involve specific regulation mechanisms that are unique to these species. However, to date, these mechanisms remain largely understudied, especially with approaches involving the use of high-throughput sequencing technology. In this study, the transcriptome of cork-producing and non-cork-producing *Quercus cerris* × *suber* hybrids was analyzed in order to elucidate the differences between the two groups of trees displaying contrasting phenotypes for cork production. The results revealed the presence of a significant number of genes exclusively associated with cork production, in the trees that developed cork. Moreover, several gene ontology subcategories, such as cell wall biogenesis, lipid metabolic processes, metal ion binding and apoplast/cell wall, were only detected in the trees with cork production. These results indicate the existence, at the transcriptome level, of mechanisms that seem to be unique and necessary for cork production, which is an advancement in our knowledge regarding the genetic regulation behind cork formation and production.

Keywords Cork oak · Transcriptome · Cork production · Hybrids

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s12298-018-0526-3) contains supplementary material, which is available to authorized users.

Brígida Meireles and Ana Usié have contributed equally to this work.

✉ António Marcos Ramos
marcos.ramos@cebal.pt

1 Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo (CEBAL), Instituto Politécnico de Beja (IPBeja), Beja, Portugal

2 Instituto de Ciências Agrárias e Ambientais Mediterrânicas (ICAAM), Universidade de Évora, Évora, Portugal

3 Faculdade de Ciências de Lisboa, Biosystems and Integrative Sciences Institute (BIOISI), Universidade de Lisboa, Lisbon, Portugal

4 Instituto Nacional de Investigação Agrária e Veterinária, I.P, Quinta do Marquês, 2780-159 Oeiras, Portugal

5 Centro de estudos Florestais, Instituto Superior de Agronomia, Universidade de Lisboa, Tapada da Ajuda, 1349-017 Lisbon, Portugal

6 Instituto Superior de Agronomia da Universidade de Lisboa (ISA), Tapada da Ajuda, 1349-017 Lisbon, Portugal

7 Present Address: Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB101SA, UK

Introduction

Quercus suber L., commonly known as cork oak, is an evergreen broad-leaved tree that belongs to the genus *Quercus* (oaks) of the *Fagaceae* family. Cork oak is widely distributed throughout the Western Mediterranean and North Africa regions, including Algeria, France, Italy, Morocco, Portugal, Spain and Tunisia, the islands of Corsica, Sardinia and Sicily and very limited areas on the islands of Majorca and Minorca (Bugalho et al. 2011). It is mostly recognized for its thick cork layer production (or phellem) that works as a protective multilayered dead tissue, composed of suberized cells that are located within the outer face of the periderm of the bark system and are produced towards the outside (Pereira 2007; Teixeira et al. 2014; Verdaguer et al. 2016). It is a renewable natural

resource of high commercial value. Its extraction is carried out by peeling off the cork layer from adult trees at regular intervals of at least 9 years, constituting a valuable raw material for cork industries and a very important source of income for Portugal. It is due to the commercialization of cork that this species is among the most important tree species with a great economic and social relevance in the countries where it is naturally distributed.

The formation of the cork tissue is the end result of the meristematic activity of a specialized phellogen tissue, or cork cambium, followed by cell expansion and an extensive cell wall deposition of suberin and waxes and, ultimately, an irreversible program of senescence ending in cell death (Soler et al. 2007). The phellogen usually originates in the subepidermis and differentiates during the first year, giving rise to phelloderm cells towards inside and to phellem (cork) cells to the outside. Hence, this meristematic layer is responsible for a sustainable production of the cork cells that form a thick continuous layer of cells covering the tree trunk, branches and roots. Every year, a new thick layer of suberized cork cells is produced and accumulated to that of the previous year in the form of annual rings (Caritat et al. 2000). This tissue is located between the inner bark (xylem or woody tissue) and cork. The phellogen activity period ranges from April to October (Graça and Pereira 2004). In the beginning of the spring, the phellogen is highly active and initiates the production of new cells. At the end of the season, by the time winter starts, the production of cells ceases (Silva et al. 2005).

The cork tissue produced by *Q. suber* has peculiar properties, defined by the chemical composition of cork cell walls, which is already well known and characterized by the presence of suberin as the main structural component associated to lignin, the second most important component (Pereira 2007; Marques and Pereira 2013; Lourenço et al. 2016). Suberin structures have been described as a complex heteropolymers consisting of polyaliphatic polyester (41%) and polyaromatic (22%) domains (Kolattukudy 1981; Pereira 1988; Silva et al. 2005; Pollard et al. 2008). At the chemical level, studies have demonstrated that the polyaliphatic components comprise mainly ω-hydroxyacids, α,ω-diacids, long and very long chain fatty acids (LCFA, VLCFAs), oxygenated fatty acids, unsubstituted fatty acids, primary fatty alcohols and glycerol (Graça and Santos 2007; Serra et al. 2009; Graça 2015; Vishwanath et al. 2015) and the polyaromatic domain is a lignin-like polymer containing mostly hydroxycinnamic acids (Ranathunge et al. 2011). At the molecular level, differential expression studies related to cork formation and suberin biosynthesis have revealed that four main secondary metabolic pathways are involved, including the synthesis of acyl-lipids, phenylpropanoids, isoprenoids, and flavonoids (Soler et al. 2007) and that the biosynthesis of the two main components derives from the general phenylpropanoid pathway and shares the same basic reactions (Rahantamalala et al. 2010). Molecular approaches performed using distinct plant tissues (roots, periderm, seeds) of the species *Arabidopsis thaliana, Solanum tuberosum* and, more recently, *Q. suber*, have already identified a list of key genes and its importance in the biosynthetic pathway of aliphatic suberin (Soler et al. 2007; Teixeira et al. 2014). The enzymes identified to date include β-ketoacyl-CoA synthases, fatty acyl reductases, long-chain acyl-CoA synthetases, cytochrome P450 monooxygenases, glycerol 3-phosphate acyltransferases and phenolics acyl transferases (Vishwanath et al. 2015).

Although some genomic studies have been published (Soler et al. 2007; Pereira-Leal et al. 2014; Teixeira et al. 2014), additional knowledge and information about suberin biosynthesis is required to obtain a clear understanding of cork development. A comprehensive analysis of the cork oak transcriptome analyzed in multiple tissues, developmental stages and physiological conditions was recently published (Pereira-Leal et al. 2014). One of these studies was focused on a population of *Quercus cerris* × *suber* hybrids with different cork production levels, where annual stems from two different developmental stages separately collected during the phellogenium period (April and October) were used as biological samples. The aims of this study were to increase the knowledge on the molecular mechanisms underlying the formation and development of cork and to identify transcripts differentially expressed between cork-producing and non-producing hybrids. Normalized and non-normalized cDNA libraries were generated and sequenced using the 454 GS-FLX Titanium technology, in order to further investigate the differences at the transcriptome level between the two types of hybrids.

Methods

Plant material

Annual stems from *Quercus cerris* × *suber* (*Quercus* × hispanica "Lucombeana") hybrids were collected twice during April/May, when the phellogen is highly active and cork growth is initiated, and in September/October, when the cell divisions become scarce. The two different collection dates allowed us to cover different phellogen developmental stages. This material was collected from 30 years old trees located at Mata Nacional do Vimeiro, Portugal (390 49′N; 90 19′W). A total of six hybrid trees belonging to the same open-pollinated family (half-sibs), and displaying contrasting phenotypes for cork formation (three trees with cork production and three trees without cork production) were selected. All the samples

were immediately frozen in liquid nitrogen and stored at −80 °C until use for subsequent analysis.

RNA extraction and cDNA preparation

Frozen samples were ground to a fine powder in liquid nitrogen with a pre-cooled electric mill and always under freezing conditions. Total RNA from each sample was isolated as previously described by Reid et al. (2006) with some minor modifications. RNA isolation was performed separately by developmental stage according to the two distinct dates of collection and by *Quercus cerris* × *suber* hybrid tree. Genomic DNA was eliminated using Ambion TURBO DNA-free DNase kit (Thermo Fisher Scientific, USA) according to the manufacturer's instructions and the total RNA was purified using the RNeasy MinElute Cleanup kit (Qiagen, Germany). The quality of total RNA was verified on an Agilent 2100 Bioanalyzer with the RNA 6000 Pico Kit (Agilent Technologies, Waldbronn, Germany) and the quantity assessed by fluorimetry with the Quant-iT RiboGreen RNA kit (Invitrogen, CA, USA). For cDNA production, the RNA of the two developmental stages was merged per individual. Equal concentrations of RNA were combined to create two specific RNA pools, one (Pool 1) for the trees with cork production and the other (Pool 2) for the trees without cork production. From each RNA pool, two libraries were constructed, one normalized (to optimize transcript detection and improve the general characterization of the transcriptome) and one non-normalized (to enable detection of differential expression). Double stranded cDNA was produced from 2 µg of total RNA using the SMART methodology (Zhu et al. 2001) and normalized according to the Evrogen Duplex-Specific nuclease (DSN)-based technology (Zhulidov et al. 2005). For the non-normalized cDNA libraries, the same two pools of RNA were used. From 50 µg of total RNA, mRNA was isolated with the MicroPoly(A) Purist kit (Ambion, USA). Then, 200 ng of mRNA was fragmented and used as template for double stranded cDNA production according to the procedure of the cDNA Synthesis System Kit (Roche, Germany), followed by adaptor ligation. Pyrosequencing of the two normalized and the two non-normalized libraries was performed in the 454 GS FLX Titanium according to standard manufacturer's instructions (Roche-454 Life Sciences, Brandford, CT, USA) at Biocant (Cantanhede, Portugal) resulting in a total of four EST libraries. Sequence reads were deposited in the NCBI Sequence Read Archive (SRA) under the accession numbers ERX143070 and ERX143071, for the normalized libraries, and SRX2677031 and SRX2677030, for the non-normalized libraries.

Reads pre-processing, de novo assembly and mapping

The raw reads were filtered using Sequence Cleaner (https://sourceforge.net/projects/seqclean/) and Mothur (Schloss et al. 2009) to remove barcodes, sequencing adaptors, reads with low average quality, poly A and poly T tails and sequences shorter than 100 nucleotides. After preprocessing, the reads that passed pre-processing were de novo assembled using MIRA version 4.9.5_2 (Chevreux et al. 2004). This assembly was subsequently used as the reference transcriptome against which mapping of preprocessed reads derived from the non-normalized libraries was performed, which was done for each library independently, using BWA MEM (Li and Durbin 2009). The mapped reads were filtered and the unique mapped reads (UMRs) were extracted with Samtools (Li et al. 2009). The reads denoted as unique mapped reads were the reads that had just one best hit and a map quality (MAPQ) $\geq$ 10.

Differential expression and variant calling analysis

Differentially expressed genes (DE genes) were identified as genes showing significant expression levels in stems tissues with cork production versus stems tissues without cork production using EdgeR (Robinson et al. 2010). Genes with low counts were excluded and a trimmed mean of M-values (TMM) to normalize the data was performed.

In accordance with the edgeR user's guide, since the data used in this study had no biological replicates to estimate the Biological Coefficient Variation (BCV), a reasonable dispersion value of the variability between the two conditions was estimated testing some values of the BCV, and for the final analysis, BCV was set to 0.075. The final list of DE genes was generated after correcting for multiple testing and filtering with a False Discovery Rate (FDR) value $\leq$ 0.05.

For SNP identification, a variant calling analysis was done using the Genome Analysis Toolkit (GATK)—Unified Genotyper (McKenna et al. 2010). The GATK detects all the differences between the reads and the reference genome, commonly named variants. Since not all variants can be considered valid SNPs, due to quality and coverage requirements, the initial output was filtered using three parameters, which included genotype coverage, genotype quality and SNP quality. Two approaches with different levels of stringency were carried out. The first was a moderate stringency, which used as thresholds a minimum variant quality score of 30, depth coverage of 7 and a phred-scaled quality score of 20, while the second approach was a high stringency, for which minimum thresholds of variant quality score of 60, depth coverage of 15 and a phred-scaled quality score of 40 were applied.

Validation of the differential expression profiles by RT-qPCR

Reverse transcription quantitative real-time PCR (RT-qPCR) of a set of 11 DE genes was carried out to validate the expression profile obtained by 454 sequencing. All cDNAs were synthesized from 1 μg of the same total RNA samples used for the 454 sequencing with the QuantiTect Reverse Transcription Kit (Qiagen, Germany) following manufacturer instructions. Specific primers were designed using Primer3Plus software (Untergasser et al. 2007). PCR experiments were then carried out in an iQ5 real-time PCR system (BioRad, USA). The reaction mixtures were performed in a final volume of 20 μL, containing 10 μL of SsoAdvanced Universal SYBR Green Supermix (BioRad), 250 nM of each gene-specific primers (Table S1) and 100 ng of cDNA as template. PCR efficiency was performed using triplicate reactions from a dilution series of cDNA and a no template control (NTC) was used for every primer pair. The amplification program was the same for all genes tested and was as follows: after an initial period of 10 min at 95 °C, each of the 45 PCR cycles consisted of a denaturation of 10 s at 95 °C, an annealing step of 15 s at 60 °C and an extension of 15 s at 72 °C. For each PCR reaction a melting curve was obtained to distinguish specific from nonspecific products and primer dimers. The change fold was calculated by means of the mathematical model described by Pfaffl (2001) using four reference genes involved in the cytoskeleton structure [Actin (Act), ß-Tubulin (ß-Tub)], in the translational elongation [Elongation factor-1alpha (EF-1α)] and in the intracellular protein transport [Clathrin adaptor complexes medium subunit family (CACs)], previously reported for *Q. suber* (Marum et al. 2012). For comparing RNA-Seq and qPCR results, a Pearson correlation was calculated using the Log_2 of the normalized expression values.

Transcriptome annotation

Candidate coding regions within transcript sequences were analyzed using TransDecoder (Haas et al. 2013) and protein homologues were annotated using NCBI BLASTp algorithms (e-value was set as 1e−5). The predicted genes identified from TransDecoder were used for transcriptome annotation. This procedure was performed using InterProScan (Jones et al. 2014). The protein domains, the Gene Ontology (GO) terms, enzymatic protein codes and KEGG pathways were identified. The interactions between the predicted genes associated with KEGG pathways and the GO terms were also analyzed with Cytoscape 3.3 (Shannon et al. 2003) using BinGO and EnrichmentMap plugins (Maere et al. 2005; Merico et al. 2010). BinGO was used for the identification of the GOs overrepresented over the set of genes up- or down-regulated, and an enrichment map analysis was performed, using those results, with the Enrichment Map plugin. Gene Ontology (GO) analysis was performed by running CateGOrizer (Hu et al. 2008) queries against the plant GOSlim database, which provided information related to three ontologies, including biological process (BP), cellular component (CC) and molecular function (MF).

Results

Sequencing and de novo assembly of *Q. suber* stems transcriptome

The sequencing of cDNA from two non-normalized and two normalized libraries comprising samples (Pool 1 and Pool 2) from two developmental stages and stems tissues of *Quercus cerris* × *suber* hybrids resulted in a total of 2,377,493 raw reads ranging from 48 to 1201 bp. Before assembling, reads undergone several quality control checks and after pre-processing 1,961,325 high-quality sequences were retained and used in the assembly and mapping steps. The final average length of the reads ranged between 350–340 and 389–352 bp for the normalized and non-normalized libraries, respectively. A summary of the metrics for the generated sequencing data is presented in Table 1. The transcriptome assembly produced by MIRA yielded 184,781 contigs, of which 183,345 had length larger than 200 bp. The mean contig size was 854 bp.

Read mapping

The mapping stats for the reads of each non-normalized library is presented in Table 2. BWA-mem was selected due its suitability for alignment and mapping long and single-end reads. A total of 971,769 were mapped, which represented 96.4% of the total. The number of UMRs was 723,348, which accounted to 71.7% of the total.

Differential expression and variant calling

A total of 378 genes were differentially expressed (FDR value ≤ 0.05), of which 291 genes were down-regulated (highly expressed in trees with cork production) and 87 were up-regulated (highly expressed in trees without cork production). The list containing all the genes for which evidence of differential expression was identified is indicated in Electronic Supplementary File 1. Additionally, the genes included in the discussion, and/or involved in one of the pathways highlighted in this study, are represented in Table 3.

Table 1 Sequencing metrics for the libraries of cork oak stems tissues before and after pre-processing

	Normalized cDNA libraries		Non-normalized cDNA libraries	
	Pool 1	Pool 2	Pool 1	Pool 2
Raw reads	597,215	591,877	659,901	528,500
Total number of bases	211,616,383	203,136,849	262,549,977	191,305,649
Average read length	350	340	389	352
Pre-processed reads	480,891 (80.5%)	472,126 (79.8%)	577,254 (87.5%)	431,054 (81.6%)

Table 2 Number of mapped reads, unique mapped reads and their percentages for each non-normalized library

Sample	Total reads	Mapped reads (%)	UMRs (%)
Non-normalized (Pool 1)	577,254	556,848 (96.5%)	397,893 (68.9%)
Non-normalized (Pool 2)	431,054	414,921 (96.3%)	325,455 (75.5%)
Total	1,008,308	971,769 (96.4%)	723,348 (71.7%)

As a result of the variant calling analysis, the total number of putative variants identified by the GATK-Unified Genotyper was 27,097. However, after applying the filtering strategies, the number of SNPs identified decreased to 1610 and 119, using the moderate and high stringency strategies, respectively. The SNPs identified were further inspected, in order to confirm if some of the SNPs were located in genes that had also displayed differential expression. Using the moderate stringency SNP set, a total of 19 DE genes were observed containing at least one SNP. Regarding the putative impact of each SNP, a total of 12, 8 and 9 SNPs presented low, moderate and modifier impacts, respectively.

qPCR validation

In order to validate the results from the RNAseq data, 11 DE genes were selected to assess their expression by RT-qPCR. The RT-qPCR assay for the selected transcripts displayed a similar expression pattern to the one obtained by RNAseq data analysis. The expression pattern of the transcripts tested was maintained, although some variation was detected for a few transcripts, like OPR3, protein exordium-like 2 and ethylene responsive TF 2 (Fig. 1).

A correlation between the Log_2 of the expression values of both RNAseq and qPCR data was performed and showed a strong positive correlation ($r = 0.78$) between the values obtained by these two different approaches (Fig. S1), which supports the fact that the RNAseq assay was accurate and allows the use of the data for further gene expression analysis.

Functional annotation

The total number of contigs that resulted from the de novo transcriptome assembly step was 184,781. Transdecoder found 101,869 contigs that contained candidate coding regions. These were annotated using BLAST against the non-redundant NCBI plants, which resulted in a total of 90,667 functionally annotated genes (89%), with an e-value $\leq$ 1e−05. A total of 923 distinct species were found in the annotation, and the most frequent BLAST hit species were *Vitis vinifera, Theobroma cacao* and *Prunus persica* (Fig. S2).

The functional gene annotation was performed by InterProScan, including Gene Ontology (GO) and KEGG pathway analysis. From the 90,667 predicted genes, a total of 49,773 (55%) genes covered all three GO categories with most of the genes being associated with multiple GO terms. A total of 32,143 genes were found associated to Biological Process (BP) category, 43,674 genes to Molecular Function (MF) category and 13,534 genes to Cellular Component (CC) category.

The web tool CateGOrizer was used for each category of GOs in order to organize them in subcategories making use of the GOSlim-Plants. Such approach was applied to all GOs that were associated with genes in the transcriptome i.e. genes highly expressed in trees with cork production (DE genes down-regulated) and genes highly expressed in trees without cork production (DE genes up-regulated). As highlighted in Fig. 2, the functions of the identified DE genes cover a vast set of biological processes, molecular functions and cellular components.

The distribution of the DE genes included 78 subcategories through Biological Process, 19 subcategories through Cellular Component and 110 subcategories throughout Molecular Function. The distribution of gene ontology categories shows a general trend of a larger number of genes more highly expressed in trees with cork production, with some GOs subcategories identified exclusively in this type of trees and associated with more than one gene. For example, in the Molecular Function category, the subcategory metal ion binding was associated with 10 down-regulated genes and in the Biological

Table 3 List of genes included in the discussion section and/or involved in one pathways described above

Acession number	GeneID	Functional annotation	LogFC	Exclusive expression
CAB36910.1	FN_c1210|m.1843	HSP17	− 9.19	YES
BAC55114.1	FN_c18383|m.21770	*S*-adenosylmethionine decarboxylase	− 6.88	YES
ABG49114.1	FN_c6459|m.8754	Peroxidase	− 4.61	NO
XP_002284346.1	FN_c285|m.451	UDP-glucose 6-dehydrogenase	− 4.53	NO
BAI66465.1	FN_c1512|m.2307	Chalcone synthase (CHS)	− 2.72	NO
XP_004508035.1	FN_c3416|m.4926	Sucrose synthase	− 7.87	NO
AGW82424.1	FN_c3400|m.4908	3-Hydroxy-3-methylglutaryl coenzyme A reductase	− 6.96	YES
XP_006385389.1	FN_c916|m.1397	Beta-amylase family protein	− 3.56	NO
XP_006373519.1	FN_c3859|m.5530	Phosphogluconate dehydrogenase family protein	3.95	NO
AHC69780.1	FN_c716|m.1084	Thaumatin-like protein	− 4.27	NO
AHC69781.1	FN_c271|m.433	Chitinase	− 1.79	NO
XP_007018362.1	FN_c7907|m.10436	Sucrose proton symporter	− 7.12	YES
XP_007034442.1	FN_c417|m.644	Chalcone and stilbene synthase family protein	1.22	NO
XP_007040068.1	FN_c48|m.82	Leucoanthocyanidin dioxygenase isoform 2	− 2.15	NO
XP_007046764.1	FN_c4055|m.5806	Ferulic acid 5-hydroxylase 1	− 7.57	YES
KDO80366.1	FN_c2246|m.3351	3-Ketoacyl-CoA synthase	− 2.94	NO
XP_008220823.1	FN_c5293|m.7403	Isoflavone reductase	− 7.63	YES
XP_008223483.1	FN_c387|m.601	SAM synthase 5	− 1.55	NO
XP_008438436.1	FN_c610|m.930	Beta-amylase 1, chloroplastic-like	− 1.78	NO
XP_008448804.1	FN_c288|m.456	Ethylene	− 8.65	YES
XP_009395973.1	FN_c8879|m.11549	2,3-Bisphosphoglycerate-independent phosphoglycerate mutase-like	− 7.04	YES
XP_010036726.1	FN_c24|m.36	Fructose-bisphosphate aldolase, cytoplasmic isozyme	− 1.86	NO
XP_010090135.1	FN_c1360|m.2073	OPR3	− 8.54	YES
XP_010097930.1	FN_c3494|m.5041	*S*-methyltransferase	− 3.90	NO
XP_010106073.1	FN_c140|m.225	Allene oxide synthase	− 2.18	NO
KHN12344.1	FN_c3070|m.4451	Adenosylhomocysteinase	− 4.42	NO
XP_011471031.1	FN_c1224|m.1863	Diacylglycerol kinase 2	− 5.23	NO
XP_012074088.1	FN_c726|m.1098	Xyloglucan endotransglucosylase	− 7.57	YES
NP_001292934.1	FN_c30|m.49	Phospholipase D alpha 1	− 1.17	NO
XP_012475021.1	FN_c185|m.294	SAM synthase 2	− 3.47	NO
AKT71851.1	FN_c492|m.743	Chalcone isomerase	2.77	NO
KVI10094.1	FN_c6432|m.8723	Enolase	− 7.04	YES
XP_015577664.1	FN_c3123|m.4519	Glucan endo-1,3-beta-glucosidase 14	− 2.94	NO
XP_015584653.1	FN_c15378|m.18572	Endoglucanase	4.20	NO

The genes highly expressed only in trees with cork formation are highlighted in light grey

Process, the cellular glucan metabolic process, lipid metabolic process and cell wall biogenesis subcategories were associated with 5 down-regulated genes. The group of trees without cork production also contained specific ontology subcategories, such as the molecular function, structural constituent of cytoskeleton and the biological process, nucleosome assembly. The number of genes identified for each subcategory varied greatly, from a single gene (actin cytoskeleton and ln80 complex, cellular component) to 38 genes (ATP binding, molecular function). A comparison of the GOs associated with the up- and down-regulated genes revealed the presence of some GOs that were exclusively represented over the set of down-regulated genes (highly expressed in trees with cork production) such as metal ion binding, cell wall, apoplast, lipid metabolic process, cellular glucan metabolic process and cell wall biogenesis (Table 4).

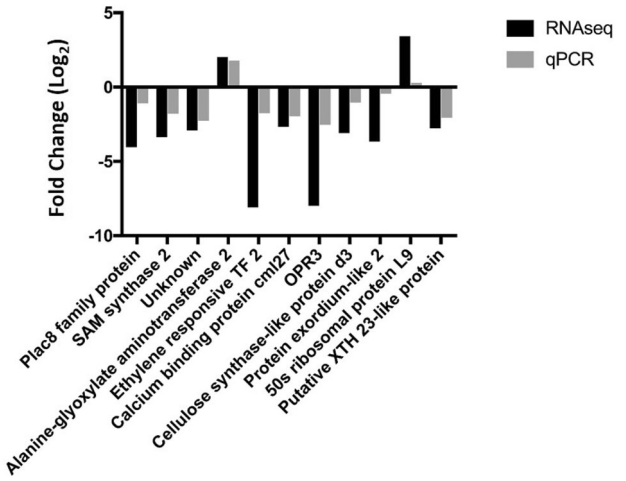

Fig. 1 Fold change for each of the transcripts selected for validation. The values are represented as the Log_2 of the normalized expression value of both RNAseq and qPCR data

Pathway analysis during cork production

Cytoscape was used for the analysis of KEGG pathways in order to identify the significant pathways associated with the DE genes. There were 11 up-regulated and 41 down-regulated pathways (Fig. 3). The most represented pathways for trees producing cork were cysteine and methionine metabolism, starch and sucrose metabolism, and gluconeogenesis. Regarding the trees that do not produce cork, the pathways containing the higher number of genes was flavonoid biosynthesis. This analysis showed that some down-regulated genes were related to terpenoid backbone biosynthesis and methane metabolism, two important biosynthetic mechanisms related to cork production.

For a more detailed analysis, two specific Cytoscape plugins, BinGO and Enrichment Map, were used for the visualization and interpretation of the gene ontology. From all the GO datasets, only the down-regulated genes associated to molecular function and cellular component GO categories showed an enrichment map of GOs (Figs. 4, 5), which included kinase activity, nucleotide binding and transcription factor activity, for the molecular function subcategory, and cell wall, extracellular region and external encapsulating structure, for the cellular component subcategory.

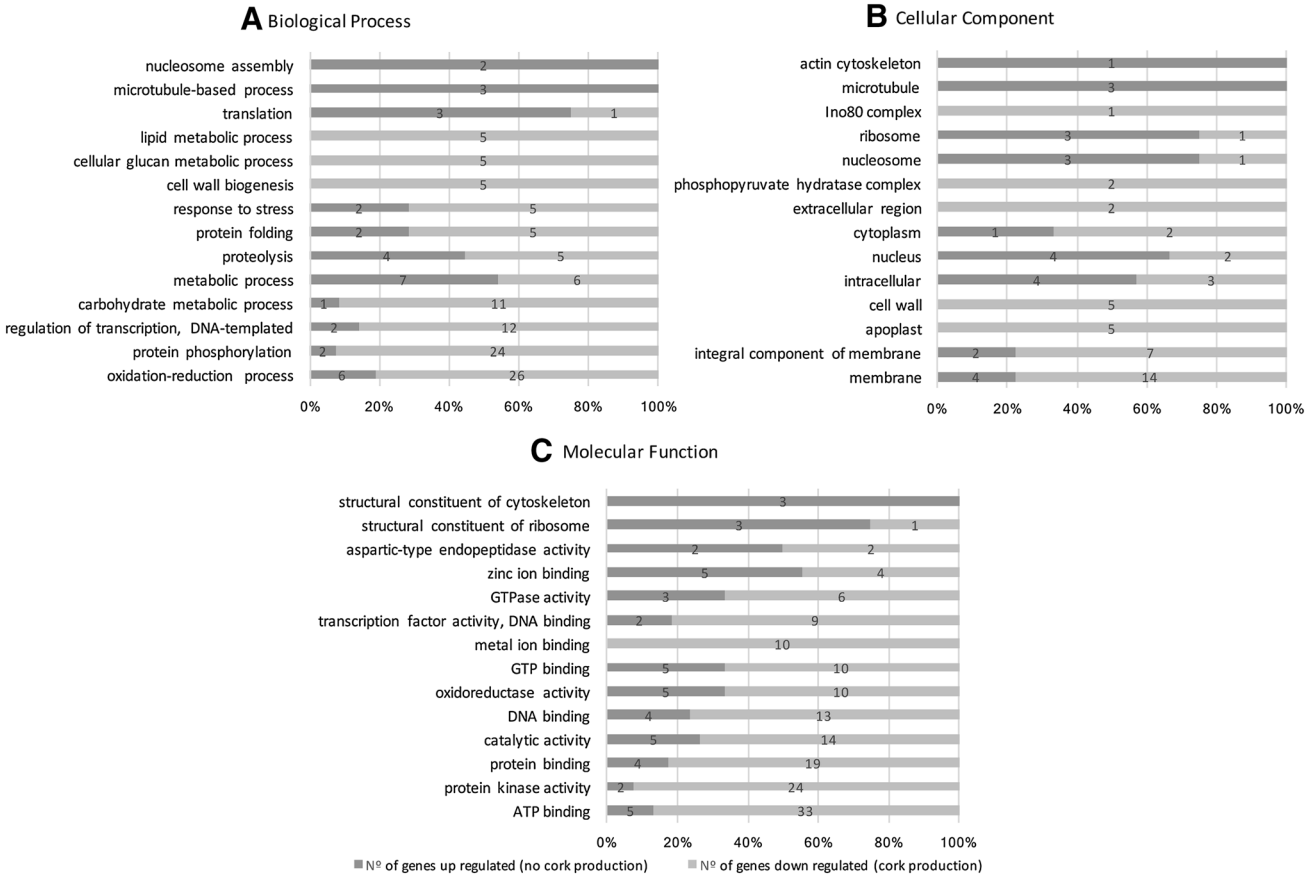

Fig. 2 Distribution of the DE genes throughout the gene ontology categories. Percentages of annotation for biological process (**a**), cellular component (**b**) and molecular function (**c**) are separated in dark grey for hybrids trees without cork production and light grey for hybrids trees with cork production. The numbers in the graphs correspond to the number of DE genes in each sub-category

Table 4 Gene ontologies by categories and subcategories exclusive to genes down-regulated (negative logFC) or genes up-regulated (positive logFC)

Acession number	logFC	Annotation	Functional category
Biological process			
ABB72441.1	− 1.89	Xyloglucan endotransglucosylase	Cell wall biogenesis cellular/glucan metabolic process
KHF98175.1	− 2.45	Putative XTH 23-like protein	Cell wall biogenesis cellular/glucan metabolic process
XP_012440766.1	− 4.42	XTH 22-like protein	Cell wall biogenesis cellular/glucan metabolic process
ACX55829.1	− 7.51	Aspartic proteinase 1	Lipid metabolic process
AHJ79158.1	− 2.78	Plastid fatty acid desaturase	Lipid metabolic process
XP_002273774.1	− 6.88	Omega-3 fatty acid desaturase, chloroplastic	Lipid metabolic process
XP_009346803.1	− 6.78	Glycerophosphodiester phosphodiesterase gde1-like	Lipid metabolic process
XP_009364630.1	− 7.19	Probable glycerophosphoryl diester phosphodiesterase 2	Lipid metabolic process
ACN40426.1	8.68	Unknown	Microtubule-based process
XP_006433708.1	8.96	Hypothetical protein CICLE_v100015522mg, partial	Microtubule-based process
AAN37904.1	2.55	Histone H1D	Nucleosome assembly
XP_006383131.1	7.07	Histone H1-3 family protein	Nucleosome assembly
Molecular function			
ABP04052.1	− 7.04	2,3-Bisphosphoglycerate-independent phosphoglycerate mutase-like	Metal ion binding
AAZ79662.1	− 4.03	Putative lipoxygenase	Metal ion binding
XP_006369132.1	− 3.36	Hypothetical protein POPTR_0001s16780g	Metal ion binding
XP_007040346.1	− 2.76	Heavy metal transport/detoxification superfamily protein isoform 1	Metal ion binding
XP_007220214.1	− 3.67	Hypothetical protein PRUPE_ppa002415mg	Metal ion binding
XP_008344802.1	− 3.68	Nardilysin-like	Metal ion binding
XP_010033729.1	− 7.19	Linoleate 13S-lipoxygenase 3-1, chloroplastic	Metal ion binding
XP_011040932.1	− 3.39	Uncharacterized protein LOC105137053	Metal ion binding
XP_015574273.1	− 1.73	Putative lactoylglutathione lyase	Metal ion binding
ACN40426.1	8.68	Unknown	Structural constituent of cytoskeleton
XP_006433708.1	8.96	Hypothetical protein CICLE_v100015522mg, partial	Structural constituent of cytoskeleton
Cellular component			
KCW50259.1	1.93	Hypothetical protein EUGRSUZ_J00048	Actin cytoskeleton
ABB72441.1	− 1.89	Xyloglucan endotransglucosylase	Apoplast/cell wall
KHF98175.1	− 2.45	Putative XTH 23-like protein	Apoplast/cell wall
XP_012440766.1	− 4.42	XTH 22-like protein	Apoplast/cell wall
CAE12163.1	− 5.09	Expansin-like protein	Extracellular region
XP_010484367.1	− 6.96	Actin-related protein 4	Ino80 complex
ACN40426.1	8.68	Unknown	Microtubule
XP_006433708.1	8.96	Hypothetical protein CICLE_v100015522mg, partial	Microtubule
KVI10094.1	− 7.04	Enolase	Phosphopyruvate hydratase complex

The interaction network analysis performed with Cytoscape 3.3 also revealed the presence of three KEGG pathways that were significantly regulated over the set of DE genes. The acyl-lipid biosynthesis pathway (Fig. 6) contained four distinct pathways points associated with three DE genes. This is one of the main pathways involved in the synthesis of the chemical components required for cork cell wall synthesis and deposition (Soler et al. 2007).

Another crucial metabolic pathway for the production of the components needed for cork synthesis is the flavonoid biosynthesis pathway (Fig. 7), for which the DE genes

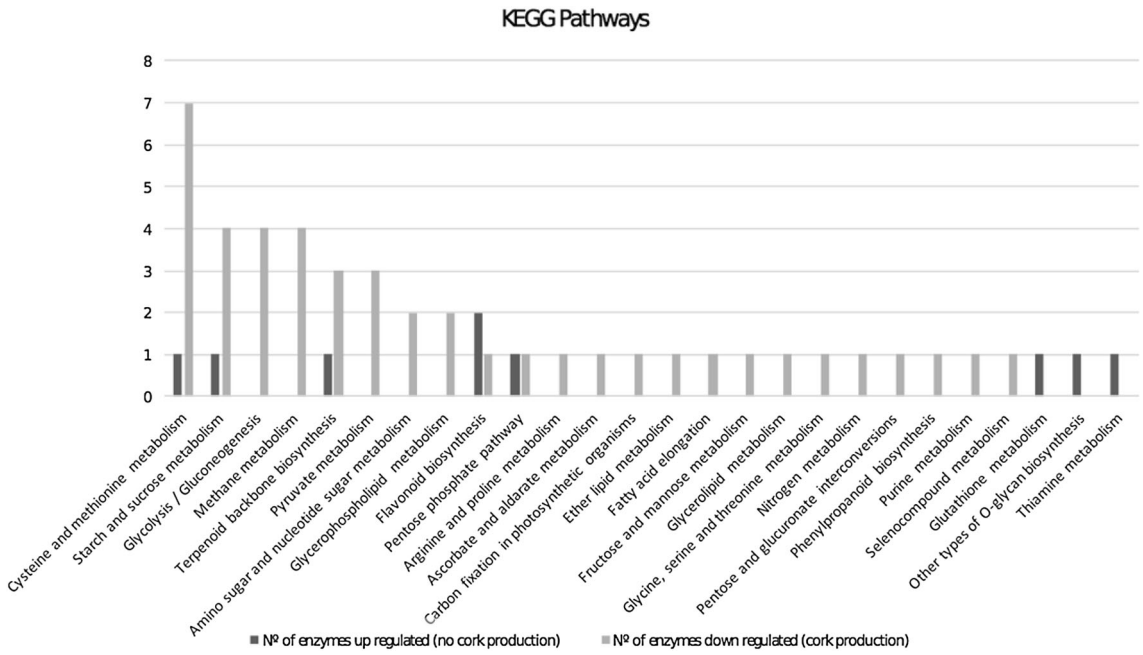

Fig. 3 The most representative KEGG pathways associated with the DE genes

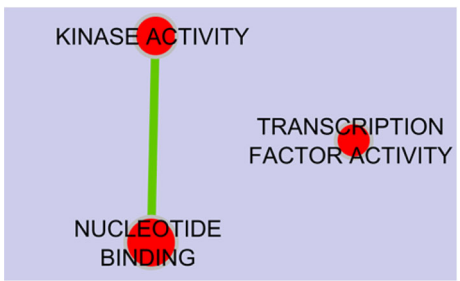

Fig. 4 Molecular function enrichment map for the down-regulated genes

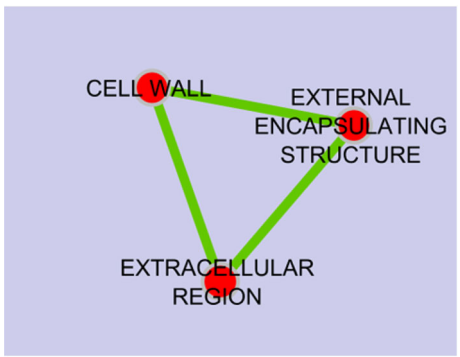

Fig. 5 Cellular component enrichment map for the down-regulated genes

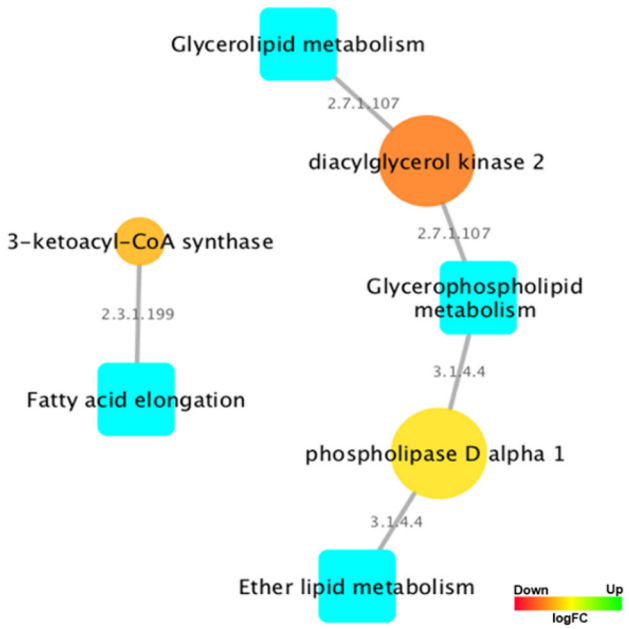

Fig. 6 Acyl-lipids biosynthesis pathway with interactions related to the DE genes. The blue square nodes represent the pathways. The rest of the nodes represent all DE genes associated to the pathways. Note that due to the complexity of representing the expression values of those genes, the RGB color scale goes from the most down-regulated in red to the most up-regulated in green (color figure online)

were related with chalcone production, which plays a key role in the formation of the flavonoid structure.

The terpenoid backbone biosynthesis pathway (Fig. 8) is involved in the synthesis of terpenoids, which represent a large class of organic chemicals made of isoprene units that in plants, are mostly related with aromatic compounds. This pathway is represented by three enzymes, two down-regulated, 3-hydroxy-3-methylglutaryl-coenzyme A reductase 1-like and 3-hydroxy-3-methylglutaryl-coenzyme A reductase and one up-regulated, 1-deoxyxylulose-

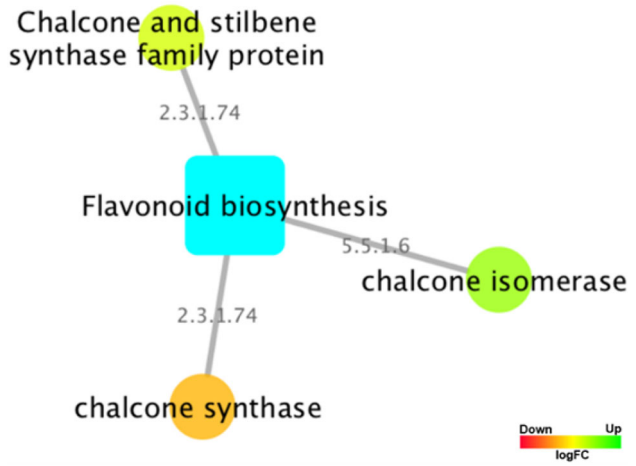

Fig. 7 Flavonoid biosynthesis pathway with interactions related to the DE genes. The blue square nodes represent the pathways. The rest of the nodes represent all DE genes associated to the pathways. Note that due to the complexity of representing the expression values of those genes, the RGB color scale goes from the most down-regulated in red to the most up-regulated in green (color figure online)

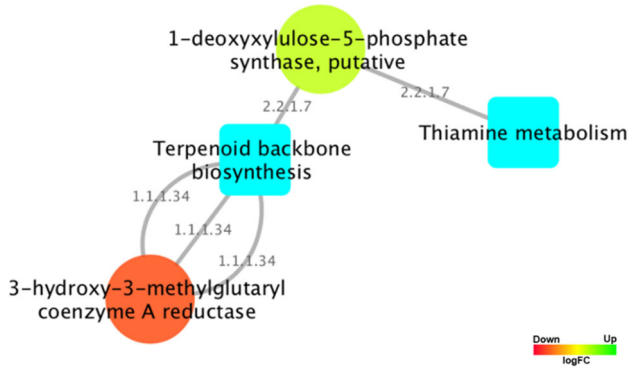

Fig. 8 Terpenoid backbone biosynthesis pathway with interactions related to the DE genes. The blue square nodes represent the pathways. The rest of the nodes represent all DE genes associated to the pathways. Note that due to the complexity of representing the expression values of those genes, the RGB color scale goes from the most down-regulated in red to the most up-regulated in green (color figure online)

5-phosphate synthase and is connected with the thiamine metabolism pathway.

Moreover, pathways such as cysteine and methionine metabolism (Fig. S3), starch and sucrose metabolism (Fig. S4) and methane metabolism pathways (Fig. S5) displayed a substantial number of interactions.

Discussion

This work was conducted to identify candidate genes involved in the production/formation of cork in *Quercus cerris* × *suber* hybrid trees. In order to obtain samples that allowed such comparison, contrasting cork production phenotypes (presence/absence of cork tissue) were used.

The amount of information regarding the genomic resources available for the *Quercus* genus, including *Quercus suber,* is scarce, a scenario that is also common for other species of the *Quercus* genus. Consequently, few proteins are annotated and available in public databases, and is the reason for the low number of homologies to *Quercus* species, as demonstrated by the lack of hits to *Quercus* sequences in the BLAST results. In addition, it will be very difficult to annotate cork oak specific genes displaying little to no homology to any gene/transcript included in the commonly used databases. Hence, a broader characterization of the cork oak transcriptome, in particular of the genes involved in cork formation is needed and studies are already emerging towards that goal (Pereira-Leal et al. 2014).

A KEGG pathway analysis focused on the KEGG Orthology (KO) revealed that the most represented KOs were the nucleotide metabolism, carbohydrate metabolism, amino acid metabolism and energy metabolism, suggesting that the transcriptome of the *Quercus cerris* × *suber* hybrid trees was programmed for the production of carbohydrates, which are one of the major forms of energy and the fundamental metabolites involved in the production of the cell-wall biomass.

Out of all the annotated transcripts, three different proteins, heat shock protein 17 (HSP17), thaumatin-like protein and chitinase were annotated with homology to *Quercus suber*. Genes codifying the HSP17, important for the acquisition of thermotolerance (Wang et al. 2004; Huang and Xu 2008), were identified to be highly expressed in trees with cork production and in trees without cork production. Previous studies demonstrated that under heat stress conditions, cork oak trees increased the translation of HSPs (Ghouil et al. 2003; Correia et al. 2013). Due to the fact that both types of trees grow under hot summers, one can speculate that the high expression of the HSPs might be a response towards the hot weather abiotic stress (Pla et al. 1998). The identification of genes more highly expressed in trees with cork production than in trees without cork production is expectable, considering that cork development confers heat tolerance (Correia et al. 2014) amongst other stress induced conditions as a protective tissue function.

Cork production may be related with defensive mechanisms involving reprogramming of primary and secondary metabolisms

The genes codifying for the thaumatin-like protein and chitinase, were highly expressed in trees with cork formation, being both associated with defense signaling

pathways. The thaumatin-like protein is related with the defense by displaying antimicrobial (Ebadzad and Cravador 2014) and antifugical activity, inhibiting the growth of pathogens (Wang and Ng 2002). The enzyme chitinase was also identified in the synthesis of suberin, a relevant metabolic pathway involved in the formation of cork cells (Pereira-Leal et al. 2014), which is in accordance with the cork-producing phenotype displayed by this group of trees. Genes associated directly or indirectly with the production of suberin are important for a better understanding of some cork properties (Soler et al. 2007).These results are indicative that the genes expression pattern that is related with the tree defense mechanisms is exclusively detected in the trees with cork production.

In this study, other genes codifying peroxidase, that were previously associated with the production of suberin and defense mechanisms (Ricardo et al. 2011), were highly expressed in trees with cork formation. The products from the reactions catalyzed by the peroxidase in the phenylpropanoid biosynthesis are necessary to initiate the lignin biosynthesis pathway that leads to the synthesis of the suberin compounds, Another very important gene for the synthesis of suberin is the 3-ketoacyl-CoA synthase gene (Teixeira et al. 2014). It was found highly expressed in trees with cork production, and its importance in the production of suberin derives from its involvement in the elongation of suberin precursors and in the synthesis of suberin monomers (Soler et al. 2007). Previously, this gene was found to be differentially expressed between good and bad quality cork samples, with a pattern of higher expression in good quality cork (Teixeira et al. 2014).

The cork chemical composition studies conducted so far, have shown that four main secondary metabolic pathways are involved to produce the different components necessary for cork development: acyl-lipids biosynthesis, flavonoid biosynthesis, phenylpropanoid biosynthesis and terpenoid backbone biosynthesis (Pereira 1988; Silva et al. 2005; Koes et al. 2005; Soler et al. 2007). In this study, genes involved in the acyl-lipids and flavonoid pathways were significantly up- regulated, which would be expected considering the group of cork-producing trees analyzed.

On the other hand, the chalcone synthase (CHS) is a fundamental enzyme of the flavonoid biosynthesis including flavonols, anthocyanins and tannins (Baxter and Stewart 2013). It plays an important role as a mechanism of defense under stress conditions, such as UV light, bacterial or fungal infections (Dao et al. 2011), and it is a possible response to heat shock (Coberly and Rausher 2003). The fact that the CHS gene was highly expressed in trees producing cork, likely indicates the higher metabolic requirements regarding the availability of the key components of cork chemical composition.

Interestingly, among the several SNPs identified in DE genes, one was identified in the leucoanthocyanidin dioxygenase isoform 2 gene. This enzyme is expressed in the trees with cork production and catalyzes the conversion of leucoanthocyanidins in anthocyanidins in the flavonoids biosynthesis (Gollop et al. 2001), an important via in the production of cork.

The importance of the phenylpropanoid biosynthesis during the cork formation of cork can be stressed by observing that genes codifying for the isoflavone reductase and for ferulic acid 5-hydroxylase 1 were exclusive highly expressed in trees capable of producing cork. The ferulic acid 5-hydroxylase, characterized as a cytochrome P-450-dependent, is involved in the synthesis of lignin precursor monomers, and it is an important player in the regulation of lignification (Grand 1984). The ferulic acid is an important component involved in rigidity and strength in increasing of suberized cell walls. Despite knowing that it is linked to glycans and serves as an initiation site for lignification, its role is still poorly understood (Serra et al. 2010). In this study, genes codifying for *S*-adenosylmethionine synthase were identified to be highly expressed in trees with cork production. This enzyme was also identified in a previous work (Teixeira et al. 2014) and it is responsible for catalyzing *S*-adenosyl-L-methionine, (Ravanel et al. 2004) needed for ethylene and polyamine biosynthesis (Teixeira et al. 2014). The polyamines have relevant roles in plant growth, development, senescence and adversity stress tolerance. A key enzyme for the synthesis of polyamines, *S*-adenosylmethionine decarboxylase, was codified by a gene found to be highly expressed only in trees with cork formation. The expression of this enzyme suggests that the synthesis of polyamines, which confers tolerance to abiotic stresses (Alcázar et al. 2006), was possibly needed for cork formation.

Several compounds and metabolic pathways that were proved to be related with cork production were present in the transcriptome but not in the DE genes. A possible cause is combination of the hybridization of *Quercus suber* and *Quercus cerris* and the need of these pathways for the plant development. As an example, we can highlight the biosynthesis of cutin, suberin and wax, which are crucial for cork formation. Significant enzymes in these pathways involved in cork formation, such as diacylglycerol O-acyltranferase, long-chain-alcohol O-fatty-acytransferase and plant seed peroxygenase were identified in the transcriptome, but did not show a pattern of differential expression.

Sugar metabolism as possible relevant metabolic super pathway for cork development

The GO annotation analysis showed that most of the genes highly expressed in trees with cork production were associated with the methane metabolism pathway, important in

the carbohydrate degradation. Most of the identified genes are also involved in starch and sucrose metabolism. The other pathways identified in this work that also displayed a significant number of interactions with DE genes were terpenoid backbone biosynthesis and cysteine and methionine metabolism.

Some of these genes, that were only observed highly expressed in trees with cork production, were identified in these pathways, indicating that they are important for cork formation. For example, the methane metabolism, glycolysis/gluconeogenesis, and the glycine, serine and threonine metabolism were connected by two enzymes codified by genes expressed only in trees with cork production, namely the 2,3-bisphosphoglycerate-independent phosphoglycerate mutase-like and enolase genes. The role of these proteins is still unclear and further studies are required for a better understanding of their involvement in the formation of cork.

As mentioned above, both the starch and sucrose metabolism can be relevant for the formation of cork. A gene codifying for a sucrose proton symporter was identified to be expressed only in trees with cork formation and is necessary for an efficient transport of sucrose (Srivastava et al. 2009). Genes codifying for sucrose synthase were also identified to be highly expressed in trees with cork production. The sucrose synthase cleaves the sucrose, which can be converted into secondary compounds, enabling plants to combat pathogens or other environmental challenges (Sturm 1999). This enzyme is putatively involved in the synthesis of UDP-glucose (described below) and ADP-glucose that are associated to cellulose and starch biosynthesis (Miguel et al. 2015).

Trees producing cork express different genes involved in cell wall formation and modification with putative involvement of jasmonic acid and ethylene

The GO analysis of the DE genes between the two conditions, revealed some differences in the molecular function and cellular component categories. Both categories had an enrichment map for the condition with trees that produce cork, which demonstrates that the major differences found in molecular function are nucleotide binding, kinase activity and transcription factor activity, while for cellular component the main differences are related with cell wall, external encapsulation structure and extracellular region.

Several DE genes codifying for proteins related with cell wall construction were identified in trees without and with cork production, such as the endoglucanase and the UDP-glucose 6-dehydrogenase genes, respectively. They are involved in cell wall assembly and formation. The UDP-glucose 6-dehydrogenase gene is also involved in the biosynthesis of UDP-glucuronic acid (UDP-GlcA), providing nucleotide sugars for cell-wall polymers (Mizrachi et al. 2012). Despite the high expression of genes related with cell wall formation and modification in both set of trees, it is important to note that the Cellular Component subcategory cell wall was associated only with down-regulated genes (Fig. 3), most of them codifying for a xyloglucan endotransglucosylase (XEH). This enzyme is responsible for the cleavage and binding of xyloglucan polymers, indispensable constituents of the primary cell wall, and participates in cell wall construction of growing tissues (Park and Cosgrove 2015).

In our study, some genes related with jasmonic acid synthesis were highly expressed in trees with cork production. The allene oxide synthase and the 12-oxophytodienoate reductase 3 were associated with the jasmonic acid (JA) synthesis which in turn is associated with the response to cell wall damage (CWD), with the production of lignin as compensatory response mechanism. The interaction of jasmonic acid and reactive oxygen species regulate the lignin biosynthesis in response to CWD (Denness et al. 2011). In this study, the pattern of expression of genes codifying for these proteins could be related with cork formation, and not with CWD.

Genes associated directly or indirectly with the production of suberin are important for a better understanding of cork properties (Soler et al. 2007). In this study, novel and possibly very important proteins for the formation of cork were identified. Until now some proteins identified in this study, such as the ferulic acid 5-hydroxylase 1, xyloglucan endotransglucosylase and sucrose proton symporter had not been directly associated with cork formation mechanisms. A pattern of expression of genes related with the plant defense mechanisms exclusively expressed in the trees with cork production were also detected, implying a connection between the plant's defense and the specific processes for cork formation. For example, genes codifying for specific ethylene-related proteins and kinase proteins were identified to be expressed only in trees with cork production. These proteins influence plant growth and development and can also induce defense-related genes (Teixeira et al. 2014).

Conclusions

The first detailed comparison, at the transcriptome level, between cork producing and non-cork producing trees, was performed and described in this study. It revealed several mechanisms exclusively associated with cork production and fundamental enzymes, such as CHS, that take part in pathways that are already associated with cork. The

substantial number of genes and pathways related with this phenotype suggest that the cork layer development is a very complex and involve unique molecular process. Most of the genes highly expressed identified in this work were associated with the defense mechanism, which has led to the hypothesis that these genes have dual functions, being also involved in the unique processes needed for the cork formation. Several genes highly expressed that participate in the primary cell wall formation and modification were identified. Most of these genes were highly expressed or only highly expressed in trees with cork production, indicating that the cell wall construction was a crucial process for the formation of cork.

DE genes that displayed a pattern of expression only in the cork producing trees that participate in the methane metabolism or in the starch and sucrose metabolism were identified indicating that reprogramming of carbohydrate metabolism is required for cork production.

Acknowledgements This study was funded by Fundação para a Ciência e a Tecnologia (FCT) projects Cork Oak EST Consortium SOBREIRO/0017/2009 and UID/AGR/00115/2013. Financial support for AMR, AU and PB was provided by Investigador FCT project IF/00574/2012/CP1209/CT0001: "Genetic characterization of national animal and plant resources using next-generation sequencing". Financial support for AMF was provided by projects PEst-OE/BIA/UI4046/2011 and FCT Investigator IF/00169/2015.

Author contribution This study was conceived by IC, RC, AMF, RT, SG and FN (coordinator). Collection and identification of field material was performed by IC, RC and FN. Sample preparation and nucleic acid isolation were performed by RC and FN. qPCR validations were executed by AF and SG. Bioinformatics data analyses were conducted by BM, AU, PB, IC and AMR. Biological interpretation of the results was conducted by BM, AU, AMR, AMF, IC and FN. The manuscript was written by BM, AU, FN and AMR. All authors read and approved the final manuscript.

Data availability Sequence reads were deposited in the NCBI Sequence Read Archive (SRA) under the accession numbers ERX143070 and ERX143071, for the normalized libraries, and SRX2677031 and SRX2677030, for the non-normalized libraries.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest.

References

Alcázar R, Marco F, Cuevas JC et al (2006) Involvement of polyamines in plant response to abiotic stress. Biotechnol Lett 28:1867–1876. https://doi.org/10.1007/s10529-006-9179-3

Baxter HL, Stewart CN (2013) Effects of altered lignin biosynthesis on phenylpropanoid metabolism and plant stress. Biofuels 4:635–650. https://doi.org/10.4155/bfs.13.56

Bugalho MN, Caldeira MC, Pereira JS et al (2011) Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. Front Ecol Environ 9:278–286. https://doi.org/10.1890/100084

Caritat A, Gutiérrez E, Molinas M (2000) Influence of weather on cork-ring width. Tree Physiol 20:893–900

Chevreux B, Pfisterer T, Drescher B et al (2004) Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Res 14:1147–1159

Coberly LC, Rausher MD (2003) Analysis of a chalcone synthase mutant in Ipomoea purpurea reveals a novel function for flavonoids: amelioration of heat stress. Mol Ecol 12:1113–1124

Correia B, Valledor L, Meijón M et al (2013) Is the interplay between epigenetic markers related to the acclimation of cork oak plants to high temperatures ? PLoS ONE 8:e53543. https://doi.org/10.1371/journal.pone.0053543

Correia B, Rodriguez JL, Valledor L et al (2014) Analysis of the expression of putative heat-stress related genes in relation to thermotolerance of cork oak. J Plant Physiol 171:399–406. https://doi.org/10.1016/j.jplph.2013.12.004

Dao TTH, Linthorst HJM, Verpoorte R (2011) Chalcone synthase and its functions in plant resistance. Phytochem Rev 10:397–412. https://doi.org/10.1007/s11101-011-9211-7

Denness L, McKenna JF, Segonzac C et al (2011) Cell wall damage-induced lignin biosynthesis is regulated by a reactive oxygen species- and jasmonic acid-dependent process in *Arabidopsis*. Plant Physiol 156:1364–1374. https://doi.org/10.1104/pp.111.175737

Ebadzad G, Cravador A (2014) Quantitative RT-PCR analysis of differentially expressed genes in *Quercus suber* in response to *Phytophthora cinnamomi* infection. SpringerPlus 3:613

Ghouil H, Montpied P, Epron D et al (2003) Thermal optima of photosynthetic functions and thermostability of photochemistry in cork oak seedlings. Tree Physiol 23:1031–1040

Gollop R, Farhi S, Perl A (2001) Regulation of the leucoanthocyanidin dioxygenase gene expression in *Vitis vinifera*. Plant Sci 161:579–588

Graça J (2015) Suberin: the biopolyester at the frontier of plants. Front Chem. https://doi.org/10.3389/fchem.2015.00062

Graça J, Pereira H (2004) The periderm development in *Quercus suber*. IAWA J 25:325–335. https://doi.org/10.1163/22941932-90000369

Graça J, Santos S (2007) Suberin: a biopolyester of plants' skin. Macromol Biosci 7:128–135. https://doi.org/10.1002/mabi.200600218

Grand C (1984) Ferulic acid 5-hydroxylase: a new cytochrome P-450-dependent enzyme from higher plant microsomes involved in lignin synthesis. FEBS Lett 169:7–11

Haas BJ, Papanicolaou A, Yassour M et al (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8:1494–1512. https://doi.org/10.1038/nprot.2013.084

Hu Z-L, Bao J, Reecy JM (2008) CateGOrizer: a web-based program to batch analyze gene on-tology classification categories. Online J Bioinform 9:108–112

Huang B, Xu C (2008) Identification and characterization of proteins associated with plant tolerance to heat stress. J Integr Plant Biol 50:1230–1237. https://doi.org/10.1111/j.1744-7909.2008.00735.x

Jones P, Binns D, Chang H-Y et al (2014) InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Koes R, Verweij W, Quattrocchio F (2005) Flavonoids: a colorful model for the regulation and evolution of biochemical pathways. Trends Plant Sci 10:236–242. https://doi.org/10.1016/j.tplants.2005.03.002

Kolattukudy PE (1981) Structure, biosynthesis, and biodegradation of cutin and suberin. Annu Rev Plant Physiol 32:539–567. https://doi.org/10.1146/annurev.pp.32.060181.002543

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352

Lourenço A, Rencoret J, Chemetova C et al (2016) Lignin composition and structure differs between xylem, phloem and phellem in *Quercus suber* L. Front Plant Sci 10:10. https://doi.org/10.3389/fpls.2016.01612

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

Marques AV, Pereira H (2013) Lignin monomeric composition of corks from the barks of *Betula pendula*, *Quercus suber* and *Quercus cerris* determined by Py–GC–MS/FID. J Anal Appl Pyrolysis 100:88–94. https://doi.org/10.1016/j.jaap.2012.12.001

Marum L, Miguel A, Ricardo CP, Miguel C (2012) Reference gene selection for quantitative real-time PCR normalization in *Quercus suber*. PLoS ONE 7:e35113. https://doi.org/10.1371/journal.pone.0035113

McKenna A, Hanna M, Banks E et al (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20:1297–1303. https://doi.org/10.1101/gr.107524.110

Merico D, Isserlin R, Stueker O et al (2010) Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. PLoS ONE 5:e13984. https://doi.org/10.1371/journal.pone.0013984

Miguel A, de Vega-Bartol J, Marum L et al (2015) Characterization of the cork oak transcriptome dynamics during acorn development. BMC Plant Biol. https://doi.org/10.1186/s12870-015-0534-1

Mizrachi E, Mansfield SD, Myburg AA (2012) Cellulose factories: advancing bioenergy production from forest trees. New Phytol 194:54–62. https://doi.org/10.1111/j.1469-8137.2011.03971.x

Park YB, Cosgrove DJ (2015) Xyloglucan and its interactions with other components of the growing cell wall. Plant Cell Physiol 56:180–194. https://doi.org/10.1093/pcp/pcu204

Pereira H (1988) Chemical composition and variability of cork from *Quercus suber*. Wood Sci Technol 22:211–218

Pereira H (2007) Cork: biology, production and uses, 1st edn. Elsevier, London

Pereira-Leal JB, Abreu IA, Alabaça CS et al (2014) A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. BMC Genom 15:371

Pfaffl M (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29:e45

Pla M, Huguet G, Verdaguer D et al (1998) Stress proteins co-expressed in suberized and lignified cells and in apical meristems. Plant Sci 139:49–57

Pollard M, Beisson F, Li Y, Ohlrogge JB (2008) Building lipid barriers: biosynthesis of cutin and suberin. Trends Plant Sci 13:236–246. https://doi.org/10.1016/j.tplants.2008.03.003

Rahantamalala A, Rech P, Martinez Y et al (2010) Research article coordinated transcriptional regulation of two key genes in the lignin branch pathway-CAD and CCR-is mediated through MYB-binding sites. BMC Plant Biol 10(1):130

Ranathunge K, Schreiber L, Franke R (2011) Suberin research in the genomics era—new interest for an old polymer. Plant Sci 180:399–413. https://doi.org/10.1016/j.plantsci.2010.11.003

Ravanel S, Block MA, Rippert P et al (2004) Methionine metabolism in plants: chloroplasts are autonomous for de novo methionine synthesis and can import *S*-adenosylmethionine from the cytosol. J Biol Chem 279:22548–22557. https://doi.org/10.1074/jbc.M313250200

Reid KE, Olsson N, Schlosser J et al (2006) An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. BMC Plant Biol 6:27

Ricardo CPP, Martins I, Francisco R et al (2011) Proteins associated with cork formation in *Quercus suber* L. stem tissues. J Proteomics 74:1266–1278. https://doi.org/10.1016/j.jprot.2011.02.003

Robinson MD, McCarthy DJ, Smyth GK (2010) edgeR: a bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140. https://doi.org/10.1093/bioinformatics/btp616

Schloss PD, Westcott SL, Ryabin T et al (2009) Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09

Serra O, Soler M, Hohn C et al (2009) CYP86A33-targeted gene silencing in potato tuber alters suberin composition, distorts suberin lamellae, and impairs the periderm's water barrier function. Plant Physiol 149:1050–1060. https://doi.org/10.1104/pp.108.127183

Serra O, Figueras M, Franke R et al (2010) Unraveling ferulate role in suberin and periderm biology by reverse genetics. Plant Signal Behav 5:953–958

Shannon P, Markiel A, Ozier O et al (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Silva SP, Sabino MA, Fernandes EM et al (2005) Cork: properties, capabilities and applications. Int Mater Rev 50:256. https://doi.org/10.1179/174328005X41168

Soler M, Serra O, Molinas M et al (2007) A genomic approach to suberin biosynthesis and cork differentiation. Plant Physiol 144:419–431. https://doi.org/10.1104/pp.106.094227

Srivastava AC, Dasgupta K, Ajieren E et al (2009) *Arabidopsis* plants harbouring a mutation in *AtSUC2*, encoding the predominant sucrose/proton symporter necessary for efficient phloem transport, are able to complete their life cycle and produce viable seed. Ann Bot 104:1121–1128. https://doi.org/10.1093/aob/mcp215

Sturm A (1999) Invertases. Primary structures, functions, and roles in plant development and sucrose partitioning. Plant Physiol 121:1–8

Teixeira RT, Fortes AM, Pinheiro C, Pereira H (2014) Comparison of good- and bad-quality cork: application of high-throughput sequencing of phellogenic tissue. J Exp Bot 65:4887–4905. https://doi.org/10.1093/jxb/eru252

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JA (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35(suppl_2):W71–W74

Verdaguer R, Soler M, Serra O et al (2016) Silencing of the potato *StNAC103* gene enhances the accumulation of suberin polyester and associated wax in tuber skin. J Exp Bot 67:5415–5427. https://doi.org/10.1093/jxb/erw305

Vishwanath SJ, Delude C, Domergue F, Rowland O (2015) Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. Plant Cell Rep 34:573–586. https://doi.org/10.1007/s00299-014-1727-z

Wang H, Ng TB (2002) Isolation of an antifungal thaumatin-like protein from kiwi fruits. Phytochemistry 61:1–6

Wang W, Vinocur B, Shoseyov O, Altman A (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. Trends Plant Sci 9:244–252. https://doi.org/10.1016/j.tplants.2004.03.006

Zhu YY, Machleder EM, Chenchik A et al (2001) Reverse transcriptase template switching: a SMART^{TM} approach for full-length cDNA library construction. Biotechniques 30:892–897

Zhulidov PA, Bogdanova EA, Shcheglov AS et al (2005) A method for the preparation of normalized cDNA libraries enriched with full-length sequences. Russ J Bioorganic Chem 31:170–177