

Molecular regulation of secondary growth: searching for *SHORT-ROOT* function in cork formation

Andreia Lúcia Campos dos Santos Ferreira Miguel



Dissertation presented to obtain the Ph.D degree in Biology
Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras,
June, 2014



INSTITUTO
DE TECNOLOGIA
QUÍMICA E BIOLÓGICA
ANTÓNIO XAVIER / UNL

Knowledge Creation



Molecular regulation of secondary growth: searching for *SHORT-ROOT* function in cork formation

Andreia Lúcia Campos dos Santos Ferreira Miguel

Dissertation presented to obtain the Ph.D degree in Biology
Instituto de Tecnologia Química e Biológica António Xavier | Universidade Nova de Lisboa

Oeiras, June, 2014

WORK PERFORMED AT:

Forest Biotechnology Laboratory

Instituto de Biologia Experimental e Tecnológica

Instituto de Tecnologia Química e Biológica

Universidade Nova de Lisboa

Av. da República

2780-157 Oeiras

Portugal

SUPERVISOR:

Dr. Célia Maria Romba Rodrigues Miguel

Principal Investigator, IBET and ITQB-UNL

Aos meus pais

"Learn from yesterday, live for today, hope for tomorrow. The important thing is to not stop questioning"

Albert Einstein

ACKNOWLEDGEMENTS

I would like to thank Célia Miguel for the opportunity she offered me to be a member of her lab. For being not only my supervisor and advisor but also for her friendship throughout these years. I am truly grateful for all you have taught me, for your good advices, your encouragement and for always believing and making me believe that I would be able to fulfill my goals. Thanks for all the availability during the writing of the thesis, for all the suggestions, improvements, for always being optimistic and for all your dedication. In a few words, Thank You Very Much!

To Brian Jones for his availability and good advice. For being critical when necessary and for all the words of encouragement.

To Professor Pinto Ricardo and Professor Margarida Oliveira for having accepted to be part of my PhD thesis committee. Thank you for the availability, for the words of encouragement and for always caring about my work.

To all my colleagues who worked with me. To Liliana Marum for all the support, for her good mood and friendship. Thank you for being my partner in the discovery of the “Cork oak world”. No one like you to understand the difficulties of working with this so peculiar tree species. To Ana Milhinhos, for all the debates and scientific discussions in the “Universe of *Populus*”. Thanks for all the companionship and friendship and for your dedication. To José de Vega-Bartol for being my partner in the “transcriptomics”. Thank you for all the patience in the bioinformatics and for the availability.

To all current and former members of the Forest Biotech lab. To Andreias: Andreia Matos, my partner at the bench and at the laminar flow chamber,

thanks for making the endless hours at the flow chamber a bit shorter in your company and for your friendship; to Andreia Rodrigues for your companionship and friendship. Thank you for always caring about my work. To Inês Chaves, Inês Modesto, Ilanit, Raissa, Mariagrazia, Ana Maria, Marta Simões, Marta Madeira. Thank you all for your companionship, good mood and for the team spirit.

To everyone at the GPS lab, for all your kindness and for always being helpful.

To Eugénia, Pilar, Sónia and all the staff at the washing rooms. Without your help the *Populus* subcultures would have taken so much longer.

To everyone at the ITQB and IBET that in some way helped me over the years and to Ana Maria Portocarrero and Fátima Madeira for all the help, especially in the last steps of thesis preparation and delivery.

Um especial obrigada aos meus pais por sempre me terem apoiado. Não há palavras que descrevam a importância do vosso apoio. Obrigada pela paciência e pelas palavras de conforto, por todo o incentivo e confiança. O vosso amor e os vossos ensinamentos foram, são e sempre serão fundamentais.

Ao Ricardo por ser o meu melhor amigo e companheiro. Obrigada por toda a compreensão, companheirismo e paciência. Obrigada por teres estado sempre presente, por todo o carinho e incentivo. Esta é mais uma das etapas ultrapassadas a dois!

A toda a minha família. Em especial aos meus avós por todo o amor, carinho e preocupação, ao meu irmão e cunhada por todo o apoio e incentivo e por

me terem dado dois sobrinhos lindos que me têm proporcionado verdadeiros momentos de alegria e de distração.

Aos pais do Ricardo, à irmã e avós, por todo o incentivo e palavras de conforto. Senti-me como uma filha, irmã e neta.

Aos meus amigos por me terem proporcionado momentos de descontração. Obrigada por terem estado presentes nesta fase.

Um muito obrigada a todos, sem vocês teria sido muito mais difícil!

LIST OF ABBREVIATIONS

ACT	<i>ACTIN</i>
At	<i>Arabidopsis thaliana</i>
BCIP	5-bromo-4-chloro-3'-indolylphosphate
bp	Base pair
CAC	<i>CLATHRIN ADAPTOR COMPLEX</i>
CaMV 35S	Cauliflower mosaic virus 35S promoter
cDNA	Complementary DNA
CEI	Cortex/endodermal initial
CK	Cytokinin
CKX	<i>CYTOKININ OXIDASE</i>
cm	Centimetre
Cq	Quantification cycle
CODB	Cork oak database
COG	Clusters of Orthologous Groups
CV	Coefficient of variation
DAG	Days after germination
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
DE	Differentially expressed
DEG	Differentially expressed gene
E	Efficiency
EN	Internode
EST	Expressed Sequence Tag
FAA	Formaldehyde, Acetic acid, Ethanol
FDR	False Discovery Rate
g	Gram
GO	Gene ontology
GUS	β -Glucuronidase

h	Hour
HK	Housekeeping
HD-ZIP III	Class III homeodomain leucine zipper
kb	Kilobase
KEGG	Kyoto Encyclopedia of Genes and Genomes
M	Stability value
Mbp	Mega base pair
min	Minute
miRNA	MicroRNA
mm	Millimeter
mM	Millimolar
mRNA	Messenger RNA
NBT	Nitro blue tetrazolium
nM	Nanomolar
NF	Normalization factor
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
Pg	Phellogen
Ph	Phellem
QC	Quiescent center
Qi	<i>Quercus ilex</i>
Qs	<i>Quercus suber</i>
RAM	Root apical meristem
RG	Reference gene
RNA	Ribonucleic acid
RNA-Seq	RNA sequencing
RR7	A-type response regulator; Cytokinin primary response gene
RT-qPCR	Reverse Transcription Quantitative Polymerase Chain Reaction
s	second

SAM	Shoot apical meristem
SCR	SCARECROW
<i>scr</i>	<i>scarecrow</i> mutant
SHR	SHORT-ROOT
<i>shr</i>	<i>short-root</i> mutant
TF	Transcription factor
<i>uidA</i>	Encodes the β -glucuronidase enzyme
VC	Vascular cambium
$V_{n/n+1}$	Pairwise variation
WT	Wild-type
X-Gluc	5-bromo-4-chloro-3-indolyl β -D-glucuronide
$^{\circ}\text{C}$	Celsius degrees
μg	Microgram
μl	Microliter
μm	Micrometer
μM	Micromolar
ϕ	Diameter

SUMMARY

The analysis of molecular regulators involved in controlling the maintenance and function of plant meristems has been the subject of many studies. Some master regulators of these processes have been identified in *Arabidopsis* benefiting from the array of tools available for genetic and molecular analysis in this model plant. However, aspects such as secondary growth that are more extensively observed in woody plants, have been less studied. Secondary growth is responsible for the enlargement of the plant stems and roots and results from the activity of the lateral (secondary) meristems, vascular cambium and cork cambium (phellogen), which produce two important renewable natural resources, wood and cork, respectively. Based on studies conducted in the root and shoot apical meristems and the vascular cambium, overlapping or related molecular mechanisms have been found to regulate both primary and secondary meristems. However, very few studies have focused on phellogen activity and therefore, the molecular mechanisms regulating this secondary meristem are not yet known. *SHORT-ROOT (SHR)* is an example of a gene involved in the regulation of both primary meristems and vascular cambium. One of the main aims of this thesis was to investigate the role of *SHR* gene in order to provide a deeper knowledge of the mechanisms regulating secondary growth and search for a possible involvement in the regulation of the phellogen activity. The model tree *Populus* was used for this purpose, but additional studies have been pursued in *Quercus suber* (cork oak) due to its relevance in terms of cork production ability. In the *Populus* genome three *SHR*-like genes have been previously identified, *PtSHR1*, *PtSHR2A* and *PtSHR2B*. In this work, hybrid aspen plants were genetically manipulated to investigate the expression of *SHR*-like genes in *Populus* plants. *PtSHR2B* promoter was found to lead expression of the *GUS* reporter gene in several tissues of *in vitro* propagated plants including root tip, shoot apex, primary xylem and outer cell layers of the stem while in

greenhouse grown plants with well developed secondary growth, the promoter was active in the phellogen. Given the interesting localization of *SHR2B* transcripts in the phellogen, its function was further investigated through an overexpression approach. Evaluation of several growth parameters and anatomic analysis carried out in *PtSHR2B* overexpressing plants showed a negative effect in terms of plant growth and changes in stem anatomy. Furthermore, elements of the cytokinin metabolism and response were altered in the transgenic plants. The obtained results suggest that *PtSHR2B* is a negative regulator of plant growth and meristem activity, like its homolog *PtSHR1*, and its function is possibly mediated through the control of cytokinin homeostasis. Since cork oak is perhaps the most peculiar tree species concerning the ability to produce phellem (cork) as a result of phellogen activity, and the interest in acquiring basic knowledge on the regulation of phellogen in *Quercus suber*, two putative cork oak *SHR* homologs, *QsSHR1* and *QsSHR2*, were identified and further characterized. While *QsSHR1* transcripts, more similar to the *Arabidopsis SHR* (*AtSHR*) and *PtSHR1*, were found localized in the vascular cambium, *QsSHR2* transcripts were localized in the phellogen, indicating putative roles during secondary growth. For a proper measurement of the expression levels of each gene in different tissues, several putative cork oak reference genes were analysed to verify their suitability for quantitative PCR normalization. Seasonal regulation during periderm development was observed for both cork oak *SHR* genes. Nonetheless, *QsSHR2* seemed more expressed in the periderm than *QsSHR1* that was highly expressed in leaves. The localization of *QsSHR2* in the phellogen and its higher expression in the periderm of branches collected during a period of phellogen inactivity suggests that, like *PtSHR1* and possibly *PtSHR2B*, *QsSHR2* acts as a negative regulator of phellogen activity. Despite the differences in expression levels, no major variation was found in the expression patterns of both genes in periderm tissues of cork oak and holm oak (*QiSHR1/2*), a closely related species that does not produce a thick cork

layer. To further understand the role of the cork oak *SHR* genes, functional studies were performed in *Arabidopsis*. None of the cork oak gene sequences fully complemented the *Arabidopsis shr2* mutant phenotype, thus indicating that different *SHR*-like genes may have evolved different functions (neofunctionalization) or the function of the original gene may have been distributed between the copies (sub-functionalization) after genome duplication following the divergence of *Populus/Quercus* and *Arabidopsis* lineages. Altogether, the results obtained in the experiments conducted in *Populus* and cork oak, as well as previously published data, point to the involvement of *SHR* genes in the regulation of secondary growth, where *SHR1* appears to act in the regulation of vascular cambium and *SHR2* in the regulation of the phellogen.

Another aim of this thesis was the transcriptomic analysis of cork oak fruit development as part of a Portuguese effort to obtain a reference transcriptome of the cork oak. A fruit developmental staging system was established and RNA-Seq libraries were prepared from several fruit and embryo developmental stages. The analysis of sequencing data allowed to obtain a dynamic view of gene expression along development and to identify differentially expressed genes (DEGs) focusing on specific processes related to response to water, including water transport and water deprivation as well as on transcriptional regulation and carbohydrate metabolism. Nearly a quarter of the DEGs code for putative transcription factors associated to several biological processes reflecting a very active and tightly regulated transcriptional activity during fruit development. While the DEGs putatively coding for transcription factors were found almost equally distributed throughout the analysed developmental stages, the transcripts involved in response to water or carbohydrate metabolism were over-represented in particular stages. This data set represents a very useful tool towards understanding several aspects of the cork oak biology, including reproduction

and early development, that may reveal crucial for the implementation of successful regeneration programs for the species.

Altogether, the results here obtained provided novel insights into the regulation of specific post-embryonic developmental processes in cork oak, focusing on secondary growth and particularly the cork formation process, but also on fruit development, contributing new tools for a better knowledge of cork oak biology.

SUMÁRIO

A análise de reguladores moleculares envolvidos no controlo da manutenção e função dos meristemas das plantas tem sido alvo de diversos estudos. Em *Arabidopsis* já se identificaram alguns reguladores chave destes processos beneficiando de um conjunto de ferramentas disponíveis para análise genética e molecular nesta planta modelo. Contudo, aspetos como o crescimento secundário que é mais amplamente observado em plantas lenhosas têm sido menos estudados. O crescimento secundário é responsável pelo alargamento dos caules e raízes, resultando da atividade dos meristemas laterais (secundários), o câmbio vascular e o câmbio súbero-felodérmico (felogénio), que produzem dois importantes recursos naturais renováveis, a madeira e a cortiça, respetivamente. Com base em estudos conduzidos nos meristemas apicais radicular e caulinar assim como no câmbio vascular, tem-se verificado alguma sobreposição ou semelhança entre os mecanismos moleculares que regulam tanto os meristemas primários como os secundários. No entanto, poucos são os estudos que se têm centrado no funcionamento do felogénio e, por conseguinte, os mecanismos moleculares que regulam este meristema secundário não são ainda conhecidos. O gene *SHORT-ROOT (SHR)* é um exemplo de um regulador molecular de ambos os meristemas primários assim como do câmbio vascular. Um dos principais objetivos desta tese foi investigar a função do gene *SHR* de modo a obter um conhecimento mais detalhado dos mecanismos que regulam o crescimento secundário e analisar o possível envolvimento na regulação da atividade do felogénio. A árvore modelo *Populus* foi utilizada nestes estudos, mas foi também conduzido trabalho adicional em *Quercus suber* (sobreiro) devido à sua relevância em termos de capacidade de produção de cortiça. No genoma do *Populus* foram previamente identificados três genes muito semelhantes ao *SHR*, *PtSHR1*, *PtSHR2A* e *PtSHR2B*. Neste trabalho, foram geneticamente manipuladas

plantas de choupo (híbrido *Populus tremula* L. × *Populus tremuloides* Michx.) para investigar a expressão dos genes do tipo *SHR*. Observou-se que o promotor do gene *PtSHR2B* conduz a expressão do gene repórter *GUS* em vários tecidos de plantas propagadas *in vitro*, nomeadamente na extremidade da raiz, no ápice caulinar, no xilema primário e nas camadas celulares exteriores do caule, enquanto em plantas com crescimento secundário bem patente (crescimento em estufa), o promotor mostrou-se ativo no felogénio. Dada a interessante localização dos transcritos *SHR2B* no felogénio, a sua função foi investigada através de uma abordagem de sobre-expressão. A avaliação de vários parâmetros de crescimento e de aspetos anatómicos realizada em plantas com sobre-expressão do gene *PtSHR2B* revelou um efeito negativo em termos do crescimento da planta e alterações na anatomia do caule. Além disso, elementos envolvidos no metabolismo e resposta às citocininas também sofreram alterações nas plantas transgénicas. Os resultados obtidos sugerem que o gene *PtSHR2B* é um regulador negativo do crescimento e da atividade meristemática, como o seu homólogo *PtSHR1*, e a sua função é possivelmente mediada através do controlo da homeostase das citocininas. Uma vez que o sobreiro é talvez a espécie mais peculiar no que respeita à capacidade de produzir felema (cortiça) como resultado da atividade do felogénio, e dado o interesse em adquirir conhecimentos básicos sobre a regulação do felogénio nesta espécie, foram identificados e caracterizados dois possíveis homólogos do gene *SHR* em sobreiro, *QsSHR1* e *QsSHR2*. Enquanto os transcritos do gene *QsSHR1*, mais semelhante ao gene em *Arabidopsis* (*AtSHR*) e ao *PtSHR1*, foram localizados no câmbio vascular, os transcritos do gene *QsSHR2* foram localizados no felogénio, indicando possíveis funções durante o crescimento secundário. Para uma medição correta dos níveis de expressão de cada gene em diferentes tecidos do sobreiro, foram analisados vários possíveis genes de referência de modo a verificar a sua adequação para normalizar os dados provenientes da quantificação de transcritos pela técnica de PCR em tempo-real. Ambos os

genes *SHR* de sobreiro demonstraram ter uma regulação sazonal durante o desenvolvimento da periderme. Todavia, o *QsSHR2* aparentou estar mais expresso na periderme do que o *QsSHR1* cuja expressão mais elevada foi observada nas folhas. A localização do gene *QsSHR2* no felogénio e a sua expressão mais elevada na periderme de ramos colhidos durante o período de inatividade do felogénio sugere que, tal como o *PtSHR1* e possivelmente o *PtSHR2B*, também o gene *QsSHR2* atua como regulador negativo da atividade do felogénio. Apesar da diferença nos níveis de expressão, não foi registada grande variação nos padrões de expressão de ambos os genes nos tecidos da periderme de sobreiro e azinheira, uma espécie estreitamente relacionada que não produz cortiça. Para uma melhor compreensão da função dos genes *SHR* de sobreiro, foram realizados estudos funcionais em *Arabidopsis*. Nenhuma das sequências de sobreiro complementou na totalidade o fenótipo do mutante *shr2* de *Arabidopsis*, indicando que diferentes genes do tipo *SHR* podem ter desenvolvido diferentes funções (neo-funcionalização) ou a função do gene original pode ter sido distribuída entre as várias cópias (sub-funcionalização) após a duplicação do genoma e a divergência das linhagens do *Populus/Quercus* e *Arabidopsis*. De um modo global, os resultados obtidos nas experiências realizadas em *Populus* e sobreiro, assim como dados já publicados, apontam para o envolvimento dos genes *SHR* na regulação do crescimento secundário, em que o *SHR1* parece intervir na regulação do câmbio vascular e o *SHR2* na regulação do felogénio. Um outro objetivo desta tese foi a análise transcritômica do desenvolvimento do fruto do sobreiro como parte de uma iniciativa Portuguesa para obter o transcrito de referência do sobreiro. Após o estabelecimento de um sistema para classificação do fruto em vários estádios de desenvolvimento prepararam-se bibliotecas de RNA-Seq correspondendo a diferentes estádios. A análise dos dados de sequenciação permitiu obter uma visão dinâmica da expressão dos genes ao longo do desenvolvimento e a identificação de genes diferencialmente expressos dando ênfase a processos específicos

relacionados com a resposta à água, incluindo o transporte de água e a privação de água, bem como com a regulação da transcrição e o metabolismo dos hidratos de carbono. Aproximadamente um quarto dos genes diferencialmente expressos codificam para possíveis fatores de transcrição associados a vários processos biológicos refletindo uma atividade transcricional muito ativa e fortemente regulada durante o desenvolvimento do fruto. Enquanto estes genes se encontram distribuídos quase igualmente ao longo de todos os estádios de desenvolvimento analisados, os transcritos envolvidos na resposta à água ou no metabolismo dos hidratos de carbono estavam sobre-representados em estádios específicos. Este conjunto de dados representa uma ferramenta muito útil para a compreensão de vários aspetos da biologia do sobreiro, nomeadamente ao nível da reprodução e desenvolvimento inicial, que se podem revelar cruciais para a implementação de programas de regeneração de sobreiro.

Em resumo, os resultados obtidos nesta tese contribuíram para um maior conhecimento dos mecanismos de regulação do desenvolvimento pós-embriónico em sobreiro, centrando-se no crescimento secundário, em particular na formação da cortiça, mas também no desenvolvimento do fruto, disponibilizando novas ferramentas para o conhecimento da biologia do sobreiro.

TABLE OF CONTENTS

<i>Acknowledgements</i>	v
<i>List of abbreviations</i>	ix
<i>Summary</i>	xiii
<i>Sumário</i>	xvii
Chapter I	
General Introduction	1
Chapter II	
A <i>SHORT-ROOT</i> -like gene (<i>PtSHR2B</i>) is involved in <i>Populus</i> phellogen activity	63
Chapter III	
Reference gene selection for quantitative real-time PCR normalization in <i>Quercus suber</i>	95
Chapter IV	
<i>SHORT-ROOT</i> -like genes are differentially regulated during secondary growth in <i>Quercus suber</i>	131
Chapter V	
Transcriptomic profiling during acorn development in cork oak: a contribute towards a reference transcriptome	175
Chapter VI	
Concluding Remarks and Future Perspectives	261

Chapter I

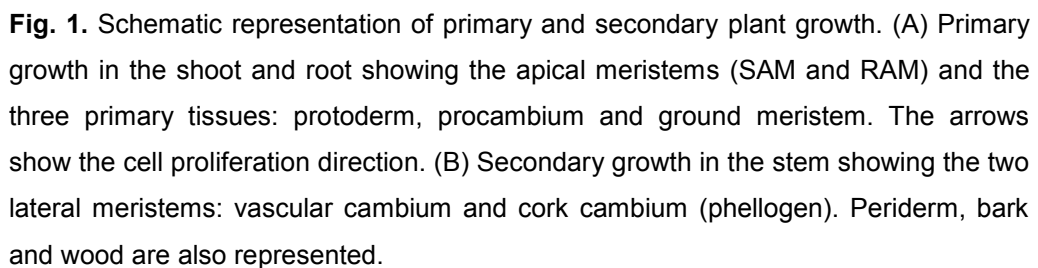
General Introduction

Plant meristems

Plants have an open body in the sense that they are able to form new organs and highly specialized tissues by post-embryonic development. This ability depends on the activity of specialized structures called meristems, in which pluripotent stem cells are localized. By exhibiting the capacity for self-renewal and for producing progeny that follow various differentiation pathways, stem cells are crucial for the development of multicellular organisms. While in animals stem cells are used mainly to compensate for cell loss during post-embryonic development, in plants they are used in a continuous way to generate new structures during the entire plant's life-span. These new structures include organs like roots, leaves, flowers, fruits and stems, but also specialized tissues formed during secondary growth such as wood and cork.

Based on their position in the plant body, meristems can be either apical or lateral (Esau, 1977). The apical meristems comprise the shoot and root apical meristems, termed as SAM and RAM, respectively, and are responsible for plant elongation through cell division that provides new cells for expansion, tissue differentiation or initiation of new organs (stems, leaves and roots). The cells produced in the apical meristems differentiate into primary tissues: protoderm, procambium and ground meristem (Fig. 1A). In the shoots the procambium will form the vascular tissues and the ground meristem will form the pith, whereas in the roots the procambium gives rise to the stele and the ground meristem to the cortex. The protoderm is responsible for the formation of the epidermis in both shoots and roots.

All plants undergo primary growth derived from the apical meristems but not all plants exhibit secondary growth. Secondary growth refers to the increase in girth (thickening) of plant stems and roots and results from the activity of the two lateral (or secondary) meristems, the vascular cambium and the cork cambium (or phellogen). Vascular cambium is involved in the formation of secondary vascular tissues, while cork cambium is responsible for



The development of vascular tissues was a major evolutionary novelty that made possible the colonization of land by plants. The specialized cells of

vascular tissues allow the transport of water and nutrients along the plant making them less dependent on a very humid environment. Structural support is another essential function of the plant vasculature allowing growth in height as a way to compete for light. While the xylem transports water and minerals, the phloem is associated with the transport of organic nutrients. In addition, the vascular tissues provide a long-distance communication system, delivering information in the form of hormones and other molecules, for the coordination of developmental and physiological processes. During primary shoot growth the phloem is typically positioned parallel to the xylem and both conducting tissues are organized into vascular bundles.

Primary xylem and phloem develop from the activity of the procambium, a primary meristem formed during early embryogenesis, with narrow and cytoplasm dense cells organized in continuous strands (Scarpella and Meijer, 2004). The vascular cambium derives from the procambium within the vascular bundles (fascicular cambium) and from the parenchyma between the vascular bundles (interfascicular cambium), forming a continuous cylinder around the plant stem. Two types of cell divisions occur in the vascular cambium. The anticlinal divisions produce new cambial initials and the periclinal divisions of cambial initials and its derivatives produce secondary vascular tissues through differentiation into secondary xylem (wood) to the inside of the plant stem or root, and secondary phloem to the outside (Lachaud *et al.*, 1999; Scarpella and Meijer, 2004), thus promoting growth in lateral directions (Esau, 1977). Cambial activity is seasonal and related to changes in temperature and humidity, and the length of the period of cambial activity depends on the species (Lachaud *et al.*, 1999; Evert, 2006a).

Xylem consists of both parenchyma and sclerenchyma cells. The xylem parenchyma cells have a storage function while the sclerenchyma cells provide mechanical support and are involved in defense and water transport. The sclerenchyma cells include the conducting tracheary elements (xylem tracheids and vessels) and xylem fibers.

Xylogenesis includes cell division and enlargement, cell wall thickening, lignification and programmed cell death (Fukuda, 1996; Myburg and Sederoff, 2001). Protoxylem cells are the first xylem cells to differentiate within the vascular bundles, developing in the innermost position in the shoot and next to the pericycle in the root, whereas the metaxylem cells differentiate later. Both xylem cell types can be distinguished based on their secondary cell wall. Protoxylem cells have ring-like (annular) or helical (spiral) cell wall thickenings, that can be stretched, making it possible to elongate in the growth direction, while metaxylem cells appear as net-like (reticulate) or porous (pitted) being more uniform and thicker than protoxylem and providing more support. During secondary cell wall thickening the deposition of substances like cellulose, lignin and hemicellulose occur.

Phloem is composed of sieve elements (sieve tubes and sieve cells) which are the conducting elements, and of non conducting cells such as fibers and sclereids, and parenchyma cells. Unlike the tracheary elements, the sieve elements are living cells and during differentiation sieve tubes lose most of their organelles, including nucleus (Raven *et al.*, 2005; Schuetz *et al.*, 2012). Fibers and sclereids are responsible for the mechanical support function and exhibit a rigid and very difficult to damage secondary cell wall.

Periderm development

While the vascular cambium is responsible for the formation of wood, the phellogen is responsible for periderm formation, which holds great importance as a protective barrier. During primary growth epidermis provides a barrier against dehydration but enables the exchange of gases, such as CO₂ and O₂, essential for photosynthesis and respiration. The cuticle, a waxy layer of cutin-containing polymers secreted from the epidermal cells, functions against water loss and reduces the risk of predation, whereas the stomata are essential for

the regulation of gas exchanges (Lendzian, 2006; Glover, 2010). An analogy between the cuticle-epidermis-stomata and the periderm has been made (Lendzian, 2006). During secondary growth the epidermis is replaced in the stems and roots by the periderm, resulting in an increased thickness of these organs (Evert, 2006*b*). It is composed by the phellogen or cork cambium, phellem, usually termed as cork, and phelloderm (Fig. 2). The function of stomata and cuticle are replaced by the lenticels and phellem cells, respectively (Lendzian, 2006).

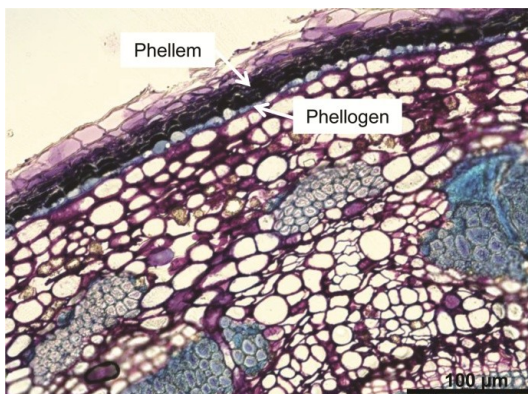


Fig. 2. Periderm development in a two year-old stem of hybrid aspen, *Populus tremula* L. × *Populus tremuloides* Michx.; Clone T89. The phellogen and the phellem cells are highlighted by arrows.

The first periderm of stems and roots usually appears during the first year of growth and subsequent periderms, if present, can initiate later in the same year or many years after, depending on the species. Environmental conditions can influence the initiation of the first or subsequent periderms. Species like *Betula*, *Fagus*, *Abies* and *Quercus* can keep the first periderm during their entire life or for many years (Evert, 2006*b*), but others, like *Norway spruce*, present a layered periderm since the phellogen functions only for one season and subsequent phellogen cells arise (Hejnowicz, 2007). In *Quercus suber* a new layer of cork (phellem) is added each year through the activity of the phellogen meristematic activity (Caritat *et al.*, 2000).

Phellogen

The phellogen has a relatively simple structure consisting in only one cell type appearing rectangular in a transverse section, flattened radially and polygonal in tangential sections (Pereira, 2007). The first cell division as a precursor of phellogen formation is derived from a periclinal division that gives rise to two apparently similar cells. The inner daughter cell differentiates into phelloderm and no further divides while the outer cell undergoes a second periclinal division resulting again in the formation of two similar cells. The outer one differentiates into a phellem cell whereas the inner cell is the phellogen initial and continues to divide (Evert, 2006*b*; Pereira, 2007). However, more phellem cells are formed as a result of the division of the phellogen cells than phelloderm cells (Pereira, 2007). While most of the divisions are periclinal, some divisions have to occur in a radial anticlinal plane due to the increase in the circumference of the stems (Evert, 2006*b*). The number of phellem cell layers formed varies among species, and can be very large in the case of *Quercus suber* (Pereira, 2007).

Phellogen initiation may occur at different distances from the apex (Yadun and Liphshitz, 1989) and in different sites depending on the plant species, initiating in the subepidermal or epidermal cell layer as in *Quercus suber* (Graça and Pereira, 2004), in the cortex as in *Norway spruce* (Hejnowicz, 2007), within the phloem as in *Vitis* (Evert, 2006*b*) or in the subepidermal layers or in the outer cortical cells as in *Populus deltoides* (Waisel and Liphshitz, 1975). In addition, the seasonal activity of the phellogen also differs among species (Waisel and Liphshitz, 1975; Liphshitz *et al.*, 1984; Fialho *et al.*, 2001; Costa *et al.*, 2002, 2003; Yáñez-Espinosa *et al.*, 2010) and may vary as well between branches of different ages (Waisel and Liphshitz, 1975). Phellogen activity in some species only lasts two to four months (Waisel and Liphshitz, 1975; Liphshitz *et al.*, 1984; Yáñez-Espinosa *et al.*, 2010) but in other species such as *Quercus suber* it can last at least

seven months (Oliveira *et al.*, 1994; Caritat *et al.*, 1996; Fialho *et al.*, 2001; Costa *et al.*, 2002, 2003; Silva *et al.*, 2005).

The annual rhythm of the two lateral meristems can overlap or can occur independently, also depending on the species (Waisel *et al.*, 1967; Yáñez-Espinosa *et al.*, 2010). The activity of the phellogen is considered relatively slow under normal conditions when compared to the activity of vascular cambium (Waisel *et al.*, 1967; Liphshitz *et al.*, 1984). Consequently, the stem radial growth has been mostly associated with the vascular cambium activity and to a lesser extent with the phellogen activity (Yáñez-Espinosa *et al.*, 2010). The lifespan of the phellogen activity varies between species from less than one year to several years and the first phellogen remains functional for the entire tree life in only a few species. In most species other periderms form after the first one due to the development of a new phellogen (Pereira, 2007).

Phellem and phelloderm cells

Phellem or cork cells are often rectangular prisms but at a tangential plane can be irregular and they can be elongated at a vertical, radial or tangential plane (Evert, 2006*b*). They are arranged in a compact way without intercellular spaces, except in lenticels, and in radial rows to the outside of the phellogen initial. At maturity they are dead cells that typically have a relatively thick layer of suberin deposited internally to the primary cell wall (Evert, 2006*b*; Pereira, 2007). The chemical composition of the cork cells can differ within the same species (Pereira, 1988, 2013; Silva *et al.*, 2005; Teixeira and Pereira, 2010) due to external factors (Silva *et al.*, 2005) and among different species, such as *Quercus suber*, *Quercus ilex*, *Betula pendula*, *Populus tremula* and *Castanea sativa* (Holloway, 1983; Pereira, 1988, 2013; Gandini *et al.*, 2006; Şen *et al.*, 2010; Miranda *et al.*, 2012). Suberization of these cells provides very little permeability to water (Soliday *et al.*, 1979; Vogt *et al.*, 1983) but

confers resistance to pathogens (Lulai and Corsini, 1998). The number of phellem layers and their thickness also vary among species. In the genus *Quercus*, some species have an extensive production of phellem, as it is the case of *Quercus suber*, *Quercus cerris* and *Quercus variabilis* (Pereira *et al.*, 1992; Graça and Pereira, 2004; Şen *et al.*, 2011b; Miranda *et al.*, 2012) whereas others, like *Quercus faginea*, produce a much thinner phellem layer (Quilhó *et al.*, 2013). *Quercus faginea* and *Quercus cerris* usually produce 2-5 and 6-12 layers of phellem, respectively, but as a discontinuous layer around the trunk (Şen *et al.*, 2011a,b; Quilhó *et al.*, 2013). On the other hand, *Quercus suber* produces a continuous layer of 3 to 8 phellem layers surrounding the stems of one year-old branches and their numbers keep increasing reaching an annual production of 10-20 phellem layers (Graça and Pereira, 2004). In contrast, for example in *Populus* the thickness of the periderm seems to be constant in younger and older trees consisting of about 6 to 10 cell layers independently of the plant age (Kaufert, 1937).

Phelloderm cells are similar to the cortical or phloem parenchyma cells and can be distinguished from those due to their arrangement in radial rows like phellem cells. They are living cells with non-suberized walls and present intercellular spaces (Evert, 2006b; Pereira, 2007). The number of phelloderm layers also differs depending on the tree species. In *Quercus cerris* and *Quercus faginea* this tissue has one to three cell layers (Evert, 2006b; Şen *et al.*, 2011a; Quilhó *et al.*, 2013), while some plants do not even have phelloderm (Evert, 2006b). On the contrary, tropical trees seem to have a very thin phellem layer but thick phelloderm as the main protective tissue (Roth, 1981; Evert, 2006b). Moreover, according to Roth (1981), plants that have a very thick cork (phellem), usually do not have a very thick phelloderm, and vice versa (Roth, 1981).

Lenticels

Lenticels, also termed as lenticular channels, are special structures through which gas exchange can occur in the bark. They develop under the stomata (Evert, 2006b; Lendzian, 2006) and these regions are characterized by an intense phellogen activity (Graça and Pereira, 2004; Evert, 2006b) producing a tissue with numerous intercellular spaces as well as the lenticel phellogen. The lenticel phellogen is continuous with the other phellogen but bends inward and it forms filling tissue to the outside, that protrude above the surface through a fissure in the periderm, and phelloderm to the inside, which is usually thicker than usual (Evert, 2006b). The difference between phellem and the filling tissue varies among species but it is mainly due to the presence of intercellular spaces or to different cell shape (Evert, 2006b).

The presence of lenticels in the cork tissue contributes to its porosity and is a detrimental characteristic for industrial quality since large and abundant lenticular channels change cork properties and contributes to poor quality cork (Pereira *et al.*, 1996; Anjos *et al.*, 2008).

Regulatory mechanisms in plant meristems

Mechanisms regulating the apical meristems

Plant meristems are dynamic structures that are regulated to keep a tight balance between stem cell proliferation and differentiation. This balance ensures the maintenance of stem cell identity while keeping meristem size.

Although the RAM and SAM of *Arabidopsis* are organized in a different way (Laux, 2003) it has been shown that stem cell maintenance signaling in both meristems employs related regulators (Sarkar *et al.*, 2007). In the SAM, *WUSCHEL* (*WUS*) gene functions in the organizing center (OC). It is required to keep the stem cells undifferentiated and induces the expression of

CLAVATA3 (*CLV3*), thus promoting organ initiation (Schoof *et al.*, 2000; Sarkar *et al.*, 2007). *CLV3*, in turn, binds to the receptor-like kinase *CLV1/2* and represses *WUS* at the transcriptional level, acting as a negative regulator, and maintaining the size of the SAM (Fig. 3A) in a self-regulatory way (Mayer *et al.*, 1998; Schoof *et al.*, 2000; Brand *et al.*, 2000; Ogawa *et al.*, 2008). This negative-feedback loop is further controlled by other mechanisms including epigenetic factors and receptor systems (Perales and Reddy, 2012). *SHOOTMERISTEMLESS* (*STM*) is required for maintaining stem cells undifferentiated, antagonizing the function of *CLV* genes (Clark *et al.*, 1996), while *Class III HD-ZIP* genes are known to negatively regulate the level of *WUS* mRNA, restricting SAM activity (Green *et al.*, 2005; Williams and Fletcher, 2005). Proper meristem functioning involve the interaction between the *CLV/WUS* network and cytokinin (CK) signaling pathway (Leibried *et al.*, 2005). Moreover, CK signaling regulates *WUS* expression through a *CLV* dependent or independent pathway (Gordon *et al.*, 2009). In the root, *WUSCHEL-RELATED HOMEBOX 5* (*WOX5*) functions similarly to *WUS* in the shoot, by expressing in the quiescent center (QC) and maintaining the stem cells undifferentiated through a non-cell-autonomous manner (Haecker *et al.*, 2004), suggesting a common regulatory mechanism for stem cell maintenance both in the shoot and root. A member of the *CLE* gene family, *CLAVATA3/ESR-RELATED* (*CLE40*), the closest homolog of the stem cell restricting signal *CLV3*, alters the expression of *WOX5* and promotes stem cell differentiation (Stahl and Simon, 2009b). Similarly to what happens in the shoot, *CLE40* acts through *ACR4*, that encodes a receptor-like kinase of the CRINKLY4 family, to regulate *WOX5* expression (Stahl and Simon, 2009a,b), which is in turn dependent on the induction of the *MONOPTEROS* (*MP*) mediated auxin signaling and on the activity of two transcription factors, *SHORT-ROOT* (*SHR*) and *SCARECROW* (*SCR*) (Sarkar *et al.*, 2007). Two other genes, *PLETHORA1* and 2 (*PLT1* and *PLT2*) are also required for stem cell specification and maintenance of the RAM acting in a parallel way to the

SHR/SCR in the establishment of the QC and stem cell position (Aida *et al.*, 2004). *PLT* transcription is elevated by auxin accumulation and is dependent on *AUXIN RESPONSE FACTORS* (ARF) transcription factors, *MP* or the *MP* homolog *NON-PHOTOTROPIC HYPOCOTYL4* (*NPH4*), that bind to the auxin-responsive elements (Aida *et al.*, 2004) (Fig. 3B). Briefly, *WUS* and *WOX5*, as well as *CLV3/CLV1* and *CLE40/ACR4*, seem to be functionally equivalent in stem cell control in the shoot and root, respectively.

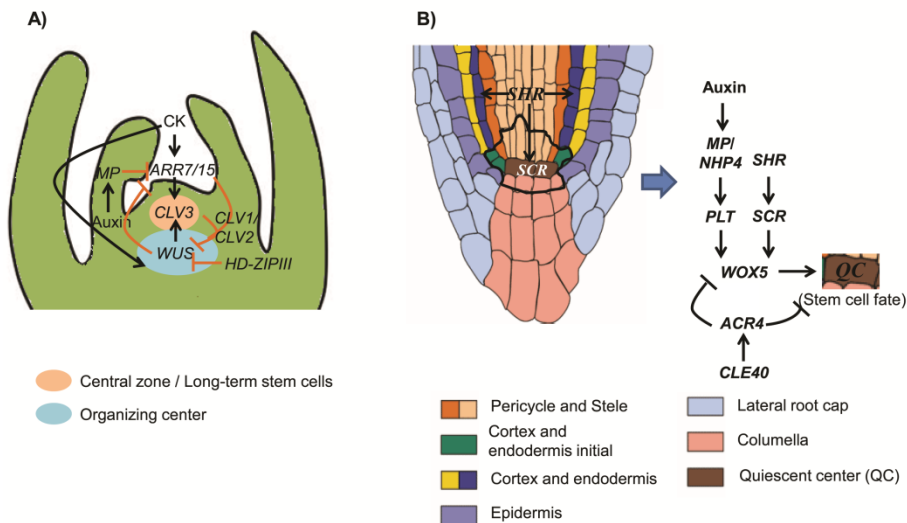


Fig. 3. Schematic representation of equivalent regulatory mechanisms in the shoot (A) and root apical meristem (B). The thick black line in B) surrounds the quiescent center and the root stem cells. (A) Adapted from (Williams and Fletcher, 2005; Zhao *et al.*, 2010) and (B) adapted from (Dinnyeny and Benfey, 2008; Stahl and Simon, 2009b).

Mechanisms regulating the vascular cambium

The control of vascular stem cell fate involves a system similar to the one described for the regulation of stem cell homeostasis in the SAM and RAM, based on the interaction between small mobile peptides (CLEs) and receptor-like kinases (RLKs) (Miyashima *et al.*, 2013). This system plays a determinant role in the fate of vascular stem cells in a non-cell autonomous manner and is

composed by TDIF (TRACHEARY ELEMENT DIFFERENTIATION INHIBITORY FACTOR) and its receptor TDR (TRACHEARY ELEMENT DIFFERENTIATION INHIBITORY FACTOR RECEPTOR), also termed PHLOEM INTERCALATED WITH XYLEM (PXY) (Hirakawa *et al.*, 2008). While TDIF is a CLV3/ENDOSPERM SURROUNDING REGION (CLE)-family peptide produced by the activity of the *CLE41* or *CLE44* genes, TDR is a leucine-rich repeat receptor kinase, belonging to the subclass XI of LRR-RLK, that also includes *CLV1* (Hirakawa *et al.*, 2008). TDIF was shown to promote proliferation of procambial cells while suppressing xylem differentiation (Fig. 4) (Hirakawa *et al.*, 2008). TDR or PXY, first described as essential in maintaining the polarity within the vascular meristems (Fisher and Turner, 2007), was also associated with induction/initiation of secondary growth in the interfascicular region (Etchells and Turner, 2010; Agusti *et al.*, 2011). Moreover, the *WUSCHEL-related HOMEODOMAIN 4* (*WOX4*) functions during differentiation and/or maintenance of procambial cell fate in *Arabidopsis* (Ji *et al.*, 2010). In addition, *WOX4*, expressed in the procambium and cambium stem cell niche, was found to be a key target of the TDIF-TDR signaling pathway and it is required for promoting the proliferation of procambial/cambial stem cells but not to repress xylem differentiation (Hirakawa *et al.*, 2010) (Fig. 4). In *Populus tremula* it was reported that the orthologs of the *WOX4* and *TDR*, *PttHB3* and *PttRLK3*, respectively, are both expressed in the vascular cambium with similar patterns (Schrader *et al.*, 2004), supporting an important role for TDIF-TDR-*WOX4* signaling in the maintenance of the vascular meristem organization during secondary growth (Hirakawa *et al.*, 2010).

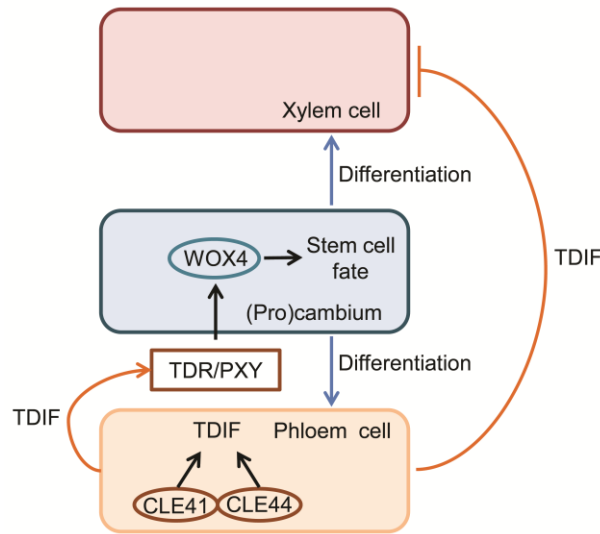


Fig. 4. Schematic representation of the TDIF-TDR-WOX4 signaling mechanism regulating the fate of procambial cells. Adapted from (Hirakawa *et al.*, 2008, 2010; Miyashima *et al.*, 2013).

In the *Arabidopsis* shoots the formation of the interfascicular cambium is initiated through cell divisions in parenchyma cells between primary bundles and in the starch sheath, the innermost cell layer of the cortex (Altamura *et al.*, 2001; Sehr *et al.*, 2010; Agusti *et al.*, 2011). Identification of genes involved in the transition from primary to secondary growth in *Arabidopsis* demonstrated that stem weight triggers the transition from primary to secondary growth (Ko *et al.*, 2004) but it is not sufficient to induce interfascicular cambium initiation, and instead it modulates the dynamics of secondary growth initiation and cambium activity (Ko *et al.*, 2004; Sehr *et al.*, 2010).

Transcriptional regulators are described as playing central roles in the regulation of secondary growth. In a first attempt to obtain more information about the regulatory networks in secondary meristems, a microarray analysis was performed covering a large cambial region of *Populus tremula* (Schrader *et al.*, 2004). Several genes were identified as being transcribed in specific stages of xylem and phloem differentiation, and others as potential regulators

of the cambial activity. Some known regulator genes implicated in the development of the apical meristem were also found expressed in the *Populus* cambial zone (Schrader *et al.*, 2004). One such example is the *Populus WUS*-like genes and *CLV1* that are also expressed during secondary growth (Schrader *et al.*, 2004). This suggests an overlapping of molecular mechanisms and the presence of evolutionary conserved processes in the functioning of plant meristems (Schrader *et al.*, 2004; Groover, 2005; Baucher *et al.*, 2007; Spicer and Groover, 2010).

One of the first expressing transcriptional regulators in the preprocambial cells is the *Class III HD-ZIP* gene *ATHB-8* (Donner *et al.*, 2009). It is expressed in the procambium (Baima *et al.*, 1995) induced by auxin, and acts as a positive regulator of proliferation and differentiation modulating the activity of procambial and cambial cells to differentiate (Baima *et al.*, 2001). Other *Class III HD-ZIP* genes with a function in SAM and RAM (Otsuga *et al.*, 2001; McConnell *et al.*, 2001; Emery *et al.*, 2003; Hawker and Bowman, 2004; Prigge and Otsuga, 2005; Barton, 2010) are also expressed in the cambial region of *Populus* (Schrader *et al.*, 2004), regulating cambium initiation and secondary vascular tissues patterning, like *Populus REVOLUTA* (Ko *et al.*, 2005; Robischon *et al.*, 2011), while *KANADY* genes acts antagonistically to the *Class III HD-ZIP* (Emery *et al.*, 2003; Hawker and Bowman, 2004; Schrader *et al.*, 2004; Ilegems *et al.*, 2010). *SHR* is also expressed in the cambial region of *Populus* (Schrader *et al.*, 2004; Wang *et al.*, 2011) and mediates, in a non-cell autonomous manner, *Class III HD-ZIP* expression and xylem patterning in the *Arabidopsis* root (Carlsbecker *et al.*, 2010).

Two *Populus* Class 1 *KNOX* homeobox genes, *ARBORKNOX1* and 2, orthologs of the *Arabidopsis* *SHOOTMERISTEMLESS* (*STM*) and *BREVIPEDICELLUS* (*BP*), respectively, are known to be involved in the regulation of both primary and secondary meristems (Long *et al.*, 1996; Mele *et al.*, 2003; Groover *et al.*, 2006; Du *et al.*, 2009). *ARK1* and *ARK2* are

involved in the regulation of vascular cambium promoting meristem identity (Groover *et al.*, 2006; Du *et al.*, 2009). While *ARK1* promotes meristematic cell fate and delays differentiation of daughter cells derived from the cambium, *ARK2* is positively related with the timing of cambium formation, width of the cambial region and inhibition of cambial cells differentiation during secondary growth (Groover *et al.*, 2006; Du *et al.*, 2009) and both *ARK1* and *ARK2* regulate the expression of other genes involved in secondary growth. More recently, it was shown that the *Populus LATERAL ORGAN BOUNDARY DOMAIN (LBD)* regulates secondary phloem development in *Populus* stems suppressing meristem cell identity, activating phloem differentiation, promoting lateral stem growth and one of the main regulatory routes for its action is through modulation of *ARK1* and *ARK2* expression (Yordanov *et al.*, 2010).

All these lines of evidence suggest that the apical meristems (RAM and SAM) and the lateral meristems may share common regulators. A relation between the regulation of secondary growth processes promoted by the two lateral meristems has also been briefly addressed (Soler, 2008). Soler (2008) reported that some genes involved in wood formation such as *Class III HD-ZIP*, *KNOX* and *KANADY*, are also expressed in cork tissues, indicating a possible role in cork formation.

Complex molecular mechanisms controlling the vascular tissue differentiation comprehend different genes and hormones (Aloni, 1987; Ye, 2002; Baucher *et al.*, 2007; Du and Groover, 2010; Schuetz *et al.*, 2012). Regulation of cambial activity has long been associated with phytohormones (Jacobs, 1952; Aloni, 1987; Little and Savidge, 1987; Savidge, 1988; Ugglä *et al.*, 1996; Lachaud *et al.*, 1999; Ye, 2002), and an appropriate hormone balance is required to maintain proper cambium and secondary tissue development. Several recent reviews focus on the roles of auxin, cytokinin, ethylene, gibberellin and brassinosteroids as being implicated in the maintenance and activity of cambial stem cells (Lachaud *et al.*, 1999; Fukuda, 2004; Dettmer *et al.*, 2009; Elo *et al.*, 2009; Caño-Delgado *et al.*, 2010;

Vanstraelen and Benková, 2012; Ursache *et al.*, 2013; Milhinhos and Miguel, 2013). Auxin has been shown essential for (pro)cambium initiation and activity (Snow, 1935; Nilsson *et al.*, 2008; Donner *et al.*, 2009; Baba *et al.*, 2011) and for fiber and xylem differentiation (Aloni, 1987; Tuominen *et al.*, 1997; Zhong and Ye, 1999, 2001; Björklund *et al.*, 2007; Bishopp *et al.*, 2011). Cytokinins maintain stem cell pools by regulating cell differentiation, promoting cell proliferation in the vascular stem cells during secondary growth (Nieminen *et al.*, 2008) and having a role in cambium formation and activity (Ye and Varner, 1994; Mähönen *et al.*, 2000; Nieminen *et al.*, 2008; Matsumoto-Kitano *et al.*, 2008; Hejálko *et al.*, 2009) and in xylem differentiation/specification (Mähönen *et al.*, 2006; Bishopp *et al.*, 2011) as well as fiber differentiation and programmed cell death (Aloni, 1982; Carimi *et al.*, 2003, 2004; Zottini *et al.*, 2006; Vescovi *et al.*, 2012; Kunikowska *et al.*, 2013a,b). Ethylene has a positive effect on cambial cell divisions and on xylem differentiation (Miller *et al.*, 1984; Savidge, 1988; Love *et al.*, 2009). Gibberellin has a role in xylem differentiation and cambial cell division (Ridoutt *et al.*, 1996; Eriksson *et al.*, 2000; Björklund *et al.*, 2007). Brassinosteroids are involved in xylem development by inducing xylem differentiation in *Zinnia* cell cultures (Yamamoto *et al.*, 1997, 2001; Ohashi-Ito *et al.*, 2002, 2005; Ohashi-Ito and Fukuda, 2003) and in *Arabidopsis* (Caño-Delgado *et al.*, 2004; Ibañes *et al.*, 2009).

Some essential genes and processes involved in vascular development seem conserved across different plant species, from herbaceous dicots to woody dicots and monocots, such as *Arabidopsis*, *Populus* and *Oryza*, respectively, making the transfer of knowledge among different plant species plausible at some extent (Xu *et al.*, 2013).

SHORT-ROOT: a gene involved in the regulation of primary and secondary growth

SHORT-ROOT controls radial patterning and stem cell maintenance in the Arabidopsis root

SHORT-ROOT (SHR) is a transcription factor of the GRAS family, a plant specific protein family (Helariutta *et al.*, 2000; Bolle, 2004; Gallagher and Benfey, 2010), whose members play important roles in very diverse processes (Bolle, 2004). The first study reporting the identification of the *short-root* (*shr*) mutant was made in *Arabidopsis* through a genetic analysis of root development, where mutants with abnormal morphogenesis were found. This mutant was characterized by its determinate root growth (Benfey *et al.*, 1993) and loss of one ground tissue layer, the endodermis, but whose cortex attributes were retained (Benfey *et al.*, 1993; Helariutta *et al.*, 2000). The roots of these mutants cease growth after a short period and cells become differentiated. As the root increases in length, the meristematic and elongation zones diminish in size, suggesting a gradual loss of cells entering in differentiation. The authors explained the loss of the meristem ability to maintain root growth by lack of replacement of the initials. Therefore, growth only occurs until its cell-division potential is exhausted (Benfey *et al.*, 1993). The absence of endodermis in the mutants indicates that *SHR* controls ground tissue patterning during root development in *Arabidopsis*. Thus, the *SHR* gene appears to be necessary both for asymmetric cell division of the cortex/endodermal initial (CEI) and for endodermis specification (Benfey *et al.*, 1993; Helariutta *et al.*, 2000). *SHR* function is not limited to the primary root and is also required for the initiation and patterning of the *Arabidopsis* lateral root primordia and for maintaining its indeterminate growth (Lucas *et al.*, 2011). Surprisingly, *SHR* expression was found in the root stele but not in the ground tissue, indicating that *SHR* controls radial organization of ground tissue

in a non-cell autonomous manner (Helariutta *et al.*, 2000). However, the SHR protein moves from the stele, where it is localized in the cytoplasm and nuclei, to all the cells immediately adjacent to the stele, which include quiescent center (QC), CEI and daughter cells as well as endodermis, where it enters the nucleus (Nakajima *et al.*, 2001). The evidence that SHR protein is able to move from the stele provided a mechanism by which it could act as a positional signal in radial patterning (Helariutta *et al.*, 2000; Nakajima *et al.*, 2001). SHR was also identified as an upstream and positive regulator of SCARECROW (SCR) (Helariutta *et al.*, 2000), another GRAS family member. SCR expression is localized in CEI, QC and endodermis (Scheres *et al.*, 1995; Laurenzio *et al.*, 1996; Sabatini *et al.*, 2003). It is involved in the regulation of the ground tissue patterning in the *Arabidopsis* root (Scheres *et al.*, 1995; Laurenzio *et al.*, 1996) and, together with SHR regulates the asymmetric cell division of the CEI daughter cell. The SHR movement to the adjacent cell layer activates SCR transcription and endodermis specification. Nonetheless, SHR activation of SCR is necessary for cell divisions which are essential for a correct radial patterning (Nakajima *et al.*, 2001). The *scr* mutant also presents a single cell layer between the epidermis and pericycle but, unlike the *shr* mutant, this layer has differentiated attributes of both cortex and endodermis. This suggests a role of SCR in the regulation of the asymmetric cell division of the CEI instead of the specification of the cell identity (Laurenzio *et al.*, 1996). In fact, SHR expression in the *scr* mutants is still confined to the stele indicating that the establishment of the SHR expression pattern is not dependent on SCR activity and, therefore, can explain the retention of endodermal differentiation in the ground tissue of the *scr* mutant (Helariutta *et al.*, 2000). Analogous to their functioning in the ground tissue patterning both SHR and SCR are required for QC identity (Sabatini *et al.*, 2003). SCR expression is required cell-autonomously for QC identity and for maintaining the surrounding stem cells in a non-cell autonomous manner (Sabatini *et al.*, 2003). In the *shr* mutant the morphology of the QC and columella region is

irregular and the correct patterning is not restored when *SCR* is re-expressed in the QC region, suggesting that *SHR* has a role in QC function and stem cell maintenance, independent of the *SCR* transcription in this region (Sabatini *et al.*, 2003). Overall, *SHR* and *SCR* are key regulators of root radial patterning (Laurenzio *et al.*, 1996; Helariutta *et al.*, 2000) and stem cell maintenance (Sabatini *et al.*, 2003).

SHR protein movement and localization

Unlike diffusible animal morphogens, which form a gradient across multiple cell layers, *SHR* movement is limited to one cell layer. A possible role for *SCR* in restricting *SHR* movement was first described almost simultaneously by Sena *et al.* (2004) and Heidstra *et al.* (2004). It was evidenced that *SHR* intercellular trafficking is both regulated and targeted (Gallagher *et al.*, 2004). The authors observed that cytoplasmic localization is required for *SHR* movement but it is not sufficient, and that its localization at the cytoplasm enables it to interact with proteins that facilitate its movement (Gallagher *et al.*, 2004). The substitution of a specific threonine amino acid at the VHIID C-terminal domain (T289) of the GRAS protein family makes *SHR* uniquely localized in the cytoplasm of the stele and, in turn, the protein loses its capability to move, suggesting that *SHR* is not diffusible (Gallagher *et al.*, 2004). This amino acid is not specifically required for *SHR* movement, instead it revealed to be required for proper *SHR* nuclear localization which is in turn required for proper movement (Gallagher and Benfey, 2010), acting as a possible regulator of the activity and movement of *SHR* (Gallagher *et al.*, 2004; Gallagher and Benfey, 2010). Other GRAS domains beyond the VHIID are also required for proper movement (Gallagher and Benfey, 2010). The finding that promoting either nuclear or cytoplasmic location over the other inhibits movement, strongly suggested that *SHR* movement requires both

cytoplasmic and nuclear localization, favoring a balance between nuclear import and export (Gallagher and Benfey, 2010).

A mechanism in which SCR tightly restricts SHR movement was proposed, describing that SCR sequesters SHR into the nucleus through protein complex formation preventing further SHR movement (Cui *et al.*, 2007). Another gene belonging to a plant specific family of zinc finger proteins, *JACKDAW* (*JKD*), has a role in the control of radial pattern formation in the root meristem by restricting the range of SHR action, limiting its movement into the cortex and beyond the QC, by protein-protein interaction forming nuclear complexes in the cells where they are coexpressed and sequestering SHR in the nucleus (Welch *et al.*, 2007).

A novel endosome-associated protein, SHORT-ROOT INTERACTING EMBRYONIC LETHAL (*SIEL*), is associated with intercellular trafficking and promotes movement of the SHR protein from the stele into the endodermis (Koizumi *et al.*, 2011). This protein interacts with other non-cell-autonomous transcription factors and has a role in the movement of multiple different families of transcription factors. Whereas in the endodermis SHR up-regulates the expression of *SCR* and *JKD*, which in turn restrict SHR movement, it also promotes *SIEL* expression, suggesting that SHR may actively promote its own movement.

Very early it was suggested that SHR movement might occur through plasmodesmata (Nakajima *et al.*, 2001). More recently, it was found that callose biosynthesis has a central role on the regulation of symplastic trafficking through the plasmodesmata and in mediating cell-to-cell signaling. The gain-of-function mutations of *CALLOSE SYNTHASE 3* (*CALS3*) increased the accumulation of callose (β -1,3-glucan) at the plasmodesmata, decreasing its aperture and affecting root development by the reduction and impairment of intercellular trafficking (Vatén *et al.*, 2011). By inducing callose production, SHR movement from the stele to the endodermis is inhibited (Vatén *et al.*, 2011).

SHR regulates a large transcriptional network

Levesque *et al.* (2006) combined several microarray experiments to analyse global expression profiles after modulating SHR activity (Levesque *et al.*, 2006). This study confirmed SCR as a direct target of SHR and revealed seven more putative SHR direct targets that are positively regulated: a SCARECROW-like 3 gene (*SCL-3*), two C2H2 zinc finger transcription factor genes, *MAGPIE* (*MGP*) and *NUTCRACKER* (*NUC*), a gene that encodes a receptor-like kinase (*RLK*), two genes that encode metabolic enzymes putatively involved in tropane alkaloid synthesis and in brassinosteroid biosynthesis (*TRI* and *BR6ox2*, respectively), and *SNEEZY/SLEEPY2* (*SNE*) thought to play a role in gibberellin signaling. Many indirect target genes were also identified including transcription factors such as the *Class III HD-ZIP* genes *PHABULOSA* (*PHB*) and *PHAVOLUTA* (*PHV*) and genes involved in the response to hormones such as auxin, most of them regulated in a repressive manner. These results point to a major role of SHR in root development through regulation of a large transcription factor network and of hormonal and signaling pathways using receptor-like kinases.

Beyond the mechanisms revealed to be implicated in proper patterning, Sozzani *et al.* (2010) provided the first evidence of a transcriptional regulatory relationship between SHR/SCR and specific components of the cell-cycle machinery (Sozzani *et al.*, 2010). SHR and SCR had been previously reported as regulators of formative cell divisions in the ground tissue (Benfey *et al.*, 1993; Scheres *et al.*, 1995; Laurenzio *et al.*, 1996; Helariutta *et al.*, 2000). A D-type cyclin, *CYCD6;1*, was identified by Sozzani *et al.* (2010) as being regulated by SHR and SCR and specifically involved in formative cell divisions within the CEI cells, needed for proper ground tissue patterning, instead of being required for proliferative cell divisions. Other genes, particularly two cyclin-dependent protein kinases, *CDKB2;1* and

CDKB2;2, together with *CYCD6;1* seem to be involved in specific formative cell divisions downstream of the SHR/SCR network (Sozzani *et al.*, 2010).

SHR function in the Arabidopsis root vascular system and in the regulation of leaf growth

Although much has been known about the mechanisms by which SHR regulates ground tissue patterning, its role in the control of other aspects of root development has more recently started to be revealed. The localization of SHR transcript and protein in the stele is suggestive of a function in the development of the vascular tissues in the root. In fact, Levesque *et al.* (2006) observed that mutation of *SHR* results in a decrease in the number of stele initials, affecting their specification and the differentiation of the phloem. These data strongly supports a role for SHR in cell division and specification within the stele (Levesque *et al.*, 2006), and thus in the development of the vascular tissues. Additional developmental defects in the root vascular system of *shr* mutants include aberrant specification and/or patterning in both protophloem and protoxylem elements (Yu *et al.*, 2010). In the *shr* mutants the number of cells in the xylem strand is reduced and protoxylem is not even present, indicating that SHR specifically regulates the specification and/or patterning of protoxylem, but not metaxylem, affecting stele development in a non-cell-autonomous manner (Yu *et al.*, 2010). Indeed, in *shr* and *scr* mutants metaxylem differentiates ectopically in place of protoxylem (Carlsbecker *et al.*, 2010).

A correct patterning of the central vascular cylinder occurs through a crosstalk between the vascular cylinder and the surrounding endodermis and this is mediated by cell-to-cell movement of the SHR in one direction and microRNAs (*miR165a* and *miR166b*) in the opposite direction (Carlsbecker *et al.*, 2010). A correct xylem patterning was shown to require both SHR and SCR presence in the endodermis. SHR directly activates SCR and together

trigger the miR165/6, and it is through this activation mechanism that SHR/SCR function non-cell-autonomously in xylem patterning. Upon activation, miR165/6 moves into the stele to restrict *HD-ZIP III* mRNA domains, thereby specifying protoxylem cell fate. While *PHABULOSA* (*PHB*) is the primary determinant for xylem patterning through its role on metaxylem specification, the other *Class III HD-ZIP* genes have a role as well in the determination of xylem type (Carlsbecker *et al.*, 2010).

Gene regulation in the pericycle and vascular domains was also found directly related with SHR and novel SHR targets were found (Cui *et al.*, 2011). Whereas in the root of *shr* the phloem and phloem-associated pericycle are enlarged, xylem and xylem-associated pericycle are reduced, indicating that SHR controls the balance between these vascular tissues and has a role in vascular patterning (Cui *et al.*, 2011).

Through a genome-wide analysis of SHR in the development of the *Arabidopsis* root, Levesque *et al.* (2006) had suggested that SHR controls root development through the regulation of three different modules, one of which related to the regulation of hormone signaling and biosynthesis (Levesque *et al.*, 2006). Recently, other authors further reinforced this idea by finding that SHR regulates vascular patterning in the *Arabidopsis* root through cytokinin (CK) homeostasis (Cui *et al.*, 2011; Hao and Cui, 2012). CK plays an important role in vascular tissue differentiation (Mähönen *et al.*, 2000, 2006; Mähönen, 2005; Bishopp *et al.*, 2011) and it was verified that the application of exogenous CK confers a phenotype similar to the *shr* mutant and that the mutants have higher CK levels (Cui *et al.*, 2011). Indeed, *CYTOKININ OXIDASE 3* (*CKX3*), a gene responsible for CK inactivation, was identified as a SHR target, and its overexpression results in a reduction of CK levels that reverses the vascular patterning defect in *shr* (Cui *et al.*, 2011; Hao and Cui, 2012). In addition, it is suggested that SHR regulates *miR165A* and *miR166B* indirectly through its effect on CK homeostasis (Hao and Cui, 2012).

Inhibition of leaf growth is also observed in the *shr* and *scr* mutants (Benfey *et al.*, 1993; Laurenzio *et al.*, 1996; Dhondt *et al.*, 2010), suggesting that SAM is partially inhibited (Dhondt *et al.*, 2010). This effect was not due to the compromised root development but it is instead caused by a reduced cell division rate and early exit of the proliferation phase in the leaves (Dhondt *et al.*, 2010). Both *SHR* and *SCR* are expressed in leaf proliferating cells and in the vascular system and, similarly to what happens in the root, *SCR* acts downstream of *SHR* in the shoot. In the leaf, *SHR* is activated simultaneously with the *Class III HD-ZIP* gene *ATHB-8* and their expression domains coincide, suggesting that synchronous activation of their expression defines a reproducible cell state that announces vein appearance (Gardiner *et al.*, 2011). Similar to the behavior observed in the *Arabidopsis* root, also in leaf development *SHR* moves to the contiguous cell layer, a region of non-vascular cells that surrounds leaf veins, suggesting activities in procambium-precursor cells beyond vein formation (Gardiner *et al.*, 2011).

In addition to the previously described roles, both *SHR* and *SCR* regulate the radial organization of the shoot axial organs in *Arabidopsis* and are essential for normal shoot gravitropism, which in turn is dependent on a normal endodermis formation (Fukaki *et al.*, 1998).

SHR acts in a dosage-dependent manner

As the root develops the abundance of *SHR* protein changes dynamically (Koizumi *et al.*, 2012). During maturation of the wild-type *Arabidopsis* root a third layer of ground tissue forms and assumes cortex characteristics, termed as middle cortex. The timing of this transition is regulated by *SCR* and gibberellin and *SHR* is required for it (Paquette and Benfey, 2005). A complete block of *SHR* movement results in the loss of ground tissue but in contrast, a decrease in its movement induces periclinal divisions in the endodermis and increases the number of ground tissue layers (Koizumi *et al.*, 2012).

Therefore, the pattern of cell division within the endodermis is sensitive to the SHR concentration and high levels of SHR prevent the formation of middle cortex whereas intermediate levels promote its formation (Koizumi *et al.*, 2012).

Another example of the SHR mode of action through a dosage-dependent manner was reported in the aerial part of the *Populus* tree (Wang *et al.*, 2011). The partial silencing of the ortholog of the *Arabidopsis* SHR (*AtSHR*) gene in *Populus*, *PtSHR1*, revealed an overall increase in plant size, with an enhancement of the primary (height) and secondary (girth) growth rates. Furthermore, by reducing the *Arabidopsis* SHR activity, rather than completely suppressing it, similar effects to those found in *Populus* were achieved.

Taken together, these results indicate that SHR acts in a concentration-dependent manner to regulate plant growth and development through the regulation of cell division and meristem activity, both in root and shoot.

SHR function in other plant species

Very few studies have been reported on the analysis of SHR in species other than *Arabidopsis*, but two of them were reported in tree species. In *Populus*, three SHR-like genes were identified and *PtSHR1*, the ortholog of the *Arabidopsis* SHR, was found expressed in the root stele, like in *Arabidopsis*, and in the vascular cambium of the *Populus* stem. *Populus* SHR1 was referred as a regulator of cell division and as a negative regulator of the vascular cambium activity (Wang *et al.*, 2011). In *Pinus radiata* only one SHR gene, *PrSHR*, was identified and considered as a putative ortholog of the *Arabidopsis* SHR gene (Solé *et al.*, 2008). Expression analysis during vegetative development showed that the *P. radiata* SHR is predominantly expressed in roots, followed by hypocotyls, shoot apex and cotyledons. *In situ*

hybridization analysis showed that it is also expressed in the root primordia and in the cambial region of hypocotyl cuttings. Its expression pattern suggests that *PrSHR*, like the *AtSHR* gene (Helariutta *et al.*, 2000), has a function in root initiation and root development and a potential role in organs with active cell division (Solé *et al.*, 2008). Furthermore, the authors suggest that *SHR* plays a role in the development of the vascular tissues possibly regulating asymmetric cell division in the hypocotyl cambial region or in the development and maturation of the vascular system (Solé *et al.*, 2008). Whilst adventitious root formation in pine is dependent on the application of exogenous auxin, increased levels of *SHR* were observed in the cambial region and root competent cells of hypocotyl cuttings independently of the presence or absence of exogenous auxin (Solé *et al.*, 2008).

In the model legume *Medicago truncatula*, *SHR* (*MtSHR*) was mainly expressed in the root tip and, like in *P. radiata*, its expression did not increase in response to exogenous auxin (Imin *et al.*, 2007). In rice, *SHR* (*OsSHR1*) was found expressed during stomata development and its expression in the root was not restricted to vascular tissues but also occurred in the endodermis and some cortex cells. Therefore, the authors suggested that *OsSHR1* may also be involved in root development but, due to the differential expression patterns, it may act in a somehow different way as compared to *Arabidopsis* (Kamiya *et al.*, 2003). In a transcriptomic study in *Solanum tuberosum* to identify genes involved in periderm response to heat stress, *SHR* was found down-regulated in the skin (phellem) when compared to the phelloderm (Ginzberg *et al.*, 2009).

***Populus* as a model system for studies on secondary growth**

In the last few years many of the studies addressing the regulation of secondary growth were performed in *Populus*, which is considered as a more

complex system when compared to *Arabidopsis*. However, genes required for secondary growth are not exclusive of woody plants and are present in herbaceous species, such as *Arabidopsis* (Groover, 2005).

Arabidopsis has long been considered as a model species for molecular studies in dicotyledonous plants. This is not only due to the fact that this was the first plant species to have its genome sequenced, but also to other features like its small size, short life cycle, easy manipulation including genetic transformation, and the large collection of knock-out mutations (Pang and Meyerowitz, 1987; Goodman *et al.*, 1995; Taylor, 2002). Under appropriate conditions *Arabidopsis* can exhibit secondary growth at some extent in the hypocotyls as a result of the activity of the vascular cambium and cork cambium, resembling secondary growth in trees (Chaffey *et al.*, 2002). However, the use of *Arabidopsis* presents some limitations as it lacks some unique characteristics of the perennial growth in trees, becoming unsuitable for studies related to the cambial activity, which is seasonal and comprises periods of activity and dormancy (Bradshaw Jr *et al.*, 2000; Chaffey *et al.*, 2002). Moreover, when comparing *Arabidopsis* and *Populus*, ray parenchymatic cells are completely absent from the *Arabidopsis* xylem and cambial and wood cells are smaller (Chaffey *et al.*, 2002).

Populus has become largely accepted as a model woody plant. It was the first tree species with a sequenced genome (Tuskan *et al.*, 2006), relatively small in size (450-550 Mbp). Moreover, it is fast growing, shows extensive secondary growth, it is easy to manipulate genetically and a number of genetic maps are available as well as molecular and genomic tools (Bradshaw Jr *et al.*, 2000; Taylor, 2002). *Populus* can also be easily propagated vegetatively allowing the production of large amounts of clonal material for experiments (Taylor, 2002). The large size and radial diameter and organization of the cambial zone allows its detailed study and even the harvesting of cells from the cambium at different specification stages (Uggla *et al.*, 1996; Hertzberg *et al.*, 2001; Schrader *et al.*, 2004; Nieminen *et al.*, 2008).

The disadvantages of using *Populus* as a model include its size as it can achieve an appreciable dimension not very suitable for some studies and its dioecy, making impossible self and back-cross manipulations (Bradshaw Jr *et al.*, 2000; Taylor, 2002). Nevertheless, it is still highly attractive when compared to other tree species where genetic and genomic tools are lacking (Groover, 2005).

Cork oak: a peculiar species with major ecological and economic impact

The Fagaceae is a large plant family that comprises more than 900 species belonging to 8-10 genera spread throughout the Northern hemisphere, from the tropical to the boreal regions (Kremer *et al.*, 2012). These species have a high socio-economic value, providing wood, biomass, food, timber, and are considered as keystone species in their native ecosystems. About half of these plants belong to the *Quercus* family.

Quercus suber L. or cork oak is an evergreen long-living species native from the Western Mediterranean region and North Africa (Bugalho *et al.*, 2011). It is a monoecious wind-pollinated species with asynchronous development of its sexual organs in each tree, extending from April until the end of May, with the male flowers developing first followed by the female flowers, thus promoting a high degree of self-incompatibility (Boavida *et al.*, 1999). Fertilization occurs in July and several ovules can be fertilized but only one continues development. Embryo development occurs during autumn and a monospermic seed matures giving rise to the acorn (Boavida *et al.*, 1999), which is an elongated fruit with a short pedicel and covered by a half-spherical cupule.

Phenological aspects of *Quercus suber* flowering have been described mainly focusing on external development of the male and female floral

structures evidencing the reproductive process and pollination phase (Varela and Valdivieso, 1996). A peculiar characteristic in this species is the variability in its reproductive cycle, enabling the existence of two different seed cycles giving rise to annual and biennial acorns (Elena-Rosello *et al.*, 1993). Annual biotypes require one season to complete its reproductive cycle while seed maturation in the biennial acorns occurs only one year after pollination (Elena-Rosello *et al.*, 1993; Díaz-Fernández *et al.*, 2004). This feature may represent an adaptive strategy to external conditions such as climate (Elena-Rosello *et al.*, 1993; Díaz-Fernández *et al.*, 2004; Pons and Pausas, 2012), and a single tree can exhibit both types of acorn maturation in the same year (Varela and Valdivieso, 1996; Díaz-Fernández *et al.*, 2004), demonstrating the plasticity of this species.

Cork oak woodlands, also termed “montados”, have a high social, ecological and economic value and are considered as reservoirs of biological diversity due to their richness in flora and fauna (Gil; Pereira *et al.*, 2008). Its social and economic importance derives from the cork exploitation activity. Cork oak is the only plant species able of renewable production of cork with properties suitable for industrial applications (Şen *et al.*, 2011b).

Cork is stripped from the tree by a debarking procedure that is environmentally friendly because the tree is not cut down and the periderm renews itself for the next harvesting. This process is periodical and occurs for the first time when the tree is about 25 years-old and has a circumference over bark at breast height of 70 cm or higher, and then subsequently at 9-year intervals. Virgin cork is the first cork stripped from the cork oak tree and has an irregular structure, thickness and density (Silva *et al.*, 2005). The second cork stripped, termed first reproduction cork, is harvested when the tree is about 34 years-old and, despite being more regular, it has not yet enough quality for cork stoppers (Pereira *et al.*, 1987). From this point on, when the tree reaches about 43 year-old, the cork removed is of great quality and is termed as reproduction cork or “amadia” (Silva *et al.*, 2005).

Cork is 100% recyclable and reusable and its uniqueness, given by its chemical, mechanical and physical properties makes it a material of choice for innumerable applications, of which the cork stoppers in wine and champagne industry have a high relevance. Additional applications include pavements, coverings and insulation, automobile industry, musical instruments, fashion accessories among others (Gil; APCOR, 2012).

Periderm development in cork oak

In cork oak the phellogen is described as arising under the epidermis and it is produced during the first year of growth forming a continuous layer around the stem and having already some phellem cells (Graça and Pereira, 2004). The formation of the periderm in cork oak is also derived from periclinal divisions, first in small regions around the stem perimeter (Pereira, 2007). In a tangential section, cork oak phellogen cells have a polygonal shape and are very similar to each other only varying slightly in dimension (Pereira, 2007). The phellogen in this species can be functional during many years and probably during the entire tree's life. When the initial phellogen is no longer active due to climate conditions, fire, wounding or by the debarking process, a traumatic phellogen differentiates in the outer phloem followed by an enhanced meristematic activity in the following years (Pereira, 2007). The process of cork removal exposes the phellogen to the atmosphere resulting in the dying of the phellogen cells which will promote the formation of a new phellogen (Pereira *et al.*, 1992). The new phellogen restores its activity in the next 25-35 days (Machado, 1935) and this cyclic process allows cork exploitation. Cork is considered the component that most contributes to the cork oak tree diameter increment in the case of adult trees under production (Costa *et al.*, 2001). The phellogen activity in cork oak is seasonal, comprehending about seven months, from April to October/November (Fialho *et al.*, 2001; Costa *et al.*, 2003; Silva *et al.*, 2005; Pereira, 2007) and its maximum activity is achieved in

June and July when the phellogen is swollen and the thin newly formed phellem cells allow an easy rupture and separation from the underlying tissues (Fialho *et al.*, 2001; Costa *et al.*, 2003; Pereira, 2007). A dormancy period was reported for the remaining months (Natividade, 1950; Fialho *et al.*, 2001; Costa *et al.*, 2003; Silva *et al.*, 2005; Pereira, 2007) or a low growth rate from early autumn to the next spring (Oliveira *et al.*, 1994; Costa *et al.*, 2003). The phellogen activity, and consequently cork growth, is dependent on external and internal factors, related to climatic conditions (Caritat *et al.*, 1996, 2000; Costa *et al.*, 2002; Pereira, 2007) and cork quality inherent to the tree and related to the cellular structure (Gonzalez-Adrados and Pereira, 1996; Pereira *et al.*, 1996; Gonzalez-Adrados *et al.*, 2000; Pereira, 2007).

One year-old cork oak stems already present a continuous ring of periderm with a uniform thickness, composed by three to six phellem cell layers to the outside of the phellogen and one phelloderm cell layer to its interior, with few and small lenticels. At this stage, the division of the phellogen only occurs after suberization of the previously formed phellem cell (Graça and Pereira, 2004). During the following year of growth the periderm increases in thickness and in four year-old branches a higher number of irregularities of the periderm is observed due to the increase of the lenticels. From the fifth to the seventh years of growth, phellem cells acquire the characteristics of “adult” cork cells showing empty lumens and thin suberized cell walls. At this point the activity of the phellogen increases producing 10-20 phellem cell layers each year with regular radial arrangement (Graça and Pereira, 2004).

Cork oak cells have no intercellular spaces (Graça and Pereira, 2004) and tangentially they are seen as polygons, analogous to a honeycomb arrangement. Transversely they are regularly arranged and rectangular. The cellular structure and chemical composition of cork are determinant for its characteristic properties such as low density, reduced permeability to liquid and gases, elasticity and resilience as well as heat and acoustic insulation (Pereira, 1988). The cork formation process may influence cell dimensions,

cell wall characteristics and cellular structure, therefore affecting cork properties and quality. It has been reported that the chemical composition of cork depends on several factors, such as climate and soil as well as genetic origin, growth conditions and geographic location, tree dimensions and age (Silva *et al.*, 2005). However, variation among trees seems to account for larger differences in cork properties than factors such as geographic location (Pereira, 2013).

The width of a cork ring is highly dependent on the phellogen age. The cork boards withdrawn from the cork oak tree have different thickness as result of differences in cork growth rates, and they are classified by caliber. Different cork boards thicknesses tend to have different properties like density and porosity (Natividade, 1950; Pereira *et al.*, 1992).

Previous work on periderm and cork formation

In addition to the reports describing the composition and structure of periderm tissues in *Eucalyptus* (O'Gara *et al.*, 2009), *Quercus suber* (Pereira *et al.*, 1987; Graça and Pereira, 2004) and other *Quercus* (Şen *et al.*, 2010, 2011a,b; Quilhó *et al.*, 2013), several studies have focused on the development of the potato tuber periderm, which is considered a good model of suberization as the cells accumulate a substantial amount of suberin in the cell walls. In the potato tuber, like in *Quercus suber*, potato excoriation (or debarking in the cork oak) damages the phellogen and increased tissue susceptibility is observed in the immature tuber periderm where the active phellogen cells have thin and fragile walls (Lulai and Freeman, 2001). A transcriptomic profile to identify genes involved in the potato periderm response to heat stress has been performed and suggests that the periderm, consistently with its protective role, accelerates its development and accumulates suberized phellem layers due to an increase in phellogen activity under high temperature (Ginzberg *et al.*, 2009).

The analysis of candidate genes for suberin biosynthesis and periderm formation, performed through a suppression subtractive hybridization library between the potato tuber skin (phellem) and tuber parenchyma flesh tissue (Soler *et al.*, 2011), identified a set of genes that can be used in molecular studies focusing on suberin and periderm formation and regulation. More recently, genes involved in cell cycle and encoding structural cell wall proteins have been suggested to be involved in wound periderm formation (Neubauer *et al.*, 2012).

A genomic approach to cork formation has been reported, however it mostly focused on suberin biosynthesis and investigation of gene expression patterns was performed solely in phellem tissues (Soler *et al.*, 2007). Based on these results, some of the identified candidate genes representing metabolic pathways involved in cork biosynthesis and regulatory transcription factors were selected to analyse its seasonal variation in cork tissues. Structural genes of suberin pathways and the regulatory genes in analysis showed highest transcript accumulation in June, a crucial month for cork development (Soler *et al.*, 2008). Proteomic analysis of cork formation was also performed in cork oak and several of the identified proteins had also been detected in studies performed in the potato tuber tissues (Ricardo *et al.*, 2011).

Molecular and genomic tools in cork oak

The unique and peculiar characteristics of the cork produced from the cork oak tree make this species quite attractive. Nevertheless, genomic resources in cork oak have been lacking and it has been only in the last couple of years that an increasing interest of the scientific community and a few cork industries has led to the generation of such resources by Portuguese consortia. The availability of next generation sequencing technologies and the

need to obtain more information on cork oak have been decisive in the launching of a large scale transcriptomic project targeting the release of a reference transcriptome of cork oak. In this project (CorkOak EST Consortium) the sequencing of Expressed Sequence Tags (ESTs) in cork oak covered a wide range of tissues, developmental stages and stress conditions, allowing the assembly of the transcriptome (Pereira-Leal *et al.*, 2014) and the creation of a database (CorkOakDB, <http://corkoakdb.org/>). A cork oak genome sequencing project is currently ongoing (GenoSuber project: <http://www.genosuber.com/index.html>) and it is expected to release a draft genome by the end of 2015. These novel tools will surely open new avenues for studying this peculiar tree allowing to further dissect the regulatory networks controlling cork formation, stress responses, reproductive biology among other processes.

Transcriptomic (Derory *et al.*, 2006; Durand *et al.*, 2010; Ueno *et al.*, 2010, 2013; Santamaría *et al.*, 2011) and genomic (Faivre Rampant *et al.*, 2011; Lesur *et al.*, 2011) data generated from other oak species have also become available in recent years and provide a helpful resource for ongoing molecular studies in cork oak. *Quercus robur* (pedunculate oak) and *Quercus petraea* (sessile oak) have been in the center of the studies in oaks. The development of an expressed gene sequence database resulted from the first large scale analysis of the oaks transcriptome and data were generated from mRNA extracted from several tissues of the two oak species, some of which subjected to abiotic stresses (Ueno *et al.*, 2010). The analysis of orthologous relationships between oak and other species evidenced a higher number of orthologous gene pairs between oak and *Vitis*.

A database with genomic data from species of the Fagaceae family is also available (<http://www.fagaceae.org/>) and includes information on genetic and physical maps, transcriptomic analyses and functional analyses among other resources.

Gene expression analysis

In order to further characterize genes and gene networks identified in transcriptomic studies, accurate protocols for gene expression analysis need to be established.

Reverse transcription quantitative PCR (RT-qPCR) is a standard technique that allows the determination of gene transcript levels in different tissues and it is crucial in the identification of cross-regulation between genes in complex networks during plant development. This technique is widely used to validate the expression data obtained by deep sequencing experiments. However, for an accurate use of this method a reliable choice of internal control genes to normalize the obtained data is of outmost importance and is even essential for the success of the technique. Ideally, a reference gene (RG) should have a constant expression level in all cells of different tissues and under different experimental conditions, otherwise the determination of the expression levels of the target genes gives erroneous results. Therefore, no gene can act as a universal RG, as assumed in some older studies. Although the best RGs for gene expression experiments can be the same independently of the tissue and species in analysis (Jarosová and Kundu, 2010; Yang *et al.*, 2010; Li *et al.*, 2010), a selection of the most adequate RGs in each experiment should be made. In fact, there are many studies where the best RGs for gene expression experiments may vary between different species, or even within the same species between different tissues or applied treatments (Brunner *et al.*, 2004; Czechowski *et al.*, 2005; Remans *et al.*, 2008; Gutierrez *et al.*, 2008; Artico *et al.*, 2010; Huis *et al.*, 2010; Die *et al.*, 2010; Hong *et al.*, 2010; Migocka and Papierniak, 2011; De Oliveira *et al.*, 2012). Indeed, the traditional genes used for RT-qPCR normalization, such as *ACTIN* (*ACT*), α - and β -*TUBULIN* (α and β -*TUB*), *ELONGATION FACTOR-1 α* (*EF-1 α*), *UBIQUITIN* (*UBQ*), *GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* (*GAPDH*) and *18S rRNA* (Czechowski *et al.*, 2005;

Guénin *et al.*, 2009; Hong *et al.*, 2010), are no longer used in a pre-established way, and some reports evidenced that their expression levels can be variable (Remans *et al.*, 2008; Gutierrez *et al.*, 2008; Huis *et al.*, 2010; Die *et al.*, 2010; Li *et al.*, 2010; Hong *et al.*, 2010; Migocka and Papierniak, 2011). Thus, they need to be tested in the desired samples and their stability values need to be evaluated. Moreover, the use of a single RG for expression data normalization leads to misleading results (Vandesompele *et al.*, 2002; Die *et al.*, 2010), and therefore the use of multiple RGs has become the most common and accurate methodology to normalize the expression data (Basa *et al.*, 2009; Artico *et al.*, 2010; Huis *et al.*, 2010; Die *et al.*, 2010; Xu *et al.*, 2011; Migocka and Papierniak, 2011). Also, the minimum number of genes required to calculate a reliable normalization factor varies within different experiments and need to be verified in each case study (Vandesompele *et al.*, 2002). Statistical methods have been developed to identify the most stable genes in a group of potential RGs, and the most used were GeNorm (Vandesompele *et al.*, 2002), NormFinder (Andersen *et al.*, 2004), Coefficient of variation (CV) method (Hellemans *et al.*, 2007) and BestKeeper (Pfaffl *et al.*, 2004). Nowadays, the great majority of studies reporting the identification of adequate RGs for a given experiment uses at least two of those statistical methods (Artico *et al.*, 2010; Huis *et al.*, 2010; Jarosová and Kundu, 2010; Maroufi *et al.*, 2010; Demidenko *et al.*, 2011; Migocka and Papierniak, 2011). While several studies reporting RG selection have been made on different plant species, only few were made on trees, such as *Populus*, *Eucalyptus*, *Picea*, *Dimocarpus*, *Platycladus* and *Aquilaria* (Brunner *et al.*, 2004; Gutierrez *et al.*, 2008; Phillips *et al.*, 2009; Lin and Lai, 2010; Xu *et al.*, 2011; Chang *et al.*, 2012; De Oliveira *et al.*, 2012; Gao *et al.*, 2012) and many of them are quite recent. A lack of studies on recalcitrant tissues, like wood, bark or cork is evident. Nonetheless, two studies were made on cork oak, one reporting seasonal variation of specific genes where RGs were not submitted to an extensive analysis of genes expression stability (Soler *et al.*, 2008) and, more

recently, a work focusing on the expression of DNA methyltransferases in cork of different qualities described that the cork oak DNA methyltransferase *DOMAIN REARRANGED METHYLTRANSFERASE 2*, *QsDRM2*, was the most stable gene in the active phellogen and derived cells (Ramos *et al.*, 2013).

Overall, the combination of molecular and genomic data retrieved from cork oak, or model plant species like *Populus*, will contribute to elucidate important aspects of its biology such as the regulation of secondary growth with an impact on cork and periderm formation, or molecular features of acorn development with a potential impact on the establishment of strategies to improve plant perpetuation and regeneration.

Research objectives and thesis layout

The peculiar and amazing biologic characteristics of cork oak, which are the drivers of its socio-economic relevance and recognition as tree of national interest, have also been the inspiration for this thesis. The general aim of the work here described was to provide novel insights into the regulation of specific post-embryonic developmental processes in cork oak, focusing on secondary growth and particularly the cork formation process, but also contributing new tools for the dissection of other processes with an emphasis on fruit development.

The plant specific transcription factor *SHORT-ROOT (SHR)* had been previously characterized in the functioning of the root apical meristem and the root development process of the model plant species *Arabidopsis*. Strong evidences also pointed to a possible function in other meristems such as the lateral meristems at the origin of secondary growth. We proposed to investigate the role of *SHR* gene in order to provide a deeper knowledge of

the mechanisms regulating secondary growth and search for a possible role in the regulation of the phellogen activity. The model tree *Populus* was used for this purpose, but additional studies have been pursued in *Quercus suber* (cork oak).

Another aim of this study was to perform a transcriptomic analysis of cork oak developing fruits as part of a larger effort to obtain a reference transcriptome of the cork oak, thus contributing to the understanding of important aspects of its reproductive biology which may be important for the implementation of successful regeneration programs.

The specific objectives of this work were:

1. To understand the role of *SHR*-like genes during the secondary growth of *Populus tremula* L. x *Populus tremuloides* Michx.; Clone T89 (hybrid aspen), particularly *PtSHR2B*, by analyzing specific expression patterns and altering transcript levels to study the effects on plant growth and stem anatomy. (Chapter II)
2. To establish basic procedures for accurate transcript quantification in cork oak tissues through identification of adequate reference genes. (Chapter III)
3. To characterize the cork oak *SHR* homologs and their putative roles during secondary growth by analyzing their expression patterns along periderm development and phellogen dormancy/activity periods. Additional gene characterization approaches included comparative studies with a related species (holm oak) that does not produce a thick phellem layer, and functional studies in *Arabidopsis*. (Chapter IV)
4. To provide a broad view of cork oak fruit development using a transcriptomic approach, thus contributing for the release of the cork

oak reference transcriptome and allowing the identification of genes differentially regulated between different fruit developmental stages. Genes related to carbohydrate metabolism, response to water and transcriptional control were specifically targeted. (Chapter V)

This thesis is in the form of articles, from Chapters II to V, and the work followed the steps described in Figure 5. Final conclusions and future perspectives are discussed in Chapter VI.

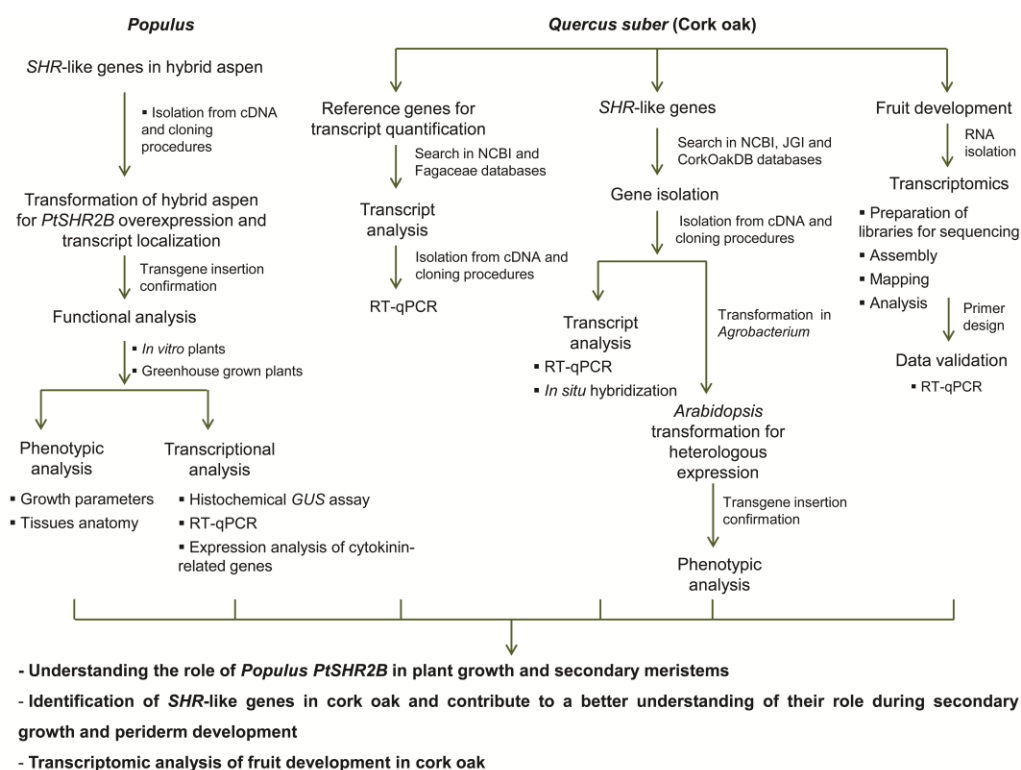


Fig. 5. General organization of the research and thesis, highlighting the main techniques used during thesis studies.

References

Agusti J, Lichtenberger R, Schwarz M, Nehlin L, Greb T. 2011. Characterization of transcriptome remodeling during cambium formation identifies *MOL1* and *RUL1* as opposing regulators of secondary growth. *PLoS Genetics* **7**, e1001312.

Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, Nussaume L, Noh Y, Amasino R, Scheres B, et al. 2004. The *PLETHORA* genes mediate patterning of the *Arabidopsis* root stem cell niche. *Cell* **119**, 109–120.

Aloni R. 1982. Role of cytokinin in differentiation of secondary xylem fibers. *Plant Physiology* **70**, 1631–1633.

Aloni R. 1987. Differentiation of vascular tissues. *Plant Physiology* **38**, 179–204.

Altamura MM, Possenti M, Matteucci A, Baima S, Ruberti I, Morelli G. 2001. Development of the vascular system in the inflorescence stem of *Arabidopsis*. *New Phytologist* **151**, 381–389.

Andersen CL, Jensen JL, Ørntoft TF. 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets normalization of real-time quantitative reverse. *Cancer Research* **64**, 5245–5250.

Anjos O, Pereira H, Rosa ME. 2008. Effect of quality, porosity and density on the compression properties of cork. *Holz als Roh- und Werkstoff* **66**, 295–301.

APCOR. 2012. Cork 2012. , 1–76.

Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M. 2010. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biology* **10**, 49.

Baba K, Karlberg A, Schmidt J, Schrader J, Hvidsten TR, Bako L. 2011. Activity – dormancy transition in the cambial meristem involves stage-specific modulation of auxin response in hybrid aspen. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 3418–3423.

Baima S, Nobili F, Sessa G, Lucchetti S, Ruberti I, Morelli G. 1995. The expression of the *Athb-8* homeobox gene is restricted to provascular cells in *Arabidopsis thaliana*. *Development* **121**, 4171–4182.

Baima S, Possenti M, Matteucci A, Wisman E, Altamura MM, Ruberti I, Morelli G, Vegetale B, Sapienza L, Moro PA. 2001. The *Arabidopsis* ATHB-8 HD-Zip protein

acts as a differentiation-promoting transcription factor of the vascular meristems. *Plant Physiology* **126**, 643–655.

Barton MK. 2010. Twenty years on: the inner workings of the shoot apical meristem, a developmental dynamo. *Developmental Biology* **341**, 95–113.

Basa B, Solti Á, Sárvári É, Tamas L. 2009. Housekeeping gene selection in poplar plants under Cd-stress: comparative study for real-time PCR normalisation. *Functional Plant Biology* **36**, 1–9.

Baucher M, El Jaziri M, Vandeputte O. 2007. From primary to secondary growth: origin and development of the vascular system. *Journal of Experimental Botany* **58**, 3485–3501.

Benfey PN, Linstead PJ, Roberts K, Schiefelbein JW, Hauser MT, Aeschbacher RA. 1993. Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* **119**, 57–70.

Bishopp A, Help H, El-Showk S, Weijers D, Scheres B, Friml J, Benková E, Mähönen AP, Helariutta Y. 2011. A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots. *Current Biology* **21**, 917–926.

Björklund S, Antti H, Uddestrand I, Moritz T, Sundberg B. 2007. Cross-talk between gibberellin and auxin in development of *Populus* wood: gibberellin stimulates polar auxin transport and has a common transcriptome with auxin. *The Plant Journal: for cell and molecular biology* **52**, 499–511.

Boavida LC, Varela C, Feijó JA. 1999. Sexual reproduction in the cork oak (*Quercus suber* L.). I. The progametic phase. *Sexual Plant Reproduction* **11**, 347–353.

Bolle C. 2004. The role of GRAS proteins in plant signal transduction and development. *Planta* **218**, 683–692.

Bradshaw Jr HD, Ceulemans R, Davis J, Stettler R. 2000. Emerging Model Systems in Plant Biology: Poplar (*Populus*) as A Model Forest Tree. *Journal of Plant Growth Regulation* **19**, 306–313.

Brand U, Fletcher J, Hobe M, Meyerowitz E, Simon R. 2000. Dependence of Stem Cell Fate in *Arabidopsis* on a Feedback Loop Regulated by CLV3 Activity. *Science* **289**, 617–619.

Brunner AM, Yakovlev IA, Strauss SH. 2004. Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biology* **4**, 14.

Bugalho MN, Caldeira MC, Pereira JS, Aronson J, Pausas JG. 2011. Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. *Frontiers in Ecology and the Environment* **9**, 278–286.

Caño-Delgado A, Yin Y, Yu C, Vafeados D, Mora-García S, Cheng J-C, Nam KH, Li J, Chory J. 2004. BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in *Arabidopsis*. *Development* **131**, 5341–5351.

Caño-Delgado A, Lee J-Y, Demura T. 2010. Regulatory mechanisms for specification and patterning of plant vascular tissues. *Annual review of Cell and Developmental Biology* **26**, 605–637.

Carimi F, Zottini M, Formentin E, Terzi M, Lo Schiavo F. 2003. Cytokinins: new apoptotic inducers in plants. *Planta* **216**, 413–421.

Carimi F, Terzi M, De Michele R, Zottini M, Lo Schiavo F. 2004. High levels of the cytokinin BAP induce PCD by accelerating senescence. *Plant Science* **166**, 963–969.

Caritat A, Molinas M, Gutierrez E. 1996. Annual cork-ring width variability of *Quercus suber* L. in relation to temperature and precipitation (Extremadura, southwestern Spain). *Forest Ecology and Management* **86**, 113–120.

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

Carlsbecker A, Lee J-Y, Roberts CJ, Dettmer J, Lehesranta S, Zhou J, Lindgren O, Moreno-Risueno MA, Vatén A, Thitamadee S, et al. 2010. Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* **465**, 316–321.

Chaffey N, Cholewa E, Regan S, Sundberg B. 2002. Secondary xylem development in *Arabidopsis*: a model for wood formation. *Physiologia Plantarum* **114**, 594–600.

Chang E, Shi S, Liu J, Cheng T, Xue L, Yang X, Yang W, Lan Q, Jiang Z. 2012. Selection of reference genes for quantitative gene expression studies in *Platycladus orientalis* (Cupressaceae) using real-time PCR. *PLoS One* **7**, e33278.

Clark SE, Jacobsen SE, Levin JZ, Meyerowitz EM. 1996. The *CLAVATA* and *SHOOT MERISTEMLESS* loci competitively regulate meristem activity in *Arabidopsis*. *Development* **122**, 1567–1575.

Costa A, Pereira H, Oliveira A. 2001. A dendroclimatological approach to diameter growth in adult cork-oak trees under production. *Trees* **15**, 438–443.

Costa A, Pereira H, Oliveira Â. 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.

Costa A, Pereira H, Oliveira A. 2003. Variability of radial growth in cork oak adult trees under cork production. *Forest Ecology and Management* **175**, 239–246.

Cui H, Levesque MP, Vernoux T, Jung JW, Paquette AJ, Gallagher KL, Wang JY, Blilou I, Scheres B, Benfey PN. 2007. An evolutionarily conserved mechanism delimiting SHR movement defines a single layer of endodermis in plants. *Science* **316**, 421–425.

Cui H, Hao Y, Kovtun M, Stolc V, Deng X-W, Sakakibara H, Kojima M. 2011. Genome-wide direct target analysis reveals a role for SHORT-ROOT in root vascular patterning through cytokinin homeostasis. *Plant Physiology* **157**, 1221–1231.

Czechowski T, Stitt M, Altmann T, Udvardi MK. 2005. Genome-wide identification and testing of superior reference genes for transcript normalization. *Plant Physiology* **139**, 5–17.

Demidenko N V, Logacheva MD, Penin AA. 2011. Selection and validation of reference genes for quantitative real-time PCR in buckwheat (*Fagopyrum esculentum*) based on transcriptome sequence data. *PLoS ONE* **6**, e19434.

Derory J, Léger P, Garcia V, Schaeffer J, Hauser M-T, Salin F, Luschign C, Plomion C, Glössl J, Kremer A. 2006. Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *The New Phytologist* **170**, 723–738.

Dettmer J, Elo A, Helariutta Y. 2009. Hormone interactions during vascular development. *Plant Molecular Biology* **69**, 347–360.

Dhondt S, Coppens F, De Winter F, Swarup K, Merks RMH, Inzé D, Bennett MJ, Beemster GTS. 2010. SHORT-ROOT and SCARECROW regulate leaf growth in Arabidopsis by stimulating S-phase progression of the cell cycle. *Plant Physiology* **154**, 1183–1195.

Díaz-Fernández PM, Climent J, Gil L. 2004. Biennial acorn maturation and its relationship with flowering phenology in Iberian populations of *Quercus suber*. *Trees* **18**, 615–621.

Die J V, Román B, Nadal S, González-Verdejo CI. 2010. Evaluation of candidate reference genes for expression studies in *Pisum sativum* under different experimental conditions. *Planta* **232**, 145–153.

Dinneny JR, Benfey PN. 2008. Plant stem cell niches: standing the test of time. *Cell* **132**, 553–557.

Donner TJ, Sherr I, Scarpella E. 2009. Regulation of preprocambial cell state acquisition by auxin signaling in *Arabidopsis* leaves. *Development* **136**, 3235–3246.

Du J, Mansfield SD, Groover AT. 2009. The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. *The Plant Journal* **60**, 1000–1014.

Du J, Groover A. 2010. Transcriptional regulation of secondary growth and wood formation. *Journal of Integrative Plant Biology* **52**, 17–27.

Durand J, Bodénès C, Chancerel E, Frigerio J-M, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn H-P, Villani F, et al. 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* **11**, 570.

Elena-Rosello J, Rio J, Valdecantos JG, Santamaria I. 1993. Ecological aspects of the floral phenology of the cork-oak (*Q suber* L.): why do annual and biennial biotypes appear? *Annals of Forest Science* **50**, 114–121.

Elo A, Immanen J, Nieminen K, Helariutta Y. 2009. Stem cell function during plant vascular development. *Seminars in Cell & Developmental Biology* **20**, 1097–1106.

Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL. 2003. Radial patterning of *Arabidopsis* shoots by class III HD-ZIP and KANADI genes. *Current Biology* **13**, 1768–1774.

Eriksson ME, Israelsson M, Olsson O, Moritz T. 2000. Increased gibberellin biosynthesis in transgenic trees promotes growth, biomass production and xylem fiber length. *Nature Biotechnology* **18**, 784–788.

Esau K. 1977. *Anatomy of seed plants*. New York: John Wiley & Sons.

Etchells JP, Turner SR. 2010. The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. *Development* **137**, 767–774.

Evert R. 2006a. Vascular cambium. *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body – Their Structure, Function and Development*. New Jersey: John Wiley & Sons, Inc., 323–355.

Evert R. 2006b. Periderm. *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development*. New Jersey: John Wiley & Sons, Inc., 427–445.

Faivre Rampant P, Lesur I, Boussardon C, Bitton F, Martin-Magniette M-L, Bodénès C, Le Provost G, Bergès H, Fluch S, Kremer A, et al. 2011. Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. *BMC Genomics* **12**, 292.

Fialho C, Lopes F, Pereira H. 2001. The effect of cork removal on the radial growth and phenology of young cork oak trees. *Forest Ecology and Management* **141**, 251–258.

Fisher K, Turner S. 2007. PXY, a receptor-like kinase essential for maintaining polarity during plant vascular-tissue development. *Current Biology* **17**, 1061–1066.

Fukaki H, Wysocka-Diller J, Kato T, Fujisawa H, Benfey PN, Tasaka M. 1998. Genetic evidence that the endodermis is essential for shoot gravitropism in *Arabidopsis thaliana*. *The Plant Journal* **14**, 425–430.

Fukuda H. 1996. Xylogenesis: Initiation, Progression, and Cell Death. Annual review of Plant Physiology and Plant Molecular Biology **47**, 299–325.

Fukuda H. 2004. Signals that control plant vascular cell differentiation. Nature reviews. Molecular cell biology **5**, 379–391.

Gallagher KL, Paquette AJ, Nakajima K, Benfey PN, Carolina N. 2004. Mechanisms Regulating SHORT-ROOT Intercellular Movement. *Current Biology* **14**, 1847–1851.

Gallagher KL, Benfey PN. 2010. Both the conserved GRAS domain and nuclear localization are required for SHORT-ROOT movement. *The Plant Journal* **57**, 785–797.

Gandini A, Pascoal Neto C, Silvestre AJD. 2006. Suberin: A promising renewable resource for novel macromolecular materials. *Progress in Polymer Science* **31**, 878–892.

Gao Z-H, Wei J-H, Yang Y, Zhang Z, Zhao W-T. 2012. Selection and validation of reference genes for studying stress-related agarwood formation of *Aquilaria sinensis*. *Plant Cell Reports* **31**, 1759–1768.

Gardiner J, Donner TJ, Scarpella E. 2011. Simultaneous activation of *SHR* and *ATHB8* expression defines switch to preprocambial cell state in *Arabidopsis* leaf development. *Developmental Dynamics* **240**, 261–270.

geNorm software manual.

Gil L. Cork as a building material. 1–66.

Ginzberg I, Barel G, Ophir R, Tzin E, Tanami Z, Muddarangappa T, De Jong W, Fogelman E. 2009. Transcriptomic profiling of heat-stress response in potato periderm. *Journal of Experimental Botany* **60**, 4411–4421.

Glover BJ. 2010. Epidermis: Outer Cell Layer of the Plant. Encyclopedia of Life Sciences (ELS). Chichester: John Wiley & Sons.

Gonzalez-Adrados J, Pereira H. 1996. Classification of defects in cork planks using image analysis. Wood Science and Technology **30**, 207–215.

Gonzalez-Adrados JR, Lopes F, Pereira H. 2000. Quality grading of cork planks with classification models based on defect characterisation. Holz als Roh- und Werkstoff **58**, 39–45.

Goodman HM, Ecker JR, Deanl C. 1995. The genome of *Arabidopsis thaliana*. Proceedings of the National Academy of Sciences of the United States of America **92**, 10831–10835.

Gordon SP, Chickarmane VS, Ohno C, Meyerowitz EM. 2009. Multiple feedback loops through cytokinin signaling control stem cell number within the *Arabidopsis* shoot meristem. Proceedings of the National Academy of Sciences of the United States of America **106**, 16529–16534.

Graça J, Pereira H. 2004. The periderm development in *Quercus suber*. IAWA Journal **25**, 325–335.

Green KA, Prigge MJ, Katzman RB, Clark SE. 2005. CORONA, a member of the class III homeodomain leucine zipper gene family in *Arabidopsis*, regulates stem cell specification and organogenesis. The Plant Cell **17**, 691–704.

Groover AT. 2005. What genes make a tree a tree? Trends in Plant Science **10**, 210–214.

Groover AT, Mansfield SD, DiFazio SP, Dupper G, Fontana JR, Millar R, Wang Y. 2006. The *Populus* homeobox gene ARBORKNOX1 reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. Plant Molecular Biology **61**, 917–932.

Guénin S, Mauriat M, Pelloux J, Van Wuytswinkel O, Bellini C, Gutierrez L. 2009. Normalization of qRT-PCR data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. Journal of Experimental Botany **60**, 487–93.

Gutierrez L, Mauriat M, Guénin S, Pelloux J, Lefebvre J-F, Louvet R, Rusterucci C, Moritz T, Guérineau F, Bellini C, et al. 2008. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. Plant Biotechnology Journal **6**, 609–618.

Haecker A, Gross-Hardt R, Geiges B, Sarkar A, Breuninger H, Herrmann M, Laux T. 2004. Expression dynamics of *WOX* genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development* **131**, 657–668.

Hao Y, Cui H. 2012. SHORT-ROOT regulates vascular patterning, but not apical meristematic activity in the *Arabidopsis* root through cytokinin homeostasis. *Plant Signaling & Behavior* **7**, 314–317.

Hawker NP, Bowman JL. 2004. Roles for class III HD-Zip and KANADI genes in *Arabidopsis* root development. *Plant Physiology* **135**, 2261–2270.

Heidstra R, Welch D, Scheres B. 2004. Mosaic analyses using marked activation and deletion clones dissect *Arabidopsis* SCARECROW action in asymmetric cell division. *Genes & Development* **18**, 1964–1969.

Hejatko J, Ryu H, Kim G-T, Dobesova R, Choi S, Choi SM, Soucek P, Horak J, Pekarova B, Palme K, et al. 2009. The histidine kinases CYTOKININ-INDEPENDENT1 and ARABIDOPSIS HISTIDINE KINASE2 and 3 regulate vascular tissue development in *Arabidopsis* shoots. *The Plant Cell* **21**, 2008–2021.

Hejnowicz A. 2007. Anatomy, embryology, and karyology. Bud structure and shoot development. In: Tjoelker MG, Boratynski A, Bugala W, eds. *Biology and Ecology of Norway Spruce*. Dordrecht, Netherlands: Elsevier, 57.

Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN. 2000. The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* **101**, 555–567.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**, R19.

Hertzberg M, Aspeborg H, Schrader J, Andersson A, Erlandsson R, Blomqvist K, Teeri TT, Lundeberg J, Nilsson P, Bhalerao R, et al. 2001. A transcriptional roadmap to wood formation. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 14732–14737.

Hirakawa Y, Shinohara H, Kondo Y, Inoue A, Nakanomyo I, Ogawa M, Sawa S, Ohashi-Ito K, Matsubayashi Y, Fukuda H. 2008. Non-cell-autonomous control of vascular stem cell fate by a CLE peptide/receptor system. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 15208–15213.

Hirakawa Y, Kondo Y, Fukuda H. 2010. TDIF Peptide Signaling Regulates Vascular Stem Cell Proliferation via the *WOX4* Homeobox Gene in *Arabidopsis*. *The Plant Cell* **22**, 2618–2629.

Holloway PJ. 1983. Some variations in the composition of suberin from the cork layers of higher plants. *Phytochemistry* **22**, 495–502.

Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH. 2010. Identification and testing of superior reference genes for a starting pool of transcript normalization in *Arabidopsis*. *Plant & Cell Physiology* **51**, 1694–1706.

Huis R, Hawkins S, Neutelings G. 2010. Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). *BMC Plant Biology* **10**, 71.

Ibañez M, Fàbregas N, Chory J, Caño-Delgado AI. 2009. Brassinosteroid signaling and auxin transport are required to establish the periodic pattern of *Arabidopsis* shoot vascular bundles. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 13630–13635.

Ilegems M, Douet V, Meylan-Bettex M, Uyttewaal M, Brand L, Bowman JL, Stieger P a. 2010. Interplay of auxin, KANADI and Class III HD-ZIP transcription factors in vascular tissue formation. *Development* **137**, 975–984.

Imin N, Nizamidin M, Wu T, Rolfe BG. 2007. Factors involved in root formation in *Medicago truncatula*. *Journal of Experimental Botany* **58**, 439–451.

Jacobs WP. 1952. The role of auxin in differentiation of xylem around a wound. *American Journal of Botany* **39**, 301–309.

Jarosová J, Kundu JK. 2010. Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR. *BMC Plant Biology* **10**, 146.

Ji J, Shimizu R, Sinha N, Scanlon MJ. 2010. WOX4 Promotes Procambial Development. *Plant Physiology* **152**, 1346–1356.

Kamiya N, Itoh J-I, Morikami A, Nagato Y, Matsuoka M. 2003. The *SCARECROW* gene's role in asymmetric cell divisions in rice plants. *The Plant Journal* **36**, 45–54.

Kaufert F. 1937. Factors influencing the formation of periderm in aspen. *American Journal of Botany* **24**, 24–30.

Ko J, Han K, Park S, Yang J. 2004. Plant body weight-induced secondary growth in *Arabidopsis* and its transcription phenotype revealed by whole-transcriptome profiling. *Plant Physiology* **135**, 1069–1083.

Ko J, Prassinos C, Han K. 2005. Developmental and seasonal expression of *PtaHB1*, a *Populus* gene encoding a class III HD-Zip protein, is closely associated

with secondary growth and inversely correlated with the level of microRNA (*miR166*). *New Phytologist*, 469–478.

Koizumi K, Wu S, MacRae-Crerar A, Gallagher KL. 2011. An essential protein that interacts with endosomes and promotes movement of the SHORT-ROOT transcription factor. *Current Biology* **21**, 1559–1564.

Koizumi K, Hayashi T, Wu S, Gallagher KL. 2012. The SHORT-ROOT protein acts as a mobile, dose-dependent signal in patterning the ground tissue. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13010–13015.

Kremer A, Abbott AG, Carlson JE, Manos PS, Plomion C, Sisco P, Staton ME, Ueno S, Vendramin GG. 2012. Genomics of Fagaceae. *Tree Genetics & Genomes* **8**, 583–610.

Kunikowska A, Byczkowska A, Doniak M, Kaźmierczak A. 2013a. Cytokinins résumé: their signaling and role in programmed cell death in plants. *Plant Cell Reports* **32**, 771–780.

Kunikowska A, Byczkowska A, Kaźmierczak A. 2013b. Kinetin induces cell death in root cortex cells of *Vicia faba* ssp. *minor* seedlings. *Protoplasma* **250**, 851–861.

Lachaud S, Catesson AM, Bonnemain JL. 1999. Structure and functions of the vascular cambium. *Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie* **322**, 633–650.

Laurenzio L Di, Wysocka-diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. 1996. The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* **86**, 423–433.

Laux T. 2003. The stem cell concept in plants: a matter of debate. *Cell* **113**, 281–283.

Leibried A, To JPC, Busch W, Stehling S, Kehle A, Demar M, Kieber JJ, Lohmann JU. 2005. WUSCHEL controls meristem function by direct regulation of cytokinin-inducible response regulators. *Nature* **438**, 1172–1175.

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

Lesur I, Durand J, Sebastiani F, Gyllenstrand N, Bodénès C, Lascoux M, Kremer A, Vendramin GG, Plomion C. 2011. A sample view of the pedunculate oak (*Quercus robur*) genome from the sequencing of hypomethylated and random genomic libraries. *Tree Genetics & Genomes* **7**, 1277–1285.

Levesque MP, Vernoux T, Busch W, Cui H, Wang JY, Blilou I, Hassan H, Nakajima K, Matsumoto N, Lohmann JU, et al. 2006. Whole-genome analysis of the SHORT-ROOT developmental pathway in *Arabidopsis*. *PLoS Biology* **4**, e143.

Li Q-F, Sun SSM, Yuan D-Y, Yu H-X, Gu M-H, Liu Q-Q. 2010. Validation of candidate reference genes for the accurate normalization of real-time quantitative RT-PCR data in rice during seed development. *Plant Molecular Biology Reporter* **28**, 49–57.

Lin YL, Lai ZX. 2010. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Science* **178**, 359–365.

Liphschitz N, Lev-Yadun S, Rosen E, Waisel Y. 1984. The annual rhythm of activity of the lateral meristems (cambium and phellogen) in *Pinus Halepensis* Mill. and *Pinus Pinea* L. *IAWA Bulletin* **5**, 263–274.

Little CHA, Savidge RA. 1987. The role of plant growth regulators in forest tree cambial growth. *Plant Growth Regulation* **6**, 137–169.

Long JA, Moan EI, Medford JI, Barton MK. 1996. A member of the KNOTTED class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* **379**, 66–69.

Love J, Björklund S, Vahala J, Hertzberg M, Kangasjärvi J, Sundberg B. 2009. Ethylene is an endogenous stimulator of cell division in the cambial meristem of *Populus*. *Proceedings of the National Academy of Sciences of the United States of America* **106**, 5984–5989.

Lucas M, Swarup R, Paponov IA, Swarup K, Casimiro I, Lake D, Peret B, Zappala S, Mairhofer S, Whitworth M, et al. 2011. Short-Root regulates primary, lateral, and adventitious root development in *Arabidopsis*. *Plant Physiology* **155**, 384–398.

Lulai E, Corsini D. 1998. Differential deposition of suberin phenolic and aliphatic domains and their roles in resistance to infection during potato tuber (*Solanum tuberosum* L.) wound-healing. *Physiological and Molecular Plant Pathology* **53**, 209–222.

Lulai E, Freeman TP. 2001. The importance of phellogen cells and their structural characteristics in susceptibility and resistance to excoriation in immature and mature potato tuber (*Solanum tuberosum* L.) periderm. *Annals of Botany* **88**, 555–561.

Machado D. 1935. Contribuição para o estudo da formação da cortiça no sobreiro. *Revista Agronómica* **23**, 75–104.

Mähönen AP, Bonke M, Kauppinen L, Riikonen M, Benfey PN. 2000. A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes & Development* **14**, 2938–2943.

Mähönen AP. 2005. Cytokinins Regulate Vascular Morphogenesis in the *Arabidopsis thaliana* root. PhD thesis, University of Helsinki.

Mähönen AP, Bishopp A, Higuchi M, Nieminen KM, Kinoshita K, Törmäkangas K, Ikeda Y, Oka A, Kakimoto T, Helariutta Y. 2006. Cytokinin signaling and its inhibitor AHP6 regulate cell fate during vascular development. *Science* **311**, 94–98.

Maroufi A, Van Bockstaele E, De Loose M. 2010. Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Molecular Biology* **11**, 15.

Matsumoto-Kitano M, Kusumoto T, Tarkowski P, Kinoshita-Tsujimura K, Václavíková K, Miyawaki K, Kakimoto T. 2008. Cytokinins are central regulators of cambial activity. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20027–20031.

Mayer KF, Schoof H, Haecker a, Lenhard M, Jürgens G, Laux T. 1998. Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95**, 805–815.

McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK. 2001. Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. *Nature* **411**, 709–713.

Mele G, Ori N, Sato Y, Hake S. 2003. The knotted1-like homeobox gene *BREVIPEDICELLUS* regulates cell differentiation by modulating metabolic pathways. *Genes & Development* **17**, 2088–2093.

Migocka M, Papierniak A. 2011. Identification of suitable reference genes for studying gene expression in cucumber plants subjected to abiotic stress and growth regulators. *Molecular Breeding* **28**, 343–357.

Milhinhos A, Miguel CM. 2013. Hormone interactions in xylem development: a matter of signals. *Plant Cell Reports* **32**, 867–883.

Miller AR, Pengelly WL, Roberts LW. 1984. Introduction of xylem differentiation in lactuca by ethylene. *Plant Physiology* **75**, 1165–1166.

Miranda I, Gominho J, Pereira H. 2012. Cellular structure and chemical composition of cork from the Chinese cork oak (*Quercus variabilis*). *Journal of Wood Science* **59**, 1–9.

Miyashima S, Sebastian J, Lee J-Y, Helariutta Y. 2013. Stem cell function during plant vascular development. *The EMBO Journal* **32**, 178–193.

Myburg A, Sederoff R. 2001. Xylem Structure and Function. *Encyclopedia of life sciences*. Wiley Interscience Publisher, 1–9.

Nakajima K, Sena G, Nawy T, Benfey PN. 2001. Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* **413**, 307–311.

Natividade J. 1950. *Subericultura*. Ministério da Economia, Pescas e Alimentação, Direcção Geral das Florestas e Aquícolas, Lisboa.

Neubauer JD, Lulai EC, Thompson AL, Suttle JC, Bolton MD. 2012. Wounding coordinately induces cell wall protein, cell cycle and pectin methyl esterase genes involved in tuber closing layer and wound periderm development. *Journal of Plant Physiology* **169**, 586–595.

Nieminen K, Immanen J, Laxell M, Kauppinen L, Tarkowski P, Dolezal K, Tahtiharju S, Elo A, Decourteix M, Ljung K, et al. 2008. Cytokinin signaling regulates cambial development in poplar. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20032–20037.

Nilsson O, Aldén T, Sitbon F, Anthony Little CH, Chalupa V, Sandberg G, Olsson O. 1992. Spatial pattern of cauliflower mosaic virus 35S promoter-luciferase expression in transgenic hybrid aspen trees monitored by enzymatic assay and non-destructive imaging. *Transgenic Research* **1**, 209–220.

Nilsson J, Karlberg A, Antti H, Lopez-Vernaza M, Mellerowicz E, Perrot-Rechenmann C, Sandberg G, Bhalerio RP. 2008. Dissecting the molecular basis of the regulation of wood formation by auxin in hybrid aspen. *The Plant Cell* **20**, 843–855.

O’Gara E, Howard K, Colquhoun IJ, Dell B, McComb J, Hardy GSEJ. 2009. The development and characteristics of periderm and rhytidome in *Eucalyptus marginata*. *Australian Journal of Botany* **57**, 221.

Ogawa M, Shinohara H, Sakagami Y, Matsubayashi Y. 2008. *Arabidopsis* CLV3 peptide directly binds CLV1 ectodomain. *Science* **319**, 294.

Ohashi-Ito K, Demura T, Fukuda H. 2002. Promotion of transcript accumulation of novel *Zinnia* immature xylem-specific HD-Zip III homeobox genes by brassinosteroids. *Plant & Cell Physiology* **43**, 1146–1153.

Ohashi-Ito K, Fukuda H. 2003. HD-zip III homeobox genes that include a novel member, ZeHB-13 (*Zinnia*)/ATHB-15 (*Arabidopsis*), are involved in procambium and xylem cell differentiation. *Plant & Cell Physiology* **44**, 1350–1358.

Ohashi-Ito K, Kubo M, Demura T, Fukuda H. 2005. Class III homeodomain leucine-zipper proteins regulate xylem cell differentiation. *Plant & Cell Physiology* **46**, 1646–1656.

Oliveira G, Correia O, Martins-Ioud MA, Catarino FM. 1994. Phenological and growth patterns of the Mediterranean oak *Quercus suber* L. *Trees* **9**, 41–46.

De Oliveira LA, Breton MC, Bastolla FM, Camargo SDS, Margis R, Frazzon J, Pasquali G. 2012. Reference genes for the normalization of gene expression in eucalyptus species. *Plant & Cell Physiology* **53**, 405–422.

Otsuga D, DeGuzman B, Prigge MJ, Drews GN, Clark SE. 2001. *REVOLUTA* regulates meristem initiation at lateral positions. *The Plant Journal* **25**, 223–236.

Pang PP, Meyerowitz EM. 1987. *Arabidopsis Thaliana*: A Model System for Plant Molecular Biology. *Nature* **5**, 1177–1181.

Paquette AJ, Benfey PN. 2005. Maturation of the ground tissue of the root is regulated by gibberellin and *SCARECROW* and requires *SHORT-ROOT*. *Plant Physiology* **138**, 636–640.

Perales M, Reddy GV. 2012. Stem cell maintenance in shoot apical meristems. *Current Opinion in Plant Biology* **15**, 10–16.

Pereira H, Rosa ME, Fortes M. 1987. The cellular structure of cork from *Quercus suber* L. *IAWA Bulletin* **8**, 213–218.

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

Pereira H, Graça J, Baptista C. 1992. The effect of growth rate on the structure and compressive properties of cork. *IAWA Bulletin* **13**, 389–396.

Pereira H, Lopes F, Graça J. 1996. The Evaluation of the Quality of Cork Planks by Image Analysis. *Holzforschung - International Journal of the Biology, Chemistry, Physics and Technology of Wood* **50**, 111–115.

Pereira H. 2007. *Cork: Biology, Production and Uses*. Amsterdam: Elsevier.

Pereira H. 2013. Variability of the Chemical Composition of Cork. *BioResources* **8**, 2246–2256.

Pereira JS, Bugalho MN, Caldeira M da C. 2008. *From the Cork Oak to cork: a sustainable system (APCOR)*.

Pereira-Leal JB, Abreu IA, Alabaça CS, Almeida MH, Almeida P, Almeida T, Amorim MI, Araújo S, Azevedo H, Badia A, et al. A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. 2014. BMC Genomics **15**, 371.

Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP. 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnology Letters **26**, 509–515.

Phillips MA, D'Auria JC, Luck K, Gershenzon J. 2009. Evaluation of candidate reference genes for real-time quantitative PCR of plant samples using purified cDNA as template. Plant Molecular Biology Reporter **27**, 407–416.

Pons J, Pausas JG. 2012. The coexistence of acorns with different maturation patterns explains acorn production variability in cork oak. Oecologia **169**, 723–731.

Prigge MJ, Otsuga D. 2005. Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in Arabidopsis development. The Plant Cell **17**, 61–76.

Quilhó T, Sousa V, Tavares F, Pereira H. 2013. Bark anatomy and cell size variation in *Quercus faginea*. Turkish Journal of Botany **37**, 561–570.

Ramos M, Rocheta M, Carvalho L, Inácio V, Graça J, Morais-Cecilio L. 2013. Expression of DNA methyltransferases is involved in *Quercus suber* cork quality. Tree Genetics & Genomes **9**, 1481–1492.

Raven P, Evert R, Eichhorn S. 2005. *Biology of plants*. New York: WH Freeman and Company.

Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A. 2008. Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. Planta **227**, 1343–1349.

Ricardo CPP, 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.

Ridoutt BG, Pliaris RP, Sands R. 1996. Fibre length and gibberellins A_i and A₂₀ are decreased in *Eucalyptus globulus* by acycyclohexanedione injected into the stem. Physiologia Plantarum **96**, 559–566.

- Robischon M, Du J, Miura E, Groover A.** 2011. The *Populus* class III HD ZIP, *popREVOLUTA*, influences cambium initiation and patterning of woody stems. *Plant Physiology* **155**, 1214–1225.
- Roth I.** 1981. *Structural patterns of tropical barks*. Berlin: Gebruder Borntraeger Verlagsbuchhandlung, Science Publishers.
- Rozen S, Skaletsky H.** 2000. Primer3 on the WWW for general users and for Biologist Programmers. *Methods in Molecular Biology* **132**, 365–386.
- Sabatini S, Heidstra R, Wildwater M, Scheres B.** 2003. SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes & Development* **17**, 354–358.
- Santamaria ME, Rodríguez R, Cañal MJ, Toorop PE.** 2011. Transcriptome analysis of chestnut (*Castanea sativa*) tree buds suggests a putative role for epigenetic control of bud dormancy. *Annals of Botany* **108**, 485–498.
- Sarkar AK, Luijten M, Miyashima S, Lenhard M, Hashimoto T, Nakajima K, Scheres B, Heidstra R, Laux T.** 2007. Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* **446**, 811–814.
- Savidge RA.** 1988. Auxin and ethylene regulation of diameter growth in trees. *Tree Physiology* **4**, 401–414.
- Scarpella E, Meijer AH.** 2004. Pattern formation in the vascular system of monocot and dicot plant species. *New Phytologist* **164**, 209–242.
- Scheres B, Laurenzio L Di, Willemsen V, Hauser M, Janmaat K, Weisbeek P, Benfey PN.** 1995. Mutations affecting the radial organisation of the *Arabidopsis* root display specific defects throughout the embryonic axis. *Development* **121**, 53–62.
- Schoof H, Lenhard M, Haecker A, Mayer KF, Jürgens G, Laux T.** 2000. The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* **100**, 635–644.
- Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M.** 2004. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *The Plant Cell* **16**, 2278–2292.
- Sehr EM, Agusti J, Lehner R, Farmer EE, Schwarz M, Greb T.** 2010. Analysis of secondary growth in the *Arabidopsis* shoot reveals a positive role of jasmonate signalling in cambium formation. *The Plant journal : for cell and molecular biology* **63**, 811–822.

Şen A, Miranda I, Santos S, Graça J, Pereira H. 2010. The chemical composition of cork and phloem in the rhytidome of *Quercus cerris* bark. *Industrial Crops and Products* **31**, 417–422.

Şen A, Quilhó T, Pereira H. 2011a. Bark anatomy of *Quercus cerris* L. var. *cerris* from Turkey. *Turkish Journal of Botany* **35**, 45–55.

Şen A, Quilhó T, Pereira H. 2011b. The cellular structure of cork from *Quercus cerris* var. *cerris* bark in a materials' perspective. *Industrial Crops and Products* **34**, 929–936.

Sena G, Jung JW, Benfey PN. 2004. A broad competence to respond to SHORT ROOT revealed by tissue-specific ectopic expression. *Development* **131**, 2817–2826.

Silva SP, Sabino MA, Fernandes EM, Correlo VM, Boesel LF, Reis RL. 2005. Cork: properties, capabilities and applications. *International Materials Reviews* **50**, 345–365.

Snow R. 1935. Activation of cambial growth by oure hormones. *The New Phytologist* **XXXIV**, 347–360.

Solé A, Sánchez C, Vielba JM, Valladares S, Abarca D, Díaz-Sala C. 2008. Characterization and expression of a *Pinus radiata* putative ortholog to the *Arabidopsis* *SHORT-ROOT* gene. *Tree Physiology* **28**, 1629–39.

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.

Soler M. 2008. Molecular genetics of cork formation. PhD thesis, Girona University.

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, Fluch S, Molinas M, Figueras M. 2011. A potato skin SSH library yields new candidate genes for suberin biosynthesis and periderm formation. *Planta* **233**, 933–945.

Soliday CL, Kolattukudy PE, Davis RW. 1979. Chemical and ultrastructural evidence that waxes associated with the suberin polymer constitute the major diffusion barrier to water vapor in potato tuber (*Solanum tuberosum* L.). *Planta* **146**, 607–614.

Sozzani R, Cui H, Moreno-Risueno MA, Bush W, Van Norman JM, Vernoux T, Brady SM, Dewitte W, Murray JAH, Benfey PN. 2010. Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. *Nature* **466**, 128–132.

Spicer R, Groover A. 2010. Evolution of development of vascular cambia and secondary growth. *The New Phytologist* **186**, 577–592.

Stahl Y, Simon R. 2009a. Is the *Arabidopsis* root niche protected by sequestration of the CLE40 signal by its putative receptor ACR4? *Plant Signaling & Behavior* **4**, 634–635.

Stahl Y, Simon R. 2009b. A signaling module controlling the stem cell niche in *Arabidopsis* root meristems. *Current Biology* **19**, 909–914.

Taylor G. 2002. *Populus*: *Arabidopsis* for Forestry. Do We Need a Model Tree? *Annals of Botany* **90**, 681–689.

Teixeira RT, Pereira H. 2010. Suberized cell walls of cork from cork oak differ from other species. *Microscopy and microanalysis: the official journal of Microscopy Society of America, Microbeam Analysis Society, Microscopical Society of Canada* **16**, 569–575.

Tuominen H, Puech L, Fink S, Sundberg B. 1997. A Radial Concentration Gradient of Indole-3-Acetic Acid Is Related to Secondary Xylem Development in Hybrid Aspen. *Plant Physiology* **115**, 577–585.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604.

Ueno S, Le Provost G, Léger V, Klopp C, Noirot C, Frigerio J-M, Salin F, Salse J, Abrouk M, Murat F, et al. 2010. Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* **11**, 650.

Ueno S, Klopp C, Leplé JC, Derory J, Noirot C, Léger V, Prince E, Kremer A, Plomion C, Le Provost G. 2013. Transcriptional profiling of bud dormancy induction and release in oak by next-generation sequencing. *BMC Genomics* **14**, 236.

Uggla C, Moritz T, Sandberg G, Sundberg B. 1996. Auxin as a positional signal in pattern formation in plants. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 9282–9286.

Ursache R, Nieminen K, Helariutta Y. 2013. Genetic and hormonal regulation of cambial development. *Physiologia Plantarum* **147**, 36–45.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, RESEARCH0034.1–RESEARCH0034.11.

Vanstraelen M, Benková E. 2012. Hormonal interactions in the regulation of plant development. *Annual Review of Cell and Developmental Biology* **28**, 463–487.

Varela M, Valdivieso T. 1996. Phenological phases of *Quercus suber* L. flowering. *Forest Genetics* **3**, 93–102.

Vatén A, Dettmer J, Wu S, Stierhof Y-D, Miyashima S, Yadav SR, Roberts CJ, Campilho A, Bulone V, Lichtenberger R, et al. 2011. Callose biosynthesis regulates symplastic trafficking during root development. *Developmental Cell* **21**, 1144–1155.

Vescovi M, Riefler M, Gessuti M, Novák O, Schmülling T, Lo Schiavo F. 2012. Programmed cell death induced by high levels of cytokinin in *Arabidopsis* cultured cells is mediated by the cytokinin receptor CRE1/AHK4. *Journal of Experimental Botany* **63**, 2825–2832.

Vogt E, Schönherr J, Schmidt HW. 1983. Water permeability of periderm membranes isolated enzymatically from potato tubers (*Solanum tuberosum* L.). *Planta* **158**, 294–301.

Waisel Y, Liphschitz N, Arzee T. 1967. Phellogen activity in *Robinia pseudacacia* L. *New Phytologist* **66**, 331–335.

Waisel Y, Liphschitz N. 1975. Sites of Phellogen Initiation. *Botanical Gazette* **136**, 146–150.

Wang J, Andersson-Gunnerås S, Gaboreanu I, Hertzberg M, Tucker MR, Zheng B, Leśniewska J, Mellerowicz EJ, Laux T, Sandberg G, et al. 2011. Reduced expression of the *SHORT-ROOT* gene increases the rates of growth and development in hybrid poplar and *Arabidopsis*. *PLoS ONE* **6**, e28878.

Welch D, Hassan H, Blilou I, Immink R, Heidstra R, Scheres B. 2007. *Arabidopsis* JACKDAW and MAGPIE zinc finger proteins delimit asymmetric cell division and stabilize tissue boundaries by restricting SHORT-ROOT action. *Genes & Development* **21**, 2196–2204.

Williams L, Fletcher JC. 2005. Stem cell regulation in the *Arabidopsis* shoot apical meristem. *Current Opinion in Plant Biology* **8**, 582–586.

Xu M, Zhang B, Su X, Zhang S, Huang M. 2011. Reference gene selection for quantitative real-time polymerase chain reaction in *Populus*. *Analytical Biochemistry* **408**, 337–339.

Xu P, Kong Y, Li X, Li L. 2013. Identification of molecular processes needed for vascular formation through transcriptome analysis of different vascular systems. *BMC Genomics* **14**, 217.

Yadun SL, Liphshitz N. 1989. Sites of first phellogen initiation in conifers. *IAWA Bulletin* **10**, 43–52.

Yamamoto R, Demura T, Fukuda H. 1997. Brassinosteroids induce entry into the final stage of tracheary element differentiation in cultured *Zinnia* cells. *Plant & Cell Physiology* **38**, 980–983.

Yamamoto R, Fujioka S, Demura T, Takatsuto S, Yoshida S, Fukuda H. 2001. Brassinosteroid levels increase drastically prior to morphogenesis of tracheary elements. *Plant Physiology* **125**, 556–563.

Yáñez-Espinosa L, Terrazas T, López-Mata L. 2010. Phenology and radial stem growth periodicity in evergreen subtropical rainforest trees. *IAWA Journal* **31**, 293–307.

Yang Y, Hou S, Cui G, Chen S, Wei J, Huang L. 2010. Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza*. *Molecular Biology Reports* **37**, 507–513.

Ye ZH, Varner JE. 1994. Expression of an auxin- and cytokinin-regulated gene in cambial region in *Zinnia*. *Proceedings of the National Academy of Sciences of the United States of America* **91**, 6539–6543.

Ye Z-H. 2002. Vascular tissue differentiation and pattern formation in plants. *Annual Review of Plant Biology* **53**, 183–202.

Yordanov YS, Regan S, Busov V. 2010. Members of the LATERAL ORGAN BOUNDARIES DOMAIN transcription factor family are involved in the regulation of secondary growth in *Populus*. *The Plant Cell* **22**, 3662–3677.

Yu N-I, Lee SA, Lee M-H, Heo J-O, Chang KS, Lim J. 2010. Characterization of SHORT-ROOT function in the *Arabidopsis* root vascular system. *Molecules and Cells* **30**, 113–119.

Zhao Z, Andersen SU, Ljung K, Dolezal K, Miotk A, Schultheiss SJ, Lohmann JU. 2010. Hormonal control of the shoot stem-cell niche. *Nature* **465**, 1089–1092.

Zhong R, Ye ZH. 1999. *IFL1*, a gene regulating interfascicular fiber differentiation in *Arabidopsis*, encodes a homeodomain-leucine zipper protein. *The Plant Cell* **11**, 2139–2152.

Zhong R, Ye ZH. 2001. Alteration of auxin polar transport in the *Arabidopsis ifl1* mutants. *Plant Physiology* **126**, 549–563.

Zottini M, Barizza E, Bastianelli F, Carimi F, Lo Schiavo F. 2006. Growth and senescence of *Medicago truncatula* cultured cells are associated with characteristic mitochondrial morphology. *The New Phytologist* **172**, 239–247.

Chapter II

A *SHORT-ROOT*-like gene (*PtSHR2B*) is involved in *Populus* phellogen activity



Miguel A, Milhinhos A, Jones B, Miguel CM. A *SHORT-ROOT*-like gene (*PtSHR2B*) is involved in *Populus* phellogen activity. (*under revision*)

In this research paper Andreia Miguel participated in the experimental design, laboratory experiments, analysis of the results and paper writing.

A *SHORT-ROOT*-like gene (*PtSHR2B*) is involved in *Populus* phellogen activity

Abstract

SHORT-ROOT (SHR) is a GRAS family transcription factor that was first characterized for its role in the specification of the stem cell niche and radial patterning of *Arabidopsis thaliana* (*Arabidopsis*) roots. Three *SHR*-like genes have been identified in *Populus trichocarpa* (*Populus*). *PtSHR1* and the *Arabidopsis SHR*, *AtSHR*, share high similarity over the entire length of the coding sequence. *PtSHR1* has been implicated in the regulation of secondary, radial growth in *Populus* stems. The two other *Populus SHR*-like genes, *PtSHR2A* and *PtSHR2B*, are 5' truncated compared to *AtSHR*. This study reports the functional characterization of *PtSHR2B*. Unlike *PtSHR1*, that is expressed throughout the cambial zone, *PtSHR2B**prom:uidA* expression was detected in the outer cell layers of the stem and phellogen. *PtSHR1* and *PtSHR2B* promoters both drive *uidA* expression in the *Populus* shoot apex and in the roots, but the GUS staining differed markedly between the two. Ectopic expression of *PtSHR2B* in hybrid aspen reduced tree growth and decreased the relative proportion of phellem compared to the wild-type. Elements of the cytokinin metabolism and response mechanism were altered in the transgenic plants, particularly in the bark. The results indicate that *PtSHR2B* is a negative regulator of plant growth and meristem activity possibly acting through modulation of cytokinin activity. The localization of *PtSHR2B* in the phellogen and the decrease in phellem thickness reveal for the first time a regulatory role in this lateral meristem and indicate a functional diversification of *SHR* after the divergence of the *Populus* and *Arabidopsis* lineages.

Keywords

Cytokinin, lateral meristem, phellogen, *Populus tremula* × *Populus tremuloides*, secondary growth, *SHORT-ROOT*

Introduction

Plant growth occurs from specialized regions called meristems. Mitotic divisions in the meristems produce the cells that eventually differentiate into the organs and tissues that comprise the body of the plant. Four main meristems exist in woody perennials, the root and shoot apical meristems, that provide cells for shoot and root growth, respectively, and the vascular cambium and cork cambium (phellogen) that generate cells for the secondary, or radial, growth of the stem, branches and roots. Several lines of evidence have indicated that although the structure of the meristems differs, there are commonalities in the molecular mechanisms underlying their function (Schrader *et al.*, 2004; Baucher *et al.*, 2007; Du and Groover, 2010). Examples of genes known to be involved in the regulation of different meristems include the Class III HD-ZIP (Otsuga *et al.*, 2001; McConnell *et al.*, 2001; Emery *et al.*, 2003; Hawker and Bowman, 2004; Schrader *et al.*, 2004; Prigge and Otsuga, 2005; Du *et al.*, 2011; Robischon *et al.*, 2011; Zhu *et al.*, 2013), the *Populus* Class I KNOX homeobox genes *ARBORKNOX1* and 2 (*ARK1* and *ARK2*) which are orthologs of *Arabidopsis* *SHOOTMERISTEMLESS* (*STM*) and *BREVIPEDICELLUS* (*BP*), respectively (Long *et al.*, 1996; Mele *et al.*, 2003; Groover *et al.*, 2006; Du *et al.*, 2009), *SHORT-ROOT* (*SHR*) gene (Benfey *et al.*, 1993; Fukaki *et al.*, 1998; Helariutta *et al.*, 2000; Schrader *et al.*, 2004; Dhondt *et al.*, 2010; Wang *et al.*, 2011), and the *Arabidopsis* *WUSCHEL* (*WUS*) (Mayer *et al.*, 1998; Schoof *et al.*, 2000; Brand *et al.*, 2000) and *WUSCHEL-RELATED HOMEODOMAIN 5* (*WOX5*) (Haecker *et al.*, 2004; Stahl and Simon, 2009a,b; Schrader *et al.*, 2004; Baucher *et al.*, 2007; Sarkar *et al.*, 2007; Tucker and Laux, 2007). Although considerable attention has been paid to the functioning of the apical meristems and the vascular cambium, little attention has been paid to the phellogen, despite its critical importance in establishing a protective layer on stems, branches and roots. Given the existing evidence, it can be

hypothesized that the mechanisms underlying phellogen function will overlap those of the other meristems (Soler, 2008).

The SHR transcription factor belongs to the GRAS family of plant specific proteins, that are characterized by a variable N-terminal domain, but a highly conserved C-terminal domain (Helariutta *et al.*, 2000; Bolle, 2004). The *Arabidopsis* SHR (AtSHR) has been well characterized, and has been shown to be a key regulator, along with the related GRAS protein, SCARECROW (SCR), of radial patterning and stem cell niche specification in the roots (Benfey *et al.*, 1993; Laurenzio *et al.*, 1996; Helariutta *et al.*, 2000; Nakajima *et al.*, 2001). AtSHR is essential for the asymmetric cell divisions of the cortex/endodermal initial (CEI), that give rise to the cortex and endodermal cell lineages (Benfey *et al.*, 1993; Helariutta *et al.*, 2000; Nakajima *et al.*, 2001), and for the periclinal divisions of cortex cells in a maturing root (Paquette and Benfey, 2005). AtSHR acts non-cell-autonomously in the establishment of root radial patterning (Helariutta *et al.*, 2000) and its loss of function leads to the differentiation of stem cells and the loss of the endodermis (Benfey *et al.*, 1993; Helariutta *et al.*, 2000). AtSHR function is not limited to the root. Fukaki *et al.* (1998) demonstrated that *AtSHR* is also involved in radial patterning in the shoot. More recently, Dhondt *et al.* (2010) demonstrated that, similarly to the *Arabidopsis* root, SHR functions in association with its downstream target, SCR, in the regulation of cell proliferation and vascular differentiation in leaves. The SHR/SCR mechanism therefore appears to have been co-opted to regulate cell proliferation and differentiation in multiple organs.

SHR modulates the expression of genes involved in a wide range of processes during *Arabidopsis* root development, including transcriptional regulation, signaling and response to hormones, and in the regulation of cell-cycle genes (Levesque *et al.*, 2006; Sozzani *et al.*, 2010). In *Arabidopsis*, correct patterning of the central vascular cylinder is mediated through movement of the SHR protein from the stele into the endodermis (Nakajima *et al.*, 2001), where it activates its target, SCR, that together activate *miR165a*

and *miR166b* (Carlsbecker *et al.*, 2010). The regulation of vascular patterning by SHR in the *Arabidopsis* root involves the modulation of cytokinin (CK) homeostasis through the direct regulation of the cytokinin degrading enzyme, *CYTOKININ OXIDASE 3* (*CKX3*) (Cui *et al.*, 2011; Hao and Cui, 2012).

SHR has also been studied in tree species such as *Pinus radiata*, where it was suggested to have roles in root meristem formation and maintenance, and in the cambial region of hypocotyls (Solé *et al.*, 2008). The putative *Populus* ortholog of *AtSHR*, *PtSHR1*, is expressed in the cambial zone (Schrader *et al.*, 2004; Wang *et al.*, 2011) and functions as a regulator of cell division and meristem activity in the shoots (Wang *et al.*, 2011). Partial suppression of the *PtSHR1* transcript in transgenic lines leads to taller trees with a larger vascular cambium due to an increase in cell proliferation in the cambial zone (Wang *et al.*, 2011). In both *Arabidopsis* and *Populus* it has been suggested that *SHR* regulates growth through the control of cell divisions, in a concentration-dependent manner (Paquette and Benfey, 2005; Wang *et al.*, 2011; Koizumi *et al.*, 2012).

Whereas there is only one *SHR* gene in *Arabidopsis*, three *SHR*-like genes have been identified in the *Populus* genome, *PtSHR1*, *PtSHR2A* and *PtSHR2B* (Wang *et al.*, 2011). Based on sequence similarity and on functional studies with the *PtSHR1* coding sequence, driven by the *AtSHR* promoter, the *PtSHR1* is considered the putative ortholog of the *Arabidopsis* *SHR* (Wang *et al.*, 2011). *PtSHR2A* and *PtSHR2B* are 5' truncated compared to *AtSHR* and *PtSHR1*. In this study, *PtSHR2B* was investigated in order to better understand its role in hybrid aspen. The work indicated that *PtSHR2B* is involved in the control of several aspects of plant growth, and acts, at least partially, through the regulation of CK homeostasis. Its localization in the phellogen points to a regulatory role in this important meristem during secondary growth.

Materials and Methods

Plant material

Hybrid aspen (*Populus tremula* L. × *Populus tremuloides* Michx.; Clone T89) was propagated *in vitro* on half-strength basal MS salt medium (Murashige and Skoog, 1962), maintained in a growth chamber at 21° C in a 16 h light/8 h dark photoperiod. For greenhouse experiments, *in vitro*-established wild-type and transgenic plants were transferred to a soil:peat:perlite (1:3:1) potting mix and acclimatized in a growth chamber, gradually decreasing the humidity from 95% to 70% over five weeks before transferring the plants to the greenhouse, where they were grown for a minimum of 10 weeks prior to analysis. The position of all pots within the greenhouse was changed weekly to minimize positional bias experimental error.

After 10 weeks in the greenhouse, the five youngest fully-expanded leaves were collected (Supplementary Fig. S1A). Bark was isolated by peeling off the stem tissues external to the vascular cambium. The remaining stem tissues (vascular cambium, developing xylem, mature xylem and pith), hereafter termed 'wood', were collected together. All samples were immediately frozen in liquid nitrogen and stored at -80° C until further processing. Intact stem sections were also collected and fixed in FAA (5% formaldehyde, 5% acetic acid, 50% ethanol) for anatomical analysis.

Gene constructs and genetic transformation

To detect promoter activity of *PtSHR1*, *PtSHR2A* and *PtSHR2B*, transgenic lines were generated using constructs incorporating ~2.5 Kbp sequence upstream of the start codon of each of the genes, fused to the coding sequence of *uidA* (encoding the β -glucuronidase enzyme (GUS)), (*PtSHR_nprom:uidA*), using the pKGWFS7,0 binary vector (Karimi *et al.*, 2002). The sequences (eugene3.01860017, eugene3.00070144 and

eugene3.00640143, for *PtSHR1*, *PtSHR2A* and *PtSHR2B*, respectively) were retrieved from the *Populus trichocarpa* genome version v1.1 (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html) and previously described by Wang *et al.* (2011). In *Phytozome* v9.1 (<http://www.phytozome.net/>), the corresponding *loci* are Potri.007G063300, Potri.007G132000 and Potri.017G019900 for *PtSHR1*, *PtSHR2A* and *PtSHR2B*, respectively. To ectopically express *PtSHR2B*, its coding sequence was cloned downstream of the 35S cauliflower mosaic virus promoter in the pK7WG2,0 vector (35S:*PtSHR2B*) (Karimi *et al.*, 2002). Hybrid aspen stem sections were transformed according to Nilsson *et al.* (1992). Ten transgenic lines were generated for each construct. Representative *PtSHR1prom:uidA* and *PtSHR2Bprom:uidA* lines and three ectopic expression lines, 2B_7, 2B_8 and 2B_12, were chosen for further analysis. In this study, the nomenclature and the gene sequence information followed Wang *et al.* (2011).

Histochemical GUS assay in transgenic lines

GUS assays were performed in transformed hybrid aspen leaves, stems, roots and shoot apex collected from six-week-old *in vitro*-grown shoots, six-month-old and one year-old stems of greenhouse-grown trees. Tissues were placed in ice-cold 90% acetone for 30 min and then washed in water prior to immersion in the GUS staining solution (10 mM sodium phosphate buffer pH 7.0, 0.5% Triton X-100, 2 mM potassium ferricyanide and 2 mM X-Gluc (5-bromo-4-chloro-3-indolyl β -D-glucuronide)), vacuum infiltrated and incubated overnight in the dark at 37° C. After washing in water, tested leaves, roots, shoot apex and stems of *in vitro* plants were gradually dehydrated to 70% ethanol and stem sections of greenhouse grown plants were fixed in FAA and then included in Technovit 7100 resin (Heraeus Kulzer), according to the manufacturer's instructions, with minor modifications: after vacuum infiltration, samples were left for two days at 4° C in the pre-infiltration solution. The

solution was then replaced and samples left for another seven days at 4° C. The material was subsequently placed in the infiltration solution and left for one to three weeks at 4° C, followed by polymerization at room temperature.

Stereomicroscope observations were performed with a Nikon SMZ800 and images were captured using an Olympus SC30 camera and software. Microscope observations were made with a Nikon Inverted Microscope Eclipse TE300 and images taken with a Nikon DS-Fi1 camera using the NIS-Elements F3.0 software.

Anatomical analysis and growth measurements

Stem pieces of one- and two-year old shoots of wild-type hybrid aspen were collected and fixed in ice-cold FAA, as previously described, vacuum infiltrated and left overnight in a desiccator at 4° C. After gradual dehydration to 100% ethanol, tissues were embedded in resin as described above, and 6 to 8 µm thick sections stained with Toluidine Blue O. Several growth parameters were analysed in 10 week-old greenhouse-grown *35S:PtSHR2B* trees (Supplementary Fig. S1A). Tree height and total length between the 10th (EN10) and the 17th (EN17) internodes from the shoot tip, were recorded. Stem diameters were measured at the reference internode, (EN14, the internode chosen for equal comparison between transgenic and wild-type plants), and at the stem base internode (ENbase). Measurements of distances between different stem tissues were taken at a minimum of four positions around the circumference of the stem using ImageJ software (Abràmoff *et al.*, 2004; Schneider *et al.*, 2012) (Supplementary Fig. S1B). The total lamina area of the five leaves surrounding EN14 was determined using a leaf area meter (LI-3000A, LI-COR Inc.).

Reverse Transcription Quantitative PCR (RT-qPCR)

The first five fully expanded leaves down from the shoot tip, and bark and wood from the ENbase were collected and ground to a fine powder (Supplementary Fig. S1A) using either a mortar and pestle (leaves and bark) or a grinder mill (M 20 Universal mill, Ika). Isolation of bark and wood tissues was done as described in “Plant material” section. Total RNA from leaves and bark was isolated as described by Reid *et al.* (2006) with minor modifications (Marum *et al.*, 2012). Total RNA from wood samples was extracted using the protocol as described in Chang *et al.*, (1993) and all RNA samples were treated with TURBO DNase (Ambion) according to manufacturer’s instructions. cDNA synthesis was performed from 1.5 µg of DNase-treated RNA, using Transcriptor High Fidelity cDNA Synthesis Kit (Roche) with anchored-oligo(dT)₁₈ primers. Quantitative real-time PCR (qPCR) was carried out in 96-well plates in a LightCycler 480 (Roche) using SYBR Green I Master Mix (Roche). Primers for amplifying a transcript fragment of *PtSHR2B*, 5’F-CAGCAATACCCTTTGCACACAG-3’ and 5’R-ACCCAGTCCTTCCTTTGTG-3’, were designed using the *Populus trichocarpa* genome version 1.1 (http://genome.jgi-psf.org/Poptr1_1/Poptr1_1.home.html) and gene sequence (eugene3.00640143) described by Wang *et al.* (2011). For amplification of *PtCKX3* (Potri.006G152500.1) transcripts, the specific primers 5’F-TCAGATCCAAACCCTTGATTTC-3’ and 5’R-CAGTAAAAGGGGTGTAGTT-3’ were designed using *Populus trichocarpa* genome version 3, from Phytozome (<http://www.phytozome.net/>). For *PtRR7* (Potri.016G038000.1) the primers were as previously described (Nieminen *et al.*, 2008). *PtCYP2* (Potri.004G168800.1) was used as a reference gene as previously described (Brunner *et al.*, 2004; Milhinhos *et al.*, 2013). The PCR program used was 95° C for 10 min, 45 cycles of 10 s at 95° C, 20 s at 60° C for *PtRR7* and *PtCKX3* or 20 s at 63° C for *PtSHR2B*, and 10 s at 72° C. The annealing temperature for the reference gene primers was 60 or 63° C, depending on the experiment.

Three technical replicates were used for each of the three biological samples in each experiment. To normalize values obtained from different plates, a calibrator sample consisting of cDNA synthesized from RNA from leaves of a transgenic line was used in each plate. Normalized relative quantities were obtained through the $\Delta\Delta C_T$ method (Pfaffl, 2001; Livak and Schmittgen, 2001; Hellemans *et al.*, 2007) and the amplification efficiency determined using Real-Time PCR Miner (Zhao and Fernald, 2005).

Statistical analysis

The assessment of significance in transcript profiles and phenotypic parameters, was carried out using non-parametric analysis, Mann-Whitney *U*-test. A significance level of $p=0.05$ was used. Statistics were performed using the Statistica (StatSoft Inc., <http://www.statsoft.com>) software package.

Results

In this work, the *Populus* *SHR*-like gene, *PtSHR2B*, was characterized and compared to the putative *Populus* ortholog of the *Arabidopsis* *SHR* gene, *PtSHR1* (Wang *et al.*, 2011). *AtSHR* and *PtSHR1* genes have previously been implicated in the regulation of primary, apical, meristems and vascular cambium (VC) activity. It is unknown whether *SHR* is involved in the regulation of phellogen (Pg) activity. Transverse sections of hybrid aspen stems at different developmental stages showed a distinct phellogen meristematic layer already present at the end of the first year of growth (Supplementary Fig. S1C). Periclinal divisions and the characteristic layer of suberized phellem cells in the periderm could be observed. The periderm in two year-old stems had only a slight increase in phellem layer thickness compared to one year-old stems (Supplementary Fig. S1C, D).

PtSHR1 and PtSHR2B promoters are active in different tissues

While tissues from wild-type controls were always negative to GUS histochemical assay (Fig. 1A-E), analysis of hybrid aspen plants carrying either the *PtSHR1* or *PtSHR2B* promoter driving *uidA* expression (*PtSHR_nprom:uidA*) indicated different patterns of promoter activity. In *in vitro*-grown hybrid aspen plantlets, *PtSHR1prom*-driven GUS staining was found in the cambial zone (Fig. 1F), in accordance with previous reports (Wang *et al.*, 2011). Promoter activity was also analysed in greenhouse-grown trees that had undergone substantial secondary growth. In these plants, *GUS* expression was observed throughout the cambial zone and in xylem rays (Fig. 1G, H). In stem sections of young tissue (3rd internode from tip) similarly to the *in vitro* plants, *GUS* expression was also observed in the primary xylem (Supplementary Fig. S2A). *SHR* has previously been shown to play a role in the development of protoxylem in *Arabidopsis* (Carlsbecker *et al.*, 2010; Yu *et al.*, 2010). In older tissues (9th internode), *GUS* staining seemed restricted to the cambial zone (Supplementary Fig. S2B). Additionally, *GUS* expression was found in the leaf vasculature and in the root stele of *in vitro* plants (Fig. 1I, J), similarly to its homolog in *Arabidopsis* (*AtSHR*) (Helariutta *et al.*, 2000; Dhondt *et al.*, 2010; Wang *et al.*, 2011). In the shoot apex, *GUS* staining was observed in the apical meristem and vasculature (Fig. 1K). *GUS* staining in *in vitro*-grown *PtSHR2Bprom:uidA* plants was detected in the primary xylem and in the outer cell layers of the stem (Fig. 1L-N). Unlike *PtSHR1* promoter, in *PtSHR2Bprom:uidA* trees, *GUS* staining appeared strongly localized in the phellogen cell layer (Fig. 1O, P), suggesting a specific function for the modified version of *SHR* in this meristem. In *in vitro* leaves *GUS* expression under the *PtSHR2B* promoter was not detected (Fig. 1Q) but in *in vitro* roots *GUS* staining showed a stark contrast to that observed with the *PtSHR1* promoter with *GUS* expression being observed at the root tip in *PtSHR2Bprom:uidA* plants (Fig. 1R). Differences between the expression

driven by each of the two promoters were also found in *in vitro* developing shoot tip where *PtSHR2Bprom:uidA* staining was restricted to apical and axillary meristems (Fig. 1S). Since no GUS signal was ever detected in any of the analysed tissues from plants carrying the *PtSHR2Apro:uidA* construct, the study only proceeded with the analysis of *PtSHR2B*.

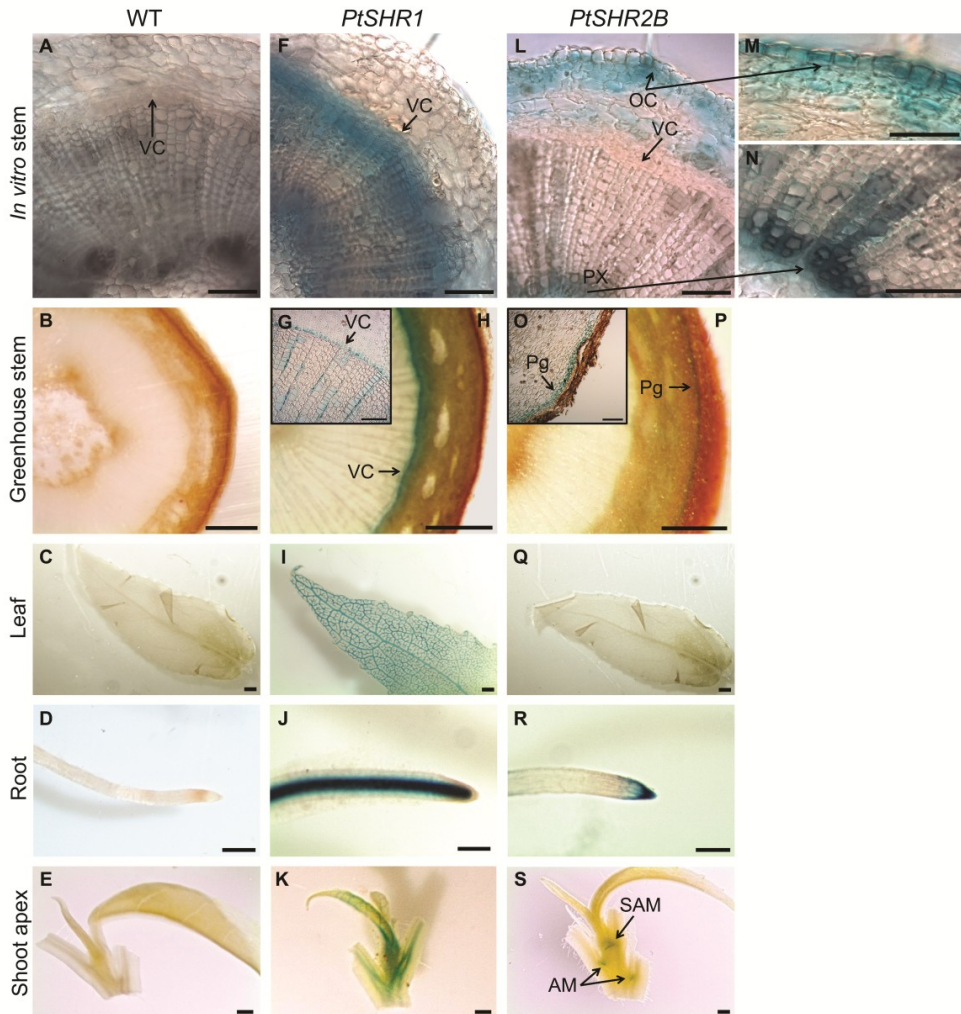


Fig. 1. Localization of *PtSHR1* and *PtSHR2B* promoter-driven *uidA* expression. (A, C-E, F, I-K, L-N, Q-S) GUS histochemical assay in six-week-old hybrid aspen *in vitro*-grown plants. (B, G, H, O, P) GUS histochemical assay in stems from six month-old greenhouse-grown hybrid aspen. (A, F, G, L-N, O) Images were taken under the ►

◀microscope (Bars = 100 μm). (B-E, H-K, P-S) Images were taken under the stereomicroscope (Bars = 475 μm). AX, axillary meristem; OC, outer cell layers of the stem; Pg, phellogen; PX, primary xylem; SAM, shoot apical meristem; VC, vascular cambium.

Ectopic expression of PtSHR2B reduces overall tree growth

To explore the role of *PtSHR2B* in hybrid aspen, the *PtSHR2B* coding region was isolated and transformed into hybrid aspen under the control of the constitutive, CaMV 35S, promoter. No obvious phenotype could be observed in *in vitro*-grown plants (data not shown). In 10-week old greenhouse-grown trees, however, the transgenic trees showed a reduced growth compared to the wild-type (Fig. 2). Tree height in all of the transgenic lines was significantly reduced when compared to the wild-type plants (Fig. 2A, B). Although no change could be observed in the reference internode (EN14), stem diameter was reduced compared to the wild-type at the base of the transgenic trees (Fig. 2C). The reduction in height in the transgenic trees was primarily the result of a reduced internode length, as the mean stem length between internodes 10 and 17 was significantly shorter in the transgenic trees compared to the wild-type (Fig. 2D). The total number of internodes was slightly reduced, but was only significant in one of the transgenic lines (Fig. 2E). The total average fresh weight and lamina area of the five leaves surrounding the reference internode was also reduced in the transgenic trees compared to the wild-type (Fig. 2F, G).

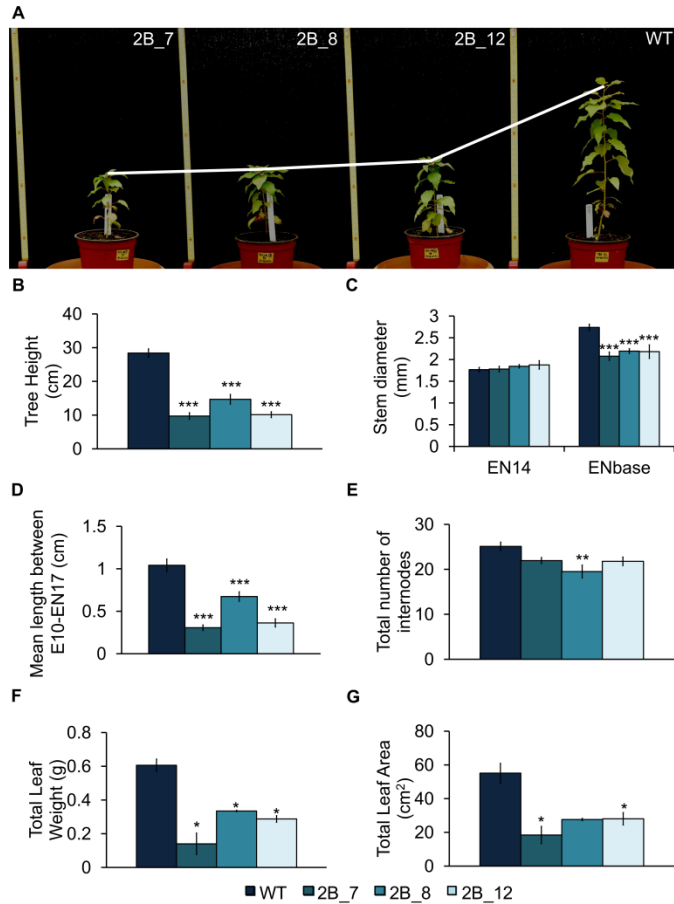


Fig. 2. Phenotypic characterization of transgenic *35S:PtSHR2B* hybrid aspen lines, grown in the greenhouse for 10 weeks. (A) Wild-type (WT) and plants from independent transgenic lines (2B_7; 2B_8 and 2B_12). (B) Tree height. Values are means $\pm$ SE of at least 8 biological replicates. (C) Stem diameter at the reference internode (EN14) and at the stem base (ENbase). Values are means $\pm$ SE of at least 6 biological replicates. (D) Mean length between internodes 10 and 17 (EN10 and EN17). Values are means $\pm$ SE of at least 8 biological replicates. (E) Total number of internodes. Values are means $\pm$ SE of at least 8 biological replicates. (F) Total weight of the 5 leaves surrounding EN14. Values are means $\pm$ SE of at least 3 biological replicates. (G) Total area of the leaves surrounding EN14. Values are means $\pm$ SE of at least 3 biological replicates. Asterisks indicate the significance level between each individual line and the wild-type (* p <0.05, ** p <0.01 and *** p <0.001, Mann-Whitney U test).

PtSHR2B transcript profiling

Profiling of *PtSHR2B* transcript levels in greenhouse-grown tissues revealed highest levels in the bark, with considerably lower levels in wood and leaf tissues (Fig. 3A). The lower levels of transcripts found in leaves corroborates the GUS staining results, while the significant levels of transcripts in the wood tissues indicates that expression is predominantly in the Pg, but not restricted to it (Fig. 1O, P, 3A). In the ectopic expression lines, *35S:PtSHR2B*, RT-qPCR confirmed the ectopic expression of *PtSHR2B* in leaves and stem tissues of *in vitro* plants, (Fig. 3B), however, in the greenhouse-grown trees, the increased accumulation of *PtSHR2B* transcript was observed in leaves and bark, but not in wood tissues (Fig. 3C).

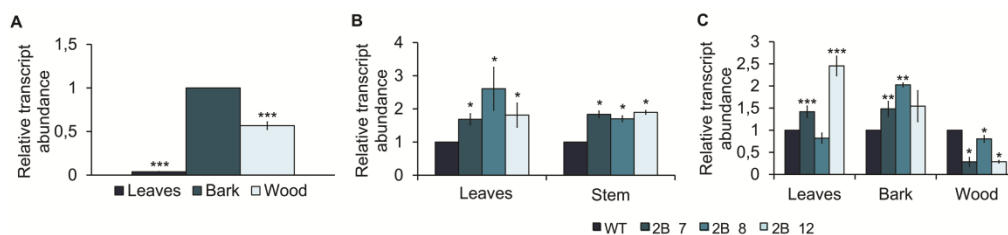


Fig. 3. Relative expression levels of *PtSHR2B* in tissues of hybrid aspen. (A) Wild-type (WT) plants grown in a greenhouse for 10 weeks. Results are expressed relative to the expression in the bark and values are means $\pm$ SE of at least 4 biological, except for wood where a mix of different wild-type plants were used, and 3 technical replicates. (B) *PtSHR2B* transcript levels in *in vitro*-grown *35S:PtSHR2B* plants with about four-week-old. Values are means $\pm$ SE of 3 biological and 3 technical replicates. (C) *PtSHR2B* transcript levels in *35S:PtSHR2B* trees grown in a greenhouse for 10 weeks. Values are means $\pm$ SE of at least 2 biological and 2 technical replicates. Asterisks indicate the significance of the difference between each tissue and the bark (A) or between each individual line and the wild-type, (B, C) (* p <0.05, ** p <0.01 and *** p <0.001, Mann-Whitney U test).

Altered stem anatomy in 35S:PtSHR2B ectopic expression plants

Transverse sections taken from stems of the 35S:PtSHR2B transgenic trees were analysed by light microscopy. The amount of wood produced in the transgenic trees (radial distance from the pith side of the lignified xylem to the cambial zone (Supplementary Fig. S1B)) was significantly reduced compared to the wild-type in two of the three transgenic lines (Fig. 4A). The distance from the cambial zone to the phellem (bark layer) was larger in the transgenic stems than in the wild-type in all of the transgenic lines (Fig. 4A). This bias towards non-wood tissues was more evident at the base of the stem, where secondary growth was more extensive (Fig. 4A, B). The ratio between phellem and the stem diameter was reduced in two out of the three transgenic lines (Fig. 4C, D). Considerable variation existed between the independent transformant lines.

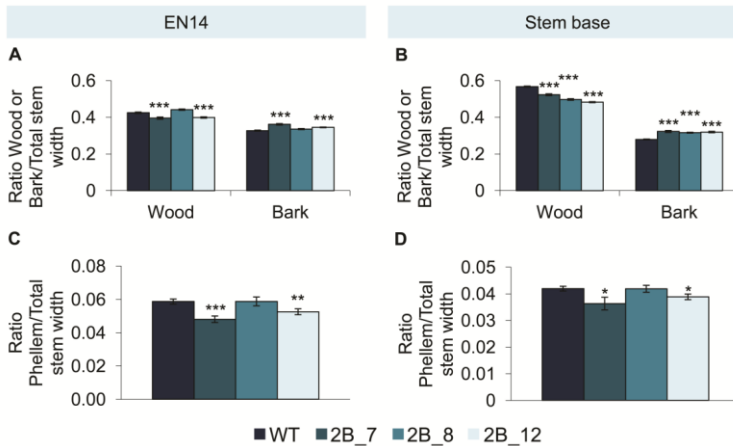


Fig. 4. Anatomical characterization of the 35S:PtSHR2B hybrid aspen stems at the 14th internode (EN14) and at the stem base internode (ENbase). (A and B) Ratio between the width of the wood or

the bark layer and the stem radius at the EN14 (A) and ENbase (B). (C and D) Ratio between the width of the phellem layer and the stem radius at the EN14 (C) and ENbase (D). Values are means $\pm$ SE of at least 3 biological replicates and asterisks indicate the significance of the difference between each individual line and the wild-type (* p <0.05, ** p <0.01 and *** p <0.001, Mann-Whitney U test). For each tree cross-section, the measurements were taken at a minimum of four equidistant positions around the circumference of the stem.

Genes involved in cytokinin metabolism are altered in transgenic plants

Cytokinin (CK) has been linked to SHR function in *Arabidopsis* roots where SHR controls vascular patterning through its effect on CK homeostasis (Cui *et al.*, 2011). We analysed the transcript levels of a central CK primary response gene, the A-type response regulator, *PtRR7* (Nieminen *et al.*, 2008; Ramírez-Carvajal *et al.*, 2008), in bark and wood tissues of the *35S:PtSHR2B* transgenic trees. *PtRR7* transcript levels were higher in bark tissues in all *PtSHR2B* over-expression lines compared to the wild-type (Fig. 5A), indicating altered CK signaling in this tissue.

In *Arabidopsis* roots, SHR is known to act, at least partially, by directly regulating the expression of *CYTOKININ OXIDASE 3*, *AtCKX3*, (Cui *et al.*, 2011). We also analysed transcript levels for the putative *Populus AtCKX3* ortholog, *PtCKX3*, in bark and wood tissues of the transgenic plants. Transcripts for *PtCKX3* were more abundant in the bark tissues of the transgenic trees, compared to the wild-type (Fig. 5B). Increased transcript levels for these genes in the bark indicate that *PtSHR2B* levels are important for the regulation of CK metabolism in this tissue. Transcript levels for both genes were also up-regulated in wood tissue (Fig. 5A, B), but the results were less clear because of variation between the independent transgenic lines.

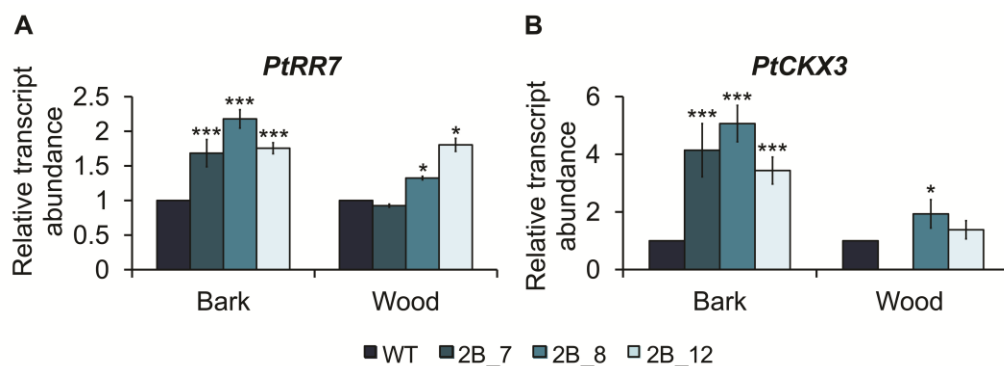


Fig. 5. Expression of genes involved in cytokinin function, determined at the stem base internode (ENbase), in the *35S:PtSHR2B* hybrid aspen plants grown in a greenhouse for 10 weeks. (A, B) Relative transcript levels of (A) the cytokinin ►

◀primary response gene *PtRR7* and (B) the cytokinin oxidase gene, *PtCKX3*, in bark and wood tissues of 35S:*PtSHR2B* trees. Values are means $\pm$ SE of at least 3 biological replicates and 2 technical replicates and asterisks indicate the significance of the difference between each individual line and the wild-type (* $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$, Mann-Whitney U test).

Discussion

The vascular cambium and the development of wood tissues have been the subject of numerous reports (Baucher *et al.*, 2007; Du and Groover, 2010; Schuetz *et al.*, 2012; Sanchez *et al.*, 2012). By contrast, despite the important role that the phellogen plays in providing cells for the development of the protective layer on stems, branches and roots it has received minimal attention. Under our growth conditions, a surrounding phellogen meristematic and phellem (cork) layer had formed in one year old hybrid aspen stems. The thickness of this layer was almost unchanged between one and two year-old hybrid aspen stems. Consistency in the phellem layer thickness during the life-span of the close relative *Populus tremuloides* Michx. has long been known (Kaufert, 1937), indicating that in *Populus* species phellogen cell divisions are matched by a shedding of cork layer cells.

Populus has long been used as a model species in the study of angiosperm tree function. It has a high annual rate of secondary growth, and like other model species has the advantages of being relatively easily genetically modified and cultured. There are also now ample molecular, genomic and bioinformatics resources available for various *Populus* species (Stettler and Bradshaw Jr, 1996; Bradshaw Jr *et al.*, 2000). Importantly, the *Populus trichocarpa* genome has been fully sequenced (Tuskan *et al.*, 2006). The sequence data indicate that two whole-genome duplication events occurred in ancestors of the species (Tuskan *et al.*, 2006). Although

duplicated genes are often lost over evolutionary time, higher gene retention is often found, particularly for specific classes of genes, such as: (i) genes with regulatory functions; transcription factors and developmental regulators (Blanc and Wolfe, 2004; Seoighe and Gehring, 2004; Carretero-Paulet and Fares, 2012); and (ii) for genes derived from a previous round of genome duplication (Seoighe and Gehring, 2004). The biased retention is most likely because multiple copies of the retained genes impart specific beneficial effects for the organism (Seoighe and Gehring, 2004; Carretero-Paulet and Fares, 2012). One of the ways in which multiple genes can be beneficial is through speciation, leading to divergent function and expression patterns (Blanc and Wolfe, 2004; Tuskan *et al.*, 2006; Rodgers-Melnick *et al.*, 2012). In *Populus*, it has been hypothesized that after duplication, gene preservation is influenced by a combination of sub-functionalization and selection favouring retention of genes that encode proteins with a large number of interactions (Rodgers-Melnick *et al.*, 2012). The existence of three *SHR*-like genes in the *Populus* genome indicates that it fits these criteria. In *Arabidopsis*, *SHR* is involved in radial patterning of the roots, hypocotyls and aerial organs (Scheres *et al.*, 1995; Fukaki *et al.*, 1998; Wysocka-Diller *et al.*, 2000; Helariutta *et al.*, 2000; Wang *et al.*, 2011). In the gymnosperm tree species, *Pinus radiata*, *SHR1* is expressed in the roots, in the shoot apex and in the cambial region of hypocotyls (Solé *et al.*, 2008). Similarly, in the angiosperm, *Populus*, *PtSHR1* and *PtSHR2B* promoters drive expression of the marker gene, *uidA*, in the shoot apex of *in vitro*-grown hybrid aspen. The GUS staining profiles differed markedly between the two genes in both the shoot apex and the roots. Particularly in the roots, the difference in the GUS staining profiles indicates that the two genes have markedly different functions. Commensurate with *AtSHR* expression in *Arabidopsis*, GUS staining in the roots of the *PtSHR1prom:uidA* plants was confined to the stele. By contrast, the *PtSHR2B* promoter activity was detected in the root tip. In *in vitro*-grown stems, GUS expression under the control of the *PtSHR1* was observed in the vascular

cambium. On the other hand, GUS staining appears to be located in the primary xylem and in the outer cell layers of *in vitro*-grown *PtSHR2Bprom:uidA* plant stems, rather than in the vascular cambium. In older, greenhouse-grown material with extensive secondary growth, GUS staining was strongly associated with the lateral meristems, the vascular cambium and phellogen in the *PtSHR1prom:uidA* and *PtSHR2Bprom:uidA* plants, respectively. *PtSHR1* has previously been reported to be expressed in the vascular cambium, and to regulate its activity (Schrader *et al.*, 2004; Wang *et al.*, 2011). *AtSHR* and *PtSHR1* have been shown to have broad activity in meristems in the roots and shoots (Helariutta *et al.*, 2000; Schrader *et al.*, 2004; Wang *et al.*, 2011). The association of *PtSHR2Bprom* activity with the phellogen suggests that *PtSHR2B* fulfills a function similar to the other SHR proteins in this lateral meristem.

PtSHR2B function was further explored by ectopically expressing it in hybrid aspen trees behind the constitutive CaMV 35S promoter. Ectopic expression of *PtSHR2B* affected both primary and secondary growth. Compared to the wild-type, overall growth was reduced in the transgenic trees. Wang *et al.* (2011) have shown that partial suppression of *PtSHR1* in hybrid aspen enhanced overall tree growth, suggesting that the protein acts as a dose-dependent negative regulator of meristem activity (Wang *et al.*, 2011). *AtSHR* has also been shown to be a dose-dependent regulator of ground tissue patterning in *Arabidopsis* (Koizumi *et al.*, 2012).

The vascular cambium and phellogen meristems contribute to the radial development of stems, branches and roots. Ectopic expression of *PtSHR2B* resulted in a reduction in wood production relative to the bark. It also led to a slight reduction in the relative proportion of the phellem layer in the transgenic plants compared to the wild-type. The reduction in phellem layer thickness in the ectopic expression lines further suggests an optimum level of *PtSHR2B* is required for proper phellogen function. This is also suggested by a study in *Solanum tuberosum* L. under heat stress where

Ginzberg *et al.*, (2009) found that there is an increase in the tuber periderm due to the accumulation of suberized skin-cell layers (phellem) and, interestingly, the transcriptomic profiling of the tuber periderm in these plants, revealed that *SHR* was down-regulated in the skin (phellem) when compared to the phelloderm (Ginzberg *et al.*, 2009).

Finally, over-expression of *PtSHR2B* affected the expression of the CK-related genes, *PtCKX3* and *PtRR7*, with the direction of change indicating an overall decrease in the CK response in the transgenic plants. It has previously been shown in *Populus*, that over-expression of a CK degrading *CYTOKININ OXIDASE* gene, that would be expected to decrease CK levels and responses, resulted in a reduction in both apical and radial growth in transgenic trees (Nieminen *et al.*, 2008). The authors were able to demonstrate the importance of CK signaling in the regulation of cambial activity in *Populus* stems (Nieminen *et al.*, 2008). CKs have also been shown to be positive regulators of radial growth in *Arabidopsis* (Matsumoto-Kitano *et al.*, 2008). Although it is difficult to interpret the results of ectopic expression experiments, the data obtained in this study, together with the previous studies (Nieminen *et al.*, 2008; Matsumoto-Kitano *et al.*, 2008) strongly suggest that the over-expression of *PtSHR2B* altered the relative proportion of phellem at least partially through its effect on CK metabolism in the phellogen. *PtRR7* transcript levels are positively correlated with the amount of CK present across the stems of hybrid aspen trees (Nieminen *et al.*, 2008). Similarly, CK oxidase gene expression has been shown to be up-regulated in response to increased CK levels (Motyka *et al.*, 1996; Jiao *et al.*, 2003). It is possible that the up-regulation of *PtCKX3* and *PtRR7* observed in the *PtSHR2B* ectopic expression lines is similarly associated with increased CK levels in the phellem and wood tissues. It also presents the possibility that, like *AtSHR* in *Arabidopsis* (Levesque *et al.*, 2006; Cui *et al.*, 2011), *PtSHR2B* is involved in the regulation of CK levels and responses in the phellogen, phellem, and other tissues in *Populus*.

In *Arabidopsis*, AtSHR is a key growth regulator, controlling a large transcriptional regulatory network (Levesque *et al.*, 2006), and through this, it is critical for growth and developmental processes in both the shoot and the root. PtSHR1 has been shown to regulate the function of the vascular cambium, and therefore, of secondary xylem and phloem production in *Populus* stems (Wang *et al.*, 2011). The data presented here suggest that *PtSHR2B* is also involved in lateral meristem functioning. It should be noted that responses varied between individual transgenic lines, making statistically significant differences more difficult to achieve, but this variation is consistent with previous work with transgenic Poplars (Robischon *et al.*, 2011; Milhinhos *et al.*, 2013). Although it cannot be ruled out that PtSHR2B plays a role alongside PtSHR1 in regulating vascular cambium activity, the data presented here suggest that speciation and functional diversification has led to these two AtSHR homologs playing different roles in *Populus* stems. PtSHR2B appears to function principally in the phellogen and therefore in the regulation of phellem and periderm formation. Further research will clarify the connection between PtSHR2B and CK responses.

Acknowledgments

We thank Dr. Paula Scotti Campos (Instituto Nacional de Investigação Agrária e Veterinária, I. P., Oeiras, Portugal) for making available the equipment of leaf area meter. This work was supported by Fundação para a Ciência e Tecnologia, through projects PEst-OE/EQB/LA0004/2011 and PTDC/AGR-GPL/098369/2008 and through 'grant SFRH/BD/44474/2008' (to Andreia Miguel) and 'grant SFRH/BD/30074/2006' (to Ana Milhinhos).

References

- Abràmoff MD, Magalhães PJ, Ram SJ.** 2004. Image Processing with ImageJ. Biophotonics International.
- Baucher M, El Jaziri M, Vandeputte O.** 2007. From primary to secondary growth: origin and development of the vascular system. *Journal of Experimental Botany* **58**, 3485–3501.
- Benfey PN, Linstead PJ, Roberts K, Schiefelbein JW, Hauser MT, Aeschbacher RA.** 1993. Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* **119**, 57–70.
- Blanc G, Wolfe KH.** 2004. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *The Plant Cell* **16**, 1679–1691.
- Bolle C.** 2004. The role of GRAS proteins in plant signal transduction and development. *Planta* **218**, 683–692.
- Bradshaw Jr HD, Ceulemans R, Davis J, Stettler R.** 2000. Emerging model systems in plant biology: poplar (*Populus*) as a model forest tree. *Journal of Plant Growth Regulation* **19**, 306–313.
- Brand U, Fletcher J, Hobe M, Meyerowitz E, Simon R.** 2000. Dependence of stem cell fate in *Arabidopsis* on a feedback loop regulated by *CLV3* activity. *Science* **289**, 617–619.
- Brunner AM, Yakovlev IA, Strauss SH.** 2004. Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biology* **4**, 14.
- Carlsbecker A, Lee J-Y, Roberts CJ, Dettmer J, Lehesranta S, Zhou J, Lindgren O, Moreno-Risueno MA, Vatén A, Thitamadee S, et al.** 2010. Cell signalling by microRNA165/6 directs gene dose-dependent root cell fate. *Nature* **465**, 316–21.
- Carretero-Paulet L, Fares MA.** 2012. Evolutionary dynamics and functional specialization of plant paralogs formed by whole and small-scale genome duplications. *Molecular Biology and Evolution* **29**, 3541–3551.
- Chang S, Puryear J, Cairney J.** 1993. A simple and efficient method for isolating RNA from pine trees. *Plant Molecular Biology Reporter* **11**, 113–116.
- Cui H, Hao Y, Kovtun M, Stolc V, Deng X-W, Sakakibara H, Kojima M.** 2011. Genome-wide direct target analysis reveals a role for *SHORT-ROOT* in root vascular patterning through cytokinin homeostasis. *Plant Physiology* **157**, 1221–1231.

Dhondt S, Coppens F, De Winter F, Swarup K, Merks RMH, Inzé D, Bennett MJ, Beemster GTS. 2010. SHORT-ROOT and SCARECROW regulate leaf growth in *Arabidopsis* by stimulating S-phase progression of the cell cycle. *Plant Physiology* **154**, 1183–1195.

Du J, Mansfield SD, Groover AT. 2009. The *Populus* homeobox gene *ARBORKNOX2* regulates cell differentiation during secondary growth. *The Plant Journal* **60**, 1000–1014.

Du J, Groover A. 2010. Transcriptional regulation of secondary growth and wood formation. *Journal of Integrative Plant Biology* **52**, 17–27.

Du J, Miura E, Robischon M, Martinez C, Groover A. 2011. The *Populus* Class III HD ZIP transcription factor *POPCORONA* affects cell differentiation during secondary growth of woody stems. *PLoS One* **6**, e17458.

Emery JF, Floyd SK, Alvarez J, Eshed Y, Hawker NP, Izhaki A, Baum SF, Bowman JL. 2003. Radial patterning of *Arabidopsis* Sshoots by class III HD-ZIP and KANADI genes. *Current Biology* **13**, 1768–1774.

Fukaki H, Wysocka-Diller J, Kato T, Fujisawa H, Benfey PN, Tasaka M. 1998. Genetic evidence that the endodermis is essential for shoot gravitropism in *Arabidopsis thaliana*. *The Plant Journal* **14**, 425–430.

Ginzberg I, Barel G, Ophir R, Tzin E, Tanami Z, Muddarangappa T, De Jong W, Fogelman E. 2009. Transcriptomic profiling of heat-stress response in potato periderm. *Journal of Experimental Botany* **60**, 4411–4421.

Groover AT, Mansfield SD, DiFazio SP, Dupper G, Fontana JR, Millar R, Wang Y. 2006. The *Populus* homeobox gene *ARBORKNOX1* reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. *Plant Molecular Biology* **61**, 917–932.

Haecker A, Gross-Hardt R, Geiges B, Sarkar A, Breuninger H, Herrmann M, Laux T. 2004. Expression dynamics of *WOX* genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. *Development* **131**, 657–668.

Hao Y, Cui H. 2012. SHORT-ROOT regulates vascular patterning, but not apical meristematic activity in the *Arabidopsis* root through cytokinin homeostasis. *Plant Signaling & Behavior* **7**, 314–317.

Hawker NP, Bowman JL. 2004. Roles for class III HD-Zip and KANADI genes in *Arabidopsis* root development. *Plant Physiology* **135**, 2261–2270.

Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN. 2000. The *SHORT-ROOT* gene controls radial patterning of the *Arabidopsis* root through radial signaling. *Cell* **101**, 555–567.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**, R19.

Jiao S, Hantke S, Zinselmeier C, Roessler JA, Niu X, Jones RJ, Habben JE, International PH, Traits A, Avenue NW, et al. 2003. Cytokinin oxidase gene expression in maize is localized to the vasculature, and is induced by cytokinins, abscisic acid, and abiotic stress. *Plant Physiology* **132**, 1228–1240.

Karimi M, Inzé D, Depicker A. 2002. Gateway vectors for *Agrobacterium*-mediated plant transformation. *Trends in Plant Science* **7**, 193–195.

Kaufert F. 1937. Factors influencing the formation of periderm in aspen. *American Journal of Botany* **24**, 24–30.

Koizumi K, Hayashi T, Wu S, Gallagher KL. 2012. The *SHORT-ROOT* protein acts as a mobile, dose-dependent signal in patterning the ground tissue. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 13010–13015.

Laurenzio L Di, Wysocka-diller J, Malamy JE, Pysh L, Helariutta Y, Freshour G, Hahn MG, Feldmann KA, Benfey PN. 1996. The *SCARECROW* gene regulates an asymmetric cell division that is essential for generating the radial organization of the *Arabidopsis* root. *Cell* **86**, 423–433.

Levesque MP, Vernoux T, Busch W, Cui H, Wang JY, Blilou I, Hassan H, Nakajima K, Matsumoto N, Lohmann JU, et al. 2006. Whole-genome analysis of the *SHORT-ROOT* developmental pathway in *Arabidopsis*. *PLoS Biology* **4**, e143.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta(\Delta CT)}$ Method. *Methods* **25**, 402–408.

Long JA, Moan EI, Medford JI, Barton MK. 1996. A member of the *KNOTTED* class of homeodomain proteins encoded by the *STM* gene of *Arabidopsis*. *Nature* **379**, 66–69.

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.

Matsumoto-Kitano M, Kusumoto T, Tarkowski P, Kinoshita-Tsujimura K, Václavíková K, Miyawaki K, Kakimoto T. 2008. Cytokinins are central regulators of cambial activity. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20027–20031.

Mayer KF, Schoof H, Haecker A, Lenhard M, Jürgens G, Laux T. 1998. Role of *WUSCHEL* in regulating stem cell fate in the *Arabidopsis* shoot meristem. *Cell* **95**, 805–15.

McConnell JR, Emery J, Eshed Y, Bao N, Bowman J, Barton MK. 2001. Role of *PHABULOSA* and *PHAVOLUTA* in determining radial patterning in shoots. *Nature* **411**, 709–713.

Mele G, Ori N, Sato Y, Hake S. 2003. The *knotted1*-like homeobox gene *BREVIPEDICELLUS* regulates cell differentiation by modulating metabolic pathways. *Genes & Development* **17**, 2088–2093.

Milhinhos A, Prestele J, Bollhöner B, Matos A, Vera-Sirera F, Rambla JL, Ljung K, Carbonell J, Blázquez MA, Tuominen H, et al. 2013. Thermospermine levels are controlled by an auxin-dependent feedback loop mechanism in *Populus* xylem. *The Plant Journal* **75**, 685–698.

Motyka V, Faiss M, Strand M, Kamínek M, Schmülling T. 1996. Changes in cytokinin content and cytokinin oxidase activity in response to derepression of *ipt* gene transcription in transgenic tobacco calli and plants. *Plant Physiology* **112**, 1035–1043.

Murashige T, Skoog F. 1962. A revised medium for rapid growth and bio assays with tobacco tissue cultures. *Physiologia Plantarum* **15**, 473–497.

Nakajima K, Sena G, Nawy T, Benfey PN. 2001. Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* **413**, 307–311.

Nieminen K, Immanen J, Laxell M, Kauppinen L, Tarkowski P, Dolezal K, Tahtiharju S, Elo A, Decourteix M, Ljung K, et al. 2008. Cytokinin signaling regulates cambial development in poplar. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 20032–20037.

Nilsson O, Aldén T, Sitbon F, Anthony Little CH, Chalupa V, Sandberg G, Olsson O. 1992. Spatial pattern of cauliflower mosaic virus 35S promoter-luciferase expression in transgenic hybrid aspen trees monitored by enzymatic assay and non-destructive imaging. *Transgenic Research* **1**, 209–220.

Otsuga D, DeGuzman B, Prigge MJ, Drews GN, Clark SE. 2001. *REVOLUTA* regulates meristem initiation at lateral positions. *The Plant Journal* **25**, 223–236.

Paquette AJ, Benfey PN. 2005. Maturation of the ground tissue of the root is regulated by gibberellin and *SCARECROW* and requires *SHORT-ROOT*. *Plant Physiology* **138**, 636–640.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**, e45.

Prigge MJ, Otsuga D. 2005. Class III homeodomain-leucine zipper gene family members have overlapping, antagonistic, and distinct roles in *Arabidopsis* development. *The Plant Cell* **17**, 61–76.

Ramírez-Carvajal GA, Morse AM, Davis JM. 2008. Transcript profiles of the cytokinin response regulator gene family in *Populus* imply diverse roles in plant development. *The New Phytologist* **177**, 77–89.

Reid KE, Olsson N, Schlosser J, Peng F, Lund ST. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology* **6**, 27.

Robischon M, Du J, Miura E, Groover A. 2011. The *Populus* class III HD ZIP, *popREVOLUTA*, influences cambium initiation and patterning of woody stems. *Plant Physiology* **155**, 1214–1225.

Rodgers-Melnick E, Mane SP, Dharmawardhana P, Slavov GT, Crasta OR, Strauss SH, Brunner AM, Difazio SP. 2012. Contrasting patterns of evolution following whole genome versus tandem duplication events in *Populus*. *Genome Research* **22**, 95–105.

Sanchez P, Nehlin L, Greb T. 2012. From thin to thick: major transitions during stem development. *Trends in Plant Science* **17**, 113–121.

Sarkar AK, Luijten M, Miyashima S, Lenhard M, Hashimoto T, Nakajima K, Scheres B, Heidstra R, Laux T. 2007. Conserved factors regulate signalling in *Arabidopsis thaliana* shoot and root stem cell organizers. *Nature* **446**, 811–814.

Scheres B, Laurenzio L Di, Willemsen V, Hauser M, Janmaat K, Weisbeek P, Benfey PN. 1995. Mutations affecting the radial organisation of the *Arabidopsis* root display specific defects throughout the embryonic axis. *Development* **121**, 53–62.

Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**, 671–675.

Schoof H, Lenhard M, Haecker A, Mayer KF, Jürgens G, Laux T. 2000. The stem cell population of *Arabidopsis* shoot meristems is maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. *Cell* **100**, 635–644.

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M. 2004. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *The Plant Cell* **16**, 2278–2292.

Schuetz M, Smith R, Ellis B. 2012. Xylem tissue specification, patterning, and differentiation mechanisms. *Journal of Experimental Botany* **64**, 11–31.

Seoighe C, Gehring C. 2004. Genome duplication led to highly selective expansion of the *Arabidopsis thaliana* proteome. *Trends in genetics* **20**, 461–464.

Solé A, Sánchez C, Vielba JM, Valladares S, Abarca D, Díaz-Sala C. 2008. Characterization and expression of a *Pinus radiata* putative ortholog to the *Arabidopsis SHORT-ROOT* gene. *Tree Physiology* **28**, 1629–1639.

Soler M. 2008. Molecular genetics of cork formation. PhD thesis, Girona University.

Sozzani R, Cui H, Moreno-Risueno MA, Bush W, Van Norman JM, Vernoux T, Brady SM, Dewitte W, Murray JAH, Benfey PN. 2010. Spatiotemporal regulation of cell-cycle genes by SHORTROOT links patterning and growth. *Nature* **466**, 128–132.

Stahl Y, Simon R. 2009a. Is the *Arabidopsis* root niche protected by sequestration of the CLE40 signal by its putative receptor ACR4? *Plant Signaling & Behavior* **4**, 634–635.

Stahl Y, Simon R. 2009b. A signaling module controlling the stem cell niche in *Arabidopsis* root meristems. *Current Biology* **19**, 909–914.

Stettler RF, Bradshaw Jr HD. 1996. Part I-Evolution, Genetics, and Genetic manipulation. In: Stettler RF, Bradshaw Jr. HD, Heilman PE, Hinckley TM, eds. *Biology of Populus and its implications for management and conservation*. Ottawa, Ontario, Canada: NRC Research Press, 1–6.

Tucker MR, Laux T. 2007. Connecting the paths in plant stem cell regulation. *Trends in Cell Biology* **17**, 403–410.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604.

Wang J, Andersson-Gunnerås S, Gaboreanu I, Hertzberg M, Tucker MR, Zheng B, Leśniewska J, Mellerowicz EJ, Laux T, Sandberg G, et al. 2011. Reduced expression of the *SHORT-ROOT* gene increases the rates of growth and development in hybrid poplar and *Arabidopsis*. *PLoS ONE* **6**, e28878.

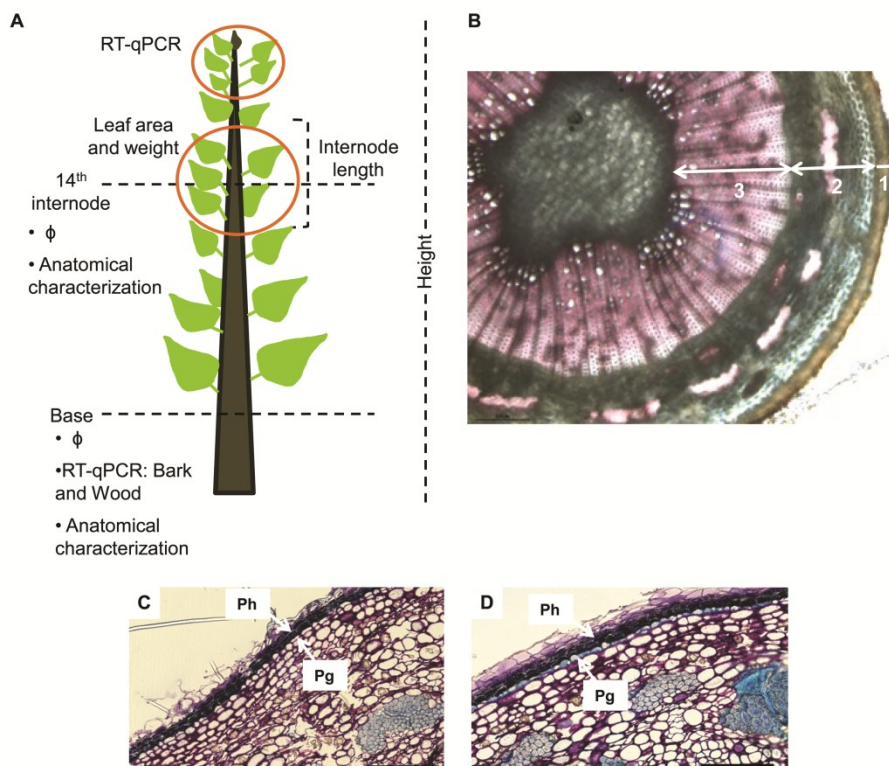
Wysocka-Diller JW, Helariutta Y, Fukaki H, Malamy JE, Benfey PN. 2000. Molecular analysis of SCARECROW function reveals a radial patterning mechanism common to root and shoot. *Development* **127**, 595–603.

Yu N-I, Lee SA, Lee M-H, Heo J-O, Chang KS, Lim J. 2010. Characterization of *SHORT-ROOT* function in the *Arabidopsis* root vascular system. *Molecules and Cells* **30**, 113–119.

Zhao S, Fernald RD. 2005. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology* **12**, 1047–1064.

Zhu Y, Song D, Sun J, Wang X, Li L. 2013. *PtrHB7*, a class III HD-Zip gene, plays a critical role in regulation of vascular cambium differentiation in *Populus*. *Molecular Plant* **6**, 1331–1343.

Supporting information



Supplementary Fig. S1. Stem tissues from hybrid aspen trees. (A) Schematic representation of tissue sampling. The first five fully expanded leaves at the top, and bark and wood at the stem base were collected for RT-qPCR experiments. The five leaves surrounding the reference internode, 14th, were collected to measure leaf area and weight. Measurements of stem diameter were performed at the 14th (EN14) and stem base (ENbase) internodes. Total tree height and length between internodes 10 and 17 (EN10 to EN17) were also evaluated. (B) Schematic representation of the stem section measurements, at the EN14 and at the ENbase. 1, phellem; 1+2, bark; 3, wood. (C-D) Anatomical aspects of hybrid aspen stem stained with toluidine blue. (C) Transverse section of a stem collected at the end of the first year of growth. (D) Transverse section of a stem collected in the second year of growth. Pg, phellogen; Ph, phellem. Bars = 100 μ m.



Supplementary Fig. S2. GUS histochemical assay of stems of greenhouse-grown *PtSHR1prom:uidA* hybrid aspen at different developmental stages. (A) 3rd internode of a recently formed shoot. (B) 9th internode from a lignified shoot. Bars = 475 μm. PX, primary xylem; VC, vascular cambium.

Chapter III

Reference gene selection for quantitative real-time PCR normalization in *Quercus suber*



Marum L¹, Miguel A¹, Pinto Ricardo C, Miguel CM. (2012). Reference gene selection for quantitative real-time PCR normalization in *Quercus suber*. PLoS ONE 7, e35113. (doi:10.1371/journal.pone.0035113). ¹Equal contribution

Andreia Miguel participated in the experimental design, performing laboratory experiments, analyzing the results and writing the paper.

Reference gene selection for quantitative real-time PCR normalization in *Quercus suber*

Abstract

The use of reverse transcription quantitative PCR technology to assess gene expression levels requires an accurate normalization of data in order to avoid misinterpretation of experimental results and erroneous analyses. Despite being the focus of several transcriptomics projects, oaks, and particularly cork oak (*Quercus suber*), have not been investigated regarding the identification of reference genes suitable for the normalization of real-time quantitative PCR data. In this study, ten candidate reference genes (*ACT*, *CACs*, *EF-1 α* , *GAPDH*, *HIS3*, *PsaH*, *SAND*, *PP2A*, *β -TUB* and *UBQ*) were evaluated to determine the most stable internal reference for quantitative PCR normalization in cork oak. The transcript abundance of these genes was analysed in several tissues of cork oak including leaves, reproduction cork and periderm from branches at different developmental stages (1, 2 and 3 year-old) or collected in different dates (active growth period *versus* dormancy). The three statistical methods (geNorm, NormFinder and CV method) used in the evaluation of the most suitable combination of reference genes identified *ACT* and *CACs* as the most stable candidates when all the samples were analysed together, while *β -TUB* and *PsaH* showed the lowest expression stability. However, when different tissues, developmental stages and collection dates were analysed separately, the reference genes exhibited some variation in their expression levels. In this study, and for the first time, we have identified and validated reference genes in cork oak that can be used for quantification of target gene expression in different tissues and experimental conditions and will be useful as a starting point for gene expression studies in other oaks.

Keywords

Cork oak, normalization, quantitative PCR, reference gene,

Introduction

The use of reverse transcription quantitative PCR (RT-qPCR) to assess transcript level has been widespread in plant biology. RT-qPCR is a sensitive, precise, easy and cost-effective method allowing the detection of low abundant mRNAs and slight variations in gene expression. It has also become the preferred method for the validation of microarray results.

To avoid bias, the use of reliable internal controls for RT-qPCR analysis is essential (Bustin *et al.*, 2009). Genes required for the maintenance of basic cellular functions, such as *ACTIN*, β -*TUBULIN*, *ELONGATION FACTOR-1 α* and *18S rRNA* are commonly used as reference genes (RGs) or internal controls. In theory, a RG is a gene with a constant level of expression in all cell types and under every experimental condition which may include developmental stages and biotic/abiotic stresses. However, a universal RG does not exist. In fact several studies reported that, according to the experimental conditions and species used, the level of expression of the commonly used RG can often be variable (Gutierrez *et al.*, 2008; Huis *et al.*, 2010; Hong *et al.*, 2010; Migocka and Papierniak, 2011), showing that these genes are differentially regulated among experimental conditions and plant species. Furthermore, it has been shown that the conventional use of a single gene for normalization may lead to relatively large errors in a significant proportion of samples (Vandesompele *et al.*, 2002; Die *et al.*, 2010). Currently, the use of multiple internal control genes is considered as an essential approach for an accurate normalization of data (Maroufi *et al.*, 2010; Demidenko *et al.*, 2011; Xu *et al.*, 2011; Migocka and Papierniak, 2011). Such an approach relies on the comparison of the mean variation of each gene relative to the mean variation of the other RGs in order to obtain the best normalization factor. Statistical algorithms, such as geNorm (Vandesompele *et al.*, 2002) and NormFinder (Andersen *et al.*, 2004), were developed to facilitate the

evaluation of potential RG expression stability under different experimental conditions. More recently, Hellemans *et al.*, (2007) also proposed the Coefficient of variation (CV) method as another powerful indicator of gene stability. Still, geNorm is the only tool that allows us to determine the minimum number of genes to be applied in normalization factor.

While the evaluation of expression stability of potential RGs has been addressed under specific conditions for species such as *Arabidopsis* (Czechowski *et al.*, 2005; Remans *et al.*, 2008; Hong *et al.*, 2010), barley, wheat and oat (Jarosová and Kundu, 2010), rice (Li *et al.*, 2010), cotton (Artico *et al.*, 2010), pea (Die *et al.*, 2010), flax (Huis *et al.*, 2010), medicago (Kakar *et al.*, 2008), tomato (Expósito-Rodríguez *et al.*, 2008) and tobacco (Schmidt and Delaney, 2010), in tree species only a few studies have been reported in poplar, spruce and longan (Brunner *et al.*, 2004; Gutierrez *et al.*, 2008; Phillips *et al.*, 2009; Lin and Lai, 2010; Xu *et al.*, 2011). Moreover, there is a general lack of information regarding the suitability of commonly used RGs in the RT-qPCR analysis of target genes expressed in recalcitrant hardwood tissues such as wood, bark or cork.

Cork oak (*Quercus suber*), an evergreen tree characteristic of the Western Mediterranean (Portugal, Spain, Southern France, Italy, North Africa), has a remarkable capacity to produce suberose tissue, the phellem or cork (Silva *et al.*, 2005), with unique properties that make it an excellent material for industrial applications. Due to the ecological and socio-economic significance of this species, large scale transcriptomic projects have been recently launched, targeting specific stress tolerance mechanisms and developmental processes such as cork differentiation (Marum *et al.*, 2011; Miguel *et al.*, 2011; Paiva *et al.*, 2011; Ramos *et al.*, 2011). Therefore, the need for RT-qPCR approaches to determine, as accurately as possible, the transcript abundance of specific genes is evident. The expression level of target transcripts in cork, as measured by

RT-qPCR, has already been studied by Soler *et al.* (2008), but a thorough evaluation of the expression stability of RG was not reported.

In order to select the most suitable RG for gene expression quantification by RT-qPCR, we analysed several tissues of cork oak including leaves, reproduction cork and periderm from branches at different developmental stages or collected in alternate seasons. Ten potential RGs involved in different biological roles, such as cytoskeleton structure [*ACTIN* (*ACT*), *β-TUBULIN* (*β-TUB*)], translational elongation [*ELONGATION FACTOR-1alpha* (*EF-1α*)], carbohydrate metabolism [*GLYCERALDEHYDE-3-PHOSPHATE DEHYDROGENASE* (*GAPDH*)], chromosome organization, biogenesis and nucleosome assembly [*HISTONE 3* (*HIS3*)], chloroplast constitution [*Photosystem I psaH* (*PsaH*)], vesicle trafficking and endocytosis [*SAND family* (*Sand*)], protein modification process [*UBIQUITIN* (*UBQ*)], protein kinase cascade [*SERINE/THREONINE PROTEIN PHOSPHATASE* (*PP2A*)] and intracellular protein transport [*CLATHRIN ADAPTOR COMPLEXES MEDIUM SUBUNIT FAMILY* (*CACs*)], were assessed using several statistical approaches for the normalization of data.

Material and Methods

Plant material

Cork oak leaves, periderm tissues isolated from branches and reproduction cork were used for sampling at these locations. While reproduction cork was harvested from 3 trees growing in Coruche and São Brás de Alportel (Portugal), leaves and branches were collected from a single donor tree at Instituto Superior de Agronomia (Lisboa, Portugal). Young leaves and 1 to 3-year-old branches were collected during the active growth period in May 2010. Three-year-old branches were also collected during the active growth period in April and July 2010 and during the dormancy period in January

2010. Periderm tissues were isolated from branches by peeling off the external bark with a sterile scalpel. Reproduction cork was harvested during the debarking period in July 2009 and 2010. All the harvested tissues were immediately frozen by immersion in liquid nitrogen and stored at -80° C until further use.

Total RNA extraction and purification

Frozen samples were ground to a fine powder in liquid nitrogen using a mortar and pestle. Total RNA was extracted following a protocol developed for grapevine (Reid *et al.*, 2006) with minor modifications: (1) all centrifugations were performed at 13200 rpm; (2) after the addition of isopropanol the recovered nucleic acids were dissolved in 375 µl Tris-EDTA buffer (pH 7.5); (3) total RNA was precipitated with 140 µl of 8 M LiCl overnight at 4° C and (4) the final pellet was dissolved in 50 µl DEPC-treated water.

To remove any traces of genomic DNA contamination after RNA extraction, two different DNase treatments, DNase I (Qiagen) and TURBO DNase (Ambion), were tested according to the manufacturer's instructions. The RNA samples treated with DNase I were also purified using the RNeasy MinElute Cleanup (Qiagen). The integrity of the RNA samples was assessed by 1% (w/v) agarose gel electrophoresis with ethidium bromide staining. RNA concentration and the 260/280 and 260/230 nm absorbance ratios were determined using a ND-1000 Spectrophotometer (NanoDrop Technologies Inc.). The automated micro-capillary electrophoresis systems currently provide accurate resolution and sensitivity for analysis of RNA quality. By this fact, the first cork oak RNA samples extracted by the methodology described before were analysed by BioAnalyzer 2100 Agilent. However, as a routine procedure, the integrity of the majority of the RNA

samples was assessed by 1% (w/v) agarose gel electrophoresis with ethidium bromide staining.

Absence of genomic DNA contamination was confirmed by performing PCR amplification using total RNA as template and primers designed for amplification of a 1069 bp DNA fragment (5'-GGAGGCGTGGAAAGTGTTTA-3'; 5'-R-ACTCAAACCCCAACGTAGCA-3') from the *GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 5* gene (*GPAT5*) coding sequence (GenBank accession number: JN819185).

First-strand cDNA synthesis and quality controls

cDNA was synthesized from 1.5 µg of total RNA using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) with the anchored-oligo(dT)₁₈ primers according to the manufacturer's instructions. To standardize each biological replicate, the products from different cDNA synthesis reactions of the same RNA sample were combined.

In order to ensure reliable results in further steps, the RNA integrity was also checked by amplifying fragments in the 5' and 3' regions of *SHORT-ROOT* (*SHR*; GenBank accession number: JN819303) cDNA by qPCR. Primers were designed in the 5' and 3' regions of the *SHR* cDNA to amplify fragments of 177 and 193 bp, respectively (*SHR_5'F*: ATGGATACC TTGTTTAGGC; *SHR_5'R*: GGTTCAAGTCCAATTTTCGTTC; *SHR_3'F*: GTAGTGTCTAGAAGAAGACG; *SHR_3'R*: GCGTGTAAGAAGGTACGC). The 3'/5' ratio was determined according to the following equation:

$$(1 + E_{SHR3'})^{Cq_{SHR3'}} / (1 + E_{SHR5'})^{Cq_{SHR5'}} \quad (\text{Nolan et al., 2006}).$$

Primer design

Ten candidate RGs were evaluated in this study: *ACT*, *CACs*, *EF-1α*, *GAPDH*, *HIS3*, *PsaH*, *SAND*, *PP2A*, *β-TUB* and *UBQ*. The RGs were chosen based on their previous use as internal controls in gene expression

studies of hardwood species such as *Q. suber*, *Populus* species and *V. vinifera*, and based on their consistent PCR amplification. RG sequences were obtained from the *Fagaceae* database, Fagaceae Genome Web (<http://www.fagaceae.org/>) and from GenBank. Primers were designed using Primer3 software (Rozen and Skaletsky, 2000) and PCR Primer Stats (Stothard, 2000) taking into account the following criteria: annealing temperature of 60° C, GC content of 42-55% and primer length of 19-21 bp. The sequence accession numbers, the closest *Arabidopsis* homolog, as well as the primer sequences and amplicon size, are described in Table 1. To confirm the specificity of primer annealing the amplicons obtained after PCR amplification were sequenced, with the exception of *ACT* and β -*TUB* already available in Genbank (EU697020 and EE743717, respectively). The amplicon sequences are presented in the supplementary data (Supplementary Table S1).

qPCR conditions and PCR efficiency

The experiments were carried out in 96-well plates with a LightCycler 480 (Roche) using SYBR Green I Master (Roche) to monitor the PCR amplification. Reaction mixtures contained 10 μ l of 2x SYBR Green I Master, 400nM of each primer and 1.5 μ l of cDNA as template, in a total volume of 20 μ l. The following amplification program was used in all PCR reactions: 95° C for 10 min, 45 cycles of 10 s at 95° C, 10 s at 60° C and 10 s at 72° C. The specificity of each amplification reaction was verified by a dissociation curve (melting curve) analysis after the 45 cycles, by heating the amplicon from 65° C to 97° C. No-template controls were included for each primer pair.

For all RGs studied 2 biological samples were used and the expression levels in each sample were based on 3 technical replicates. Leaf

samples were used as calibrator to normalize the values between different plates.

Two different approaches were tested to determine the amplification efficiencies of the RGs, using leaves as sample: a standard curve with a three dilution series calculated according to the equation $(1+E) = 10^{\text{slope}}$ and the statistical algorithm Real-time PCR Miner. PCR efficiencies (E) for all the other samples were estimated with the Real-time PCR Miner algorithm (Zhao and Fernald, 2005) using the raw fluorescence data as input. Mean efficiency values were obtained for each biological replicate and were used to adjust quantification cycle (Cq) values for subsequent analysis.

Experimental design and data analysis

Several RT-qPCR experiments were performed to analyse transcript levels in cork oak leaves, cork, periderm from 1, 2 and 3 year-old branches and periderm from 3 year-old dormant branches. For clarity, samples corresponding to periderm from 1, 2 and 3 year-old branches together with cork will be referred to as developmental stage set and periderm from 3-year-old branches collected during active growth (May) *versus* dormancy (January) will be referred to as the seasonal growth sample set. Comparative analyses of all cork oak tissue samples, as well as individual analyses of the several sample types, were also performed.

RG transcript abundance in all the samples was determined by the Cq value or the number of cycles needed to reach a specific threshold level of detection in the exponential phase of the PCR reaction. Three statistical approaches were used to determine the stability of the candidate RGs: geNorm v3.5 (Vandesompele *et al.*, 2002), NormFinder (Andersen *et al.*, 2004) and CV method (Hellemans *et al.*, 2007). The Cq values were converted into relative quantities to be used as input data for geNorm and

NormFinder (only Cq<40 were used for analysis). The conversion was performed through the formula $Q = E^{\Delta Cq}$, where E is the efficiency of the gene amplification for each primer pair (in each tissue) and ΔCq is the lowest Cq value as calibrator (which corresponds to the sample with the highest expression) minus the Cq value of the sample tested. The data obtained from each biological replicate were analysed in separate. For the CV method the relative quantities were first transformed into normalized relative quantities (formula 15 as in Hellemans *et al.* (2007)) and the CV was calculated using the standard error (through the formulas 17-19 in Hellemans *et al.* (2007)). Finally, the normalization factor (NF) was based on the geometric mean of the best RG selected.

Validation of RG analysis

One gene of interest putatively coding for a *GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 5* (*GPAT5*) (GenBank accession number: EE743865), was used to validate the selected RGs. Primers were designed using Primer3 software (Rozen and Skaletsky, 2000) (*GPAT5_5`F*: GCTAGAGCGGTCTTGACAAAG-3`; *GPAT5_5`R*: GACCTCATCAGCTCG CAAAT-3`). The relative expression level of the target gene was determined in periderm tissues from 3-year-old branches collected in April and July 2010. The experimental procedure was the same as used in the selection of RGs. For comparative purposes, the relative expression of the target gene was calculated with different normalization factors based on the geometric mean of the two most stable genes [lower M value, NF2(S)] and the two most unstable genes [higher M value, NF2(U)].

Table 1. Description of the 10 candidate reference genes and primer sequences for RT-qPCR.

Gene abbreviation	Gene description	Accession number	<i>Arabidopsis</i> homolog locus	Primer sequences (forward/reverse)	Amplicon length (bp)
<i>ACT</i>	ACTIN	EU697020	At5g09810	GCTGGTCGTGATCTAACTG/ CTTTGCAGTCTCCAACTCCT	153
<i>CACs</i>	CLATHRIN ADAPTOR COMPLEXES medium subunit family protein	ID6728500	At5g46630	TCTGGGAGAAGAGTGGCTACA/ GAGCCACCATTCAAAATCCT	175
<i>EF-1α</i>	ELONGATION FACTOR 1- α	ID2007208	At5g60390	TTGTGCCGTCCTCATTATTGACT/ TCACGGGTCTGACCATCCTT	75
<i>GAPDH</i>	GLYCERALDEHYDE-3-PHOSPHATE DEHYDRIGENASE	ID6744996	At3g04120	ACCGACTTCATTGGTGACAG/ AGATGCGATGTGGACAATCA	150
<i>HIS3</i>	HISTONE 3	ID6923956	At1g09200	GCTCTTCGAGGACACCAATC/ TAAGCCCTCTCGCCTCTGAT	102
<i>PsaH</i>	Photosystem I psaH protein	ID6967182	At1g52230	CAGTTGCTCTGAAACCAAGG/ CACAGCACCACTCCTGAACT	163
<i>PP2A</i>	Serine/threonine protein phosphatase 2A	ID6966573	At3g25800	GAGCCACTCTATCCGATTGC/ GTCGTCATTGTTCTCGCTGA	162
<i>SAND</i>	SAND family protein	ID6961629	At2g28390	AGGATTGCAGGATTCGTATTG/ GACCACCAATGCCAACAAA	204
<i>B-TUB</i>	β -TUBULIN	EE743717	At2g29550	AAGAACATGATGTGCGCTGCT/ TCCACCTCCTTGGTGCTCA	92
<i>UBQ</i>	UBIQUITIN	ID3028522	At4g05320	CGAAGATCCAGGACAAGGAG/ CAGGGCTTTTCACTCCTCAG	172

Results

Data normalization using a set of reference genes (RGs) is nowadays a current and crucial procedure when analysing the expression levels of target transcripts by RT-qPCR in different tissues or under different conditions. In the present study, the transcript abundance of 10 potential RGs was assessed in cork oak by qPCR. A total of 36 cDNA samples including several tissue types and periderm tissues from branches under different developmental stages or collected during dormancy *versus* active growth period, were analysed.

RNA quality

The assessment of RNA quality encompassed both its purity, characterized by the absence of protein and DNA contamination, and its integrity. All samples were analysed spectrophotometrically and showed absorbance ratios at 260/280 and 260/230 nm above 1.8. Total RNA samples were also analysed in agarose gels showing well defined bands corresponding to the rRNA and absence of nucleic acid degradation. To confirm the absence of contaminating genomic DNA, positive and no RT controls were used in *GPAT* amplification. The DNase I (Qiagen) treatment proved inefficient for the complete removal of genomic DNA from total RNA extracted from cork oak tissues. However, the Turbo DNase (Ambion) proved effective in the removal of DNA contamination, since gene amplification was obtained only from reverse transcribed samples.

The RNA integrity was checked by performing a 3':5' assay according to Nolan *et al.* (2006). This assay gives an indication of the mRNA integrity, since in most cases the RNA degradation starts in the 5'-end region. In general, while a 3':5' ratio close to 1 means that the percentage of full-length transcripts in the sample is high, a ratio higher than 5 suggests degradation in the RNA samples. In our study the obtained 3':5'

ratios were close to 1 with primer efficiencies of 1.8 and 1.9. These results showed that total RNA samples used for RT-qPCR analyses were pure, non-degraded and free of DNA contamination.

qPCR experiments and PCR efficiency

Specificity of amplification of the several transcripts was supported by the analysis of melting curves and by gel electrophoresis, showing a single PCR amplification product with the expected size for each gene (Supplementary Fig. S1) and further confirmed by amplicon sequencing. The PCR efficiency (E) of each primer pair was first calculated in cork oak leaves through the standard curve method and then compared with the E value obtained through the statistical algorithm PCR Miner. According to Czechowski *et al.* (2005) both methods give similar amplification efficiencies. In our study using cork oak leaves, the E values obtained by both methods were similar (Supplementary Table S2). However, the standard curve method is time consuming, requiring the production of repeatable and reliable standards (Pfaffl, 2001), with no errors from contamination or sample dilution. Moreover, this method relies on the assumption that the PCR efficiency of each amplicon is constant in all samples, which rarely can be achieved in real experiments (Zhao and Fernald, 2005), strongly influencing the C_q analysis (Kamphuis *et al.*, 2001; Ramakers *et al.*, 2003). Therefore, Real-time PCR Miner algorithm (Zhao and Fernald, 2005), using the single raw fluorescence data as an input, was the chosen method to calculate the PCR efficiency for each primer pair in each tissue type (Table 2).

The 10 potential RGs tested (*ACT*, *β-TUB*, *EF-1α*, *GAPDH*, *HIS3*, *PsaH*, *SAND*, *UBQ*, *PP2A* and *CACs*) were successfully amplified in cork oak tissues. The efficiency values (E) were calculated as the mean values obtained from the technical and biological replicates (Table 2), and used to adjust C_q values for subsequent analysis. Primer efficiencies were higher

than 1.9 for all the experiments, except for *UBQ* in cork and for *PsaH* in periderm from 1 and 2-year-old branches and cork, where the values varied between 1.84 and 1.89 (Table 2). Altogether, these results confirm that the selected primers accurately amplify the potential RGs.

Table 2. PCR amplification efficiency of each primer pair.

Genes	Tissue/stage					
	Leaves	1 st B	2 nd B	3 rd B	3 rd DB	Cork
<i>ACT</i>	1.93 ± 0.03	1.96 ± 0.02	1.94 ± 0.02	1.90 ± 0.02	1.91 ± 0.02	1.94 ± 0.03
<i>CACs</i>	1.93 ± 0.02	1.97 ± 0.02	1.94 ± 0.01	1.94 ± 0.02	1.97 ± 0.02	1.95 ± 0.02
<i>EF-1α</i>	1.92 ± 0.02	1.95 ± 0.02	1.92 ± 0.02	1.96 ± 0.02	1.96 ± 0.02	1.93 ± 0.02
<i>GAPDH</i>	1.98 ± 0.02	1.96 ± 0.01	1.91 ± 0.02	1.92 ± 0.03	1.90 ± 0.01	1.94 ± 0.02
<i>HIS3</i>	1.91 ± 0.03	1.95 ± 0.01	1.94 ± 0.02	1.94 ± 0.02	1.95 ± 0.03	1.92 ± 0.03
<i>PP2A</i>	1.95 ± 0.01	1.92 ± 0.02	1.97 ± 0.02	1.96 ± 0.02	1.91 ± 0.04	1.97 ± 0.02
<i>PsaH</i>	1.93 ± 0.04	1.89 ± 0.02	1.89 ± 0.03	1.92 ± 0.03	1.94 ± 0.03	1.84 ± 0.00
<i>UBQ</i>	1.94 ± 0.02	1.91 ± 0.02	1.92 ± 0.01	1.93 ± 0.02	1.94 ± 0.03	1.86 ± 0.01
<i>SAND</i>	1.92 ± 0.02	1.92 ± 0.02	1.95 ± 0.03	1.91 ± 0.01	1.91 ± 0.03	1.93 ± 0.05
<i>β-TUB</i>	1.99 ± 0.03	1.99 ± 0.02	1.97 ± 0.02	1.95 ± 0.01	1.95 ± 0.03	1.96 ± 0.04

Efficiency values obtained after the amplification of each candidate RG (*ACT*, *β-TUB*, *EF-1α*, *GAPDH*, *HIS3*, *PsaH*, *SAND*, *UBQ*, *PP2A*, *CACs*) in leaves, periderm from 1, 2 and 3-year-old branches (1stB, 2ndB and 3rdB), periderm from 3-year-old branches in the dormancy period (3rdDB) and cork, estimated with the Real-Time PCR Miner algorithm.

The calculation of mean Cq values for the ten RGs in all cDNA samples (Fig. 1) showed a range of variation from 14.8 to 31.0. *GAPDH* displayed the most abundant transcript level, while *PsaH* was the less abundant. Based on the interquartile range (25-75% percentiles) for Cq values, the

lower Cq dispersion was observed for *ACT*, *CACs*, *EF-1 α* , followed by *UBQ*, *SAND* and *PP2A*.

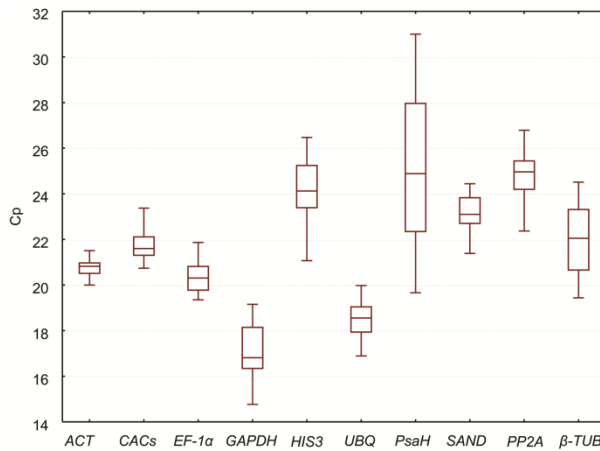


Fig. 1. Range of Cq values of the candidate reference genes obtained for all cDNA samples. Each box corresponding to *ACT*, *CACs*, *EF-1 α* , *GAPDH*, *HIS3*, *UBQ*, *PsaH*, *SAND*, *PP2A* and β -*TUB* indicates the 25% and 75% percentiles. Whiskers represent the maximum and minimum

values. The median is depicted by the line across the box.

Stability analysis

In order to identify and rank the most suitable RGs based on their expression stability, three different statistical approaches, geNorm, NormFinder and CV method, were tested. For all analyses, the Cq values were transformed into relative quantities using the Δ Cq method, and the amplification efficiencies of the RG were calculated by PCR Miner algorithm.

When using geNorm algorithm the candidate RGs were ranked according to their expression stability measure (M), which represents the average pairwise variation of a particular gene with all other control genes. The stability values are reached after stepwise exclusion of the worst-scoring RG. Considering the data obtained from all samples, *ACT* and *CACs* were the most stable genes (lowest M value of 0.462), followed by *EF-1 α* (M value of 0.525) (Fig. 2). Hellemans *et al.* (2007) recommended a stability measure threshold lower than 1 to ensure the most stable genes are selected. In our study six of the genes showed an M value lower than 1. The highest M value (2.203) was observed for *PsaH*, the most unstable

gene, which can be explained by the lower expression levels observed for this gene in some of the tissues, namely cork and a few periderm samples. A similar ranking of the tested genes was obtained when expression stability was analysed through the NormFinder algorithm taking into account all cDNA samples (Fig. 2). The three most stable (*ACT*, *CACs* and *EF-1 α*) and the two least stable (*PsaH* and *HIS3*) candidate genes were the same as identified by geNorm. When using the CV method, *ACT* and *CACs* were also the best performing genes, but a different ranking of the remaining candidate RGs was obtained (Fig. 2).

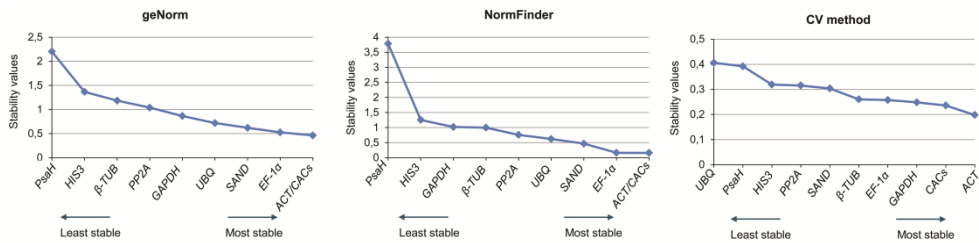


Fig. 2. Stability values of candidate reference genes (RGs) calculated by different statistical methods using all cDNA samples. Ranking of each RG (*ACT*, *CACs*, *EF-1 α* , *GAPDH*, *HIS3*, *UBQ*, *PP2A*, *PsaH*, *SAND*, β -*TUB*), calculated by geNorm, NormFinder and CV method, for all tested samples [leaves, branches (1st, 2nd, 3rd active, 3rd dormancy) and cork].

The analysis of the developmental stage data set by geNorm and NormFinder also revealed a similar ranking of the tested RGs, while in the seasonal growth data set analysed by NormFinder and CV method, the most and less stable genes were similar but the intermediate rank positions differed (Supplementary Fig. S2). Combining the three statistical approaches in the analysis of the developmental stage and seasonal growth data sets, the *ACT/CACs/EF-1 α /PP2A* and *ACT/CACs/GAPDH* clusters represented the most stable genes in each set, respectively (Fig. 3).

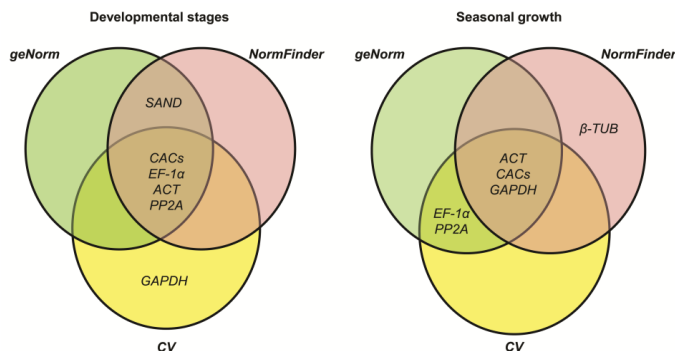


Fig. 3. Venn diagram showing the most stable genes identified by the geNorm, NormFinder and CV method. The most stable genes were identified using data from the developmental stage and seasonal growth sample sets.

However, when performing separate analysis of specific tissues, developmental stages or active growth *versus* dormancy, less uniform results were obtained (Table 3). For instance, in the active growth *versus* dormancy, *CACs/GAPDH*, and *EF-1 α /SAND* can be used as good RG according to geNorm, while NormFinder selected *CACs/GAPDH* and *UBQ* and CV method selected *CACs* and *ACT* as the most adequate for data normalization.

On the other hand, according to Hellemans *et al.* (2007), the mean stability values (M) and the mean CV for heterogeneous sample panels should be within the $M \leq 1$ and $CV \leq 0.5$ ranges. Taking into account all cDNA samples, our results completely match these criteria. The obtained M values ranged from 0.462 to 0.866, except for the four less stable genes (*PP2A*, *β -TUB*, *HIS3* and *PsaH*), while CV values were in the range of 0.198-0.406. For homogeneous sample panels Hellemans *et al.* (2007) consider a different range, $M \leq 0.5$ and $CV \leq 0.25$, for selecting the best RG. The M and CV values for the most stable RG selected from the separate analysis of the developmental stages set, seasonal growth set and the individual analysis of each sample type, also fits the proposed value ranges (Table 3; Supplementary Fig. S2).

Table 3. Stability values for the candidate RGs in individual sample types.

Method	Tissue/ stage	ACT	CACs	EF-1 α	GAPDH	HIS3	UBQ	PsaH	SAND	PP2A	B-TUB
GeNorm	Leaves	0.023	0.157	0.008	0.008	0.485	0.065	0.130	0.252	0.302	0.013
	1 st B	0.030	0.087	0.009	0.194	0.138	0.068	0.002	3.38E-04	0.272	3.38E-04
	2 nd B	0.098	0.276	0.022	0.255	0.145	0.059	0.843	0.209	0.432	0.22
	3 rd B	0.167	0.050	0.218	0.050	0.288	0.415	1.086	0.590	0.080	0.686
	3 rd DB	0.458	0.396	0.047	0.194	1.126	0.631	1.967	0.047	1.508	0.528
NormFinder	Cork	0.548	0.608	0.679	0.472	0.233	0.308	0.214	0.086	0.052	0.052
	Leaves	0.065	0.317	0.105	0.096	0.841	0.037	0.311	0.197	0.207	0.118
	1 st B	0.093	0.006	0.041	0.309	0.176	0.006	0.014	1.17E-04	0.402	1.17E-04
	2 nd B	0.034	0.009	0.144	0.009	0.039	0.247	1.722	0.031	0.840	0.173
	3 rd B	0.036	0.017	0.036	0.017	0.239	0.433	1.858	0.840	0.024	0.878
CV method	3 rd DB	0.467	0.571	1.190	0.946	1.153	0.116	2.616	1.152	1.609	0.163
	Cork	0.353	0.467	0.658	0.338	0.521	0.090	0.517	0.292	0.174	0.224
	Leaves	0.145	0.038	0.131	0.186	0.382	0.422	0.352	0.280	0.323	0.264
	1 st B	0.182	0.253	0.193	0.115	0.289	0.273	0.249	0.219	0.319	0.201
	2 nd B	0.141	0.182	0.159	0.163	0.129	0.216	0.440	0.170	0.320	0.216
	3 rd B	0.167	0.055	0.211	0.257	0.127	0.377	0.444	0.365	0.207	0.352
	3 rd DB	0.107	0.182	0.356	0.285	0.442	0.167	0.470	0.325	0.356	0.157
	Cork	0.345	0.377	0.395	0.299	0.298	0.325	0.485	0.348	0.227	0.207



◀ The stability values for *ACT*, *β-TUB*, *EF-1α*, *GAPDH*, *HIS3*, *PsaH*, *SAND*, *UBQ*, *PP2A*, CACs in leaves, periderm from 1, 2 and 3 years-old branches and cork (1stB, 2ndB, 3rdB) or from 3-year-old branches collected in alternate seasons corresponding to active growth *versus* dormancy (3rdB, 3rdDB) and cork, were calculated by geNorm, NormFinder and CV method. The values in bold refer to the most stable genes.

The optimal number of RGs used for data normalization was determined through the pairwise variation ($V_{n/n+1}$), using the geNorm algorithm (Vandesompele *et al.*, 2002). This is calculated between the two sequential normalization factors, NF_n and NF_{n+1} for all the samples under analysis and reveals the effect of adding an $(n+1)^{th}$ gene, indicating whether the inclusion of an extra reference gene adds stability to the normalization factor. A small variation means no significant effect on the normalization by the addition of another gene. Our study reveals that all $V_{n/n+1}$ values were below 0.15, except when the analysis was performed using the data from all the samples (Fig. 4). Although Vandesompele *et al.* (2002) recommended a cut-off value of 0.15 (below which the addition of new gene is not required), this should not be considered as a strict threshold (“geNorm software manual”) and several subsequent studies have reported higher cut-off values of $V_{n/n+1}$ (Kuijk *et al.*, 2007; Silveira *et al.*, 2009; Maroufi *et al.*, 2010). Our data shows a small variation between $V_{2/3}$ (0.169) and $V_{3/4}$ (0.162), when all the samples were analysed together, suggesting that the addition of a third gene has no significant effect on the normalization factor.

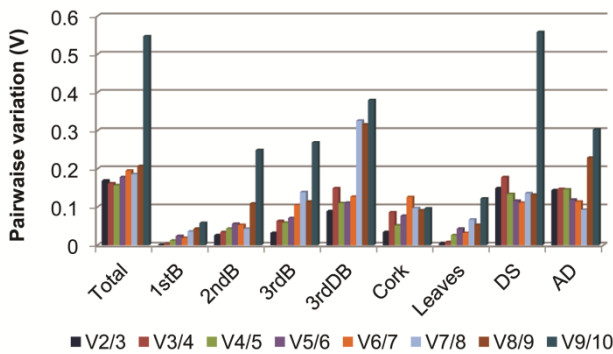


Fig. 4. Determination of the optimal number of reference genes for normalization according to geNorm software. Pairwise variation ($V_{n/n+1}$) analysis between the normalization factors NF_n and NF_{n+1} , carried out for all the samples (Total), individual

samples [leaves, periderm from 1-year-old (1stB), 2-year-old (2ndB) and 3-year-old branches during active growth (3rdB) or dormancy (3rdDB) and cork], developmental stage sample set (DS) and seasonal growth sample set (SG).

RG validation

The conclusions from the analyses described above were applied to quantify the transcript level of a gene of interest, *GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE 5* (*GPAT5*). The quantification of transcripts was performed in periderm tissues from 3-year-old branches collected in the spring (April) and summer (July), to validate the RGs selected by the different statistical methods. The seasonal variation in *GPATs* transcript level had been previously evaluated in cork oak by Soler *et al.* (2008) using RT-qPCR.

Different RGs combinations were tested in order to assess whether the use of the different normalization factors (NF) obtained by geNorm, NormFinder and CV method had an impact on the transcript quantification results. The several NF were calculated taking into account the stability of the RG as inferred by different statistical methods (geNorm, NormFinder and CV method) and the type of data used for the analysis. Thus, the NF were determined using the two most stable, NF2(S), and the two most unstable genes, NF2(U), identified by geNorm and NormFinder, when

analysing (1) data from all the samples [NF2(S) (*ACT*; *CACs*) and NF2(U) (*HIS3*; *PsaH*)] and (2) data from periderm tissues of 3-year-old branches [NF2(S) (*CACs*; *GAPDH*) and NF2(U) (β -*TUB*; *PsaH*)] (Fig. 5A, B). The NF were also calculated using the two most stable as well as the two most unstable genes identified by the CV method, when analysing data from all the samples [NF2(S) (*ACT*; *CACs*) and NF2(U) (*PsaH*; *UBQ*)] or data from periderm of 3-year-old branches [NF2(S) (*CACs*; *GAPDH*) and NF2(U) (*UBQ*; *PsaH*)] (Fig. 5C, D).

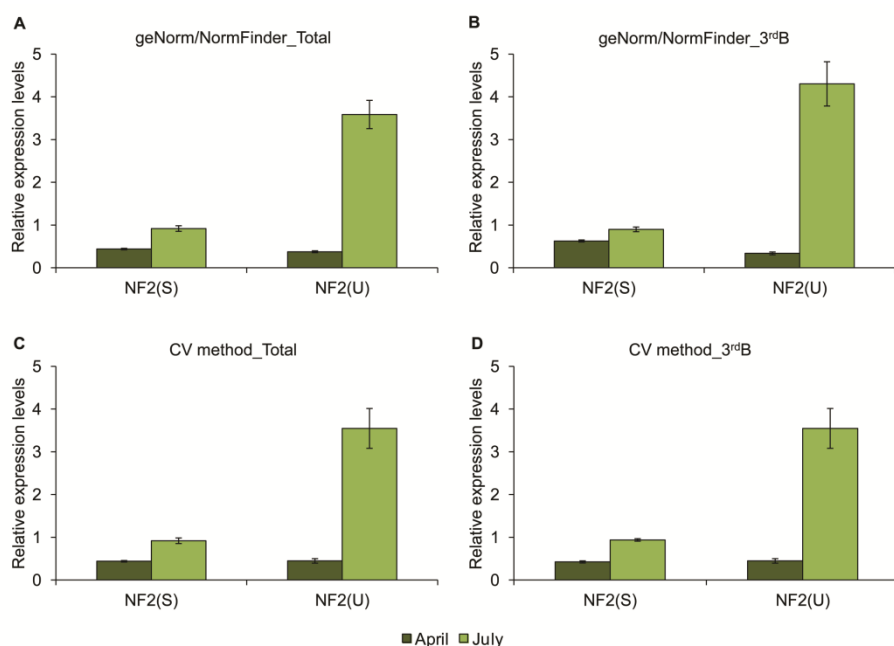


Fig. 5. Validation of the reference genes (RGs). Relative expression levels of *GPAT5* in periderm from 3-year-old branches collected in spring (April) and summer (July). Normalization factors were calculated with RGs obtained in the analysis of data from all samples (Total) by geNorm/NormFinder (A) and CV method (C) or data from periderm of 3-year-old branches (3rdB), also in geNorm/NormFinder (B) and CV method (D). The normalization factors were based on the geometric means of the two most stable genes [NF2(S)] and the two most unstable [NF2(U)].

When the *GPAT* expression level was calculated with the NF2(S) obtained by the several statistical methods, a small variation (<0.5-fold) between the two seasons (spring and summer) was observed. However, important changes in the relative expression levels (≥ 3 fold) were obtained when the several NF2(U) were used.

The different NF used also enabled us to calculate the average gene specific variation based on the most stable and unstable genes identified from the two type of data used for the analyses. The smallest average gene-specific variation (22.66%) was obtained with NF2(S) (*ACT*; *CACs*), calculated by geNorm and NormFinder, for the data from periderm samples (Fig. 6). The highest gene-specific variation, 89% and 96%, was obtained with the NF calculated with the most unstable genes from geNorm, NormFinder and CV method, using data from all the samples and from periderm samples, respectively (Fig. 6).

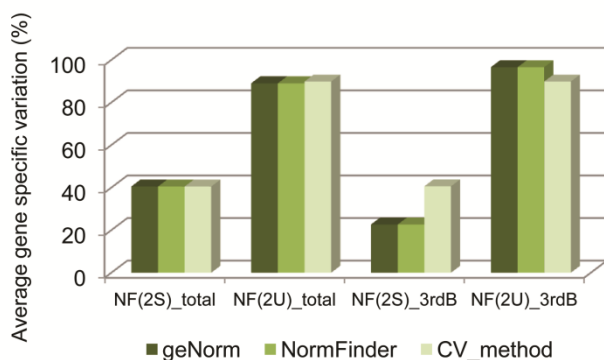


Fig. 6. Average *GPAT5* gene specific variation. Determination of the coefficient of variation in percentage, for the two most stable, NF(2S), and the two most unstable genes, NF(2U). This analysis was performed for the RGs selected from all tested samples (total) and from

periderm from 3-year-old branches (3rdB), using three statistical methods (geNorm, NormFinder and CV method).

Discussion

Gene expression can vary across tissues or cell types as well as developmental and physiological stages. However, the genes commonly

referred to as housekeeping function genes (HK) required to maintain basic cellular functions, are expressed in all metabolically active cells or tissues, being critical to the activities that must be carried out for successful completion of cell cycle (Warrington *et al.*, 2000). Molecular characterization studies have even pointed to characteristic features of HK genes such as a lower degree of conservation of their promoters when compared to those of non-HK genes, and a higher density of SSRs in their 5'-UTRs (Farré *et al.*, 2007; Lawson and Zhang, 2008). Due to their role, HK genes have been widely chosen as valuable controls (like RGs) in gene expression analyses. Although HK genes continue to be often used, several studies showed that the HK genes are not necessarily expressed at the same level in all tissues (Zhang and Li, 2004; Kouadjo *et al.*, 2007; Huis *et al.*, 2010). Therefore, it is recognized that a careful choice of RGs supported by experimental evidence is essential to obtain a reliable normalization of gene expression data for accurate quantification of transcript levels.

Although some authors still use RGs from model plants, several studies have demonstrated that such genes are frequently unstable. For example, the *GAPDH* gene, involved in basic cellular functions and often assumed to have a uniform expression pattern, is one of the most stable genes in barley, oat and grapevine (Reid *et al.*, 2006; Jarosová and Kundu, 2010) but in *N. tabacum* it proved to be much less stable (Schmidt and Delaney, 2010). Furthermore, it has been advocated that even for the same type of samples the RG can vary between experiments and laboratories, which can lead to misleading results. Nevertheless, the experimentally induced variations (samples, operator and instruments) can be strongly reduced by implementing a robust methodology.

In this study, the evaluation of ten candidate RGs in cork oak samples encompassing different tissues, developmental and seasonal growth stages, was performed following a careful experimental design using several controls for checking RNA quality and integrity and PCR efficiency

and specificity. Furthermore, the consistency of the best-scoring RG was tested by several statistical approaches available for this purpose. One of the critical steps addressed in the experimental procedures was to assure the high quality of total RNA isolated from cork oak recalcitrant tissues such as periderm and cork, containing highly suberized cells and high contents of secondary metabolites. This issue was successfully overcome by the use of an optimized RNA extraction protocol and a specific DNase treatment. The RNA isolation method described here had been previously optimized for cork oak tissues and its effectiveness checked by several available methods including analysis in the Agilent 2100 Bioanalyzer (unpublished results). Although, automated micro-capillary electrophoresis systems currently provide higher resolution and sensitivity for analysis of RNA quality than agarose electrophoresis, the combination of several procedures namely spectrophotometric analysis in the NanoDrop, gel electrophoresis and the 3':5'-assay used in this work, was adequate for checking the quality of RNA samples isolated through this well established isolation procedure.

As expected, the variation in expression stability of the ten RGs tested in our study suggests there is no single RG that can be used for a diversity of cork oak samples. The use of multiple RGs that are stable under a given experimental condition was the chosen method for the normalization of RT-qPCR, in agreement with previous reports in species such as *Arabidopsis*, wheat, barley, oat, tomato, tobacco, pea, cucumber and poplar (Expósito-Rodríguez *et al.*, 2008; Schmidt and Delaney, 2010; Die *et al.*, 2010; Demidenko *et al.*, 2011; Xu *et al.*, 2011; Migocka and Papierniak, 2011), where a similar strategy has proved efficient for the relative quantification of a target gene.

Many of the RGs used until now have been selected based on data compiled from microarray databases. However, Czechowski *et al.* (2005) showed that, depending on the specificity of the database used to assess the expression stability, the ranking of the RGs can change. In order to find

adequate RGs, several analysis tools have been developed (Vandesompele *et al.*, 2002; Andersen *et al.*, 2004; Pfaffl *et al.*, 2004; Hellemans *et al.*, 2007) but the question of which procedure is the most suitable remains open. From the different statistical approaches used in this work to analyse gene expression stability, the geNorm and NormFinder algorithms generated a similar RG ranking when considering data from all the samples (heterogeneous sample panel), while the ranking obtained with the CV method was slightly different. Although CACs and ACT were within the best-scoring RGs by any of the methods, *EF-1 α* was ranked as one of the most stable genes by geNorm and NormFinder but not by the CV method. ACT and *EF-1 α* have been traditionally used as RGs namely in *N. tabacum*, *V. vinifera* and *E. ulmoides*. The high stability of CACs has also been described in *Arabidopsis* time-course experiments, in *Cucumis sativus* subjected to abiotic stress and exogenously applied growth regulators (Migocka and Papierniak, 2011), in different plant structures of *Fagopyrum esculentum* (Demidenko *et al.*, 2011), and in vegetative and reproductive organs of *Vaccinium spp.* (Vashisth *et al.*, 2011). In fact, this gene is involved in a number of essential cellular processes, including membrane trafficking, protein sorting and endocytosis, and it has been identified in *Arabidopsis* as being among the five most stable genes. On the other hand, *EF-1 α* , one of the genes with a lower M value in our study, was shown to be quite unstable in *Arabidopsis* (Czechowski *et al.*, 2005) and in *Salvia miltiorrhiza* (Yang *et al.*, 2010) when geNorm algorithm was used, and also in barley, oat and wheat (Demidenko *et al.*, 2011) when three different algorithms were applied. These results confirm that a universal reference gene does not exist, highlighting the need to evaluate commonly used RGs for a particular species or condition. When the expression stability was analysed separately for each sample set, the ranking of the RG stability was not uniform. Some of the variation in expression levels may be due to the role of the RG in specific tissues. For example, the role of *PsaH* in the

chloroplast as a component of photosystem I could explain the low expression levels found for this gene in some of the tested tissues such as cork. On the contrary, β -*TUB* was one of the most stable genes in cork. In fact, β -*TUB* had already been used as a RG in the quantification of cork transcripts by Soler *et al.* (2008).

Despite the differences we found in the RG ranking when using different statistical approaches, there is a general agreement among the methods for the selection of the most stable genes. The geNorm algorithm is based on the geometric averaging of multiple genes and it has been the most used method (Gutierrez *et al.*, 2008; Artico *et al.*, 2010; Huis *et al.*, 2010; Yang *et al.*, 2010; Schmidt and Delaney, 2010; Die *et al.*, 2010; Hong *et al.*, 2010) due to its simplicity and robustness in the calculation of the NF. The CV method is another powerful indicator which represents the variation of the normalized quantities across the tested samples while the NormFinder allows estimation, not only of the overall variation of the candidate gene, but also of the variation among sample subgroups. According to Andersen *et al.* (2004), NormFinder is more effective in avoiding the effect of gene co-regulation because it takes into account the intra and inter-group variation. The geNorm software has the advantage of indicating the minimal number of RGs required for data normalization through the pairwise variation tool. The use of an increased number of RGs in the normalization can improve the reliability of a study, but it is time-consuming and more expensive and thus, a trade-off between the gain in accuracy and the costs and time involved needs to be carefully balanced. In the analysis of our data, the determination of the pairwise variation of two sequential normalization factors ($V_{n/n+1}$) using the geNorm software, indicated that 2 was the minimum number of RGs to be included in the normalization for all the analysed sample sets. The inclusion of a third gene does not add any significant contribute to the calculation of the normalization factor.

Based on previous studies for determination of seasonal variance in transcript abundance of *GPAT5* in cork tissues, we chose this gene as a target to validate the RGs selected in this study. *GPAT5* is involved in suberin biosynthesis, one of the main compounds of cork (Beisson *et al.*, 2007) and it has been identified in a EST collection showing a high and specific expression in the suberin-rich phellem of the cork oak tree (Soler *et al.*, 2007). When *GPAT5* expression was measured in periderm tissues from 3 year-old cork oak branches no significant variation in transcript abundance was found between April and July, which was in accordance with previous results reported by Soler *et al.* (2008). The importance of selecting the most stable genes to calculate the NF was evidenced by the huge difference observed in the relative expression levels when NF was calculated with the most stable genes *versus* the most unstable ones. The genes selected by geNorm and NormFinder, when data from the periderm of 3-year-old branches were used, seem to be more reliable to quantify the relative expression levels when comparing to the CV method, as judged by the lowest average gene-specific variation obtained with NF2(S) (*ACT*; *CACs*).

This study is the first attempt to identify RGs in several cork oak tissues. We concluded that *ACT* and *CACs* were the most stable genes even when considering heterogeneous sample sets, and these were further validated in the transcript quantification of a target gene. These results should be a solid starting point to analyse the expression levels of genes of interest in cork oak or even in other oaks for which large transcriptomics and genomics programs are being developed.

Acknowledgements: Dr. Sónia Gonçalves (CEBAL, Portugal) is acknowledged for providing the cork tissue samples used in this work.

References

- Andersen CL, Jensen JL, Ørntoft TF.** 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets normalization of real-time quantitative reverse. *Cancer Research* **64**, 5245–5250.
- Artico S, Nardeli SM, Brilhante O, Grossi-de-Sa MF, Alves-Ferreira M.** 2010. Identification and evaluation of new reference genes in *Gossypium hirsutum* for accurate normalization of real-time quantitative RT-PCR data. *BMC Plant Biology* **10**, 49.
- 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*. *The Plant Cell* **19**, 351–368.
- Brunner AM, Yakovlev IA, Strauss SH.** 2004. Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biology* **4**, 14.
- Bustin SA, Benes V, Garson JA, Hellemans J, Huggett J, Kubista M, Mueller R, Nolan T, Pfaffl MW, Shipley GL, et al.** 2009. The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry* **55**, 611–622.
- Czechowski T, Stitt M, Altmann T, Udvardi MK.** 2005. Genome-wide identification and testing of superior reference genes for transcript normalization. *Plant Physiology* **139**, 5–17.
- Demidenko N V, Logacheva MD, Penin AA.** 2011. Selection and validation of reference genes for quantitative real-time PCR in buckwheat (*Fagopyrum esculentum*) based on transcriptome sequence data. *PLoS ONE* **6**, e19434.
- Die JV, Román B, Nadal S, González-Verdejo CI.** 2010. Evaluation of candidate reference genes for expression studies in *Pisum sativum* under different experimental conditions. *Planta* **232**, 145–153.
- Expósito-Rodríguez M, Borges AA, Borges-Pérez A, Pérez JA.** 2008. Selection of internal control genes for quantitative real-time RT-PCR studies during tomato development process. *BMC Plant Biology* **8**, 131.
- Farré D, Bellora N, Mularoni L, Messeguer X, Albà MM.** 2007. Housekeeping genes tend to show reduced upstream sequence conservation. *Genome Biology* **8**, R140.
- geNorm software manual.**

Gutierrez L, Mauriat M, Guénin S, Pelloux J, Lefebvre J-F, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C, et al. 2008. The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotechnology Journal* **6**, 609–618.

Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J. 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**, R19.

Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH. 2010. Identification and testing of superior reference genes for a starting pool of transcript normalization in *Arabidopsis*. *Plant & Cell Physiology* **51**, 1694–1706.

Huis R, Hawkins S, Neutelings G. 2010. Selection of reference genes for quantitative gene expression normalization in flax (*Linum usitatissimum* L.). *BMC Plant Biology* **10**, 71.

Jarosová J, Kundu JK. 2010. Validation of reference genes as internal control for studying viral infections in cereals by quantitative real-time RT-PCR. *BMC Plant Biology* **10**, 146.

Kakar K, Wandrey M, Czechowski T, Gaertner T, Scheible W-R, Stitt M, Torres-Jerez I, Xiao Y, Redman JC, Wu HC, et al. 2008. A community resource for high-throughput quantitative RT-PCR analysis of transcription factor gene expression in *Medicago truncatula*. *Plant Methods* **4**, 18.

Kamphuis W, Schneemann A, Beek LM Van, Smit AB, Hoyng PFJ, Koya E. 2001. Prostanoid receptor gene expression profile in human trabecular meshwork: a quantitative real-time PCR approach. *Investigative Ophthalmology & Visual Science* **42**, 3209–3215.

Kouadjo KE, Nishida Y, Cadrin-Girard JF, Yoshioka M, St-Amand J. 2007. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8**, 127.

Kuijk EW, Du Puy L, Van Tol HTA, Haagsman HP, Colenbrander B, Roelen BAJ. 2007. Validation of reference genes for quantitative RT-PCR studies in porcine oocytes and preimplantation embryos. *BMC Developmental Biology* **7**, 58.

Lawson MJ, Zhang L. 2008. Housekeeping and tissue-specific genes differ in simple sequence repeats in the 5'-UTR region. *Gene* **407**, 54–62.

Li Q-F, Sun SSM, Yuan D-Y, Yu H-X, Gu M-H, Liu Q-Q. 2010. Validation of candidate reference genes for the accurate normalization of real-time quantitative RT-PCR data in rice during seed development. *Plant Molecular Biology Reporter* **28**, 49–57.

- Lin YL, Lai ZX.** 2010. Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Science* **178**, 359–365.
- Maroufi A, Van Bockstaele E, De Loose M.** 2010. Validation of reference genes for gene expression analysis in chicory (*Cichorium intybus*) using quantitative real-time PCR. *BMC Molecular Biology* **11**, 15.
- Marum L, Miguel A, Ricardo PC, Miguel C.** 2011. Identification of GPAT acyltransferases in cork oak. *BMC Proceedings* **5**, P69.
- Migocka M, Papierniak A.** 2011. Identification of suitable reference genes for studying gene expression in cucumber plants subjected to abiotic stress and growth regulators. *Molecular Breeding* **28**, 343–357.
- Miguel A, Ricardo PC, Jones B, Miguel C.** 2011. Identification of a putative molecular regulator of cork cambium. *BMC Proceedings* **5**, P70.
- Nolan T, Hands RE, Bustin SA.** 2006. Quantification of mRNA using real-time RT-PCR. *Nature protocols* **1**, 1559–1582.
- Paiva JA, Fevereiro P, Marques P, Rodrigues J, Le Provost G, Plomion C, Grima-Pettenati J, Bouchez O, Klopp C, Berges H, et al.** 2011. Deciphering cork formation in *Quercus suber*. *BMC Proceedings* **5**, P172.
- Pfaffl MW.** 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**, e45.
- Pfaffl MW, Tichopad A, Prgomet C, Neuvians TP.** 2004. Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. *Biotechnology Letters* **26**, 509–515.
- Phillips MA, D'Auria JC, Luck K, Gershenzon J.** 2009. Evaluation of candidate reference genes for real-time quantitative PCR of plant samples using purified cDNA as template. *Plant Molecular Biology Reporter* **27**, 407–416.
- Ramakers C, Ruijter JM, Deprez RHL, Moorman AF.** 2003. Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data. *Neuroscience Letters* **339**, 62–66.
- Ramos M, Rocheta M, Carvalho L, Graça J, Morais-Cecilio L.** 2011. Expression analysis of DNA methyltransferase and co-repressor genes in *Quercus suber* phellogen: an attempt to correlate with cork quality. *BMC Proceedings* **5**, P169.

Reid KE, Olsson N, Schlosser J, Peng F, Lund ST. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology* **6**, 27.

Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A. 2008. Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. *Planta* **227**, 1343–1349.

Rozen S, Skaletsky H. 2000. Primer3 on the WWW for general users and for Biologist Programmers. *Methods in Molecular Biology* **132**, 365–386.

Schmidt GW, Delaney SK. 2010. Stable internal reference genes for normalization of real-time RT-PCR in tobacco (*Nicotiana tabacum*) during development and abiotic stress. *Molecular Genetics and Genomics* **283**, 233–241.

Silva SP, Sabino MA, Fernandes EM, Correlo VM, Boesel LF, Reis RL. 2005. Cork: properties, capabilities and applications. *International Materials Reviews* **50**, 345–365.

Silveira ED, Alves-Ferreira M, Guimarães LA, Da Silva FR, Carneiro VTDC. 2009. Selection of reference genes for quantitative real-time PCR expression studies in the apomictic and sexual grass *Brachiaria brizantha*. *BMC Plant Biology* **9**, 84.

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.

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.

Stothard P. 2000. The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* **28**, 1102–1104.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, RESEARCH0034.1–RESEARCH0034.11.

Vashisth T, Johnson LK, Malladi A. 2011. An efficient RNA isolation procedure and identification of reference genes for normalization of gene expression in blueberry. *Plant Cell Reports* **30**, 2167–2176.

Warrington JA, Nair A, Mahadevappa M, Tsyganskaya M. 2000. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. *Physiological Genomics* **2**, 143–147.

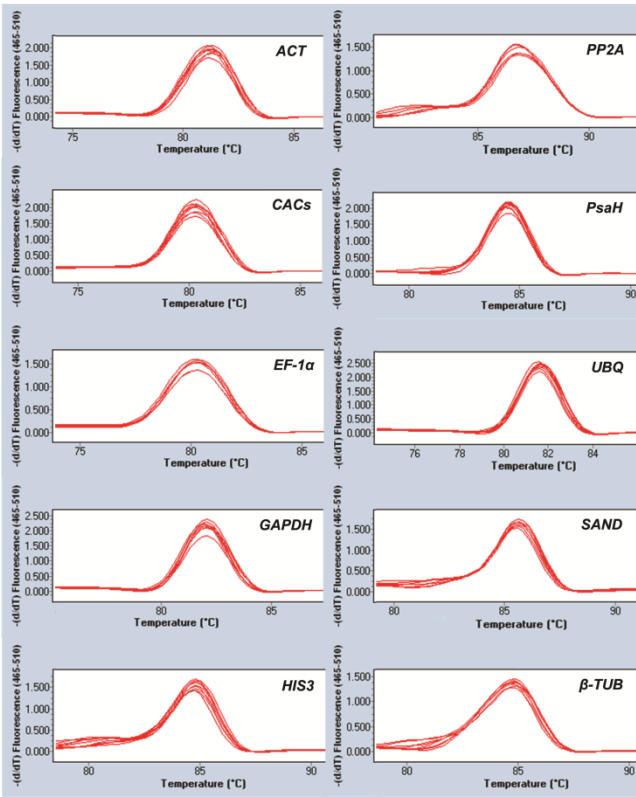
Xu M, Zhang B, Su X, Zhang S, Huang M. 2011. Reference gene selection for quantitative real-time polymerase chain reaction in *Populus*. *Analytical Biochemistry* **408**, 337–339.

Yang Y, Hou S, Cui G, Chen S, Wei J, Huang L. 2010. Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza*. *Molecular Biology Reports* **37**, 507–513.

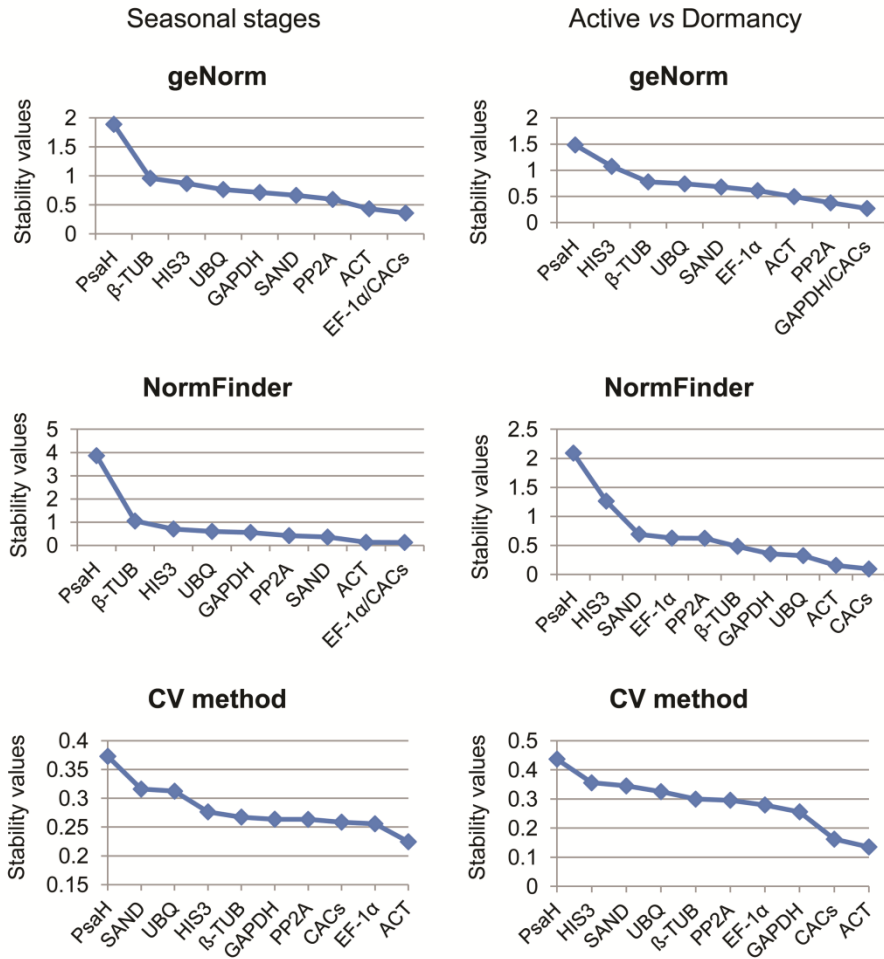
Zhang L, Li W-H. 2004. Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular Biology and Evolution* **21**, 236–239.

Zhao S, Fernald RD. 2005. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology* **12**, 1047–1064.

Supporting information



Supplementary Fig. S1. Melting curves generated for all amplicons.



Supplementary Fig. S2. Stability values of candidate reference genes calculated by different statistical methods using two data sets. The ranking of reference genes (*ACT*, *CACs*, *EF-1 α* , *GAPDH*, *HIS3*, *UBQ*, *PP2A*, *PsaH*, *SAND*, *β -TUB*) was calculated by geNorm, NormFinder and CV method using the developmental stage and seasonal growth data sets.

Supplementary Table S1. Amplicon sequences of the 8 candidate reference genes (RGs).

Gene abbreviation	Amplicon sequences
CACs	TCTGGGAGAAGAGTGGCTACAATACTGTTGAGTGGGTCCGTTATATT ACCAAAGCTGGCTCTTATGAGATTAGGTGCTAGAGACTGGAAATGTT GCTGAGACTATAGATGCCTTAAAATGGGTGGCGATTGAAATATTTGT ACGTCTATTTTTCTAAGGATTGAAATGGTGGCTC
EF-1α	TTGTGCCGTCCTCATTATTGACTCCACCACTGGAGGTTTTGAAGCTG GTATTTCTAAGGATGGTCAGACCCGTGA
GAPDH	ACCGACTTCATTGGTGACAGCAGGTCTAGTATATTTGATGCCAAGGC TGGAATTGCATTGAATGACAATTTTGTGAACTTGTCTCTTGGTATGA CAACGAGTGGGGCTACAGTTCCTCGTGTGTTGACCTGATTGTCCAC ATCGCATCT
HIS3	GCTCTTCGAGGACACCAATCTGTGCGCCATTACGCCAAGAGAGTC ACCATCATGCCCAAGGATATCCAGCTCGCTCGGAGGATCAGAGGCG AGAGGGCTTA
PsaH	CAGTTGCTCTGAAACCAAGGCCATGGCTTCTCTAGCAACCTTAGCTG CTGTTCAACCAGTCAACATCAAGGGCCTTGGTGAAGCTCCCTAACA GGAACAAAGCTTGCTATCAAGCCCACTCGCCAGAGCCTAAGGTCCA AAAACCTCAGGACTGGTGCTGTG
PP2A	GAGCCACTCTATCCGATTGCTGTCTTAATTGATGAGCTTAAAAATGAA GATATTCAGCTCCGGCTGAACTCGATCCGCCGGCTCTCTACGATTGC GCGTGCGCTTGAGAGGAGAGGACCAGGAAGGAGCTGATTCTTTT CTCAGCGAGAACAATGACGAC
SAND	AGGATTGCAGGATTTCGATTGAAAGTGGTCCTTTTGAAGTCAAATGTT CTTAGCGAAGTTCAGAGATCCATGCTAGATGGAGGGATGCATGTTGA AGATTTGCCTACCGATCCATTACCTCGTTCTGGAACCTTATCTCCACA TCTGGGSSAACCCAGAGATTCTCTTGAGAGTCTCAAAGAACCATTG TTGGCATTGGTGGTC
UBQ	CGAAGATCCAGGACAAGGAGGGGATCCACCGGACCAGCAGAGGT TGATCTTTGCAGGAAAGCAGCTGGAGGATGGCCGCACTCTTGCTGA CTACAACATCCAGAAGGAGTCCACCCTTCACCTTGTCTCCGTCTCC GCGGTGGTGCTTTCTGAGGAATGAAAAAGCCCTG

Supplementary Table S2. Amplification efficiencies of the 10 candidate RGs measured using the standard curve method.

RG/ Tissue	<i>ACT</i>	<i>CACs</i>	<i>EF-1α</i>	<i>GAPDH</i>	<i>HIS3</i>	<i>PP2A</i>	<i>PsaH</i>	<i>UBQ</i>	<i>SAND</i>	<i>β-TUB</i>
Leaves	1.85	2.02	1.95	1.90	1.99	1.89	1.95	1.92	1.96	1.94

Chapter IV

SHORT-ROOT-like genes are differentially regulated during secondary growth in Quercus suber



Miguel A, Marum L, Miguel CM. *SHORT-ROOT-like genes are differentially regulated during secondary growth in Quercus suber. (submitted)*

Andreia Miguel participated in the experimental design, performing laboratory experiments, analyzing the results and writing the paper.

SHORT-ROOT-like genes are differentially regulated during secondary growth in *Quercus suber*

Abstract

The cork oak (*Quercus suber* L.) tree exhibits a rare ability to produce a thick phellem layer, also known as cork, with properties suited for industrial applications. Phellem is formed during secondary growth as a result of the activity of the phellogen (cork cambium). Here we report the identification and characterization of two putative homologs of the *Arabidopsis* and *Populus* *SHR* genes in cork oak (*QsSHR1* and *QsSHR2*) and in the related species holm oak (*Quercus ilex*) lacking the cork-producing ability, indicating their involvement in the functioning of the lateral meristems during secondary growth. Seasonal variation in the expression of *SHR1* and *SHR2* in the periderm of branches at different ages differed in the two species, but it was not possible to correlate *SHR* expression to cork production in the cork oak tree. While *SHR1* was highly expressed in one year-old periderm harvested in June and July, *SHR2* was more expressed in January, during the phellogen dormancy period. Moreover, while *QsSHR1* expression was localized in the vascular cambium, *QsSHR2* transcripts appeared prevalent in the phellogen. Functional studies in the *Arabidopsis shr2* mutant revealed that none of these gene functions fully recovers the wild-type phenotype, when tested separately. These data suggest a role for *QsSHR2* in regulating periderm development, possibly through a negative control of phellogen activity. Genome duplication events after separation of the *Arabidopsis* and cork oak lineages together with the failure in the *shr2* mutant complementation, indicates a putative functional divergence between the *Arabidopsis* and cork oak *SHR* proteins.

Keywords

Cork oak, lateral meristem, periderm, phellogen, *SHORT-ROOT* (*SHR*), vascular cambium

Introduction

The cork oak tree (*Quercus suber* L.) is an evergreen and long living species of the Fagaceae family. It is distributed along the western Mediterranean region and North Africa (Aronson et al., 2009; Bugalho *et al.*, 2011) where it has a great socio-economic and ecological impact. Cork oak has the rare ability of producing cork highly suited for industrial applications. Cork is a protective multilayer of suberized dead cells that in cork oak can achieve an appreciable thickness, and has unique properties that derive from its cellular structure, chemical composition and mechanical and physical properties (Silva *et al.*, 2005). Cork is formed during secondary growth as a result of the activity of the cork cambium or phellogen. The phellogen is a lateral meristem that, like the vascular cambium, is formed by a continuous cylinder of meristematic cells around the plant stem and roots. However, unlike the vascular cambium that differentiates into secondary xylem (wood) and phloem, the phellogen differentiates into phelloderm to the inside and phellem (cork) to the outside of the stem (Evert, 2006). Together, the phellogen, phelloderm and phellem form the periderm that replaces the epidermis in woody species and has an important role in the protection from biotic and abiotic stresses. Phellogen can derive from cells of the epidermis, subepidermis or the phloem/cortex (Waisel and Liphschitz 1975; Evert, 2006), and in cork oak it is described to develop under the epidermis (Graça and Pereira 2004). The removal of the cork layer from the cork oak tree (debarking) destroys the phellogen, but a traumatic phellogen differentiates in the inner bark (Pereira *et al.*, 1992) restoring its activity after approximately 25-35 days (Machado, 1935) and allowing sustained cork exploitation. The activity of the phellogen is seasonal and occurs in the period from April to October (Costa *et al.*, 2002; Costa *et al.*, 2003; Silva *et al.*, 2005). Usually, the debarking process happens when phellogen reaches its maximum activity, in June and July (Costa *et al.*, 2003), allowing an easy

separation of the cork layer from the tree. Other *Quercus* species, such as *Quercus cerris* and *Quercus variabilis*, also produce substantial amounts of cork tissues but the first does not form a continuous cork layer around the trunk (Şen *et al.*, 2011a) and both have differences in cork cellular structure and chemical composition contributing to cork of lower industrial quality (Şen *et al.* 2010, 2011a, 2011b; Miranda *et al.* 2012). On the other hand, *Quercus ilex* (holm oak), a species that coexists with cork oak in the same environments, does not produce exploitable cork but only a thin periderm.

The influence of environmental factors on phellogen activity and on radial growth has been addressed by several authors (Pereira *et al.*, 1992; Caritat *et al.*, 1996, 2000; Fialho *et al.*, 2001; Costa *et al.*, 2002; Costa *et al.*, 2003). Studies on periderm formation were also performed using the potato tuber periderm as experimental system (Ginzberg *et al.*, 2009; Chaves *et al.*, 2009; Soler *et al.*, 2011; Neubauer *et al.*, 2012), paying special attention to heat stress responses and wounding, and searching for candidate genes for suberin biosynthesis, involved in defense and periderm/potato skin regulation. Despite the importance of cork for industrial applications (Silva *et al.*, 2005; Pereira, 2007; Mestre and Gil, 2011), studies on the molecular regulation of cork formation and phellogen activity are scarce. A genomic approach to cork formation has been reported (Soler *et al.*, 2007), but it mostly focused on suberin biosynthesis and differential gene expression analyses were based on the comparison between wood and phellem tissues. The first molecular characterization study of one of those candidate genes, *QsMYB1*, was made recently (Almeida *et al.*, 2013). A histological study of the development of the first cork oak periderm (Graça and Pereira, 2004), a proteomic study of cork formation (Ricardo *et al.*, 2011) and, more recently, a relationship between cork quality and DNA methylation in active phellogen differentiation tissue (Ramos *et al.*, 2013) have also been reported.

In the *Arabidopsis* root, *SHR* is involved in root patterning, cell division and specification/differentiation (Benfey *et al.*, 1993; Helariutta *et al.*, 2000; Nakajima *et al.*, 2001; Sabatini *et al.*, 2003), maintenance of the quiescent center (QC) and stem cell niche (Benfey *et al.*, 1993; Helariutta *et al.*, 2000; Sabatini *et al.*, 2003) and also as a regulator of cell division in leaves (Dhondt *et al.*, 2010). Studies conducted in *Populus* showed that *PtSHR1*, described as an ortholog of the *Arabidopsis SHR* (*AtSHR*) gene (Wang *et al.*, 2011), and the *SHR*-like homolog *PtSHR2B*, probably derived through genome duplication events after divergence of *Populus* and *Arabidopsis* lineages, may be involved in the regulation of different plant meristems. In fact, *PtSHR1* is expressed in the root stele and vascular cambium of poplar stems being involved in the control of primary and secondary growth (Wang *et al.*, 2011), whereas *PtSHR2B* is expressed in the root tip and in the phellogen (Miguel *et al.* submitted).

In this study we focused on the role of cork oak *SHR*-like genes as potential regulators of the lateral meristems, giving special attention to the regulation of phellogen. We have identified two *SHR*-like genes in cork oak transcriptome and in a related oak species, holm oak, lacking the ability of extensive cork production. Expression patterns of both cork oak genes, *QsSHR1* and *QsSHR2*, in different tissues were examined and a comparative analysis of *SHR* transcript levels in the periderm from cork oak and holm oak branches of increasing age was performed. We found that the cork oak *SHR*-like genes were expressed in the lateral meristems and neither of the two gene functions fully complemented the *Arabidopsis shr2* mutant phenotype. A possible role for *QsSHR2* gene during phellogen development is discussed.

Materials and Methods

Plant material

Cork oak and holm oak fully expanded young leaves were collected in June. One to three year-old branches were harvested from a single tree at Tapada da Ajuda (Instituto Superior de Agronomia, Lisboa, Portugal) during the growth (April, June and July) and the dormancy (October and January) periods in 2010. Reproduction cork, the cork harvested from the second stripping onwards (Silva *et al.*, 2005), was harvested in July 2009 and July 2010 from trees growing in Coruche and São Brás de Alportel (Portugal), and newly formed phellem tissue was collected from the inner part of the cork planks with sterile scalpels. Developing and mature acorns were collected from different cork oak trees at Quinta da Serra, Azeitão (Portugal). Cork oak seedlings were also used to collect hypocotyls, cotyledons and radicles at different days after germination (DAG), main roots and lateral roots, young leaves and stems. Periderm from branches was isolated by peeling off the external bark with a sterile scalpel. Embryos were carefully isolated under a stereomicroscope and for the mature acorns the pericarp was removed and only the half acorn near the embryo was used. Samples were immediately frozen in liquid nitrogen and stored at -80° C until further use. For *in situ* hybridization one year-old branches were collected from six year-old potted plants.

Arabidopsis shr2 mutant and wild-type Columbia (Col-0) were used in functional studies. *Arabidopsis* seeds were surface sterilized, germinated on MS medium and transferred to soil after 1 month. Plants were grown in a chamber with 16 h light / 8 h dark photoperiod at 22° C / 18° C, respectively.

RNA isolation and cDNA synthesis

Cork oak and holm oak tissues were ground in liquid nitrogen using a mortar and pestle but frozen acorns and cotyledons were first ground in a mill (M 20 Universal mill, Ika). Total RNA was extracted using the protocol developed by Reid *et al.* (2006) with the modifications described previously (Marum *et al.*, 2012). All samples were treated with TURBO DNase (Ambion), according to manufacturer's instructions, to remove any genomic DNA contamination. RNA integrity was evaluated by 1% (w/v) agarose gel electrophoresis and some RNA samples were analysed in the BioAnalyzer 2100 Agilent. Total RNA amount was determined using a ND-1000 Spectrophotometer (NanoDrop Technologies Inc.) and only RNA samples with $A_{260/280} > 1.8$ were used for further experiments. Reverse Transcription Polymerase Chain Reaction was performed using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) with the anchored-oligo(dT)₁₈ primers, according to the manufacturer's instructions, and cDNA was synthesized from 1.5 µg of total RNA. Products from different cDNA synthesis reactions generated from the same RNA sample were joined together.

Gene cloning

A search in the public databases NCBI (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and JGI-*Populus trichocarpa* v1.0 (<http://genome.jgi-psf.org/Poptr1/Poptr1.home.html>) for the *Arabidopsis SHR* homologs, retrieved several sequences that were aligned and used for designing degenerate primers (5'-ATGGATACMTTGTTTAGRCTAGTYAG-3' and 5'-ACACCACCGGYTGKTCYTTCCA-3') to clone partial fragments of the putative cork oak and holm oak *SHR1* genes, *QsSHR1* and *QiSHR1*, respectively. PCR products obtained from cork oak and holm oak samples were cloned into pCR2.1 – TOPO and pCRII vector (TA Cloning Kit – Invitrogen), respectively, and sequenced.

The complete *QsSHR1* as well as the other cork oak *SHR* homologous sequence, *QsSHR2*, were identified in EST libraries generated from cork oak embryos and fruits at different developmental stages (<http://www.corkoakdb.org/>). The partial coding sequence of the holm oak *QiSHR2* was obtained by amplification with the primers 5'-ggggacaagtttgtaaaaaagcaggcttcACCCACATGGACATAACTCTTTTC-3' and 5'-ggggaccactttgtacaagaaagctgggtcTTATGGTTTCCATGCTGAAGCCC-3', which were designed for cloning *QsSHR2* through the GATEWAY cloning technology (see below).

Phylogenetic analysis

Phylogenetic analysis was performed using the phylogeny platform phylogeny.fr (Dereeper *et al.*, 2008). Full-length *SHR* protein sequences were used for the construction of the phylogenetic tree, except for *Quercus ilex* (Qi_KF692544 and Qi_KF692545) and one of the *Citrus sinensis* (Csi_orange1.1g047247m) sequences which were partial. Protein sequences from species with a sequenced genome were retrieved from phytozome v9.1 (Goodstein *et al.*, 2011) and through a GenBank Blast search at NCBI (<http://www.ncbi.nlm.nih.gov/>). Sequences were aligned using Muscle v3.7 (Edgar, 2004) and poorly aligned positions and divergent regions were removed with Gblocks v0.91b (Castresana, 2000) to have a more stringent selection. The following default settings were used for Gblocks: ten as a minimum length of a block after gap cleaning; no gap positions were allowed in the final alignment; all segments with contiguous non conserved positions bigger than eight were rejected; 85% was established as the minimum number of sequences for a flank position. The phylogenetic analysis was performed using the PhyML method.

Semi-quantitative RT-PCR

A semi-quantitative RT-PCR was performed to evaluate the expression of both *QsSHR1* and *QsSHR2* in different cork oak seedling tissues, embryos and fruits. Gene specific primers were designed in the N-terminal region of both cork oak *SHR* genes, and were used to amplify fragments of 178 bp (*QsSHR1_5`F*-CAGCTCTAGATCCTCCAGAC-3`; *QsSHR1_5`R*-GGTGGTGGGAGTGAGAGTAG-3`) and 196 bp (*QsSHR2_5`F*- TGGACATAACTCTTTTCACTCCG-3`; *QsSHR2_5`R*-GCACACTCCTTGAGAAGTCTT-3`). The PCR mixtures for a total volume of 20 µl contained 200 µM dNTPs, 0.2 µM of each primer, 1 µl of cDNA, 1× GoTaq Buffer, 2.5 mM MgCl₂ and 0.05 U/µl of GoTaq (Promega). The amplification program included an initial denaturation at 95° C for 2 min, followed by 34 cycles of 50 s at 95° C, 45 s at 60° C and 30 s at 72° C and a final extension step of 5 min at 72° C. Amplified products were obtained in incremental PCR cycles (cycle 24, 27 and 35) and *ACTIN* (*ACT*) was used as reference gene.

Quantitative Real-Time PCR

The reverse transcription quantitative PCR (RT-qPCR) was performed in 96-well plates with a LightCycler 480 (Roche) using SYBR Green I Master (Roche). Reaction mixtures contained 10 µl of 2x SYBR Green I Master, 400 nM of each primer and 1.5 µl of cDNA as template, in a final volume of 20 µl. The amplification program was the same for the two cork oak and the holm oak *SHR* genes: 95° C for 10 min, 45 cycles of 10 s at 95° C, 20 s at 60° C and 10 s at 72° C. The gene specific primers were those used for the semi-quantitative assay. To confirm the presence of a single amplicon, a dissociation curve was analysed in all amplifications and no-template controls were also performed for each primer pair. Three biological replicates were analysed and for each sample three technical replicates

were performed. To normalize values obtained in different plates, cDNA from leaves was used as calibrator in each plate.

The amplification efficiency of each primer pair in each tissue was determined through the statistical algorithm Real-Time PCR Miner (Zhao and Fernald, 2005) using the raw fluorescence data as an input and a mean efficiency value was determined for each biological replicate. Normalized relative quantities were calculated using an improvement from the classic delta-delta-Ct method adapted for expression measurements with multiple

reference genes. The formula $NRQ = \frac{E_{goi}^{\Delta Ct, goi}}{f \sqrt{\prod_0^f E_{ref,0}^{\Delta Ct, ref, 0}}}$ was used, where E is the efficiency of the amplification for each primer pair in each tissue, goi and ref are the gene of interest and the reference gene respectively, ΔCt is the Ct of the calibrator minus the Ct value of the sample in test and f the number of reference genes used to normalize the data (Pfaffl, 2001; Livak and Schmittgen, 2001; Hellemans *et al.*, 2007). According to the stability analysis of reference genes in different cork oak tissues (Marum *et al.*, 2012) *ACTIN* (*ACT*) and *CLATHRIN ADAPTOR COMPLEX* (*CAC*) were used as reference genes in the geometric mean. The calibrator sample used to determine the expression patterns of *QsSHR1/2* and *QiSHR1/2* was the periderm from three year-old branches collected in July.

In situ Hybridization

One year old-branches were collected and 100 μ m cross sections were obtained with a vibratome (Series 1000 Plus Sectioning System, Vibratome). The sectioned tissue was placed in ice-cold FAA (10% formaldehyde, 5% acetic acid, 50% ethanol) in a desiccator for about 15 min. The samples were then placed in a fresh FAA solution for approximately 4 h at 4° C. The fixative was replaced and samples washed 2 times with 30% ethanol for 5 min and then twice with water. Before the

hybridization step the sections were submitted to a protease and acetylation treatment (Jackson, 1991; Karlgren *et al.*, 2009), followed by 2 washes in water and once with 50% formamide at room temperature, prior to the hybridization step.

To generate *QsSHR1* and *QsSHR2* specific probes, one fragment of 350 bp amplified with the primers *QsSHR1_5'*F-CTCTGGAGGAGAGC TTTTC-3' and *QsSHR1_5'*R-AGCGTGTAAGAAGGTACGCACC-3', and another fragment of 358 bp amplified with the primers *QsSHR2_5'*F-GAGTCAAAGGGGCTCCATAACCC-3' and *QsSHR2_5'*R-CCGTGGCCTT ACAAATAGAGCTTG-3', were cloned into pCRII vector (Invitrogen) for the synthesis of both sense and antisense RNA probes using the DIG RNA Labelling Kit SP6/T7 (Roche), according to the manufacturer's instructions. Probes in 50% formamide were denatured by heating at 80° C for 2 min followed by cooling on ice. For hybridization a ratio of 4:1 volumes of hybridization solution:probe (Jackson, 1991; Karlgren *et al.*, 2009) were added corresponding to a final concentration of 0.4 ng probe/μl of hybridization solution (Harding *et al.*, 2011). All hybridizations were performed overnight (approximately 14 h) at 55° C in PCR tubes with a total volume of 200 μl. The following steps were performed as previously described (Groover *et al.*, 2006). Colour development, using 5 mg/ml NBT and 1.875 mg/ml BCIP, was performed monitoring the time needed for the detection of the alkaline phosphatase activity along several hours.

Sections were then fixed again in FAA (5% formaldehyde, 5% acetic acid, 50% ethanol) and vacuum infiltrated during at least two periods of 30 min each. Samples were left in fixative overnight and afterwards dehydrated in an ethanol series (50%, 70%, 80%, 90% and two times 100%) (Takechi *et al.*, 1999). Tissues were then embedded in Technovit 7100 (Heraeus Kulzer), according to manufacturer's instructions with some modifications. Samples were placed in pre-infiltration solution and vacuum infiltrated. In the next day the pre-infiltration solution was renewed and left for

approximately one week at 4° C. Infiltration was performed during one, two or three weeks at 4° C followed by polymerization.

Plasmid construction and Arabidopsis transformation

To produce transgenic plants carrying *AtSHRprom:QsSHR1* or *AtSHRprom:QsSHR2*, approximately 2.2 kb upstream the ATG start codon of the *Arabidopsis SHR* was amplified by PCR and cloned into pCRII (Invitrogen). The primers used for promoter isolation include the enzyme restriction site followed by the promoter sequence (pSHR_5'F-AAGCTTGGAGAGTTATGTAATGTAGG-3' and pSHR_5'R-ACTAGTAATGAATAAGAAAATGAATAGAAG-3', this last one described by Yu *et al.* (2010)). The destination vector, pK7WG2.0 (Karimi *et al.*, 2002), was digested with HindIII and SpeI (New England Biolabs) at 37° C for 3 h and then ligated to the promoter using T4 DNA Ligase from pCRII vector TA Cloning Kit (Invitrogen) in a 1:3 ratio of vector:insert, overnight at 14° C. The coding regions of *QsSHR1* and *QsSHR2* were PCR amplified with primers including attB regions to be suitable for further steps (attBQsSHR1_5'F-ggggacaagttgtacaaaaaagcaggcttcATGGATACCTTGTTTAGGC-3'; attBQsSHR1_5'R-ggggaccactttgtacaagaaagctgggtcTCAAGGCCTCCATGC ACTG-3'; attBQsSHR2_5'F-ggggacaagttgtacaaaaaagcaggcttcACCCACAT GGACATAACTCTTTTC-3' and attBQsSHR2_5'R-ggggaccactttgtacaagaaagctgggtcTTATGGTTTCCATGCTGAAGCCC-3'), cloned into pCRII and sequenced. GATEWAY technology (Invitrogen) was then used to clone *QsSHR1* and *QsSHR2* coding sequences into the entry vector pDONR221, and then into the destination vector pK7WG2.0 already containing the *Arabidopsis SHR* promoter. These plasmid constructs were used separately for transformation of *Arabidopsis shr2* mutant plants by the floral-dip method (Clough and Bent, 1998).

Statistical analysis

To assess for significant differences in gene expression we employed a Factorial Anova followed by the Tukey HSD Post-hoc test. The non-parametric Mann-Whitney U-test was carried out for phenotypic parameters. A significance level of $p=0.05$ was used. Statistics were performed using the Statistica (StatSoft Inc., <http://www.statsoft.com>) software package.

Results

Two SHR-like genes identified in cork oak and holm oak

A comparative sequence analysis was done to identify putative homologs of the *Arabidopsis SHR* gene in the cork oak transcriptome. A *SHR*-like sequence with 1432 bp, named as *QsSHR1*, was amplified from cork oak cDNA using degenerate primers designed after multiple alignment of the *SHR* deduced amino acid sequences from *Arabidopsis thaliana*, *Populus trichocarpa*, *Vitis vinifera* and *Medicago truncatula* (Fig. 1).

SHR-like genes during secondary growth in *Quercus*

PtSHR1	MDTLFRLVSLQQQ---SEQSFNST---SRTSSSSSRSSRQNNHHHHYQOEDEECFNFFM
VviSHR1	MDTLFRLVSLQ---SDQSYNS---SRTSSSSSRSSRQ---HHHHYQOEDEECFNFFM
MtrSHR1	MDTLFRLVSLFQPHQQQEDQSLNSTTSRTTSSSSSRSSQON-YHQYYNPFQVEVEECFNFFM
AtSHR	MDTLFRLVSLQQQQQ-SDSIITNQSSLRSTSTTTTGSPQTAYHYNFPQNDVVEECFNFFM
PtSHR1	DEEDFSSSSS-----KHYYPPYHHNQQQHQQHTTTTTPTTTTNTSTPSTHHVL
VviSHR1	DEEDFSSSSS-----KHYYP-----YQQQTHPSTTTTP--TTTNTSTP--THHAF
MtrSHR1	DEEDFSSSSS-----KHYYP-----YNQPQTSNTITNTPTTSLTPPF-----
AtSHR	DEEDFSSSSSHNNHHNNHNNPNTYYSPTTPTQYHPATSTSPSTAAAAALASPYSSSGHH
PtSHR1	DSADFSFSPSHDLNFEF---SGKWVTDILLESAAHAIADKNSARLQQLMWMLNELGSPYG
VviSHR1	EPTEFSFSPARDLHLEFNSSASGKWASEILLETARAIADKNSARVQQLMWMLNELSSPYG
MtrSHR1	-----PTADFTEFEL---TCRWANNILLETARAIATKNSTRLOQLMWMLNELSSPYG
AtSHR	NDPSAFSIPQTPPSFDFS--ANAKWADSVLLEAARAFSDKDTARAQQLLWTLNELSSPYG
PtSHR1	DTEQKLASYFLQALFSRMNDSGERCYRTLAS--ASEKTCSEFSTRKMVLKFQEVSPWTFE
VviSHR1	DTDQKLAAAYFLQALFSRMTDSGERCYRTLIS--ASEKTCSEFSTRKMVLKFQEVSPWTFE
MtrSHR1	DIEQKLAAAYFLQALFSRMTEACTRTERTLAS--ASEKTCSEFSTRKMVLKFQEVSPWTFE
AtSHR	DTEQKLASYFLQALENRMTGSGERCYRTMVIAAAATEKTCSEFSTRKMVLKFQEVSPWATF
PtSHR1	GHVSCNGAIMEAFEGESKLHIIDISNTYCTQWPTLLEALATRTDEPHLKLTTTVASKSS
VviSHR1	GHVACNGAIMEALEGESKLHIIDISNTYCTQWPTLLEALATRTDEPHLRLTTTVTSKAG
MtrSHR1	GHVACNGAILEAFEGESKLHIIDISNTYCTQWPTLFEALATRADTTPHLRLTTTVTAGGS
AtSHR	GHVAANGAILEAVDGEAKIHIIDISSTFCTQWPTLLEALATRSDDTTPHLRLTTTVVANKF
PtSHR1	GNNIGLTS TGGLASVHKVMKEIGNRMEKFARLMGVPFKFENVIIHAGDLCDLNLAELDVKD
VviSHR1	-----TGGMAPVQKLMKEIGNRMEKFARLMGVPFKFENVLHSGDLSHLNLAELDTKD
MtrSHR1	-----VQKVMKEIGARMEKFARLMGVPFKFENVIIHSGDLSDLNFLDLDTKE
AtSHR	VN-----DQTASHRMMEKEIGNRMEKFARLMGVPFKFENIIHHVGDLSDFDLNELDVKP
PtSHR1	DEALAINCVGALHSITPAS-----RRRDYVISSFRTHQPRIITVVEEADL----DG--
VviSHR1	DEALAINCVGALHSVTAFG-----NRRDITVSSFRRLHPRIITVVEEADLDVGVGDG--
MtrSHR1	DEALAINCVNALHSVTVCGNGNGNRRDSLTASLIALRPRIVTVVEEADLNFENEG--
AtSHR	DEVLAINCVGAMHGTASRG-----SPRDAVISSFRRLRPRIVTVVEEADLVGEEGGF
PtSHR1	-LDFVKGFQECRLWFRVYFESLDESFPRTSNEQLMLERAAGRAIVDLVACPPSDSIERRE
VviSHR1	-FDFVKGFQECRLWFRVYLESLDESFPRTSNERLMLERAAGRAIVDLVACPPSESIERRE
MtrSHR1	-YEFVNGFECLRWFRVYFEALESFPKTSNERLMLERBAGRGIVDLVACAPAESIERRE
AtSHR	DDEFLLRGFGECLRWFRVCFESWESFPRTSNERLMLERAAGRAIVDLVACEPSDSTERRE
PtSHR1	TATRWSGRLHSCGFSPFIIFSDEVCDDVRALLRRYKEG-WSMTQCGD-AGIFLCWKQPQVV
VviSHR1	TATRWGQRLHASGFSPVVSFDEVCDDVRALLRRYKEG-WSMTQSSD-AGIFLSWKQPQVV
MtrSHR1	TAVRWSRRLHGRGENTVAFSEFVCDDVRALLRRYKEG-WSMIRCND-AGIFLTWKQPQVV
AtSHR	TARKWSRRMRNSGFGAVGYSDEVADDVRALLRRYKEGVWSMVOCPEAGIFLCWRDQPQVV
PtSHR1	WASAWRE-
VviSHR1	WTSAWKP-
MtrSHR1	WASAWRE-
AtSHR	WASAWRET



Fig. 1. Multiple alignment of SHR deduced amino acid sequences from *Populus trichocarpa* (PtSHR1: XP_002332364.1), *Vitis vinifera* (VviSHR1: XP_002267068.1), *Medicago truncatula* (MtrSHR1: XP_003611578.1) and the *Arabidopsis thaliana* SHR protein (AtSHR: NP_195480.1) using ClustalW2 (Larkin *et al.*, 2007) for the alignment and Jalview (Waterhouse *et al.*, 2009) for graphic representation. Regions of identity are represented with different shades (the darker the colour the higher the degree of identity of the amino acid residues between species). Underlined amino acids show the regions from which the degenerate primers were designed.

The full coding sequences of the two putative cork oak *SHR*-like genes, *QsSHR1* (JN819303) and *QsSHR2* (KF692546), were subsequently obtained from expressed sequence tag (EST) data generated from embryos and fruits, available in the CorkOakDB (<http://www.corkoakdb.org/>). Sequence analysis revealed a complete *QsSHR1* coding sequence of 1455 bp and a *QsSHR2* coding sequence of 1305 bp. Like in *Arabidopsis* (Helariutta *et al.*, 2000) and *Populus* (Goodstein *et al.*, 2011), both cork oak *SHR* genes lack introns (data not shown). The putative amino acid sequence of *QsSHR1* showed 82% identity to a predicted *Vitis vinifera* SHR protein (XP_002267068.1), and 77% and 74% to the putative SHR sequences from *Eucalyptus grandis* (BAH80549.1) and *Populus trichocarpa* (XP_002332364.1), respectively. *QsSHR2* displayed 86% identity with two poplar SHR-like sequences (XP_002327997.1 and XP_002309782.1) and 82% identity to a putative SHR sequence from *Vitis vinifera* (XP_002272196.1).

The alignment of the two predicted *QsSHR* amino acid sequences revealed characteristic domains of the conserved GRAS protein family (Fig. 2), the plant specific transcription factor family to which SHR proteins belong.

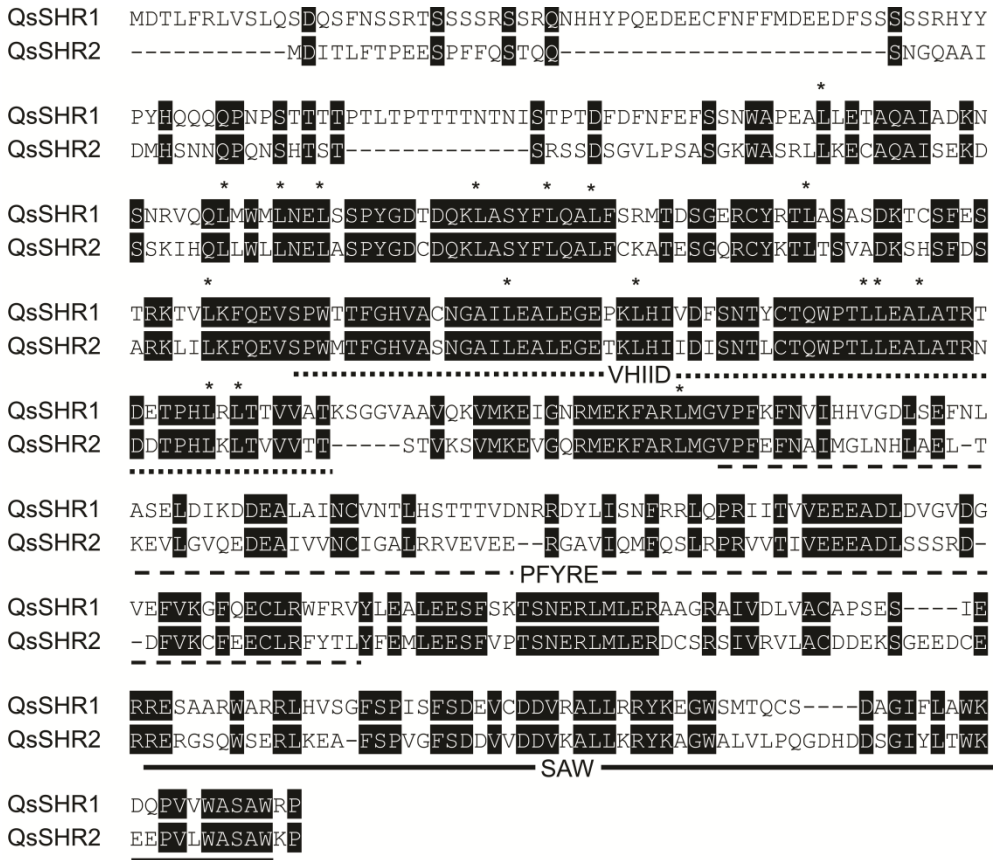


Fig. 2. Alignment of the deduced amino acid sequences of the two cork oak SHR-like sequences, QsSHR1 (AET37154) and QsSHR2 (KF692546), using ClustalW2 (Larkin *et al.*, 2007) and Jalview (Waterhouse *et al.*, 2009), and highlighting the functional domains. Regions of identity are highlighted in black. GRAS conserved domains in the carboxyl-terminal region are underlined and asterisks indicate leucine rich regions.

Like the GRAS proteins (Bolle 2004, Gallagher and Benfey 2010), cork oak SHR, QsSHR1 and QsSHR2, also showed significant homology in the C-terminal domains whereas the N-terminal region is variable. Five different motifs can be distinguished in the GRAS sequences by the following order: leucine heptad repeat I (LHRI), VHIID motif, leucine heptad repeat II (LHRII), PFYRE motif and the SAW motif (Pysh *et al.*, 1999; Bolle,

2004). We were able to distinguish two leucine-rich regions which flank the VHIID motif (Fig. 2), however as in most cases (Bolle, 2004), they do not seem to occur as heptad repeats. In the VHIID motif the characteristic P-N-H-D-Q-L residues were present (Fig. 2) and the spacing between them is the same as in other GRAS proteins (Pysh *et al.*, 1999). However, according to Bolle, among the GRAS family members only the histidine and aspartic acid are considered absolutely conserved (Bolle, 2004). The PFYRE motif, which is less conserved at the sequence level compared to the VHIID and the SAW motifs, is present but not in the same order (Fig. 2). At the C-terminal the SAW motif can also be recognized, wherein three pairs of conserved residues, R-E, W(X)₇G and W(X)₁₀W (Pysh *et al.*, 1999; Bolle, 2004), were present.

Previous reports showed that multiple regions within the GRAS domains are required for movement, localization, stability and activity of the SHR protein (Gallagher *et al.*, 2004; Gallagher and Benfey, 2010). In *Arabidopsis*, movement of SHR was found to require both nuclear and cytoplasmic localization and a point mutation of a threonine amino acid at the VHIID domain was found to disrupt SHR movement (Gallagher *et al.*, 2004; Gallagher and Benfey 2010). The conserved threonine at the VHIID domain was also present in the two cork oak SHR-like proteins but other GRAS domains previously identified as needed for proper SHR movement (Gallagher and Benfey, 2010) were not completely conserved in the cork oak proteins.

We have also identified two putative *SHR*-like sequences in *Quercus ilex* (holm oak), a cork oak related species. *QiSHR1* (KF692544) was retrieved by amplification with the degenerate primers designed to clone *QsSHR1*. A fragment of 1432 bp was amplified, cloned into the pCRII vector (Invitrogen) and sequenced. The alignment of the deduced amino acid sequence showed 98.74% identity to *QsSHR1* (AET37154) (Supplementary Fig. S1) followed by 82% and 80% identity with two *SHR*-like isoforms in

Vitis vinifera (XP_002267068.1 and XP_003632486.1). The other putative SHR sequence, QiSHR2 (KF692545), was obtained with specific primers designed to clone the full-length QsSHR2 gene. A fragment of 1206 bp was obtained and showed 99.25% identity to the QsSHR2 gene (KF692546) (Supplementary Fig. S2). This sequence also showed 86% identity with two poplar SHR-like sequences (XP_002327997.1 and XP_002309782.1), 85% to a GRAS family transcription factor from *Prunus persica* (EMJ22811.1) and 84% to a putative SHR sequence from *Vitis vinifera* (XP_002272196.1).

Phylogenetic analysis groups the two cork oak SHR-like proteins in different clusters

A phylogenetic analysis was performed to search for evolutionary relationships between cork oak SHR-like proteins and SHR proteins from other plant species, including monocots and dicots from different families, gymnosperms and an ancestral vascular plant (Fig. 3).

The phylogenetic tree shows two major clusters identified as SHR1 and SHR2 (Fig. 3), but other clusters are present which include the gymnosperms and the ancestral vascular plant. It is furthermore shown that QsSHR proteins are more closely related to the SHR proteins from other dicots than to monocots or gymnosperms. Moreover, the two putative cork oak SHR-like proteins are clustered in different groups, with QsSHR1 clustering with the *Arabidopsis* AtSHR and *Populus* PtSHR1 (Fig. 3, dark pink boxes), and QsSHR2 clustering with the other two poplar SHR-like proteins (Fig. 3, light pink boxes) which present more differences when compared to the *Arabidopsis* SHR (Fig. 3). Sequences coding for putative QiSHR1 and QiSHR2 proteins seem to be phylogenetically most similar to the cork oak proteins.

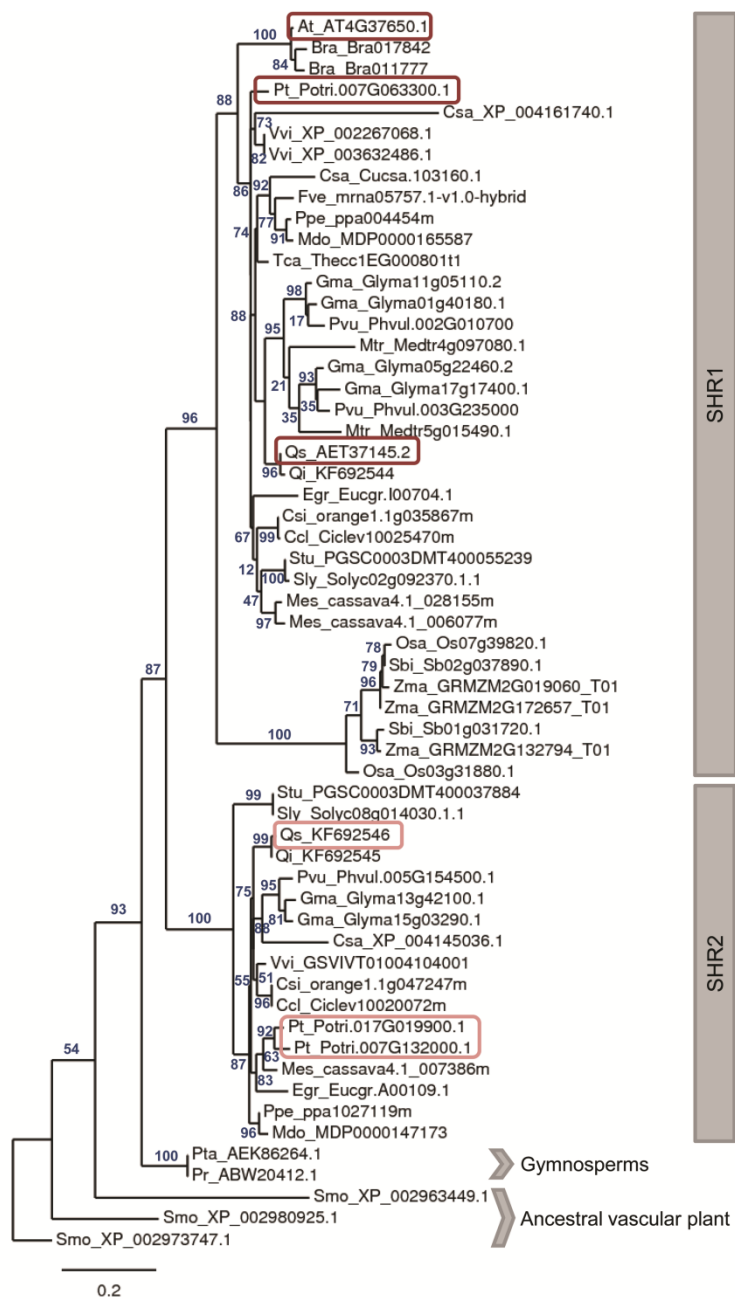


Fig. 3. Phylogenetic relationship among SHR-like proteins. SHR from *Ath* (*Arabidopsis thaliana*), *Bra* (*Brassica rapa*), *Csa* (*Cucumis sativus*), *Csi* (*Citrus sinensis*), *Ccl* (*Citrus clementina*), *Egr* (*Eucalyptus grandis*), *Fve* (*Fragaria vesca*), *Gma* (*Glycine max*), *Mdo* (*Malus domestica*), *Mes* (*Manhiot esculenta*), *Mtr* ►

◀(*Medicago truncatula*), Osa (*Oryza sativa*), Ppe (*Prunus persica*), Pr (*Pinus radiata*), Pta (*Pinus taeda*), Ptr (*Populus trichocarpa*), Pvu (*Phaseolus vulgaris*), Qi (*Quercus ilex*), Qs (*Quercus suber*), Sly (*Solanum lycopersicum*), Smo (*Selaginella moellendorffii*), Stu (*Solanum tuberosum*), Tca (*Theobroma cacao*), Vvi (*Vitis vinifera*), Zma (*Zea mays*). The boxes in a darker pink highlight the SHR1 proteins from *Populus*, *Quercus suber* and *Arabidopsis* and boxes drawn with a lighter pink highlight the SHR2 proteins from *Populus* and *Quercus suber*.

QsSHR1 and QsSHR2 are expressed in seedling tissues

The analysis of *QsSHR1* and *QsSHR2* transcript levels by semi-quantitative RT-PCR showed some differences among tissues (Fig. 4). At cycle 27 radicles and hypocotyls collected two days after germination (2DAG), as well as lateral roots from one month-old seedlings showed the strongest *QsSHR1* amplification signal revealing the predominance of these transcripts in such tissues (Fig. 4A, B). On the other hand, the highest *QsSHR2* expression levels were detected in hypocotyls collected at different stages of germination (Fig. 4A, B). In the last amplification cycle (35th), similar amplification signal of *QsSHR1* transcripts appeared in all analysed tissues, cotyledons and mature acorns showing less intense bands, while *QsSHR2* transcripts were less represented in the acorns and especially in the cotyledons (Fig. 4A, B). In conclusion, both *QsSHR1* and *QsSHR2* transcripts were present in seedling tissues, but *QsSHR1* seemed more expressed in the radicular tissues, whereas *QsSHR2* showed higher expression levels in the hypocotyls (Fig. 4A, B).

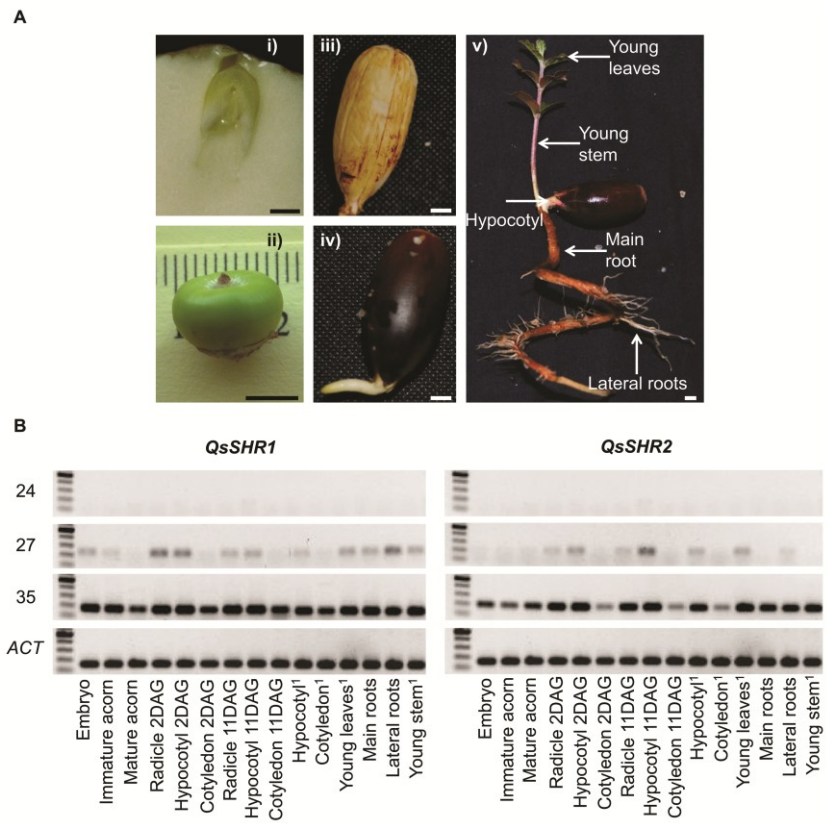


Fig. 4. Expression of cork oak *SHR*-like genes in acorns and seedling tissues of cork oak. (A) Cork oak tissues used for the semi-quantitative RT-PCR: i) Embryo, ii) Immature acorn, iii) Mature acorn, iv) Radicle emerging from a germinating acorn, v) One month-old seedling. The scale bar corresponds to 1 mm in i) and 5 mm in ii) to v). (B) Gene expression analysis in cork oak tissues using semi-quantitative RT-PCR. Samples were taken at PCR cycles 24, 27 and 35 and *ACTIN* was used as reference gene. Hypocotyl¹, cotyledon¹, young leaves¹ and young stem¹ mean that the tissues were collected from one month-old seedlings.

Quercus SHR-like genes are seasonally regulated in periderm tissues

Due to the interesting features of the cork oak periderm and to our previous observations in *Populus* which pointed to the specific localization of *PtSHR2B* transcripts in the phellogen (Miguel *et al.* submitted), we

proceeded with the analysis of the expression patterns of cork oak *SHR*-like genes during periderm development. The analysis of *QsSHR1* and *QsSHR2* transcript levels revealed a seasonal regulation of the expression of both genes but different patterns were observed (Fig. 5). The expression of *QsSHR1* was highest in the periderm of young one year-old branches collected in June and July (Fig. 5A). On the other hand, the highest expression level of *QsSHR2* was observed in the periderm of both one and two year-old branches collected in January (Fig. 5B). Focusing only on the periderm, both *SHR*-like genes presented highest transcript abundance in the younger branches. During the periods of highest expression *QsSHR1* transcript abundance generally decreased with the increasing age of the branches (June and July) while *QsSHR2* showed no significant differences in the periderm of one and two year-old branches (January) (Fig. 5A, B). In addition, the results also revealed different levels of transcript abundance for the two cork oak *SHR*-like transcripts either in periderm from branches of increasing age or in different seasons. *QsSHR2* transcripts were more abundant in all the periderm tissues collected over the year with the exception of periderm from one year-old branches collected in June. However, *QsSHR1* was more represented (two-fold more abundant) in leaves (Fig. 5A, B). With respect to the reproduction cork, the expression of both *QsSHR1* and *QsSHR2* was negligible.

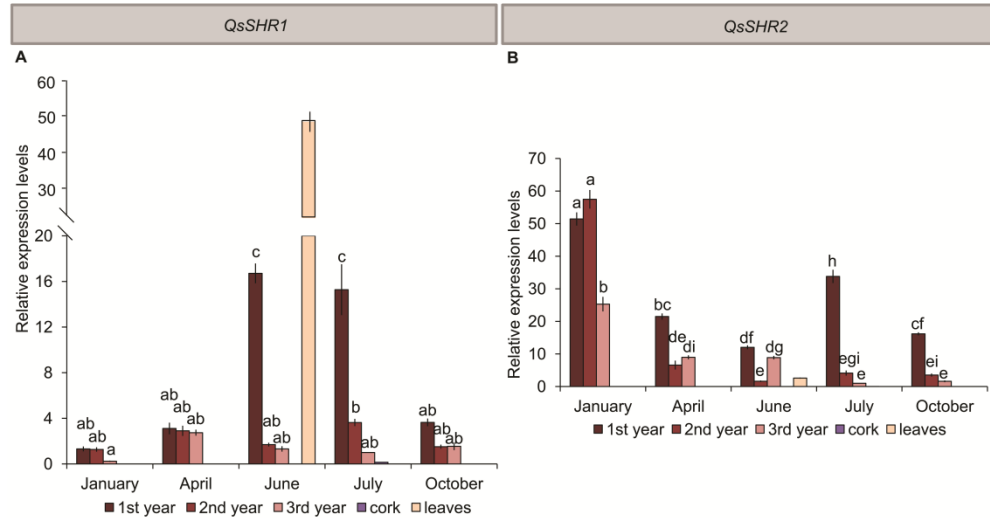


Fig. 5. Expression patterns of (A) *QsSHR1* and (B) *QsSHR2* during annual growth in the periderm from branches of different ages, in reproduction cork and in leaves of cork oak. Results are expressed relative to the expression in the periderm of three year-old branches collected in July and the data was normalized by a geometric mean of *ACT* and *CAC*. Values are means $\pm$ SE and different letters indicate statistical significance ($p < 0.05$, Tukey test) between different collection dates and periderm from branches of different age.

Relative transcript abundance of the two putative *SHR*-like genes in holm oak was also analysed. The stability of *ACTIN* and *CAC* expression in periderm tissues and leaves of holm oak was evaluated with the statistical algorithms GeNorm (Vandesompele *et al.*, 2002) and NormFinder (Andersen *et al.*, 2004). As previously reported for cork oak (Marum *et al.*, 2012), *ACTIN* and *CAC* also demonstrated to be adequate reference genes for these tissues in holm oak (Supplementary Table S1). The expression was analysed in periderm from one to three year-old branches collected in January, July and October showing that the transcription level of *SHR*-like genes is seasonally regulated also in this oak species (Fig. 6). Both genes were highly expressed in the periderm of one year-old branches regardless

of the collection date. The transcript abundance of *QiSHR1* was similar in periderm from branches of different ages and from different collection dates, except for the periderm of one year-old branches where the expression was similar in January and October and in July and October, with the highest expression levels observed in July (Fig. 6A). *QiSHR2* was differentially expressed in the periderm of one year-old branches collected in January comparing with the tissues collected in July or October (Fig. 6B), similarly to *QsSHR2* (Fig. 5B). As observed for the cork oak *SHR*-like genes, also *QiSHR2* transcripts were more abundant than *QiSHR1* in the periderm tissues collected in the different months and *QiSHR1* was significantly more transcribed in leaves than in periderm (Fig. 6A, B).

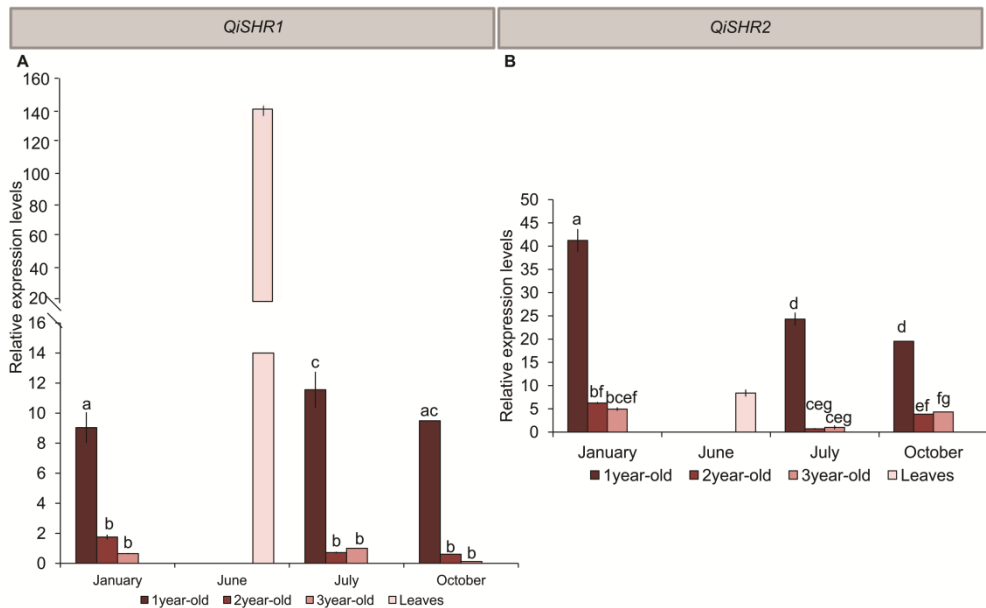


Fig. 6. Expression patterns of (A) *QiSHR1* and (B) *QiSHR2* during annual growth, in periderm from branches of different ages and in leaves of holm oak. Results are expressed relative to the expression in the periderm of three year-old branches collected in July and the data was normalized by a geometric mean of *ACT* and *CAC*. Values are means $\pm$ SE and different letters indicate statistical significance ($p < 0.05$, Tukey test) between different collection dates and periderm from branches of different age.

QsSHR1 and QsSHR2 are expressed in the lateral meristems

In situ hybridization was performed to determine the spatial expression patterns of both cork oak *SHR*-like genes. Despite the difficulties in carrying out this technique in woody samples, such as those collected from cork oak which required some modifications to previously described protocols, we succeeded in detecting positive hybridization signals. In one year-old branches *QsSHR1* appeared specifically localized in the vascular cambium (Fig. 7A). A visible staining was detected in the bark, but it was considered as background signal since it was also observed in the tissues hybridized with the sense probe (Fig. 7A, B). Regarding *QsSHR2*, the transcripts appeared localized in the phellogen, the other secondary meristem present in the analysed tissues, but a faint signal was also detected in the vascular cambium, as observed for *QsSHR1* (Fig. 7C, D).

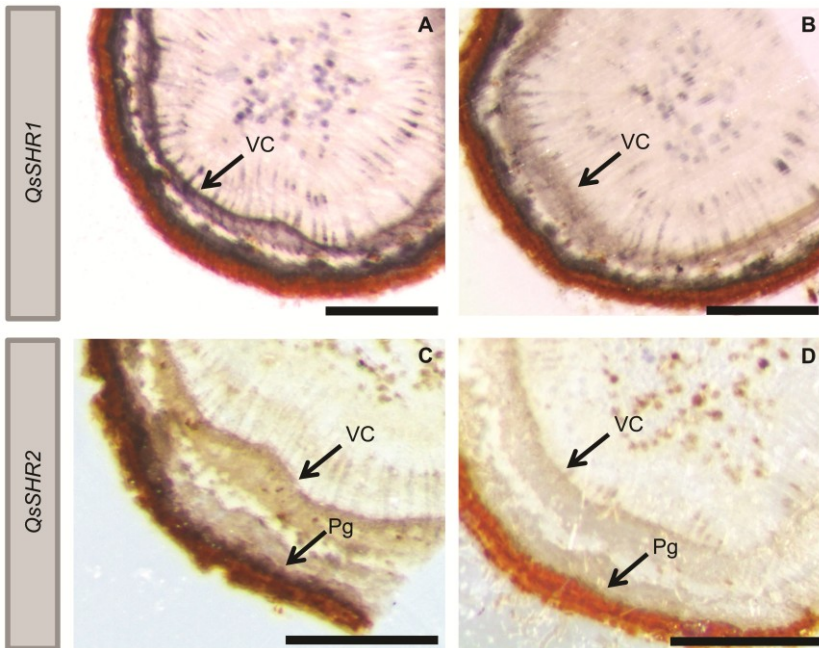


Fig. 7. *In situ* hybridization of *QsSHR1* and *QsSHR2* transcripts in one year-old branches of cork oak. Transverse section of the branch hybridized with (A) antisense *QsSHR1* RNA probe, (B) sense *QsSHR1* RNA probe as a control, (C) ►

◀antisense QsSHR2 RNA probe and (D) sense QsSHR2 RNA probe as a control. Pg, phellogen; VC, vascular cambium. Bars = 450 µm.

QsSHR1 or QsSHR2 functions do not fully complement the Arabidopsis shr2 mutant phenotype

To determine if the cork oak *SHR*-like genes play functions similar to the *Arabidopsis SHR*, transformation vectors containing the *Arabidopsis SHR* promoter in fusion to the complete coding region of *QsSHR1* or *QsSHR2* (*AtSHRprom:QsSHR1* or *AtSHRprom:QsSHR2*) were prepared. From the transformation of *Arabidopsis shr2* mutant, two and five transformants carrying the *AtSHRprom:QsSHR1* and *AtSHRprom:QsSHR2* construct, respectively, were selected for further studies. All the seedlings growing on selective medium exhibited the expected amplification product of the *Arabidopsis SHR* promoter and of the inserted cork oak cDNA. The complementation with either *QsSHR1* or *QsSHR2* did not recover the wild-type phenotype of the root (Fig. 8A). At the time of transfer of transgenic seedlings to soil, we did not observe differences in the rosette, but differences were noticed in the root length when compared to the wild-type. The number and size of the leaves in the rosette were similar in the transformed plants, wild-type and mutant and it should be pointed out that this situation is commonly observed when seedlings are grown on nutrient agar medium (Benfey *et al.*, 1993). However, the root length was smaller both in the transformed and mutant plants comparing to the wild-type (Fig. 8A).



Fig. 8. Phenotypic aspects of *Arabidopsis* wild-type (WT), *shr2* mutant (*shr2*) and *shr2* plants transformed with *AtSHRprom:QsSHR1* or *AtSHRprom:QsSHR2*. (A) Aspect of the roots and rosettes of plants before transfer to soil. (B) Aspects of the aerial region: Qs1 for the complementation with *AtSHRprom:QsSHR1* and Qs2 for the complementation with *AtSHRprom:QsSHR2*. Bars = 0.9 cm.

Detailed analysis of the transformed plants on soil for approximately one month showed that the rosette of the transformed plants complemented with either the *QsSHR1* or the *QsSHR2* gene seemed similar to the rosette of the *shr2* mutant (Fig. 8B). This indicates that the shoot apical meristem was still partially inhibited and that the wild-type phenotype of the root and rosette was not recovered.

Other developmental and growth parameters such as the height of the plants, number of leaves and number of branches were also recorded (Fig. 9A, B). The *Arabidopsis shr2* mutant transformed with *QsSHR1* or *QsSHR2* under the *Arabidopsis* SHR promoter did not show differences in

the number of branches when compared to the wild-type (Fig. 9A, B), but this number was higher than in the mutant plants and these differences were statistically significant suggesting partial restoration of wild-type functions. However, the height of the transformed plants and leaf size were similar to the mutant and smaller than the wild-type plants (Fig. 9A, B).

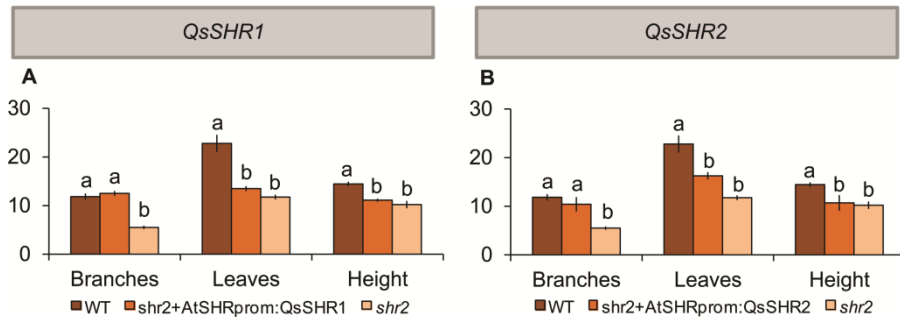


Fig. 9. Morphological parameters of the *Arabidopsis* wild-type, *shr2* mutant and *shr2* transgenic plants carrying (A) *AtSHRprom:QsSHR1* and (B) *AtSHRprom:QsSHR2*, approximately one month after transfer to soil, except for the number of leaves that were counted in plants after 15 days on soil. Values are means $\pm$ SE of 2 biological replicates for *shr2* plants with the *AtSHRprom:QsSHR1* insertion, 5 with the *AtSHRprom:QsSHR2* insertion and at least 6 biological replicates for *shr2* mutants and WT plants. Different letters indicate statistical significance in each parameter ($p < 0.05$, Mann-Whitney U test).

Discussion

SHR is a major regulator of several developmental processes in *Arabidopsis* and putative *SHR* orthologs have been found in other plants, including tree species such as *Populus* or *Pinus* (Solé *et al.*, 2008; Wang *et al.*, 2011). Functional studies in *Populus* revealed that the *Arabidopsis* *SHR* ortholog is expressed in the vascular cambium being involved in the control of secondary growth (Wang *et al.*, 2011).

To our knowledge, this is the first report on the characterization of *SHR*-like genes in the Fagaceae. In this work we identified and characterized a *SHR* gene, as well as a putative paralogous sequence, both in cork oak and in the related species holm oak, which does not produce a thick cork layer. The identified sequences presented high identity with *SHR* sequences from other plant species and, as expected, the analysis of the deduced amino acid sequences placed them in the GRAS family, the plant-specific protein family where the *SHR* proteins are included. The sequences share common characteristics of the GRAS proteins, which have a high conserved C-terminal region (Bolle, 2004). The finding of more than one *SHR*-like sequences both in cork oak and in holm oak has been reported also for other species (Wang *et al.*, 2011). This fact most probably results from genome duplication events occurred after the divergence of the *Arabidopsis* and other lineages, including cork oak, holm oak and poplar (Simillion *et al.*, 2002; Tuskan *et al.*, 2006; Velasco *et al.*, 2007; Jaillon *et al.*, 2007; Soltis *et al.*, 2009; Ueno *et al.*, 2010). In fact, it has been reported that the oak genome has gone through at least two series of whole genome duplication (Ueno *et al.*, 2010). In *Populus* three *SHR*-like genes have been found, *PtSHR1*, *PtSHR2A* and *PtSHR2B* (Wang *et al.*, 2011). The phylogenetic analysis performed in our study grouped one of the cork oak and holm oak *SHR* proteins, QsSHR1 and QiSHR1 respectively, in the same cluster as the *Arabidopsis* *SHR* and its *Populus* ortholog *PtSHR1*. QsSHR2 and QiSHR2 clustered with the other two putative *SHR*-like proteins from *Populus*, *PtSHR2A* and *PtSHR2B* (Wang *et al.*, 2011), as well as with *SHR* proteins from other plant species, which were less similar to that of *Arabidopsis*.

Although the expression of the cork oak *SHR*-like genes was not tissue-specific as revealed by the analyses performed in several seedling tissues, different expression patterns were noticed. The slightly higher expression levels of *QsSHR1* in the radicle tissues is in accordance with

transcript localization in the root observed for the putative orthologs in *Arabidopsis* and *Populus* (Benfey *et al.*, 1993; Helariutta *et al.*, 2000; Nakajima *et al.*, 2001; Wang *et al.*, 2011). On the other hand, the expression of *QsSHR2* was prevalent in the aerial tissues such as hypocotyls. Interestingly, previous data in *Pinus radiata* showed that *SHR* has a role not only in the root meristem formation but also in the cambial zone of hypocotyl cuttings (Solé *et al.*, 2008). Similarly to the observations in *P. radiata*, the weakest expression for both cork oak *SHR*-like transcripts was observed in the cotyledons. In *Populus*, both *PtSHR1* and *PtSHR2B* are expressed in roots and in the aerial part of the plant but, while *PtSHR1* expression is found in the root stele and in vascular cambium (Wang *et al.*, 2011), *PtSHR2B* is found in the root tip and in the phellogen (Miguel *et al.*, submitted). This prompted us to investigate the expression patterns of both cork oak *SHR* genes in the periderm of branches of different ages and along the year. The experiments revealed that although both transcripts, *QsSHR1* and *QsSHR2*, were detected in the periderm tissues, *QsSHR2* was significantly more abundant, suggesting that it plays a specific role in the periderm. On the other hand, *QsSHR1* was more expressed in the leaves suggesting it may have a role similar to the *Arabidopsis SHR* and the *Populus PtSHR1* (Wang *et al.*, 2011; Miguel *et al.*, submitted).

The results point *QsSHR2* as the putative ortholog of *PtSHR2B*, as suggested by sequence phylogenetic analysis and localization of the transcripts in the phellogen. Its higher expression in January than in June or July is interesting since it is known that the meristematic activity of the phellogen follows a seasonal pattern. Between November and February the phellogen is usually considered to be dormant (Caritat *et al.*, 1996; Fialho *et al.*, 2001; Costa *et al.*, 2002; Costa *et al.*, 2003), although it has also been reported to have a low level of activity from early autumn to the next spring (Oliveira *et al.*, 1994). In contrast, in June and July the phellogen activity reaches its maximum (Caritat *et al.*, 1996; Fialho *et al.*, 2001; Costa *et al.*,

2002; Costa *et al.*, 2003). The seasonal pattern of expression and the higher expression levels of *QsSHR2* in the periderm from branches at different ages collected in January suggests it negatively regulates meristem function. In the *Arabidopsis* root SHR functions in the maintenance of the meristematic cell layer (Benfey *et al.*, 1993; Helariutta *et al.*, 2000; Nakajima *et al.*, 2001; Sabatini *et al.*, 2003) and it has been recently shown that it acts in a dose-dependent way in the patterning of cell division within the endodermis, and as a negative regulator of meristem activity (Wang *et al.*, 2011; Koizumi *et al.*, 2012). The lack of *QsSHR2* expression in the cork tissues is probably due to the absence (or low number) of intact phellogen cells, and therefore *QsSHR2* transcripts, in the isolated phellem tissues. In fact, during debarking the separation of the cork plank (phellem) from the cork oak tree occurs at the level of the phellogen (Machado, 1935; Caritat *et al.*, 1996; Lulai and Freeman, 2001; Oliveira and Costa, 2012). As a result the phellogen is damaged leading to the development of a traumatic phellogen (Machado, 1935; Pereira *et al.*, 1992; Caritat *et al.*, 1996; Costa *et al.*, 2002; Pereira, 2007; Oliveira and Costa, 2012). Therefore, the tissues harvested from the cork plank contain mainly phellem cells (Almeida *et al.*, 2013). *QsSHR1* expression was also detected in the periderm tissues but at much lower levels than those of *QsSHR2*, and not high enough to be detected by *in situ* hybridization. It may be possible that while peeling off the periderm from the branches, some underlying tissue may have been collected leading to transcript amplification by the sensitive technique of reverse transcription quantitative PCR (RT-qPCR). A positive *QsSHR1 in situ* hybridization signal was instead detected in the vascular cambium which is in accordance with the reports describing the same transcript localization for the *Populus* putative ortholog, *PtSHR1* (Schrader *et al.*, 2004; Wang *et al.*, 2011). Our results also showed that *QsSHR2* transcript abundance was higher in the periderm of one year-old branches than in older branches. In cork oak, the phellogen meristematic

cell layer is initiated in the first year of growth after initiation of the vascular cambium activity, and extends throughout as a continuous layer with some phellem cells already formed (Graça and Pereira, 2004). The higher expression at this stage may be related to the meristem initiation process but further studies are necessary to elucidate this age-dependent variation in *QsSHR2* expression.

An interesting question was whether the differences in *SHR2* expression patterns between cork oak and holm oak could be somehow related to differences in cork production ability. Although the expression levels of *SHR2* were higher in cork oak than in holm oak, our results did not allow establishing any relation to the cork production ability in these two species. Seasonal variation in expression was also observed for holm oak *SHR* transcripts, which followed a trend similar to the one observed in cork oak. Higher expression levels were found in the periderm of one year-old branches and *QiSHR1*, like *QsSHR1*, was more expressed in July, whereas *QiSHR2* showed higher transcript abundance in one year-old periderm collected in January, similarly to *QsSHR2*.

Functional complementation studies in the *Arabidopsis shr2* mutant transformed with *QsSHR1* and *QsSHR2* coding sequences did not allow recovering the wild-type root and rosette phenotypes. The anomalies of the root in the transgenic *Arabidopsis* lines indicate that the regular morphology and organization of the quiescent center and the maintenance of the stem cell niche (Benfey *et al.*, 1993; Helariutta *et al.*, 2000; Sabatini *et al.*, 2003; Levesque *et al.*, 2006) is not recovered by the function of the *QsSHR1* or *QsSHR2*. The failure in leaf size recovery also indicates that the role of *SHR* in the regulation of proliferative cell divisions is not re-established (Dhondt *et al.*, 2010). Furthermore, the height of the transformed plants was similar to the mutants but the number of branches appeared similar to the wild-type plants. Taken together, these results revealed that neither of the two cork oak *SHR*-like sequences was able to fully complement the *Arabidopsis shr2*

mutant phenotype. This finding indicates that although the deduced protein sequences of QsSHR1 and QsSHR2 genes are 69% and 54% identical to the *Arabidopsis* SHR, respectively, there are relevant differences influencing their functions. Identically to what was suggested in other studies (Hörnblad *et al.*, 2013), a different subcellular localisation and thus, mislocalisation and/or different conformation of the QsSHR proteins when expressed in *Arabidopsis*, compared to AtSHR, may contribute to the failure in the full complementation of the mutant. The apparent partial complementation can, alternatively, be a consequence of gene duplication where the function of the original gene may have been distributed between the two copies (sub-functionalization) or the genes may have evolved different functions (neo-functionalization).

Overall, the localization of the cork oak *SHR*-like transcripts in the lateral meristems, like their putative orthologs in *Populus*, suggest that these genes are involved in the regulation of secondary growth. The localization of the *QsSHR2* together with its expression pattern during periderm development indicates that this gene, like the *Populus PtSHR2B*, might be involved in the regulation of phellogen activity during secondary growth but probably acting as a negative regulator, like *SHR1* in *Arabidopsis* and *Populus*. Our results also suggest a functional divergence between the *Arabidopsis* and cork oak SHR. These data can contribute to elucidate basic aspects of the molecular processes involved in the regulation of lateral meristems, and specifically in cork formation.

Acknowledgments

We thank Dr. Sónia Gonçalves (CEBAL, Portugal) for providing the cork tissue samples used in this work and Prof. C. Pinto Ricardo for all the help in harvesting the cork oak branches.

References

- Almeida T, Menéndez E, Capote T, Ribeiro T, Santos C, Gonçalves S.** 2013. Molecular characterization of *Quercus suber* MYB1, a transcription factor up-regulated in cork tissues. *Journal of Plant Physiology* **170**, 172–178.
- Andersen CL, Jensen JL, Ørntoft TF.** 2004. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets normalization of real-time quantitative reverse. *Cancer Research* **64**, 5245–5250.
- Aronson J, Pereira JS, Pausas JG.** 2009. *Cork Oak Woodlands on the Edge: Ecology, Adaptive Management, and Restoration* (J Aronson, JS Pereira, and JG Pausas, Eds.). Washington DC: Island Press.
- Benfey PN, Linstead PJ, Roberts K, Schiefelbein JW, Hauser MT, Aeschbacher RA.** 1993. Root development in *Arabidopsis*: four mutants with dramatically altered root morphogenesis. *Development* **119**, 57–70.
- Bolle C.** 2004. The role of GRAS proteins in plant signal transduction and development. *Planta* **218**, 683–692.
- Bugalho MN, Caldeira MC, Pereira JS, Aronson J, Pausas JG.** 2011. Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. *Frontiers in Ecology and the Environment* **9**, 278–286.
- Caritat A, Molinas M, Gutierrez E.** 1996. Annual cork-ring width variability of *Quercus suber* L. in relation to temperature and precipitation (Extremadura, southwestern Spain). *Forest Ecology and Management* **86**, 113–120.
- Caritat A, Gutiérrez E, Molinas M.** 2000. Influence of weather on cork-ring width. *Tree Physiology* **20**, 893–900.
- Castresana J.** 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* **17**, 540–552.
- Chaves I, Pinheiro C, Paiva JA, Planchon S, Sergeant K, Renaut J, Graça JA, Costa G, Coelho AV, Ricardo CPP.** 2009. Proteomic evaluation of wound-healing processes in potato (*Solanum tuberosum* L.) tuber tissue. *Proteomics* **9**, 4154–4175.
- Clough SJ, Bent AF.** 1998. Floral dip: a simplified method for *Agrobacterium*-mediated transformation of *Arabidopsis thaliana*. *The Plant Journal* **16**, 735–743.

Costa A, Pereira H, Oliveira Â. 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.

Costa A, Pereira H, Oliveira A. 2003. Variability of radial growth in cork oak adult trees under cork production. *Forest Ecology and Management* **175**, 239–246.

Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard J-F, Guindon S, Lefort V, Lescot M, et al. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Research* **36**, W465–W469.

Dhondt S, Coppens F, De Winter F, Swarup K, Merks RMH, Inzé D, Bennett MJ, Beemster GTS. 2010. SHORT-ROOT and SCARECROW regulate leaf growth in *Arabidopsis* by stimulating S-phase progression of the cell cycle. *Plant Physiology* **154**, 1183–1195.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792–1797.

Etchells JP, Turner SR. 2010. The PXY-CLE41 receptor ligand pair defines a multifunctional pathway that controls the rate and orientation of vascular cell division. *Development* **137**, 767–774.

Evert R. 2006. Periderm. *Esau's Plant Anatomy: Meristems, Cells, and Tissues of the Plant Body: Their Structure, Function, and Development*, Third Edition. John Wiley & Sons, Inc., 427–445.

Fialho C, Lopes F, Pereira H. 2001. The effect of cork removal on the radial growth and phenology of young cork oak trees. *Forest Ecology and Management* **141**, 251–258.

Gallagher KL, Paquette AJ, Nakajima K, Benfey PN, Carolina N. 2004. Mechanisms regulating SHORT-ROOT intercellular movement. *Current Biology* **14**, 1847–1851.

Gallagher KL, Benfey PN. 2010. Both the conserved GRAS domain and nuclear localization are required for SHORT-ROOT movement. *The Plant Journal* **57**, 785–797.

Ginzberg I, Barel G, Ophir R, Tzin E, Tanami Z, Muddarangappa T, De Jong W, Fogelman E. 2009. Transcriptomic profiling of heat-stress response in potato periderm. *Journal of Experimental Botany* **60**, 4411–4421.

Goodstein DM, Shu S, Howson R, Neupane R, Hayes RD, Fazo J, Mitros T, Dirks W, Hellsten U, Putnam N, et al. 2011. Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40**, D1178–D1186.

- Graça J, Pereira H.** 2004. The periderm development in *Quercus suber*. IAWA Journal **25**, 325–335.
- Groover AT, Mansfield SD, DiFazio SP, Dupper G, Fontana JR, Millar R, Wang Y.** 2006. The *Populus* homeobox gene *ARBORKNOX1* reveals overlapping mechanisms regulating the shoot apical meristem and the vascular cambium. Plant Molecular Biology **61**, 917–932.
- Harding SA, Tsai C, Cseke LJ.** 2011. Localization of RNA Expression. Handbook of Molecular and Cellular Methods in Biology and Medicine. CRC Press, 279–295.
- Helariutta Y, Fukaki H, Wysocka-Diller J, Nakajima K, Jung J, Sena G, Hauser MT, Benfey PN.** 2000. The SHORT-ROOT gene controls radial patterning of the *Arabidopsis* root through radial signaling. Cell **101**, 555–567.
- Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J.** 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biology **8**, R19.
- Hörnblad E, Ulfstedt M, Ronne H, Marchant A.** 2013. Partial functional conservation of IRX10 homologs in *Physcomitrella patens* and *Arabidopsis thaliana* indicates an evolutionary step contributing to vascular formation in land plants. BMC Plant Biology **13**, 1–14.
- Jackson D.** 1991. In situ hybridization in plants.pdf. In: Gurr S, McPherson M, BowlesDJ, eds. in Molecular Plant Pathology: A Practical Approach. Oxford University Press, 163–174.
- Jaillon O, Aury J-M, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, et al.** 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature **449**, 463–467.
- Karimi M, Inzé D, Depicker A.** 2002. Gateway vectors for *Agrobacterium*-mediated plant transformation. Trends in Plant Science **7**, 193–195.
- Karlgrén A, Carlsson J, Gyllenstrand N, Lagercrantz U, Sundström JF.** 2009. Non-radioactive in situ hybridization protocol applicable for Norway spruce and a range of plant species. Journal of Visualized Experiments: JoVE, 1–13.
- Koizumi K, Hayashi T, Wu S, Gallagher KL.** 2012. The SHORT-ROOT protein acts as a mobile, dose-dependent signal in patterning the ground tissue. Proceedings of the National Academy of Sciences of the United States of America **109**, 13010–13015.

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, et al. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948.

Levesque MP, Vernoux T, Busch W, Cui H, Wang JY, Blilou I, Hassan H, Nakajima K, Matsumoto N, Lohmann JU, et al. 2006. Whole-genome analysis of the SHORT-ROOT developmental pathway in *Arabidopsis*. *PLoS Biology* **4**, e143.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-(\Delta\Delta CT)}$ Method. *Methods* **25**, 402–408.

Lulai E, Freeman TP. 2001. The importance of phellogen cells and their structural characteristics in susceptibility and resistance to excoriation in immature and mature potato tuber (*Solanum tuberosum* L.) periderm. *Annals of Botany* **88**, 555–561.

Machado D. 1935. Contribuição para o estudo da formação da cortiça no sobreiro. *Revista Agronómica* **23**, 75–104.

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.

Mestre A, Gil L. 2011. Cork for sustainable product design. *Ciência & Tecnologia dos Materiais* **23**, 52–63.

Miranda I, Gominho J, Pereira H. 2012. Cellular structure and chemical composition of cork from the Chinese cork oak (*Quercus variabilis*). *Journal of Wood Science* **59**, 1–9.

Nakajima K, Sena G, Nawy T, Benfey PN. 2001. Intercellular movement of the putative transcription factor SHR in root patterning. *Nature* **413**, 307–311.

Neubauer JD, Lulai EC, Thompson AL, Suttle JC, Bolton MD. 2012. Wounding coordinately induces cell wall protein, cell cycle and pectin methyl esterase genes involved in tuber closing layer and wound periderm development. *Journal of Plant Physiology* **169**, 586–595.

Oliveira G, Correia O, Martins-loud MA, Catarino FM. 1994. Phenological and growth patterns of the Mediterranean oak *Quercus suber* L. *Trees* **9**, 41–46.

Oliveira G, Costa A. 2012. How resilient is *Quercus suber* L. to cork harvesting? A review and identification of knowledge gaps. *Forest Ecology and Management* **270**, 257–272.

Pereira H, Graça J, Baptista C. 1992. The effect of growth rate on the structure and compressive properties of cork. *IAWA Bulletin* **13**, 389–396.

Pereira H. 2007. *Cork: Biology, Production and Uses*. Amsterdam: Elsevier.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**, e45.

Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, Benfey PN. 1999. The GRAS gene family in *Arabidopsis*: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *The Plant Journal* **18**, 111–119.

Ramos M, Rocheta M, Carvalho L, Inácio V, Graça J, Morais-Cecilio L. 2013. Expression of DNA methyltransferases is involved in *Quercus suber* cork quality. *Tree Genetics & Genomes* **9**, 1481–1492.

Reid KE, Olsson N, Schlosser J, Peng F, Lund ST. 2006. An optimized grapevine RNA isolation procedure and statistical determination of reference genes for real-time RT-PCR during berry development. *BMC Plant Biology* **6**, 27.

Ricardo CPP, 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.

Sabatini S, Heidstra R, Wildwater M, Scheres B. 2003. SCARECROW is involved in positioning the stem cell niche in the *Arabidopsis* root meristem. *Genes & Development* **17**, 354–358.

Schrader J, Nilsson J, Mellerowicz E, Berglund A, Nilsson P, Hertzberg M. 2004. A high-resolution transcript profile across the wood-forming meristem of poplar identifies potential regulators of cambial stem cell identity. *The Plant Cell* **16**, 2278–2292.

Şen A, Miranda I, Santos S, Graça J, Pereira H. 2010. The chemical composition of cork and phloem in the rhytidome of *Quercus cerris* bark. *Industrial Crops and Products* **31**, 417–422.

Şen A, Quilhó T, Pereira H. 2011a. Bark anatomy of *Quercus cerris* L. var. *cerris* from Turkey. *Turkish Journal of Botany* **35**, 45–55.

Şen A, Quilhó T, Pereira H. 2011b. The cellular structure of cork from *Quercus cerris* var. *cerris* bark in a materials' perspective. *Industrial Crops and Products* **34**, 929–936.

Silva SP, Sabino MA, Fernandes EM, Correlo VM, Boesel LF, Reis RL. 2005. Cork: properties, capabilities and applications. *International Materials Reviews* **50**, 345–365.

Simillion C, Vandepoele K, Van Montagu MCE, Zabeau M, Van de Peer Y. 2002. The hidden duplication past of *Arabidopsis thaliana*. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 13627–13632.

Solé A, Sánchez C, Vielba JM, Valladares S, Abarca D, Díaz-Sala C. 2008. Characterization and expression of a *Pinus radiata* putative ortholog to the *Arabidopsis* *SHORT-ROOT* gene. *Tree Physiology* **28**, 1629–1639.

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.

Soler M, Serra O, Fluch S, Molinas M, Figueras M. 2011. A potato skin SSH library yields new candidate genes for suberin biosynthesis and periderm formation. *Planta* **233**, 933–945.

Soltis DE, Albert V a, Leebens-Mack J, Bell CD, Paterson AH, Zheng C, Sankoff D, Depamphilis CW, Wall PK, Soltis PS. 2009. Polyploidy and angiosperm diversification. *American Journal of Botany* **96**, 336–348.

Takechi K, Sakamoto W, Katsuhara M, Murata M, Motoyoshi F. 1999. In situ RNA hybridization using Technovit resin in *Arabidopsis thaliana*. *Plant Molecular Biology Reporter* **17**, 43–51.

Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N. 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science* **313**, 1596–1604.

Ueno S, Le Provost G, Léger V, Klopp C, Noirot C, Frigerio J-M, Salin F, Salse J, Abrouk M, Murat F, et al. 2010. Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. *BMC Genomics* **11**, 650.

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biology* **3**, RESEARCH0034.1–RESEARCH0034.11.

Velasco R, Zharkikh A, Troggio M, Cartwright DA, Cestaro A, Pruss D, Pindo M, Fitzgerald LM, Vezzulli S, Reid J, et al. 2007. A high quality draft consensus sequence of the genome of a heterozygous grapevine variety. *PLoS One* **2**, e1326.

Waisel Y, Liphschitz N. 1975. Sites of Phellogen Initiation. *Botanical Gazette* **136**, 146–150.

Wang J, Andersson-Gunnerås S, Gaboreanu I, Hertzberg M, Tucker MR, Zheng B, Leśniewska J, Mellerowicz EJ, Laux T, Sandberg G, et al. 2011. Reduced expression of the *SHORT-ROOT* gene increases the rates of growth and development in hybrid poplar and Arabidopsis. *PLoS ONE* **6**, e28878.

Waterhouse AM, Procter JB, Martin DM, Clamp M, Barton GJ. 2009. Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191.

Yu N-I, Lee SA, Lee M-H, Heo J-O, Chang KS, Lim J. 2010. Characterization of *SHORT-ROOT* function in the Arabidopsis root vascular system. *Molecules and Cells* **30**, 113–119.

Zhao S, Fernald RD. 2005. Comprehensive algorithm for quantitative real-time polymerase chain reaction. *Journal of Computational Biology* **12**, 1047–1064.

Supporting information

QsSHR1	MDTLFRLVSLQSDQSFNSSRTSSSSSRSSRQNHHPQEDEECFNFFMDEEDFSSSSSRHY
QiSHR1	MDTLFRLVSLQSDQSFNSSRTSSSSSRSSRQNHHPQEDEECFNFFMDEEDFSSSSSKHY
QsSHR1	PYHQQQQPNPSTTTTPTLTPTTTTNTNISTPTDFDFNFEFSSNWAPEALLETQAIAADKN
QiSHR1	PYHQQQQPNPSTTTTPTLTPTTTTNTNISTPTDFDFNFEFSSNWAPEALLETARAIAADKN
QsSHR1	SNRVQQLMWMLNELSSPYGDTDQKLASYFLQALFSRMTDSGERCYRTLASASDKTCSFES
QiSHR1	SNRVQQLMWMLNELSSPYGDTDQKLASYFLQALFSRMTDSGERCYRTLASASDKTCSFES
QsSHR1	TRKTVLKFQEVSPWTTFGHVACNGAILEALEGEPKLHIVDFSNTYCTQWPTLLEALATRT
QiSHR1	TRKTVLKFQEVSPWTTFGHVACNGAILEALEGEPKLHIVDFSNTYCTQWPTLLEALATRT
QsSHR1	DETPHLRLTTVVATKSGGVAAVQKVMKEIGNRMEKFARLMGVPFKFNVIHHVGDLEFNL
QiSHR1	DETPHLRLTTVVATKSRGVAAVQKVMKEIGNRMEKFARLMGVPFKFNVIHHVGDLEFNL
QsSHR1	ASELDIKDEALAINCVNTLHSTTTVDNRRDY LISNFRRLQPRIITVVEEADLDVGVDG
QiSHR1	ASELDIKDEALAINCVNTLHSTTAVDNRRDY LISNFRRLQPRIITVVEEADLDVGVDG
QsSHR1	VEFVKGFQOECLRWFRVYLEALEESFSKTSNERLMLERAAGRAIVDLVACAPSESIERRES
QiSHR1	VEFVNGFQOECLRWFRVYLEALEESFSKTSNERLMLERGAAGRAIVDLVACAPSESIERRES
QsSHR1	AARWARRLHVSGFSPISFSDEVCDV RALLRRYKEGWSMTQCSDAGIFLAWKDQPVVWAS
QiSHR1	AARWARRLHVSGFSPISFSDEVCDV RALLRRYKEGWSMTQCSDAGIFLAWKDQPVV---
QsSHR1	AWRP
QiSHR1	----

Supplementary Fig. S1. Alignment of the deduced amino acid sequences of SHR1 in *Q. ilex* (KF692544) and *Q. suber* (AET37154), using ClustalW2 (Larkin *et al.*, 2007) and Jalview (Waterhouse *et al.*, 2009). Regions of identity are shaded.

```

QsSHR2 MDITLFTPEESPFFQSTQQSNQGAAIDMHSNNQPQNSHTSTSRSSDSGVLPASGKWASR
QiSHR2 -----QQSNQGAAIDMHSNNQPQNSHTSTSRSSDSGVLPASGKWASR

QsSHR2 LLKECAQAISEKDSSKIHQLLWLLNELASPYGDCDQKLASYFLQALFCKATESGQRCYKT
QiSHR2 LLKECAQAISEKDSSKIHQLLWMLNELASPYGDCDQKLASYFLQALFCKATESGQRCYKT

QsSHR2 LTSVADKSHSFDSARKLILKFQEVSPWMTFGHVASNGAILEALEGETKLHIIDISNTLCT
QiSHR2 LTSVADKSHSFDSARKLILKFQEVSPWMTFGHVASNGAILEALEGETKLHIIDISNTLCT

QsSHR2 QWPTLLEALATRNDDTPHLKLTVVVTSTVKSVMKEVGQRMKEFARLMGVPPFENAIMGL
QiSHR2 QWPTLLEALATRNDDTPHLKLTVVVTSTVKSVMKEVGQRMKEFARLMGVPPFENAIMGL

QsSHR2 NHLAELTKEVLGVQEDEAIVVNCIGALRRVEVEERGAVIQMFQSLRPRVVTIVEEEADLS
QiSHR2 NHLAELTKEVLGVQEDEAIVVNCIGALRRVEVEERGAVIQMFQSLRPRVVTIVEEEADLS

QsSHR2 SSRDDFVKCFEECLRFYTLTYFEMLEESFVPTSNERLMLERDCSRISIVRVLACDDEKSCEE
QiSHR2 SSRDDFVKCFEECLRFYTLTYFEMLEESFVPTSNERLMLERECSRSIVRVLACDDEKSCEE

QsSHR2 DCERRERGSQWSERLKEAFSPVGFSDDVDDVKALLKRYKAGWALVLPQGDHDDSGIYLT
QiSHR2 DCERRERGSQWSERLKEAFSPVGFSDDVDDVKALLKRYKAGWALVLPQGDHDDSGIYL-

QsSHR2 WKEEPVLWASAWKP
QiSHR2 -----

```

Supplementary Fig. S2. Alignment of the deduced amino acid sequences of SHR2 in *Q. ilex* (KF692545) and *Q. suber* (KF692546) using ClustalW2 (Larkin *et al.*, 2007) and Jalview (Waterhouse *et al.*, 2009). Regions of identity are shaded.

Supplementary Table S1. Stability values of candidate reference genes in *Q. ilex* calculated by different statistical methods, GeNorm and NormFinder.

	Stability values	
	GeNorm	NormFinder
ACT	0.462	0.110
CAC	0.475	0.086
EF	0.548	0.183

Chapter V

Transcript profiling during acorn development in cork oak: a contribute towards a reference transcriptome



Miguel A¹, de Vega-Bartol J¹, Marum L, Miguel CM. Transcript profiling during acorn development in cork oak. (*in preparation*) ¹Equal contribution

Andreia Miguel participated in the experimental design, performing laboratory experiments, analyzing the expression results and writing the paper.

The data here obtained was also included in the following publication:

Pereira-Leal JB, Abreu IA, Alabaça CS, Almeida MH, Almeida P, Almeida T, Amorim MI, Araújo S, Azevedo H, Badia A, Batista D, Bohn A, Capote T, Carrasquinho I, Chaves I, Coelho AC, Costa MM, Costa R, Cravador A, Egas C, Faro C, Fortes AM, Fortunato AS, Gaspar MJ, Gonçalves S, Graça J, Horta M, Inácio V, Leitão JM, Lino-Neto T, Marum L, Matos J, **Miguel A**, Miguel C, Morais-Cecílio L, Neves I, Nóbrega F, Oliveira MM, Oliveira R, Pais MS, Paiva JÁ, Paulo OS, Pinheiro M, Raimundo JÁ, Ramalho JC, Ribeiro AI, Ribeiro T, Rocheta M, Rodrigues AI, Rodrigues JC, Saibo NJ, Santo TE, Santos AM, Sá-Pereira P, Sebastiana M, Simões F, Sobral RS, Tavares R, Teixeira R, Varella C, Veloso MM, Ricardo CP. *A comprehensive assessment of the transcriptome of cork oak (Quercus suber) through EST sequencing*. 2014. BMC Genomics **15**, 371. (doi: 10.1186/1471-2164-15-371)

Andreia Miguel contributed for the experimental design, participated in the preparation of plant material and laboratory experiments.

Transcript profiling during acorn development in cork oak: a contribute towards a reference transcriptome

Abstract

Cork oak (*Quercus suber* L.) has a natural distribution across Mediterranean regions and is one of the keystone forest tree species in these ecosystems. Fruiting and germination phases are especially critical for its regeneration but the molecular mechanisms underlying the biochemical and physiological changes during oak fruit (acorn) development are still poorly understood. In this study, the transcriptome of the cork oak acorn was characterized along five stages of development in order to identify the dominant processes in each stage as well as important regulatory genes. 80,357 expressed sequence tags (ESTs) were *de novo* assembled from RNA-Seq libraries representative of the several developmental stages. The analysis of expression profiles along development identified 2,285 differentially expressed (DE) transcripts, which were clustered into six groups according to their expression profiles. Of these, 21.8% were transcripts putatively coding for transcription factors (TF) which were almost equally distributed throughout acorn development, highlighting their major roles in controlling this process. Carbohydrate metabolism was the level-2 pathway most represented during cork oak fruit development but especially in late stages as evidenced by enrichment analysis in the expression clusters. Our work also shows that genes related to response to water were mainly represented during the early (S2) and late stages (S8) of acorn development, when tolerance to water desiccation is most needed. The obtained results provide novel insights into the developmental biology of acorns and, to our knowledge this work represents the first report of fruit development transcriptomics in oaks.

Keywords

Quercus suber, fruit, transcriptomics, transcription factors, response to water, carbohydrate metabolism

Introduction

The Fagaceae family comprises more than one thousand species, half of which belong to the *Quercus* genus and are commonly known as oaks. The oaks have adapted to extremely variable habitats and are widely distributed throughout the northern hemisphere in an almost continuous pattern, having a high socio-economical and ecological impact. Cork oak (*Quercus suber* L.) is the native oak species in the Western Mediterranean and North Africa regions (Bugalho *et al.*, 2011) characterized by hot and dry summers but benign winters. Cork oak has been considered a keystone forest tree species in the ecosystems where it grows (Aronson *et al.*, 2009; Ramírez-Valiente *et al.*, 2009a), but it is mostly recognized for producing cork. The removal of cork planks from adult cork oak trees every 9 years sustains highly profitable cork industries that are crucial for the economy of some Mediterranean countries such as Portugal (Silva *et al.*, 2005). Given the long lifespan of cork oak, breeding activities are hardly conducted and natural regeneration is the most common way of cork oak propagation (Boavida *et al.*, 1999). Seed development and germination are thus critical for the successful maintenance of the cork oak growing regions (Belletti, 2001; Belletti *et al.*, 2004; Pausas *et al.*, 2009).

During development, the acorn undergoes many biochemical and physiological changes which likely confer the ability to survive the severe drought periods and high temperatures to which cork oak is typically exposed to. A peculiar characteristic of cork oak reproductive strategy is the production of annual and biennial acorns (Merouani *et al.*, 2003; Díaz-Fernández *et al.*, 2004; Pons and Pausas, 2012), often present simultaneously in the same tree (Díaz-Fernández *et al.*, 2004). The significance of this strategy is not clear but it possibly provides physiological plasticity enabling a rapid switch between fruiting types as an adaptation to variable climate conditions (Elena-Rosello *et al.*, 1993; Pons and Pausas,

2012). The acorn size has also been related to adaptive traits related to drought conditions (Moles and Westoby, 2004; Ramírez-Valiente *et al.*, 2009b; Bonito *et al.*, 2011) and seed germination ability (Leishman *et al.*, 2000; Gómez, 2004; Moles and Westoby, 2004; Bonito *et al.*, 2011).

Given their importance for human consumption, fruits of several species such as tomato (Moore, 2002; Fei *et al.*, 2004), grape (Goes *et al.*, 2005; Peng *et al.*, 2007; Sweetman *et al.*, 2012), apple (Park *et al.*, 2006; Janssen *et al.*, 2008), date palm (Yin *et al.*, 2012; Zhang *et al.*, 2012), blueberry (Rowland *et al.*, 2012), cucumber (Ando *et al.*, 2012) sweet orange (Yu *et al.*, 2012) and maize (Teoh *et al.*, 2013), have been subjected to transcriptomic analyses during development. Genes related to fruit ripening have been extensively studied in tomato, grape and sweet orange (Moore, 2002; Fei *et al.*, 2004; Sweetman *et al.*, 2012; Yu *et al.*, 2012) and fruit specific genes have been also pointed out in apple (Park *et al.*, 2006) and date palm (Zhang *et al.*, 2012). In oaks, the few studies that have been conducted in fruits have focused on morphological, physiological and phenological aspects of the acorns (Elena-Rosello *et al.*, 1993; Merouani *et al.*, 2003; Bonito *et al.*, 2011). Within the *Quercus* species a few reports exist on aspects of the male and female flower development (Varela and Valdivieso, 1996; Boavida *et al.*, 1999, 2001) and flower/fruit anatomy (Borgardt and Nixon, 2003). To the best of our knowledge, transcriptomic studies of fruit and/or embryo development have not been reported in the Fagaceae family. Nonetheless, several transcriptomic (Derory *et al.*, 2006; Durand *et al.*, 2010; Ueno *et al.*, 2010, 2013; Santamaría *et al.*, 2011) and genomic (Faivre Rampant *et al.*, 2011; Lesur *et al.*, 2011) studies have been conducted in oaks and a Fagaceae Genomics database is available (<http://www.fagaceae.org/>). A Cork Oak ESTs database (CODB; <http://www.corkoakdb.org/>) has also been recently released providing information on ESTs present in different tissues and conditions, including embryogenesis and fruit development (Pereira-Leal *et al.*, 2014).

Here we present the dynamics of the transcriptome of cork oak acorns along five stages of development defined according to morphological characters from early development to maturation. The obtained data were analysed in order to gain insights into the molecular mechanisms underlying differentiation and morphogenesis of the developing cork oak acorn and to identify transcripts putatively associated with adaptive traits. Our approach identified genes with putative roles on key developmental processes during cork oak acorn development and further focused on the identification of transcripts and biological processes related with water response, including water transport and deprivation, as well as on transcription factors. Carbohydrate metabolism was also discussed mainly focusing on transcripts putatively involved in controlling sugar status changes along acorn development. As result of this work novel resources for studying cork oak biology are made available and this knowledge may be useful to other Fagaceae species.

Materials and Methods

Plant material

Acorns were collected between mid June and late November 2009 from cork oak trees growing in six different locations in the South and Centre of Portugal: Quinta da Serra (Vila Nogueira de Azeitão), Alter do Chão, São Brás de Alportel, Monchique, Calhariz (Santarém) and Abrantes. The term acorn is here used as referring to the whole structure consisting of the pericarp and all the tissues enclosed by the pericarp including the seed. Acorns in the early stages of development, from stage S1 to S4 (Fig. 1A), were collected from several trees in each location and were immediately frozen in liquid nitrogen. To obtain the samples at S5 developmental stage

and subsequent stages, branches bearing the acorns were kept at 4° C for up to 24 h and acorns were afterwards washed before freezing in liquid nitrogen. In S8 stage samples approximately 1/3 of the acorn part opposite to the embryo radicle was removed to minimize the presence of polysaccharides that could compromise the purity of isolated RNA. Additionally, some acorns from stage S3, S4, S5 and S8 were opened for embryo isolation; embryos from S8 stage were isolated by excising the embryo axis but excluding most cotyledonal tissue. In each collection, acorns and isolated embryos were carefully observed for evaluation and documentation of the developmental stage. No distinction was made between acorns with annual and biennial maturation.

RNA extraction

Frozen samples were first ground in a mill (M 20 Universal mill, Ika) previously cooled with liquid nitrogen and then transferred to a cooled mortar to be reduced to a fine powder. Total RNA was isolated following a protocol described by Reid *et al.* (2006) with some minor modifications (Marum *et al.*, 2012). RNA isolation was performed separately by developmental stage, date of collection and donor tree.

Total RNA was purified using the RNeasy MinElute Cleanup kit (Qiagen) with on-column DNase I treatment (Qiagen RNase-Free DNase Set) and only samples with $A_{260/280} > 1.8$ were used for further steps. RNA integrity was assessed in 1% (w/v) agarose gels after ethidium bromide staining and for a rigorous assessment of RNA quality, the RNA samples were run on a RNA Pico6000 chip in Agilent 2100 Bioanalyzer RNA (Agilent). Additionally, each sample was quantified by fluorescence with the Quant-iT Ribogreen RNA Assay kit (Invitrogen).

Preparation of cDNA libraries and RNA-Seq

Two normalized and 5 non-normalized cDNA libraries were prepared. Normalized libraries were prepared with RNA isolated from acorn tissues (FR) or from isolated embryos (EM). In each library, a pool of 2 µg total RNA containing equal amounts of RNA extracted from the several acorn (from S1 to S5, S7 and S8) or embryo developmental stages (from S3 to S5 and S8) was prepared. Double stranded cDNA was obtained using SMART technology (Zhu *et al.*, 2001) and the normalization was performed with the Duplex-Specific Nuclease (DSN) technology (Zhulidov *et al.*, 2004).

The 5 non-normalized cDNA libraries were prepared corresponding to different acorn developmental stages (S1, S2, S3+S4, S5 and S8). For each library, a pool of total RNA was prepared containing 50 µg of total RNA. MicroPoly(A)Purist kit (Ambion) was used to isolate mRNA from each total RNA pool and 200 ng of mRNA were fragmented and used as template for double stranded cDNA production using cDNA Synthesis System Kit (Roche) followed by adaptor ligation.

Pyrosequencing of the normalized and non-normalized libraries was performed in the Titanium GS-FLX (454-Roche) at Biocant (Cantanhede, Portugal).

The data were deposited in the European Nucleotide Archive (ENA) under the accession number PRJEB6178/ERP005652. For each non-normalized library the accession numbers are the following: ERX455655 for S1, ERX455656 for S2, ERX455657 for S3S4, ERX455658 for S5 and ERX455659 for S8. For the normalized libraries ERX455660 and ERX455661 are the accession numbers of the FR and EM libraries, respectively.

Reads pre-processing, de novo assembly and transcriptome annotation

The full workflow is schematized in the Supplementary Fig. S1. First, the raw reads were filtered by SeqTrimNext (Falgueras *et al.*, 2010) to remove adapter sequences and low quality/complexity sequences, which included fragments (window of 15 nts) with a quality value lower than 20, more than an 80% of indeterminations, or 75% of polyA or polyT sequences. Fragments with an E-value $< 1e^{-10}$ and 85% identity to contaminants such as plastids, mitochondria, ribosome and virus/bacteria sequences, were trimmed. Final sequences shorter than 40 nts were also excluded. RNA-Seq data was *de novo* assembled using MIRA version 3.4.0 (Chevreux *et al.*, 2004) and Newbler version 2.6 (Margulies *et al.*, 2005). MIRA was executed with the default 454 settings and without clipping steps. Newbler was executed with the default parameters. The individual assemblies were merged with CAP3 with default options and an identity threshold of 95%.

Transcripts were compared with the Uniprot and Trembl databases using NCBI Blastx with an E-value of $1e^{-6}$. Only full-length plant proteins were included in the target database. Full Lengther Next scripts (www.scbi.uma.es/) were used to compare the aligned regions in query and target, in order to determine the right translation frame and classify the transcripts as complete, internal or terminal. These translated proteins constitute the *Q. suber* proteome used in later comparisons. For those transcripts without any alignment, the program runs an Open Reading Frame (ORF) prediction step. Novel ORFs with a result higher than 0.7, the default threshold, were annotated as novel transcripts. Transcripts were compared with the NCBI non-redundant (nr) and *Arabidopsis* TAIR protein databases using NCBI Blastx with an E-value of $1e^{-10}$. Results were imported in Blast2GO (Conesa *et al.*, 2005) to annotate the gene ontology terms, enzymatic protein codes and KEGG pathways. The reads of each

transcript belonging to the same level-2 KEGG pathway were summed up. For each pathway, the number of reads in each stage was transformed in Z-scores, clustered, and plotted in a heatmap using Mayday (Battke *et al.*, 2010). Gene ontology terms and NCBI COGs (Clusters of Orthologous Groups of proteins) associated to each *Arabidopsis* gene were downloaded from TAIR (www.arabidopsis.org), and associated back to the original *Q. suber* transcript.

The transcriptomes of other Fagaceae species were downloaded from NCBI and Fagaceae Genomics Project (www.fagaceae.org). The proteome for each of them was built in a similar way as for *Q. suber* by comparison with the Uniprot and TrEMBL full plant proteins using the Full Lengther Next scripts. Proteomes were compared by pairs using NCBI Blastp. Proteins in a query species were considered as having an orthologous in a target species if they shared both a minimal identity and coverage of 70%.

Expression analysis and differentially expressed gene (DEG) clustering

The CLC Genomic Workbench was used to quantify the expression of the RNA-Seq data in four steps. First, the reads from each of the five non-normalized libraries were aligned to the 80,357 contigs in the *Q. suber* assembly using the default scoring values and ignoring reads not uniquely mapping. Second, the number of aligned reads in each library was normalized by quartile normalization to take into account the different total number of reads per library. Third, a statistical analysis that compares the expected versus observed proportions of mapped reads by Kal's z-test between consecutive acorn developmental stages was used to identify the differentially expressed genes (DEGs). Finally, p-values were False

Discovery Rate (FDR) corrected. Transcripts with a FDR value lower than 0.01 were considered as differentially expressed (DE).

DEGs were divided in 6 clusters according to the normalized number of aligned reads in each stage by Neural Gas clustering base on Euclidian correlation using Mayday. The list of transcripts in each cluster was used in Blast2GO to identify the enriched GO terms. Blast2GO enrichment analysis was based on a F-fisher test ($FDR < 0.05$). The relation among GO terms was assigned using REVIGO with the Resvik algorithm option (Supek *et al.*, 2011) and R treemap library. DEGs annotated as related to response to water, including water transport and water deprivation were identified. DEG sequences were aligned to the Plant Transcription Factor database (<http://plntfdb.bio.uni-potsdam.de/>) using Blastx and an E-value of $1e^{-10}$, to identify transcription factors.

Quantitative RT-PCR analysis

Reverse transcription quantitative real-time PCR (RT-qPCR) was carried out to validate a set of the DEGs (Table 1). RNA samples were first treated with TURBO DNase (Ambion) and afterwards all cDNAs were synthesized from 1.5 µg of total RNA using the Transcriptor High Fidelity cDNA Synthesis Kit (Roche) with the anchored-oligo(dT)₁₈ primers. Specific primers were designed using Geneious software (Kearse *et al.*, 2012). Quantitative real-time PCR experiments were then performed in LightCycler 480 (Roche) using SYBR Green I Master (Roche) and 96-well plates. For the genes tested, 3 biological replicates were used and the reaction mixtures were performed in a final volume of 16 µl containing 8 µl of 2× SYBR Green I Master, 400 nM of each primer and 1.5 µl of cDNA as template. The amplification program was the same for all genes tested: 95° C for 10 min, 45 cycles of 10 s at 95° C, 20 s at 60° C and 10 s at 72° C, except for *β-AMYLASE 1 (BAM1)*, *GALACTINOL SYNTHASE 2 (GoIS2)* and *CALCIUM*

DEPENDENT PROTEIN KINASE 10 (CPK10) for which the annealing temperature was 62° C. A calibrator sample was used in each plate to normalize the values obtained and the potential differences among plates. Normalization was carried out with two reference genes *ACTIN (ACT)* and *CLATHRIN ADAPTOR COMPLEXES (CAC)* (Marum *et al.*, 2012). Normalized relative quantities were obtained through the formula $NRQ =$

$$\frac{E_{goi}^{\Delta Ct, goi}}{\sqrt[f]{\prod_0^f E_{ref_0}^{\Delta Ct, ref_0}}},$$

where E is the efficiency of the amplification for each primer pair in each tissue, f the number of reference genes used to normalize the data, goi and ref are the gene of interest and the reference gene, respectively, and ΔCt is the Ct of the calibrator minus the Ct value of the sample in test (Pfaffli, 2001; Livak and Schmittgen, 2001; Hellemans *et al.*, 2007). The data obtained from the RNA-Seq experiment and the RT-qPCR were compared. From the RNA-Seq a logarithmic ratio of base 2 between the counts of a gene in each developmental stage and the mean counts of the same gene in all developmental stages were made. A similar approach was followed for the data obtained by RT-qPCR by doing a logarithmic ratio of base 2 between the normalized quantities of the gene of interest in each developmental stage and the mean normalized quantities of the same gene in all developmental stages in analysis. For genes where the RNA-Seq values were zero in some of the developmental stages, a value of 1 was added to all the RNA-Seq results of those genes to avoid indetermination (Stewart *et al.*, 2013).

Table 1. Specific primers used for quantitative real-time PCR. Cluster, according to their expression profiles each transcript belong to a different cluster; Gene abbreviation, corresponds to the homologous in *Arabidopsis thaliana*.

Cluster	Gene abbreviation	Gene description	Transcript name	At Locus	Primer sequences (forward/reverse)
A	MRP10, ABCC14	ABC transporter C family member 14	Contig3296	AT3G62700	GCTGCCCTTGGCCCACT/ TGGAAGAGCCTTGAACGCTGCC
A	CPK10, CDPK1	Calcium-dependent protein kinase	Contig4133	AT1G18890	GCTCACCTCCCTCGCAACT/ AGGCGGAGGTGGCGATAAT
B	ASPG1	Protein aspartic protease in guard cell	Contig19387	AT3G18490	ATGCAGCTCGCTCGACGTG/ TGTCGTGTCCGCAACCCAGA
B	ABF2	Abscissic acid responsive elements-binding factor 2	Contig22999	AT1G45249	CGGCCTGCTTTTCCGCAACT/ AGTCAGCTGCAAGGTCACGAGC
B	AVP1	Pyrophosphate-energized vacuolar membrane proton pump	Contig 8793	AT1G15690	CGCACTTGAGAACGACGCT/ TGC GCGGTCTCGGAATCAT
B	HSP90-7	Chaperone protein htpG family protein	Qs-dev_rep_c72408	AT4G24190	CAGCATCAGCTTCAGCCTCGGT/ AGCGGCTTTCACGCTTCTGA
B	BAM1, BMY7, TR-BAMY	β -amylase 1	Contig 25491	AT3G23920	AACGCGAGTTTGCAGGCGGT/ CGACGTTTCCGCCGCAATTGA
B	GoS2	Galactinol synthase 2	Contig 17943	AT1G56600	TGACGCCATAACATGGCCAGCA/ TTGCCAGCAGTGCCCTGACA
C	ALDH7B4	Aldehyde dehydrogenase 7B4	Contig 9124	AT1G54100	AGGGTGCTCCAACGACTCCA/ ACGCACAGCTAGGCCGATG
C	OPR3	Oxophytodienoate-reductase 3	Contig 16616	AT2G06050	TTGGTCTCTAACACAGCCGCCG/ CACGCCGACATGCCATA

Table 1. (Continued)

Cluster	Gene abbreviation	Gene description	Transcript name	At Locus	Primer sequences (forward/reverse)
C	PAL1	Phenylalanine ammonia-lyase 1	Contig 9321	AT2G37040	TGCTAACTGGCCGCCCAAT/ GCCAGAACCAACAGCTGTGCCA
D	ABI3, SIS10	Abscisic acid insensitive 3	Contig 19318	AT3G24650	GCAGTGGCCGTGGTGCAAT/ ATGGCGAGGCAAAAGCGGT
D	PHS1	Glycosyl transferase, family 35	Contig 10143	AT3G29320	TGGCTTGAGATGGGCAACCCT/ TTTGGTGTGCTCCCCGGCAT
D	SIR	Sulfite reductase	Contig 19883	AT5G04590	TGCAATGGCATGCCCAGCCT/ ACTGGGGTTTCCACCAAGCCA
D	PIP2B	Aquaporin PIP2-4	Contig 24062	AT2G37170	GCTTGGGCCCTTTGGTGGA/ TGGCTCCACCACCGTACTTGCT
F	COR414-TM1	Cold-regulated 314 inner membrane 1	Contig 19349	AT1G29395	TGGTGAGCTTGAGCTGCCGT/ TGCAAGCTGCCTGCTCGTGA
F	DI21	Late embryogenesis abundant protein	Contig 16865	AT4G15910	AGCGTGCTCTCACCTTGTGGT/ TGC GGCTGCATCACATGCCT
F	JMT	Jasmonic acid carboxyl methyltransf.	Contig 17776	AT1G19640	TGGCTATCTTCGCCGCCCTT/ TGCCGTGGTTTGACACTACTTCG
F	GSTU25	Glutathione S-transferase TAU 25	Contig 17981	AT1G17180	ACCACCTCGTCCGCCATTG/ AGCCTGACCGACTCCATGGCAA
F	LTP3	Lipid transfer protein 3	Contig 19993	AT5G59320	TGCCAGGCCACCAATACATGC/ TGTGGTCTTGGCCATGCCGT

Results

Categorization of cork oak acorns into different developmental stages

Although the fruits are often defined as structures derived from a mature ovary containing the seeds, many structures that might be defined as fruit are in fact composed of different tissue types (Seymour *et al.*, 2013). Other definitions have been proposed such as the one by Van der Pijl (1982) that considers the fruit as the dispersal unit. As previously mentioned, in this work we use the term acorn for simplicity, referring to all the tissues enclosed by the pericarp including the seed. Cork oak acorns were collected from June to November in order to cover all developmental stages, from early development to full maturation. A staging system was established based on several morphological aspects (Fig. 1 and Table 2). Since the dimensions of the acorns were variable in the same collection date among trees in different locations, additional morphological parameters were used to establish developmental classes. These included the presence of a distinct embryo(s), multiple embryos or a dominant embryo within the developing seed, covering of the acorn by the cupule and colour of the pericarp (Table 2). According to this system, eight stages of acorn development were established (S1-S8, Fig. 1A, B).

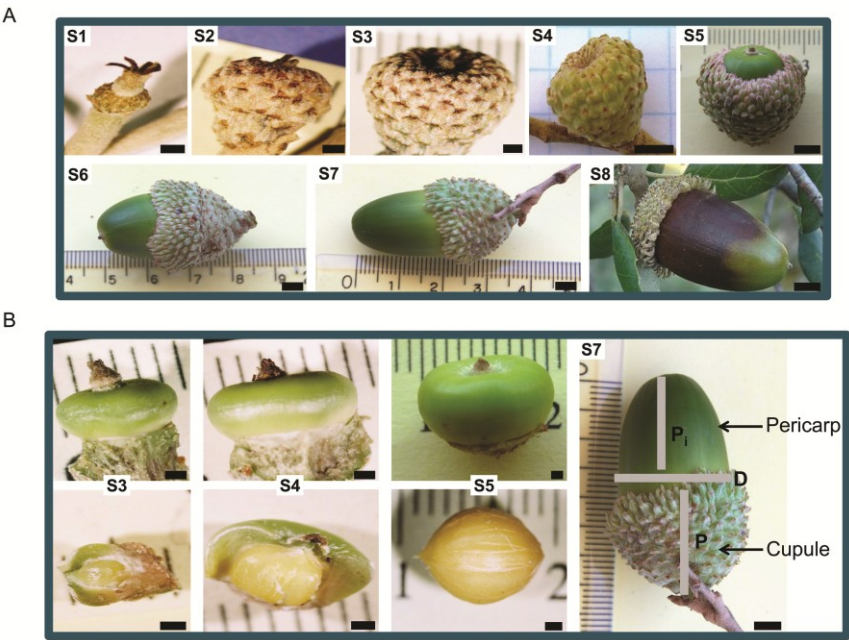


Fig. 1. Developmental stages established for the cork oak acorn. (A) Cork oak fruits collected at different developmental stages (S1-S8). The scale bar corresponds to 1 mm in S1 to S3 and to 5 mm in S4 to S8. (B) Cork oak fruits at stages S3-S5 after removal of the cupule (above), or cupule and pericarp (below) exposing the seed and acorn measurements used in the establishment of the staging system (S7). D , maximum diameter of the acorn; P_i , portion of the acorn outside the cupule; P , acorn portion covered by the cupule. The scale bar corresponds to 1 mm in S3 to S5 and to 5 mm in S7.

Table 2. Criteria used for categorizing the cork oak acorn into different developmental stages and representation of each stage in the normalized (N) and non-normalized (nN) cDNA libraries.

Developmental stage	Max Ø with cupule (mm)	Max Ø without cupule (mm)	Isolated embryos	Other observations/ library type
S1	2 - 3	nd	no	no embryos visible; N and nN
S2	3 - 5	nd	no	multiple embryos, some aborted; N and nN

Table 2. (Continued)

Developmental stage	Max Ø with cupule (mm)	Max Ø without cupule (mm)	Isolated embryos	Other observations/ library type
S3	5 - 8	nd	yes (EM3)	dominant embryo; N and nN
S4	8 - 12	nd	yes (EM4)	acorn completely covered by the cupule; N and nN
S5	12 - 17	7 - 11	yes (EM5)	acorn already visible out of the cupule; N and nN
S6	17 - 20	11 - 16	no	approx. half of the acorn out of the cupule
S7	> 20	> 16	no	acorn out of the cupule and still green; N
S8	> 20	> 16	yes (EM8)	acorn out of the cupule and mostly brown; N and nN

nd: not determined

Sequencing and assembly of the cork oak acorn and embryo transcriptome

The sequencing of the five non-normalized libraries corresponding to samples from stages S1, S2, S3+S4, S5 and S8 aimed at gene expression analysis during acorn development. In addition, two normalized libraries prepared from RNA of cork oak acorns or from isolated embryos were sequenced to favour the detection of rare transcripts and thereby facilitate the assembly. The total number of raw reads was 2,534,447, which had an average length of 513 bp. After pre-processing, 2,088,335 high-quality sequences were conserved and used in the assembly and mapping steps. The final average length of the reads was 215 and 400 bp for the non-normalized and normalized libraries, respectively (Table 3).

Table 3. Cork oak sequenced libraries and number of reads per library before and after pre-processing.

		Non-normalized cDNA libraries					Normalized cDNA libraries	
		Fruit stages					FR	EM
		S1	S2	S3+S4	S5	S8	S1 to S8	EM3 to EM5+EM8
Raw reads	Total	111,703	373,962	200,862	302,253	102,250	738,266	705,151
	Range (bp)	51-1,200	52-1,200	47-1,200	50-1,200	55-1,200	50-1,201	52-1,201
	Mode (bp)	520	521	519	522	519	516	516
	Mean (bp)	503.7	515.8	510.7	524.8	513	538.5	538.7
	Size (Mbp)	47.5	162.5	86.1	133.6	44.3	314.7	298.8
Valid reads	Total	100,081	332,674	179,588	267,171	88,261	572,665	547,895
	Range (bp)	40-568	40-579	40-571	40-565	40-546	40-576	40-616
	Mode (bp)	209	213	220	217	214	409	380
	Mean (bp)	254.1	255.2	253.6	253.5	258.1	304.3	289.2
	Size (Mbp)	27.6	92.7	49.5	74.4	25.5	201.0	183.0

The seven libraries were assembled by MIRA and Newbler (Table 4). MIRA assembly contained 104,862 contigs, 52.2% of which were longer than 500 bp. Newbler assembly contained 33,034 contigs, 79.6% of which were longer than 500 bp. The merging of the MIRA and the Newbler assembly using CAP3 resulted in 80,357 contigs that were deposited in ENA (the accession number of the *de novo* transcriptome is HABZ01000000 and the accession numbers of the contigs are HABZ01000001-HABZ01080357). 62.5% of these contigs were longer than 500 bp.

Table 4. *De novo* transcriptome assemblies and classification of the assembled cork oak transcripts.

	MIRA	Newbler 2.6	Final (Merged)
Contigs	104,862	33,034	80,357
Contigs > 500 bp	54,764 (52.2%)	26,313 (79.6%)	50,197 (62.5%)
Contigs < 200 bp	6,306 (6%)	481 (1.5%)	4,510*
Contigs with homologous in Uniprot	73,103 (69.7%)	26,951 (81.6%)	56,517 (70.3%)
Unique Uniprot IDs	28,154 (38.5%)	16,047 (48.6%)	24,474 (43.3%)
Complete contigs	16,149 (15.4%)	12,953 (39.2%)	19,146 (23.8%)
C-terminus contigs	15,818 (15.1%)	5,342 (16.2%)	11,410 (14.2%)
N-terminus contigs	15,368 (14.7%)	4,064 (12.3%)	10,108 (12.6%)
Internal contigs	25,432 (24.2%)	4,575 (13.8%)	15,509 (19.3%)
Misassembled contigs	336 (0.32%)	17 (0.05%)	344 (0.43%)
Contigs without UNIPROT homologous	31,759 (30.3%)	6,083 (18.40%)	23,840 (29.6%)
Novel genes	5,388 (5.14%)	1,320 (4%)	4,658 (5.8%)
Complete ORF	2,758 (2.6%)	664 (2.01%)	2,318 (2.9%)
Partial ORF	2,630 (2.5%)	656 (1.99%)	2,340 (2.9%)
Unknown contigs	26,346 (25.1%)	4,757 (14.4%)	19,163 (23.8%)
Unknown contigs < 200 bp	5,471 (5.2%)	407 (1.2%)	0*
Putative ncRNAs	25 (0.02%)	6 (0.02%)	19 (0.02%)
Reads mapped**	2,020,921 (96.8%)	1,909,842 (91.5%)	2,009,759 (96.2%)
Unique mapped reads	1,703,996 (84.3%)	1,549,803 (74.2%)	1,491,131 (71.4%)
Duplicated mapped reads	273,619 (13.1%)	270,027 (14.1%)	367,486 (17.6%)
Mapping more than two times	43,306 (2.1%)	90,012 (4.7%)	151,142 (7.2%)
Average coverage	7.8	14.2	8

*Contigs shorter than 200 bp were filtered out before analyzing.

**Over 2,088,230 total reads

The assembled transcripts were classified as complete, terminal, internal or novel by comparison with the complete plant proteins in UniprotKB database (Table 4 and Supplementary File S1). 23,840 contigs did not have any homologous sequence in the tested database (Complete plant Uniprot proteins). However, it was possible to predict a clear ORF for 4,658 of them, and they were classified as novel.

Completeness of the Q. suber transcriptome by comparison with other oaks

The proteins in the *Q. suber* assembly were compared to the proteins from other four *Quercus spp.*, two *Castanea spp.* and a *Fagus sp.*. For this purpose, we obtained the assembled transcriptomes from the *Fagaceae* project (www.fagaceae.org) or NCBI (for *Q. robur* and *Q. petraea*). The transcripts in each transcriptome were classified as complete, terminal, internal or novel by comparison with the complete plant proteins in UniprotKB database, as we had previously done for *Q. suber* (Supplementary Table S1). Our *Q. suber* assembly had the higher number of complete proteins (19,146) and the second higher number of total proteins (56,517).

On average, 94.2% of the proteins from any of the tested species could be found in our *de novo* *Q. suber* transcriptome when it was used as the target database (Supplementary Table S2). On the other hand, when the *Q. suber* proteins were used as query, 66.6% of the *Q. suber* proteins were found in the other transcriptomes. Furthermore, approximately 81% of the *Q. suber* proteins were found in *Castanea mollissima* and *Castanea dentata*. Only 1 protein (Contig 20020) from our assembly was not found in at least one of the other transcriptomes.

Functional annotation of the Q. suber transcriptome

All 80,357 transcripts were compared with the NCBI non-redundant (nr) protein database using Blastx with an E-value of $1e^{-10}$, which resulted in 53,670 sequences with a significant alignment (Supplementary Fig. S2). 19,757 transcripts had the best match to *Vitis vinifera* sequences, followed by 9,329, 8,636 and 5,324 transcripts that matched to *Ricinus communis*, *Populus trichocarpa* and *Glycine max* sequences, respectively. The alignments obtained among the Fagaceae family were very low: 122, 101, 97 and 75 sequences matched with sequences of *Castanea sativa*, *Fagus sylvatica*, *Quercus suber* and *Castanea mollissima*, respectively (Supplementary Fig. S2A). This is mainly due to the limited amount of data available at the GenBank database for non-sequenced species. Most of the alignments showed a similitude between 75 and 90% (Supplementary Fig. S2B).

50,228 transcripts were annotated with at least one Gene Ontology (GO) term (Supplementary File S2). There was a direct relation between the sequence length and percentage of annotated sequences and over 75% of the sequences longer than 1 Kb could be annotated (Supplementary Fig. S2C).

49,945 *Q. suber* transcripts had a homologous in the *A. thaliana* genome (Blastx E-value $< 1e^{-10}$). Each transcript was annotated with the GO terms of its *Arabidopsis* homologous gene. Additionally, each *A. thaliana* gene was annotated with its NCBI COGs, if any exists, and this annotation was also associated backwards to the original *Q. suber* transcript (Supplementary File S3). In order to compare our *de novo* transcriptome and identify COGs specific to *Q. suber*, a similar approach was done for *Q. petraea*, *Q. robur* and the cork oak ESTs database (CODB). 44,300, 59,572 and 51,916 transcripts from *Q. petraea*, *Q. robur* and CODB were homologous to genes from the *A. thaliana* genome, respectively. The

distribution of protein clusters is summarized in a Venn diagram (Supplementary Fig. S3). 2,254 of a total of 3,110 COGs were presented in all the species and 221 COGs were specific to *Q. suber*. Of these 221 COGs, 12% were involved in *replication, recombination and repair*, 6% in *RNA processing and modification and in translation, ribosomal structure and biogenesis*, 5.4% in *cell division and chromosome partition*, 5% in *post-transcription* modifications and 5% in *transcription*. Finally, 20.8% of the 221 COGs were *unknown* or *poorly characterized* (Supplementary File S3 and Supplementary Table S3).

Pathway analysis during cork oak acorn development

15,612 sequences were annotated according to their homology with known enzymes that belonged to 140 KEGG level-3 pathways and all 14 KEGG level-2 pathways (Supplementary File S4). The *carbohydrate metabolism* was the level-2 pathway most represented, which also includes several of the more represented level-3 pathways, such as *starch and sucrose*, *glycolysis and gluconeogenesis*, *amino sugar and nucleotide sugar*, *pyruvate*, and *galactose* metabolic pathways. The second most represented level-2 pathway was *amino acid metabolism*, which includes *phenylalanine* metabolism. When the number of different enzymes is considered, the more relevant pathways were *glycine, serine and threonine*; *arginine and proline* and *cysteine and methionine* pathways. The third most represented pathways were *lipid metabolism* and *energy metabolism*. Among the most represented level-3 pathways were also *purine* and *pyrimidine* metabolism, *methane* metabolism, and *phenylpropanoid biosynthesis* (Supplementary File S4).

The reads from the five non-normalized libraries were mapped to the transcripts in assembly to quantify the expression in each stage. The number of mapped reads of the transcripts belonging to the same pathway

was summed up to determine the expression of each pathway on time (Supplementary File S4). The normalized expression values for the level 2 pathways were represented in a heatmap (Fig. 2). Carbohydrate, nucleotide, glycan and energy metabolic pathways were up-regulated during the middle stages of development (S2 to S5). Up-regulation of the amino acid metabolism was evident in S3S4. Signal transduction pathways were up-regulated at S2, while the lipid, cofactors and vitamins metabolic pathways were specifically up-regulated at S8.

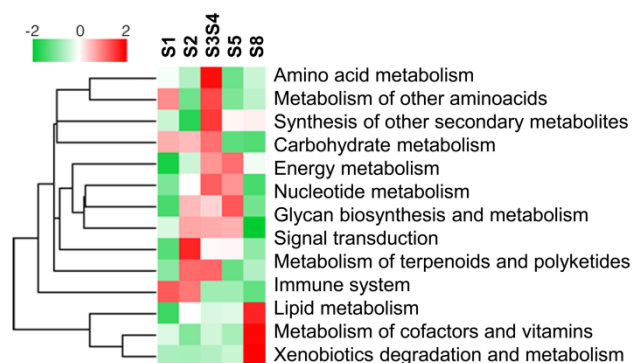


Fig. 2. Heatmap of the expression levels of the KEGG level 2 pathways at the analysed stages of acorn development. The expression levels were normalized in Z scores, with signals from red (higher expression) to green (lower expression).

Differentially expressed genes (DEGs) during cork oak acorn development

From the mapping of the reads of the five non-normalized libraries to the assembly, 58,839 genes were identified as expressed during any of the developmental stages, 7,824 transcripts were expressed in all the stages and 22,802 (38.75%) were specific to one stage. The total number of transcripts present in each stage was 23,104, 37,501, 30,035, 33,367 and 17,310, respectively from S1 to S8 (Supplementary Fig. S4).

Of the 58,839 transcripts expressed during acorn development, 2,285 were considered DE (Supplementary File S5). From those 710, 475,

685 and 1,078 transcripts were DE between stages S2 and S1, S3S4 and S2, S5 and S3S4, and S8 and S5, respectively. 568 transcripts (24.9%) were DE in more than one stage transition (Fig. 3).

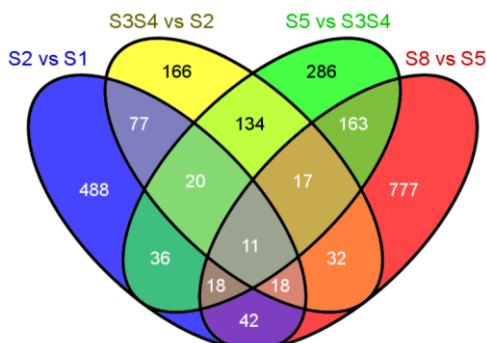


Fig. 3. Venn diagram illustrating the number of transcripts differentially expressed between two consecutive stages of development.

An enrichment analysis by F-fisher test ($FDR < 0.05$) comparing the set of 2,285 DEGs *versus* the complete transcriptome evidenced that 466 GO terms were over-represented (Supplementary File S6). One third of the DEGs were involved in responses to abiotic stimulus, one fifth in carbohydrate catabolism, and other fifth in the catabolism and generation of energetic compounds. GO terms related with transport process, such as water and auxin polar transport, or development and growth were also significantly represented (Supplementary Fig. S5).

DEGs were clustered in six groups according to their expression profile on time (Fig. 4 and Supplementary File S6). An enrichment analysis ($FDR < 0.01$) of the genes in each cluster *versus* the complete transcriptome evidenced the dominant processes in each stage (Supplementary File S6). Eight GO terms were over-represented at S1, including osmotic and salt stress, and hexose transmembrane transport. Forty-one GO terms were over-represented at S2, including osmotic, heat, high temperature and salt stresses, as well as response to water, water

deprivation, water and fluid transport, and transmembrane transport. Seventy-three GO terms were over-represented during S3 and S4, including the previous terms related with response, plus meristem growth and development. Thirty-one GO terms were over-represented at S5, plus 68 GO terms that were over-represented during S5 and S8, including glycogen synthesis and metabolism, carbohydrates (glucose, hexose, pyruvate and glucan) metabolism, as well as starch synthesis and xylem development. Fifty GO terms were over-expressed at S8, including chitin binding and metabolism and aminoglycan, amino sugar and glucosamine catabolism.

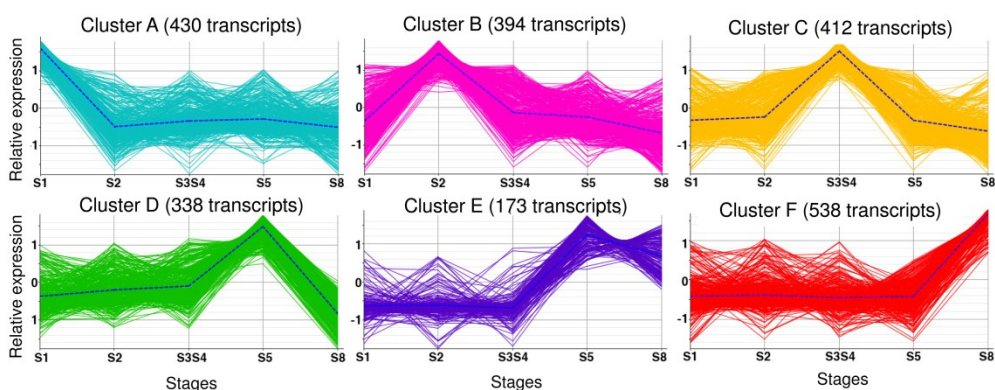


Fig. 4 Clustering analysis of differentially expressed genes (DEGs) according to their expression profiles at different stages of acorn development in cork oak.

Genes related to response to water, water deprivation or water transport

From the total DEGs, 211 were related to water response, deprivation and transport, and distributed across all the developmental stages but over-represented in S2, followed by the last developmental stage (S8) (Supplementary Table S4). About 2/3 of DEGs were expressed in a specific developmental transition, corresponding to approximately 27% in the

transition from S5 to S8, followed by 18% in S1 to S2 and 10% both for transitions from S2 to S3S4 and S3S4 to S5.

In Table 5 a shortlist of these transcripts is presented that includes only those that are specific of a single cluster (except *GoIS2*), are stage-specific, lack expression in one or more stages or show a considerable higher expression in a single stage. Homologous to the *Arabidopsis* *GALACTINOL SYNTHASE 2* (*GoIS2*) was represented in all the acorn developmental stages except S8. Mostly up-regulated in the early stage (S1) and specific of cluster A, was a member of the NAC family, a homolog of *RESPONSIVE TO DESICCATION 26* (*RD26*). In the S2 and S3S4 stages homologs to the *Arabidopsis* β -*AMYLASE 1* (*BAM1*) and to the *PHENYLALANINE AMMONIA-LYASE* (*PAL1*), respectively, were also found highly up-regulated, the latter lacking expression in the full mature acorns (S8). A putative homolog of the *ABSCISIC ACID INSENSITIVE 3* (*ABI3*), exclusive of cluster D, was mostly expressed in the last stages of development, especially in S5, and lacked expression in the early and middle stages of the acorn development. Transcripts putatively encoding for members of Late Embryogenesis Abundant protein family (*LEA* and *DI21*) and for a Lipid Transfer Protein (*LTP3*) were mostly expressed in the last stage of the acorn development (Table 5 and Supplementary Table S4).

Table 5. Shortlist of differentially expressed transcripts annotated as involved in response to water. From the total of 211 DEGs in this category, the ones here represented were selected according to the following criteria: the transcript is either specific of a single cluster (except *Go/S2*), it is stage-specific, it lacks expression in one or more stages or it shows the highest expression in a single stage. * correspond to stages where the transcript is differentially expressed.

Cluster	Transcript name	At Locus	At homologous	Annotation	S1 counts	S2 counts	S3S4 counts	S5 counts	S8 counts
A	Contig16452	AT4G27410	RD26	NAC domain transcriptional regulator superfamily protein	608.8*	184.2	138.6*	32	11
B	Contig18768		BAM1	β -Amylase	1.2	29.6*	0.6	1.2	2.2
	Contig25491				303.6	557.6*	608.8	175.8*	4.6
	Qs-dev_c30725	AT3G23920			0	23.8*	0	9.2	0
	Qs-dev_rep_c76744				0.4	54.2*	0.2	8.4	0.8
C	Contig9321	AT2G37040	PAL1	Phenylalanine ammonia-lyase	26.2	40.4	81.2*	22.2*	0
D	Contig19318	AT3G24650	ABI3	AP2/B3-like transcriptional factor family protein	0	0	7.2	66.8*	42.2
F	Contig16865		DI21	Late embryogenesis abundante protein Lea5; drought-induced 21	48.2*	0	8.2	19.8	371.6*
	Contig16935				1.2	0	3.6	18.8	179.2*
	Contig18905	AT4G15910			25.8	69.8*	7.2	41.6*	312.6*
	Qs-dev_c42419				0	0	0	0	16.4*
F	Contig18471	AT5G59320	LTP3	Lipid transfer protein 3	0	0.2	0	5.6	38.6*
	Contig19993				0	0	0	9.4	56.8*

Table 5. (Continued)

Cluster	Transcript name	At Locus	At homologous	Annotation	S1 counts	S2 counts	S3S4 counts	S5 counts	S8 counts
A	Qs- dev_rep_c77063	AT1G56600	GolS2	Galactinol synthase	43*	0	0.2	0.8	0.8
B	Contig17943				39.2	101.6*	90.4	27.6*	2
B	Contig21230				3	81.2*	92.2	38.6*	2.2
B	Contig21232				8.8	21.4*	12.8	14.6*	0
B	Contig21234				39.6	86.4*	63.4	3.4*	0
B	Contig5853				1.6	30.4*	21	9	0
D	Contig18148				1.2	17.8	19.4	18.8*	0
C	Contig21240				20	24.6	42.8	37.8*	0

Transcription factors differentially expressed during acorn development

Transcription factors have important roles in gene expression due to their ability to bind specific DNA sequences and control transcription by acting as transcriptional activators or repressors. From the 2,285 DEGs during acorn development a total of 498 were annotated as TFs (Supplementary Table S5). These transcripts were almost equally distributed among the different clusters, with cluster A (up-regulation in S1 stage) showing slightly higher number (22,5%) of putative transcription factor transcripts and cluster F (up-regulation in S8 stage) showing the lower number (15,9%). A total of 70% were preferentially up-regulated in a given developmental stage.

A list of selected genes annotated as transcriptional regulators is presented in Table 6 and includes genes that are specific of a single cluster, are stage-specific or belong to specific TF families with well characterized roles in plant development. NAC, bHLH, and OLEOSIN are some of the TF families most represented during the cork oak acorn development. Up-regulation in the early stages of development (S1 and S2) was observed for class II *KNOTTED1*-like homeobox genes. During the late stages of the acorn development transcripts putatively coding for *OLEOSIN* (*OLEO*) were up-regulated. Transcripts coding for NAC and bHLH transcription factor families were found DE across all the studied developmental stages (Table 6 and Supplementary Table S5).

Table 6. Shortlist of differentially expressed transcripts putatively coding for transcription factors. Transcripts were selected according the following criteria: transcript is either specific of a single cluster, it is stage-specific or belongs to well known families. * correspond to stages where the transcript is differentially expressed.

Cluster	Transcript name	At Locus	At homologous	Annotation	S1 counts	S2 counts	S3S4 counts	S5 counts	S8 counts
A	Contig23988	AT5G25220	KNAT3	KNOTTED1-like homeobox	35.6*	0.8	3.8	11.6	13.8
B	Contig17627	AT5G11060	KNAT4		1.2	34*	27.4	5.6	2
E	Contig24728	AT4G25140	OLEO1	Oleosin	0	0	3	69.2*	58.5
	Contig20473				0	0	8.2	428.2*	192.4
E	Contig23293				0	0	0	71*	25.8
F	Qs-dev_rep_c103605	AT3G27660	OLEO4		0.4	0	0	9.2	39.6*
F	Qs-dev_c70574	AT5G40420	OLEO2		0.2	0	0	2.2	49.8*
A	Contig5781	AT4G00870		Basic helix-loop-helix (bHLH) DNA binding superfamily protein	40.6	25.6	18.4*	0.4	2.2
B	Contig1681	AT5G08130	BIM1		0	22.8*	0	1.2	8.8
D	Contig25637	AT1G01260	JAM2		5.2	16.6	7.4	21.6*	0
D	Contig20612	AT1G72210			8.4	27.8	21.6	32*	6.8
E	Contig18945	AT5G41315	GL3, MYC6.2		0	4.6	0.6	0	0
A	Contig4505	AT1G56010	NAC1		34.2*	2.2	0.6	0	0
A	Contig23086	AT1G34190	NAC017	NAC domain	20.2*	0	0	11	4.6
B	Contig23085				0	27.8*	22.8	18.4	7
A	Contig16452	AT4G27410	RD26		608.8*	184.2	138.6*	32	11
B	Contig1296	AT4G35580	NTL9		45.8	47	39.4	41.6*	13
D	Contig26025	AT3G15510	NAC2		1.2	1.6	12	19.8*	0

Validation of the differential expression profiles by RT-qPCR

Several genes were selected to validate the data obtained through the pyrosequencing with 454 Technology (Table 1). Twenty DEGs related to water responses, seven of which also annotated as TFs, were chosen for the validation of expression profiles by reverse transcription quantitative real-time PCR (RT-qPCR). Two transcripts belong to cluster A, six to cluster B, three to C, four transcripts belong to D and five to F. Correlation between the gene expression levels and the profiles obtained in RNA-Seq was demonstrated by Pearson correlation (Fig. 5) with most of the genes showing strong or moderately strong correlation (Chan, 2003).

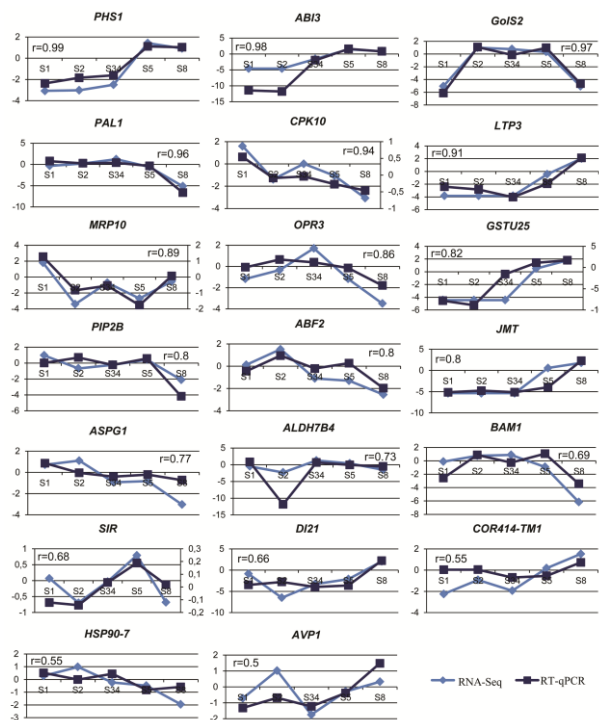


Fig. 5. Validation of the RNA-Seq transcript profiles. Comparison of transcripts expression patterns from RNA-Seq data and from reverse transcription quantitative real-time PCR (RT-qPCR). In the y-axis it is represented the Log_2 of the relative expression level in each developmental stage and the five acorn developmental stages are represented in the x-axis. The numbers above the graphics correspond to the values obtained with the Pearson correlation.

Discussion

Seed protection and dispersal are the main functions of the fruits and adaptations in these traits are important drivers of colonization of new niches, therefore playing a role in plant evolution. In this work we provide a dynamic view of gene expression in developing cork oak acorns using next generation sequencing technology and identified a set of 2,285 differentially expressed genes with roles in a range of biological processes. We then focused our analysis on aspects related to carbohydrate metabolism and on groups of transcripts with putative functions in transcriptional regulation and traits likely relevant to seed survival and dispersal, such as mechanisms related to water response including water transport and water deprivation. Also, it should be pointed out that at maturity, most of the acorn mass consists of seed tissues, mainly cotyledons. Therefore, transcripts identified during late development stages are mostly derived from the seed.

A de novo transcriptome of cork oak acorn and embryo tissues

A *de novo* transcriptome assembly with the data here generated allowed us to identify the transcripts expressed during the acorn developmental process, some of which classified as novel. This *de novo* assembly facilitated the mapping of reads in unique positions since it was not necessary to allow mismatches between reads and reference. In fact, we discarded the marginal number of reads mapping in several positions. Assemblers of 454 transcriptome data have been systematically compared using real and simulated datasets. In such reviews, Newbler (Margulies *et al.*, 2005) and MIRA (Chevreux *et al.*, 2004) outperformed other assemblers (Weber *et al.*, 2007; Kumar and Blaxter, 2010; Garg *et al.*, 2011; Mundry *et al.*, 2012). Newbler usually assembles longer contigs that often cover more than the 80% of the reference sequences. MIRA joins reads in a more conservative way than Newbler, which prevents chimeric contigs and

generates bigger assemblies using more bases and containing higher number of contigs, but some of them are redundant. Kumar et al. (2010) proposed an assembly strategy that was used for the *de novo* assembly of pyrosequencing data from chickpea (Garg *et al.*, 2011), by merging individual assemblies using a traditional Overlap-Layout-Consensus assembler, such as CAP3 (Huang and Madan, 1999). Merged datasets aligned better to reference datasets and were more consistent in the total span and number and size of contigs than individual assemblies. In our case, the number of complete contigs (19,146) was higher in the merged assembly than in the individual ones. On the other hand, the percentage of C-terminal and N-terminal contigs was smaller in the merged assembly than in any of the original assemblies, which supports that several contigs from the same transcript were merged. When compared with other Fagaceae transcriptomes, the number of total and complete proteins in this *Q. suber* transcriptome assembly evidenced the advantages of this strategy.

Carbohydrate metabolism throughout cork oak acorn development

Carbohydrate metabolism was the level-2 pathway most represented during cork oak fruit development. The enrichment analysis performed in the different clusters evidenced also the timing when a specific metabolic process appears prevalent. Using this approach, carbohydrates metabolism as well as starch synthesis, were found over-represented in late stages of fruit development, both S5 and S8. However, specific processes like hexose transmembrane transporters were found over-represented in early stages of acorn development, where actively dividing cells contribute to a rapid growth of the fruit. In general, hexoses favour cell division and expansion, whereas sucrose favours differentiation and maturation (Koch, 2004). Cork oak transcripts homologous to the *Arabidopsis* hexose transporter genes *STP13* and *STP1* were up-regulated at early stages of fruit development, and

almost absent at complete maturation. The same was observed for a homolog to the *Arabidopsis* *GLUCOSE 6-PHOSPHATE/PHOSPHATE TRANSLOCATOR 1* (*GPT1*), which is essential for pollen maturation and embryo sac development in response to glucose and sucrose stimulus (Gómez *et al.*, 2006; Andriotis *et al.*, 2010a). However, a transcript homologous to *GPT2*, was up-regulated during S8. Interestingly, while *GTP1* is the major GPT responsible for the transport of Glc6P into plastids of heterotrophic tissues in *Arabidopsis* (Niewiadomski *et al.*, 2005), Dyson *et al.* (2014) found that *GPT2* plays a crucial role in determining the response of seedlings to exogenous sugars during their establishment. These authors identified *GPT2* as a transporter involved in sugar sensing and signaling to ensure appropriate responses to changes in sugar status, supporting that *GPT2* is necessary for the correct responses to a number of changes in carbon metabolism in *Arabidopsis* (Kunz *et al.*, 2010). Therefore, if the cork oak transcripts have a role similar to their *Arabidopsis* homologs, our results are consistent with expected changes in sugar status along fruit development.

During the middle stages of acorn development, several DEGs homologous to *SUCROSE SYNTHASES* were identified. Sucrose represents the main form of carbon transported in plants, while starch is an important carbon reserve. In *Arabidopsis*, a multigene family encodes six *SUCROSE SYNTHASES* (*SUS*) isoforms required for cellulose and starch metabolism (Baroja-Fernández *et al.*, 2012), where *SUS3* is fundamental for late seed maturation (Tai *et al.*, 2005; Falcone *et al.*, 2007). In cork oak, several DEGs homologous to *SUCROSE SYNTHASE 3* (*SUS3*) and *SUCROSE SYNTHASE 4* (*SUS4*) were up-regulated between S2 and S5, highlighting the relevance of carbon reserves synthesis during the middle developmental stages.

Additionally, transcripts putatively coding for amylases were also highly represented during fruit development. Amylases are Carbohydrate-

Active Enzymes (CAZymes) involved in starch degradation. While some of the amylase transcripts, e.g. homologous to the *A. thaliana* chloroplast β -AMYLASEs *BAM1/BAMY7*, were up-regulated from the early to middle stages, other transcripts with homology to *BAM5* or *BAM6* were up-regulated in later stages up to S8 being almost absent in earlier stages. The significance of these divergent expression patterns for different homologs is not currently clear. In poplar, amylases were predominant in storage tissues and seeds (Geisler-lee *et al.*, 2006) while *BAM1* was found to be involved in *Arabidopsis* pollen germination (Wang *et al.*, 2008) and oilseed embryo development (Andriotis *et al.*, 2010b). During rice grain germination and seedling growth a tight temporal and spatial regulation of α -amylase expression controls rates of sugar production in embryos and endosperms, which is balanced between energy supply (source) and seedling development (sink) (Chen *et al.*, 2006). However, it is not known if β -amylases are also involved in these control mechanisms. In any case, alterations in the availability of soluble sugars are crucial in the regulation of plant development and growth and amylases play a role in this process. Several evidences have shown that sugars interact with phytohormone signalling during plant development (Gibson, 2004) and one of the first known examples of these interactions showed that expression of some α -amylase genes are induced by gibberellins (reviewed in Bethke *et al.* (1997)). Additionally, the accumulation of soluble sugars plays an important role in the acquisition of desiccation tolerance in seeds (Moore *et al.*, 2009).

Response to water across acorn development

The natural shedding of cork oak acorns coincides with complete maturity and acorns left on the ground after shedding will either germinate or lose their viability as a result of desiccation (Merouani *et al.*, 2003). The seeds contain very little solid endosperm, while the embryo is large and fleshy and

has a high water content. Since acorns are shed at high moisture content (Merouani et al 2003), tolerance to desiccation may play an important role in cork oak regeneration success, especially when considering the climate changes predicted for the Mediterranean region in the near future (Christensen *et al.*, 2007; Giorgi and Lionello, 2008). We have identified the DEGs annotated as being related to water responses during acorn development although most of them are not fruit specific, as also observed in *Arabidopsis* (Bray, 2004; Le *et al.*, 2010). A high number of DEGs in this category were identified at the last stage of acorn development corresponding to maturity, probably reflecting a reduction in water content at this stage, but also during the early developmental stages (S2). In addition to the above referred transcripts homologous to β -*AMYLASE* genes such as *BAM1*, described as contributing to osmoprotection during desiccation (Kaplan and Guy 2004), other transcripts strongly related to desiccation were identified in early fruit development. For instance, an homolog of *RESPONSIVE TO DESICCATION 26 (RD26)*, which is a dehydration responsive gene identified as highly responsive to abiotic stress (Tran *et al.*, 2004; Fujita *et al.*, 2004; Singh *et al.*, 2013), was found highly expressed in the early and middle stages of the acorn development and specifically associated to one of the expression profiles (cluster A).

In the late stages of fruit development, well characterized genes related to acquisition of tolerance to seed desiccation were found, such as an homolog to the *ABA INSENSITIVE 3 (ABI3)* essentially up-regulated at S5, encoding a TF expressed in seeds that mediates abscisic acid (ABA) responses. In *Arabidopsis*, *ABI3* is necessary during seed maturation for the accumulation of seed storage proteins, onset and dormancy maintenance and for the acquisition of seed desiccation tolerance (Parcy *et al.*, 1994; Liu *et al.*, 2013). Our results seem in agreement with a tight regulation of the mechanisms controlled by *ABI3* during cork oak seed development towards maturation, such as ABA responsiveness and synthesis and accumulation

of proline in order to increase stress tolerance in the embryo (Hare *et al.*, 1999). In the recalcitrant seeds (desiccation-sensitive) of *Quercus robur* the concentration of ABA, shown to play a role in the acquisition of desiccation tolerance and reserve accumulation, increases toward the end of embryogenesis and reaches its maximum during the intermediate stage of maturation phase, followed by a decrease in the late maturation phase (Finch-Savage and Clay, 1994; Prewin *et al.*, 2005). The expression profile of *ABI3* in cork oak acorns, with no expression in the early stages of development, low expression levels in S3S4 and maximum expression during S5 followed by a decrease in S8, seems consistent with alterations in ABA content and corresponding ABA-induced processes along development as described for *Q. robur*.

During the last stage of acorn development (S8, cluster F) two transcripts protrude from the list of DEGs annotated as related to water responses, an homolog of *LIPID TRANSFER PROTEIN 3* (*LTP3*) and an homolog of the *DROUGHT-INDUCED 21* (*DI21*). Their specificity to the late stages of acorn development is in accordance with the literature. The *Arabidopsis LTP3* was highly expressed in mature siliques and induced by drought, thus positively regulating the plant response to drought stress through the transcriptional activation of the cuticular wax biosynthesis as a response to cellular dehydration (Seo *et al.*, 2011; Guo *et al.*, 2013). The protein family to which *DI21* belongs, the Late Embryogenesis Abundant (LEA) proteins, is characteristically accumulated in the last stages of seed development when tolerance to desiccation is required (Olvera-Carrillo *et al.*, 2010). Additionally, *DI21* is known to be up-regulated under abiotic stresses (Hundertmark and Hinch, 2008). Accumulation of LEA proteins during seed maturation and in response to altered water status was previously observed in oak species (Sunderlíková and Wilhelm, 2002; Prewin *et al.*, 2005; Sunderlíková *et al.*, 2009). However, their effect in

increasing drought tolerance in oaks is not as extensive as in non-recalcitrant seeds (Finch-Savage and Clay, 1994).

Another worth mentioning transcript putatively related to desiccation tolerance is a homolog of the *GALACTINOL SYNTHASE* (*GoIS2*), which codes for a key enzyme in the synthesis of galactinol. This protein is the first committed enzyme in the biosynthesis of raffinose family oligosaccharides (RFOs), which is a highly specialized metabolic event in higher plants implicated in attenuating the effects of environmental stresses (Downie *et al.*, 2003). Accumulation of galactinol and RFOs was observed during seed maturation in *Arabidopsis*, soybean and maize, indicating that these sugars play a role in the desiccation tolerance of seeds as osmoprotectants (Blackman *et al.*, 1992; Brenac *et al.*, 1997; Taji *et al.*, 2002). As in the soybean (Blackman *et al.*, 1992), also during the development of the cork oak acorn galactinol synthase decreases as seeds mature being absent in stage S8.

Transcription factors during acorn development

Transcriptional regulators are crucial for plant developmental processes through their function in the regulation of gene expression and fruit development is no exception. In *Arabidopsis* which, like cork oak, has dry fruits, it has been revealed that the core and extended genetic network controlling fruit development consists entirely of interactions among transcription factors (Seymour *et al.*, 2013). In our analysis, from the 498 DE putative TF genes identified during acorn development, we highlighted a few transcripts due to their specificity to a developmental stage or expression profile or the well characterized involvement during this phase of plant lifecycle.

Two putative class II *KNOTTED1*-like homeobox (*KNOX2*) genes were up-regulated in the early stages of acorn development, one in S1

(*KNAT3*) and the other one in S2 (*KNAT4*). *KNAT3* was previously reported to have a role in seed development, specifically in embryo sac development and during megagametogenesis (Pagnussat *et al.*, 2007). In a more recent study conducted in the moss *Physcomitrella patens* by Sakakibara *et al.* (2013) KNOX2 TFs were shown to have a critical role in establishing an alternation of generations in land plants by preventing the haploid-specific body plan from developing in the diploid plant body, which appears consistent with the up-regulated expression of cork oak putative *KNOX2* genes in early acorn developmental stages.

On the other hand, *OLEOSINS* (*OLEO1* and *OLEO4*) were only up-regulated in the last stages of the acorn development (S5 and S8). This is in accordance with the reported function of these plant specific genes in other plant species. *OLEOSINS* have a role in the control of oil body structure and accumulation of seed reserves, affecting seed germination and embryo phenotypes (Siloto *et al.*, 2006; Miquel *et al.*, 2014). Recent work in different species showed an increase in their accumulation during seed development (Li and Fan, 2009; Popluechai *et al.*, 2011; Cao *et al.*, 2014).

Several members of two major families of TFs were present in all stages of the acorn development, bHLH and NAC. bHLH proteins are the second largest class of TFs in plants and in *Arabidopsis* they are known to be involved in the regulation of fruit dehiscence, carpel and anther development, flavonoid biosynthesis, stress responses, among others (Feller *et al.*, 2011). As an example of a putative bHLH TF gene, a homolog of the *Arabidopsis* bHLH *GLABRA3* (*GL3*), was up-regulated during the late stages of the acorn development (S5 and S8) when the acorn pericarp starts to become visible out of the cupule and turning brown. If the function of this transcript is conserved, then it is tempting to speculate that its up-regulation during late acorn development is related to the regulation of anthocyanin biosynthesis (Zhang *et al.*, 2003; Feller *et al.*, 2011) that occurs during this phase.

Several transcripts putatively belonging to the NAC family were also present along different stages of the acorn development. This family of TFs plays important roles in responses to plant biotic and abiotic stress (Nakashima *et al.*, 2012; Nuruzzaman *et al.*, 2013) but also in other developmental processes such as seed and embryo development (Olsen *et al.*, 2005; Nuruzzaman *et al.*, 2013). Our data revealed three different up-regulated transcripts coding for the NAC protein family at the S1 developmental stage, two at S2 and S5 and one in both S5 and S8. One of the transcripts up-regulated in S1, *RESPONSIVE TO DESSICATION 26* (*RD26*), was previously described as being induced by drought but also by abscisic acid (ABA) functioning as a transcriptional activator in ABA-inducible gene expression under abiotic stress in plants (Tran *et al.*, 2004; Fujita *et al.*, 2004).

Conclusions

In summary, our analysis allowed to cluster transcripts differentially expressed along acorn development in different profiles showing up-regulation in specific stages of development. While the DE transcripts putatively coding for transcription factors associated to several biological processes were found almost equally distributed throughout the analysed developmental stages, other transcripts involved in specific processes such as response to water or carbohydrate metabolism were over-represented in particular stages. Future functional analysis of genes of interest identified in this work will be important to devise successful strategies for regeneration and breeding of this important species. Additionally, this dataset significantly contributes to increase the still scarce information on cork oak genomics providing tools for further molecular dissection of cork oak biology.

References

- Ando K, Carr KM, Grumet R.** 2012. Transcriptome analyses of early cucumber fruit growth identifies distinct gene modules associated with phases of development. *BMC Genomics* **13**, 518.
- Andriotis VME, Pike MJ, Bunnewell S, Hills MJ, Smith AM.** 2010a. The plastidial glucose-6-phosphate/phosphate antiporter GPT1 is essential for morphogenesis in *Arabidopsis* embryos. *The Plant Journal* **64**, 128–139.
- Andriotis VME, Pike MJ, Kular B, Rawsthorne S, Smith AM.** 2010b. Starch turnover in developing oilseed embryos. *The New Phytologist* **187**, 791–804.
- Aronson J, Pereira JS, Pausas JG.** 2009. *Cork Oak Woodlands on the Edge: Ecology, Adaptive Management, and Restoration*. In: (Aronson J, Pereira JS, and Pausas JG, eds.). Washington DC: Island Press.
- Baroja-Fernández E, Muñoz FJ, Li J, Bahaji A, Almagro G, Montero M, Etcheberria E, Hidalgo M, Sesma MT, Pozueta-Romero J.** 2012. Sucrose synthase activity in the *sus1/sus2/sus3/sus4 Arabidopsis* mutant is sufficient to support normal cellulose and starch production. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 321–326.
- Battke F, Symons S, Nieselt K.** 2010. Mayday--integrative analytics for expression data. *BMC Bioinformatics* **11**, 121.
- Belletti P.** 2001. Seed and seedling management in cork-oak. The research of European community. *Management and long term storage of cork-oak acorns. FAIR5-CT97-3480*. 1–9.
- Belletti P, Monteleone I, Cartarasa M.** 2004. Cork-oak and stone-oak seed management aimed at implementing nursery production of seedlings. In: Ciccarese L, Lucci S, Mattsson A, eds. *Nursery production and stand establishment of broad-leaves to promote sustainable forest management*. Rome, 21–24.
- Bethke PC, Schuurink R, Jones RL.** 1997. Hormonal signalling in cereal aleurone. *Journal of Experimental Botany* **48**, 1337–1356.
- Blackman SA, Obendorf RL, Leopold AC.** 1992. Maturation proteins and sugars in desiccation tolerance of developing soybean seeds. *Plant Physiology* **100**, 225–230.
- Boavida LC, Varela C, Feijó JA.** 1999. Sexual reproduction in the cork oak (*Quercus suber* L.). I. The progamic phase. *Sexual Plant Reproduction* **11**, 347–353.

Boavida LC, Silva JP, Feijó JA. 2001. Sexual reproduction in the cork oak (*Quercus suber* L.). II . Crossing intra- and interspecific barriers. Sexual Plant Reproduction **14**, 143–152.

Bonito A, Varone L, Gratani L. 2011. Relationship between acorn size and seedling morphological and physiological traits of *Quercus ilex* L. from different climates. Photosynthetica **49**, 75–86.

Borgardt S, Nixon K. 2003. A comparative flower and fruit anatomical study of *Quercus acutissima*, a biennial-fruitle oak from the *Cerris* group (Fagaceae). American Journal of Botany **90**, 1567–1584.

Bray EA. 2004. Genes commonly regulated by water-deficit stress in *Arabidopsis thaliana*. Journal of Experimental Botany **55**, 2331–2341.

Brenac P, Horbowicz M, Downer SM, Dickerman AM, Smith ME, Obendorf RL. 1997. Raffinose accumulation related to desiccation tolerance during maize (*Zea mays* L.) seed development and maturation. Journal of Plant Physiology **150**, 481–488.

Bugalho MN, Caldeira MC, Pereira JS, Aronson J, Pausas JG. 2011. Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. Frontiers in Ecology and the Environment **9**, 278–286.

Cao H, Zhang L, Tan X, Long H, Shockey JM. 2014. Identification, classification and differential expression of oleosin genes in tung tree (*Vernicia fordii*). PLoS ONE **9**, e88409.

Chan YH. 2003. Biostatistics 104: Correlational Analysis. Singapore Medical Journal **44**, 614–619.

Chen P-W, Chiang C-M, Tseng T-H, Yu S-M. 2006. Interaction between rice MYBGA and the gibberellin response element controls tissue-specific sugar sensitivity of alpha-amylase genes. The Plant Cell **18**, 2326–2340.

Chevreur B, Pfisterer T, Drescher B, Driesel AJ, Müller WEG, Wetter T, Suhai S. 2004. Using the miraEST assembler for reliable and automated mRNA transcript assembly and SNP detection in sequenced ESTs. Genome Research **14**, 1147–1159.

Christensen JH, Hewitson B, Busuioc A, Chen A, Gao X, Held I, Jones R, Kolli RK, Kwon WT, Laprise R, et al. 2007. Regional Climate Projections. In: Fu C, Giorgi F, eds. *Climate change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 872–879.

- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M.** 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676.
- Derory J, Léger P, Garcia V, Schaeffer J, Hauser M-T, Salin F, Luschign C, Plomion C, Glössl J, Kremer A.** 2006. Transcriptome analysis of bud burst in sessile oak (*Quercus petraea*). *The New Phytologist* **170**, 723–738.
- Díaz-Fernández PM, Climent J, Gil L.** 2004. Biennial acorn maturation and its relationship with flowering phenology in Iberian populations of *Quercus suber*. *Trees* **18**, 615–621.
- Downie B, Gurusinghe S, Dahal P, Thacker RR, Snyder JC, Nonogaki H, Yim K, Fukunaga K, Alvarado V, Bradford KJ.** 2003. Expression of a *GALACTINOL SYNTHASE* gene in tomato seeds is up-regulated before maturation desiccation and again after imbibition whenever radicle protrusion is prevented. *Plant Physiology* **131**, 1347–1359.
- Durand J, Bodénès C, Chancerel E, Frigerio J-M, Vendramin G, Sebastiani F, Buonamici A, Gailing O, Koelewijn H-P, Villani F, et al.** 2010. A fast and cost-effective approach to develop and map EST-SSR markers: oak as a case study. *BMC Genomics* **11**, 570.
- Dyson BC, Webster RE, Johnson GN.** 2014. GPT2: a glucose 6-phosphate/phosphate translocator with a novel role in the regulation of sugar signalling during seedling development. *Annals of Botany* **113**, 643–652.
- Elena-Rosello J, Rio J, Valdecantos JG, Santamaria I.** 1993. Ecological aspects of the floral phenology of the cork-oak (*Q suber* L): why do annual and biennial biotypes appear? *Annals of Forest Science* **50**, 114–121.
- Faivre Rampant P, Lesur I, Boussardon C, Bitton F, Martin-Magniette M-L, Bodénès C, Le Provost G, Bergès H, Fluch S, Kremer A, et al.** 2011. Analysis of BAC end sequences in oak, a keystone forest tree species, providing insight into the composition of its genome. *BMC Genomics* **12**, 292.
- Falcone A, Nelissen H, Fleury D, Van Lijsebettens M, Bitonti MB.** 2007. Cytological investigations of the *Arabidopsis thaliana elo1* mutant give new insights into leaf lateral growth and elongator function. *Annals of Botany* **100**, 261–270.
- Falgueras J, Lara AJ, Fernández-Pozo N, Cantón FR, Pérez-Trabado G, Claros MG.** 2010. SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. *BMC Bioinformatics* **11**, 38.
- Fei Z, Tang X, Alba RM, White JA, Ronning CM, Martin GB, Tanksley SD, Giovannoni JJ.** 2004. Comprehensive EST analysis of tomato and comparative genomics of fruit ripening. *The Plant Journal* **40**, 47–59.

Feller A, Machemer K, Braun EL, Grotewold E. 2011. Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *The Plant Journal* **66**, 94–116.

Finch-Savage WE, Clay HA. 1994. Evidence that ethylene, light and abscisic acid interact to inhibit germination in the recalcitrant seeds of *Quercus robur* L. *Journal of Experimental Botany* **45**, 1295–1299.

Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, Tran L-SP, Yamaguchi-Shinozaki K, Shinozaki K. 2004. A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *The Plant Journal* **39**, 863–876.

Garg R, Patel RK, Jhanwar S, Priya P, Bhattacharjee A, Yadav G, Bhatia S, Chattopadhyay D, Tyagi AK, Jain M. 2011. Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. *Plant Physiology* **156**, 1661–1678.

Geisler-lee J, Geisler M, Coutinho PM, Segerman B, Nishikubo N, Takahashi J, Aspeborg H, Djerbi S, Master E, Andersson-gunnera S, et al. 2006. Poplar carbohydrate-active enzymes. Gene identification and expression analyses. *Plant Physiology* **140**, 946–962.

Gibson SI. 2004. Sugar and phytohormone response pathways: navigating a signalling network. *Journal of Experimental Botany* **55**, 253–264.

Giorgi F, Lionello P. 2008. Climate change projections for the Mediterranean region. *Global and Planetary Change* **63**, 90–104.

Goes F, Iandolino A, Al-kayal F, Bohlmann MC, Cushman MA, Lim H, Ergul A, Figueroa R, Kabuloglu EK, Osborne C, et al. 2005. Characterizing the grape transcriptome. Analysis of expressed sequence tags from multiple *Vitis* species and development of a compendium of gene expression during berry development. *Plant Physiology* **139**, 574–597.

Gómez JM. 2004. Bigger is not always better: conflicting selective pressures on seed size in *Quercus ilex*. *Evolution; International Journal of Organic Evolution* **58**, 71–80.

Gómez LD, Baud S, Gilday A, Li Y, Graham IA. 2006. Delayed embryo development in the *ARABIDOPSIS* *TREHALOSE-6-PHOSPHATE SYNTHASE 1* mutant is associated with altered cell wall structure, decreased cell division and starch accumulation. *The Plant Journal* **46**, 69–84.

Guo L, Yang H, Zhang X, Yang S. 2013. *Lipid transfer protein 3* as a target of MYB96 mediates freezing and drought stress in *Arabidopsis*. *Journal of Experimental Botany* **64**, 1755–1767.

- Hare PD, Cress WA, Van Staden J.** 1999. Proline synthesis and degradation: a model system for elucidating stress-related signal transduction. *Journal of Experimental Botany* **50**, 413–434.
- Hellemans J, Mortier G, De Paepe A, Speleman F, Vandesompele J.** 2007. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**, R19.
- Huang X, Madan A.** 1999. CAP3: A DNA sequence assembly program. *Genome Research* **9**, 868–877.
- Hundertmark M, Hinch DK.** 2008. LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* **9**, 118.
- Janssen BJ, Thodey K, Schaffer RJ, Alba R, Balakrishnan L, Bishop R, Bowen JH, Crowhurst RN, Gleave AP, Ledger S, et al.** 2008. Global gene expression analysis of apple fruit development from the floral bud to ripe fruit. *BMC Plant Biology* **8**, 16.
- Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al.** 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.
- Koch K.** 2004. Sucrose metabolism: regulatory mechanisms and pivotal roles in sugar sensing and plant development. *Current Opinion in Plant Biology* **7**, 235–246.
- Kumar S, Blaxter ML.** 2010. Comparing de novo assemblers for 454 transcriptome data. *BMC Genomics* **11**, 571.
- Kunz HH, Häusler RE, Fettke J, Herbst K, Niewiadomski P, Gierth M, Bell K, Steup M, Flügge U-I, Schneider A.** 2010. The role of plastidial glucose-6-phosphate/phosphate translocators in vegetative tissues of *Arabidopsis thaliana* mutants impaired in starch biosynthesis. *Plant Biology* **12 Suppl 1**, 115–128.
- Le BH, Cheng C, Bui AQ, Wagmaister JA, Henry KF, Pelletier J, Kwong L, Belmonte M, Kirkbride R, Horvath S, et al.** 2010. Global analysis of gene activity during *Arabidopsis* seed development and identification of seed-specific transcription factors. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 8063–8070.
- Leishman MR, Wright IJ, Moles AT, Westoby M.** 2000. The Evolutionary Ecology of Seed Size. *Plant ecological strategies: some leading dimensions*. Sydney, 31–58.

Lesur I, Durand J, Sebastiani F, Gyllenstrand N, Bodénès C, Lascoux M, Kremer A, Vendramin GG, Plomion C. 2011. A sample view of the pedunculate oak (*Quercus robur*) genome from the sequencing of hypomethylated and random genomic libraries. *Tree Genetics & Genomes* **7**, 1277–1285.

Li DD, Fan YM. 2009. Cloning, characterisation, and expression analysis of an oleosin gene in coconut (*Cocos nucifera* L.). *Journal of Horticultural Science & Biotechnology* **84**, 483–488.

Liu X, Zhang H, Zhao Y, Feng Z, Li Q, Yang H-Q, Luan S, Li J, He Z-H. 2013. Auxin controls seed dormancy through stimulation of abscisic acid signaling by inducing ARF-mediated *ABI3* activation in *Arabidopsis*. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 15485–15490.

Livak KJ, Schmittgen TD. 2001. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-(\Delta\Delta CT)}$ Method. *Methods* **25**, 402–408.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380.

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.

Merouani H, Apolinário L, Almeida M, Pereira J. 2003. Morphological and physiological maturation of acorns of cork oak (*Quercus suber* L.). *Seed Science and Technology* **31**, 111–124.

Miquel M, Trigui G, D'Andréa S, Kelemen Z, Baud S, Berger A, Deruyffelaere C, Trubuil A, Lepiniec L, Dubreucq B. 2014. Specialization of oleosins in OB dynamics during seed development in *Arabidopsis thaliana* seeds. *Plant Physiology* **164**, 1866–1878.

Moles AT, Westoby M. 2004. Seedling survival and seed size: a synthesis of the literature. *Journal of Ecology* **92**, 372–383.

Moore S. 2002. Use of genomics tools to isolate key ripening genes and analyse fruit maturation in tomato. *Journal of Experimental Botany* **53**, 2023–2030.

Moore JP, Le NT, Brandt WF, Driouich A, Farrant JM. 2009. Towards a systems-based understanding of plant desiccation tolerance. *Trends in Plant Science* **14**, 110–117.

Mundry M, Bornberg-Bauer E, Sammeth M, Feulner PGD. 2012. Evaluating characteristics of de novo assembly software on 454 transcriptome data: a simulation approach. *PLoS ONE* **7**, e31410.

Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K. 2012. NAC transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta* **1819**, 97–103.

Niewiadomski P, Knappe S, Geimer S, Fischer K, Schulz B, Unte US, Rosso MG, Ache P, Flu U. 2005. The *Arabidopsis* plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. *The Plant Cell* **17**, 760–775.

Nuruzzaman M, Sharoni AM, Kikuchi S. 2013. Roles of NAC transcription factors in the regulation of biotic and abiotic stress responses in plants. *Frontiers in Microbiology* **4**, 248.

Olsen AN, Ernst HA, Leggio L Lo, Skriver K. 2005. NAC transcription factors: structurally distinct, functionally diverse. *Trends in Plant Science* **10**, 79–87.

Olvera-Carrillo Y, Campos F, Reyes JL, Garcíarrubio A, Covarrubias AA. 2010. Functional analysis of the group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis*. *Plant Physiology* **154**, 373–390.

Pagnussat GC, Yu H-J, Sundaresan V. 2007. Cell-fate switch of synergid to egg cell in *Arabidopsis* eostre mutant embryo sacs arises from misexpression of the BEL1-like homeodomain gene *BLH1*. *The Plant Cell* **19**, 3578–3592.

Parcy F, Valon C, Raynal M, Gaubier-Comella P, Delseny M, Giraudat J. 1994. Regulation of gene expression programs during *Arabidopsis* seed development: roles of the *ABI3* locus and of endogenous abscisic acid. *The Plant Cell* **6**, 1567–1582.

Park S, Sugimoto N, Larson MD, Beaudry R, Nocker S Van. 2006. Identification of genes with potential roles in apple fruit development and biochemistry through large-scale statistical analysis of expressed sequence tags. *Plant Physiology* **141**, 811–824.

Pausas JG, Marañón T, Caldeira M, Pons J. 2009. Natural Regeneration. In: Aronson J, Pereira JS, Pausas JG, eds. *Cork Oak Woodlands on the Edge*. Island Press, 115–125.

Peng FY, Reid KE, Liao N, Schlosser J, Lijavetzky D, Holt R, Martínez Zapater JM, Jones S, Marra M, Bohlmann J, et al. 2007. Generation of ESTs in *Vitis vinifera* wine grape (Cabernet Sauvignon) and table grape (Muscat Hamburg) and

discovery of new candidate genes with potential roles in berry development. *Gene* **402**, 40–50.

Pereira-Leal JB, Abreu IA, Alabaça CS, Almeida MH, Almeida P, Almeida T, Amorim MI, Araújo S, Azevedo H, Badia A, et al. A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. 2014. *BMC Genomics* **15**, 371.

Pfaffl MW. 2001. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Research* **29**, e45.

Pons J, Pausas JG. 2012. The coexistence of acorns with different maturation patterns explains acorn production variability in cork oak. *Oecologia* **169**, 723–731.

Popluechai S, Froissard M, Jolivet P, Breviario D, Gatehouse AMR, O'Donnell AG, Chardot T, Kohli A. 2011. *Jatropha curcas* oil body proteome and oleosins: L-form *JcOle3* as a potential phylogenetic marker. *Plant Physiology and Biochemistry* **49**, 352–356.

Prewein C, Endemann M, Reinöhl V, Salaj J, Sunderlikova V, Wilhelm E. 2005. Physiological and morphological characteristics during development of pedunculate oak (*Quercus robur* L.) zygotic embryos. *Trees* **20**, 53–60.

Ramírez-Valiente JA, Lorenzo Z, Soto A, Valladares F, Gil L, Aranda I. 2009a. Elucidating the role of genetic drift and natural selection in cork oak differentiation regarding drought tolerance. *Molecular ecology* **18**, 3803–3815.

Ramírez-Valiente JA, Valladares F, Gil L, Aranda I. 2009b. Population differences in juvenile survival under increasing drought are mediated by seed size in cork oak (*Quercus suber* L.). *Forest Ecology and Management* **257**, 1676–1683.

Rowland LJ, Alkharouf N, Darwish O, Ogden EL, Polashock JJ, Bassil N V, Main D. 2012. Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. *BMC Plant Biology* **12**, 46.

Sakakibara K, Ando S, Yip HK, Tamada Y, Hiwatashi Y, Murata T, Deguchi H, Hasebe M, Bowman JL. 2013. KNOX2 genes regulate the haploid-to-diploid morphological transition in land plants. *Science* **339**, 1067–1070.

Santamaría ME, Rodríguez R, Cañal MJ, Toorop PE. 2011. Transcriptome analysis of chestnut (*Castanea sativa*) tree buds suggests a putative role for epigenetic control of bud dormancy. *Annals of Botany* **108**, 485–498.

- Seo PJ, Lee SB, Suh MC, Park M-J, Go YS, Park C-M.** 2011. The MYB96 transcription factor regulates cuticular wax biosynthesis under drought conditions in *Arabidopsis*. *The Plant Cell* **23**, 1138–1152.
- Seymour GB, Østergaard L, Chapman NH, Knapp S, Martin C.** 2013. Fruit development and ripening. *Annual Review of Plant Biology* **64**, 219–241.
- Siloto RMP, Findlay K, Lopez-villalobos A, Yeung EC, Nykiforuk CL, Moloney MM.** 2006. The accumulation of oleosins determines the size of seed oilbodies in *Arabidopsis*. *The Plant Cell* **18**, 1961–1974.
- Silva SP, Sabino MA, Fernandes EM, Correlo VM, Boesel LF, Reis RL.** 2005. Cork: properties, capabilities and applications. *International Materials Reviews* **50**, 345–365.
- Singh AK, Sharma V, Pal AK, Acharya V, Ahuja PS.** 2013. Genome-wide organization and expression profiling of the NAC transcription factor family in potato (*Solanum tuberosum* L.). *DNA Research* **20**, 403–423.
- Stewart R, Rascón CA, Tian S, Nie J, Barry C, Chu L-F, Ardalani H, Wagner RJ, Probasco MD, Bolin JM, et al.** 2013. Comparative RNA-seq analysis in the unsequenced axolotl: the oncogene burst highlights early gene expression in the blastema. *PLoS Computational Biology* **9**, 1–15.
- Sunderlíková V, Salaj J, Kopecky D, Salaj T, Wilhem E, Matusíková I.** 2009. Dehydrin genes and their expression in recalcitrant oak (*Quercus robur*) embryos. *Plant Cell Reports* **28**, 1011–1021.
- Sunderlíková V, Wilhelm E.** 2002. High accumulation of legumin and *Lea*-like mRNAs during maturation is associated with increased conversion frequency of somatic embryos from pedunculate oak (*Quercus robur* L.). *Protoplasma* **220**, 97–103.
- Supek F, Bošnjak M, Škunca N, Šmuc T.** 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS ONE* **6**, e21800.
- Sweetman C, Wong DC, Ford CM, Drew DP.** 2012. Transcriptome analysis at four developmental stages of grape berry (*Vitis vinifera* cv. Shiraz) provides insights into regulated and coordinated gene expression. *BMC Genomics* **13**, 691.
- Tai HH, Tai GCC, Beardmore T.** 2005. Dynamic histone acetylation of late embryonic genes during seed germination. *Plant Molecular Biology* **59**, 909–925.
- Taji T, Ohsumi C, Iuchi S, Seki M, Kasuga M, Kobayashi M, Yamaguchi-Shinozaki K, Shinozaki K.** 2002. Important roles of drought- and cold-inducible

genes for galactinol synthase in stress tolerance in *Arabidopsis thaliana*. The Plant Journal **29**, 417–426.

Teoh KT, Requesens DV, Devaiah SP, Johnson D, Huang X, Howard JA, Hood EE. 2013. Transcriptome analysis of embryo maturation in maize. BMC Plant Biology **13**, 19.

Tran LP, Nakashima K, Sakuma Y, Simpson SD, Fujita Y, Maruyama K, Fujita M, Seki M, Shinozaki K, Yamaguchi-shinozaki K. 2004. Isolation and functional analysis of Arabidopsis stress-inducible NAC transcription factors that bind to a drought-responsive cis-element in the early responsive to dehydration stress 1 promoter. The Plant Cell **16**, 2481–2498.

Ueno S, Klopp C, Leplé JC, Derory J, Noirot C, Léger V, Prince E, Kremer A, Plomion C, Le Provost G. 2013. Transcriptional profiling of bud dormancy induction and release in oak by next-generation sequencing. BMC Genomics **14**, 236.

Ueno S, Le Provost G, Léger V, Klopp C, Noirot C, Frigerio J-M, Salin F, Salse J, Abrouk M, Murat F, et al. 2010. Bioinformatic analysis of ESTs collected by Sanger and pyrosequencing methods for a keystone forest tree species: oak. BMC Genomics **11**, 650.

Van der Pijl L. 1982. *Principles of dispersal in higher plants*. Springer Berlin Heidelberg.

Varela M, Valdivieso T. 1996. Phenological phases of *Quercus suber* L. flowering. Forest Genetics **3**, 93–102.

Wang Y, Zhang W-Z, Song L-F, Zou J-J, Su Z, Wu W-H. 2008. Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis. Plant Physiology **148**, 1201–1211.

Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB. 2007. Sampling the Arabidopsis transcriptome with massively parallel pyrosequencing. Plant Physiology **144**, 32–42.

Yin Y, Zhang X, Fang Y, Pan L, Sun G, Xin C, Ba Abdullah MM, Yu X, Hu S, Al-Mssallem IS, et al. 2012. High-throughput sequencing-based gene profiling on multi-staged fruit development of date palm (*Phoenix dactylifera*, L.). Plant Molecular Biology **78**, 617–626.

Yu K, Xu Q, Da X, Guo F, Ding Y, Deng X. 2012. Transcriptome changes during fruit development and ripening of sweet orange (*Citrus sinensis*). BMC Genomics **13**, 10.

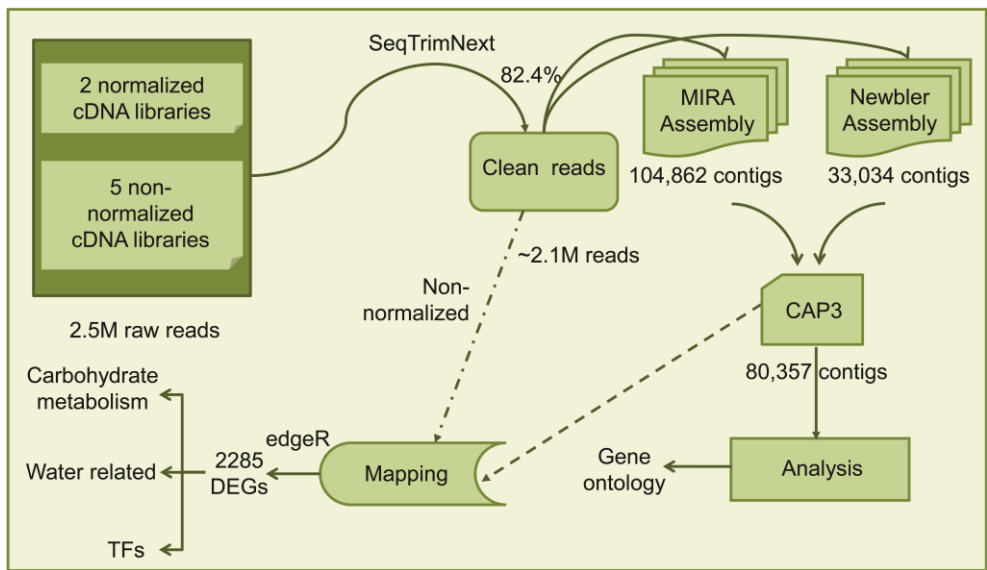
Zhang F, Gonzalez A, Zhao M, Payne CT, Lloyd A. 2003. A network of redundant bHLH proteins functions in all TTG1-dependent pathways of *Arabidopsis*. *Development* **130**, 4859–4869.

Zhang G, Pan L, Yin Y, Liu W, Huang D, Zhang T, Wang L, Xin C, Lin Q, Sun G, et al. 2012. Large-scale collection and annotation of gene models for date palm (*Phoenix dactylifera*, L.). *Plant Molecular Biology* **79**, 521–536.

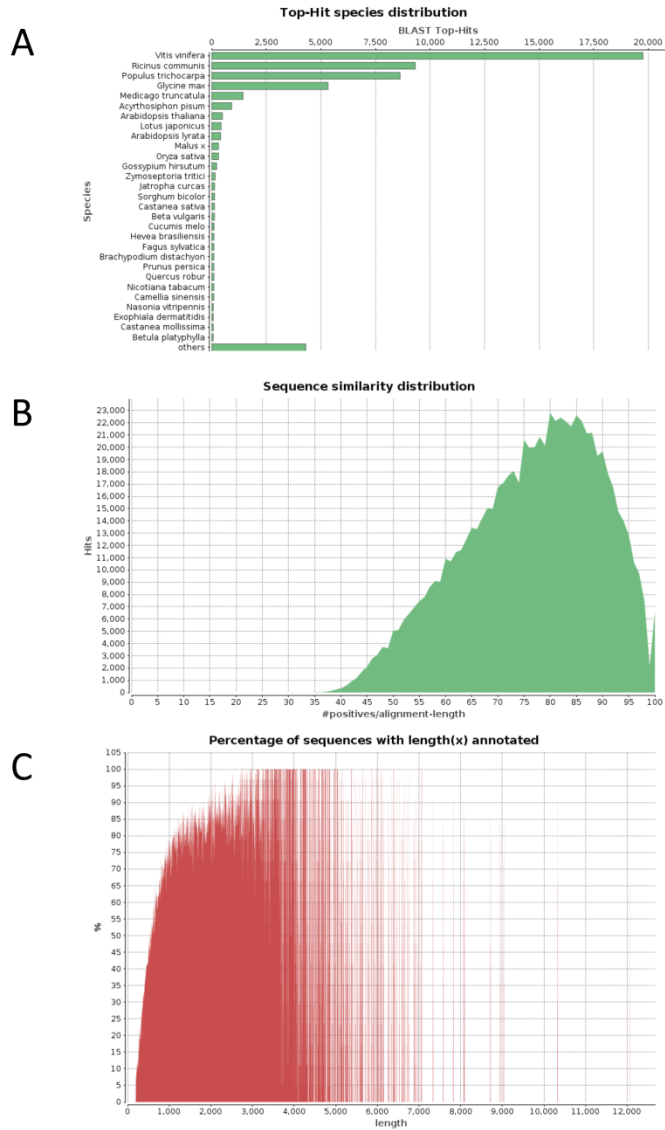
Zhu YY, Machleder EM, Chenchik A, Li R, Siebert PD. 2001. Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. *BioTechniques* **30**, 892–897.

Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB, Matz M V, Meleshkevitch E, Moroz LL, Lukyanov SA, et al. 2004. Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Research* **32**, e37.

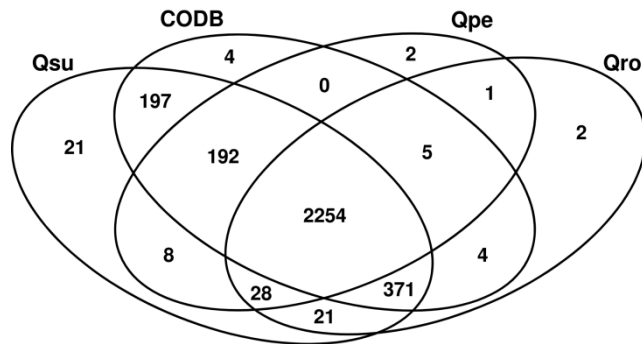
Supporting information



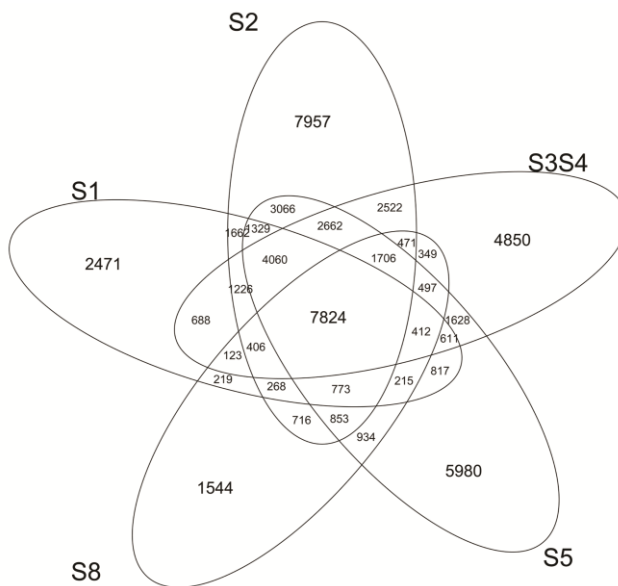
Supplementary Fig. S1. Workflow of the 454 sequencing data analyses. After sequencing, the raw reads obtained were pre-processed to remove non desired sequences, including low quality sequences and contaminants. The clean reads from all libraries were then assembled using two different assemblers and then joined using CAP3. The obtained contigs were then associated to gene ontology (GO) terms and searched for transcription factors. After mapping the differentially expressed genes (DEGs) along cork oak acorn development were identified.



Supplementary Fig. S2. Quality assessment of the gene annotation. (A) Species with the best BLASTX alignment of each query sequence. (B) Values of similitude from Blastx alignments. (C) Number of gene ontology (GO) terms per sequence length.

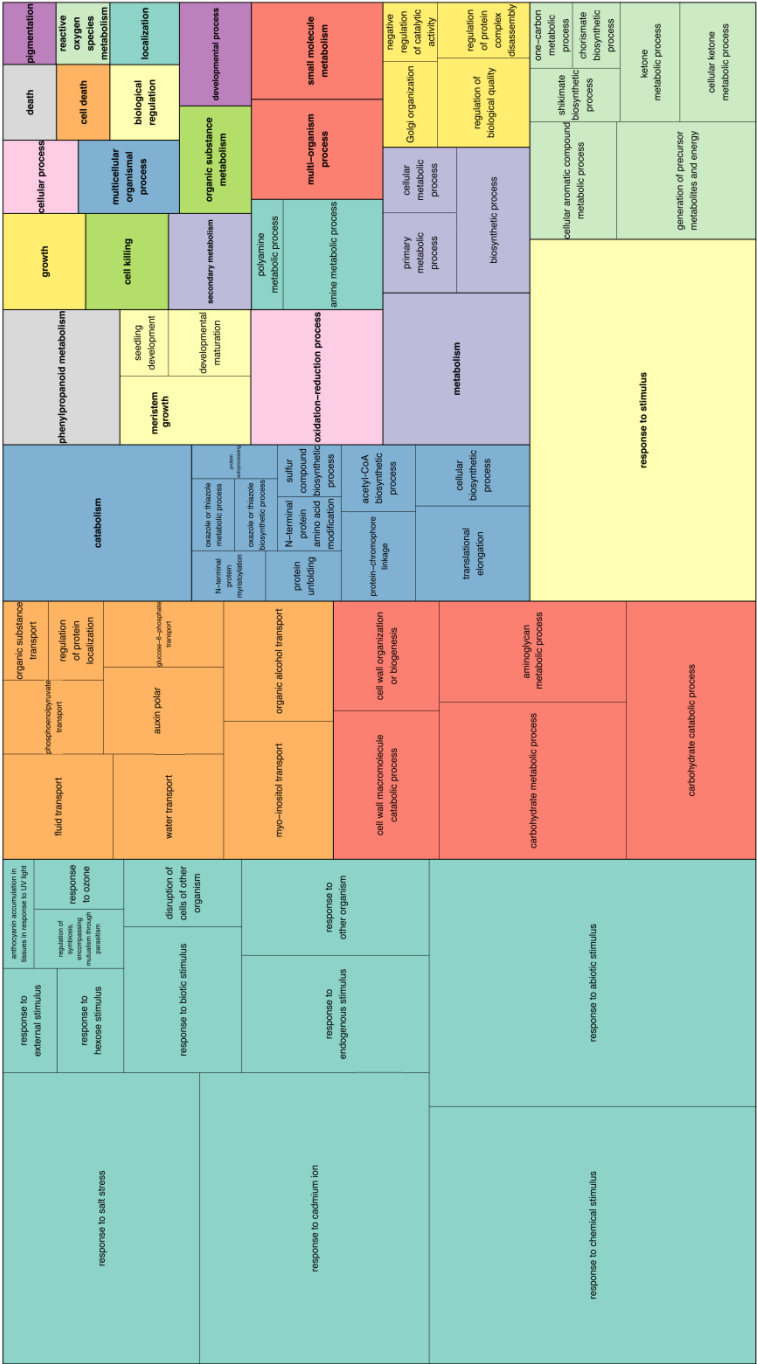


Supplementary Fig. S3. Venn diagram representing the distribution of Clusters of Orthologs Groups (COGs) between different species of the *Quercus* genus: *Q. suber* (Qsu), *Q. petraea* (Qpe), *Q. robur* (Qro) and the data from the Cork Oak Database (CODB).



Supplementary Fig. S4 Venn diagram showing the number of transcripts expressed in the different fruit developmental stages.

Supplementary Fig. S5. Treemap of the GO terms from the DEGs. The area of each cell is proportional to the number of genes annotated with that term.



Supplementary Table S1. Comparison of the transcriptomes of several Fagaceae tree species with full length plant proteins in UniprotKB. Transcriptomes were retrieved from the Fagaceae project (www.fagaceae.org) or NCBI for *Q. robur* and *Q. petraea*.

	<i>Q. suber</i>	<i>Q. robur</i>	<i>Q. petraea</i>	<i>Q. rubra</i>	<i>Q. alba</i>	<i>F. sylvatica</i>	<i>C. mollissima</i>	<i>C. dentata</i>
Total contigs	80,357	81,671	58,230	28,041	22,102	31,309	48,335	45,288
Translated proteins	56,517 (70.3)	65,030 (79.6)	47,567 (81.7)	20,186 (72.0)	16,200 (73.3)	27,765 (88.7)	31,907 (66.0)	31,338 (69.2)
Unique Uniprot homologous	24,474 (30.5)	17,565 (21.5)	14,618 (25.1)	13,804 (49.2)	11,777 (53.3)	10,264 (32.8)	19,723 (40.8)	18,814 (41.5)
Complete proteins	19,146 (33.9)	16,112 (24.8)	9,481 (19.9)	1,864 (9.2)	1,401 (8.6)	3,830 (13.8)	4,947 (15.5)	2,392 (7.6)
C-terminus proteins	11,410 (20.2)	21,881 (33.6)	16,121 (33.9)	4,477 (22.2)	3,843 (23.7)	10,544 (38.0)	7,252 (22.7)	6,668 (21.3)
N-terminus proteins	10,108 (17.9)	14,922 (22.9)	12,361 (26.0)	3,570 (17.7)	2,650 (16.4)	7,299 (26.3)	5,314 (16.7)	5,970 (19.1)
Internal proteins	15,509 (27.4)	12,091 (18.6)	9,584 (20.1)	10,218 (50.6)	8,258 (51.0)	6,086 (21.9)	14,200 (44.5)	16,252 (51.9)
Misassembled	344 (0.6)	24 (0.04)	20 (0.04)	57 (0.3)	48 (0.3)	6 (0.02)	194 (0.6)	56 (0.2)

Supplementary Table S2. Identification of shared proteins among some member of the Fagaceae family. Sequences from a query species (rows) with homologous in a different target species (column). Proteins were considered homologous between the query and the target species if they share a minimal identity and coverage of 70%.

% and total	Total *	<i>Q. suber</i>	<i>Q. robur</i>	<i>Q. petraea</i>	<i>Q. rubra</i>	<i>Q. alba</i>	<i>F. sylvatica</i>	<i>C. mollissima</i>	<i>C. dentata</i>	Average **
<i>Q. suber</i>	56,163	100 56,163	67.8 38,071	61.8 34,704	65.6 36,821	59.0 33,145	50.7 28,480	80.9 45,433	80.5 45,239	66.6
<i>Q. robur</i>	65,006	94.4 61,349	99.9 64,925	84.2 54,713	83.1 54,048	79.3 51,536	68.2 44,355	92.1 59,865	91.7 59,628	84.7
<i>Q. petraea</i>	47,547	96.2 45,718	88.1 41,873	99.9 47,492	82.6 39,288	79.3 37,681	69.1 32,854	92.9 44,177	91.5 43,519	85.7
<i>Q. rubra</i>	20,129	94.7 19,061	67.2 13,529	60.3 12,145	99.9 20,113	60.3 12,131	48.0 9,654	86.2 17,360	84.1 16,922	71.5
<i>Q. alba</i>	16,152	95.2 15,384	70.4 11,379	63.9 10,327	72.0 11,628	99.9 16,134	50.5 8,150	87.1 14,074	80.8 13,053	74.3
<i>F. sylvatica</i>	27,759	93.0 25,815	77.7 21,566	74.0 20,545	75.9 21,082	71.8 19,920	100 27,747	85.7 23,786	82.0 22,765	80.0
<i>C. mollissima</i>	31,713	92.4 29,289	61.6 19,537	55.4 17,55	59.0 18,726	51.5 16,219	42.8 13,574	99.9 31,686	73.2 23,219	62.3
<i>C. dentata</i>	31,282	93.7 29,308	63.8 19,971	56.5 17,688	62.2 19,452	53.5 16,741	44.4 13,901	84.9 26,550	99.9 31,244	65.6
Average **		94.2	70.9	65.2	71.5	65.0	53.4	87.1	83.4	

*Translated proteins minus misassembled proteins; ** Average excluding species own value.

Supplementary Table S3. Biological domain of the Clusters of Orthologous Groups of proteins (COGs) specific to *Q. suber*.

Group	COGs	%
A - RNA processing and modification	14	6.33
B - Chromatin structure and dynamics	4	1.81
C - Energy production and conversion	4	1.81
D - Cell cycle control, cell division, chromosome partitioning	12	5.43
E - Amino acid transport and metabolism	4	1.81
F - Nucleotide transport and metabolism	3	1.36
G - Carbohydrate transport and metabolism	2	0.90
H - Coenzyme transport and metabolism	5	2.26
I - Lipid transport and metabolism	11	4.98
J - Translation, ribosomal structure and biogenesis	13	5.88
K - Transcription	11	4.98
L - Replication, recombination and repair	27	12.22
M - Cell wall/membrane/envelope biogenesis	2	0.90
O - Post-translational modification, protein turnover, and chaperones	11	4.98
P - Inorganic ion transport and metabolism	2	0.90
Q - Secondary metabolites biosynthesis, transport, and catabolism	1	0.45
R - General function prediction only	29	13.12
S - Function unknown	46	20.81
T - Signal transduction mechanisms	8	3.62
U - Intracellular trafficking, secretion, and vesicular transport	6	2.71
Y - Nuclear structure	3	1.36
Z - Cytoskeleton	3	1.36
	221	100.00

Supplementary Table S4. Differentially expressed genes related to water. Cluster annotation, Uniprot Locus, association with *Arabidopsis* homologs, normalized counts in each acorn developmental stage and fold-change between consecutive stages.

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts					Fold-change				
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4	S6/S5	
A	Contig15851	B9HJ50	AT1G60010			19.8	0	0.6	0.4	0	-INF				
A	Contig16112	O80725	AT2G47000	MDR4, PGP4, ABCB4, ATPGP4	ATP binding cassette subfamily B4	35.6	3	1.6	0	7	-11.87				
A	Contig23781	O48651	AT1G58440	XFL1, SOE1	FAD/NAD(P)-binding oxidoreductase family protein	31.6	1.6	0	0.4	4.6	-19.75				
A	Contig24336	Q6C5U1	AT1G27320	AHK3, HK3	Histidine kinase 3	76	1.6	6.2	4.2	7	-47.5				
A	Contig25679	O04057	AT1G11910	APA1, ATAPA1	Aspartic proteinase A1	80.8	2.2	14.2	3.6	16	-36.73				
A	Contig32396	B8SAF4	AT3G62700	ATMRP10, MRP10, ABCG14	Multidrug resistance-associated protein 10	24.2	2.2	7.4	3	0	-11				
A	Contig4133	Q8MPV8	AT1G18990	ATCDPK1, CPK10, CDPK1, ACPK10	Calcium-dependent protein kinase 1	100.4	9.4	8.8	16.2	0	-10.68				-INF
A	Contig553	Q70DU8	AT1G44170	ALDH3H1, ALDH4	Aldehyde dehydrogenase 3H1	42.8	1.6	0.8	0	0	-26.75				
A	Contig9531	A5BTC8	AT1G13930			43	0	0.2	0.8	0.8	-INF				
A	Qs-dev_1ep_277063	Q8FXB2	AT1G56600	AGS2, GOS2	Galactinol synthase 2	25	0	1.2	0	1.2	-INF				
A	Qs-dev_1ep_364235	Q8LS97	AT3G17000	UBC32	Ubiquitin-conjugating enzyme 32	23.2	2.4	0.2	5.4	1.2	-9.67				
A	Qs-dev_1ep_365023	Q8PGJ6-2	AT2G37770	AKR4C9	NAD(P)-linked oxidoreductase superfamily protein	79.2	68.4	53.2	64.8	5.2					
A	Qs-dev_c10451	Q03194	AT2G18960	AHA1, PMA, OST2, HA1	H(+)-ATPase 1	179.2	50.8	17	22.4	4.2	-3.53				-12.46
A	Contig16356	P31168	AT3G50970	LT30, XERO2	Dehydrin family protein	608.8	184.2	138.6	32	11	-3.31				4.33
A	Contig16452	Q83VY3	AT4G27410	RD26	NAC domain transcriptional regulator superfamily protein	71.6	46	23.8	48.2	17.2					-2.80
A	Contig17464	Q8SM09	AT1G78900	VHA-A	Vacuolar ATP synthase subunit A	17.8	0	12.6	15.6	4.2	-INF				
A	Contig18250	P46519	AT2G44060		Late embryogenesis abundant protein	41.8	0.8	44	17.6	25.6	-52.25				55
A	Contig19089	O85554	AT4G30960	CIPK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	102.6	28.8	42.8	48.8	40.6	-3.56				
A	Contig19419	P48534	AT1G07890	APX1, MEE6, CSI.1, ATAPX1, ATAPX01	Ascorbate peroxidase 1	56.8	20.4	5.2	5.8	18.2	-2.76				
A	Contig3706	B8TV63	AT2G17940	ERD7	Senescence/oxidation-associated protein-related	51.8	7.4	23.4	15	5.2	-7				
A	Contig3766	Q8FHH8	AT5G57660	ATCOL5, COL5	Zinc finger protein CONSTANS like 5	32.6	5.2	2.2	12	26.6	-6.27				
A	Qs-dev_c42583	Q8STE9	AT1G70560	AOAT2, GGT2	Alanine 2-oxoglutarate aminotransferase 2	61.6	12.2	23.6	30	10.4	-5.05				
A	Qs-dev_1ep_27262	P36585	AT1G20693	HMG2, HMG BETA.1, NFD2, NFD02	High mobility group B2	39	8.2	0.2	9.2	36.2	-4.76				3.93
A	Qs-dev_1ep_273055	Q8FDW1	AT4G37260	MYB73, ATMYB73	myb domain protein 73	21.4	0.6	17.8	1.6	5.8	-35.67				
A	Qs-dev_1ep_276493	P31414	AT1G15860	AVP1, ATAVP3, AVP-3, AVHP1.1	Inorganic H pyrophosphatase family protein	19.2	0	1.4	9.2	1.6	-INF				
A	Qs-dev_1ep_28219	P17407	AT5G62350		Plant invertase/pectin methyltransferase inhibitor superfamily protein										

Supplementary Table S4. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts								Fold-change	
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5S3S4	S6/S5	
B	Contig120	Q83Y54	AT5G06530	ABCQ22	ABC-2 type transporter family protein	8.2	31.2	1.6	0.4	0		-19.5			
B	Contig12009	Q8F-X62	AT1G56600	AlGdS2, Gds2	Galactinol synthase 2	39.2	101.6	90.4	27.6	2	2.59		-3.28	-13.8	
B	Contig14342	Q24370	AT3G45140	LOX2, ATLOX2	Lipoxygenase 2	28.8	61	36.2	5	0			-7.24		
B	Contig14864	P48534	AT1G07890	APX1, MEE6, CS1, ATAPX1, ATAPX01	Ascorbate peroxidase 1	39.2	57.6	42.4	44.8	4.2				-10.67	
B	Contig16034	Q8FN09	AT5G40390	SIP1	Raffinose synthase family protein	173	281.2	394.6	107.4	0	1.63		-3.59	-INF	
B	Contig16899	P43287	AT2G37170	PIP2B, PIP2.2	Plasma membrane intrinsic protein 2	27.4	92	83.4	36.8	27	3.36		-2.27		
B	Contig17810	Q39011	AT4G18710	BIN2, DWF12, UCU1, ATSK21, SK21	Protein kinase superfamily protein	14	86.8	17.2	5.6	8.8	6.2	-5.05			
B	Contig17849	Q38890	AT5G49720	ATGH8K1, TSD1, DEC, KOR, RSN2, IRX2, KOR1, GH8K1	Glycosyl hydrolase 9A1	15.6	57.8	8.2	14.4	29.6	3.71	-7.05			
B	Contig17843	Q8F-X62	AT1G56600	AlGdS2, Gds2	Galactinol synthase 2	0	65.4	55.6	40.8	0	INF			-INF	
B	Contig18088	P43296	AT4G39090	RD16, RD18A	Cysteine protease family	29.4	65.6	12.6	26.8	23.6		-5.2			
B	Contig18151	Q39196	AT4G00430	TMP-C, PIP1.4, PPI1E	Plasma membrane intrinsic protein 1,4	83.4	189.4	103.8	61	0	1.9			INF	
B	Contig18192	Q41951	AT3G16240	DELTA-TIP, TIP2.1, DELTA-TIP-1, AQP1, ATTIP2.1	Delta tonoplast integral protein	12	43.8	35.2	20.4	6.6	3.63				
B	Contig18473	Q82598	AT4G01470	GAMMA-TIP3, TIP1.3, ATTIP1.3	Tonoplast intrinsic protein 1,3	8.4	46.4	39.6	12.2	0	5.52				
B	Contig18768	Q8LIR6	AT3G23920	BAM1, BMV7, TR-BAMY	Beta-amylase 1	1.2	29.6	0.6	1.2	2.2	24.67	49.33			
B	Contig18884	B8HH88	AT1G47128	RD21, RD21A	Cysteine protease family protein	128.8	93.2	31.8	30.6	27.6		-2.93			
B	Contig19315	P40267	AT2G18050	HIS1-3	Histone H1-3	44	69.8	5.2	15.6	17.6		-13.42			
B	Contig19387	Q8LS40	AT3G18490	ASPG1	Aspartic protease family protein	29.6	38.6	9.4	10.2	2.2		4.11			
B	Contig19452	Q8VMZ7	AT2G45570	CYP76C2	Cytochrome P450, family 76, subfamily C, polypeptide 2	0	30	33.8	2	0	INF		-16.9		
B	Contig19458	Q8FN09	AT5G40390	SIP1	Raffinose synthase family protein	8.4	76.2	25.4	2	2.2	9.07	-3	-12.7		
B	Contig19793	Q8ZT5	AT3G63060	EDL3	ED1-like 3	3	24.6	12.6	0	0	8.2				
B	Contig21230	Q8F-X62	AT1G56600	AlGdS2, Gds2	Galactinol synthase 2	3	81.2	92.2	38.6	2.2	27.07		-2.39	-17.55	
B	Contig21232	Q8F-X62	AT1G56600	AlGdS2, Gds2	Galactinol synthase 2	8.8	21.4	12.8	14.6	0				-INF	
B	Contig21234	Q8F-X62	AT1G56600	AlGdS2, Gds2	Galactinol synthase 2	39.6	86.4	63.4	3.4	0	2.18		-18.65		
B	Contig21412	Q84980	AT5G52300	RD28B, LT165	CAR160 protein	8.8	23.6	0	2	2.2		-INF			
B	Contig21559	Q24301	AT4G02260	SUS3, ATUS3	Sucrose synthase 3	0	20.4	5.2	5.2	15.6	INF				
B	Contig21661	Q8M111	AT4G02260	SUS3, ATUS3	Sucrose synthase 3	1.2	44	5.2	29.4	19.8	36.67	-8.46	5.05		
B	Contig22701	P35007	AT4G13940	HOG1, ENB1395, SAHH1, MEE56, ATSAHH1	S-adenosyl-L-homocysteine hydrolase	140	214.6	86.6	106.6	81	1.53	-2.18			
B	Contig22899	Q8M7Q4	AT1G45249	ABF2	Abscic acid responsive elements-binding factor 2	14	36.8	6	5.2	2.2		-6.13			

Supplementary Table S4. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts						Fold-change		
						S1	S2	S3S4	S5	S8	S2/S1	S3/S4	S5/S3S4	S8/S5
B	Contig23109	O04057	AT1G11910	APA1, ATAPA1	Aspartic proteinase A1	84.6	92.2	44.4	70.6	74.6			-2.08	
B	Contig2384	Q6SC24	AT3G51550	FER		10.4	27.8	8.2	23.6	2.2				-10.73
B	Contig23659	Q39196	AT4G00430	TMP-C, PIP1-4, PIP1E	Maleicillin/ceceptor-like protein kinase family protein	5	93.8	37.8	27.8	2.2	18.76	-2.48		-12.64
B	Contig24038	Q7XSO9	AT4G00430	TMP-C, PIP1-4, PIP1E	Plasma membrane intrinsic protein 1;4	498	371.6	207.8	207.8	43.8	-1.34	-1.79		
B	Contig24061	Q8JTM6	AT2G37170	PIP2B, PIP2-2	Plasma membrane intrinsic protein 2	3	29.8	22.8	20.4	37.4	9.93			
B	Contig24434	P26585	AT1G20693	HMG82, HMG BETA 1, NFD2, NFD02	High mobility group B2	8.8	16.6	0	4.2	4.6		-INF		
B	Contig25183	P49608	AT2G05710	ACO3	Aconitase 3	6.8	58.6	35.4	29.8	30	8.62			
B	Contig25211	Q648U0	AT1G64060	ATRBOH F, ATRBOHF, RBOHAP108, RBOHF, RBOH F	Respiratory burst oxidase protein F	31.8	67.6	56.8	56	2.2				-25.45
B	Contig25242	Q41142	AT3G15730	PLDALPHA1, PLD	Phospholipase D alpha 1	216.6	189.2	134.8	107.8	55.8				-1.93
B	Contig25491	Q8JLR6	AT3G32920	BAM1, BMV7, TR-BAMY	Beta-amylase 1	303.6	557.6	608.8	175.8	4.6	1.84			-3.46
B	Contig25581	Q6MT02	AT3G19290	ABF4	ABRE binding factor 4	26.2	20.4	10.8	16.8	0				-INF
B	Contig25946	Q8QOQ6	AT5G56030	HSP81-2	Heat shock protein 81-2	93	87	103	50.8	46.4				-2.03
B	Contig26056	B8FV68	AT4G30960	CIPK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	27.8	153.8	42	86.2	27.8	5.53	-3.66		-3.1
B	Contig26069	P43296	AT4G39090	RD19, RD19A	Cysteine protease family	117.8	103	107.4	101	30				-3.37
B	Contig26432	Q7XSO9	AT1G01620	PIP1C, TMP-B, PIP1-3	Plasma membrane intrinsic protein 1C	44.4	99.2	84.2	98	30	2.23			-3.27
B	Contig3549	Q8MEB8	AT3G14440	NCED3, ATNCECD3, STO1, SIS7	Nine-cis-epoxycarotenoid dioxygenase 3	103.8	384.6	112.6	36.2	0	3.71	-3.42	-3.11	-INF
B	Contig3821	Q8LY87	AT5G14420	RGL52	RING domain ligase2	5.2	29.2	8.6	19.8	7.4	5.62			
B	Contig5183	P49598	AT3G11410	ATPP2CA, AHG3, PP2CA	Protein phosphatase 2CA	38.6	82.4	48.8	63.2	9.6	2.13			-6.88
B	Contig5234	Q8SRP5	AT2G38470	WRKY33, ATWRKY33	WRKY DNA-binding protein 33	32	55.6	40.6	10.4	7.4				-3.9
B	Contig5473	P46897	AT3G61890	ATHB-12, ATHB12, HB-12	Homeobox 12	40.4	88.4	49.8	12	2.2	2.19			-4.15
B	Contig5553	Q8FXB2	AT1G56600	AGD42, GOS2	Galactinol synthase 2	1.6	30.4	21	9	0	19			
B	Contig7694	P36908	AT5G24090	ATCHA, CHIA	Chitinase A	7.4	26.6	8.8	16.2	0				-INF
B	Contig7776	Q8LMA9	AT1G19180	JAZ1, TEY10A	Jasmonate-2m-domain protein 1	87.4	89.2	96.4	35.4	9.6				-2.72
B	Contig8279	P17407	AT5G82350		Plant invertase/pectin methylesterase inhibitor superfamily protein	9.2	18.2	9.8	18.4	0				-INF
B	Contig8793	P21616	AT1G15890	AVP1, ATAVP3, AVP-3, AVHP1;1	Inorganic H pyrophosphatase family protein	44.8	149.6	22	60.4	92.2	3.34	-6.8	2.75	
B	Contig8927	P55034	AT4G38630	RPN10, MCB1, ATMCB1, MBP1	Regulatory particle non-ATPase 10	13.2	32	5.2	17.8	5.2				-6.15
B	Contig9985	B9ID06	AT5G52060	ATBAG1, BAG1	BCL2-associated athanogene 1	7.4	33	22.2	16.2	0	4.45			-INF
B	Qs-dev_c30725	Q8JLR6	AT3G23920	BAM1, BMV7, TR-BAMY	Beta-amylase 1	0	23.8	0	9.2	0	INF	-INF		

Supplementary Table S4. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts						Fold-change	
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4
B	Os_dev_6836	Q68024-4	AT5G60410	ATSI21, SZ1	DNA-binding protein with MIZ/SP-RING zinc finger, PHD-finger and SAP domain	5.8	29.4	17.8	7	28.8	5.07		
B	Os-dev_j_rep_272408	P35016	AT4G24190	SHD, HSP90.7, AHsp90.7, AHsp90-7	Chaperone protein in hspG family protein	96.8	157.4	67.6	56.8	20.6		-2.33	
B	Os-dev_j_rep_272753	O04057	AT1G11910	APA1, ATAPA1	Aspartic proteinase A1	17.2	69.2	41.2	54	24.6	4.02		
B	Os-dev_j_rep_272815	Q03878	AT2G21660	ATGRF7, GCR2, GR-RBP7, GRP7	Glycine-rich RNA-binding protein	4.2	33.2	24.6	7.4	34.8	7.9		4.7
B	Os-dev_j_rep_273270	B876V8	AT4G30960	CIPK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	21.4	45	2.2	20	3.4	-20.45		
B	Os-dev_j_rep_273395	O24370	AT3G45140	LOX2, ATLOX2	Lipoxygenase 2	19.2	91	29.2	9.8	0.8	4.74	-3.12	
B	Os-dev_j_rep_274595	Q42711	AT5G03630	ATHMR2	Pyridine nucleotide-disulphide oxidoreductase family protein	2.2	35.2	0.2	3.2	5.8	16	-176	
B	Os-dev_j_rep_276744	QJLR8	AT3G23920	BAM1, BMV7, TR-BAMY	Beta-amylase 1	0.4	54.2	0.2	8.4	0.8	135.5	-271	
B	Contig16565	Q39613	AT2G21130		Cyclophilin-like peptidyl-prolyl cis-trans isomerase family protein	165.6	192.4	101	170.4	86	-1.9	1.89	-2
B	Contig16456	B91491	AT5G52300	RD28B, LT165	CAP160 protein	12	37.8	0.6	5	21	-63		
C	Contig16616	B9R8B8	AT2G06050	OPR3	Oxaphydriolate-reductase 3	10	17.8	73.4	10	2	-3.65	4.12	-7.34
C	Contig17367	Q08298	AT5G25810	RD22, ATRD22	Dehydration-responsive protein RD22	46	12.6	42.4	14.4	25.4	3.37		
C	Contig19544	Q8L1B8	AT5G47100	CBL9, ATCBL9	Caldeurain B-like protein 9	0	0	46.4	12.2	0	INF	-3.8	
C	Contig19684	P31414	AT1G15960	AVP1, ATAVP3, AVP-3, AVHP1.1	Inorganic H pyrophosphatase family protein	4.6	6.2	44.4	5.6	2.2	7.16	-7.93	
C	Contig20487	O24370	AT3G45140	LOX2, ATLOX2	Lipoxygenase 2	14	0.2	21.6	2	0	108	-10.8	
C	Contig22431	F6HLV2	AT5G55120	VTC5	GDP-D-glucose phosphorylases	3	5.6	75.4	6.6	2.2	13.46	-11.42	
C	Contig22432	F6HLV2	AT5G55120	VTC5	GDP-D-glucose phosphorylases	5	8.8	75.8	11	2.2	8.61	-6.89	
C	Contig24282	P48534	AT1G07960	APX1, MEE6, CSI, ATAPX1, ATAPX01	Ascorbate peroxidase 1	3	1.6	48.2	0.4	2.2	30.13	-120.5	
C	Contig25263	P35016	AT4G24190	SHD, HSP90.7, AHsp90.7, AHsp90-7	Chaperone protein in hspG family protein	8.8	4.6	173	49.8	33.8	37.61	-3.47	
C	Contig26805	Q06289	AT1G27730	SZ2, ZAT10	Salt tolerant zinc finger	1.2	0	18.4	5.8	0	INF		
C	Os-dev_c15832	Q8LJ46	AT4G23400	PIP1D, PIP1F.5	Plasma membrane intrinsic protein 1.5	1.6	11	33.2	3	0		-11.07	
C	Os-dev_j_rep_275764	Q7X5Q9	AT4G00430	THMP-C, PIP1F.4, PIP1E	Plasma membrane intrinsic protein 1.4	4.2	2	43.8	9.8	3.6	21.9	-4.47	
C	Os-dev_j_rep_276379	Q94B08	AT1G47126	RD21, RD21A	Cysteine protease family protein	8	3	93	20.8	12.6	31	-4.47	
C	Os-dev_j_rep_277033	Q08298	AT5G25810	RD22, ATRD22	Dehydration-responsive protein RD22	0.4	0	34.4	0	3.6	INF	-INF	
C	Os-dev_j_rep_289669	P35016	AT4G24190	SHD, HSP90.7, AHsp90.7, AHsp90-7	Chaperone protein in hspG family protein	9.6	0.6	47	15	14.8	78.33	-3.13	
C	Contig17758	Q941L0	AT5G05170	CESA3, IXP1, ATCESA3, ATH-B, CEV1	Cellulose synthase family protein	39.4	37.8	74	30.4	40.4		-2.43	
C	Contig18046	B05781	AT1G19160	JAZ1, TIFY10A	Jasmonate-zim-domain protein 1	3	17.8	35.2	2	0		-17.6	
C	Contig18864	11KS55	AT4G13940	HOG1, EMB1395, SAHH1, MEE58, ATSAHH1	S-adenosyl-L-homocysteine hydrolase	192.4	141.6	351.6	120.6	58.2	2.48	-2.92	-2.07

Supplementary Table S4. (Continued)

Cluster	Transcript name	Unigrot	At Locus	At homologous	At homologous function	Normalized counts						Fold-change		
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4	S6/S5
C	Contig23722	Q24370	AT3G45140	LOX2, ATLX2	Lipoxygenase 2	31.6	25.6	37	9.4	0				-3.94
C	Contig25837	Q85554	AT4G30960	CIPK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	26.2	16.2	31.8	16.2	0				-NF
C	Contig4832	F6HP56	AT2G45570	CYP76C2	Cytochrome P450, family 76, subfamily C, polypeptide 2	11	17.6	25.6	3	0				-8.53
C	Contig9321	P45730	AT2G37040	PAL1, ATPAL1	Phenylalanine ammonia-lyase 1	26.6	40.4	81.2	22.2	0				-3.66
C	Contig9499	Q82598	AT4G01470	GAMMA-TIP3, TIP13, ATIP13	Tonoplast intrinsic protein 13	3.4	13.2	25.8	3	0				-8.6
C	Contig16423	AT1226	AT1G02500	SAM1, SAM1-1, MAT1, ASAM1	S-adenosylmethionine synthetase 1	21.6	57.6	86.8	53.6	36.8	2.67			
C	Contig17590	Q8LD45	AT5G47120	ATBI-1, BI-1, ATBI1, BI1	BAX inhibitor 1	23.6	94.2	113.2	81.2	53.8	3.99			
C	Contig21240	Q8F682	AT1G56600	AGS4S2, GMS2	Galactinol synthase 2	20	24.6	42.8	37.8	0				-NF
C	Qe-dev_c14813	Q42560	AT4G35830	ACO1	Acetoinase 1	1.6	24.2	36	26.4	26.4	15.13			
D	Contig10022	Q8LXV3	AT5G12860	DT1	Dicarboxylate transporter 1	1	3.8	12.6	16	0				-NF
D	Contig10143	P83536	AT3G29320	PHS1	Glycosyl transferase, family 35	9.8	10.2	14.6	222.2	156				15.22
D	Contig17288	P82983	AT3G30775	ERD5, PRODH, AT-POX, ATPOX, ATPDH, PRO1	Methylerythritolglyoxylate reductase family protein; Proline dehydrogenase 1	0	0	0	22	15				INF
D	Contig17532	Q48651	AT1G58440	XFI, SOE1	FADNAD(P)-binding oxidoreductase family protein	17.6	14.8	32.4	43.6	8.8				-4.95
D	Contig17652	Q48905	AT1G04410	CNAO-JHDH1	Lactate/malate dehydrogenase family protein	119.6	32.4	69.8	100.8	65.6	-3.69			
D	Contig18047	Q65VA5	AT2G13360	AGT, AGT1, SGAT	Alanine glyoxylate aminotransferase	6.6	6	5	20.4	2.2				-9.27
D	Contig18429	P21616	AT1G15690	AVP1, ATAVP3, AVP-3, AVHP1;1	Inorganic H pyrophosphatase family protein	62.8	75.8	140	141.6	117.8		1.85		
D	Contig18864	Q8LEV3	AT5G10860	QBSX3	Cystathionine beta-synthase (CBS) family protein	4.6	1.6	0.6	19	0				31.67
D	Contig19318	Q01593	AT3G24650	ABIS, SIS10	AP2B3-like transcriptional factor family protein	0	0	7.2	66.8	42.2				9.28
D	Contig19420	F41G5	AT1G29395	COR414-TM1, COR413M1, COR413-TM1	Cold-regulated 314 inner membrane 1	4.6	0.2	3.8	35.6	15.6				9.37
D	Contig19708	Q39196	AT1G201620	PIP1C, TMP-B, PIP1;3	Plasma membrane intrinsic protein 1C	42	33.2	49.2	103	23.8				2.09
D	Contig19883	Q82802	AT5G04590	SIR	Sulfite reductase	29.6	17.4	27.6	48.8	17.6				-2.77
D	Contig20209	QKZV51	AT5G26570	PWD, OK1, ATGW03	Catalytic carbohydrate kinases: phosphoglucon, water dikinases	60.2	47.4	241.6	218.4	9		5.1		-24.27
D	Contig20793	QBSM09	AT1G78900	VHA-A	Vacuolar ATP synthase subunit A	49.2	15.6	34	55.6	46	-3.15			
D	Contig24982	Q8ATM6	AT2G37170	PIP2B, PIP2;2	Plasma membrane intrinsic protein 2	116	37.2	51	79.8	13.8	-3.12			-5.78
D	Contig2415	P12459	AT5G12250	TUB6	Beta-4 tubulin	0	0	1.6	19.8	15.8				12.38
D	Contig4052	F6H8X2	AT1G06620		2-oxoglutarate (2OG) and Fe(II)-dependent oxygenase superfamily protein	0	0	3	15.6	0				-NF
D	Contig5531	P04796	AT1G13440	GAPC-2, GAPC2	Glyceroldehyde-3-phosphate dehydrogenase C2	152	118.4	124.8	203	107.8				1.63
D	Contig9988	Q5J907	AT3G16640	TCTP	Translationally controlled tumor protein	147.8	131	163.8	320.2	251.4				1.95

Supplementary Table S4. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts							Fold-change	
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4	
D	Os-dev_rep_c103460	Q8ALJ8	AT1G29395	COR414-TM1, COR413M1, COR413-TM1	Cdd-regulated 314 inner membrane 1	0.4	0	0	18	16.6			INF	
D	Os-dev_rep_c72718	Q8M111	AT4G02280	SUS3, ATSUS3	Sucrose synthase 3	85.2	103.8	122.6	161.2	81.4				-1.98
D	Os-dev_rep_c73066	Q8LD45	AT5G47120	ATBI-1, BI-1, ATBI1, BI1	BAX inhibitor 1	7.8	1.2	1.2	27.4	10.4			22.83	
D	Os-dev_rep_c75012	B8RAR5	AT5G14420	RGLG2	RING domain ligase2	2.2	7.2	1.2	20	0.8			16.67	-25
D	Os-dev_rep_c75610	Q76E5	AT2G38170	CAX1, ATCA1, RC4	Calcium exchanger 1	2.2	1.2	5.6	31.8	16.6			5.68	
D	Os-dev_rep_c76415	P33336	AT3G29320	PHS1	Glycyl transferase, family 35	0.4	0.6	0.2	25.4	16.6			12.7	
D	Os-dev_rep_c92139	Q841L0	AT5G05170	QESA3, IXR1, ATCEA3, ATH-B, DEV1	Cellulose synthase family protein	0.8	1.2	0.4	20	8.4			50	
D	Os-dev_rep_c92373	Q8MA41	AT1G56950	POM1, ERH2, ELPI, CTL1, ELP, HOT2, ATCTL1	Chitinase family protein	0.8	0	0.4	20.2	1.4			50.5	-14.43
D	Contig16886	Q83ZC5	AT1G13280	AOC4	Alene oxide cyclase 4	0	12.8	10.4	17.6	0				-INF
D	Contig17492	P2818	AT2G36930	GAMMA-TIP, TIP1, GAMMA-TIP1	Gamma tonoplast intrinsic protein	6.6	30.6	29.6	51	6.6	4.64			-7.73
D	Contig17577	Q8VX8	AT2G16850	PIP3B, PIP2.8	Plasma membrane intrinsic protein 2B	23.8	41.2	51	59.6	11				-5.42
D	Contig18148	Q8FX62	AT1G56600	AGDS2, GDS2	Galactinol synthase 2	1.2	17.8	19.4	18.8	0				-INF
D	Contig25637	Q8LNJ5	AT1G01260	JAM2	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein	5.2	16.6	7.4	24.2	2.2				-11
D	Contig26115	Q8SB81	AT4G21960	PRXR1	Peroxidase superfamily protein	10.8	87.4	91	93.8	15.8	8.09			-5.94
D	Contig4682	P43297	AT1G47128	RD21, RD21A	Cysteine protease family protein	5.2	12.6	24.2	21.6	0				-INF
D	Contig5705	Q8545	AT5G05700	BAM1	Leucine-rich receptor-like protein kinase family protein	26.4	20.4	30.8	33.2	0				-INF
E	Contig11296	Q8LEV3	AT5G10860	CBSX3	Cystathionine beta-synthase (CBS) family protein	1	9.6	9.2	53	77			5.76	
E	Contig14970	C8ZL25	AT3G50570	LT10, XERO2	Dehydrin family protein	4.6	2.2	29.4	43.6	50.4		13.36		
E	Contig16329	P11796	AT3G10920	MSD1, MEE3B, ATMSD1	Manganese superoxide dismutase 1	3	0	20.4	40.4	19		INF		
E	Contig19981	I2E4W7	AT2G41430	ERD15, LSR1, CD1	Dehydration-induced protein (ERD15)	0	2.2	0.6	25.8	37.2			43	
E	Contig2283	P53537	AT3G46970	ATPHS2, PHS2	Alpha-glucan phosphorylase 2	22	9.6	6	34.2	48.8			5.7	
E	Os-dev_c34533	Q7X5Q9	AT1G01620	PIP1C, TMP-B, PIP1.3	Plasma membrane intrinsic protein 1C	27	3.6	0	25.4	24.6	-7.5		INF	
E	Os-dev_c8659	P8218	AT1G65930	cICDH	Cytosolic NADP+-dependent isocitrate dehydrogenase	11.6	42.4	86.4	82.2	77.8	3.66			
E	Os-dev_rep_c74970	Q46851	AT1G58440	XFI, SDE1	FAD(MAD)P-binding oxidoreductase family protein	0.4	0.6	0.2	24.4	30.8			122	
F	Contig16865	Q39644	AT4G15910	ATD21, DI21	Late embryogenesis abundant protein; Drought-induced 21	48.2	0	8.2	19.8	371.6	-INF			18.77
F	Contig16935	Q39644	AT4G15910	ATD21, DI21	Late embryogenesis abundant protein; Drought-induced 21	1.2	0	3.6	18.8	179.2				9.53
F	Contig17540	Q02400	AT3G51810	ATEM1, GEA1, AT3, EM1	Late embryogenesis abundant protein; Stress induced protein	0	1.6	0.6	22	86.2			36.67	3.92
F	Contig17517	C8ZL25	AT3G50980	XERO1	Dehydrin zero 1	3	0.8	0.6	3.4	50.8				14.94

Supplementary Table S4. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts					Fold-change		
						S1	S2	S3S4	S5	S6	S2/S1	S3/S1	S5/S3S4
F	Contig18471	Q43129	AT5G59320	LTP3	Lipid transfer protein 3	0	0.2	0	5.6	38.6			6.9
F	Contig18508	Q01417	AT5G06760	LEA-5	Late Embryogenesis Abundant 4-5	1.2	2.2	0.6	0.4	53			132.5
F	Contig18905	Q39644	AT4G15910	ATD21, D121	Late embryogenesis abundant protein, Drought-Induced 21	25.8	69.8	7.2	41.6	312.6	2.71	-0.69	5.78
F	Contig19983	Q43129	AT5G59320	LTP3	Lipid transfer protein 3	0	0	0	9.4	56.8			6.04
F	Contig21222	Q6SVK9	AT1G05680	UGT74E2	Uridine diphosphate glucosyltransferase 74E2	0	0.8	0	4.2	33.8			8.05
F	Contig24985	Q9ZRW8	AT1G78380	ATGSTU19, GSTA, GSTU19	Glutathione S-transferase TAU 19	5	2.2	6	8.2	44.4			5.41
F	Contig5586	Q8VWZ7	AT2G45570	CYP76C2	Cytochrome P450, family 76, subfamily C, polypeptide 2	5.2	0.2	0	2.2	20.4			9.27
F	Contig6376	P91819	AT5G52840	HSP81-1, ATHSR3, HSP81-1, HSP83, ATHSP90.1, ATHsp90-1, HSP90.1	Heat shock protein 90.1	0	0.2	2	2.2	22.2			10.09
F	Contig6889	Q6SVL6	AT3G50830	COR413-PM2, ATCOR413-PM2	Cold-regulated 413-plasma membrane 2	0	1.6	0.8	1.2	36			30
F	Qs-dev_c42419	Q39644	AT4G15910	ATD21, D121	Late embryogenesis abundant protein, Drought-Induced 21	0	0	0	0	16.4			INF
F	Qs-dev_rep_c103646	Q43497	AT3G52880	ATMDAR1, MDAR1	Monodehydroascorbate reductase 1	13.2	0	0	7	61.2			8.74
F	Qs-dev_rep_c72494	Q8K6B9	AT2G24200	LAP1	Cytosol aminopeptidase family protein	42.8	20.4	16.6	33.4	102.6			3.07
F	Qs-dev_rep_c72884	Q01417	AT5G06760	LEA-5	Late Embryogenesis Abundant 4-5	4.2	0	0.2	7	32.4			4.63
F	Qs-dev_rep_c73281	Q86386	AT3G15353	MT3, ATMT3	Metallochione 3	0.4	3	0.2	4.6	75			16.3
F	Qs-dev_rep_c75701	Q8KKT8	AT5G60920	COB	COBRA-like extracellular glycosyl-phosphatidyl inositol-anchored protein family	2.2	0	2.2	0	18.8			INF
F	Qs-dev_rep_c77682	Q6ZY51	AT5G01260		Carbohydrate-binding-like fold	0.4	0	0.2	0	14.8			INF
F	Qs-dev_rep_c79475	Q86386	AT3G15353	MT3, ATMT3	Metallochione 3	0.4	0	0.2	0.8	18.8			23.5
F	Qs-dev_rep_c85371	P31414	AT1G15690	AVP1, ATAVP3, AVP-3, AVHP1;1	Inorganic P pyrophosphatase family protein	6	2	6.8	3.2	22.8			7.13
F	Qs-dev_rep_c91238	Q85TX5	AT2G04030	CR88, EMB1956, HSP90.5, Hsp98.1, AHsp90.5	Chaperone protein Hsp90 family protein	0.8	2	2.2	0.8	18.8			23.5
F	Contig17482	Q43497	AT3G52880	ATMDAR1, MDAR1	Monodehydroascorbate reductase 1	64.4	37.8	56.6	62.4	126.2			2.02
F	Contig17696	B9N841	AT5G20700		Protein of unknown function (DUF581)	23.6	9	3.6	5	37			7.4
F	Contig17776	Q6SPV4	AT1G19640	JMT	Jasmonic acid carboxyl methyltransferase	0	0	0	60.2	142.4			INF
F	Contig17906	F6HCF1	AT2G04030	CR88, EMB1956, HSP90.5, Hsp98.1, AHsp90.5	Chaperone protein Hsp90 family protein	19.8	27	53	15.4	56.6			-3.44
F	Contig17981	Q03666	AT1G17180	ATGSU25, GSTU25	Glutathione S-transferase TAU 25	0	0	0	29.8	74.2			INF
F	Contig18183	Q8SY70	AT1G35720	ANNAT1, OXY5, ATOXY5	Annexin 1	4.6	31.8	41.6	26.2	44	6.91		
F	Contig18205	Q43129	AT5G59320	LTP3	Lipid transfer protein 3	0	0.8	3.8	17.6	56.6			3.22
F	Contig19313	DTMH17	AT4G14430	IBR10, ATEC2, EC12, ECHIB, PEC12	Indole-3-butylic acid response 10	12.2	13.2	3.8	6.6	29.8			4.52
F	Contig19349	F41G55	AT1G28995	COR414-TM1, COR413M1, COR413-TM1	Cold-regulated 314 liner membrane protein	6.6	16.6	8.2	35.6	90.2			4.34

Supplementary Table S4. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts						Fold-change			
						S1	S2	S3S4	S5	S6	S8	S2/S1	S3S4/S2	S5/S3S4	S8/S5
F	Contig20364	AT3G15353	AT3G15353	MT3, ATMT3	Metallothionein 3	0	0.8	0	9.4	37.4					3.98
F	Contig25845	AT1G66930	AT1G66930	cICDH	Cytosolic NADP+-dependent isocitrate dehydrogenase	73.4	71.6	3	25.8	73.4		-23.87	8.6		2.84
F	Contig3914	AT1G74820	AT1G74820	ALDH10A8	Aldehyde dehydrogenase 10A8	1.4	0.2	0	11	42.8					3.89
F	Contig4590	AT1G10760	AT1G10760	SEX1, SOP1, SOP, GWD1, GWD	Pyruvate phosphate dikinase, PEP/pyruvate binding domain	42.2	44.4	42.8	52.2	113.8					2.18
F	Qs-dev_c50492	AT1G32230	AT1G32230	RCD1, CEO, CEO1, ATP8, AIRCD1	WWE protein-protein interaction domain protein family	2	7.2	0	0.8	18.4					23
F	Qs-dev_rep_c74668	AT3G44750	AT3G44750	HD2A, ATHD2A, HDX3, HDT1	Histone deacetylase 3	4.2	8.2	5.6	0.8	18.8					23.5
F	Qs-dev_rep_c75887	AT5G12250	AT5G12250	TUB6	Beta-6 tubulin	8	21.6	2.2	10.4	39.8					3.83

INF: expression ratios with zero counts in the numerator or denominator were annotated as minus infinitive (-INF) and infinitive (INF), respectively.

Supplementary Table S5. Differentially expressed genes annotated as transcription factors. Cluster annotation, Uniprot and *Arabidopsis* Locus, association with *Arabidopsis* homologs, normalized counts in each acorn developmental stage and fold-change between consecutive stages.

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts						Fold-change	
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4
A	Contig23379	A6M/X8	AT1G76810		Eukaryotic translation initiation factor 2 (eIF-2) family protein	27.8	3	12.8	14.8	7	-9.27		
A	Contig25150	Q84A40	AT1G82020		Cotumer alpha subunit	112.6	24.2	31.8	61.2	13.8	-4.65		-4.43
A	Contig17909	Q24076	AT1G18080	ATARCA, RACK1A, AT_RACK1A	Transducin WD40 repeat-like superfamily protein	35.4	27.8	3	3	23.6		-9.27	7.87
A	Contig18054	F6H0V9	AT1G78070		Transducin WD40 repeat-like superfamily protein	29.4	4.6	3.8	7.8	0	-6.39		
A	Qs-dev_rep_c104053	Q8RWY3	AT3G06400	CHR11	Chromatin-remodeling protein 11	17	0	1.2	8.4	0.8	-INF		
A	Contig24155	Q8L9J7	AT1G21460	SWEET1, ASWEET1	Nodulin MIN3 family protein	33.8	5.6	1.6	5.2	9.2	-6.04		
A	Contig9359	Q84A17	AT1G15750	WSIP1, TPL	Transducin family protein / WD-40 repeat family protein	96.4	46.8	56	31.8	26.4	-2.06		
A	Contig17751	B8SD87	AT5G58320	NET4A	Kinase interacting (KIP1-like) family protein	23.8	0	0.6	4.2	2	-INF		
A	Contig2868	P0C3R0	AT5G43820		Pentatricopeptide repeat (PPR) superfamily protein	24.2	0.2	1.6	1.2	0	-121		
A	Qs-dev_c16004	Q93WK5	AT5G02810	PRR7, APRR7	Pseudo-response regulator 7	22.8	0	0	1.4	0	-INF		
A	Contig6278	B8IL99	AT5G14280		DNA binding storekeeper protein-related	28.4	0.8	3	1.2	0	-36.5		
A	Isotig12822	Q6ZVC9	AT2G27110	FRS3	FAR1-related sequence 3	24.6	3	19	13.6	5.2	-8.2		
A	Contig23988	O04136	AT5G25220	KNAT3	KNOTTED1-like homeobox gene 3	35.6	0.8	3.8	11.6	13.8	-44.5		
A	Qs-dev_rep_c73055	Q9FDW1	AT4G37260	MYB73, ATM7B73	Myb domain protein 73	39	8.2	0.2	9.2	36.2	-4.76		3.93
A	Contig8294	Q84L30	AT5G38470	RAO23D	Rad23 UV excision repair protein family	30.2	28.4	3	24.2	28.4		-9.47	8.07
A	Contig6713	Q39117	AT1G76880		Duplicated homeodomain-like superfamily protein	47.2	29.8	31.2	28.8	0			-INF
A	Contig26900	Q8LDR0	AT5G06050	QBP4	QBF binding protein 4	18.4	0	4.2	2.2	0	-INF		
A	Contig17748	Q2V9B0	AT1G74840		Homeodomain-like superfamily protein	29.2	0	9.2	0	13.2	-INF		
A	Contig16714	Q10M00	AT5G12440		CCCH-type zinc fingerfamily protein with RNA-binding domain	23.6	10.8	3.6	23.2	0			-INF
A	Contig7566	Q8FHH8	AT5G57660	ATCOL5, COL5	Zinc finger protein CONSTANS-like 5	51.8	7.4	23.4	15	5.2	-7		
A	Contig24135	I1NG96	AT1G81770		Chaperone DnaJ-domain superfamily protein	63.4	7.4	16.2	8.2	25.8	-8.57		
A	Contig2448	Q0WKY2	AT1G73875		DNAse -like superfamily protein	37.8	2.2	0	3.4	4.6	-17.18		
A	Contig19360	Q64937	AT5G06030		GTP binding Elongation factor Tu family protein	89.2	49.8	18.2	7.4	60			8.11
A	Qs-dev_rep_c74679	P20973	AT2G30110	ATUBA1, MOSS5, UBA1	Ubiquitin-activating enzyme 1	21.4	0.6	4.6	3.8	0.8	-36.67		
A	Contig8483	Q04980	AT3G54110	ATJ3, ATJ	DNAJ homologue 3	428.2	203	161.2	281.2	281.2	-2.11		1.74

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts					Fold-change					
						S1	S2	S3S4	S5	S8	SZS1	S3S4S2	S5S3S4	S6S5		
A	Qs-dev_rep_c73054	F6HXQ0	AT3G14200		Chaperone DnaJ-domain superfamily protein	54.6	0.6	9	9.8	24.6	-91					
A	Qs-dev_c8566	B9SGT3	AT3G55020		YnfRab-GAP domain of gyp1p superfamily protein	23	0	2.2	1.6	0.6	-INF					
A	Contig22703	P31166	AT1G27450	APT1	Adenine phosphoribosyl transferase 1	60.4	21.4	11.6	3.4	4.6	-2.82					
A	Contig18126	Q9SVI8	AT4G13850	ATGRP2, GR-RBP2, GRP2	Glycine-rich RNA-binding protein 2	19.8	0.8	8.2	5.2	6.6	-24.75					
A	Contig16854	Q9FJH6	AT5G08790	ATGCN1, GCN1	ABC transporter family protein	325	142.4	79.8	94	117.6	-2.28	-1.78				
A	Contig4664	A5AR63	AT4G26600		S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	56.8	7	12	18.4	7.4	-8.11					
A	Contig25303	Q3M1H3	AT3G45440	ATGCN4, GCN4	General control non-repressible 4	75.4	32.8	29.8	33.2	27.8	-2.3					
A	Contig26214	B9PLJ2	AT3G28480		Oxoglutarate/iron-dependent oxygenase	33.8	5.6	3.8	21.6	11.6	-6.04					
A	Qs-dev_rep_c54887	Q8VZD5	AT5G26751	ATSK11, SK 11	Shaggy-related kinase 11	21.4	2	3.4	8.4	3.8	-10.7					
A	Contig1344	Q3M088	AT4G17180		O-Glycosyl hydrolases family 17 protein	33.2	6	8	2	4.2	-5.53					
A	Qs-dev_rep_c52484	B9RT70	AT4G34480		O-Glycosyl hydrolases family 17 protein	21.4	0.6	0.4	6.2	1.4	-35.67					
A	Contig9890	Q86S20	AT1G71695		Peroxidase superfamily protein	87.4	21	22.2	9.6	7.4	-4.16					
A	Contig18579	ATNY33	AT5G83390		Peroxidase superfamily protein	33.8	6.8	19.4	5.2	6.6	-4.97					
A	Qs-dev_c10343	Q86UP8	AT4G33865		Ribosomal protein S14pS29e family protein	42.8	5.2	0	9.2	5.2	-8.23					
A	isalg03391	Q8SHY2	AT1G85730	YSL7	YELLOW STRIPE like 7	19	0.2	0.8	1.2	0	-95					
A	Contig17888	Q8SAG3	AT2G31200	ADP6, ATADP6	Actin depolymerizing factor 6	39.4	5.6	11.6	5.6	11	-7.04					
A	Contig7205	Q8ER00	AT3G04760		Pentatricopeptide repeat (PPR-like) superfamily protein	81.2	1.6	4.2	3	0	-50.75					
A	Contig1108	Q8J163	AT5G40230	UMAMIT37	Nodulin MN21 / EranA-like transporter family protein	22.6	1.6	3	8.2	0	-14.13					
A	Contig19089	Q85554	AT4G30960	CRPK6, SIP3, SNRKG3.14, ATCIPK6	SCG3-interacting protein 3	41.8	0.8	44	17.6	25.6	-52.25	55				
A	Contig1133	Q8M9V8	AT1G18890	ATCDPK1, CPK10, CDPK1, AICPK10	Calcium-dependent protein kinase 1	24.2	2.2	7.4	3	0	-11					
A	Contig1565	Q8LGD4	AT5G45840		Leucine-rich repeat protein kinase family protein	109.4	0	0.6	0.4	0	-INF					
A	Contig23434	Q85609	AT5G03730	CTR1, SIS1, AICTR1	Protein kinase superfamily protein	46.4	8	3.8	5.2	4.6	-5.8					
A	Contig22	D7TL46	AT3G03770		Leucine-rich repeat protein kinase family protein	27.8	17.4	17.4	14.6	0						-INF
A	Contig4272	Q8SGP2	AT1G28440	HSL1	HAESA-like 1	76.4	36.2	13.2	1.2	2.2	-2.17					
A	Contig11748	P08770	AT3G42170	DAYSLEEPER	BED zinc finger: NAT family dimerisation domain	48.2	2.2	3.6	5.6	4.2	-21.91					
A	Qs-dev_c10451	Q03194	AT2G18860	AHA1, PMA, OST2, HA1	Hy(+)-ATPase 1	79.2	68.4	53.2	64.8	5.2						-12.46
A	Contig13325	Q85TE1	AT4G21300		Tetratricopeptide repeat (TPR)-like superfamily protein	17.6	0	0	0	0	-INF					
A	Qs-dev_rep_c76052	Q8S4B3	AT1G11650	ATRBP45B, RBP45B	RNA-binding (RRMRBD/RNP motifs) family protein	39	0	0.2	5.2	34.8	-INF					6.69

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts						Fold-change			
						S1	S2	S3S4	S5	S8	S2/S1	S3S4/S2	S5/S3S4	S8/S5	
A	Qs-dev_c17608	P42731	AT1G49760	PAB8, PABP8	Poly(A) binding protein 8	113.2	79.8	44.8	96.4	101.6				2.15	
A	Qs-dev_frp_84406	QRLB4	AT1G11650	ATRBP4B, RBP4B	RNA-binding (RRM/RDRP motifs) family protein	21.4	0	4.6	4.8	8.2	-INF				
A	Qs-dev_c16370	P93736	AT1G14610	TWN2, VALRS	Vai1-RNA synthetase / valine-RNA ligase (VALRS)	44.8	5.2	8.8	8.2	10.2	-8.62				
A	Qs-dev_c42593	Q6S7E9	AT1G70580	AOAT2, GST2	Alanine-2-oxoglutarate aminotransferase 2	32.6	5.2	2.2	12	26.6	-8.27				
A	Contig2864	Q10CQ8	AT5G35840	PHYC	Phytochrome C	46.4	11	19.8	23.8	17.8	-4.22				
A	Contig12665	F6GUQ3	AT5G58610		PHD finger transcription factor	17.6	0.2	0	0.4	0	-88				
A	Contig4505	Q6AT68	AT1G56010	NAC1, ANAC022	NAC domain containing protein 1	34.2	2.2	0.6	0	0	-15.55				
A	Contig9941	Q24646	AT2G40950	B2IP17	Basic-leucine zipper (bZIP) transcription factor family protein	57.2	15	5.2	21.6	7.4	-3.81				
A	Qs-dev_c3899	Q6CAJ0-2	AT1G72770	HAB1	HYPERSENSITIVE TO ABA1	30.6	5	14.4	19.4	5.6	-6.12				
A	Contig5781	Q23090	AT4G00870		Basic helix-loop-helix (bHLH) DNA-binding superfamily protein	40.6	25.6	18.4	0.4	2.2				-46	
A	Qs-dev_c17741	B9GW66	AT1G72650	TRFL6	TRF-like 6	28.4	0.4	0	2.2	5.2	-71				
A	Contig24336	Q6C5U1	AT1G27320	AKH3, HK3	Histidine kinase 3	31.6	1.6	0	0.4	4.6	-19.75				
A	Contig25823	F6HD66	AT2G01060		Myb-like HTH transcriptional regulator family protein	24.2	1.6	17.6	5.2	9.4	-15.13				
A	Contig24748	Q6YK4	AT3G53340	NF-YB10	Nuclear factor Y, subunit B10	33.8	0.2	1.6	14.8	4.6	-69				
A	Contig1236	Q6LJT8-2	AT2G38960	TK11	TSL kinase interacting protein 1	19.6	17.8	3.6	17.6	0				-INF	
A	Contig18421	Q6ZSY8	AT4G29080	PAP2, IAA27	Phytochrome-associated protein 2	111.8	98.6	35.2	42.2	25.6			-2.8		
A	Contig23086	Q6K00	AT1G34190	ana2017, NAC017	NAC domain containing protein 17	20.2	0	0	11	4.6	-INF				
A	Contig17915	Q71H73	AT5G10980	H3.3	Histone superfamily protein	138.6	0	36.8	35.2	19.2	-INF				
A	Contig25381	Q6MTQ2	AT3G19290	ABF4	ABRE binding factor 4	26.2	20.4	10.8	16.8	0				-INF	
A	Contig5114	P46607	AT1G79840	GL2	HD-ZIP IV family of homeobox-leucine zipper protein with lipid-binding START domain	38.6	27.8	18.8	21.6	2.2				-9.82	
A	Contig10613	A4PBL4	AT3G54280	RGD3	DNA repair and recombination protein	39.2	3.8	0.6	1.2	8.8	-10.32				
A	Contig22404	B6GY13	AT1G30970	SUF4	zinc finger (C2H2 type) family protein	27.8	0.8	11.6	5.6	4.6	-34.75				
A	Qs-dev_c25830	Q7XA39	AT1G30970	SUF4	zinc finger (C2H2 type) family protein	22.8	0.4	0.8	0	3	-57				
A	Contig4686	B6GV12	AT1G72220		RING/U-box superfamily protein	26.4	4.2	11	2.2	0	-6.28				
A	Qs-dev_c3005	P93736	AT4G10320		RNA synthetase class I (I, L, M and V) family protein	59.6	20.4	10	13.8	0.6	-2.92				
A	Contig3931	B6S3Y9	AT4G03090	ATNDX, NDX	sequence-specific DNA binding	67.6	11.6	0.6	6.8	16	-5.83				
A	Contig5146	Q6632	AT4G13180		NAD(P)-binding Rossmann-fold superfamily protein	32	1.6	0	2.2	7.4	-20				
A	Contig20395	Q6MZY9	AT5G57620	MYB36, AMYB36	MYB domain protein 36	24.2	0	0	0	0	-INF				

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts						Fold-change	
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4
A	Qs-dev_20887	Q700C2	AT1G20980	SPL14, FBR6, SPL1R2, AT SPL14	Squamosa promoter binding protein-like 14	19	0.8	4.6	9.8	5.2	-23.75		
A	Config4757	P33543	AT3G24660	TMKL1	Transmembrane kinase-like 1	42.2	11.6	4.2	6.8	0	-3.64		
A	Qs-dev_rep_082540	Q6LR11	AT3G56370		Leucine-rich repeat protein kinase family protein	28.4	2	0.2	0	0.8	-14.2		
A	Config3384	Q6FE20	AT5G18610		Protein kinase superfamily protein	69.8	4.2	17.6	3	2.2	-16.62		
A	Qs-dev_rep_072662	P26595	AT1G20693	HMG82, HMG BETA 1, NFD2, NFD02	High mobility group B2	61.6	12.2	23.6	30	10.4	-5.05		
A	Config16253	Q2V980	AT5G47390	MYBH	Myo-like transcription factor family protein	99.2	80.8	33.8	58.2	59.6		2.39	
A	Config8386	Q6VYT5	AT5G07670		RN1-like superfamily protein	18.8	0	2	0	0	-INF		
A	Config17491	Q38827	AT5G65670	IAA9	Indole-3-acetic acid inducible 9	30.8	9	29.6	25.6	2			-12.8
A	Config2558	F6HGL7	AT1G71780	SKIP	Chromatin protein family	24.2	3	1.6	9.4	9.4	-8.07		
A	Config22177	Q23894	AT3G05530	RPT5A, AT56A.2	Regulatory particle tripe-A ATPase 5A	31.6	12.6	3.8	29.4	13.2			7.74
A	Config835	Q8SEX2	AT4G28000		P-loop containing nucleoside triphosphate hydrolases superfamily protein	52	14	23.4	9	0	-3.71		
A	Qs-dev_rep_077078	P54774	AT3G09840	CD48, ATCD48, CDC48A	Cell division cycle 48	30.6	4.2	1.2	2.2	3.6	-7.29		
A	Qs-dev_rep_0101936	Q6FXT9	AT1G53750	RPT1A	Regulatory particle tripe-A 1A	24.8	2	0	0	3.4	-12.4		
A	Qs-dev_04329	Q8SEX2	AT2G27600	SKO1, VPS4, ATSKO1	AAA-type ATPase family protein	30.6	1.2	0.8	3	0	-25.5		
A	Config2831	P0D11	AT5G50670	SPL13B, SPL13	Squamosa promoter-binding protein-like (SBP domain) transcription factor family protein	26	1.6	0	0	0	-16.25		
A	Config9105	Q6JIM3	AT5G62000	ARF2, ARF1-BP, HSS, ORE14	Auxin response factor 2	95.8	84.2	80.8	60.4	22.6			-2.07
A	Qs-dev_20380	Q6BDA0	AT5G04240	ELF6	Zinc finger (C2H2 type) family protein	32.4	0	0.8	3	10.4	-INF		
A	Qs-dev_rep_075208	P12333	AT2G34430	LHB181, LHC31.4	Light-harvesting chlorophyll-protein complex II subunit B1	32.8	0	16.6	0.8	0.8	-INF		INF
A	Config12525	Q40392.2	AT1G72890		Disease resistance protein (TIR-NBS class)	17.4	0	0	0	0	-INF		
A	Config23089	Q6R0H1	AT5G17300	RVE1	Homeodomain-like superfamily protein	20.4	1.6	0.6	1.2	2.2	-12.75		
A	Config14583	Q6ZTX9				21.6	0	0	0	0	-INF		
A	Config17537	P10979	AT2G41250		Haloacid dehalogenase-like hydrolase (HAD) superfamily protein	95	46.2	32.6	53.6	92.2	-2.06		
A	Config18322	HKDM3	AT3G15010		RNA-binding (RRMRBD/RNP motifs) family protein	53.2	17.4	20.4	19	17.6	-3.06		
A	Config16462	Q6SVY3	AT4G27410	R226	NAC domain transcriptional regulator superfamily protein	608.8	184.2	138.6	32	11	-3.31		-4.33
A	Config5635	Q6ZTX6.2	AT1G30330	ARF6	Auxin response factor 6	63.4	42.8	52.2	18.4	20.4			-2.84
A	Qs-dev_rep_073351	Q6SSA8	AT1G53910	RAP2.12	Related to AP2.12	54.8	1.2	0.2	7	22.8	-45.67		
A	Config25608	Q6SD53	AT4G31980		Unknown protein	53.8	16.2	29.8	0	0	-3.32		-INF
A	Config4609	Q6VWG3	AT1G08260	WIP3	WIP domain protein 3	34.2	2.2	1.6	1.2	0	-15.55		

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts						Fold-change		
						S1	S2	S3S4	S5	S8	S2/S1	S3S4/S2	S5/S3S4	S8/S5
A	Os-dev_rep_59576	Q9S156	AT2G35940	BLH1, EDA29	BE1-like homeodomain 1	25	3	15.6	11.4	1.6	-8.33			
A	Os-dev_c3411	Q8GY23	AT1G55860	UPL1	Ubiquitin-protein ligase 1	85	20.4	14.4	31.2	30.6	-4.17			
A	Contig28463	Q8ABU1	AT1G11390		Protein kinase superfamily protein	26.2	3.8	0.6	2.2	0	-6.89			
B	Contig3184	B9I385	AT1G60190	ATPUB19, PUB19	ARM repeat superfamily protein	14.4	38.6	25.6	16.2	0				-INF
B	Contig2614	F6HTV6	AT5G24260		Proyl oligopeptidase family protein	24.2	48.2	33.2	32	0				-INF
B	Contig20479	A6M/X8	AT1G76810		Eukaryotic translation initiation factor 2 (eIF-2) family protein	4.6	36	27.8	18.2	15.6	7.83			
B	Contig11877	Q0WT31	AT2G34300		Sadenosyl-L-methionine-dependent methyltransferases superfamily protein	1.2	33.2	0	17.6	0	27.67	-INF	INF	-INF
B	Os-dev_c6207	Q0WT31	AT2G34300		Sadenosyl-L-methionine-dependent methyltransferases superfamily protein	24.8	41.8	0	8.4	8.2		-2.0E-31		
B	Contig25320	B9DFI7	AT1G28850		Sadenosyl-L-methionine-dependent methyltransferases superfamily protein	0	15.6	9.4	8.8	2.2	INF			
B	Contig18332	F6HDT0	AT3G32790	NET1A	Knase interacting (KIP1-like) family protein	142.4	143.8	126.6	64.4	23.6				-1.97
B	Contig23381	Q8CAJ0	AT1G72770	HAB1	HYPERSENSITIVE TO ABA1	12.2	42	21.6	16.6	17.8	3.44			
B	Contig16265	Q39613	AT2G21130	ATPP2CA, AHG3, PP2CA	Cyclophilin-like peptidyl-prolyl isomerase family protein	165.6	192.4	101	170.4	85		-1.9	1.69	-2
B	Contig5193	P49598	AT3G11410		Protein phosphatase 2CA	38.6	82.4	48.8	63.2	9.6	2.13			-6.58
B	Contig18768	Q8LIR6	AT3G32920	BAM1, BMV7, TR-BAMY	Beta-amylase 1	1.2	29.6	0.6	1.2	2.2	24.67	-49.33		
B	Contig17656	O24606	AT3G20770		Ethylene insensitive 3 family protein	37.2	52.2	14.8	29	8.8				
B	Os-dev_c10209	Q9S156	AT2G35940	BLH1, EDA29	BE1-like homeodomain 1	90.6	91	40	7.4	3				-5.41
B	Contig16211	Q2HU68	AT5G02560	HTA12	Histone H2A 12	6.6	18.4	0	0	0	-INF			
B	Contig5234	Q8SRP5	AT2G38470	WRKY33, ATWRKY33	WRKY DNA-binding protein 33	32	55.6	40.6	10.4	7.4			-3.9	
B	Contig1681	Q8LEZ3	AT5G08130	BIM1	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein	0	22.8	0	1.2	8.8	INF	-INF		
B	Contig4677	B9R903	AT1G74840		Homeodomain-like superfamily protein	58.8	110.8	68.6	64.4	30.2	1.88			
B	Contig25196	B9RAK3	AT5G63320	NPX1	Nuclear protein X1	22	23.8	3	14.8	9.2			-7.93	
B	Os-dev_c1492	P48259	AT4G14960	TUA6	Tubulin/Flaz family protein	5.6	62.4	44.8	18.4	24.4	11.14			
B	Os-dev_c17408	D7TON4	AT2G37025	TRFL8	TRF-like 8	7.4	35.6	24.6	18	14.4	4.81			
B	Os-dev_rep_c73916	Q2MHE4	AT4G38470	STY46	ACT-like protein tyrosine kinase family protein	0.4	22	3.4	0.8	0.8	55			
B	Contig22608	P60300	AT5G33280		Voltage-gated chloride channel family protein	0	18.4	5.2	4.2	0	INF			
B	Contig1207	Q8ZSY2	AT5G03160	ATP58IPK, P58IPK	Homolog of mammalian P58IPK	0	15.6	7	2	0	INF			
B	Isotig13010	D7TRL1	AT2G32400		Sadenosyl-L-methionine-dependent methyltransferases superfamily protein	1.6	21	3	9.6	10	13.13			
B	Contig192	A6BNC8	AT4G30100		P-loop containing nucleoside triphosphate hydrolases superfamily protein	15.8	54.6	22.8	36	35.4				

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts						Fold-change			
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4	S6/S5	
B	Contig31	P46869	AT2G21300		ATP binding microtubule motor family protein	20.4	31.4	12	22.8	2.2					-10.36
B	Os-dev_rfp_c37878	Q8VZD5	AT5G14640	ATSK13, SK13	Slaggy-like kinase 13	4.2	38.6	2.2	7.4	3.8	9.19	-17.55			
B	Contig17610	Q39011	AT4G18710	BIN2, DWF12, UCU1, ATSK21, SK21	Protein kinase superfamily protein	14	86.8	17.2	5.6	8.8	6.2	-5.05			
B	Contig19325	Q01390	AT3G43190	SUS4, ATSU54	Sucrose synthase 4	232	351.6	184.2	95	23.8	1.52	-1.91	-1.94	-3.09	
B	Contig21561	Q8M111	AT4G02280	SUS3, ATSU53	Sucrose synthase 3	1.2	44	5.2	29.4	19.8	36.67	-8.46	5.65		
B	Contig21559	Q24301	AT4G02280	SUS3, ATSU53	Sucrose synthase 3	0	20.4	5.2	5.2	15.6	INF				
B	Isolg15724	AGP7U9	AT2G05790		O-Glycosyl hydrolases family 17 protein	5.6	18.4	0.8	3.8	10		-23			
B	Contig16729	Q8L868	AT1G32860		Glycosyl hydrolase superfamily protein	6.6	31.8	25	20.4	6.6	4.82				
B	Contig5738	BBRE5	AT4G34480		O-Glycosyl hydrolases family 17 protein	14.4	54.8	43.6	39.2	28.4	3.81				
B	Contig23886	Q86520	AT1G71695		Peroxidase superfamily protein	5	21.6	0.6	9.4	0	-36				
B	Contig17499	Q8FV11	AT5G59890	ADF4, ATADF4	Actin depolymerizing factor 4	19.8	31.2	25	18.2	0					INF
B	Contig20510	Q8S0F0	AT3G25500	APH1, FH1, AHF1, ATFH1	Fomlin homology 1	29.6	33.8	1.6	17.6	4.6	-21.13				
B	Contig17621	Q22932	AT2G30360	CIPIK11, PK55, SIP4, SNRK3.22	SOS3-interacting protein 4	37.2	78.4	41.6	23.2	2					-11.6
B	Contig26056	BBT6V8	AT4G30960	CIPIK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	27.8	153.8	42	86.2	27.8	5.53	-3.66			-3.1
B	Os-dev_rfp_c73270	BBT6V8	AT4G30960	CIPIK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	21.4	45	2.2	20	3.4	-20.45				
B	Contig21718	E3V0H9	AT3G60240	EIF4G, CUM2	Eukaryotic translation initiation factor 4G	24.2	30.6	12.8	28.4	2.2					-12.91
B	Contig6898	Q42371	AT2G26330	ER, QRP1	Leucine-rich receptor-like protein kinase family protein	11	49.2	19.8	27.4	5.2	4.47				-5.27
B	Contig20100	Q8W3M4	AT3G19320		Leucine-rich repeat (LRR) family protein	15.8	24.6	3	2	0	-8.2				
B	Contig18343	Q8W3M4	AT3G19320		Leucine-rich repeat (LRR) family protein	19.8	56.8	14	5.2	0	2.87	-4.06			
B	Os-dev_rfp_c76199	11JE11	AT4G34100	CER9, SUD1	RING-U-box superfamily protein	4.2	31	1.2	0.8	14.8	7.38	-25.83			
B	Contig518	B8S1X8	AT3G30300		O-Glucosyltransferase family protein	5.2	37	3	12.6	12	7.12	-12.33			
B	Contig9545	B8RLH9	AT5G15740		O-Glucosyltransferase family protein	19	80.4	46	36.2	16.2	4.23				
B	Contig25539	APQCB8	AT4G24530		O-Glucosyltransferase family protein	8.8	35.6	23	21.4	11.6	4.05				
B	Contig6516	Q42806	AT3G52990		Pyruvate kinase family protein	7.4	39.6	11	35.4	12	5.35	-3.6			
B	Os-dev_c30725	Q8LIR6	AT3G23920	BAM1, BMY7, TR-BAMY	Beta-amylase 1	0	23.8	0	9.2	0	INF				-INF
B	Os-dev_rfp_c76744	Q8LIR6	AT3G23920	BAM1, BMY7, TR-BAMY	Beta-amylase 1	0.4	54.2	0.2	8.4	0.8	135.5	-271			
B	Os-dev_rfp_c76810	PAZ731	AT4G34110	PAB2, PABP2, ATPAB2	Poly(A) binding protein 2	0.4	17.6	2.2	11.4	12.6	44				
B	Contig632	Q38214	AT3G07040	RPM1, RPS3	NB-ARC domain-containing disease resistance protein	1.6	16.8	0	0.6	2.2	-INF				-INF

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts					Fold-change				
						S1	S2	S3S4	S5	S8	S2/S1	S3S4/S2	S5/S3S4	S8/S5	
B	Contig25295	Q8H1D3	AT4G31820	ENP, NPY1, MAB4	Phototropic-responsive NPH3 family protein	3	22	0.6	1.2	0		-36.67			
B	Contig120	Q8Y564	AT5G06530		ABC-2 type transporter family protein	8.2	31.2	1.6	0.4	0		-19.5			
B	Os-dev_c5280	D7TB03	AT1G87210		Calmodulin-binding transcription activator protein with CG-1 and Ankyrin domains	9.6	38.6	11.2	11.2	3.2	4.02				
B	Contig3954	F6HA24	AT5G12330	LRP1	Lateral root primordium (LRP) protein-related	36	35.2	6.6	0	0		-5.33			
B	Contig17999	P25469	AT5G59870	HTA6	Histone H2A 6	12	23.6	0	0	0		-INF			
B	Contig414	Q6C9F4	AT1G68190	BBX27	B-box zinc finger family protein	1.4	23.2	22	10.4	0	16.57				
B	Os-dev_c243	B9GY63	AT1G30820		CTP synthase family protein	32.4	119.6	111.8	43.2	5.4	3.69		-2.59	-8	
B	Contig16967	Q8LB81	AT5G33370		GDSL-like Lipase/Acylhydrolase superfamily protein	29	36	31.6	2	0			-15.8		
B	Contig20184	Q8LY94	AT5G14450		GDSL-like Lipase/Acylhydrolase superfamily protein	6.6	20.4	0.6	1.2	6.8			-34		
B	Contig5473	P46897	AT3G61890	ATHB-12, ATHB12, HB-12	Homeobox 12	40.4	88.4	49.8	12	2.2	2.19		-4.15		
B	Contig6190	Q7XA39	AT3G14470		NB-ARC domain-containing disease resistance protein	7.4	30.8	0.6	19.2	0		-51.33	32	-INF	
B	Contig21578	Q6SSA8	AT1G53910	RAP2.12	Related to AP2 12	33.8	67.6	39.6	46.4	9.2				-5.04	
B	Contig21580	Q6SSA8	AT1G53910	RAP2.12	Related to AP2 12	51	107.4	79	76	73.2	2.11				
B	Contig23985	Q8K00	AT1G34190	anao2017, NAC017	NAC domain containing protein 17	0	27.8	22.8	18.4	7	INF				
B	Contig19683	Q9SL66	AT2G35940	BLH1, EDA29	BEL1-like homeodomain 1	37.4	53	38.6	17.6	0				-INF	
B	Contig15517	Q71H73	AT5G10980	H3.3	Histone superfamily protein	156	325	126.2	86.4	40	2.08	-2.58		-2.16	
B	Contig8769	QJUL1	AT1G60420	ATNRX1, NRX1	DC1 domain-containing protein	22.8	92.2	31.4	9	9.8	4.04	-2.94			
B	Contig22218	Q9FW2	AT4G32551	LUG	LisH dimerisation motif WD40/YVTN repeat-like-containing domain	1.2	22.8	21.6	11	6.8	19				
B	Contig22403	G7LC02	AT1G30970	SUF4	Zinc finger (C2H2 type) family protein	3	26.4	3	4.2	6.8	8.8	-8.8			
B	Os-dev_c8617	Q8LDA7	AT3G15260		Protein phosphatase 2C family protein	2.2	21.6	13.2	18.4	8.2	9.82				
B	Contig1296	Q8LK40	AT3G23780	NRPO2A, DRD2, NRPO2, DMS2, NRPE2	nuclear RNA polymerase D2A	45.8	47	39.4	41.6	13				-3.2	
B	Contig25368	Q9FLJ2	AT4G35690	NTL9	NAC transcription factor like 9	19.2	42	8.2	14.8	0		-5.12		-INF	
B	Contig25141	Q64517	AT1G79350	EMB1135	RING-FYVE/PHD zinc finger superfamily protein	27.8	46.4	13	9.4	24.2		-3.57			
B	Contig5250	Q23661	AT2G33860	ETT, ARF3	Transcriptional factor B3 family protein/ auxin-responsive factor AUX/IAA-related	18.4	31.6	32	22	2.2				-10	
B	Contig17627	Q04136	AT5G11060	KNAT4	KNOTTED1-like homeobox gene 4	1.2	34	27.4	5.6	2	26.33				
B	Contig9352	Q8GYM4-2	AT4G23180	CRK10, RLK4	Cysteine-rich RLK (RECEPTOR-like protein kinase) 10	1.6	25	12	15	0	15.63			-INF	
B	Contig2364	Q9SCZ4	AT3G51550	FER	Malacatin/receptor-like protein kinase family protein	10.4	27.8	8.2	23.6	2.2				-10.73	
B	Contig593	P43298	AT2G01820		Leucine-rich repeat protein kinase family protein	36	40.8	40.8	29.6	5.2				-5.7	

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts								Fold-change	
						S1	S2	S3S4	S5	S8	S2/S1	S3S4/S2	S5/S3S4		
B	Contig25266	Q8FE20	AT5G18610		Protein kinase superfamily protein	0	18.4	1.6	15.6	4.6	INF				
B	Os-dev_c25103	Q83ZS4	AT1G06080	NIK3	NSP-interacting kinase 3	1.6	20.4	2.2	18.4	3	12.75				
B	Os-dev_rep_c76751	Q8LJM4	AT1G09970	LRR XI-23, RLK7	Leucine-rich receptor-like protein kinase family protein	2.2	21.2	2.2	3.8	3.6	9.64				
B	Contig24834	P26585	AT1G20893	HMG32, HMG BETA 1, NFD2, NFD02	High mobility group B2	8.8	16.6	0	4.2	4.6	-INF				
B	Os-dev_c2705	A5C0C9	AT5G04700		Ankyrin repeat family protein	1.6	23.8	0	0	0	14.88	-INF			
B	Contig17622	B6S9Y7	AT4G13930	SHM4	Sene hydrosymethytransferase 4	88.4	117.8	58.8	25.6	15.2	-2				
B	Contig12049	O50001	AT5G16680		RINGFYVE/PHD zinc finger superfamily protein	67	113.2	62.8	77.8	27					-2.88
B	Contig22999	Q8MTQ4	AT1G45249	ABF2	abscisic acid responsive elements-binding factor 2	14	36.8	6	5.2	2.2	-6.13				
B	Contig6461	APHG32	AT4G29040	RPT2a	Regulatory particle AAA-ATPase 2A	28.4	50	35.6	6.8	28.4				-5.24	
B	Os-dev_rep_c73918	P81327	AT3G16290	EMB2083	AAA-type ATPase family protein	17.2	37.6	9	8.4	5.8	-4.18				
B	Contig25549	Q8ZUJ0	AT2G37260	TTG2, ATWRKY44, WRKY44, DSL1	WRKY family transcription factor family protein	10.4	27.8	3.8	21.4	13.8	-7.32				
B	Contig10306	Q8C5E7	AT4G25880	APUM6, PUM6	Pumilio 6	11.8	31.8	5.8	7.6	8.8	-5.48				
B	Contig18231	Q8ELU9	AT2G45950	AGL6	AGAMOUS-like 6	17.8	53.6	26.2	0	0	3.01	-INF			
B	Contig47	Q8EVK6	AT2G46020	CHR2, ATBRM, BRM, CHA2	Transcription regulatory protein SNF2, putative	27.8	64.4	51	45.4	9.6				-4.73	
B	Contig21867	Q8LV14	AT3G17880	JAZ3, JAZ3, TIFYBB	Jasmonate-zim-domain protein 3	12.2	37.2	26.4	18.4	0				-INF	
B	Contig1592	Q8SD53	AT4G31980		Unknown protein	4.6	16.2	0	0	0	-INF				
B	Contig20250	Q8SD53	AT2G44930		Plant protein of unknown function (DUF247)	1.2	22	0.6	0	0	18.33	-36.67			
B	Contig4083	P0C8S1	AT1G53350		Disease resistance protein (CC-NBS-LRR class) family	0	19.8	0	0	0	INF	-INF			
B	Contig17620	Q80763	AT1G60420	ATNRX1, NRX1	DC1 domain-containing protein	1.2	50.8	49.2	41.6	35.4	42.33				
B	Contig25136	Q8GY23	AT1G55960	UPL1	Ubiquitin-protein ligase 1	42.2	94	47.4	37.8	44.6	2.23				
B	Isdig13728	P37116	AT4G24520	ATR1, AR1	P450 reductase 1	27	54.2	12.2	34	7.4	4.44			-4.59	
B	Contig5629	Q80763	AT1G60420	ATNRX1, NRX1	DC1 domain-containing protein	38.6	113.8	69	18.4	16.2	2.95			-3.75	
C	Contig1106	C0LGV4	AT2G16250		Leucine-rich repeat protein kinase family protein	19.4	42	51.4	10.6	0				-4.85	
C	Os-dev_c8085	A5AU06	AT1G30360	ERD4	Early-responsive to dehydration stress protein (ERD4)	63.8	24.2	96.8	24.4	30.8	-2.64	4	-3.97		
C	Contig25256	Q8WT31	AT2G34300		Sadenosyl-L-methionine-dependent methyltransferases superfamily protein	16.2	0.6	34	24.2	2.2	42.5			-11	
C	Contig6985	Q8L7V3	AT5G64030		Sadenosyl-L-methionine-dependent methyltransferases superfamily protein	61.2	53.8	77.2	41.4	5.2				-7.86	
C	Contig9099	Q8C927	AT1G52360		Odontomer, beta' subunit	28.4	20.4	56	39.4	22.6	2.75				
C	Contig10291	A5BFP1	AT1G30360	NET1D	Kinase interacting (KIP1-like) family protein	19.4	35.8	55.2	27.6	4.2				-6.37	

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts						Fold-change		
						S1	S2	S3S4	S5	S8	S2/S1	S3S4/S2	S5/S3S4	S8/S5
C	Contig19653	Q8L3J7	AT1G21460	SWEET1, ASWEET1	Nodulin MIN3 family protein	12.6	12.6	19.6	1.2	0			-16.33	
C	Contig16318	Q844G9	AT4G38970	FBA2	Fructose-bisphosphate aldolase 2	10.4	10.4	20.4	0	0			-INF	
C	Os-dev_rep_c75026	Q844G9	AT4G38970	FBA2	Fructose-bisphosphate aldolase 2	17.6	17.6	27	2.2	0.8			-12.27	
C	Contig19712	Q0WV90-1	AT1G80490	TPR1	TOPILESS-related 1	14	14	54	19	25.6			-2.84	
C	Contig9354	A9PJL2	AT1G84140		Loricrin-related	137.2	137.2	428.2	169.4	88	1.74	3.12	-2.53	-1.92
C	Contig11693	B9GZS4	AT5G84650		Loricrin-related	6.8	6.8	32.4	19.8	2			4.76	
C	Contig3765	O24606	AT3G20770	EIN3, AEN3	Ethylene insensitive 3 family protein	30.2	30.2	71	27.2	42.8			-2.35	-2.61
C	Contig17968	Q940T9	AT5G24930	ATCOL4, COL4	CONSTANS-like 4	22	22	36.8	29.8	0				-INF
C	Os-dev_rep_c92250	O24407	AT3G04730	IAA16	Indoleacetic acid-induced protein 16	0	0	27.8	3.8	1.4			INF	-7.32
C	Contig9124	Q8ZPB7	AT1G54100	ALDHFB4	Aldehyde dehydrogenase 7B4	4.2	4.2	50	27.4	7.4			11.9	
C	Contig6588	Q83ZH2	AT1G72830	HAP2C, ATHAP2C, NF-YA3	Nuclear factor Y, subunit A3	4.2	4.2	33.2	5.8	18.4			7.9	-5.72
C	Contig20974	Q8GX84	AT1G01350		Zinc finger (CCHH-type/C3HC4-type RING finger) family protein	0	0	117.8	5.6	0			INF	-21.04
C	Contig25237	O22881	AT2G40540	KT2, ATK12, SHY3, KUP2, ATRUP2, TRK2	Potassium transporter 2	23	23	25.6	14.8	13.8	19.17			
C	Contig9805	Q42961	AT1G56190		Phosphoglycerate kinase family protein	14	14	53.2	9.2	22.6			3.8	-5.78
C	Contig23884	Q84937	AT5G80390		GTP binding Elongation factor Tu family protein	128.2	128.2	282.8	118.4	59.6			2.05	-2.22
C	Contig18209	Q8MA55-2	AT3G05420	ACBP4	Acy-CoA binding protein 4	9	9	28.4	19.8	0				-INF
C	Contig4207	F4IAE9	AT3G15120		P-loop containing nucleoside triphosphate hydrolases superfamily protein	4.2	4.2	20.6	0.4	0				-51.5
C	Contig3685	B9S5B9	AT5G58470	TAF15b	TBP-associated factor 15B	4.6	4.6	19.8	1.2	9.4				-16.5
C	Os-dev_rep_c73758	Q8FJH6	AT5G80790	ATGDN1, GDN1	ABC transporter family protein	44	44	52.2	15.8	32.4				-3.3
C	Isolig08786	A5BNC8	AT4G30100		P-loop containing nucleoside triphosphate hydrolases superfamily protein	6.6	6.6	33.2	5.2	10			5.03	-6.38
C	Os-dev_c13662	Q8FJRO	AT5G47010	UPF1, LEA1, ATUPF1	RNA helicase	0.8	0.8	30	17.4	26.4	-30.75		37.5	
C	Contig7921	P13708	AT3G43190	SUS4, ATISU4	Sucrose synthase 4	2.2	2.2	216.6	194.2	16.2	-61.73		98.45	-11.37
C	Contig917	Q4ZGK1	AT2G26800		Glycosyl hydrolases superfamily protein	4.6	4.6	16.6	0	5.2				-INF
C	Contig14976	F6HST8	AT5G87460		O-Glycosyl hydrolases family 17 protein	6.8	6.8	26	4.2	0				-6.19
C	Contig17654	Q8Z8P0	AT4G38520	APD6	Protein phosphatase 2C family protein	0.2	0.2	15.8	0	0				-INF
C	Os-dev_c42611	Q8SB81	AT4G21960	PRXR1	Peroxidase superfamily protein	1.8	1.8	38.2	31.8	0			21.22	-INF
C	Contig20996	Q8FJZ9	AT5G86390	PRX72	Peroxidase superfamily protein	0	0	16.2	0	0			INF	-INF
C	Os-dev_rep_c75871	Q86520	AT1G71695		Peroxidase superfamily protein	0.6	0.6	36	2.2	8.2			60	-16.36

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts								Fold-change		
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4			
C	Contig16616	BRRRB8	AT2Q06050	OPR3	Oxaphydriene-oxidase 3	10	17.8	73.4	10	2		4.12	-7.34			
C	Contig22099	Q8S0F0	AT3G25500	AFH1, FH1, AHF1, ATH1	Fomin homology 1	6.8	0.8	42.8	5.2	23.8			53.5	-8.23		
C	Contig19666	O64882	AT2G44480	BGLU17	Beta glucosidase 17	1.2	10.4	46.4	15.6	0			4.46		-INF	
C	Contig25387	O65554	AT4G30960	CPK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	26.2	16.2	31.8	16.2	0					-INF	
C	Contig19389	B8F6V8	AT4G30960	CPK6, SIP3, SNRK3.14, ATCIPK6	SOS3-interacting protein 3	12.2	5.6	32.8	27	2.2			5.86		-12.27	
C	Contig6653	Q42438	AT5G19450	CDPK19, CPK8	Calcium-dependent protein kinase 19	13	18	37.8	21.6	2.4					-9	
C	Isigig4004	P22953	AT5G02500	HSC70-1, HSP70-1, AT-HSC70-1, HSC70	Heat shock cognate protein 70-1	81.4	107.8	142.4	95.8	46.4					-2.06	
C	Contig21549	O81832	AT4G27290		S-loucin protein kinase family protein	0	0	19.8	0	0			INF	-INF		
C	Contig9866	E3V0H9	AT3G60240	EIF4G, CUM2	Eukaryotic translation initiation factor 4G	91.2	100.8	118.4	61.2	57.6					-1.93	
C	Contig25491	Q8LIR6	AT3G23920	BAM1, BMV7, TR-BAMY	Beta-amylase 1	303.6	557.8	608.8	175.8	4.6	1.84		-3.46	-38.22		
C	Contig25376	P42731	AT1G49760	PAB8, PABP8	Poly(A) binding protein 8	18.2	15.6	63.4	16.8	19.8			4.06	-3.77		
C	Contig17557	P93736	AT1G14610	TWNC, VALRS	Vaiy-RNA synthetase / valine-tRNA ligase (VALRS)	15.6	22.4	31.6	6.6	11				4.79		
C	Contig4122	A5APB9	AT3G13224		RNA-binding (RRMRBD/RNP motifs) family protein	16.2	16.6	27.8	4.6	14				-6.04		
C	Contig8453	B8RT8	AT4G25870		Unknown protein	5.2	11	34.2	3.6	3				-9.5		
C	Contig5209	Q7X7L3	AT5G50320	ELO3, HAG3, HAC8, ELP3, AELP3	Radical SAM domain-containing protein / GCN5-related N-acetyltransferase (GNAT) family protein	0	0	22	0	0			INF	-INF		
C	Contig11717	F8HEL5	AT5G55390	EDM2	ENHANCED DOWNY MILDEW 2	17.4	17.8	38.2	17.6	0					-INF	
C	Os_dev_c6871	Q40392	AT5G17860		Disease resistance protein (TIR-NBS-LRR class)	0.2	0.6	24.6	9.2	12.6			41			
C	Contig21209	Q7X440	AT3G14470		NB-ARC domain-containing disease resistance protein	8.8	9.2	17.4	0	0				-INF		
C	Contig20249	Q8LRR4	AT3G14470		NB-ARC domain-containing disease resistance protein	4.6	3	24.2	3.4	0			8.07	-7.12		
C	Contig2223	IL1945	AT2G27930		PLATZ transcription factor family protein	1.2	0.2	17.4	0	0			87	-INF		
C	Isigig2881	P24088	AT1G75390	ABCIP44, z2P44	Basic leucine zipper 44	9.4	4.2	43.6	20	3			10.38			
C	Contig7776	Q8LMA8	AT1G19180	JAZ1, TIFY10A	Jasmonate-zim-domain protein 1	87.4	89.2	96.4	35.4	9.6			-2.72	-3.69		
C	Contig18046	B8S181	AT1G19180	JAZ1, TIFY10A	Jasmonate-zim-domain protein 1	3	17.8	35.2	2	0			-17.6			
C	Contig18108	Q8LNF4	AT1G48040		Protein phosphatase 2C family protein	4.6	17.2	32.8	3.4	2.2				-9.65		
C	Os_dev_c6120	Q30460	AT5G53460	GLT1	NADH-dependent glutamate synthase 1	70.6	40.4	83	31.2	30.8			-2.66			
C	Contig1420	O64518	AT1G79350	EMB1135	RING/PPV/PHD zinc finger superfamily protein	3	0.8	49.2	19.8	21			61.5			
C	Contig2058	Q8LV46-2	AT3G24550	ATPERK1, PERK1	Proline extensin-like receptor kinase 1	12.6	23.8	32	2.2	11.6			-14.55			
C	Contig19114	Q8L196	AT5G49760		Leucine-rich repeat protein kinase family protein	1.2	5.2	36.8	3	2.2			7.08	-12.27		

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts						Fold-change			
						S1	S2	S3S4	S5	S8	S2/S1	S3/S4	S5/S3S4	S8/S5	
C	Contig25870	Q8Y665	AT3G25560	NIK2	NSP-interacting kinase 2	10.6	4.6	31.8	3	0		6.91	-10.6		
C	Contig8228	Q8ZS4	AT2G45340		Leucine-rich repeat protein kinase family protein	16.6	26	40.8	7.4	0			-5.51		
C	Contig4468	Q8LJM4	AT1G09970	LRR_XL23, RLK7	Leucine-rich receptor-like protein kinase family protein	22.4	48.8	85.6	21.6	14			-3.98		
C	Contig16006	B8C543	AT3G58160	XU, ATLXU, ATMVOS3, MYA3, XI-16	P-loop containing nucleoside triphosphate hydrolases superfamily protein	0	0	77.8	0.4	0		INF	-194.5		
C	Contig16884	Q8LDC8	AT3G23240	ERF1, ATERF1	Ethylene response factor 1	0	0.2	18.2	0.4	0		91	-45.5		
C	Contig1477	F6HVN2	AT3G18670		Ankyrin repeat family protein	0	14.2	22.6	0	0			-INF		
C	Contig26605	Q8Z89	AT1G27730	STZ, ZAT10	Salt tolerance zinc finger	1.2	0	18.4	5.8	0			INF		
C	Os-dev_rep_c100634	Q8LMA8	AT3G43740		Leucine-rich repeat (LRR) family protein	0.4	1.2	37.4	15	5.8		31.17			
C	Contig6526	D7TBL3	AT4G29000		Tasmin/TSO-like CXC domain-containing protein	3.2	4.2	31.2	0.6	2.4		7.43	-52		
C	Contig8408	D7LUZ69	AT1G02890		AAA-type ATPase family protein	36.2	22	43.6	39.4	9.8				-4.02	
C	Contig22178	Q23894	AT3G05530	RPT5A, ATS6A.2	Regulatory particle triple-A ATPase 5A	5	13.2	17.4	0.4	0			-43.5		
C	Contig21012	P84774	AT3G53230	ATDC48B	ATPase, AAA-type, CDC48 protein	22	11.4	27.8	1.2	0			-23.17		
C	Os-dev_c22204	Q8LPT7	AT1G48480	RKL1	Receptor-like kinase 1	3.4	3.6	20	1.4	0			-14.29		
C	Isdig13847	Q8LPT7	AT1G48480	RKL1	Receptor-like kinase 1	5.6	12.8	25.8	2.2	7.4			-11.73		
C	Contig17128	Q8Z500	AT1G72890		Disease resistance protein (TIR-NBS class)	0	0.8	30.6	0	0		38.25	-INF		
C	Isdig2401	D7T8M7	AT3G17860	JAZ3, JAI3, TIFY8B	Jasmonate-2m-domain protein 3	5.2	5.2	32	6	0		6.15	-5.33		
C	Contig8592	Q39117	AT1G76880		Duplicated homeodomain-like superfamily protein	13.2	0.8	21	9.6	0		26.25			
C	Contig3938	Q8SWM6	AT5G17420	IRX3, CESA7, ATCESA7, MUR10	Cellulose synthase family protein	18.4	54	84.6	9.6	0		2.93	-8.81		
C	Contig17758	Q841L0	AT5G05170	CESA3, IXR1, ATCESA3, ATH-B, CEV1	Cellulose synthase family protein	39.4	37.8	74	30.4	40.4			-2.43		
D	Contig17112	Q22193	AT4G16490		ARM repeat superfamily protein	4.6	4.6	3.6	21	0				-INF	
D	Contig23	Q8H0T9	AT5G23430		Translucidin/WD40 repeat-like superfamily protein	10.4	9.2	10.4	22	2.2			-10		
D	Contig23468	Q8A440	AT1G82020		Cotumer, alpha subunit	35.6	24.6	33.2	39.8	9.2			-4.33		
D	Os-dev_rep_c72871	Q24076	AT1G18080	ATARCA, RACK1A_AT, RACK1A	Translucidin/WD40 repeat-like superfamily protein	30.6	6.6	27.8	51	16.6		-4.64	-3.07		
D	Contig18026	Q8L9J7	AT1G21460	SWEET11, ASWEET11	Nodulin MN3 family protein	1.2	5.2	0	19.8	0			INF		
D	Contig17948	Q8ZL52	AT2G01140	ATFBA3, FBA3	Adolase superfamily protein	12	34.8	46.4	80.8	35.4			-2.28		
D	Contig23337	Q0WV90-1	AT1G80490	TPR1	TOPLESS-related 1	3	11.4	10.4	41	9.2			3.94	-4.46	
D	Contig8419	Q8L328	AT4G33270	CDC20.1	Translucidin family protein / WD-40 repeat family protein	3.4	13.2	6.6	17.2	0			-INF		
D	Os-dev_c103	D7THK7	AT4G04940		Translucidin family protein / WD-40 repeat family protein	3.4	16	20	24.4	3			-8.13		

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts								Fold-change	
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4		
D	Contig23499	K4AXG9	AT1G11420	ATDUF2, DUF2	DOMAIN OF UNKNOWN FUNCTION 724.2	5	26.4	35.2	51	7				-7.29	
D	Contig8894	Q8FM89	AT5G56420		F-box/RNAP-like domain-containing protein	0	0	0.8	18.4	0				23	-NF
D	Contig19980	P23384	AT5G15800	SEP1, AGL2	K-box region and MADS-box transcription factor family protein	3	14	14.8	29.8	0					-NF
D	Contig6518	O04658	AT1G56110	NOP56	Homolog of nuclear protein NOP56	14.8	7	5.2	24.2	0					-NF
D	Contig1258	Q38953	AT3G26860		ATP-dependent RNA helicase	41.2	53.2	33.8	68.2	21					-3.25
D	Os-dev_rep_276614	Q3M886	AT3G30250	LBD41	LOB domain-containing protein 41	0.4	7.6	9	17.4	0.8					-21.75
D	Contig25127	ABBYR6	AT4G00990		Transcription factor JUNON (JmC) domain-containing protein	46.4	61.8	60	99.2	33.8					-2.93
D	Contig21515	Q2MHE4	AT4G38470	STY46	ACT-like protein tyrosine kinase family protein	4.8	1.6	6	18.4	0					-NF
D	Contig8743	I1U877	AT4G35780	STY17	ACT-like protein tyrosine kinase family protein	16.6	7.2	12	31.8	7.4					-4.3
D	Contig25167	P93028	AT2G30110	ATUBA1, MOSS, UBA1	Ubiquitin-activating enzyme 1	1.2	2.2	26.4	38.6	17.8			12		
D	Os-dev_c4185	Q8FJRO	AT5G47010	UPF1, LBA1, ATUPF1	RNA helicase	42.8	22	16.6	50.8	10.4				3.06	-4.88
D	Contig389	F8189	AT4G39050		Kinesin motor family protein	3	15.8	12	19.8	0					-NF
D	Os-dev_rep_272718	Q8M111	AT4G02280	SUS1, ATSUS3	Sucrose synthase 3	85.2	103.8	122.6	161.2	81.4					-1.98
D	Contig9391	Q8C8I8	AT2G05790		O-Glycosyl hydrolases family 17 protein	18.8	9.4	19	28.2	5.2					-5.42
D	Contig6485	P22195	AT5G05340	PRX62	Peroxidase superfamily protein	14.8	9.8	15.6	34.8	2.4					-14.5
D	Contig26115	Q8SB81	AT4G21960	PRXR1	Peroxidase superfamily protein	10.8	87.4	91	93.8	15.8		8.09			-5.94
D	Os-dev_rep_274662	Q86520	AT1G71695		Peroxidase superfamily protein	5.8	25	10	29.6	5.8					-5.1
D	Os-dev_rep_384640	P08770	AT3G42170	DAYSLEEPER	BED zinc finger RNA family dimerisation domain	4.2	4.2	0.2	16.4	1.2			82		
D	Contig12092	Q8I028	AT2G26790		Pentapeptide repeat (PPR) superfamily protein	1.2	8.4	3.6	26.6	0				7.39	-NF
D	Contig5921	Q8SIC9	AT2G31400	GUN1	Genomes uncoupled 1	32	33.8	13.2	54	12				4.09	-4.5
D	Contig443	Q8D869	AT2G40030	NRPD1B, DRD3, ATNRPD1B, DMSS, NPPE1	Nuclear RNA polymerase D1B	16.2	34.2	37.2	53.8	12					-4.48
D	Contig16892	Q8W417	AT3G51650	CPK13	Calcium-dependent protein kinase 13	17.6	10.8	10.4	35.6	6.6					-5.39
D	Contig42	Q8LRT1	AT3G28040		Leucine-rich receptor-like protein kinase family protein	9	16.2	14.2	27.8	0					-NF
D	Contig5705	Q89545	AT5G65700	BAM1	Leucine-rich receptor-like protein kinase family protein	26.4	20.4	30.8	33.2	0					-NF
D	Contig206	Q8VYM2	AT5G43360	PHIT3, ATP7A, PHIT1.3	Phosphate transporter 1.3	8.4	12	14	23.6	0					-NF
D	Contig17542	Q8Q492	AT1G09160		Protein phosphatase 2C family protein	17.6	13.2	10.4	22	0					-NF
D	Contig8924	Q42806	AT3G52990		Pyruvate kinase family protein	26.6	42.2	30	112.6	34.2				3.75	-3.29
D	Contig25824	BSRU15	AT1G52770		Phototropic-responsive NPH3 family protein	0	0	1.6	25.8	0				16.13	-NF

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts					Fold-change				
						S1	S2	S3S4	S5	S8	S2/S1	S3/S4	S5/S3S4	S8/S5	
D	Contig1147	Q60970	AT2G14820	NPY2	Phototropic-responsive NPH3 family protein	4.6	1.4	1.6	18.8	0					-NF
D	Contig23751	BBR1T8	AT4G25670		Unknown protein	8.8	12.6	10.4	19	0					-NF
D	Contig22816	Q5ZW21	AT2G29380	HA3	Highly ABA-induced PP2C gene 3	12.2	3.8	6	29.8	7				4.97	
D	Contig5456	Q6SWF9	AT3G20230	ZFN1	Zinc finger protein 1	20.4	10.4	20.6	22.8	22					-10.36
D	Contig26880	IL4D4	AT1G82830	LDL1, SWP1, ATSWP1, ATLSO1, LSD1	LSD1-like 1	0	7	9.6	17.6	0					-NF
D	Contig4935	Q6ROH1	AT1G18330	EPRI1, RVE7	Homeodomain-like superfamily protein	9	7	7.4	19.2	0					-NF
D	Contig18515	Q6FVV1	AT1G71250		GDSL-like Lipase/Acylhydrolase superfamily protein	0	0	16	18.2	0					-NF
D	Contig16483	Q6SN23	AT3G49940	LBD38	LOB domain-containing protein 38	6.6	11.2	11.6	26.6	0					-NF
D	Contig20687	Q0D3J9	AT5G12440		CCCH-type zinc fingerfamily protein with RNA-binding domain	35.4	27	31.8	37	9					-4.11
D	Contig19708	Q39196	AT1G01620	PIP1C, TMP-B, PIP1.3	Plasma membrane intrinsic protein 1C	42	33.2	49.2	103	23.8				2.09	-4.32
D	Contig16246	Q6ROH1	AT1G01060	LHY, LHY1	Homeodomain-like superfamily protein	25.4	10.8	21.4	27.8	2					-13.9
D	Contig20668	Q61122	AT1G66340	ETR1, EIN1, ETR, AETR1	Signal transduction histidine kinase, hybrid-type, ethylene sensor	3	29.8	26.4	37.6	13.2	9.93				
D	Contig20803	BBR3K8	AT5G51760	AHG1	Protein phosphatase 2C family protein	4.6	9.6	12.6	19.8	0					-NF
D	Contig5982	F4JIF5	AT4G14770	TOX2, ATTQX2	TESMINTS01-like CXC 2	32	26.4	18.4	30	22					-13.64
D	Contig25628	Q6L999	AT5G42520	BPC6, BBR/BPC6, ATBPC6	Basic pentacyclic 6	20.4	17.6	9.4	21.4	22					-9.73
D	Contig8096	Q65020	AT3G51770	ETO1	Tetratricopeptide repeat (TPR)-containing protein	24.6	1.6	11	37.2	14.2	-15.37				
D	Contig16907	P51846	AT4G33070	ATPDC1, PDC1	Thiamine pyrophosphate dependent pyruvate decarboxylase family protein	10	7.8	19.2	106	8.8				5.52	-12.05
D	Contig18237	P51846	AT4G33070	ATPDC1, PDC1	Thiamine pyrophosphate dependent pyruvate decarboxylase family protein	17.8	0.8	3.8	24.6	6.6	-22.25				
D	Contig7078	Q03489	AT1G24260	SEP3, AGL9	K-box region and MADS-box transcription factor family protein	1.6	6.6	5.2	15	0					-NF
D	Contig4603	Q61001	AT2G32840	AIGR1, GRF1	Growth-regulating factor 1	11	1.6	9.6	16.2	0					-NF
D	Contig5499	Q94A7	AT4G35230	BSK1	BR-signaling kinase 1	5.2	14	7.4	21.6	0					-NF
D	Contig985	P43298	AT1G66150	TMK1	Transmembrane kinase 1	5.2	7.2	16.6	34	0					-NF
D	Contig22150	P43298	AT2G01820		Leucine-rich repeat protein kinase family protein	3	2.2	3	16.2	0					-NF
D	Contig378	Q6SCT4	AT3G51740	IMK2	Inflorescence meristem receptor-like kinase 2	3	3.4	6.6	16.2	0					-NF
D	Contig25637	Q6LNJ5	AT1G01260	JAM2	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein	5.2	16.6	7.4	24.2	2.2					-11
D	Contig19095	Q64W66	AT5G47670	NF-YB6, L1L	Nuclear factor Y, subunit B6	0	0	5.2	33.2	2.2				6.38	-15.09
D	Qs-dev_c14922	Q62175	AT2G32740	SUVH6	SUV(HAR)3.9 homolog 6	0	8.8	0.8	20	0				25	-NF
D	isotig08077	Q38827	AT5G85670	IAA9	Indole-3-acetic acid inducible 9	13.2	11	19	24.2	0					-NF

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts								Fold-change		
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4			
D	Contig81	F6H140	AT1G09710			3.4	10	4.2	24.2	2.8						-8.64
D	Contig2563	Q8XR0	AT5G46890		Homoedomain-like superfamily protein	0	1.6	1.6	14.8	0						-NF
D	Contig554	Q8J56	AT2G35940	BHLH071	Beta HLH protein 71	22.6	18.4	18.4	20.6	0						-NF
D	Contig25125	Q03460	AT5G53460	BLH1, EDA29	BEL-like homeodomain 1	73	84.6	74.6	113.8	29.8						-3.62
D	Contig4218	F4IAE9	AT1G05910	GLT1	NADH-dependent glutamate synthase 1	5.2	22.6	13.2	43	12						-3.58
D	Contig20641	Q8LP77	AT1G48480	RKL1	cell division cycle protein 48-related Receptor-like kinase 1	4.6	3	8.2	23.6	0						-NF
D	Contig20612	Q8C7T4	AT1G72210		Basic helix-loop-helix (bHLH) DNA-binding superfamily protein	8.4	27.8	21.6	32	6.8						-4.71
D	Contig3216	F6HOY7	AT5G13660		Unknown protein	20.4	5.2	5.2	33.2	2.2						-15.01
D	Contig6531	P04796	AT1G13440	GAPC-2, GAPC2	Glyceradehyde-3-phosphate dehydrogenase C2	152	118.4	124.8	203	107.8						-1.88
D	Contig26025	F6HZ75	AT3G15510	ATNAC2, ANAC056, NARS1, NAC2	NAC domain containing protein 2	1.2	1.6	1.2	19.8	0						-NF
D	Contig25489	A2YQ83	AT5G02810	PRR7, APRR7	Pseudo-response regulator 7	14.2	10.4	14	28.4	2.2						-12.91
D	Contig16086	Q8ZFE4	AT4G36980	FBW2	FBXO WITH WD-40	4.6	8.4	5.8	24.6	2						-12.3
D	Contig2406	Q85685	AT4G34610	BLH6	BEL-like homeodomain 6	29.8	28.4	25.6	36.2	7						-5.17
D	Contig25582	Q8T0J6	AT4G38980		FMN-linked oxidoreductases superfamily protein	10.4	9.6	10.6	29.6	4.6						-6.43
D	Contig18114	Q8Z555	AT1G69490	NAP, ANAC029, ATNAP	NAC-like, activated by AP3PI	8.2	8.4	3	16.2	0						-NF
D	Contig18504	Q23553	AT4G17090	CT-BMY, BAM3, BMY8	Chloroplast beta-amylase	31.4	64.8	53.8	81.4	6.6						-12.33
D	Contig9408	Q8YVW2	AT5G18670	BMY3, BAM9	Beta-amylase 3	47.4	78.2	89.8	85	30.6						-2.78
D	Contig11744	I1KPK2	AT3G07040	RPM1, RPS3	NB-ARC domain-containing disease resistance protein	6.2	6	5.8	16.6	0						-NF
D	Contig5138	A8AR53	AT5G38220		alpha/beta-Hydrolases superfamily protein	5.2	4.2	4.2	17	0						-NF
D	Os-dev_rep_5101226	Q23553	AT4G17090	CT-BMY, BAM3, BMY8	Chloroplast beta-amylase	7.8	10.4	5.6	26.4	0.6						-44
D	Os-dev_rep_74072	Q22286	AT2G39760	ATBPM3, BPM3	BTB/POZ/NATH-domains containing protein	4.2	0.6	2.2	27.4	0.8						-34.25
E	Contig17504	Q8LEB3	AT3G19130	ATRBP47B, RBP47B	RNA-binding protein 47B	23.6	27	15.8	98.6	96.4						6.24
E	Os-dev_rep_276112	Q8ZU52	AT2001140	ATFBA3, FBA3	Adolase superfamily protein	8	4.2	2.2	25.4	35						11.55
E	Contig25357	Q8SCU5	AT3G51630	WNK5, ZIK1, ATWNK5	Predictable serine/threonine-protein kinase WNK5	8.8	0.8	6.2	32	39						5.16
E	Contig17783	D7U2L4	AT1G11420	ATDUF2, DUF2	DOMAIN OF UNKNOWN FUNCTION724 2	17.6	17.2	1.6	22.4	15.2						14
E	Contig2731	Q84B46	AT4G14096		F-box/RN1-like superfamily protein	0	0	0	20.4	5						INF
E	Contig9892	Q8M886	AT3G02550	LBD41	LOB domain-containing protein 41	1.6	17.6	5.2	34.8	16.2						6.69
E	Contig19249	B8MTM7	AT3G10040		Sequence-specific DNA binding transcription factors	1.2	6	0	24.2	11.2						INF

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts							Fold-change			
						S1	S2	S3S4	S5	S8	S2/S1	S3/S4	S5/S3S4	S8/S5		
E	Contig13330	Q5ZQU2	AT5G10510	AIL6	ANTELEMENTAL-like 6	3	0.2	0	17.6	4.2				INF		
E	Contig19318	Q01593	AT3G24650	ABH1, SIS10	AP2/E3-like transcriptional factor family protein	0	0	7.2	66.8	42.2				9.28		
E	Contig9796	Q8LQZ2	AT1G75490		Integrase-type DNA-binding superfamily protein	0	0	0	34	44.8				INF		
E	Contig24728	Q43804	AT4G25140	OLEO1, OLE1	Oleosin 1	0	0	3	69.2	58.8				23.07		
E	Contig25168	Q20260	AT5G20280	ATSPS1F, SPS1F	Sucrose phosphate synthase 1F	6.8	19	6.2	50.6	24.2				8.16		
E	Contig22280	Q62627	AT1G32900	GBSS1	UDP-Glycosyltransferase superfamily protein	46.4	59.6	95.6	811.4	608.8				8.49	-1.33	
E	Contig22281	Q62627	AT1G32900	GBSS1	UDP-Glycosyltransferase superfamily protein	5	3.4	8.2	262.8	275				32.05		
E	Contig19423	P49332	AT1G17180	ATGSTU25, GSTU25	Glutathione S-transferase TAU 25	0	0	0.6	20.4	13.2				34		
E	Contig18945	Q9FN69	AT5G41315	GL3, MYC6.2	Basic helix-loop-helix (bHLH) DNA-binding superfamily protein	0	4.6	0.6	24.6	23.8				41		
E	Contig15917	Q93V43	AT4G18390	TCP2	TEOSINTE BRANCHED 1, cyclobeta and PCF transcription factor 2	1.2	9	5.8	46.6	23.4				8.03		
E	Contig17823	Q8L482	AT1G08970	HAP5C, NF-YC9	nuclear factor Y, subunit C9	6.6	6.8	3	25.6	11				8.53		
E	Contig25968	Q9ZWA1	AT1G03790	SOM	Zinc finger C-x8-C-x5-C-x3-H type family protein	3	0.8	0	21.4	35.6				INF		
E	Contig18022	B9RFV3	AT5G13180	ANAC083, VM2, NAC083	NAC domain containing protein 83	4.6	18.4	12.6	44	66				3.49		
E	Contig16572	A5BEM8	AT3G25530	GHBH1, ATGHBH1, GLYR1, GR1	Glyoxylate reductase 1	0	6	3.6	46.8	29.4				13		
E	Contig18213	B2ZA18	AT4G18650		Transcription factor-related	0	0	0.6	75	27.4				125	-2.74	
E	Contig8873	P49310	AT2G31900	XIF, ATXIF, ATMYO5	Myosin-like protein XIF	7.4	4.2	2	21.6	12				10.8		
E	Contig5949	Q6AWY3	AT2G32840	AIGRF1, GRF1	Growth-regulating factor 1	14.4	2.2	0.6	28.8	12				48		
E	Contig8039	P53492	AT5G09810	ACT7	Actin 7	16.6	65.2	45.8	60	69	3.93					
E	Contig23293	A5IVA6	AT3G27860	OLEO4, OLE3	Oleosin 4	0	0	0	71	25.8				INF	-2.75	
E	Contig20473	Q43804	AT4G25140	OLEO1, OLE1	Oleosin 1	0	0	8.2	428.2	192.4				52.22	-2.23	
E	Contig19985	IS3J00	AT1G56070	LOS1	Ribosomal protein S5E elongation factor G/III/V family protein	4.6	3.8	18.4	61	54.2				3.32		
E	Contig20284	C357F0	AT3G001570		Oleosin family protein	0	0	1.6	32	27.8				20		
E	Contig25958	Q6SZD4	AT4G29040	RPT2a	Regulatory particle AAA-ATPase 2A	5.2	14.2	10.8	47.2	35.6				4.37		
E	Contig11296	Q8LEV3	AT5G10860	CBX3	Cystathionine beta-synthase (CBS) family protein	1	9.6	9.2	53	77				5.76		
E	Contig9539	Q8ZV07	AT2G29200	APUM1, PUM1	Pumilio 1	32.4	53.8	22.2	60	48.2				2.7		
E	Contig10206	Q8LPR8	AT1G50420	SCL3, SCL-3	Scarecrow-like 3	1	3.2	0.6	22.4	15				37.33		
E	Os-dev_rep_572745	D8L029	AT3G16770	RAP2.3, ATEBP, ERF72, EBP	Ethylene-responsive element binding protein	0.4	9.4	4.6	42	18.4				9.13		
E	Contig9759	Q6AVY3	AT1G43800	FTM1, SAO6	Plant steroyl-acyl-carrier-protein desaturase family protein	0	6.6	5.2	50.8	16.2				9.77	-3.14	

Supplementary Table S5. (Continued)

Cluster	Transcript name	UniProt	At Locus	At homologous	At homologous function	Normalized counts								Fold-change		
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4			
E	Contig19653	P2337	AT2G43710	SSB, FAB2	Plant sterol/acyl-carrier-protein desaturase family protein	15.8	24.6	10.4	89.2	40.6			8.69	-2.2		
E	Contig9756	O6407	AT2G32280	BMV5, BAMB	Beta-amylase 6	3.2	12.2	11	45.4	72.2			4.13			
F	Os-dev_c20461	BDF17	AT1G26850		Sadenosyl-L-methionine-dependent methyltransferases superfamily protein	15	10.4	11	6	32				5.33		
F	Os-dev_c43484	BTKR5	AT1G18080	ATARCA, RACK1A_AT, RACK1A	TransducinWD40 repeat-like superfamily protein	5.6	11.6	12.2	7	37.8				5.4		
F	Os-dev_rep_76930	Q4284	AT4G15900	PRL1	Pleiotropic regulatory locus 1	6	5.2	3.4	3.8	26.8				7.05		
F	Os-dev_rep_74812	Q84019	AT5G24830	ATCOL4, COL4	CONSTANS-like 4	17.2	14	0.2	0	16.6				INF		
F	Contig25545	BRRB49	AT3G48000	ALDH2B4, ALDH2, ALDH2A	Aldehyde dehydrogenase 2B4	35.6	32.8	20.6	20.4	65.2				3.2		
F	Contig3914	Q85796	AT1G74920	ALDH10A8	Aldehyde dehydrogenase 10A8	1.4	0.2	0	11	42.8				3.89		
F	Contig10071	Q8ZIK5	AT3G13960	AGR5, GRF5	Growth-regulating factor 5	0	0	0	2	21				10.5		
F	Contig25804	Q8FY75	AT5G09400	KUP7	K ⁺ uptake permease 7	1.2	0.8	3	3.4	35.6				10.47		
F	Contig10067	A5BSW3				0	0	0	0	15				INF		
F	Contig12127	B3H615	AT1G54390	ING2	PHD finger protein-related	0	0.8	0	0	17				INF		
F	Contig25194	POCE10	AT2G01130		DEAD(H)-box RNA helicase family protein	6.8	45.6	15	27.8	49.2	6.71					
F	Contig8072	B9HW74	AT1G71000		Chaperone DnaJ-domain superfamily protein	1.6	1.6	5.2	5.2	34.2				6.58		
F	Contig1489	Q84XV2	AT2G34900	GTE01, IME1, GTE1	Transcription factor GTE6	0	2.2	0	6.6	38.2				5.79		
F	Contig8087	F6HHN8	AT5G58470	TAF15b	TBP-associated factor 15B	3.4	3.4	0	2.2	26.4				12		
F	Os-dev_c8383	P23766	AT4G36130		Ribosomal protein L2 family	0.4	2.2	0	0.8	20.6				25.75		
F	Contig5158	Q8ZVM9	AT1G54610	OKL9	Protein kinase superfamily protein	3.2	7	4.2	3.6	30.2				8.39		
F	Contig6311	P52408	AT3G57240	BG3	Beta-1,3-glucanase 3	18.4	0	0.8	5.2	54.4	-INF			10.46		
F	Contig26823	Q8SAG3	AT2G31200	ADF6, A1ADF6	Actin depolymerizing factor 6	12.6	17.8	8.4	5.8	35.6				6.14		
F	Os-dev_c29889	Q39030	AT3G08720	ATPK19, ATPSK2, SKK2, ATPK2	Selenothreosine protein kinase 2	0	0	0	0.6	20.6				34.33		
F	Contig10510	Q8STJ6-2	AT4G25390		Protein kinase superfamily protein	1	0.8	0	0	29.2				INF		
F	Os-dev_c15659	Q8FF56	AT5G06040	ATPDIL1.4, PD12, ATPD12, PDIL1.4	PD1-like 1.4	7.4	35.2	36	9.2	39.6	4.76			-3.91	4.3	
F	Contig22205	B9MSL8	AT5G23810	AA1P7	Amino acid permease 7	0	0.2	0	5.2	62.4				12		
F	Contig1878	F4HW02	AT1G08200		Transmembrane amino acid transporter family protein	4.6	0.2	0.6	2	35.4				17.7		
F	Contig19426	P10978	AT3G21000		Gag-Pol-related retrotransposon family protein	0	0	0	5.6	33.2				5.93		
F	Contig9766	P10978	AT4G23160	CRK8	Cysteine-rich RLK (RECEPTOR-like protein kinase) 8	0	0	0	5.2	46.4				8.92		
F	Contig23375	P4Z731	AT2G44710		RNA-binding (RRM/RBD/RNP motifs) family protein	26	29.8	1.6	8.8	27.8				-18.63		

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts						Fold-change		
						S1	S2	S3S4	S5	S8	S2/S1	S3S4/S2	S5/S3S4	S8/S5
F	Contig19997	Q6S7T8	AT1G47710	SERPIN1	Serine protease inhibitor (SERPIN) family protein	42	8	3.8	22.6	48.8	-5.25			
F	Contig20497	P46297	AT5G02960		Ribosomal protein S12S23 family protein	0	8	9.4	3	27.8				9.27
F	Contig14504	Q6SZL8	AT1G76320	FRS4	FAR1-related sequence 4	0	0	0	0	23.2				INF
F	Contig17981	Q03666	AT1G17180	ATGSTU25, GSTU25	Glutathione S-transferase TAU 25	0	0	0	29.8	74.2				2.49
F	Contig24085	Q9ZRW8	AT1G78380	ATGSTU19, GST8, GSTU19	Glutathione S-transferase TAU 19	5	2.2	6	8.2	44.4				5.41
F	Contig611	O49934	AT1G09570	PHYA, PHY2, FRE1, HY8	Phytochrome A	16.4	21	5.2	32.4	91.2				2.81
F	Contig239	Q6SW80	AT4G38870	BLH2, SAW1	BEL1-like homeodomain 2	3	8.8	3.8	11.6	40.6				3.5
F	Contig6882	Q9LME6-2	AT1G32210	MBD8, ATMED8	Methyl-CPG-binding domain 8	13	7.4	7.4	4.6	26.4				5.74
F	Contig17728	B9HSP8	AT3G51880	HMG81, NFD1	High mobility group B1	17.6	4.6	3.8	7.2	62				8.61
F	Os-dev_rep_c75011	O49595	AT3G51880	HMG81, NFD1	High mobility group B1	7.8	2.2	1.2	3.2	29				9.06
F	Contig17689	Q9LIJ9	AT3G27950		GDSL-like Lipase/Acylhydrolase superfamily protein	0	2.2	3.6	3	27.4				9.13
F	Contig18524	Q94CH6	AT1G75900		GDSL-like Lipase/Acylhydrolase superfamily protein	8.4	8.8	17.2	0	21.4				INF
F	Os-dev_c9570	Q7XA42	AT3G14470		NB-ARC domain-containing disease resistance protein	0.4	0	0	0.8	18.4				23
F	Os-dev_c10248	Q9LRR5				0	5.8	0.8	0	24.4				INF
F	Os-dev_rep_c74668	Q9FVE6	AT3G44750	HD2A, ATHD2A, HDA3, HDT1	Histone deacetylase 3	4.2	8.2	5.6	0.8	18.8				23.5
F	Contig17534	O81122	AT1G66940	ETR1, EIN1, ETR, AIE1R1	Signal transduction histidine kinase, hybrid-type, ethylene sensor	23.6	3	31.6	21	40.2	-7.87	10.53		
F	Contig16927	Q9FJ88	AT5G53860	AIGRF7, GRF7	Growth-regulating factor 7	1.2	0.8	0	10.8	47				4.35
F	Contig18038	E0CU2	AT5G53140		Protein phosphatase 2C family protein	4.6	14.8	11.6	16	50.8				3.175
F	Contig20038	O81970	AT3G48270	CYP71A26	Cytochrome P450, family 71, subfamily A, polypeptide 26	0	0	0	0.4	40.6				101.5
F	Contig16776	B9ID52	AT5G53420		CCT motif family protein	4.6	0.8	0	16.6	87.2				5.25
F	Contig18217	Q8ZWA1	AT1G03790	SOM	Zinc finger C-x8-C-x5-C-x3-H type family protein	0	0	1.6	9.4	62.4				6.64
F	Contig18229	D7SXU9	AT2G14910		Unknown protein	10.4	14.8	16	25.6	80				3.13
F	Contig22282	11KM96	AT3G21000		Gag-Pol-related retrotransposon family protein	0	0	0	7.4	31.6				4.27
F	Contig1894	P92948	AT1G09770	ATDC5, CDC5, ATMYBCDC5	Cell division cycle 5	23.8	27	21.6	16.2	63.4				3.91
F	Contig19233	Q9CR10	AT1G33420		RINGFYVE/PHD zinc finger superfamily protein	22	33.8	19.4	17.6	48.8				2.77
F	Contig2412	B6SZN4	AT2G12646		PLATZ transcription factor family protein	0	0.2	0	0.4	25.8				64.5
F	Contig6574	P46604	AT4G37790	HAT22	Homeobox-leucine zipper protein family	3.2	4.6	4.2	0.6	26.4				44
F	Os-dev_rep_c74268	B9T445	AT5G16880		RINGFYVE/PHD zinc finger superfamily protein	4.2	0.6	3.4	1.6	20.6				12.88

Supplementary Table S5. (Continued)

Cluster	Transcript name	Uniprot	At Locus	At homologous	At homologous function	Normalized counts							Fold-change	
						S1	S2	S3S4	S5	S6	S2/S1	S3S4/S2	S5/S3S4	S6/S5
F	Os-dev_jep_c77899	Q8LQZ7	AT1G75540	STH2	Salt tolerance homolog2	0.4	0	0.2	0	26.8				INF
F	Os-dev_jep_c86581	P29675	AT1G10200	WLM1	GATA type zinc finger transcription factor family protein	0.6	0	0.2	0.2	18.8				94
F	Os-dev_jep_c103605	AS1VA6	AT3G27860	OLEO4, OLE3	Oleosin 4	0.4	0	0	9.2	39.6				4.3
F	Os-dev_c70574	P02857	AT5G40420	OLEO2, OLE2	Oleosin 2	0.2	0	0	2.2	49.8				22.64
F	IsoligM4348	Q8C688	AT1G50680		AP2B3 transcription factor family protein	3.4	0.8	3	2.2	20.4				9.27
F	Os-dev_jep_c92297	Q6SZD4	AT4G29040	RPT2a	Regulatory particle AAA-ATPase 2A	11.6	5.8	5.6	2.2	20.6				9.36
F	Contig5432	Q6SS94	AT3G01610	CD48C, emb1354	Cell division cycle 48C	3.2	26.4	14.2	6.8	26.2	8.25			
F	Os-dev_jep_c73053	Q6FMT4	AT5G14170	CHC1	SWIMMDM2 domain superfamily protein	0.4	2.2	1.2	0	20.6				INF
F	Contig11689	Q8LXT3	AT3G58680	MBF1B, ATMBF1B	Multiprotein bridging factor 1B	28.8	10.8	15.8	10.6	38.2				3.6
F	Contig8377	Q70I13	AT5G64750	Abr-01	Integrase-type DNA-binding superfamily protein	20.6	11.6	9.8	3	22.4				7.47
F	Contig5696	Q8FT50	AT5G07900		Mitochondrial transcription termination factor family protein	5.2	0.2	3	0.4	18.2				45.5
F	Contig3793	B9HBM8	AT3G12560	TRFL9, ATTRBP2	TRF-like 9	8.8	12.6	4.2	0.4	20				50
F	Os-dev_c6836	Q6B0Q4-4	AT5G60410	ATSIZ1, SZ1	DNA-binding protein with MIZ/SP-RING zinc finger, PHD-finger and SAP domain	5.8	29.4	17.8	7	28.8	5.07			
F	Contig7766	Q64836	AT4G24770	RBP31, ATRBP31, CP31, ATRBP33	31-kDa RNA binding protein	1.6	7	2	6.8	32				4.71
F	Contig7541	Q03878	AT3G56860	UBA2A	UBP1-associated protein 2A	1.6	16.8	13.2	4.6	37.8				8.22
F	Isolig07420	Q64407	AT4G15210	ATBETA-AMY, AT-BETA-AMY, RAM1, BMY1, BAM5	Beta-amylase 5	5.4	0	5.4	2.2	22.6				10.27
F	Contig18960	Q23310	AT4G14540	NF-YB3	Nuclear factor Y, subunit B3	1.2	2.2	0	1.2	51				42.5
F	Contig26279	D6L029	AT3G16770	RAP2.3, ATEBP, ERF72, EBP	Ethylene-responsive element binding protein	8.8	6.6	0	15.6	35.6			INF	
F	Contig22769	Q8FJA2	AT5G35550	TT2, ATMYS123, MYB123, ATT72	Duplicated homeodomain-like superfamily protein	1.2	0	0.6	0.4	27.8				69.5
F	Os-dev_jep_c87001	F6H781	AT4G00960		Transcription factor junonil (jnfC) domain-containing protein	0.6	0	0.2	0.2	16.8				84
F	Contig22387	Q6QUQ2	AT1G60420	ATNRX1, NRX1	DC1 domain-containing protein	6.8	4.6	14	5.6	29.8				5.32
F	Os-dev_c39887	Q6QUQ2	AT1G60420	ATNRX1, NRX1	DC1 domain-containing protein	5.6	10	8.8	3	24.6				8.2
F	Contig22386	Q6QUQ2	AT1G60420	ATNRX1, NRX1	DC1 domain-containing protein	3	4.6	0	0	19.8				INF
F	Contig5275	Q6Z5Y2	AT1G68370	ARG1	Chaperone DnaJ-domain superfamily protein	12.8	24.2	8.8	9	35.6				3.96
F	Contig25391	D7UAN4	AT1G55690		Sec14p-like phosphatidylinositol transfer family protein	10.4	17.8	8.2	11.6	49.2				4.24

INF: expression ratios with zero counts in the numerator or denominator were annotated as minus infinitive (-INF) and infinitive (INF), respectively.

Supplementary Files

Additional files are only in the digital format due to the size of the excel tables. Supplementary Tables S6 and S7 have more detailed information about DE transcripts related to water and annotated as transcription factors and are also only presented in the digital format. Legends are described below:

Supplementary File S1. *Quercus suber* transcript classification and predicted function. Annotation and classification of the assembled transcripts using Full Lengther Next scripts and functional annotation using the Uniprot database.

Supplementary File S2. Functional annotation of *Quercus suber* transcripts. Annotation was based on the homologous in NCBI and *Arabidopsis* (TAIR) databases using Blastx with an E-value of $1e^{-10}$ and results were submitted in Blast2GO to annotate gene ontology (GO) terms.

Supplementary File S3. Classification of Orthologous Groups (COG). Analysis of the COGs present in data retrieved from *Q. suber*, CODB (Cork Oak Database), *Q. petraea* and *Q. robur* and association to each *Arabidopsis* gene.

Supplementary File S4. Number of cork oak transcripts in KEGG level 2 and 3 pathways. Expression of each pathway was quantified as the sum of the counts (mapped reads) of all the transcripts belonging to it.

Supplementary File S5. Identification of the genes differentially expressed during acorn development by Z-test. Results of the clustering, fold-change and functional annotation (NCBI, TAIR, UNIPROT) of the differentially expressed genes (DEGs) in consecutive stages of acorn development.

Supplementary File S6. Enrichment analysis of the differentially expressed genes (DEGs). Comparison of the DEGs with the complete transcriptome evidencing the gene ontology (GO) terms over-represented.

Supplementary Table S6. Differentially expressed genes related to response to water, water deprivation or water transport. Cluster annotation, transcript name, NCBI function, Biological and Molecular functions, Uniprot locus, association with *Arabidopsis* homologs, normalized counts in each acorn developmental stage and fold-change between consecutive stages.

Supplementary Table S7. Differentially expressed genes annotated as transcription factors. Cluster annotation, transcript name, NCBI function, Biological and Molecular functions, Uniprot locus, association with *Arabidopsis* homologs, normalized counts in each acorn developmental stage and fold-change between consecutive stages.

Chapter VI

Concluding Remarks and Future Perspectives

In this work it was shown for the first time that a *SHR*-like gene is involved in the regulation of the lateral meristem responsible for cork formation. In fact, this represents one of the pioneer studies focusing on the molecular regulation of the phellogen activity. Of the three *SHR*-like genes present in the *Populus* genome (*PtSHR1*, *PtSHR2A* and *PtSHR2B*), the homolog of the *Arabidopsis SHR* (*AtSHR*), *PtSHR1*, had been associated to the functioning of the vascular cambium in a previous report (Wang *et al.*, 2011). Here, we propose that *PtSHR2B* is also involved in regulation of secondary growth, but unlike *PtSHR1*, its function is associated to the phellogen. Distinct localizations were found for these two *Populus SHR*-like genes. While our studies confirmed that *PtSHR1* is mainly expressed at the cambial region of plant stems, *PtSHR2B* promoter appears to drive expression in the primary xylem and in the outer cell layers of *in vitro* plant stems becoming strongly localized in the phellogen of mature plants. Since in *Arabidopsis* *SHR* acts in a non-cell autonomous manner, it would be interesting to investigate the localization of the *SHR2B* protein in the periderm of *Populus* plants and check if it is able to move to adjacent cell layers as in *Arabidopsis*. Different functions for these genes are also expected in the root since *GUS* expression under the *PtSHR2B* promoter was found in the root tip, while *PtSHR1* promoter drives expression in the root stele. Ectopic expression of *PtSHR2B* under the control of the 35S constitutive promoter led to an overall reduction of growth in the transgenic plants. A reduction in the relative proportion of the phellem layer in transgenic plants was also observed indicating that an optimal level of *PtSHR2B* may be required for normal phellogen activity. It is plausible that, like *AtSHR* and *PtSHR1*, *PtSHR2B* acts in a dose-dependent manner. However, measurement of protein levels in each tissue is needed to clarify this issue. In the future it will be also interesting to analyse the phellogen activity in transgenic hybrid aspen or even in transgenic *Arabidopsis*. This will allow understanding if increased levels of *PtSHR2B* are responsible for

a reduction in cell division, cell expansion or cell differentiation and study its effect on cambial growth. Transgenic plants were affected in cytokinin metabolism and showed an overall decrease in the cytokinin response. This probably contributed, at least partially, to the anatomical changes detected in the transgenic plants. The relation between *PtSHR2B* and cytokinin responses is not currently clear but measuring the levels of cytokinins in the transgenic plants, as well as analysing the effects of overexpressing or silencing cytokinin oxidase genes or primary response genes should help elucidate this matter. Furthermore, a cytokinin responsiveness assay to observe how stem regeneration occurs when subjected to different cytokinin concentrations, may also be an interesting test to perform. Since *Arabidopsis* *SHR* is a key growth regulator controlling a large transcription factor network as well as hormonal and signalling pathways, transcriptional profiles of genes known to be directly or indirectly regulated by *SHR* in the *Arabidopsis* root can also be analysed in the *Populus* transgenic plants. This would allow verifying if *PtSHR2B* triggers the expression of the same genes as in the *Arabidopsis* root, by using for instance available resources such as the GeneChip Poplar Genome Array (Affymetrix).

Like in *Populus*, more than one *SHR*-like sequence was found in cork oak, *QsSHR1* and *QsSHR2*, and in the related species *Quercus ilex* (holm oak), *QiSHR1* and *QiSHR2*. Although any of the cork oak *SHR* genes was found to be tissue specific they showed an expression profile similar to the *Populus* *SHR* genes, with *QsSHR1* being expressed mainly in radicles and *QsSHR2* in hypocotyls. This fact may suggest that they play similar roles to *PtSHR1* and *PtSHR2B*. In fact, the obtained results showed that each gene is transcribed at least in one of the two lateral meristems, as observed in *Populus*. *QsSHR1* is specifically expressed in the vascular cambium and *QsSHR2* transcripts were found in the phellogen and, at a lower level, in the vascular cambium. A relevant question was whether the cork oak genes could complement the *Arabidopsis* *shr2* mutant phenotype.

A full complementation was not observed when using one or the other gene coding sequences since the transgenic plants did not exhibit a total recovery of the wild-type phenotype. We believe this is possibly a consequence of genome duplication events after divergence of the *Arabidopsis* and the cork oak lineages and a phenomenon of sub-functionalization or neo-functionalization may have occurred. It will be interesting to analyse the generated *Arabidopsis* plants grown under conditions that promote secondary growth and to generate plants in which the same expression cassette is fused to a reporter gene to find where *QsSHR2* is expressed.

The present work also led to the identification of the most stable reference genes suitable for the normalization of reverse transcription quantitative PCR results in different cork oak tissues. Different reference genes showed some variation depending on the tissues under analysis, on the developmental stage and on the time of collection. This study has already been a starting point for gene expression studies in cork oak (Almeida *et al.*, 2013a,b; Ramos *et al.*, 2013) and it can also be useful for other oaks. In fact, this has already been the case for holm oak, where the most stable genes were precisely the ones selected for the cork oak tissues.

The fact that *QsSHR2* is transcribed in the phellogen meristematic layer responsible for the formation of the periderm, suggests that it is involved in the regulation of phellogen activity in cork oak. Additionally, both cork oak *SHR* genes are expressed in the periderm of branches of increasing age but exhibit different expression patterns. Phellogen activity varies along the year showing a reduced activity in winter months like January. Because *QsSHR2* showed higher expression levels in this period we suggest that it may be a negative regulator of the phellogen meristematic activity similarly to what has been proposed for *Populus SHR*-like genes. A major role for *QsSHR2* during the formation of the cork oak phellogen rather than in phellogen maintenance seems likely as suggested by its higher expression during the initial years of periderm formation. To

clarify this proposed mode of action it would be interesting to isolate and analyse the phellogen cells by techniques such as laser capture micro dissection to verify the expression profiles during phellogen formation, maintenance and activity.

The involvement of *PtSHR2B* in the proper functioning of phellogen, and therefore in phellem and periderm formation and regulation might be a consequence of functional diversification after divergence of the *Populus* and *Arabidopsis* lineages. Similar lines of evidence were presented for *SHR*-like genes in cork oak, suggesting that these genes are involved in the regulation of secondary growth in this peculiar long living tree. Our findings can contribute to elucidate basic features of the molecular mechanisms involved in the regulation of lateral meristems. In agreement with previous reports suggesting overlapping regulatory mechanisms between primary meristems and the vascular cambium, also the phellogen may share such mechanisms of regulation.

In this thesis it is also presented for the first time a transcriptomic analysis of cork oak developing fruits that may prove relevant to understand basic aspects of the reproductive biology of this tree species. A staging system to categorize the cork oak acorns into different developmental stages was established based on several morphological aspects. A dynamic view of gene expression using next generation sequencing technologies was obtained as well as a set of genes differentially regulated between different fruit developmental stages. We then focused on aspects of fruit development regulated by genes involved in carbohydrate metabolism, in response to water, including water transport and water deprivation, and genes annotated as transcription factors. Carbohydrate metabolism was the most represented level-2 pathway and our analysis suggested that alterations in the availability of soluble sugars, which are known to be crucial in the regulation of plant development and growth, also play important roles during acorn development. Differentially expressed genes related to

response to water were mainly expressed in an early (S2) and last stage of development (S8), possibly reflecting the importance of processes related to regulation of water content in the early stages of the acorn development that may be critical to its further development as well as the relevance of the dehydration process characteristic of late development. Transcription factors were almost equally distributed throughout development representing nearly a quarter of the differentially expressed genes, highlighting the major role of these regulators during fruit development. This data collection will provide an extremely useful basis for future studies on cork oak biology. The functional analysis of differentially expressed genes will be crucial to obtain further knowledge on the reproductive biology of the cork oak tree and, hopefully, to assist in the implementation of strategies promoting species regeneration.

References

- Almeida T, Menéndez E, Capote T, Ribeiro T, Santos C, Gonçalves S.** 2013a. Molecular characterization of *Quercus suber* MYB1, a transcription factor up-regulated in cork tissues. *Journal of Plant Physiology* **170**, 172–178.
- Almeida T, Pinto G, Correia B, Santos C, Gonçalves S.** 2013b. QsMYB1 expression is modulated in response to heat and drought stresses and during plant recovery in *Quercus suber*. *Plant Physiology and Biochemistry* **73**, 274–281.
- Ramos M, Rocheta M, Carvalho L, Inácio V, Graça J, Morais-Cecilio L.** 2013. Expression of DNA methyltransferases is involved in *Quercus suber* cork quality. *Tree Genetics & Genomes* **9**, 1481–1492.
- Wang J, Andersson-Gunnerås S, Gaboreanu I, Hertzberg M, Tucker MR, Zheng B, Leśniewska J, Mellerowicz EJ, Laux T, Sandberg G, et al.** 2011. Reduced expression of the *SHORT-ROOT* gene increases the rates of growth and development in hybrid poplar and Arabidopsis. *PLoS ONE* **6**, e28878.

This work was supported by **Fundação para a Ciência e Tecnologia**, with a PhD grant (Ref. **SFRH/BD/44474/2008**) awarded to Andreia Miguel and the research projects PEst-OE/EQB/LA0004/2011, PTDC/AGR-GPL/098369/2008 and SOBREIRO/0029/2009.

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

ITQB-UNL | Av. da República, 2780-157 Oeiras, Portugal
Tel (+351) 214 469 100 | Fax (+351) 214 411 277

www.itqb.unl.pt

Oeiras, June, 2014

Molecular regulation of secondary growth: searching for *SHORT-ROOT* function in cork formation

Andreia Miguel

