Journal of Experimental Botany, Vol. 65, No. 17, pp. 4887–4905, 2014
doi:10.1093/jxb/eru252 Advance Access publication 22 June, 2014

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Comparison of good- and bad-quality cork: application of high-throughput sequencing of phellogenic tissue

Rita Teresa Teixeira[1,*], Ana Margarida Fortes[2], Carla Pinheiro[3,4], Helena Pereira[1]

[1] Centro de Estudos Florestais, Instituto Superior de Agronomia, Universidade Técnica de Lisboa, 1349-017, Portugal
[2] Center for Biodiversity, Functional and Integrative Genomics (BioFIG); Science Faculty, University of Lisbon, Campo Grande, 1749-016 Lisboa, Portugal
[3] Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Av. da República, EAN, 2780-157 Oeiras, Portugal
[4] Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal

* To whom correspondence should be addressed. E-mail: ritamiss@yahoo.com

Received 21 February 2014; Revised 3 April 2014; Accepted 12 May 2014

Abstract

Cork is one of the most valuable non-wood forest products and plays an important role in Mediterranean economies. The production of high-quality cork is dependent on both genome and environment, posing constraints on the industry because an ever-growing amount of bad-quality cork (BQC) development has been observed. In order to identify genes responsible for production of cork of superior quality we performed a comparative analysis using the 454 pyrosequencing approach on phellogenic tissue of good- and bad-quality samples. The transcriptional profiling showed a high number of genes differentially expressed (8.48%) from which 78.8% displayed annotation. Genes more highly represented in BQC are involved in DNA synthesis, RNA processing, proteolysis, and transcription factors related to the abiotic stress response. Putative stomatal/lenticular-associated genes which may be responsible for the disadvantageous higher number of lenticular channels in BQC are also more highly represented. BQC also showed an elevated content of free phenolics. On the other hand, good-quality cork (GQC) can be distinguished by highly expressed genes encoding heat-shock proteins. Together the results provide valuable new information about the molecular events leading to cork formation and provide putative biomarkers associated with cork quality that can be useful in breeding programmes.

Key words: Abiotic stress, cork quality, lenticel, phellogen, transcriptomics, *Quercus suber*.

Introduction

The Mediterranean region harbours the richest ecosystem in Europe: the cork oak stands, with the cork oak (*Quercus suber* L.) as the main species. Cork is a protective tissue composed of suberized cells that forms a thick layer covering stem, branches, and roots. As an interface between environment and tree, cork (phellem in plant anatomy) acts as a shield against drought, solar irradiation, pathogens, and fires, and its formation is believed to be a protective mechanism (Gibson *et al.*, 1981). Cork derives from a specific meristematic layer, the phellogen, and grows continuously and homogenously, giving rise to multiple layers of radially aligned cells that accumulate into annual cork rings. The phellogen has the property of self-regeneration after damage or the peeling-off of the cork layer, without apparent injury to the tree (Caritat *et al.*, 2000; Pereira, 2007). It is this property that allows successive cork debarkings from the same tree and sustainable cork production during the tree's life.

Cork is one of the most valuable non-wood forest products and the basis of a sustained forest exploitation system. Cork wine-stoppers are known worldwide and are the main products contributing to the cork industry. The production of wine stoppers requires cork raw-material of good quality in

the thickness and homogeneity of the phellem tissue (Pereira, 2007; Sanchez-González *et al.*, 2008). It is not surprising that a potential decrease in cork quality strongly concerns producers and industry, and better understanding of the differentiation between good- and bad-quality cork (BQC) is an ever-present quest.

Cork quality is dependent on both genome and environment (Pereira, 1988; Conde *et al.*, 1998). The molecular mechanisms involved in the response of trees to climate factors continues to be elucidated, but in *Q. suber* it has been reported that the phellogen activity is reduced under conditions of drought and elevated temperature (Graça and Pereira, 1997; Costa *et al.*, 2002). Industry demands that cork displays a thickness of at least 27 mm and presents no sclereid inclusions, signs of pathogen attack, or other typical cork defects (Pereira, 2007). Most cork oaks produce cork considered to be bad quality because it displays a reduced thickness that varies between 3 and 10 mm and possesses an elevated number of lenticular channels. The excessive presence of lenticels in cork planks is, therefore, a key factor for its quality downgrading (Pereira *et al.*, 1996). In *Q. suber*, lenticels are conspicuous on the surface of cork since they accumulate brown oxidized phenolics (Wutz, 1955). Cork is the site of four main secondary metabolic pathways, such as the biosynthesis of suberin that requires the synthesis of acyl-lipids, phenylpropanoids, isoprenoids, and flavonoids (Beisson *et al.*, 2007), cork lignin and suberin aromatic components (Dixon *et al.*, 2002; Boerjan *et al.*, 2003), terpenes (Laule *et al.*, 2003), and sterols (Benveniste, 2004).

Phenylpropanoid compounds encompass a wide range of structural classes and biological functions (Dixon and Paiva, 1995). Secondary phenols have been implicated in a wide range of plant environmental and developmental interactions, and are most important in biotic and abiotic stress responses. Phenylpropanoids play a major role as defensive compounds (Vogt, 2010).

The cork cell walls are chemically characterized by the presence of suberin as the main structural component, and associated to lignin as the second most important component (Bernards, 2002; Pereira, 2007). Suberin is a complex polymer characterized by an aliphatic network of inter-esterified long-chain fatty acids and alcohols with glycerol, and including aromatic units (e.g. ferulic acid) and other phenolic domains (Graça and Pereira, 1997; Bernards, 2002). The chemical composition of cork and its variability is now well established (Marques and Pereira, 2013), but when it comes to its molecular and genetic mechanisms, very little is known. Recently, genes important for cork biosynthesis and differentiation were identified using a suppression subtractive hybridization (SSH) library after obtaining cork expressed sequence tags (ESTs) (Soler *et al.*, 2007).

Hybridization-based microarrays have until recently been the dominant platform used routinely to analyse transcriptional changes in several species (Busch and Lohmann, 2007; Strable *et al.*, 2008; O'Rourke *et al.*, 2009). However, the new generation of transcriptomic data allows new insights in plant biology and genetics. For a species like *Q. suber*, for which there is no genome sequence or physical map available, information obtained from ESTs is a practical means of gene discovery. Transcriptome obtained by high-throughput or EST sequencing offers an efficient means of generating functional genomic data for non-model organisms (Parchman *et al.*, 2010) because it provides functional information that often corresponds to genes with known or predicted functions (Andersen and Lübberstedt, 2003). Out of several next-generation sequencing methods, Roche 454 pyrosequencing is the best adapted for analysing the transcriptome of both model and non-model species (Morozova and Marra, 2008; Barakat *et al.*, 2009). This is the sequencing method that we used in this study, producing more information about the molecular players during cork development.

The present work allowed us to generate quantitative data on transcript accumulation in phellogen, the meristem responsible for the formation of cork. The assembled and annotated sequences were produced and organized in a dedicated database (see http://transcriptomics.biocant.pt/suber_rt/) providing an extensive catalogue of genes expressed in good- and bad-quality cork, allowing for an exhaustive comparison between the two types of cork.

Materials and methods

Biological sample

Phellogenic tissue was collected from 10 mature cork oaks under cork production (estimated age 50–70 years) in southern Portugal [BQC was collected in the area of Coruche, Ribatejo and good-quality cork (GQC) at Serra do Caldeirão, Algarve]. Samples were taken at the time of cork harvest, in the period end July–beginning August 2011, by scratching the inside of the plank (the so-called cork belly, corresponding to the torn-out phellogen) immediately after the separation of the cork plank from the tree stem, and freezing in liquid nitrogen immediately after collection. Five samples each were taken from cork oaks producing GQC and BQC, respectively. A sample from each cork plank was also taken for chemical analysis.

The classification of cork quality was based on visual observation taking into account typical industry criteria: thickness and homogeneity, porosity (lenticular channels), and presence of lignous inclusions and other defects (Pereira, 2007).

In the laboratory, the phellogenic material was ground using an electric mortar, always under freezing conditions.

For chemical analysis, the cork-plank samples were allowed to air dry in well-ventilated conditions. A sub-sample with approximately $15 \times 15\,cm^2$ was cut and the cork back was removed. The cork sample was milled using a knife mill (Retsch SM 2000) with an output sieve of $10 \times 10\,mm^2$, and the granulated material screened using a vibratory sieving apparatus (Retsch AS 200 basic) using standard Tyles sieves with the following mesh sizes: 80 (0.180 mm), 60 (0.250 mm), 40 (0.425 mm), 20 (0.850 mm), 15 (1.0 mm), and 10 (2.0 mm). Fractions above 1 mm (15 mesh) were milled again and fractionated. For the chemical analysis, the 40–60-mesh granulometric fraction was used.

RNA extraction

RNA from 150 mg of ground phellogenic samples was extracted from each sample using the Spectrum Plant Total RNA kit from Sigma Aldrich supplemented with Plant RNA Isolation Aid from Ambion. Genomic DNA was eliminated through an on-column DNase treatment (on-column DNase I digestion set from Sigma-Aldrich). The RNA used for pyrosequencing was further purified using polyvinylpyrrolidone (PVP).

cDNA preparation

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). A fraction of 0.5–2.0 µg of each pool of total RNA was used as starting material for cDNA synthesis with the MINT cDNA synthesis kit (Evrogen, Moscow, Russia), where a strategy based on SMART double-stranded cDNA synthesis was applied. Here, a known adapter sequence was introduced to both ends of the first strand of cDNA during the amplification of the polyRNA molecules. The synthesis was also performed using a modified oligo-dT containing a restriction site for Bsgl needed to eliminate these tails, and so minimize interference of homopolymers during the 454 sequencing run.

Non-normalized cDNA was quantified by fluorescence and sequenced in 454 GS FLX Titanium according to the standard manufacturer's instructions (Roche-454 Life Sciences, Bradford, CT, USA) at Biocant (Cantanhede, Portugal). Sequence data were submitted to NCBI Short Read Archive (SRA) with the reference SRP030001.

Sequence processing assembly and annotation

Upon 454 sequencing, raw reads were processed to remove sequences with less than 100 nucleotides and low-quality regions. The ribosomal, mitochondrial, and chloroplast reads were also identified and removed from the data set. The reads were then assembled into contigs using 454 Newbler 2.6 (Roche) with the default parameters (40 bp overlap and 90% identity).

For gene identification, a three-step analysis was carried out. First, the translation frame of contigs was assessed through BLASTx searches against Swissprot (e-value = $1e^{-6}$), and the corresponding amino acid sequence translated using an in-house script. Next, the contigs without translation were submitted to FrameDP (Gouzy *et al.*, 2009) software. Finally, the remaining contigs were analysed with ESTScan (Lottaz *et al.*, 2003).Transcripts resulting from these two last identification steps, FrameDP and ESTScan, were searched using Blastp against the non-redundant NBCI (National Center for Biotechnology Information) database to translate the putative proteins.

The functional annotation of the deduced amino acid sequences was predicted through assignment into protein families and identification of protein domains using InterProScan (Hunter *et al.*, 2009). Gene ontology terms (Ashburner *et al.*, 2000) identified by InterProScan results for each translated amino acid sequence were additionally retrieved and added to classify the transcript products.

Differential gene expression between samples was determined by calculating the number of reads sequenced for each transcript. First, the contigs from both samples were clustered at 90% similarity by CD-Hit 454 (Beifang *et al.*, 2010) to eliminate redundant sequences and generate reference contigs. Then, the reads from each sample were mapped with 454 Newbler mapping 2.6 (Roche) to those references, and the number of reads contributed by each sample counted. The reads with multiple hits were discarded. The number of reads per reference contig per sample were used to build a contingency table, which was analysed with the Myrna (Langmead *et al.*, 2010) statistical analysis package. The contingency table was normalized at a 95 percentile. Differential gene expression was evaluated using a linear regression model based on a Gaussian distribution, and taking into account only contigs with a minimum of eight mapped reads and FDR < 0.05. All the results were compiled into a SQL database developed as an information management system.

Real-time RT-PCR

The cDNA prepared from individual samples and used for pyrosequencing was also used for real-time RT-PCR analysis on a StepOne™ Real-Time PCR System (Applied Biosystems, Foster City, CA) and in MicroAmp Optical 48-well reaction plates with optical covers, according to the manufacturer's instructions. PCR reactions (final volume of 25 µl) were performed with gene-specific primers (Supplementary Table S1) and the passive reference dye ROX in order to normalize fluorescence across the plate. Reaction conditions were: 95°C for 10 min; then 40 cycles of 95°C for 15 s; and 58°C for 45 s. Analysis used StepOne software 2.1. Relative quantification values and standard deviations were calculated using the standard curve method according to the manufacturer's instructions (Applied Biosystems, Foster City, CA; User Guide). Values were normalized to GQC samples and results analysed with Microsoft Excel Software. The housekeeping gene chosen was *Actin* because according to Marum *et al.* (2012) this gene was demonstrated to be the most stable tested in *Q. suber*.

MapMan analysis

Pathway analysis was carried out with MapMan (Usadel *et al.*, 2009). Differentially expressed contigs were assigned into functional categories (or bins) by Mercator (http://mapman.gabipd.org/web/guest/mercator). A pathway map was created to represent most of the contigs. The Wilcoxon rank sum test was used to identify differentially regulated bins.

Alignment and phylogenetic analysis

The amino acid sequences of *Arabidopsis thaliana* were retrieved from TAIR (http://www.arabidopsis.org). The sequences were BLASTed against the database of *Q. suber* proteins at the *Quercus* transcriptome database (project: http://transcriptomics.biocant.pt/suber_rt/). Alignment was performed using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/) software.

Phylogenetic analysis was conducted using the Neighbor-Joining method in MEGA version 5. Tests of inferred phylogenetic groups were conducted by bootstrapping test with 1000 replicates.

Determination of total phenolics

The total soluble polyphenol content (TPC) was determined in the solutions obtained by the extractions with ethanol and water. TPC was determined by spectrophotometry, using gallic acid as the standard, according to the Folin-Ciocalteu assay (Pereira, 1981). The calibration curve was obtained by preparing different standard concentrations of gallic acid within the range 0.1–0.6 mg l^{-1}. Briefly, a 100 µl aliquot of extracts, the gallic acid standard solutions (0.1–0.6 mg l^{-1}), and a blank (deionized water) were put in different tubes. Then, 4 ml of the Folin-Ciocalteu's phenol reagent diluted 1:10 was added to each tube, and the tubes were shaken and allowed to react for 5 min. After this time, 4 ml of 7.5% Na_2CO_3 solution was added. After incubation in a thermostatic bath for 15 min at 45°C, the absorbance against a blank was determined spectrophotometrically at 765 nm (Shimadzu UV-160A). Total phenolic content was expressed as milligrams of gallic acid equivalents (GAE) per 100 g dry mass.

PCA analysis

Principal Component Analysis (PCA) was carried out as described previously (Chaves *et al.*, 2009) using the R platform [version 2.13.1 (Team RDC, 2011)] and the ade4TkGUI package (Thioulouse and Dray, 2007; Team RDC, 2011).

PCA was used to discriminate bad- and good-quality samples and applied to a whole set of expressed genes (16 792) as well as to a selected set of genes associated with suberin synthesis (29 genes). Correlation coefficients (Pearson's product-moment correlation) were calculated using the Stars4 Package (Team RDC, 2011).

Statistical analysis

Unless otherwise mentioned all statistical analysis was carried out using ANOVA: Single Factor.

Results

Phenotypic analysis of good- and bad-quality cork

The phellogenic tissue was obtained by scratching the inner surface (cork belly) of the cork planks immediately after debarking (see Material and Methods). Samples of the same planks are illustrated in Fig. 1. The planks of cork of good (A) and bad (B) quality display the same age (nine years of growth).

The minimal plank thickness required by industry for the production of stoppers is 27 mm. The samples in Fig. 1A vary between 35 and 40 mm. Homogeneity is another important feature for industry, characterized by the presence of a minimal number of lenticular channels crossing the cork (Fig. 1A: black arrow in plank 4) and no other discontinuities. In planks of GQC it is possible clearly to distinguish the growth rings (Fig. 1A: *), which appear as white lines. Annual rings have both light and dark cork. The light-coloured cork is produced early in the season when cell production is more active and the cells are relatively large. Later, cork turns darker as cell production decelerates and cell becomes smaller (Soler *et al.*, 2008). The planks of BQC are displayed in Fig. 1B. The thickness of these planks varies between 15 and 20 mm, but most importantly they possess numerous lenticular channels as can be seen, for example, in plank 8 (black arrow). These samples also show several inclusions (Fig. 1B: plank 7).

454 sequences of phellogenic tissue of good- and bad-quality cork

For gene identification upon comparative analysis of contigs obtained after Roche 454-pyrosequencing of phellogen of good- and bad-quality cork, two non-normalized libraries were constructed (of good- and bad-quality cork) from *Q. Suber* phellogen. Each library resulted from a pool of total RNA extracted from five different trees (Fig. 1A, B). The two runs (good- and bad-quality cork) provided 536 042 and 555 075 reads, respectively. The mean read length was 337.3 for the good-quality library and 323.6 for the bad-quality one. The reads from the two cork types were assembled together into a single transcriptome, generating 18 411 transcripts, with a mean read length of 770 bp. These transcripts encoded 16 792 amino acid sequences. The cut-off of normalization of eight reads at a 95% percentile, with adjusted *P* value of 0.05 (see Material and Methods) applied, resulted in 1 425 genes differentially expressed (8.48%). Of these, 1 123 showed annotation (78.8%).

Functional annotation of good- and bad-quality cork

The function of the differentially expressed genes from the two cork qualities, using the GO annotation, covers a vast set of molecular functions (Fig. 2A), cellular components (Fig. 2B), and biological processes (Fig. 2C). The identification of a large number of differentially expressed genes provides us with many candidates for studies concerning the formation of cork of good quality.

Out of the 1 425 differentially expressed genes, the catalytic activity category comprises 305 genes for GQC (21.4%) and 250 genes for BQC (17.5%); the binding category accounts for 246 genes in GQC (17.2%) and 217 genes in BQC (15.2%) (Fig. 2A, Supplementary Table S2). The 100% of genes belonging to GQC present in the actin-binding and

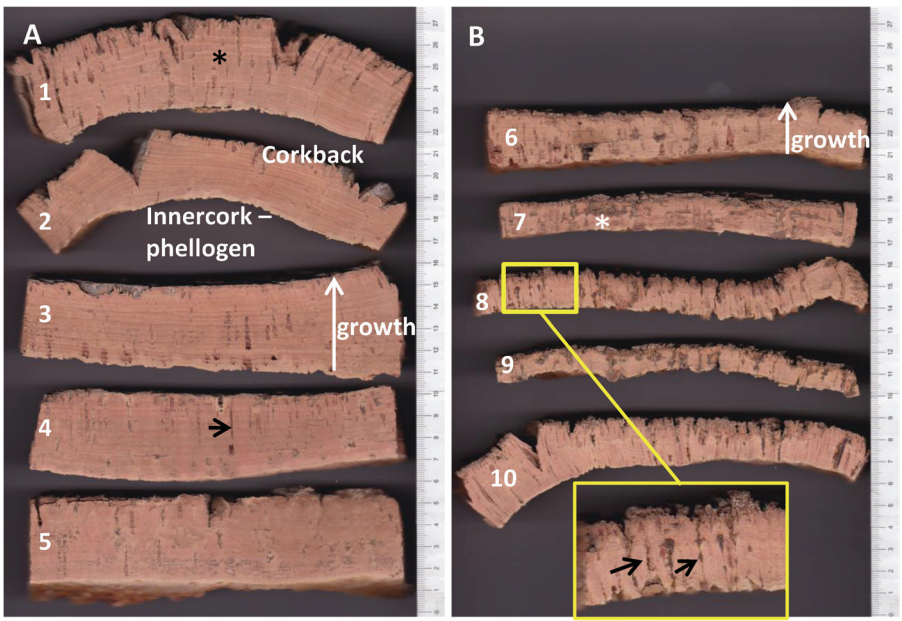

Fig. 1. Transverse cuts of cork planks where we can observe the radial growth (arrow) of the suber layer. The phellogen tissue used for pyrosequencing was removed from the plank shown in (A) and (B) immediately after debarking. (A) GQC showing a thickness of at least 28 mm displaying a reduced number of lenticels (arrow: plank 4) and no other inclusions. It is possible to clearly distinguish cork rings of growth that appear as white lines (*: plank 1). (B) BQC. Planks are thinner than in (A) with a maximum thickness of 17 mm and showing an elevated number of lenticular channels that completely cross the suber thickness. It is possible to see other cork defects such as inclusions (*: plank 7). The radial growth occurs from the inside, also known as cork belly, outwards to cork bark with its surface in contact with the environment. (This figure is available in colour at *JXB* online.)

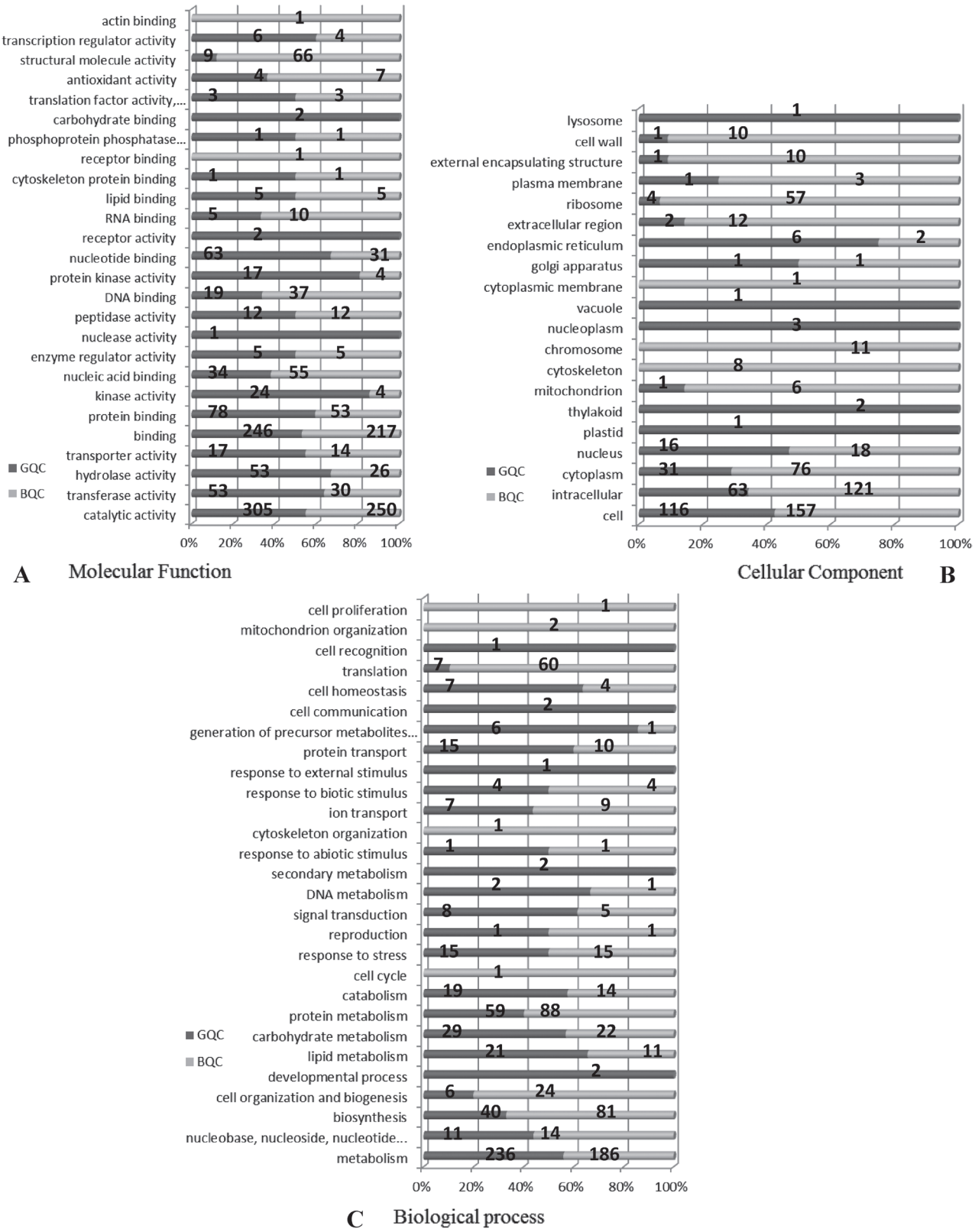

Fig. 2. Distribution of gene ontology categories. Percentage of contigs annotated for (A) molecular function, (B) cellular component, and (C) biological process are separated in dark grey (GQC) and light grey (BQC). Numbers in the graphs correspond to the number of genes in each sub-category.

receptor-binding categories is the result of just one gene in each category (Fig. 2A). The same happens for nuclease activity and carbohydrate binding, this time with one and two genes, respectively, differentially expressed in BQC.

When analysing the different sub-categories, it is possible to see that structural molecule activity, antioxidant activity, RNA binding, DNA binding, and enzyme regulatory activity show a higher number of genes that are more highly

expressed in BQC. The protein kinase activity and kinase activity sub-categories comprise more genes that are more highly expressed in GQK (Fig. 2A).

The ontology classification of cellular components shows that several sub-categories, such as plastid, vacuole, cytoplasmic membrane-bound vesicle and lysosome, are represented by a single gene (Fig. 2B). On the other hand, the most represented sub-categories are cell, intracellular, cytoplasm and nucleus. In the ontology classification of cellular components, the number of genes more highly expressed in BQC is 497 out of 753 (66%), mainly distributed among the sub-categories ribosome, extracellular region, mitochondrion, cytoplasm and intracellular bounding, cytoskeleton, cytoplasm, and intercellular categories. In contrast, the sub-categories endoplasmic reticulum and Golgi apparatus are mainly represented by more highly expressed genes in GQC (Fig. 2B).

In the biological process category, the sub-categories with 100% bias, such as cell proliferation, mitochondrion organization, cell communication, response to biotic stimulus, cytoskeleton, secondary metabolism, cell cycle and developmental process comprise only one or two genes (Fig. 2C). The sub-categories translation, protein metabolism, cell organization and biogenesis, and biosynthesis are mainly represented by genes that are highly expressed in BQC (Fig. 2C). The genes more highly expressed in GQC gather in higher numbers in the sub-categories generation of precursor metabolites and energy, DNA metabolism, and signal transduction and lipid metabolism (Fig. 2C).

When analysing all the expressed genes using the PCA, the separation that occurs along the 1st axis explains 95% of the variation (Pearson correlation coefficients, $P < 0.001$) and the separation between good and bad quality cork samples is achieved along the 2nd axis (5% of the variation; Pearson correlation coefficients, $P < 0.001$; Fig. 3, Supplementary Table S3).

Analysis of metabolic pathways with MapMan

MapMan software tools have been developed to map *Arabidopsis* transcriptome data to define functional categories, display them onto pathway diagrams, and allow

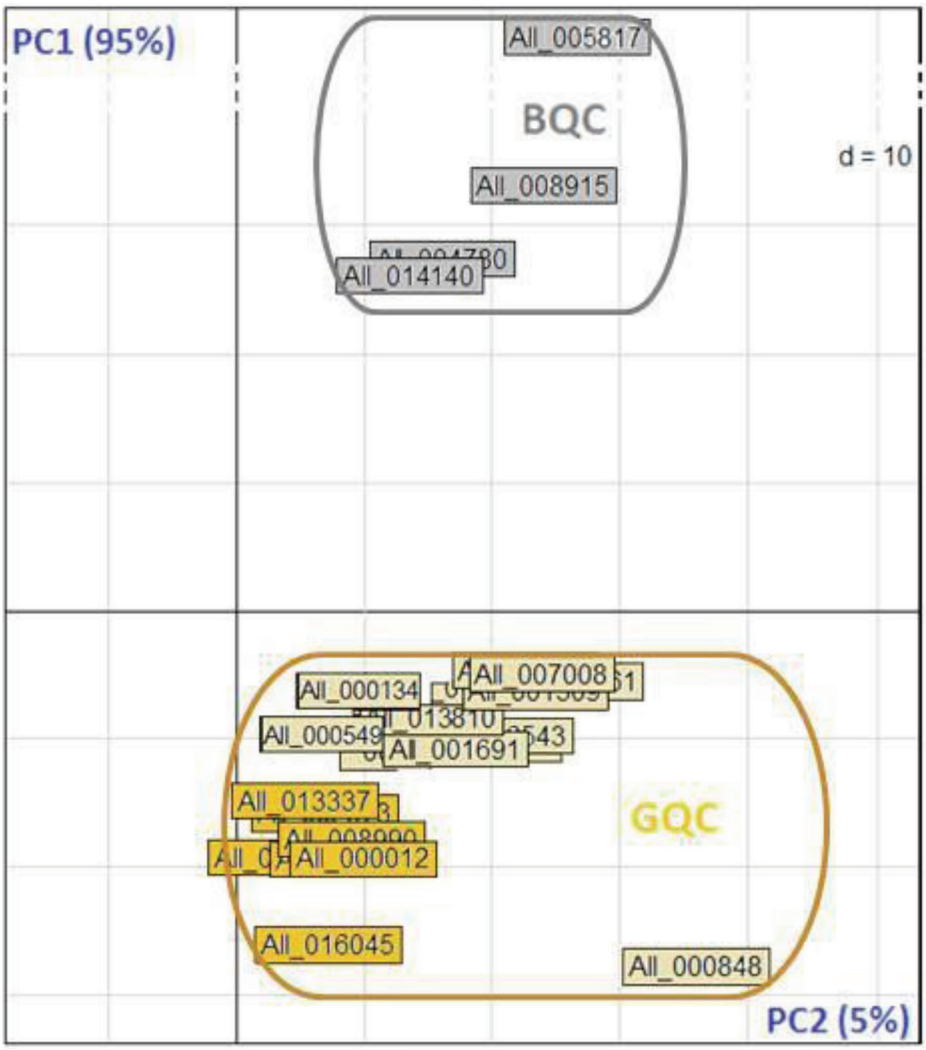

Fig. 3. Principal component analysis bi-plots (PC1-PC2) of the genes detected in BQC and GQC. Genes with higher expression in BQC (shading) or GQC (darker shading) are highlighted.

time-course analysis for functional gene groups. In this study we adapted these tools to *Q. suber* by generating unique mapping files of the *Q. suber* pyrosequencing data. The automatic annotations of the differentially expressed sequences between good- and bad-quality cork allows the assignment of the 1123 differentially expressed sequences with annotation into 36 distinct bins. We have to take into consideration the fact that the annotation performed by Mercator software is not as robust as the annotation used by the pyrosequencing biostatistical pipeline (see Material and Methods). The MapMan software tool was used to display the transcriptome data onto pathway diagrams. Three maps were chosen, giving a general distribution of the differentially expressed genes: cellular responses (Fig. 4A), different classes of stresses (Fig. 4B), and transcription factors (Fig. 4C). Because only 1123 annotated genes are differentially expressed (cut-off eight contigs and

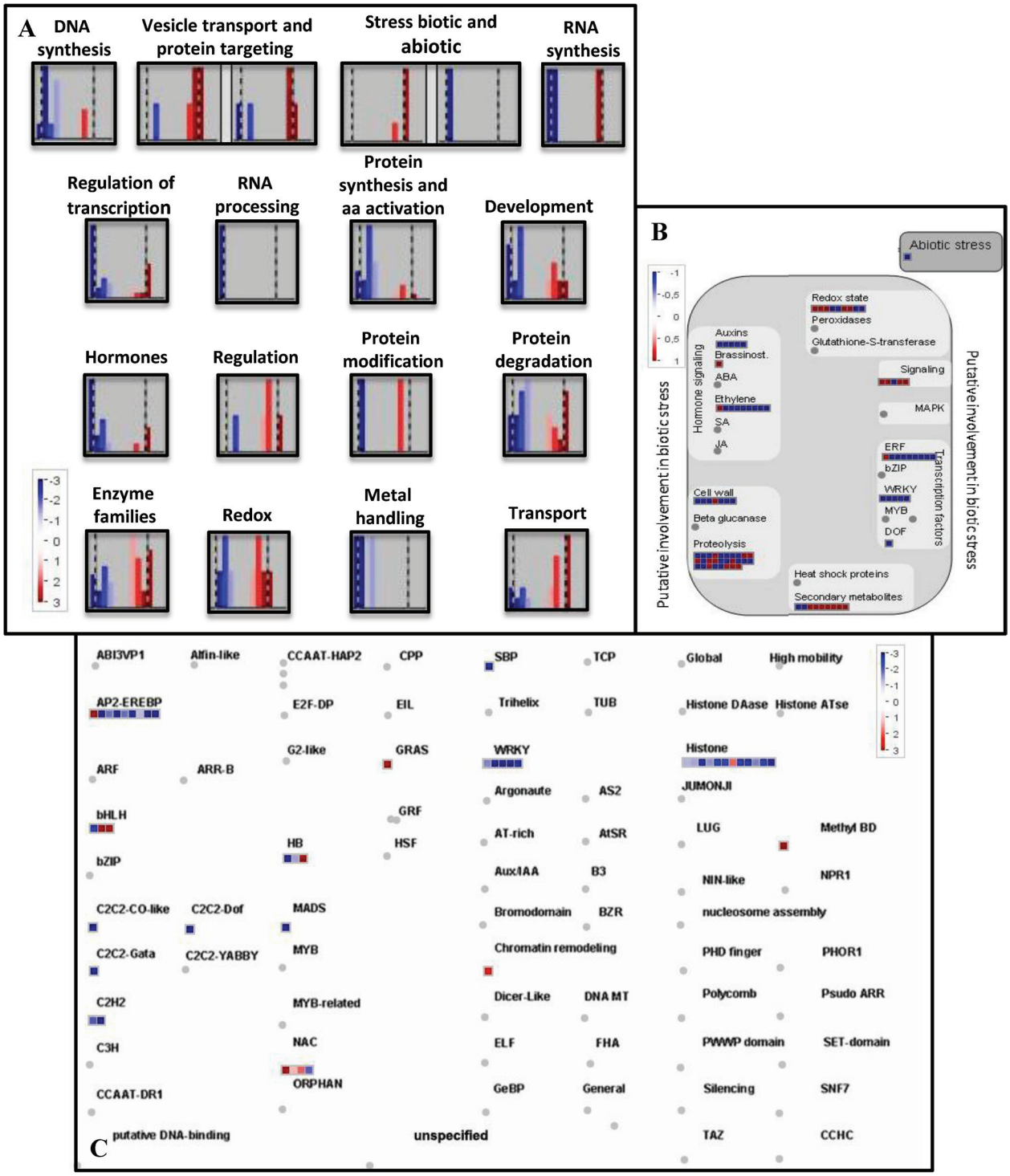

Fig. 4. Distribution of contigs according to MapMan analysis among (A) cellular responses, (B) different classes of stresses, and (c) different families of transcription factors. Contigs with values under 1 were more highly expressed in BQC and contigs with values above 1 were more highly expressed in GQC. (This figure is available in colour at *JXB* online.)

$P < 0.05$) and used for MapMan analysis, many of the pathways are only represented by a few genes.

In accordance with the gene ontology analysis (Fig. 2), with MapMan analysis, we realized that genes related to protein synthesis (RNA synthesis, RNA processing, protein synthesis, and amino acid activation) are more highly expressed in BQC (Fig. 4A). The DNA synthesis and metal-handling proteins are more highly expressed in BQC (Fig. 4A). Another cellular mechanism showing a large number of differentially expressed genes is abiotic stress. Under MapMan analysis, the number of transcription factors highly expressed in BQC is higher (Fig. 4B, C). However, this is not observed when looking at gene ontology analysis, where the number of stress genes is equivalent (Fig. 2C). The transcription factors associated with biotic stress, namely ERFs and WRKY, are more highly expressed in BQC (Table 1, Fig. 4B, C). This holds true for the list of transcription factors that are differentially expressed (Table 1, Fig. 4C). AP2, WRKY and C2R2 transcription factors, as well as histones, are more highly represented in samples of BQC (Table 1). GRAS- and NAC-family transcription factors and chromatin remodelling display more expressed elements in GQC (Fig. 4)

Real time RT-PCR

Several genes involved in pathways analysed throughout this work were chosen for quantitative real-time RT-PCR as a confirmation of the transcriptional profile (*Arginine Decarboxylase*, *Dehydrin*, *PAL*, *Sucrose Synthase*, *bHLH96*, and *VLFA* (Table 2, Fig. 5). Primers were designed based on the sequences obtained after whole-genome sequencing (primer gene in Supplementary Table S1). It was assured that each amplified a single gene by blasting the primer sequence against our database.

The expression levels of all the genes tested by real-time RT-PCR on individual samples were in accordance with the expression levels obtained after pyrosequencing, as shown in Table 2 (Fig. 5).

Despite all samples following the same tendency (higher expression in GQC or BQC), the levels of expression vary, as can be seen in *PAL* gene sample 2 (S2) when compared with the remaining samples (Fig. 5). The *Dehydrin* gene does not show a strong difference in expression between all GQC and BQC samples, with the exception of sample 4 (S4) (Fig. 5); hence this gene lacks a significant difference in expression ($P = 0.38$). In fact, the differential expression level verified in the transcriptomic data shows that out of all the genes analysed here, *Dehydrin* is the one that displays the fold-change value closest to 1 (0.65) (Table 2). The correlation coefficient R2 between the expression fold-change obtained from the qRT-PCR and the expression fold-change obtained after 454 pyrosequencing is 0.8242 (Supplementary Table S1).

Total phenolic content and phenylpropanoid pathway genes

The soluble phenolic content was determined from the same cork planks from which phellogenic tissue was removed for sequencing and shown in Fig. 1. Every time that RNA and protein extraction was performed, we observed that BQC samples were harder to handle and the final extract solution displayed a more-brownish colouration (data not shown) compared to GQC samples.

As expected, the total free phenolic content calculated in mg of GAE ml^{-1} is more than double in BQC samples (6.92 mg GAE ml^{-1}) compared with GQC (3.315 mg GAE ml^{-1}) (Fig. 6). These results led us to look more closely into genes involved in the phenylpropanoid pathway. The pathway presented here was based on data obtained from the KEGG pathway database, MapMan software and *Arabidopsis* metabolic pathway present in TAIR (Fig. 7). Based on the *Arabidopsis* enzymes identified in the shikimate and phenylpropanoid pathway, we blasted the *Arabidopsis* protein sequence against our phelogenic transcriptome database and obtained the genes listed in Table 3. Upon the blast, we identified two *Q. suber* paralogues for 3-dehydroshikimate dehydratase and phenylalanine ammonia lyase (PAL) (Table 3).

The first step in the biosynthesis of phenylpropanoid compounds is the conversion of L-phenylalanine to *trans*-cinnamic acid by PAL (Fig. 7). The second step is the hydroxylation of *trans*-cinnamic acid by a cytochrome P450 monooxygenase (cinnamic acid 4-hydroxylase, C4H). These initial pathway steps are involved in the synthesis of lignin, suberin, and phenolics. Only two genes encoding enzymes involved in the phenylpropanoid pathway are more highly expressed in GQC: *PAL* and caffeoyl CoA 3-*O*-methyltransferase (CCoAOMT). Transcripts more highly expressed in BQC are found for *trans*-C4H, cinnamoyl CoA reductase (CCR) and *O*-methyl transferase (COMT). The remaining genes for the following enzymes show no differential expression (NDE): 4-coumarate-CoA ligase (4CL); *p*-coumarate 3-hydrolase (C3H); and cinnamyl-alcohol dehydrogenase (CAD). Note that when moving to scopolin biosynthesis, one of the free phenolics that can be found in cork (Conde *et al.*, 1998), 2-oxoglutarate dependent dioxygenase (2OG), an orthologue of At3g13610, is more highly expressed in BQC (Fig. 7, Table 3). Most of the genes participating in the flavonoid pathway are not differentially expressed. Nevertheless, the synthesis of condensed tannins seems to be favoured in GQC as suggested by the upregulation in these samples of genes coding for anthocyanidin synthase (ANS) and dihydroflavonol reductase-like (BAN) (Fig. 7, Table 3).

Transcripts involved in suberin synthesis

The verification of expression levels of genes involved in suberin biosynthesis was made using the *Arabidopsis* orthologous sequence blasted against our database. Genes were chosen from the work of Soler *et al.* (2007). As can be observed in Table 4, all genes involved in suberin and wax biosynthesis, and biosynthesis of unsaturated fatty acids and fatty acid, are either higher expressed in GQC (values above 1) or show no differential expression (NDE). The genes coding for the enzymes responsible for synthesis of aromatic monomers necessary for suberin synthesis that are derived

Table 1. *Differential expression levels of sequence-specific DNA-binding transcription factor (GO: 0003700)*

EST *Q. suber* accession no.[a]	Fold change[b]	Best TAIR BLASTX	q-value	Annotation
All_003894	0.117	AT3G58780	0.313	Transcription factor, MADS-box
All_001355	0.217	AT3G54810	0.254	Zinc finger, GATA-type
All_004297	0.269	AT1G46480	0.260	Homeobox
All_001662	0.705	AT2G22430	0.003	Helix–turn–helix motif, lambda-like repressor – HD-ZIP protein ATHB-6
bZIP transcription factor family				
All_007028	0.182	AT1G75390	$6.81e^{-6}$	Basic-leucine zipper (bZIP) transcription factor
All_007036	0.227	AT4G34590	0.293	Basic-leucine zipper (bZIP) transcription factor
WRKY transcription factor family				
All_016452	0.2	AT2G24570	0.312	DNA-binding WRKY
All_006373	0.243	AT1G29280	0.056	DNA-binding WRKY
All_007469	0.259	AT5G13080	0.006	DNA-binding WRKY
All_007674	0.361	AT5G13080	0.228	DNA-binding WRKY
All_005356	0.605	AT2G46130	$2.22e^{-7}$	DNA-binding WRKY
AP2/EREBP transcription factor family				
All_005415	0.144	AT5G25190	$1.12e^{-9}$	Pathogenesis-related transcriptional factor/ERF
All_003976	0.147	AT1G19210	$6.93e^{-7}$	Pathogenesis-related transcriptional factor/ERF, DNA-binding
All_001279	0.214	AT5G47230	$3.25e^{-5}$	Pathogenesis-related transcriptional factor/ERF, DNA-binding
All_002811	0.235	AT3G16770	$2.53e^{-27}$	Pathogenesis-related transcriptional factor/ERF, DNA-binding
All_001371	0.296	AT4G37750	0.286	Pathogenesis-related transcriptional factor/ERF, DNA-binding
All_001999	0.818	AT1G53910	0.265	Pathogenesis-related transcriptional factor/ERF, DNA-binding
CCAAT-binding transcription factor family				
All_001545	1.909	AT1G72830	0.116	CCAAT-binding transcription factor, subunit B
All_002529	3.727	AT5G12840	0.046	CCAAT-binding transcription factor, subunit B
bHLH				
All_001518	0.408	AT1G72210	0.152	Helix–loop–helix DNA-binding domain
All_006430	Only in BQC	AT2G40435	0.260	Helix–loop–helix DNA-binding
All_014772	0.058	XP_002303073.1 Populus	0.286	BHLH transcription factor
All_002644	6	AT5G65640	0.328	Helix–loop–helix DNA-binding domain
All_003550	3.88	AT1G29950	0.086	Helix–loop–helix DNA-binding domain
All_001285	3.187	AT4G37850	0.027	Helix–loop–helix DNA-binding domain

Values below one mean that the genes are more highly expressed in BQC.
[a] *Q. suber* accession number refers to the accession number assigned after pyrosequencing.
[b] Fold-change is expressed as the ratio of gene expression of GQC to BQC.

Table 2. *Differential expression levels of the same genes chosen for real time RT-PCR analysis*

EST *Q. suber* accession no.[a]	Real time RT-PCR *P*-values for sample average	Fold change[b]	Best TAIR BLASTX	q-value	Annotation
All_006118	0.0244	0.32	At2g16500	0.003	*Arginine decarboxylase*
All_003162	0.3821	0.65	At1g20440	0.364	*Dehydrin*
All_001518	$3.988e^{-5}$	0.40	At1g72210	0.152	*bHLH96*
All_000012	0.00711	2.03	At3g53260	$5.30e^{-37}$	*PAL*
All_000113	0.000194	1.474	At1g04220	$7.30e^{-9}$	*VLFA*
All_000001	0.00032	1.378	At3g43190	0.001	*Sucrose synthase*

Values below one means that the genes are more highly expressed in BQC.
[a] *Q. suber* accession number refers to the accession number assigned after pyrosequencing.
[b] Fold-change is expressed as the ratio of gene expression of GQC to BQC.

from the phenylpropanoid pathway are listed in Table 4 and Fig. 7.

Suberin is a key compound for the development suber layers. For this reason, we performed another PCA analysis using the genes involved in suberin synthesis and listed in Table 4. The PCA analysis shows that these genes do not discriminate between samples of GQC and BQC. The separation between samples is achieved along the 2nd axis, but differences are not considered significant according to the Pearson correlation coefficients (Supplementary Table S1). Based on the PCA, the suberin-related genes here analysed can be associated with cork production, irrespective of cork quality.

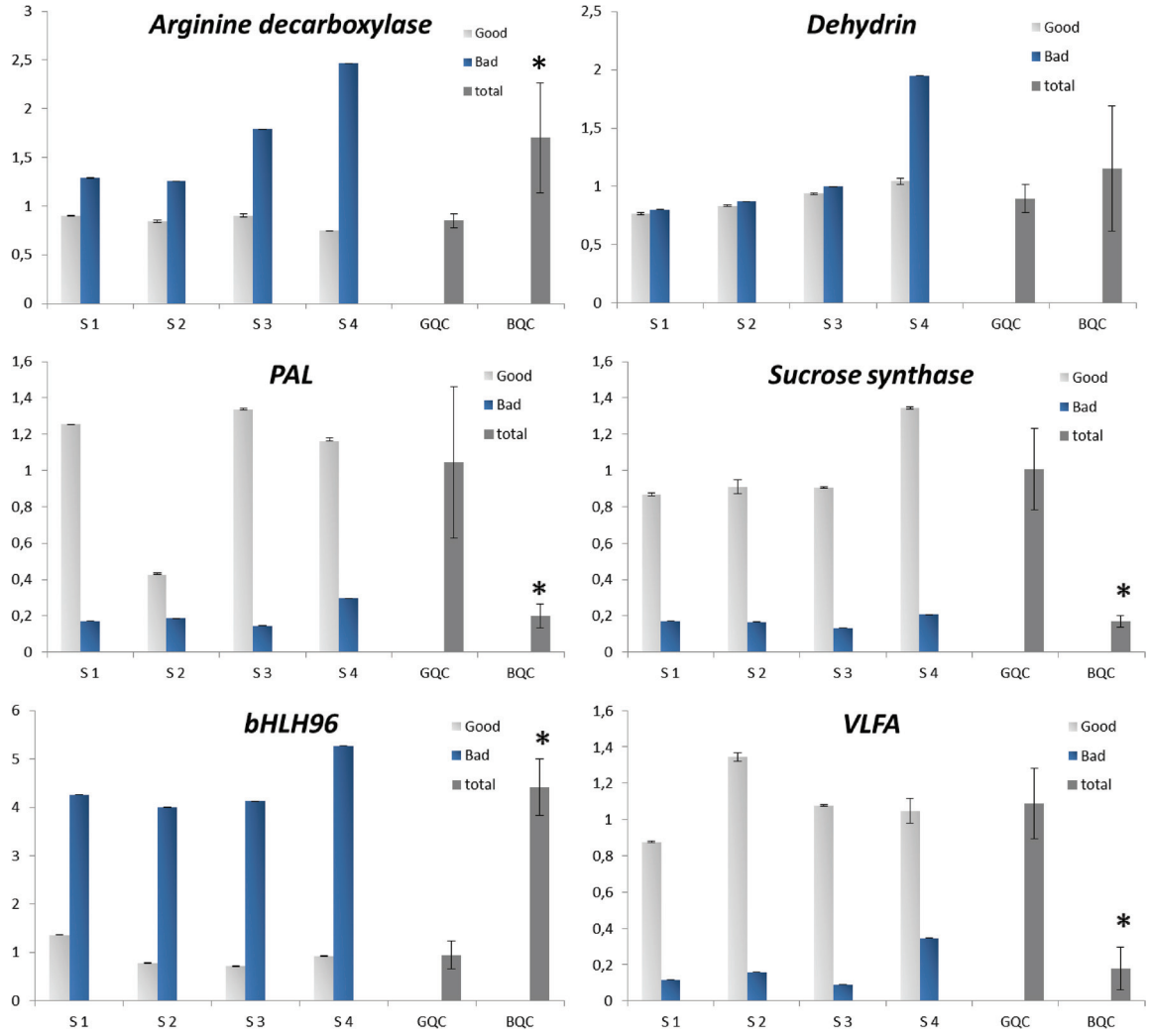

Fig. 5. Results from real-time RT-PCR amplification from the same RNA used for pyrosequencing from GQC and BQC pools. Amplifications were performed with specific primers chosen randomly from the transcriptome list. It was ensured that the primers amplified only a single gene. Standard deviation levels are shown (n = 4–6). Stars indicate significant difference at level $P < 0.05$ in comparison to GQC. The correlation coefficient R2 between the expression fold-changes obtained from the qRT-PCR and after 454 pyrosequencing is 0.8242 (Supplementary Table S1). (This figure is available in colour at *JXB* online.)

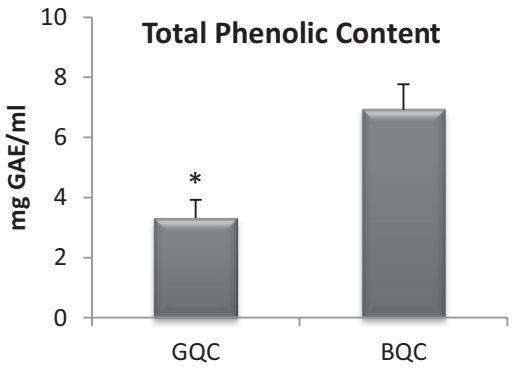

Fig. 6. Total phenolic content of extracts of GCQ and BQC. Content is expressed in milligrams of gallic acid per gram of dry cork. Asterisk indicates significant difference with $P < 0.05$ in comparison to GQC.

Analysis of putative genes involved in lenticel development

We looked for orthologues of genes that are involved in stomatal development because it is believed that stomata are the founder structures for lenticular channels (Kalachanis and Psaras, 2007). We screened the *Q. suber* phellogen transcriptome database using the *Arabidopsis* bHLH transcription factors encoded by the *FAMA*, *SPEECHLESS*, and *MUTE* genes; and *Arabidopsis* homeodomain-leucine zipper IV proteins encoded by *MERISTEM LAYER 1* (*ML1*) and *HOMEODOMAIN GLABROUS2* (*HDG2*) genes.

The protein encoded by the five *bHLH* genes aligned here (three *Arabidopsis* genes and the two *Q. suber* orthologues) consists of one conserved domain, a 65-amino acid sequence highlighted in red (Supplementary Figure S1). When alignment was performed with *AtML1*, *HDG2* and the *Q. suber* orthologue, two conserved domains were detected. The homeodomain related (99 amino acids) is only present in

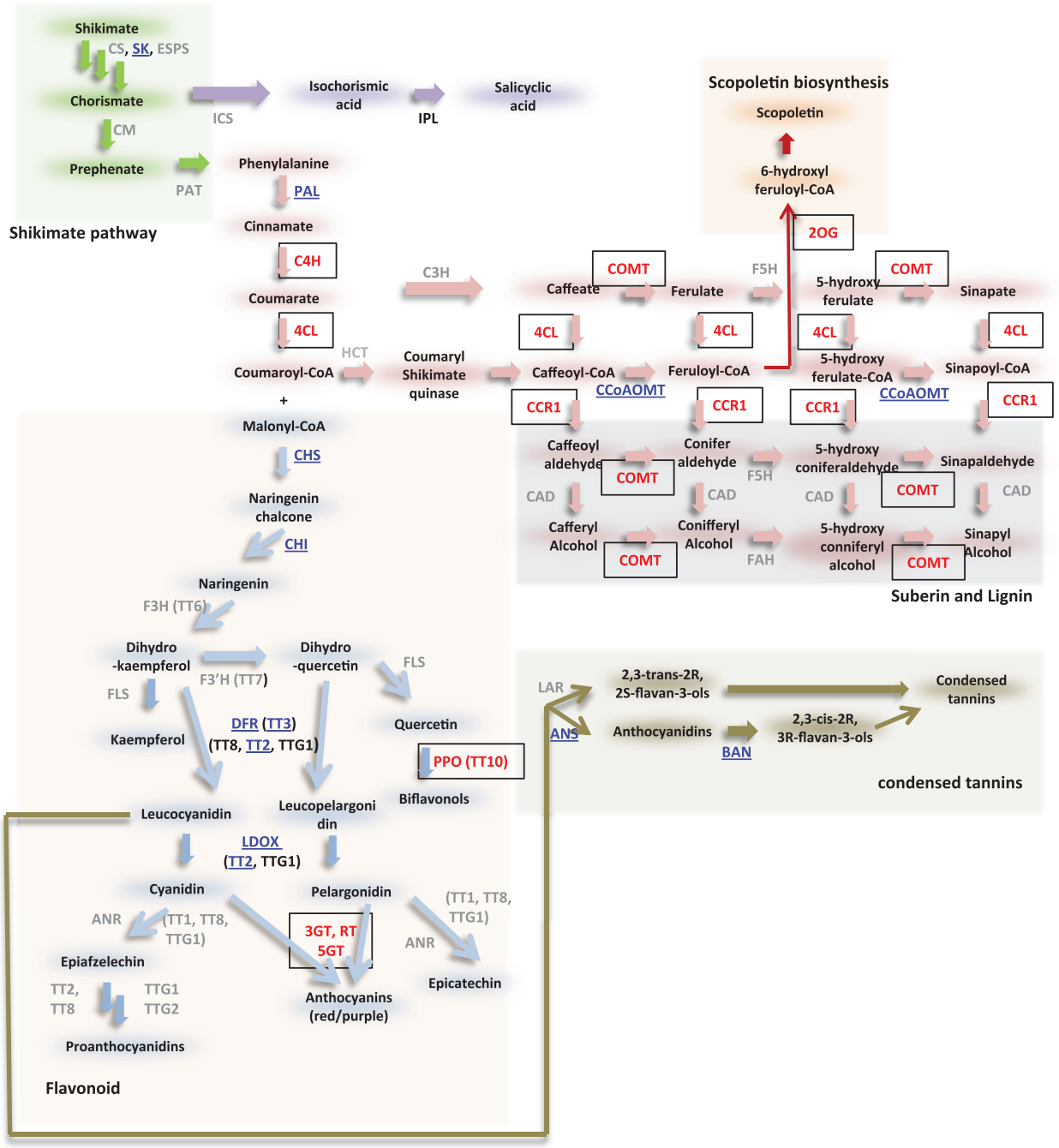

Fig. 7. Partial representation of the shikimate, phenylpropanoid, flavonoid, condensed tannin, and scopoletin biosynthetic pathways inferred from KEGG, TAIR and MapMan databases. Underlined enzymes indicate higher expression in GQC; boxes indicate upregulation in BQC. The remaining enzymes show no differential expression between the two types of cork. See Table 3 for enzyme nomenclature. (This figure is available in colour at *JXB* online.)

AtML1 and *HDG2*. The START conserved domain is present in the three genes (232 amino acids) (Supplementary Figure S2).

The two *Q. suber* genes *All_001408* and *All_001518* share 57% homology in their amino acid sequence and belong to the same clade of *SPCH*, *FAMA* and *MUTE*. Bootstrapping showed 82% replication between the two bHLH *Q. suber* identified proteins and 58% replication of their closely associated *Arabidopsis* genes. The clade comprising *HDG2*, *AtML1* and the putative HD-ZIP IV *Q. suber* orthologue shows a 100% replication of *Q. suber* proteins and the *Arabidopsis* ones (Fig. 8).

Stress-related genes

Cork tissue develops as a defence mechanism. Due to this fact, and to the result after MapMan analysis, we decided to investigate classes of genes known to be more highly

Table 3. *Differential expression levels (fold change) of genes involved in phenylpropanoid and shikimate pathways*

EST *Quercus suber* accession no.[a]	Fold change[b]	Best TAIR BLASTX	q-value	Annotation
Phenylpropanoid pathway				
All_000080	2.47	At3G06350	0.0004	3-Dehydroshikimate dehydratase
All_000285	NDE			
All_000009	1.48	At2G37040	$3.81e^{-5}$	Phenylalanine ammonia-lyase (PAL)
All_000012	2.03		$5.30e^{-37}$	
All_000500	0.86	At2G30490	0.335	*Trans*-cinnnamate 4-hydroxylase (C4H)
All_001809	0.62	At5G38120	0.001	4-Coumarate-CoA ligase (4CL)
All_001481	0.59	At1G80820	0.0002	Cinnamoyl-CoA reductase (CCR)
All_001003	NDE	At4g34230		Cinnamyl-alcohol dehydrogenase (CAD)
All_003606	2.12	At1g67990	$1.52e^{-14}$	Caffeoyl-CoA *O*-methyl transferase (CCoAOMT)
All_000148	NDE	At2g40890		*p*-Coumarate 3-hydrolase (C3H), CYP98A3
All_000829	NDE	At5g04330		Ferulate-5-hydroxylase (F5H), CYP84A
All_000980	0.64	At5g54160	$1.50e^{-6}$	*O*-methyltransferase (OMT)
All_000829	NDE	AT4G36220		Cytochrome P450 84A1 / Ferulate-5-hydroxylase (FAH)
All_001077	0.55	At3g13610	$7.246e^{-5}$	2-oxoglutarate (2OG)
Flavonoid pathway				
All_010302	NDE	AT1G34790		Transparent Testa 1 (TT1)
All_004385	4		$8.98e^{-5}$	
All_004934	NDE			
All_003703	NDE			Transparent Testa 2 (TT2) / MYB Domain Protein 123
All_005222	2.6		$1.04e^{-5}$	
All_004268	NDE	AT5G35550		
All_002113	NDE			
All_001357	2	AT5G42800	$2.87e^{-13}$	Transparent Testa 3 (TT3) /
All_001579	3.6		0.358	Dihydroflavonol 4-Reductase (DFR)
All_001381	2.6		$1.85e^{-19}$	
All_002591	NDE			
All_002592	NDE			Transparent Testa 4 (TT4) /
All_002573	2.3			Chalcone synthase (CHS)
All_004076	1.7	AT5G13930	$8.45e^{-34}$	
All_006196	NDE		0.005	
All_003758	1.6	AT3G55120	$3.45e^{-7}$	Transparent Testa 5 (TT5) / Chalcone flavanone isomerase (CFI) / Chalcone isomerase (CHI)
All_001001	NDE	AT3G51240		Transparent Testa 6 (TT6) / flavanone 3-hydroxylase (F3H)
All_000443	NDE	AT5G07990		Transparent Testa 7 (TT7) / Cytochrome P450 75B1 (CYP75B1)
All_001326	NDE	AT4G09820		Transparent Testa 8 (TT8) / BHLH42
All_000064	0.45	AT5G48100	0.257	Transparent Testa 10 (TT10) / Transparent Testa 15 (TT15) / Laccase-Like 15 (LAC15)
All_001615	NDE	AT5G24520		
All_001278	NDE			Transparent Testa Glabra 1 (TTG1)
All_001403	NDE			
All_001539	NDE	AT2G37260		Transparent Testa Glabra 2 (TTG2)
All_006986	NDE			
All_001391	NDE	AT5G08640		Flavonol synthase (FLS)
All_000418	O.59	AT5G54060	0.004	UDP-Glucose (3GT)
All_001383	0.41	–	0.002	Rhamnosyltransferase (RT)
All_000145	0.45	–	$2.61e^{-5}$	Anthocyanin 5-*O*-glucosyltransferase (5GT)
Shikimate pathway				
All_002947	NDE	AT1G48850		Chorismate synthase (CS)
All_005118	2.6	AT2G21940	0.076	Shikimate kinase (SK)
		AT4G39540		
All_002254	NDE	AT1G48860		3-Phosphoshikimate 1-carboxyvinyltransferase (ESPS)
		AT2G45300		
All_010106	NDE	AT1G18870		Isochorismate synthase (ICS)

Table 3. *Continued*

EST *Quercus suber* accession no.[a]	Fold change[b]	Best TAIR BLASTX	q-value	Annotation
All_001531	NDE	AT1G69370		Chorismate mutase (CM)
All_002232	NDE	AT3G29200		
		AT5G10870		
All_000624	NDE			Prephenate ammonia-lyase (PAT)
Condensed tannins				
All_001381	2.6	AT1G61720	$1.85e^{-19}$	Dihydroflavonol Reductase Like (BAN)
All_001105	2.9	AT4G22880	0.00	Anthocyanidin synthase (ANS)
All_003063	NDE			
All_001293	NDE	*Desmodium uncinatum*		
All_001200	NDE	(225101-NCBI)		Leucoanthocyanidin reductase (LAR)

Values below one means that the genes are more highly expressed in BQC. NDE, not differentially expressed.
[a] *Q. suber* accession number refers to the accession number assigned after pyrosequencing.
[b] Fold change is expressed as the ratio of gene expression of GQC to BQC.

expressed under abiotic stress. The GO-class annotation revealed that the number of genes involved in protein synthesis and protein degradation, such as ribosomal protein genes and ubiquitin-related genes, is substantially higher in BQC than GQC (Table 5). The same happens for histone genes that, among several roles, are involved in DNA repair during the cell cycle (Landry *et al.*, 1997). When looking at heat-shock proteins, we were only able to find differentially expressed genes in GQC (Table 5). Glutathione *S*-transferase (GST) displays an important protective function (Edwards *et al.*, 2000) and shows a higher number of genes in BQC. Superoxide dismutases are involved in oxidative stress, and the genes differentially expressed between the two types of cork and coding for these enzymes can only be detected in BQC (Table 5).

Discussion

Transcriptome comparison between good- and bad-quality cork

The present work establishes a comparison of phellogenic genes differentially expressed in GQC and BQC making use of Roche 454-pyrosequencing and further annotation performed by Newbler 2.6 (Roche). Recent studies have demonstrated highly successful *de novo* assemblies of 454 EST data for organisms with no prior genomic resources (Novaes *et al.*, 2008; Vera *et al.*, 2008; Alagna *et al.*, 2009), a feature that is observed for *Q. suber*. The annotation software used gives us a high degree of confidence because after a comparison of six annotation programmes (CAP3, CLC, MIRA, Newbler 2.3, Newbler 2.5, and SeqMan) for *de novo* assembly of transcriptome data obtained from an organism with no previous genomic resources, Newbler 2.5 had the best contig length metrics and the best alignments to related reference sequences (Kumar and Blaxter, 2010).

Suberin synthesis

Until now, very few molecular studies have examined cork development, but Soler *et al.* (2007), using a subtractive hybridization microarray system, identified specific genes putatively involved in cork formation, mainly in suberin synthesis. Other molecular studies on suberin synthesis have shown that both CYP86A1 and CYP86B1, required for the biosynthesis of very-long-chain saturated α, ϖ-bifunctional aliphatic monomers (Molina *et al.*, 2009), are associated with endoplasmatic reticulum, where suberin monomer biosynthesis seems to occur before export into the apoplast (Millar *et al.*, 1999; Höfer *et al.*, 2008). Our transcriptomic data show that the *CYP86A1* orthologue is highly expressed in GQC, a factor that could determine the higher expression of endoplasmic reticulum-related genes observed in GQC (Fig. 2B). These samples also showed a more intense lipid metabolism that may be responsible for the synthesis of precursors of the aliphatic domain of suberin, a factor that could be related to the thicker cork layers found in GQC trees. Moreover, we found higher expression of a gene coding for sucrose synthase in GQC samples (Supplementary Table S2). This enzyme may provide hexoses to be used downstream in fatty acid biosynthesis, revealing a different primary metabolism between good- and bad-quality cork.

Even though suberin synthesis is a key factor for the production of cork, the PCA analysis on suberin-related genes identified tells us that these genes do not discriminate between samples of GQC and BQC. Therefore, these genes are associated with cork production irrespective of quality.

Expression of stress-related genes during cork development

The formation of suberin-reinforced cell walls has been shown to be a plant response to drought (Hose *et al.*, 2001). In fact, cork formation and stress are related. According to the work of Ricardo *et al.* (2011), during cork development many of the proteins detected share a dual stress response and defence function. Also, in potato periderm, suberized tissue used as a model system for suberin formation, more than half of the proteins identified corresponded to stress responses (Barel and Ginzberg, 2008; Chaves *et al.*, 2009). These results strongly suggest that the development of the cork layer is as a protective tissue against environmental

Table 4. *Differential expression levels (fold change) of genes encoding proteins involved in suberin synthesis*

Suberin and wax biosynthesis.

Q. suber accession no.[a]	Fold change[b]	Best TAIR BLASTX	q-value	Annotation
All_000134	1.267	At1g01600	2,90e^{-5}	Cytochrome P450, family 86 (CYP86A4S)
		At1g63710		
		At2g45970		
		At4g00360		
All_002989	NDE	At3g10570		Cytochrome P450, family 77 (CYP77A6)
All_004601				
All_000353	NDE	At5g41040		Omega-hydroxypalmitate *O*-feruloyl transferase
All_000345	1.283		0.105	
All_000340	1.504		2.74e^{-7}	
All_000356	NDE			
All_013100	NDE	At1g70680		Caleosin-related family protein
All_013141	NDE			
All_000132	NDE	POPTR_588365		Cytochrome P450, family 94 (CYP94A1)
All_000075	NDE	At5g23190		Cytochrome P450, family 86 (CYP86B1)
All_000025		At4g16760		Long-chain-fatty-acyl-CoA reductase
All_000415	NDE	At2g35690		
All_003882				
All_005135				
--------		At1g02205		Aldehyde decarbonylase (CER1)
All_000205	2.78	At3g44540	0.179	Fatty acyl-CoA reductase
All_000301	NDE			
----------		At5g37300		Diacylglycerol O-acyltransferase
Biosynthesis of unsaturated fatty acids				
All_001650	NDE	At1g01710		Palmitoyl-CoA hydrolase
		At4g00520		
Fatty acid elongation				
All_000492	NDE	At2g33150		Acetil-CoA acyltransferase
All_005669	NDE	At5g47720		
All_003985	0.63	At5g48230	0.312	
All_006584	NDE	At3g15290		3-hydroxyacyl-CoA dehydrogenase
All_008459	2.5		0.003	
All_002945	NDE	At1g60550		Enoyl-CoA hydratase
All_004827	NDE			
All_002608	4		0.0001	
All_003110	4.2	At3g45770	0.332	*Trans*-2-enoyl-CoA reductase
All_000415	NDE	At3g51840		2-Enoyl-CoA reductase
All_006890	NDE	At3g06810		
All_000165	2.125	At1g01120	5.31e^{-9}	3-ketocyl-CoA synthase (in endoplasmic reticulum)
All_000113	1.47	At1g04220	7.3e^{-9}	
All_000221	2.62		0.0002	
All_002243	1.33	At1g67730	0.019	Very-long-chain 3-oxoacyl-CoA
All_004052	NDE	At5g10480		Very-long-chain (BR)-3-hydroxyacyl dehydratase
All_001913	4.8	At3g55360	2.87e^{-13}	Very-long-chain enoyl-CoA reductase

Values below zero mean that the genes are more highly expressed in BQC. NDE, not differentially expressed.
[a] *Q. suber* accession number refers to the accession number assigned after pyrosequencing.
[b] Fold change is expressed as the ratio of gene expression of GQC to BQC.

stress conditions. Our results reinforce this view. The higher expression of ribosomal protein genes in BQC seem to be a response to damaged ribosomes promoted by abiotic stresses that are able to induce cross-links in ribosomal RNA or between mRNA, tRNA, rRNA, and proteins (Noah *et al.*, 2000). As observed by Casati and Walbot (2003), ribosomal protein synthesis is stimulated when ribosomal genes are damaged in order to maintain correct protein synthesis in the cell. We also observed that BQC displayed a higher number of histones, ubiquitin-related genes, and DNA-binding proteins in comparison to GQC. This was also shown by Casati & Walbot (2003) as a consequence of an indirect effect of the damaged DNA caused by UV-B radiation. Ubiquitin-related genes recognize the cross-link and oxidatively damaged proteins caused by UV radiation (Gerhardt *et al.*, 1999), and target them for degradation via ubiquitination, leading

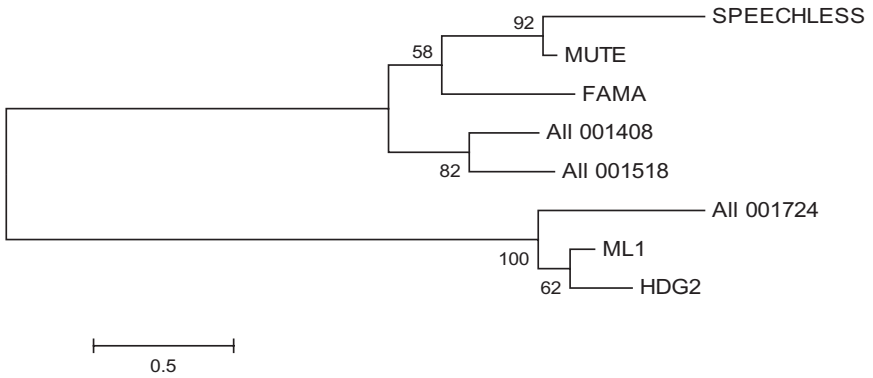

Fig. 8. Phylogenetic analysis of the three bHLH *Arabidopsis* proteins involved in stomatal formation and development (SPEECHLESS, FAMA and MUTE), two *Q. suber* orthologous (All_001408 and All_01518), and two *Arabidospsis* HD-ZIP IV proteins (ML1 and HDG2) responsible for stomatal differentiation and one *Q. suber* orthologue (All_001724). Bootstrapped values indicating the level of significance (%) by the separation of branches are shown. The branch length indicates the extent of the difference according to the scale shown. Note that SPEECHLESS, MUTE, FAMA, All_001408 and All_001518 are grouped within the same sub-clade while ML1, HDG2 and All_001724 group in distinct sub-clade.

Table 5. *Number of paralogues belonging to gene classes present in GQC and BQC: ubiquitin-related, ribosomal, histone, heat shock, glutathione* S*-transferase, and superoxide dismutases*

Gene class	GQC	BQC
Ubiquitin-related genes	10	20
Ribosomal protein genes	4	58
Histone genes	3	11
Heat-shock genes	16	0
Glutathione *S*-transferase genes	4	6
Superoxide dismutase genes	0	3

to an improvement of cellular homeostasis and maintenance of correct gene regulation.

Only GQC showed an upregulation of transcripts encoding heat-shock proteins (HSPs). Cork oak tissues, in response to heat, water deficit, UV radiation, or oxidative stress, accumulated class I HSP and HSP17 transcripts (Pla *et al.*, 1998; Puigderrajols *et al.*, 2002) conferring tolerance against heat stress (Soto *et al.*, 1999; Jofré *et al.*, 2003). In chestnut and holm oak, class I HSPs were induced during the summer (Verdaguer *et al.*, 2003; Lopez-Matas *et al.*, 2004). The upregulation of these proteins, only observed in GQC, is apparently responsible for a cell protection mechanism leading to correct phellogen development. In general, HSPs function by promoting correct folding, trafficking, maturation, and degradation of proteins, cell cycle control, and stabilization of proteins by preventing their aggregation. Based on the genes highly expressed in BQC and described earlier, these cellular mechanisms are exactly the ones compromised in BQC. In conclusion, the role of HSPs is to protect cells allowing the re-establishment of a normal protein conformation and therefore, cellular homeostasis (Buchner, 1999; Richter and Buchner, 2001; Sung *et al.*, 2001; Young *et al.*, 2001; Wang *et al.*, 2004). This important chaperone role of HSPs is not verified in BQC, which could lead to impairment of cell homeostasis and cell division.

MapMan and GO annotation analysis showed that certain classes of transcription factors were mainly highly expressed in BQC. They belong to the families of AP2-EREBP, WRKY, Histone, MADS C2H2, and bHLH. Transcription factors of the families AP2/EREBP, bZIP/HD-ZIP, and Myb, and several classes of zinc finger domains, have been implicated in plant stress responses (Chen *et al.*, 2002; Nakashima *et al.*, 2009). The ethylene-responsive element binding factors (ERFs) are more highly expressed in BQC. This class of genes plays an important role in adaptation to biotic and abiotic stress such as pathogen attack, wounding, UV irradiation, extreme temperature and drought (Penninckx *et al.*, 1996).

Lenticular channel formation and its role in the cork layer

There are no genetic studies on lenticel development, but it is believed that they have the same origin as stomata and, therefore, the same gene regulation. The trio of basic helix–loop–helix (bHLH) transcription factors, *SPEECHLESS* (*SPCH*), *MUTE*, and *FAMA*, identified in *Arabidopsis*, are essential for stomatal formation (Pillitteri *et al.*, 2007; Pillitteri *et al.*, 2008) and they form a single clade (Pires and Dolan, 2010; MacAlister and Bergmann, 2011). Two orthologues were found in our database that are included in the same bHLH clade and are more highly expressed in BQC (Fig. 4). We can consider that the genes found in *Q. suber* are orthologues of the *bHLH SPCH*, *FAMA*, and *MUTE* because within the angiosperms, putative bHLH orthologues can be identified due to the high degree of sequence conservation of its unique domain (MacAlister and Bergmann, 2011). The *HD-ZIPIV* genes *HOMEODOMAIN GLABROUS2* (*HDG2*) and *MERISTEM LAYER1* (*ML1*) are involved in stomatal differentiation (Peterson *et al.*, 2013) and the orthologue found in our database is also higher regulated in BQC. In order to prevent desiccation of the above-ground organs, plants produce protective layers such as waxes and suberin, but these prevent evaporation and limit the uptake of carbon dioxide from atmosphere. These problems are overcome by the production of stomata on epidermal structures (Pillitteri and Torii, 2007) and lenticels (Langenfeld-Heyser, 1997; Lendzian, 2006). As described in the Results section,

phellem (cork) is much more interspersed with lenticels in BQC than GQC. Surprisingly, despite lenticels being channels that facilitate gas and water exchange (Groh *et al.*, 2002), corticular tissue regions below lenticels work as a shade tissue where light does not penetrate easily (Manetas and Pfanz, 2005), due to its filling. The idea that more lenticels leads to an accumulation of free phenolics on its inner space could come from the fact that flavonoid and hydrocinnamic acid metabolites usually accumulate in the central vacuoles of guard cells and epidermal cells of stomata (Weissenbock *et al.*, 1986; Schnitzler *et al.*, 1996), the founder structures for lenticel formation, and so stimulate the accumulation of phenolics in lenticular channels. This could be one of the reasons for the higher amount of free phenolics present in BQC.

Defence mechanisms of Q. suber

Throughout evolution, each living organism has developed mechanisms that allow them to cope with unfavourable environmental conditions such as drought, high or low temperatures, and increased UV radiation. Many plants perceive UV radiation as stress leading to the synthesis of UV-absorbing compounds (Hahlbrock and Scheel, 1989) such as soluble phenols (Jansen *et al.*, 1998; Jansen, 2002; Gitz *et al.*, 2004). Our results showed an accumulation of free flavonoids and higher regulation of *CHS*, *PAL*, and *4CL* transcripts involved in the phenylpropanoid pathway in BQC samples. The phenylpropanoid pathway has several downstream branches leading to functional, distinct end-products such as auxin, tannins, suberin, and lignin (Tohge *et al.*, 2013). The first reaction of the phenylpropanoid pathway is promoted by PAL and the second step is carried out by the C4H that is induced by abiotic factors such as light, elicitors, and wounding (Batard *et al.*, 1997; Bell-Lelong *et al.*, 1997). In our transcriptome, the C4H enzyme is more highly expressed in BQC, while PAL that is induced mainly by biotic factors (Lawton and Lamb, 1987) is more highly expressed in GQC. PAL also initiates the phenylpropanoid pathway responsible for the poly-aromatic monomers of suberin and lignin. In fact, if we alter the action of PAL, the synthesis of suberin and lignin is affected (Gitz *et al.*, 2004).

It is at the end of the different sub-pathways (anthocyanins and scopolin) that the enzymes are most highly expressed in BQC. In *Arabidopsis*, hydroxycinnamic acid esters (HCAEs) are more effective UV protectants than flavonoids (Burchard *et al.*, 2000). In our data, the phenylpropanoid pathway comprising the HCAEs has key genes that are more highly expressed in BQC, such as *C4H (P450)*, *COMT* or *CCR*. With the exception of *Testa Transparent 15* (*TT15*), all other *TT* genes are highly expressed in GQC. This class of genes is involved in UV protection and they are also responsible for wax production (Caldwell, 1983), a component of suberin.

It has been demonstrated that UV-B radiation inhibits plant growth by regulating cell cycle progress (Nawkar *et al.*, 2013), which might be a protective mechanism to prevent cells with damaged DNA from dividing (Jansen, 2002). In our transcript database we do not find differentially expressed cyclin-related genes, but we do see that histone genes are highly expressed in BQC. Our gene ontology analysis shows that proteins that vary during the cell cycle, including cytoskeletal proteins, are also more highly expressed in BQC, a phenomenon also observed by Klessig *et al.* (2000) and Catterou *et al.* (2001) as defence signalling responses in *Arabidopsis* and maize (*Zea mays*).

Another defence mechanism that seems to be occurring in BQC is the upregulation of genes implicated in oxidative stress, such as GSTs and superoxide dismutases; both increased in our experiments and in the work reported by Casati and Walbot (2003) in maize in response to UV light. UV-B is able to stimulate the production of reactive oxygen species (ROS) (Dai *et al.*, 1997) and antioxidant defences (Jansen *et al.*, 1998). We can predict a higher content of ROS molecules in BQC because GST genes and especially superoxide dismutase genes are more highly expressed in cork of this quality. The role of GSTs and superoxide dismutase helps to maintain redox homeostasis in cells and also scavenge ROS (Schurmann and Jacquot, 2000), protecting cell membranes against oxidative stress and DNA against radiation (Schmidt and Dringen, 2012). The upregulation of these classes of genes in response to UV-B exposure was also reported by Arnots and Murphy (1991) and Dai *et al.* (1997). Moreover, we have observed that a gene coding for arginine decarboxylase was significantly more highly expressed in BQC (Fig. 4). This gene is involved in the biosynthesis of polyamines, an important growth regulator in response to abiotic stress (Alcázar *et al.*, 2010). In addition, three genes coding for *S*-adenosylmethionine synthetase are more highly expressed in BQC (Supplementary Table S2). This enzyme provides the precursor *S*-adenosylmethionine for ethylene and polyamine biosynthesis.

Conclusion

High-throughput analyses are opening doors to further and numerous fine-tuned developmental studies, such as that reported here. The multivariate analysis of all genes showed that only 5% of those expressed can be associated with differentiation of cork of good or bad quality. The genes associated with trees producing good- or bad-quality cork are candidate biomarkers for regulation of gene expression (transcription factors), and primary and secondary metabolism, associated with cork quality.

Supplementary material

Supplementary data can be found at *JXB* online.

Supplementary Table S1. A list of the PCR primers used for real-time RT-PCR and correlation coefficient R2 between pyrosequencing and RT-RTPCR.

Supplementary Table S2. List of genes distributed into different gene ontology categories.

Supplementary Table S3. Information on principal components.

Supplementary Figure S1. Alignment of five bHLH proteins.

Supplementary Figure S2. Alignment of three HD-ZIP IV proteins.

Funding

This work was supported by the Fundação para a Ciência e a Tecnologia, Ministério da Educação e Ciência, Portugal (project PTDC/AGRAAM/100465/2008). R. T. Teixeira and C. Pinheiro were supported by the programme Ciência2007 and A. M. Fortes by Ciência2008.

Acknowledgements

We thank Dr Isabel Miranda and Joaquina Martins for the phenol determination, Adelaide Machado for all the help in the field while collecting material. We also thank Biocant for solving all our doubts related with transcriptome analysis.

References

Alagna F, D'Agostino N, Torchia L, Servili M, Rao R, Pietrella M, Giuliano G, Chiusano ML, Baldoni L, Perrotta G. 2009. Comparative 454 pyrosequencing of transcripts from two olive genotypes during fruit development. *BMC Genomics* **10,** 399.

Alcázar R, Altabella T, Marco F, Bortolotti C, Reymond M, Koncz C, Carrasco P, Tiburcio AF. 2010. Polyamines: molecules with regulatory functions in plant abiotic stress tolerance. *Planta* **231,** 1237–1249.

Andersen JR, Lübberstedt T. 2003. Functional markers in plants. *Trends in Plant Science* **8,** 554–560.

Arnots T, Murphy TM. 1991. A comparison of the effects of a fungal elicitor and ultraviolet a radiation on ion transport and hydrogen peroxide in rose cells. *Environmental and Experimental Botany* **31,** 209–216.

Ashburner M, Ball CA, Blake JA *et al.* 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature Genetics* **25,** 25–29.

Barakat A, DiLoreto DS, Zhang Y, Smith C, Baier K, Powell WA, Wheeler N, Sederoff R, Carlson JE. 2009. Comparison of the transcriptomes of American chestnut (Castanea dentata) and Chinese chestnut (Castanea mollissima) in response to the chestnut blight infection. *BMC Plant Biology* **9,** 51.

Barel G, Ginzberg I. 2008. Potato skin proteome is enriched with plant defence components. *Journal of Experimental Botany* **59,** 3347–3357.

Batard Y, Schalk M, Pierrel MA, Zimmerlin A, Durst F, Werck-Reichhart D. 1997. Regulation of the cinnamate 4-hydroxylase (CYP73A1) in Jerusalem artichoke tubers in response to wounding and chemical treatments. *Plant Physiology* **113,** 951–959.

Beifang N, Limin F, Shulei S, Weizhong L. 2010. Artificial and natural duplicates in pyrosequencing reads of metagenomic data. *BMC Bioinformatics* **11,** 187–198.

Beisson F, Li Y, Bonaventure G, Pollard M, Ohlrogge JB. 2007. The acyltransferase GPAT5 is required for the synthesis of suberin in seed coat and root of Arabidopsis. *Plant Cell* **19,** 351–368.

Bell-Lelong DA, Cusumano JC, Meyer K, Chapple C. 1997. Cinnamate-4-hydroxylase expression in Arabidopsis. Regulation in response to development and the environment. *Plant Physiology* **113,** 729–738.

Benveniste P. 2004. Biosynthesis and accumulation of sterols. *Annual Review of Plant Biology* **55,** 429–457.

Bernards MA. 2002. Demystifying suberin. *Canadian Journal of Botany* **80,** 227–240.

Boerjan W, Ralph J, Baucher M. 2003. Lignin biosynthesis. *Annual Review of Plant Biology* **54,** 519–546.

Buchner J. 1999. Hsp90 & Co. - a holding for folding. *Trends in Biochemical Sciences* **24,** 136–141.

Burchard P, Bilger W, Weissenböck G. 2000. Contribution of hydroxycinnamates and flavonoids to epidermal shielding of UV-A and UV-B radiation in developing rye primary leaves as assessed by ultraviolet-induced chlorophyll fluorescence measurements. *Plant, Cell and Environment* **23,** 1373–1380.

Busch W, Lohmann J. 2007. Profiling a plant: expression analysis in Arabidopsis. *Current Opinion in Plant Biology* **10,** 136–141.

Caldwell MM. 1983. Internal filters: prospects for UV-acclimation in higher plants. In: Robberecht R, ed. *Physiologia plantarum* **58,** 445–450.

Caritat A, Gutiérrez E, Molinas M. 2000. Influence of weather on cork-ring width. *Tree Physiology* **20,** 893–900.

Casati P, Walbot V. 2003. Gene expression profiling in response to ultraviolet radiation in maize genotypes with varying flavonoid content. *Plant Physiology* **132,** 1739–1754.

Catterou M, Dubois F, Schaller H, Aubanelle L, Vilcot B, Sangwan-Norreel BS, Sangwan RS. 2001. Brassinosteroids, microtubules and cell elongation in Arabidopsis thaliana. II. Effects of brassinosteroids on microtubules and cell elongation in the bul1 mutant. *Planta* **212,** 673–683.

Chaves I, Pinheiro C, Paiva JA, Planchon S, Sergeant K, Renaut J, Graça JA, Costa G, Coelho AV, Ricardo CP. 2009. Proteomic evaluation of wound-healing processes in potato (Solanum tuberosum L.) tuber tissue. *Proteomics* **9,** 4154–4175.

Chen W, Provart NJ, Glazebrook J *et al.* 2002. Expression profile matrix of Arabidopsis transcription factor genes suggests their putative functions in response to environmental stresses. *The Plant Cell* **14,** 559–574.

Conde E, Cadahía E, García-Vallejo MC, de Simón BF. 1998. Polyphenolic composition of *Quercus suber* cork from diferent Spanish provenances. *Journal of Agricultural and Food Chemistry* **46,** 3166–3171.

Costa A, Pereira H, Oliveira A. 2002. Influence of climate on the seasonality of radial growth of cork oak during a cork production cycle. *Annals of Forest Science* **59,** 429–437.

Dai Q, Yan B, Huang S, Liu X, Peng S, Miranda MLM, Chavez AQ, vegara BS, Olszyk D. 1997. Response to oxidative stress defense system in rice (Oryza sativa) leaves with supplemental UV-B radiation. *Physiologia Plantarum* **101,** 301–308.

Dixon RA, Achnine L, Kota P, Liu CJ, Reddy MS, Wang L. 2002. The phenylpropanoid pathway and plant defence - a genomics perspective. *Molecular Plant Pathology* **3,** 371–390.

Dixon RA, Paiva NL. 1995. Stress-induced phenylpropanoid metabolism. *The Plant Cell* **7,** 1085–1097.

Edwards R, Dixon DP, Walbot V. 2000. Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health. *Trends in Plant Science* **5,** 193–198.

Gerhardt KE, Wilson MI, Greenberg BM. 1999. Tryptophan photolysis leads to a UVB-induced 66 kDa photoproduct of ribulose-1,5-biphosphate caboxylase/oxygenase (Rubisco) *in vitro* and *in vivo*. *Photochemistry and Photobiology* **70,** 49–56.

Gibson LJ, Easterling KE, Ashby MF. 1981. The structure and mechanics of cork. *Proceedings of the Royal Society of London* **A377,** 99–117.

Gitz DC, Liu-Gitz L, McClure JW, Huerta AJ. 2004. Effects of a PAL inhibitor on phenolic accumulation and UV-B tolerance in Spirodela intermedia (Koch.). *Journal of Experimental Botany* **55,** 919–927.

Gouzy J, Carrere S, Schiex T. 2009. FrameDP: sensitive peptide detection on noisy matured sequences. *Bioinformatics* **25,** 670–671.

Graça J, Pereira H. 1997. Cork suberin: a glyceryl based polyester. *Holzforschung* **51,** 225–234.

Groh B, Hübner C, Lendzian KJ. 2002. Water and oxygen permeance of phellems isolated from trees: the role of waxes and lenticels. *Planta* **215,** 794–801.

Hahlbrock K, Scheel D. 1989. Physiology and molecular biology of phenylpropanoid metabolism. *Annual Review of Plant Physiology and Plant Molecular Biology* **40,** 347–369.

Höfer R, Briesen I, Beck M, Pinot F, Schreiber L, Franke R. 2008. The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid omega-hydroxylase involved in suberin monomer biosynthesis. *Journal of Experimental Botany* **59,** 2347–2360.

Hose E, Clarkson DT, Steudle E, Schreiber L, Hartung W. 2001. The exodermis: a variable apoplastic barrier. *Journal of Experimental Botany* **52,** 2245–2264.

Hunter S, Apweiler R, Attwood TK *et al.* 2009. InterPro: the integrative protein signature database. *Nucleic Acids Research* **37,** D211–215.

Jansen MAK. 2002. Ultraviolet-B radiation effects on plants: induction of morphogenetic responses. *Physiologia Plantarum* **116,** 423–429.

Jansen MAK, Gaba V, Greenberg BM. 1998. Higher plants and UV-B radiation: balancing damage, repair and acclimation. *Trends in Plant Science* **3,** 131–135.

Jofré A, Molinas M, Pla M. 2003. A 10-kDa class-CI sHsp protects E.coli from oxidative and high-temperature stress. *Planta* **217,** 813–819.

Kalachanis D, Psaras GK. 2007. Structural changes in primary lenticels of Olea europaea and Cercis siliquastrum during the year. *IAWA Journal* **28,** 445–455.

Klessig DF, Durner J, Noah R *et al.* 2000. Nitric oxide and salicylic acid signaling in plant defense. *Proceedings of the National Academy of Sciences, USA* **97,** 8849–8855.

Kumar S, Blaxter ML. 2010. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* **11,** 571.

Landry LG, Stapleton AE, Lim J, Hoffman P, Hays JB, Walbot V, Last RL. 1997. An Arabidopsis photolyase mutant is hypersensitive to ultraviolet-B radiation. *Proceedings of the National Academy of Sciences, USA* **94,** 328–332.

Langenfeld-Heyser R. 1997. Physiological functions of lenticels. In: Rennenberg H, Eschrich W, Ziegler H, eds. *Trees - contributions to modern tree physiology*. Leiden, The Netherlands: Backhuys Publishers, 43–56.

Langmead B, Hansen KD, Leek JT. 2010. Cloud-scale RNA-sequencing differential expression analysis with Myrna. *Genome Biology* **11,** R83.

Laule O, Fürholz A, Chang HS, Zhu T, Wang X, Heifetz PB, Gruissem W, Lange M. 2003. Crosstalk between cytosolic and plastidial pathways of isoprenoid biosynthesis in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences, USA* **100,** 6866–6871.

Lawton MA, Lamb CJ. 1987. Transcriptional activation of plant defense genes by fungal elicitor, wounding and infection. *Molecular Cell Biology* **7,** 335–341.

Lendzian KJ. 2006. Survival strategies of plants during secondary growth: barrier properties of phellems and lenticels towards water, oxygen, and carbon dioxide. *Journal of Experimental Botany* **57,** 2535–2546.

Lopez-Matas MA, Nuñez P, Soto A, Allona I, Casado R, Collada C, Guevara MA, Aragoncillo C, Gomez L. 2004. Protein cryoprotective activity of a cytosolic small heat shock protein that accumulates constitutively in chestnut stems and is up-regulated by low and high temperatures. *Plant Physiology* **134,** 1708–1717.

Lottaz C, Iseli C, Jongeneel CV, Bucher P. 2003. Modeling sequencing errors by combining Hidden Markov models. *Bioinformatics* **19** Suppl. 2, 103–112.

MacAlister CA, Bergmann DC. 2011. Sequence and function of basic helix-loop-helix proteins required for stomatal development in Arabidopsis are deeply conserved in land plants. *Evolution and Development* **13,** 182–192.

Manetas Y, Pfanz H. 2005. Spatial heterogeneity of light penetration through periderm and lenticels and concomitant patchy acclimation of corticular photosynthesis. *Trees* **19,** 409–414.

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. *Journal of Analytical and Applied Pyrolysis* **100,** 88–94.

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.

Millar AA, Clemens S, Zachgo S, Giblin EM, Taylor DC, Kunst L. 1999. CUT1, an Arabidopsis gene required for cuticular wax biosynthesis and pollen fertility, encodes a very-long-chain fatty acid condensing enzyme. *The Plant Cell* **11,** 825–838.

Molina I, Li-Beisson Y, Beisson F, Ohlrogge JB, Pollard M. 2009. Identification of an Arabidopsis feruloyl-coenzyme A transferase required for suberin synthesis. *Plant Physiology* **151,** 1317–1328.

Morozova O, Marra M. 2008. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92,** 255–264.

Nakashima K, Ito Y, Yamaguchi-Shinozaki K. 2009. Transcriptional regulatory networks in response to abiotic stresses in Arabidopsis and grasses. *Plant Physiology* **149,** 88–95.

Nawkar GM, Maibam P, Park JH, Sahi VP, Lee SY, Kang CH. 2013. UV-induced cell death in plants. *International Journal of Molecular Sciences* **14,** 1608–1628.

Noah JW, Shapkina T, Wollenzien P. 2000. UV-induced crosslinks in the 16S rRNAs of Escherichia coli, Bacillus subtilis and Thermus aquaticus and their implications for ribosome structure and photochemistry. *Nucleic Acids Research* **28,** 3785–3792.

Novaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR, Kirst M. 2008. High-throughput gene and SNP discovery in Eucalyptus grandis, an uncharacterized genome. *BMC Genomics* **9,** 312.

O'Rourke JA, Nelson RT, Grant D, Schmutz J, Grimwood J, Cannon S, Vance CP, Graham MA, Shoemaker RC. 2009. Integrating microarray analysis and the soybean genome to understand the soybeans iron deficiency response. *BMC Genomics* **10,** 376.

Parchman TL, Geist KS, Grahnen JA, Benkman CW, Buerkle CA. 2010. Transcriptome sequencing in an ecologically important tree species: assembly, annotation, and marker discovery. *BMC Genomics* **11,** 180.

Penninckx IA, Eggermont K, Terras FR, Thomma BP, De Samblanx GW, Buchala A, Métraux JP, Manners JM, Broekaert WF. 1996. Pathogen-induced systemic activation of a plant defensin gene in Arabidopsis follows a salicylic acid-independent pathway. *The Plant Cell* **8,** 2309–2323.

Pereira H. 1981. Dosage des tanins du liège de *Quercus suber* L. *Anais Instituto Superior de Agronomia* **XL,** 9–15.

Pereira H. 1988. Chemical composition and variability of cork form *Quercus suber* L. *Wood Science and Technology* **22,** 211–218.

Pereira H. 2007. *Cork: biology, production and uses*. Amsterdam: Elsevier.

Pereira H, Lopes F, Graça J. 1996. The evaluation of the quality of cork planks by image analysis. *Holzforschung* **50,** 111–115.

Peterson KM, Shyu C, Burr CA, Horst RJ, Kanaoka MM, Omae M, Sato Y, Torii KU. 2013. Arabidopsis homeodomain-leucine zipper IV proteins promote stomatal development and ectopically induce stomata beyond the epidermis. *Development* **140,** 1924–1935.

Pillitteri LJ, Bogenschutz NL, Torii KU. 2008. The bHLH protein, MUTE, controls differentiation of stomata and the hydathode pore in Arabidopsis. *Plant Cell Physiology* **49,** 934–943.

Pillitteri LJ, Sloan DB, Bogenschutz NL, Torii KU. 2007. Termination of asymmetric cell division and differentiation of stomata. *Nature* **445,** 501–505.

Pillitteri LJ, Torii KU. 2007. Breaking the silence: three bHLH proteins direct cell-fate decisions during stomatal development. *Bioessays* **29,** 861–870.

Pires N, Dolan L. 2010. Origin and diversification of basic-helix-loop-helix proteins in plants. *Molecular Biology and Evolution* **27,** 862–874.

Pla M, Huguet G, Verdaguer D, Puigderrajols P, Llompart B, Nadal A, Molinas M. 1998. Stress proteins co-expressed in suberized and lignified cells and in apical meristems. *Plant Science* **139,** 49–57.

Puigderrajols P, Jofré A, Mir G, Pla M, Verdaguer D, Huguet G, Molinas M. 2002. Developmentally and stress-induced small heat shock proteins in cork oak somatic embryos. *Journal of Experimental Botany* **53,** 1445–1452.

Ricardo CP, Martins I, Francisco R, Sergeant K, Pinheiro C, Campos A, Renaut J, Fevereiro P. 2011. Proteins associated with cork formation in Quercus suber L. stem tissues. *Journal of Proteomics* **74,** 1266–1278.

Richter K, Buchner J. 2001. Hsp90: chaperoning signal transduction. *Journal of Cell Physiology* **188,** 281–290.

Sánchez-González M, Cañellas I, Montero G. 2008. Base-age invariante cork growth model for Spanish cork oak (*Quercus suber* L) forests. *European Journal of Forest Research* **127,** 173–182.

Schmidt MM, Dringen R. 2012. GSH synthesis and metabolism. In: Gruetter R, Choi I, eds. *Advances in neurobiology*, **Vol. 4**. New York: Springer, 1029–1050.

Schnitzler J-P, Jungblut TP, Heller W, Hutzler P, Heinzmann U, Schmelzer E, Ernst D, Langebartels C, Sandermann H. 1996. Tissue localisation of UV-B screening pigments and chalcone synthase mRNA in Scots pine (Pinus sylvestris L) needles. *New Phythologist* **132,** 247–258.

Schurmann P, Jacquot J. 2000. Plant thioredoxin systems revisited. *Annual Review of Plant Physiology and Plant Molecular Biology* **51,** 371–400.

Soler M, Serra O, Molinas M, García-Berthou E, Caritat A, Figueras M. 2008. Seasonal variation in transcript abundance in cork tissue analyzed by real time RT-PCR. *Tree Physiology* **28,** 743–751.

Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M. 2007. A genomic approach to suberin biosynthesis and cork differentiation. *Plant Physiology* **144,** 419–431.

Soto A, Allona I, Collada C, Guevara MA, Casado R, Rodriguez-Cerezo E, Aragoncillo C, Gomez L. 1999. Heterologous expression of a plant small heat-shock protein enhances Escherichia coli viability under heat and cold stress. *Plant Physiology* **120,** 521–528.

Strable J, Borsuk L, Nettleton D, Schnable PS, Irish EE. 2008. Microarray analysis of vegetative phase change in maize. *The Plant Journal* **56,** 1045–1057.

Sung DY, Kaplan K, Guy CL. 2001. Plant Hsp70 molecular chaperones: protein structure, gene family, expression and function. *Physiologia Plantarum* **113,** 443–451.

The R Development Core Team. 2011. *A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Thioulouse J, Dray S. 2007. Interactive multivariate data analysis in R with the ade4 and ade4TkGUI packages. *Journal of Statistical Software* **22,** 1–14.

Tohge T, Watanabe M, Hoefgen R, Fernie AR. 2013. Shikimate and phenylalanine biosynthesis in the green lineage. *Frontiers in Plant Science* **4,** 62.

Usadel B, Poree F, Nagel A, Lohse M, Czedik-Eysenberg A, Stitt M. 2009. A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant, Cell and Environment* **32,** 1211–1229.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. *Molecular Ecology* **17,** 1636–1647.

Verdaguer D, Aranda X, Jofré A, El Omari M, Molinas M, Fleck I. 2003. Expression of low molecular weight heat-shock proteins and total antioxidant activity in the Mediterranean tree *Quercus ilex* L. in relation to seasonal and diurnal changes in physiological parameters. *Plant, Cell and Environment* **26,** 1407–1417.

Vogt T. 2010. Phenylpropanoid biosynthesis. *Molecular Plant* **3,** 2–20.

Wang W, Vinocur B, Shoseyov O, Altman A. 2004. Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends in Plant Science* **9,** 244–252.

Weissenbock G, Hedrich R, Sachs H. 1986. Secondary phenolic products in isolated guard cell, epidermal cell and mesophyll cell protoplasts - distribution and determination. *Protoplasma* **134,** 141–148.

Wutz A. 1955. Anatomische untersuchungen uber system und periodische veranderunger der lenticellen. *Botanische Studien* **4,** 43–72.

Young JC, Moarefi I, Hartl FU. 2001. Hsp90: a specialized but essential protein-folding tool. *Journal of Cell Biology* **154,** 267–273.