INSTITUTO SUPERIOR DE AGRONOMIA
Universidade de Lisboa

QUERCUS SUBER MYB1 TRANSCRIPTION FACTOR - AN *IN VIVO* APPROACH FOR THE ELUCIDATION OF ITS REGULATORY NETWORKS

TIAGO COSTA CAPOTE

SUPERVISOR: Doctor Maria Leonor Mota Morais Cecílio
CO-SUPERVISOR: Doctor Sónia Cláudia Morgado Gonçalves

THESIS PRESENTED TO OBTAIN THE PH.D DEGREE IN BIOLOGY

2018

ULISBOA | UNIVERSIDADE DE LISBOA

INSTITUTO SUPERIOR DE AGRONOMIA
Universidade de Lisboa
Hinc Patriam Sustinet

QUERCUS SUBER MYB1 TRANSCRIPTION FACTOR - AN *IN VIVO* APPROACH FOR THE ELUCIDATION OF ITS REGULATORY NETWORKS

TIAGO COSTA CAPOTE

ORIENTADORA: Doutora Maria Leonor Mota Morais Cecílio
COORIENTADORA: Doutora Sónia Cláudia Morgado Gonçalves

TESE ELABORADA PARA OBTENÇÃO DO GRAU DE DOUTOR EM BIOLOGIA

JÚRI

Presidente: Doutora Maria Margarida Branco de Brito Tavares Tomé, Professora Catedrática, Instituto Superior de Agronomia, Universidade de Lisboa.

Vogais: Doutor Ricardo Javier Ordás Fernández, Catedrático de Universidad, Universidade de Oviedo, Espanha;

Doutora Sara Barros Queiroz Amâncio, Professora Associada com agregação Jubilada, Instituto Superior de Agronomia, Universidade de Lisboa;

Doutor Jörg Becker, Investigador Principal, Instituto Gulbenkian de Ciência, Fundação Calouste Gulbenkian;

Doutor Nelson José Madeira Saibo, Investigador Principal, Instituto de Tecnologia Química e Biológica António Xavier, Universidade Nova de Lisboa;

Doutora Maria Leonor Mota Morais Cecílio, Professora Auxiliar, Instituto Superior de Agronomia, Universidade de Lisboa.

2018

U LISBOA | UNIVERSIDADE DE LISBOA

Aos meus pais;

À minha esposa;

A todos os meus amigos;

Ao Alentejo.

Tiago Costa Capote

CEBAL – Centro de Biotecnologia Agrícola e Agro-Alimentar do Alentejo

Rua Pedro Soares, s.n – Campus IPBeja / ESAB

Apartado 6158

7801-908 Beja | Portugal

e-mail: tiago.capote@cebal.pt / capote.tiago@gmail.com

Work performed at:

Rua Pedro Soares s/n

7801-908 Beja

Portugal

Tapada da Ajuda,

1349-017 Lisboa

Portugal

Calle San Francisco

33003 Oviedo

Asturias, Spain

Av. Da República

2780-157 Oeiras

Portugal

Supervisor:

Professor Maria Leonor Mota Morais Cecílio, Doctor in Genetics

LEAF, Instituto Superior de Agronomia, Universidade de Lisboa

Co-supervisor

Doctor Sónia Cláudia Morgado Gonçalves, Doctor in Biology

AgroGenomics Group, CEBAL

“Science, my boy, is composed of errors, but errors that it is right to make, for they lead step by step towards the truth.”

— Jules Verne, Journey to the Centre of the Earth (1872)

“Everything must be taken into account. If the fact will not fit the theory - let the theory go.”

— Agatha Christie, The Mysterious Affair at Styles (1920)

The author declares to had active participation in the following papers during the practical work and writing of the present thesis:

Capote T, Gonçalves S, Morais-Cecílio L. Cork Oak somatic embryos, a model for functional studies: induction, selective agent characterization and cryopreservation. *Plant Cell, Tissue and Organ Culture*. *Submitted*

Capote T, Usié A, Barbosa P, Ramos AM, Morais-Cecílio L, Gonçalves S. Transcriptome dynamics of cork oak (*Quercus suber*) somatic embryogenesis reveals active gene players in transcription regulation and hormone homeostasis of embryo development*. Tree Genetics & Genomes*. *Submitted*

Capote T, Barbosa P, Usié A, Ramos AM, Inácio V, Ordás R, Gonçalves S, Morais-Cecílio L. ChIP-Seq reveals that QsMYB1 directly targets genes involved in lignin and suberin biosynthesis pathways in cork oak (*Quercus suber*). *BMC Plant Biology. Submitted*

ABBREVIATIONS

% SEI	Percentage of leaves producing somatic embryogenesis
2,4-D	2,4-Dichlorophenoxyacetic acid
2-MHG	2-mono(10,16-dihydroxyhexadecanoyl) glycerol
3xFLAG	triple FLAG epitope
4CL5	4-Coumarate:CoA ligase 5
A	Adenine
ABA	Abscisic acid
ABC	ATP-binding cassette
ABCC5	ATP-binding cassette transporter C family member 5
ABCG(s)	ATP-binding cassette transporter(s) of the G-subfamily
ABH1	Nuclear cap-binding protein subunit 1
ACL5	Thermospermine synthase acaulis 5
ACP	Acyl-carrier protein
ACX	Acyl-CoA oxidase
ACT	Actin
ADR	NADPH-dependent aldehyde reductase
AGL15	AGAMOUS-like 15
AHC(s)	Alkyl hydroxycinnamate(s)
AIL	Aintegumenta-like
AIL5	Aintegumenta-like 5
AIL6	Aintegumenta-like 6
ANT	Aintegumenta
AP2/ERF	APETALA2/Ethylene Responsive element binding factors
ARF(s)	Auxin response factors
ARIA	Arm repeat protein interacting with ABF2
ASFT	Aliphatic suberin feruloyl transferase
ATHB8	Arabidopsis thaliana homeobox 8
ATP	Adenosine triphosphate

AUX	Auxin
Asp	Aspargine
BAHD	refers to the BAHD acyltransferase family, named according to the first letter of each of the first four biochemically characterized enzymes of this family (BEAT, AHCT, HCBT, and DAT)
BAP	6-benzilaminopurine
BBM	BABY BOOM
BCV	Biological coefficient of variation
BDG	BODYGUARD
BED	referes to BED finger proteins, named after the Drosophila proteins BEAF and DREF discovery
bHLH	Basic helix-loop-helix
bp	Base pairs
BR	Brassinosteroid
BSA	Bovine serum albumin
BXL6	Beta-d-xylose 6
bZIP	Basic leucine zipper
C2H2	Cys(2)His(2) type zinc finger domain
CACs	Clathrin adaptor complexes
CAD	Cinnamyl alcohol dehydrogenase
CCA1	Circadian clock associated protein 1
CCHC	CysCysHisCys type zinc finger domain
CD1/GDSL1	Cutin synthase
cDNA	Complementary DNA
CESA3	Cellulose synthase a catalytic subunit 3
CHD3	Chromodomain/helicase/DNA binding domain
CHLH	Magnesium-chelatase subunit chlh
CHMP1B	Escrt-related protein CHMP1B
CK	Cytokinin
CPC	CAPRICE
CR(s)	Chromatin regulator(s)

CRF3	Cytokinin response factor 3
CUL1	Cullin 1
CUS	Cutin synthase
CYP	Cytochrome P450 enzymes
CYP86	Cytochrome P450 enzymes subfamily 86
CYP86A1	Cytochrome P450 enzymes subfamily 86A1
CYP86A8	Cytochrome P450 enzymes subfamily 86A8
ChIP	Chromatin Immunoprecipitation
ChIP-CHIP	Chromatin Immunoprecipitation followed by DNA microarray hybridization
ChIP-Seq	Chromatin Immunoprecipitation followed by Next Generation Sequencing
ChIP-qPCR	Chromatin Immunoprecipitation followed by quantitative real-time PCR
CoA	Coenzyme A
DAD1	Defective in anther dehiscence 1
DCF	Defective in cutin ferulate
DCR	Defective in cuticular ridges
DE	Differential expression
DEG(s)	Differentially expressed gene(s)
DFR-like1	Dihydroflavonol 4-reducatse-like 1
DGAT	Diacylglycerol O-acyltransferase
DNA	Deoxyribonucleic acid
Dof	DNA binding with one zinc finger
EIL	Ethylene insensitive like
EMK	EMBRYOMAKER
EP	Earlier phase
EST(s)	Expressed sequenced tag(s)
ET	Ethylene
FA(s)	Fatty acid(s)
FABI	Enoyl-[acyl-carrier-protein] reductase (NADH)
FACT	Fatty alcohol:caffeoyl-CoA caffeoyl transferase

FAE	Fatty acid elongation
FARs	Fatty acyl reductases
FAS	Fatty acid synthase
FDR	False discovery rate
FHT	Fatty ω-hydroxy acid/fatty alcohol hydroxycinnamoyl transferase
FUSCA3	B3 domain-containing transcription factor FUS3
G3P	Glycerol-3-phospate
GA	Gibberellic acid
GA2ox1	Gibberellin 2-beta-dioxygenase 1
GA3ox2	Gibberellin 3-beta-dioxygenase 2
GA3ox6	Gibberellin 3-beta-dioxygenase 6
GARP	Golgi-associated retrograde protein
GDSL	Gly-Asp-Ser-Leu domain family of esterases/acylhydrolases lipases
GN	Guanine-nucleotide exchange factor gnom
GO	Gene ontology
GPAT(s)	Glycerol 3-phosphate acyltransferase(s)
GPAT5	Glycerol 3-phosphate acyltransferase5
GPAT7	Glycerol 3-phosphate acyltransferase7
Gly	Glycine
H3K27me3	Histone H3 tri-methylated on lysine 27
H3K37me3	Histone H3 tri-methylated on lysine 37
HAP3	Heme-activated proteins
HD-ZIP	Homoeodomain-leucine zipper
HHT	Hydroxycinnamoyl-CoA: ω-Hydroxyacid *O*-Hydroxycinnamoyl Transferase
HK	Histidine kinase
HLH	Helix-turn-helix
IAA	Indole-3-acetic acid
IAA30	Indole-3-acetic acid inducible 30
IBA	Indole-3-butyric acid

IDR	Irreproducible discovery rate
IP	Immunoprecipitated
IgG	Immunoglobulin G type
JA	Jasmonic acid
KCS	β-ketoacyl-CoA synthase
KCS17	3-ketoacyl-CoA synthase17
KEGG	Kyoto Encyclopedia of Genes and Genomes
L1L	LEC1-like
LACS	Long-chain acyl-CoA synthetase
LEA	Late embryogenesis abundant protein
LEC	Leafy cotyledon
LFY	Floral meristem identity LEAFY
LHY	Late elongated hypocotyl
LIP	Triacylglycerol acylhydrolase
LP	Late phase
LTP(s)	Lipid transfer protein(s)
LTPG5	Glycosylphosphatidylinositol-anchored lipid protein transfer 5
Leu	Leucine
Lys	Lysine
MADS	or MADS-box, conserved DNA-binding motif named by the originally identified members: MCM1, AG, DEFA and SRF
MAPQ	Mapping quality
MEE40	Maternal effect embryo arrest 40
MEME	Multiple Expectation Maximization for Motif Elicitation
MP	Monopteros
mRNA	Messenger ribonucleic acid
MS	Murashige and Skoog
MSSH	refers to Murashige and Skoog micronutrients plus Schenk and Hildebrandt macronutrients
MSSH1	Primary induction medium
MSSH2	Secondary induction medium

MSWH	refers to Murashige and Skoog medium without hormones
MYB	Myeloblastosis type transcription factor
MYB0/GL1	Myeloblastosis type transcription factor 0
MYB1	Myeloblastosis type transcription factor 1
MYB107	Myeloblastosis type transcription factor 107
MYB115	Myeloblastosis type transcription factor 115
MYB118	Myeloblastosis type transcription factor 118
MYB1::3xFLAG	MYB1 fused with the triple FLAG epitope
MYB23	Myeloblastosis type transcription factor 23
MYB37/RAX1	Myeloblastosis type transcription factor 37
MYB38/RAX2	Myeloblastosis type transcription factor 38
MYB41	Myeloblastosis type transcription factor 41
MYB46	Myeloblastosis type transcription factor 46
MYB58	Myeloblastosis type transcription factor 58
MYB63	Myeloblastosis type transcription factor 63
MYB66	Myeloblastosis type transcription factor 66
MYB68	Myeloblastosis type transcription factor68
MYB85	Myeloblastosis type transcription factor 85
MYB9	Myeloblastosis type transcription factor 9
MYBL2	Myeloblastosis transcription factor like 2
NAA	1-Naphthaleneacetic acid
NAC	specific plant transcription factor family, named by the first studied members: NAM, no apical meristem; ATAF, Arabidopsis transcription activation factor; and CUC, cup-shaped cotyledon
NGS	Next generation sequencing
NMT	Phosphoethanolamine N-methyltransferase
non-tECL	transformed embryogenic cell line
nptII	Neomycin phosphotransferase II gene
nr	Non-redundant
NRQ	Normalized relative quantities
ORF(s)	Open reading frame(s)

PBS	Phosphate buffered saline
PCM	Pre-conditioning medium
PCR	Polymerase chain reaction
PED1	3-ketoacyl-CoA thiolase 2
PEM(s)	Proembryogenic masse(s)
PIN2	Auxin efflux carrier component 2
PKL	CHD3-type chromatin-remodeling factor PICKLE
PLA1	Phospholipase A1
PLDALPHA1	Phospholipase D alpha 1
PLT2	Plethora 2
PM	Plasma membrane
PPO	Class III plant peroxidase
PPR	Pentatricopeptide repeat protein
PRC	Polycomb repressive complex
PS	Polyester synthase;
PVDF	Polyvinylidene difluoride
R	repeat(s)
RGL1	Della protein
RNA	Ribonucleic acid
RNAi	RNA interference
RT-qPCR	Reverse transcription quantitative real-time PCR
ΔRWI	Relative weight increment
SA	Salicylic acid
SAC51	Supressor of acaulis 51
SAMDC4	4S-adenosylmethionine decarboxylase proenzyme 4
SBP	SQUAMOSA promoter binding protein-like
SE	Somatic embryogenesis
SEI	Somatic embryogenesis induction
SERK1	Somatic embryogenesis receptor kinase 1
SH	Schenk and Hildebrandt

SSR	Single sequence repeat
SUS	Suberin synthase
Ser	Serine
tECL	transformed embryogenic cell line
T	Thymine
T-DNA	transfer DNA
TCP	TCP family of transcription factors named by the first 4 characterized members: TEOSINTE BRANCHED1 (TB1) from maize (*Zea mays*), CYCLOIDEA (CYC) from snapdragon (*Antirrhinum majus*) as well as PROLIFERATING CELL NUCLEAR ANTIGEN FACTOR 1 (PCF1) and PCF2 from rice (*Oryza sativa*)
TCPC1	Translationally-controlled tumor protein 1
TDZ	Thidiazuron
TEM	Transmission electron microscopy
TF(s)	Transcription factor(s)
TPC1	Two pore calcium channel protein 1
TPL	Topless
TR(s)	Transcription regulator(s)
tRNA	Transfer ribonucleic acid
TRY	Triptychon
Tyr	Tyrosine
UMRs	Unique mapped reads
UTR	Untranslated region
VAL1	VP1/Abscisic acid insensitive-like1
VLCFA(s)	Very long-chain fatty acid(s)
VP1 (B3)	Related to abscisic acid insensitive3
VirG	Regulatory protein type G of the virulence group of protein from *Agrobacterium tumefaciens*
WD40-like	Structural motif of approximately 40 amino acids often terminating in a tryptophan-aspartic acid (W-D) dipeptide
WIN1/SHN1	Wax inducer 1/shine 1
WOX	WUSCHELL-related homeobox

WRKY	Transcription factors containing one or two conserved WRKY domains, about 60 amino acid residues with the WRKYGQK amino acid sequence followed by a C2H2 or C2HC zinc finger motif.
XTH24	Xyloglucan endotransglucosylase/hydrolase protein 24
Xaa	Undefined amino acid
YEP	refers to Yeast extract peptone medium
YUC1	Indole-3-pyruvate monooxygenase YUCCA 1
YUC2	Indole-3-pyruvate monooxygenase YUCCA 2
YUC4	Indole-3-pyruvate monooxygenase YUCCA 4
YUC6	Indole-3-pyruvate monooxygenase YUCCA 6
YUC10	Indole-3-pyruvate monooxygenase YUCCA 10
YUC11	Indole-3-pyruvate monooxygenase YUCCA 11
ZF	Zinc finger

ABSTRACT

The study of transcription factors (TFs) can reveal important players behind a specific developmental process. The identification of TFs DNA-targets on a genome-wide scale provides fundamental information not only on how the expression of the targeted genes are regulated but also about the molecular mechanisms behind this process. QsMYB1 is a R2R3-MYB transcription factor related to secondary growth and cork development in *Quercus suber*. Given the importance of cork quality for the cork-based industry, the study of genes involved in cork formation can provide important insights. In this context, the objective of this PhD thesis was to further study QsMYB1 gene, by identifying its DNA targets on the cork oak genome, in order to elucidate the gene regulatory mechanism in which QsMYB1 exerts its function.

The first approach was to produce *Q. suber* embryogenic cell lines. During the process large amounts of developing embryos were produce which led to the characterization of cork oak somatic embryo transcriptome over developing phases. The results revealed important gene players in embryogenesis and in cork oak embryo development, namely genes encoding for transcription factors, hormones-related, involved in germination and in root development, thus providing important knowledge for further research towards improvement of somatic embryogenesis on cork oak.

Genetic modified embryos, were selected for QsMYB1-chromatin immunoprecipitation followed by next generation sequencing, revealing the putative binding sites of QsMYB1 throughout the cork oak genome. The results evidence that gene expression regulation by QsMYB1 is directly linked with cork biosynthesis related-pathways, targeting genes encoding enzymes of the phenylpropanoid pathway and from the fatty acid and lipid metabolisms. Moreover, QsMYB1 also targets genes encoding transmenbranar lipid transporters implicated in suberin transportation and assembly. Additionally, the results showed that QsMYB1 has the potential to modulate the expression of other transcription regulators, thus amplifying its regulatory network spectrum.

Keywords: *QsMYB1*, R2-R3 Transcription Factor, ChIP-Seq, Somatic Embryogenesis, Transcriptome

Resumo

Nas células eucariotas o genoma é constituído por moléculas de ácido desoxirribonucleico (ADN) associadas a proteínas, constituindo a cromatina. Um dos mecanismos responsáveis pela expressão genética consiste na presença de sequências nucleotídicas específicas que funcionam como locais de ligação para proteínas designadas fatores de transcrição. A identificação dos locais-alvo por parte dos fatores de transcrição ao nível do genoma fornece informação fundamental não apenas acerca da regulação da expressão dos genes que lhes estão associados, mas também acerca dos mecanismos moleculares envolvidos no processo.

O gene *QsMYB1* codifica uma proteína que pertence à família dos fatores de transcrição do tipo MYB-R2R3. Este gene está envolvido no crescimento secundário em sobreiro e nos processos de formação e desenvolvimento da cortiça. A cortiça é um material vegetal único, extraído do sobreiro (*Quercus suber*) e composto por várias camadas de células de felema que sofrem um processo de suberização e morte celular programada. Devido às suas propriedades físico-químicas, a cortiça tem várias aplicações ao nível industrial, envolvendo muitas atividades e assim, promovendo o desenvolvimento socioeconómico em zonas do interior do País. No entanto, e apesar do seu valor e importância, o conhecimento científico acerca dos mecanismos regulatórios envolvidos na formação da cortiça é ainda reduzido.

Dada a importância da cortiça e o estado do conhecimento científico sobre este tema, o principal objetivo da presente Tese de Doutoramento é a identificação dos alvos do fator de transcrição QsMYB1, ao nível do genoma de sobreiro. Desta forma, pretende-se elucidar o complexo mecanismo genético pelo qual o QsMYB1 exerce a sua função regulando o metabolismo ao nível das células que irão formar a cortiça, e perceber quais as vias de biossíntese assim como os vias metabólicas que serão reguladas por si.

Neste sentido, numa primeira fase procedeu-se à indução de embriões somático de sobreiro e ao estabelecimento de linhas embriogénicas. A produção de embriões somáticos é um passo fundamental para a prossecução de estudos de análise funcional de genes, tanto através de linhas de sobre-expressão como de silenciamento em plantas não-modelo. Depois da otimização deste processo

em *Q. suber* foi possível a obtenção de grande quantidade de embriões somáticos em várias fases de desenvolvimento, permitindo o estudo transcritómico da dinâmica envolvida na embriogénese somática em sobreiro. Desta forma, é possível aumentar o conhecimento científico da temática e obter informação acerca das vias biossintéticas mais importantes e dos genes diferencialmente expressos ao longo deste processo. Assim, a disponibilidade de embriões somáticos em várias fases de desenvolvimento permitiu o isolamento de embriões somáticos em quatro fases: (1) globular, (2) coração/torpedo, (3) cotiledonar e (4) cotiledonar madura. Os respetivos transcritomas foram gerados por RNA-Seq e organizados num transcritoma *de novo*. A análise dos dados gerados permitiu a identificação de vários grupos de genes diferencialmente expressos e com relevância em vários aspetos da embriogénese e desenvolvimento embrionário em sobreiro, nomeadamente genes que codificam fatores de transcrição, relacionados com a homeostasia de fito-hormonas, envolvidos na germinação e relacionados com o desenvolvimento do sistema radicular. Para além disso, foi possível revelar as ontologias genéticas (GO) dos processos celulares mais representadas, evidenciando assim os processos aos quais os genes diferencialmente expressos estão associados. Este estudo contribuiu pela primeira vez para a caracterização e identificação de vários genes diferencialmente expressos em fases distintas do desenvolvimento de embriões somáticos de sobreiro, permitindo uma visão global da dinâmica do processo. Os resultados obtidos constituem um importante conhecimento científico e permitirão a possibilidade de futuros estudos funcionais a nível genético no sentido de esclarecer os mecanismos associados ao desenvolvimento e maturação de embriões somáticos de sobreiro. Os dados gerados abrem ainda a possibilidade de futuros estudos funcionas para o desenvolvimento de protocolos de regeneração embriogénica que poderão ser integrados em futuros programas de melhoramento e reflorestação em sobreiro.

Duas das linhas embriogénicas de sobreiro obtidas foram caracterizadas quanto à sua resistência à canamicina, escolhida como agente seletivo. Por esta metodologia, foi selecionada a linha embriogénica de sobreiro Ce1, a qual pôde ser usada para gerar embriões geneticamente modificados a sobre-expressar a proteína QsMYB1 fundida com um epítopo triplo do tipo FLAG (QsMYB1::3xFLAG). Após a geração dos embriões, estes foram transformados

por uma estratégia envolvendo a co-cultura com *Agrobacterium tumefaciens*. Após a selecção dos embriões geneticamente modificados durante um período de vinte e quatro meses, recorrendo à proliferação de embriões somáticos por embriogénese secundária, foi posteriormente confirmada a inserção do ADN transferido (T-ADN) e detetada a expressão de transcritos do gene *QsMYB1* por RT-qPCR, a presença de proteínas QsMYB1::3xFLAG por Western Blot e a sua localização no núcleo.

Com este material vegetal procedeu-se à imunoprecipitação da cromatina (ChIP) associada aos locais alvo do QsMYB1 através da aplicação da técnica ChIP-Seq, que permitiu através de uma plataforma de sequenciação de segunda geração (Seq), a identificação desses mesmos locais ao nível do genoma de sobreiro. Após tratamento bioinformático dos dados obtidos. Os resultados revelam que o fator de transcrição QsMYB1 tem vários potenciais alvos ao nível do genoma do sobreiro, tendo grande afinidade para sequências de nucleótidos do tipo [CWHCAA], [CYTCBTC] e [BKTGG], que constituem os principais motivos de ligação. A identificação dos locais de ligação mostrou que estes ocorrem ao nível dos exões (26%), intrões (19%), regiões intergénicas (18%) e dos promotores (13%). Os dados gerados e a análise bioinformática efetuada, evidenciam que os potenciais genes-alvo do QsMYB1 codificam para enzimas da via de síntese dos compostos fenilpropanóides (4-comarato-CoA ligase, cinamil álcool desidrogenase, peroxidase de plantas classe III), da biossíntese da suberina (acil-CoA ligase específica para ácido gordos de cadeia muito longa, ceto-acil sintase, citocromo P450 oxidase 86A1, citocromo P450 oxidase 86A8, glicerol-3-fosfato aciltransferase) e inseridas nos metabolismos dos ácidos gordos e lipídico (acil-CoA oxidase, ácido gordo sintase, enoil-ACP reductase, transportador de grupos acilo (ACP), aldeído redutase dependente de NADPH, diacilglicerol *O*-aciltransferase, fosfolipase A1 e fosfoetanolamina N-metiltransferase). O QsMYB1 parece ainda funcionar como um regulador da expressão de genes envolvidos no transporte lipídico através da membrana celular (ABCGs e LTPs), os quais por sua vez, estão envolvidos no transporte de monómeros e polimerização da suberina. Finalmente, os resultados evidenciam que o QsMYB1 poderá regular a expressão de genes que codificam outros reguladores da transcrição, nomeadamente membros das famílias CCHC

ZF, C2H2-type ZF e BED-type ZF, assim como genes que codificam fatores de transcrição do tipo WD40-like, AP2/ERF e bHLH.

Como conclusão, os resultados evidenciam que o fator de transcrição em estudo terá um papel importante na regulação da formação e diferenciação da cortiça. No entanto, serão necessários futuros estudos de forma a compreender e confirmar o papel específico de cada um dos genes identificados, e do próprio QsMYB1, na formação da cortiça. Ainda assim, os dados gerados representam informação fundamental, sobre as vias metabólicas envolvidas na diferenciação da cortiça. A informação associada a estes dados constituem ainda um ponto de partida para o esclarecimento aprofundado da biossíntese de suberina em sobreiro. Estas descobertas, são também o início do desvendar do complexo mecanismo molecular associado à formação da cortiça. Por último, o conhecimento obtido nesta Tese, nomeadamente acerca das enzimas e vias de biossíntese envolvidas na formação da cortiça representa uma importante e sólida base para futuros programas de melhoramento, tendo em vista parâmetros de formação e qualidade da cortiça, ou num futuro delineamento de estratégias de engenharia metabólica ao nível da produção de cortiça.

Palavras-Chave: *QsMYB1*, Fator de Transcrição R2-R3, ChIP-Seq, Embriogénese Somática, Transcritómica

TABLE OF CONTENTS

Abbreviations vii

Abstract xvii

Resumo xix

Research Objectives xxv

Chapter I - General Introduction 1

Chapter II - Cork Oak somatic embryos, a model for functional studies: induction, selective agent characterization, transformation, and cryopreservation 59

Chapter III - Transcriptome dynamics of cork oak (*Quercus suber*) somatic embryogenesis reveals active gene players in transcription regulation and hormone homeostasis of embryo development 85

Chapter IV - ChIP-Seq reveals that QsMYB1 directly targets genes involved in lignin and suberin biosynthesis pathways in cork oak (*Quercus suber*) 139

Chapter V - Conclusions and Final Remarks 187

Chapter VI - Acknowledgements 193

RESEARCH OBJECTIVES

Since the Ancient Greece times, cork has been used for a variety of applications with an associated socio-economic impact. Biologically it has generated interest by its unique tissue composition, well represented in cork oak. Although some knowledge about cork formation is available, much is still to be deeply investigated to understand the molecular mechanisms behind the cell differentiation and cork formation. To improve the knowledge about this subject it is necessary to know the genes involved in process. *QsMYB1* gene encodes a MYB transcription factor, up-regulated in organs and tissues associated with cork. Due to its transcription factor nature QsMYB1 regulates the gene expression either activating or repressing specific genes transcription.

The main objective of the present work is to *in vivo* identify the putative binding sites of the transcription factor QsMYB1 in order to obtain information about the genes regulated by this TF. Moreover, to achieve this objective, cork oak somatic embryos were generated. Embryos in various development phases of development were also selected and used for transcriptome analysis in order to contribute to the knowledge about somatic embryogenesis in *Quercus suber.*

To accomplish this purpose, it is fundamental to:

1) establish and cryopreserve embryogenic cell lines of *Quercus suber;*
2) characterize the transcription dynamics of cork oak somatic embryogenesis;
3) generate *Q. suber* genetic modified lines overexpressing QsMYB1;
4) ChIP and sequence the DNA associated with QsMYB1;
5) *in silico* discover QsMYB1 target motifs;
6) identify putative genes regulated by QsMYB1;
7) relate the target genes with biosynthetic pathway(s) associated to cork formation.

CHAPTER I

GENERAL INTRODUCTION

TABLE OF CONTENTS – CHAPTER I

General Introduction 1
1. Transcription factors 3
1.1. Plant transcription factors: the MYB transcription factors superfamily and the R2R3-MYB family evolution, structure and function 4
2. Cork oak 7
3. Cork, a unique and noble tissue 9
4. Suberin chemical composition and structure 12
5. Suberin biosynthesis 14
5.1. The aliphatic domain 15
5.2. The aromatic domain 18
5.3. Transportation and assembly of suberin monomeric units 19
5.4. Regulation of suberin biosynthesis 21
6. QsMYB1 - The R2R3-MYB transcription factor involved in suberin biosynthesis 24
7. Identifying transcription factors binding sites 25
8. ChIP-Seq as a robust technique to discover protein-DNA interactions in plants 27
9. Motif discovery from ChIP-Seq data 29
10. Somatic Embryogenesis in Plants 30
10.1 Regulatory genes in somatic embryogenesis 31
10.2 Somatic embryogenesis in *Quercus suber* 34
10.2.1 Embryos maturation, germination and plantlet conversion 35
11. RNA-sequencing, a way to discover the genes involved in somatic embryogenesis 36
12. The RNA-Sequencing methodology 37
13. Bioinformatics analysis to unravel differential genes expression 39
14. Developing a method to identify the DNA-targets of QsMYB1 in cork oak . 39
15. References 41

1. TRANSCRIPTION FACTORS

Gene expression is regulated through the integrated action of many *cis*-regulatory elements, including core promoters and promoter-proximal elements, as well as different *cis*-regulatory modules localized at greater distances from the transcription start sites such as enhancers, silencers, insulators and tethering elements (Wittkopp and Kalay 2011). Among this multitude of elements, enhancers and their associated transcription factors (TFs), which are gene expression *trans*-regulatory elements (Gilad et al. 2008), have an important role in gene expression initiation. Enhancers are fragments of deoxyribonucleic acid (DNA), commonly with a few hundred base pairs in length, that recruit TFs through short, specific DNA sequences, named motifs, in order to regulate gene transcription. The identification of DNA regions targeted by TFs on a genome-wide scale provides fundamental information not only about the regulation of target genes but also about the complex and orchestrate mode of action between TFs, and enhancer elements or other *cis*-regulatory elements (Lauria and Rossi 2011). To fully understand the mode of action of TFs as simple modulators of a specific gene transcription or in robust transcriptional programs, it is necessary to take into account several aspects of the molecular biology of the DNA environment. These include the interaction of DNA-TFs, TFs-TFs and TFs-cofactors as well as the DNA accessibility in the context of chromatin organization (Biłas et al. 2016). TFs typically target 6-12bp-long genomic regions what suggests that more complex rules, besides the simple affinity of TFs for DNA, are likely to be involved in the control of enhancer occupancy and a functional outcome (Wittkopp and Kalay 2011). In fact, TFs typical bind to enhancer regions, which contain clusters of different TF binding sites. This combinatorial binding can result in discrete and precise patterns of transcriptional activity (Wittkopp and Kalay 2011). Additionally, the activation of TFs effectors in response to signalling cascades can trigger a more restricted activity of enhancers when acting together with other TFs with overlapping expression domains. Also the TF occupancy at different developmental stages or in different conditions can often lead to diverse occupancy of enhancers (Lauria and Rossi 2011). Together, this mode of action constitute a precise mechanism of transcriptional regulation, involving several TFs with affinity for the same

enhancers which in some cases are essential for the recruitment of other TFs increasing the regulatory effect on specific biological process as plant development, differentiation and response to various environmental signals (Lauria and Rossi 2011; Biłas et al. 2016).

1.1. Plant transcription factors: the MYB transcription factors superfamily and the R2R3-MYB family evolution, structure and function

Transcription factors are commonly defined as proteins that bind to promoters in a sequence-manner to regulate expression of the target genes. Plant transcription factors are characterized by a variety of TF families with a great variety of DNA-binding specificities (Riechmann et al. 2000). According to different TFs databases, in *Arabidopsis* there are around 2000 TF genes belonging to 56 families, constituting more than 5% of all genes in this genome (Davuluri et al. 2003; Guo et al. 2005; Iida et al. 2005; Riaño-Pachón et al. 2007). Around 50% of *Arabidopsis* TFs are plant-specific and possess DNA-binding domains found only in plants, such as the APETALA2/Ethylene Responsive element binding factors (AP2/ERF), NAC, DNA binding with one zinc finger (Dof), YABBY, WRKY, Golgi-associated retrograde protein (GARP), TCP, SQUAMOSA promoter binding protein-like (SBP), related to abscisic acid insensitive 3 (VP1 (B3)), ethylene insensitive like (EIL) and the floral meristem identity LEAFY (LFY) TF families. In plants, the most represented TFs belong to myeloblastosis (MYB), AP2/ERF, basic helix-loop-helix (bHLH), NAC, MADS-box (MADS), basic leucine zipper (bZIP) and WRKY TF families (Iida et al. 2005). Amongst the most represented TFs in *Arabidopsis,* are MYB proteins integrating a superfamily of TFs that has the largest number of members of any *Arabidopsis* TF family, counting with 196 members with a highly conserved DNA-binding domain, known as MYB domain (Dubos et al. 2010). This domain is characterized by up to four imperfect amino acid sequence repeats (R) of around 52 amino acids, each forming three α-helices. The second and third helices form a helix-turn-helix (HLH) structure which permits the interaction with DNA (Dubos et al. 2010). The function of MYB proteins in plants, have been investigated in numerous plant species such as *Arabidopsis thaliana*, *Zea mays* (maize), *Oryza sativa* (rice), *Petunia hybrid* , *Vitis vinifera* (grapevine), *Populus tremuloides* (poplar), *Malus*

domestica (apple) and *Quercus suber* (cork oak), between others (Dubos et al. 2010; Almeida et al. 2013a; Almeida et al. 2013b). MYB proteins are classified into four different classes depending on the number of adjacent repeats (R1, R2, R3, R4) (Figure I.1). All the four classes occur in plants, representing the kingdom with highest diversity of MYB proteins. The smallest class with less knowledge about, is the 4R-MYB group with a single 4R-MYB protein encoded in several plant genomes (Dubos et al. 2010). The 3R-MYB group is typically encoded by five genes in higher plants genomes and have been found in most eukaryotic genomes, representing a conserved gene class with divergent roles in cell cycle control (Ito 2005). In turn, there is a heterogeneous class of single or a partial MYB repeat, designated by “MYB-related”. This class includes R3-type genes involved in the control of cellular morphogenesis and differentiation as *TRIPTYCHON (TRY)* and *CAPRICE (CPC),* respectively*,* and in secondary metabolism control as *MYBL2* which acts as negative regulator of anthocyanin biosynthesis (Dubos et al. 2008). Also included in MYB-related genes are the evolutionary older R1/R2-type, such as *CIRCADIAN CLOCK ASSOCIATED1 (CCA1) and LATE ELONGATED HYPOCOTYL (LHY)* which encodes essential components of the central circadian oscillator, or the *KANADI* and *GOLDEN2-LIKE* genes involved in organ morphogenesis and development (Dubos et al. 2010).

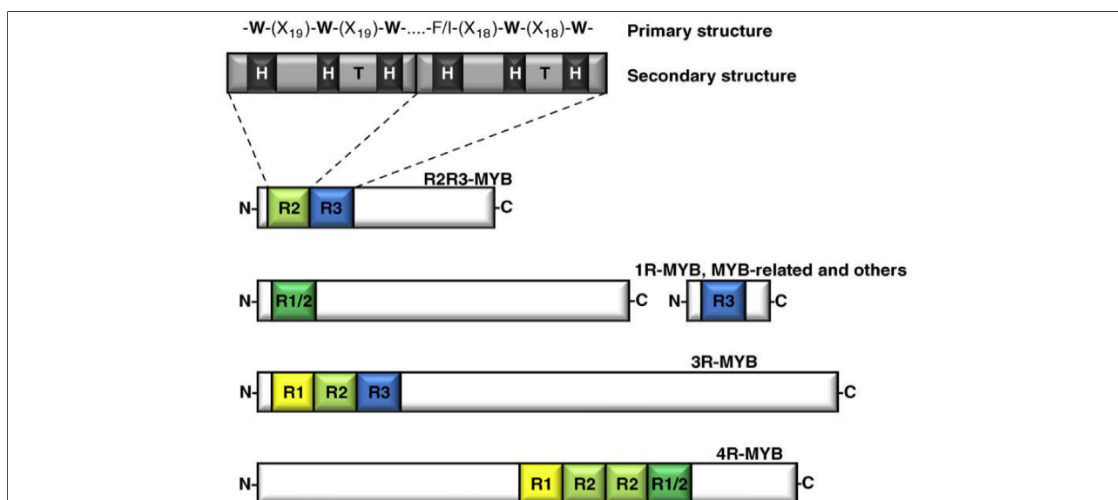


Figure I.1 | Plant MYB transcription factor classes. Different plant MYB protein classes, depending on the number of MYB repeats (R). Typical R2R3-MYB primary and secondary structures are indicated. H, helix; T, turn; W, tryptophan; X, amino acid (*Adapted from Dubos et al., 2010*)

However the majority of plant *MYB* genes encode proteins of the R2R3-MYB class which is likely to have evolved from an R1R2R3-*MYB* gene ancestor, by losing the sequence that encodes the R1 repeat and leaving to a subsequent expansion of the gene family (Ambawat et al. 2013). The R2R3-MYB TFs structure is characterized by a N terminal DNA-binding domain, the MYB domain, and an activation or repression domain typical located at the C terminus. Except for the highly conserved MYB domain, the other regions of all R2R3-MYB proteins are highly variable. Based on the amino acid motifs in the C terminal domain and on the conservation of the DNA-binding domain, the R2R3-MYB proteins can be divided into at least 25 subgroups (Dubos et al. 2010). Most of these subgroups were defined for *Arabidopsis* R2R3-MYB proteins and in some cases, must be expanded if we consider other plant species. In fact phylogenetic studies identified new R2R3-MYB subgroups in poplar, grapevine, maize, rice and cotton (*Gossypium raimondii*) suggesting that these proteins evolved to perform a specialized role in plant-specific processes (Dubos et al. 2010; Du et al. 2012; He et al. 2016).

The R2R3-MYB TFs family has been implicated in the regulation of diverse plant-specific processes. These processes include primary and secondary metabolism, as in the case of SmMYB36 that regulates the gene transcription levels of the phenylpropanoid, tyrosine, methylerythritol phosphate and tanshinone pathways in *Salvia miltiorrhiza* hairy roots (Ding et al. 2017); or as in the case of PtMYB1, PtMYB8 and PtMYB4 that are suggested to be involved in the regulation of phenylpropanoid and lignin biosynthesis in poplar (Patzlaff et al. 2003; Bomal et al. 2008); or the case of AtMYB58, AtMYB63 and AtMYB85 that activate lignin biosynthesis in fibres and vessels; or like AtMYB68 that act as negative regulator of lignin biosynthesis. Also, AtMYB41, AtMYB107 and MYB9 were recently reported to be involved in the regulation of suberin biosynthesis being AtMYB107 and AtMYB9 described as positive regulators of suberin deposition in *Arabidopsis* seed coats (Kosma et al. 2014; Lashbrooke et al. 2016; Gou et al. 2017). Several R2R3-MYB TFs are also involved in the control of cell fate and identity, namely AtMYB0/GL1 and AtMYB23 that control the determination of epidermal cell type and AtMYB66 that participates in a positive feedback loop reinforcing the cell fate establishment process (Kang et al. 2009; Tominaga-Wada et al. 2012). Some developmental processes are also regulated

by R2R3-MYB TFs. This is the case of AtMYB37/RAX1, AtMYB38/RAX2 and AtMYB84/RAX3 that are redundant regulators of axillary meristems development (Keller 2006; Muller 2006) or AtMYB68 that is a root growth specific regulator impacting the overall plant structure in case of high temperatures (Feng et al. 2004). Finally, R2R3-MYB TFs are also described as important players in the regulatory networks of responses to biotic and abiotic stresses as well as in signal transduction pathways of phytohormones. This is the case of AtMYB96 that regulates drought response together with pathogen resistance response by integrating the abscisic acid signal transduction (Seo et al. 2009); or the case of AtMYB68 whose expression is modulated by heat stress and confers heat tolerance as demonstrated by *myb68* mutants that showed a reduced ability of growing at high temperatures (Feng et al. 2004). Similarly, the expression of *QsMYB1,* a homolog of the *AtMYB68* gene in cork oak, is modulated by heat and drought stresses as well as by methyl jasmonic acid exogenous application (Almeida et al. 2013b).

2. CORK OAK

Cork oak or *Quercus suber* L. is a diploid angiosperm, member of the *Fagaceae* family and an evergreen broadleaved tree native to the Mediterranean area. Currently, cork oak flourishes in the warmer parts of the humid and sub humid western Mediterranean Basin covering around 1.5 million ha in Europe and 1 million ha in North Africa (Bugalho et al. 2011). The species has an average lifespan of more than 200 years, commonly growing to a height of 20 m, and reaching 25 m in ideal conditions (Gil and Varela 2008). It is a monoecious wind-pollinated species with a protandrous system to ensure cross-pollination (Boavida et al. 1999) . Male inflorescences are long, pedunculated and arise from the terminal and axillary buds of the previous year's branches. Female flowers appear on vigorous new growth. Depending of the regions, the flowering season extends from April onwards throughout Summer months (Elena-Rossello et al. 1993; Boavida et al. 1999). Seeds (acorns) are 2-3 cm in length in a fairly deep cup with elongated scales. Acorns maturation is observed during the year of set or in the following year depending of environmental factors and meteorological conditions (Díaz-Fernández et al. 2004). Annual maturation is more common in

the southern regions, while biennial maturation is more frequently observed in northern populations; however, some individuals display both patterns simultaneously (Elena-Rossello et al. 1993; Bugalho et al. 2011).

Cork oak is a well-adapted species to the Mediterranean climate. It prefers average temperatures of 15ºC and cannot tolerate very low temperatures, which limit its northern and altitude range. It can grow in a variety of soils but tends to prefer sandy and lightly structured ones. Remarkably, cork oaks have an extraordinary capability for growing under drought and dry climate conditions because of their extensive and deep root systems. This specialized root system allows the tree to tap water from the soil and subsoil maintaining the water status and xylem conductance above lethal levels throughout drought periods, frequent in Summer months in Mediterranean countries. In addition, this species has the ability to protect crucial organs and tissues from dehydration by closing stomata on leaves, therefore reducing water loss. Cork oak has short-lived foliage, late flushing patterns and a remarkable leaf sclerophyllous phenology designed to resist drought, high temperatures and high exposure to sunlight. The species has a unique protection against the periodic fires that occur in Mediterranean regions, its outer bark, the cork. Cork has excellent insulating properties protecting stem tissues from scorching and burning (Aronson et al. 2009; Bugalho et al. 2011).

In the Mediterranean basin, cork oak occurs in forests or open woodlands. It is frequently found in managed woodland systems named *montado* in Portugal, *dehesa* in Spain and *azaghar* in some regions of Morocco, which are characterized by a sparse cork oak cover (30-60 trees per hectare), sometimes mixed with *Quercus ilex,* conifers such as *Pinus pinaster* and *Pinus pinea*, and understorey composed of grassland for livestock, cereal crops and scrublands (Aronson et al. 2009). These agrosystems have been specifically structured around cork oak, in which trees have been selected, pruned, and shaped by people for generations. As a result of this human action, a characteristic landscape has evolved over centuries and reflects the dedicated nurture by nature of the Mediterranean people. In fact, this dynamic and co-evolutionary relationship between humans and cork oak is a remarkable example of environmental sustainability.

Nowadays, the natural distribution of cork oak in the western Mediterranean extends from the Iberian Peninsula and Morocco to the western

coast of Italy, also occurring on the Balearic Islands, Sardinia, Sicily, Corsica in the south of France and in North coast of Algeria and Tunisia (Figure I.2). Cork oak was also artificially introduced in several non-Mediterranean countries as Bulgaria, New Zealand, Australia, Chile and California with reasonable acclimatization success (Aronson et al. 2009).

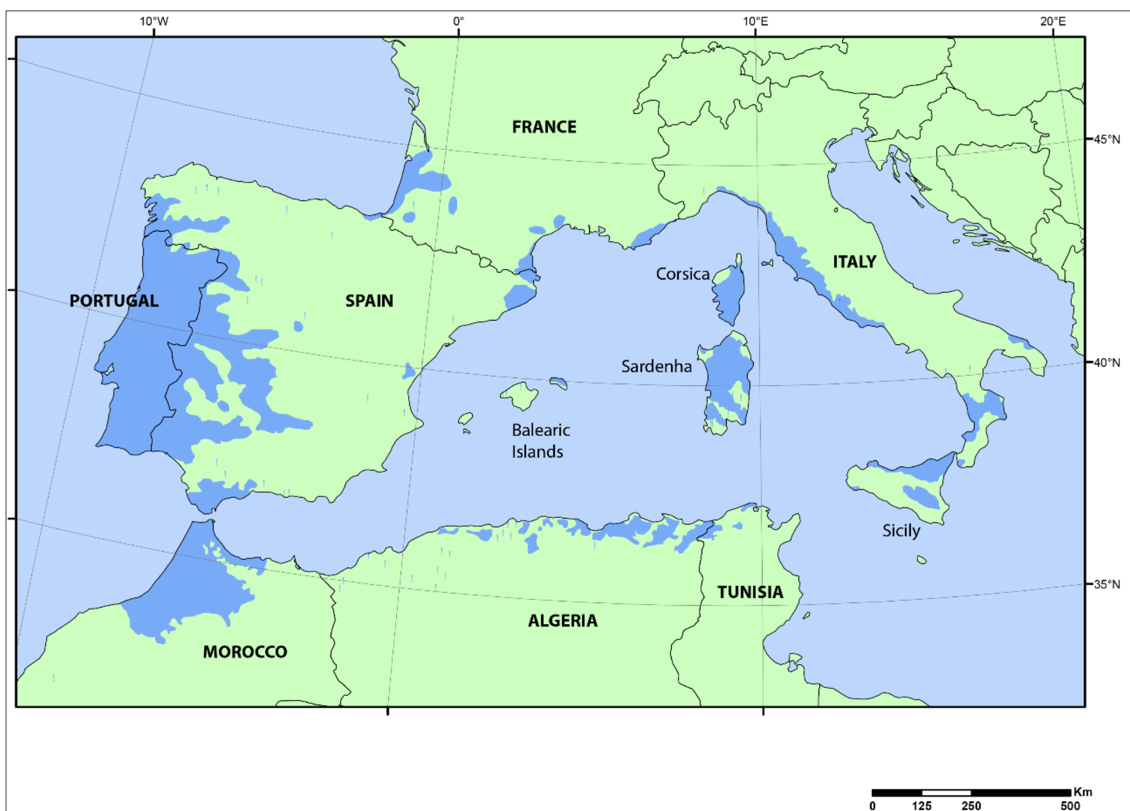


Figure I.2 | Cork oak geographic distribution in the Mediterranean basin (*Adapted from EUFORGEN 2009, www.euforgen.org*)

3. CORK, A UNIQUE AND NOBLE TISSUE

Cork oak is well-known by an almost unique characteristic in the plant kingdom, the production of cork. Cork is a natural product with remarkable characteristics: lightness, waterproof, resistance to rot, flexibility, chemical stability and retardant to fire (Pereira 2007a). Since the Ancient Greece that cork is used with a variety of applications. Today, the main value of cork crop is due to the production of wine stoppers; however cork has many other applications, including flooring, thermal insulation, aerospace industry, fashion products, design products and handicraft (Gil 2009). It is not therefore unexpected that the

exploitation of cork and cork products has an important economic and social impact. Biologically, cork has always generated great interest from scientists. In fact, thin slices of cork were the first plant tissue observed by Robert Hooke in the first microscope observations described in his book *Micrographia*. He found that the tissue was composed by small hexagonal structures he called *cells* (Figure I.3) (Hooke 1665). Interestingly, Hooke offered great attention and admiration to the *interstitia* or walls from these cells, mentioning a relation between cork properties and the cell walls. Furthermore, Hooke refers to cork as *"a kind of excrescence, or a substance distinct from the substances of the entire Tree"* (Hooke 1665). Since the publication of Hooke's observation, the study of cell walls in cork has evolved to the knowledge that in plant anatomy cork or phellem is a tissue derived from phellogen, which is part of the periderm.

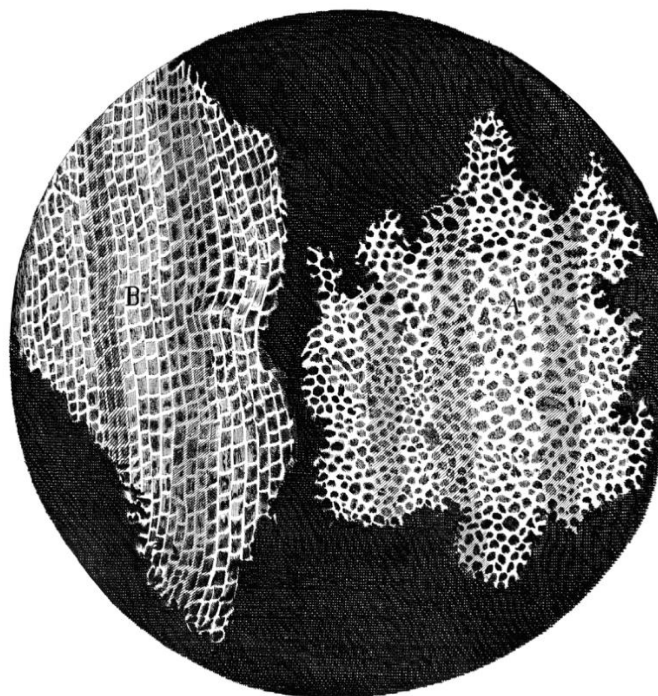


Figure I.3 | Reproduction of the original observation of Robert Hooke about slices of cork (*Adapted from Hooke 1665*)

Cork formation, results from the activity of the secondary meristem phellogen also called cork cambium. Phellogen, together with phellem and phelloderm, constitute the periderm (Figure I.4), an effective barrier for the plant against the outside environment, in the case of *Quercus suber* mainly due to the cork layer. Phellem is composed by cells whose walls are impregnated with

suberin, a lipid polyester analogous to cutin (Vishwanath et al. 2015). Phellogen cells divide periclinally, giving rise to cork cells to the outside and phelloderm cells to the interior part of the stem (Pereira 2007b; Franke and Schreiber 2007). Cork stripping is advised to be performed when the phellogen is highly active and thin-walled active cell layers are produced. If performed when the cork cambium is inactive, the stripping process may damage the tree because the inner bark is separated through the vascular cambium, which unlike the phellogen does not regenerate once exposed (Aronson et al. 2009), compromising the tree health. The rate of division of phellogen is much higher to the outside than to the inside resulting in much more phellem than phelloderm layers. The number of layers of phellem cells produced in a growth season varies from species to species. In *Quercus suber* the number is larger than in the majority of species, forming the characteristic cork oak tissue composed by a multilayer of suberized-death cells. Contrasting to the phelloderm cells, that are living cells with non-suberized walls similar to parenchyma cells of the cortex, phellem is composed by cells forming a compact tissue without empty spaces, lacking intercellular areas. The immature cork cells pass through a maturation process through deposition of a thick layer of suberin on cell walls resulting in cell death. In turn, the cell protoplasm degenerates and the lumen of the cell is replaced by air or resiniferous and tanniniferous compounds (Evert 2006).

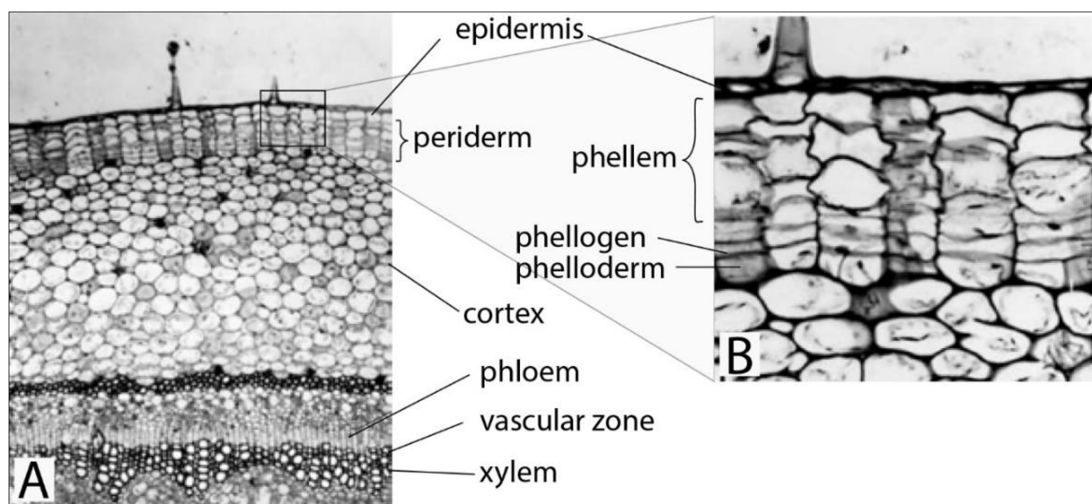


Figure I.4 | (A) Transverse section of a young stem of Pelargonium illustrating the periderm and adjacent tissue regions. Magnification x 48. (B) Enlargement of the periderm showing the organization of phellogen, phelloderm and phellem (*Adapted from Chaffey 2011*)

4. SUBERIN CHEMICAL COMPOSITION AND STRUCTURE

Suberin is the main chemical compound of cork. Chemically, it may be defined as a complex glycerol-based polymer with polymeric aliphatic and associated aromatic compounds (Graça and Santos 2007). The mechanisms behind secretion and assembly of this complex biopolymer in the apoplast as well as the biosynthetic pathways involved in suberin synthesis are complex and involve multiple enzymatic reactions (Ranathunge et al. 2011).

Suberin is present in several protection tissues at aerial or subterranean organs of plants. In addition to mature cork, high suberin contents are found in potato periderm. The rhizodermis/hypodermis of roots is also characterized by suberin depositions in cell walls (Schreiber et al. 1999). In C_4 plants suberin constitutes an impermeable layer between the bundle sheath and mesophyll cells (Franke and Schreiber 2007; Heldt and Piechulla 2011). At the cellular level, suberin is deposited inside the primary cell wall occurring in three forms: casparian strips, formed by incrustation of suberin, lignin and other cell wall components; as secondary cell wall modifications, deposited as sheets on the inner surface of cell walls just outside the plasma membrane; and deposited in the intermicrofibrillar channels in the wall of a rhizodermis (Ranathunge et al. 2011). Observations made by transmission electron microscopy (TEM) revealed that suberized cell walls have a distinctive lamellar appearance (Figure I.5) of alternating electron opaque and translucent bands, which is considered as a diagnostic of suberized cell walls (Graça 2015).

In *Quercus suber*, suberin can contribute with more than 50% of cork dry weight (Graça and Pereira 1997). The chemical identity of all constituents of the suberin polymer is unknown much because the intact suberin macromolecule is insoluble and cannot be extracted in an unaltered form for analytical studies. It is generally accepted that suberin has an aliphatic domain cross-linked with an aromatic lignin-like domain (Bernards 2002). The aliphatic domain is a cutin-like glycerol polyester esterified with phenolic compounds (Graça and Santos 2006) whereas the aromatic domain is composed by polyphenolics probably related with the aliphatic domain linked to the cell wall (Bernards and Lewis 1998). Aliphatic suberin monomers are glycerol, long-chain α,ω-diacids and long-chain ω-hydroxyacids and minor quantities of 1-alkanols and 1-alkanoic acids (Graça

and Pereira 2000), although the relative abundance of these monomeric units varies among plant species (Graça and Santos 2007). Regarding to aromatic domain, some studies had shown that suberin is a hydroxycinnamic acid-derived polymer constituted by ferulic acid and N-feruloyl tyramine (Negrel et al. 1996; Bernards and Razem 2001), while other researchers reported a significant amount of monolignols (Yan and Stark 2000).

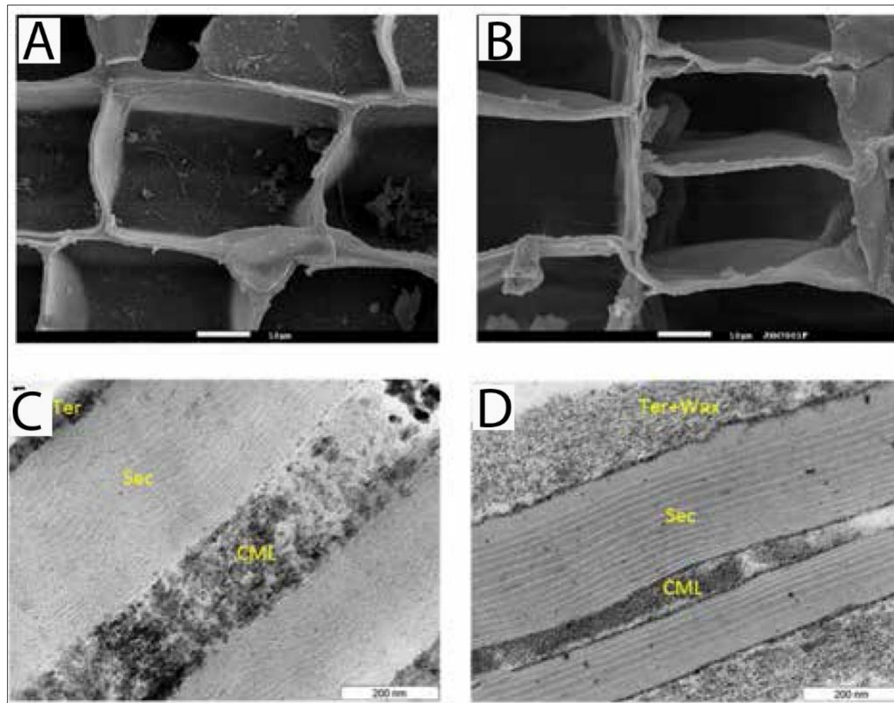


Figure I.5 | | Scanning electron microscopy of suberized cells in cork (A) and potato periderm (B) (white bars = 10 µm). Transmission electron microscopy views of cell walls in cork (C) and potato periderm (D). CML, compound middle lamella (mostly primary wall); Sec, the typical lamellar deposition in secondary cell wall; Ter, tertiary cell wall; Wax, deposited waxes (white bars = 200 nm) (*Adapted from Graça 2015*)

The macromolecular structure of suberin is not fully understood, however several models have been proposed (Kolattukudy 2001; Bernards 2002; Graça and Santos 2007). Graça and Santos (2007) hypothesized that two types of aromatic-based polymers are present in suberized cell walls: one linked to polysaccharides present in the primary (and tertiary) cell wall which may constitute a polymer similar to lignin and based on monolignols; and other, which may be a ferulic acid-based polymer associated with the aliphatic suberin domain (Figure I.6) (Graça and Santos 2007). Generally it is accepted that long-chain and very-long-chain α,ω-diacids esterified with glycerol at both ends constitute

the basic monomeric units of suberin. The biopolymer is then arranged two- and three-dimensionally by the formation of ester bonds to additional α,ω-diacids and ω-hydroxyacids, forming an insoluble polymer. Ultimately, hydroxycinnamic acids have been proposed to covalently link the aliphatic suberin polyester to the cell wall possibly involving some monolignols (Franke and Schreiber 2007).

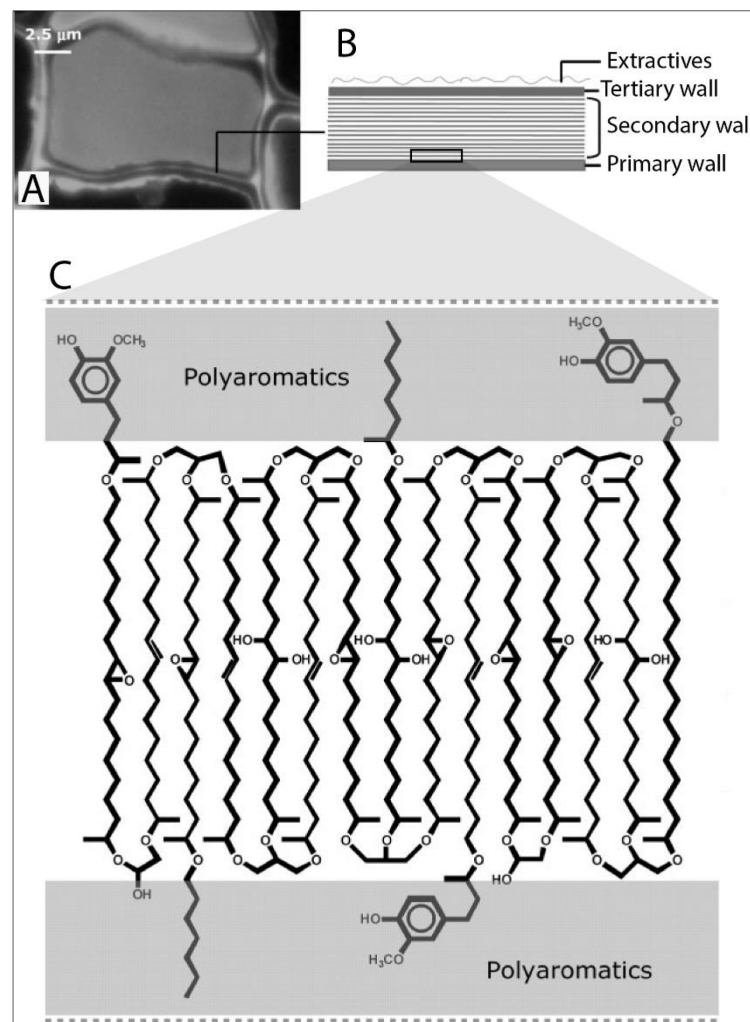


Figure I.6 | Model of the suberin macromolecular structure proposed by Graça and Santos 2007 (*Adapted from Graça and Santos 2007*)

5. Suberin biosynthesis

Suberin deposition requires the biosynthesis, transport and assembly of the monomeric units to the cell wall. As a consequence of the complex constitution of the suberin heteropolymer, precursors derive from diverse biosynthetic pathways and the various processes involved in suberin deposition

are expected to have a complex regulatory network. Although most of the enzymes related with suberin biosynthesis are associated with membranes or part of enzyme complexes which difficult their study, several have been characterized (Franke and Schreiber 2007; Pollard et al. 2008; Höfer et al. 2008; Ranathunge et al. 2011).

5.1. The aliphatic domain

The biosynthethic pathways for aliphatic suberin monomers has been hypothesized by several authors involving ω-hydroxylation of fatty acids and subsequent oxidation to α,ω-dicarboxilic acids, elongation of long-chain fatty acid precursors, activation of fatty acids to fatty acyl-CoA thioesters, reduction of fatty acyl chains to primary fatty alchohols, acylation, phenolic compound incorporation and monomeric polymerization (Figure I.7) (Kolattukudy 2001; Bernards 2002; Franke and Schreiber 2007; Vishwanath et al. 2015).

The first step in fatty acids biosynthesis is the activation of fatty acids to fatty acyl-CoA thioesters. In *Arabidopsis*, there are nine long-chain-acyl-CoA synthases (LACS) which acyl activates long-chain fatty acids (Shockey et al. 2002). Although *LACS* genes have not yet been reported in the suberin biosynthetic pathway, chemical analysis of *lacs2* mutants indicates a role of LACS genes in suberin formation (Li-Beisson et al. 2013). In addition, LACS enzymes may catalyse the production of ω-hydroxy acids and α,ω-dicacids before esterification with glycerol (Vishwanath et al. 2015).

Suberin is composed by long-chain and very long-chain fatty acids, which constitute essential compounds for the aliphatic suberin domain. Fatty acids elongation is catalysed by the fatty acid elongase complex, a multienzyme system composed of four enzymes (Samuels et al. 2008). The first enzyme in this complex is the β-ketoacyl-CoA synthase (KCS) which controls the extension of elongation of long chain fatty acids (Vishwanath et al. 2015). In *Arabidopsis* there are 21 *KCS* genes. However, it is difficult to assign a specific function to each single gene since long-chain fatty acids are common to different biosynthetic pathways (Vishwanath et al. 2015). *DAISY/AtKCS2* and *AtKCS20* are involved in C_{20} chain suberin precursors, while *kcs2* and *kcs20* mutants present a reduction in the content of C_{22} and C_{24}, very long-chain fatty acids, derivatives of aliphatic

root suberin (Lee et al. 2009). Similarly, silencing of potato *StKCS6* resulted in reduction of suberin aliphatic monomers with chain lengths of C_{28} and higher in tuber periderm, although precursors with chain lengths of C_{26} and lower are accumulated (Serra et al. 2009a).

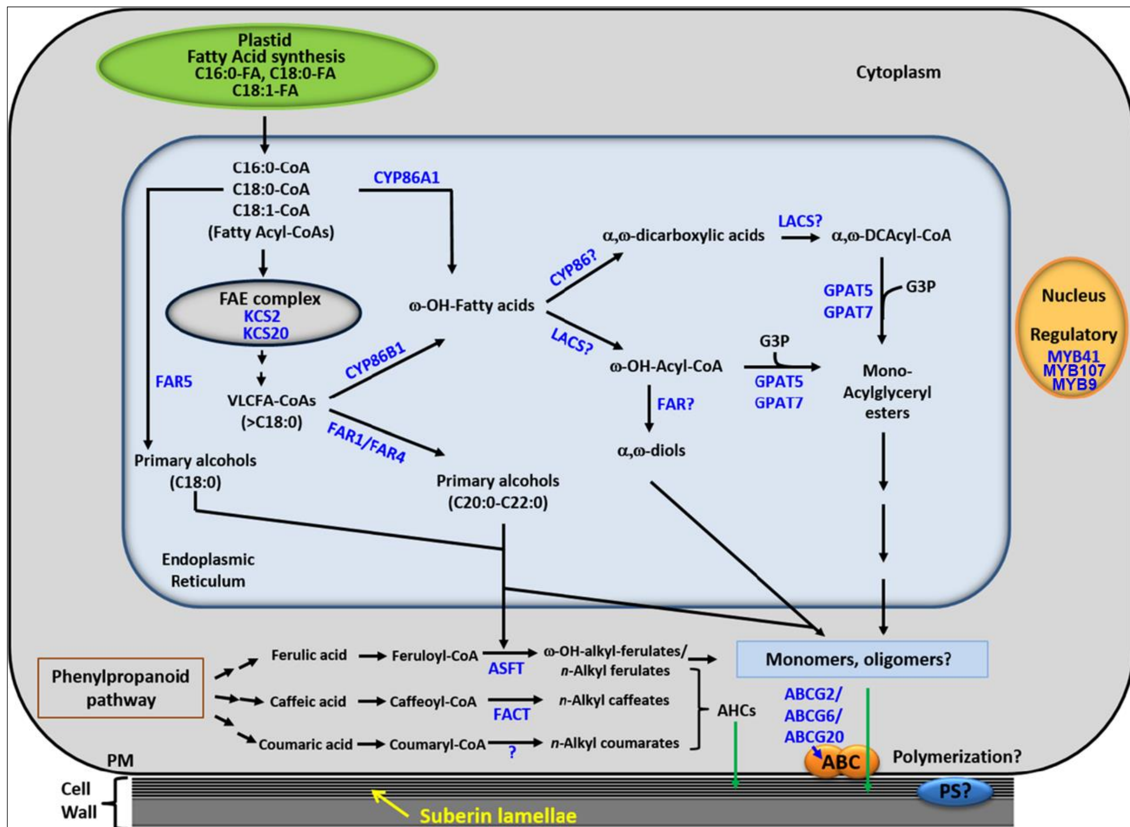


Figure I.7 | Diagram of suberin biosynthetic pathway with subsequent compounds exportation. FA, fatty acids; FAE, fatty acid elongation, KCS, β-ketoacyl-CoA synthase; VLCFAs, very long-chain fatty acids; FARs, fatty acyl reductases; CYP, cytochrome P450 enzymes; G3P, glycerol-3-phosphate; GPAT, glycerol 3-phosphate acyltransferase; LACS, long-chain acyl-CoA synthetase; ABCG, ATP-binding-cassette G type; PM, plasma membrane; PS, polyester synthase; AHCs, alkyl hydroxycinnamates. ASFT, aliphatic suberin feruloyl transferase; FACT, fatty alcohol:caffeoyl-CoA caffeoyl transferase; MYB, myeloblastosis type transcription factor *Adapted from Vishwanath et al. 2015 and Lashbrooke et al. 2016*.

These findings indicate the involvement of KCS in suberin biosynthesis, particularly in C_{28} and higher chain length precursors elongation, however it is not known the precise mechanism or if, as in *Arabidopsis,* additional factors are required (Pascal et al. 2013; Haslam et al. 2015). In plants, members of the CYP86 subfamily of cytochrome P450 monooxygenases (CYP) are responsible for hydroxylation of the terminal methyl group of aliphatic compounds. In *Arabidopsis,* CYP86A1 catalyses the formation of ω-hydroxi acids from C_{12} and

C_{18} fatty acids constituting an important role in the deposition of suberin in primary endodermis (Höfer et al. 2008). Additionally, *cyp86a1* mutants showed altered root suberin ultrastructure and decrease of around 60% in the aliphatics content of root suberin, mainly due to a significant reduction of C_{16} and C_{18} ω-hydroxy acids and α,ω-dicacids (Höfer et al. 2008). In potato, the silencing of *StCYP86A33*, the *AtCYP86A1* orthologue also reduces the content of ω-hydroxy acids α,ω-dicacids in the tuber periderm indicating an important role of this gene in biosynthesis of suberin monomers (Serra et al. 2009b). In turn, CYP86B1 is required for the formation of very long chain ω-hydroxy acids and α,ω-dicacids from C_{22} to C_{24} acids present in the suberin polyester (Compagnon et al. 2009). Moreover, *Arabidopsis* knockouts and overexpression lines confirmed that CYP86B1 is required not only for the biosynthesis of very-long-chain saturated α,ω-bifunctional aliphatic monomers in suberin but also its involvement in the formation of aliphatic monomers capable of establishing polymeric linkages in the suberin aliphatic polyester (Molina et al. 2009).

Suberin is a glycerol-based polymer where monoacylglycerols may be considered as initial suberin precursors. In addition, glycerol has been reported to be covalently bound to aliphatic and aromatic suberin domains and to be esterified to fatty acid (Graça and Pereira 2000; Graça and Santos 2006). Glycerol esterification assumes great importance in the suberin tree-dimensional network since it allows the formation of a cross-linked polymer, which was not possibly just with the aliphatic monomers of suberin that only can form linear polyesters by themselves (Graça 2015). In this context, acyl transfer reactions are particularly important since they produce monoacylglycerols by acyl-CoA dependent glycerol 3-phopshate acyltransferases (GPATs). GPATs are bifunctional enzymes that catalyse the first reaction in the pathway for the *de novo* synthesis of membranes and storage lipids, catalysing the transfer of an acyl group from acyl-CoA or acyl-ACP to the *sn*-1 or *sn-2* position of glycerol-3-phospate (G3P). In addition, GPATs are involved in the esterification of very long aliphatic compounds in *Arabidopsis* (Beisson et al. 2007). Specifically, *Arabidopsis* mutant of *GPAT5* were reported to present a reduction in C_{20}-C_{24} unsubstituted fatty acids, ω-hydroxy acids and α,ω-dicarboxilic acids in root and seed coat suberin counting for a 50% reduction in the total suberin content (Beisson et al. 2007). *GPAT7* which is close related with *GAPT5* sharing 88% of

similarity and 81% of identity in amino acid sequences, is also involved in the formation of very-long-chain monoacylglycerols from C_{22} and C_{24} free fatty acids, particularly in seed and stem waxes (Yang et al. 2012). This is similar to the phenotype found in GPAT5 ectopic expression suggesting that GPAT7 has similar acyl substrate specificity (Li et al. 2007; Yang et al. 2012). However, the polyester analysis in *gpat7* mutants in various tissues, suggests that GPAT7 may play a specialized role in suberin wounding response, only in specific cell types (Yang et al. 2012).

5.2. The aromatic domain

The aromatic suberin domain is based on phenylpropanoid metabolism which starts with the deamination of phenylalanine by phenylalanine ammonia-lyase (Kolattukudy 2001) followed by a series of hydroxylation, methylation, CoA-ligation and reduction reactions producing the hydroxycinnamic, p-coumaric, caffeic, ferulic and sinapic acids (Vishwanath et al. 2015). These are further conjugated with coenzyme A and reduced to the correspondent alcohols to form the monolignols which in addition of constituting the basic components of lignin are also referred by some authors as important constituents of suberin (Kolattukudy 2001; Bernards 2002). In addition to monolignols, the presence of several amounts of hydroxycinnamic acids and derivatives is a constant when suberin is depolymerized by ester breaking techniques (Graça 2010). Ferulic acid is the hydroxycinnamic acid presented in higher quantities among suberized tissues, and may have a role in the structure of suberized walls. Other phenolic compounds have also been found in suberin, namely tyramine, feruloyltyramine and feruloyloctopamine, however in relative small quantities (Graça 2015). The biosynthesis of the aromatic domain is characterized by the incorporation of ferulic acid in the form of ferulate in the suberin polyester. The *Arabidopsis* aliphatic suberin feruloyl transferase (ASFT) and the potato ortholog named fatty ω-hydroxy acid/fatty alcohol hydroxycinnamoyl transferase (FHT) were reported to catalyse acyl transfer of feruloyl-CoA to ω- hydroxy and fatty alcohols (Gou et al. 2009; Molina et al. 2009; Serra et al. 2010). While *asft* mutants have a complete lack of ferulate in root suberin and a decrease in ω- hydroxyl α,ω-dicarboxilic acids (Gou et al. 2009; Molina et al. 2009), silencing of *FHT* by RNA

interference (RNAi) leave to a strong reduction of ester-linked ferulic acids in tuber periderm (Serra et al. 2010). Similarly, fatty alcohol:caffeoyl-CoA caffeoyl transferases are responsible for the production of a subset of alkyl hydroxycinnamate esters, the alkyl caffeates in root waxes. Since root waxes are composed by saturated long-chain primary alcohols and hydroxycinnamic acids also present in suberin it is plausible that an identical reaction occurs in suberin biosynthesis.

5.3. Transportation and assembly of suberin monomeric units

In general, cell wall polymers are assembled by monomeric compounds exported through the plasma membrane by a vesicular pathway involving Golgi-derived vesicles from the endoplasmic reticulum (Pollard et al. 2008). However, some other mechanisms have been reported as direct membrane exportation involving ATP-binding cassette (ABC) transporters and possibly lipid transfer proteins (LTPs) (Vishwanath et al. 2015). Plasma membrane-localized ABC transporters of the G-subfamily (ABCG) are likely to be involved in suberin monomers transport since they are responsible for the exportation of cuticular and sporopollenin precursors, also present in suberin (Kang et al. 2011). In fact, ABCG2, ABCG6 and ABCG20 are half-transporters required for synthesis of an effective suberin barrier in *Arabidopsis* roots and seed coats. Moreover, the suberin of *abcg2 abcg6 abcg20* triple mutant roots and seed coats presents a distorted lamellar structure, reduction in aliphatic components as well a deficiency in alkyl hydroxycinnamate esters of root wax (Yadav et al. 2014), confirming the importance of ABCG transporters in suberin precursors translocation and deposition. Other members of the ABCG subfamily have also been reported as involved in the biosynthesis of lipid compounds and translocation across the plasma membrane. In *Arabidopsis*, ABCG11 have been associated with cutin and suberin biosynthesis (Bird et al. 2007; Panikashvili et al. 2007; Panikashvili et al. 2010), while ABCG12 (Pighin et al. 2004) and ABCG13 (Panikashvili et al. 2011) are required for the export of cuticular wax monomers.

It has been suggested that non-specific lipid transfer proteins (LTPs) could also be involved in suberin synthesis and accumulation (Edstam et al. 2013). Expression and co-expression analysis of LPTs from *Arabidopsis* and *Oryza*

sativa showed significant co-expression with genes known to be involved in suberin biosynthesis. It was the case of *AtLTPg3*, *AtLTPg4*, *AtLTPg23* and *AtLTPg26* all belonging to an expression cluster of genes correlated with suberin biosynthesis and *OsLTP1g1*, *OsLTPg2* and *OsLTPg24* that were pointed to have putative roles in suberin biosynthesis in roots (Edstam et al. 2013; Salminen et al. 2016). The loss-of-function mutant *atltpI-4* and ectopic expression of *AtLTPI-4* revealed an impact of AtLTPI-4 on extracellular lipid deposition of epidermal cells (Deeken et al. 2016). In crown gall development, AtLTPI-4 is involved in suberin formation possibly by transferring long-chain and very-long-chain fatty acids into the apoplast (Deeken et al. 2016). Thus, LTPs may have an important role in the trafficking of suberin precursors to polymerization in cell walls, however future studies are required to clarify the precise mechanism.

Suberin assembly may occur after precursors transportation throughout the plasma membrane as observed in cutin, since both biopolymers are characterized by a lipid matrix deposited on a polysaccharide scaffold with a phenolic domain of variable proportions (Domínguez et al. 2015). The first cutin synthase (CD1/GDSL1) was identified in tomato fruit cuticle by two independent groups not long ago (Girard et al. 2012; Yeats et al. 2012). CD1/GDSL is an extracellular protein of the Gly-Asp-Ser-Leu domain family of esterases/acylhydrolases, commonly known as GDSL-lipases. Both groups reported similar conclusions, namely that the GSDL protein is extracellular and is located in the cutinized region of the outer epidermal cell wall. In the study conducted by Yeats *et al.* (2012), the fruit of tomato null mutant, *cutin deficient 1 (cd1)*, was characterized by a severe decrease in levels of polymeric cutin (>95%) and accumulation of the cutin monomer 2-mono(10,16-dihydroxyhexadecanoyl) glycerol (2-MHG). In the sequence of this study, it was found that *in vitro* recombinant CD1 uses 2-MHG as substrate to produce linear cutin oligomers composed by terminal ester-linked fatty acids suggesting that cutin polymer branching and cross-linking may require additional enzymes (Yeats et al. 2014). The authors also reported *Arabidopsis thaliana* and *Physcomitrella patents* orthologs of CD1 and demonstrated that homologous genes encoding for the cutin synthase (CUS) constitute an ancient and conserved protein family among plants (Yeats et al. 2014). In the study of Girard *et al.* (2012), the suppression of *CD1/GDSL1* by RNAi lead to cuticle thickness and cuticle content reduction with

appearance of nanopores in the cutin matrix. Although one of the enzymes responsible for cutin polymerization has been identified, much more is to be discovered to fully understand the mechanistic details and other aspects of cutin polymerization. In addition to CUS proteins of the GDSL superfamily, BODYGUARD (BDG), an α/β-hydrolase family protein and DEFECTIVE IN CUTICULAR RIDGES (DCR), a BAHD acyltransferase have been hypothesized to be involved in cutin polymeratizon, however clear biochemical support is lacking (Kurdyukov et al. 2006; Panikashvili et al. 2009). In suberin polymer synthesis, suberin synthase (SUS) proteins may perform an analogous role using monoacylglycerols and hydroxyl-alkyl ferulates as substrates as the involvement of putative GDSL orthologous in suberin formation has previously been reported through the identification of candidate genes related with suberin deposition in cork and potato (Soler et al. 2007; Serra et al. 2009a; Serra et al. 2010), although the molecular mechanisms behind suberin assembly still remains to be elucidated.

5.4. Regulation of suberin biosynthesis

Suberin occurrence is limited to specific cell types and its deposition can be a response to both biotic and abiotic stresses. As described above, the expression of genes involved in suberin biosynthesis closely match the location of these cell types indicating a strict and tight control of gene transcription. *Arabidopsis* MYB41 was reported to ectopically induce aliphatic suberin synthesis and deposition of cell wall-associated suberin-like lamellae (Kosma et al. 2014). When overexpressed in leaves of *Arabidopsis* or transiently overexpressed in leaves of *Nicotiana benthamiana, AtMYB41* increases the abundance of suberin biosynthetic gene transcripts, what results in the accumulation of suberin-type aliphatic monomers (Kosma et al. 2014). In addition, the overexpression of *AtMYB41* resulted in relatively high amounts of monolignols and increased levels of phenylpropanoid and lignin biosynthetic transcripts (Kosma et al. 2014). It was also found that *AtMYB41* gene is only expressed in roots under abiotic stress conditions, such as high salt concentrations. Recently, two MYB genes were reported to be involved in the regulation of suberin biosynthesis. *AtMYB107* and At*MYB9* were identified in an

attempt to discover genes associated with the developmental control of suberin deposition through the investigation of the chemical composition and transcriptomes of suberized tomato (*Solanum lycopersicum*) and russet apple (*Malus x domestica*) fruit surface. By comparative gene co-expression analysis of apple (russet clone), tomato (SlDCR-RNAi line) fruit, *Arabidopsis* seed tissues, major organs of grapevine, potato, and in roots of rice, a total of 26 genes were identified as having a 'suberin gene expression signature' (Lashbrooke et al. 2016). A close examination of orthologous genes identified in the co-expression datasets, revealed a clade of MYB transcription factors containing MYB9 and MYB107, and a clade of GDSL-motif esterases. The latter is phylogenetically close related with the CUS clade (Yeats et al. 2012; Yeats et al. 2014). The authors proposed that these proteins may correspond to the SUS proteins referred above which may constitute an important step in the elucidation of the suberin assembly molecular mechanism. The two central elements in this gene signature - *AtMYB107* and *AtMYB9* - were further characterized (Lashbrooke et al. 2016). Seeds of *myb107* and *myb9 Arabidopsis* mutants had lower levels of aliphatic and aromatic suberin constituents, while seeds from the *myb*107 line exhibit decreased germination rates under osmotic and salt stress (Lashbrooke et al. 2016). *AtMYB107* was 2-fold up-regulated in *myb9* lines, possibly as a compensation of the *AtMYB9* down-regulation (Lashbrooke et al. 2016). These results indicate that the two transcription factors may exert their regulatory action redundantly and that MYB9 regulates negatively MYB107. From a chemical analysis of the seed coat polyester, gene expression data from seeds, as well as seed permeability assays, the authors found evidences that suggest that MYB9 and MYB107 act in tandem. Both TFs positively regulate various aspects of suberin formation, namely the fatty acid and the phenylpropanoid biosynthesis, the production of ferulate compounds, and extracellular suberin monomer transport and polymerization (Figure I.8). The hypothesis explores the MYB107 transcription regulation of dihydroflavonol 4-reducatse-like1 (DFR-like1), 4-Coumarate:CoA ligase 5 (4CL5), KCS17, GPAT5, ASFT, LTPG5, and SUS, and the transcription regulation of DFR-like1, CYP86A1 and SUS by MYB9. Recently, Gou *et al.* (2017) confirmed that the *Arabidopsis* MYB107 acts as a positive regulator of suberin biosynthetic gene expression in seed coat. *AtMYB107* is co-expressed in a temporal manner during seed development with

Hydroxycinnamoyl-CoA: ω-Hydroxyacid *O*-Hydroxycinnamoyl Transferase (HHT), FACT, Defective in Cutin Ferulate (DCF), GPATs, FARs and CYP86 genes (Gou et al. 2017). The authors conclude that the down-regulation of *MYB107* suppresses the expression of most known suberin biosynthetic genes, namely those that encode for enzymes related to synthesis and assembly of suberin aliphatic and aromatic domains in *Arabidopsis* seed coat.

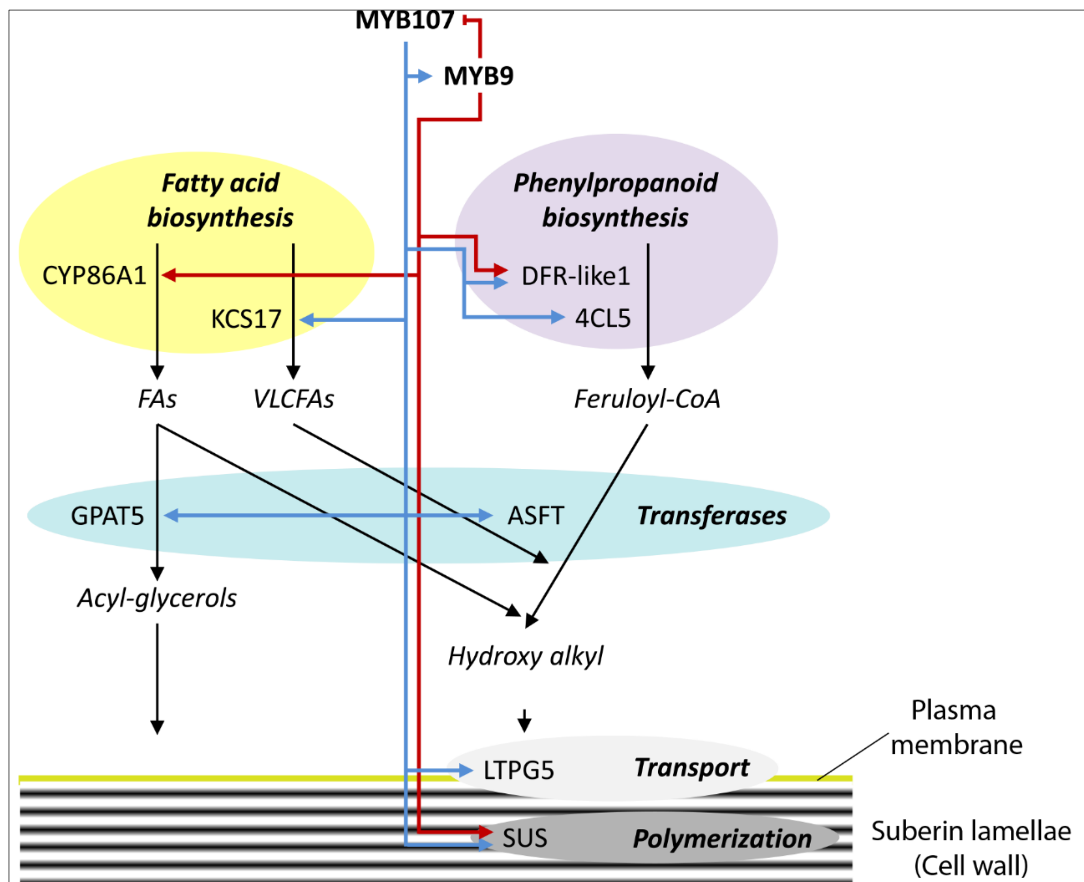


Figure I.8 I A model for suberin formation in the Arabidopsis seed coat and a putative regulatory mechanism in suberin formation by MYB9 and MYB107. MYB9 and MYB107 seem to regulate positively various aspects of suberin formation, particularly fatty acid and phenylpropanoid biosynthesis, as well the transportation of suberin monomers and its polymerization. Potential gene targets of MYB9 and MYB107 are represented with red and blue arrows, respectively. FAS, fatty acids; VLCFAs, very long chain fatty acids; GPAT5, GLYCEROL-3-PHOSPHATE ACYLTRANSFERASE5; ASFT, ALIPHATIC SUBERIN FERULOYL TRANSFERASE; KCS17, 3-KETOACYL-COA SYNTHASE17; 4CL5, 4-COUMARATE:COA LIGASE5; DFR-like1, DIHYDROFLAVONOL 4-REDUCTASE-LIKE; LTPG5, GLYCOSYLPHOSPHATIDYLINOSITOL-ANCHORED LIPID PROTEIN TRANSFER5; SUS, SUBERIN SYNTHASE (*Adapted from Lashbrooke et al., 2016*)

The suberization of *myb107* mutant seed coat is impaired with a drastic reduction of suberin aliphatic and aromatic monomers, and an alteration in suberin lamellae disposition resulting in enhanced seed coat permeability and increased abiotic

stress susceptibility of *myb107* seeds. Furthermore, the complementation with *MYB107* expression restores suberin biosynthetic gene expression, abnormal suberin synthesis seed coat permeability, and abiotic stress tolerance. The authors conclude that MYB107 directly targets the promoters of some of the studied biosynthetic genes, but not all, what may imply that MYB107 regulates suberin biosynthetic genes both directly and indirectly. It is plausible that MYB107 have a mode of action similar to TFs involved in cutin synthesis as the wax Inducer 1/Shine 1 (WIN1/SHN1) transcription factor that regulates the expression of a set of cutin and/or wax biosynthesis and directly activate cutin biosynthetic genes by targeting their promotor regions (Gou et al. 2017). Furthermore, *WIN/SHN1* is regulated by MYB106 and MYB16, two TFs involved in cuticle development constituting an hierarchical regulatory network for cutin formation (Yeats and Rose 2013). Similarly, in secondary cell wall formation, the NAC-type TFs , VASCULAR-RELATED NAC DOMAIN TRANSCRIPTION FACTOR 6/7 and SECONDARY WALL THICKENING PROMOTING FACTOR 1-3 act as master activators or repressors in the entire secondary cell wall program in *Arabidopsis*, and are able to up-regulate directly the lignin master regulator *AtMYB46*, suggesting that they also regulate the lignin pathway (Zhao and Dixon 2011). These evidences indicate that transcription factors are able to co-ordinately regulate multiple pathways involved in a particular cellular process, such as the secondary wall formation; a similar process may happen in suberin biosynthesis.

6. QsMYB1 - The R2R3-MYB transcription factor involved in suberin biosynthesis

QsMYB1 is a particular interesting TF involved in suberin biosynthesis and consequently in cork formation since it was first identified together with a group of candidate genes involved in cork differentiation and development (Soler et al. 2007). This TF was further characterized by Almeida and collaborators (Almeida et al. 2013a) who reported that the *QsMYB1* was up-regulated in cork producing tissues and suggested an involvement in cork formation. In this work, the results also indicated that the *QsMYB1* gene was alternatively spliced into two mRNA variants (*QsMYB1.1 and QsMYB1.2*) differing in length and in sequence

composition by retaining an intron at the 5'-untranslated region (UTR). Studies in *Arabidopsis* and rice have showed that the regulatory effect associated with alternative splicing variants is controlled in a developmental, tissue or subcellular specific manner (Koo et al. 2007; Zhang and Mount 2009). Other studies have also showed that the occurrence of an intron within 5' or 3'-UTRs may result in a regulatory function influencing mRNA stability, transport and translation (Parida et al. 2009). Interestingly, Almeida and collaborators (2013a) also detected the presence of a single sequence repeat (SSR) within the retained intron which led to the hypothesis that the presence of the SSR may act as a *cis*-element in the regulation of *QsMYB1* expression. The authors also conclude that *QsMYB1* is mainly expressed in tissues and organs with secondary growth, resulting from the activity of the phellogen (Almeida et al. 2013a). Interestingly, the relative transcripts abundance is higher in the phellem from the highly suberized cork than in wood which is a lignified tissue, what is in accordance with the results reported by Soler et al. (2007). In *Arabidopsis*, Feng et al. (2004) had previously found that the putative orthologous of *QsMYB1*, the *AtMYB68* is specifically expressed in root pericycle cells, at the side of the secondary root formation but not in endodermis. Moreover, lignin levels in root cultures of *myb68* mutant were increased when compared with the wild type, suggesting that the *AtMYB68* may repress directly or indirectly some aspects of lignin production (Feng et al. 2004). A similar process may occur with *QsMYB1*. *In silico* analysis of the *QsMYB1* promoter region showed the presence of putative *cis*-acting regulatory elements related to phenylpropanoid pathway, which includes several essential reactions of suberin phenolic domain synthesis as well as of the lignin synthesis (Almeida et al. 2013a). According to the results, it seems that *QsMYB1* may be involved in the regulation of metabolic pathways related with cork formation, however this hypothesis requires validation through functional studies.

7. IDENTIFYING TRANSCRIPTION FACTORS BINDING SITES

Transcription factors operate individually or in complexes with other molecules to activate or repress the recruitment of the transcriptional machinery to specific genes, and by this way modulating mRNA transcription. TFs are crucial for plant development, response to environment, reproduction, cell cycle,

intercellular signalling and metabolism. For molecular biologists it is a challenge to understand the mechanistic regulation of TFs, considering that TFs may bind several *cis* regulatory elements across a genome with different affinities and often function together with other regulatory elements (Jothi et al. 2008).

In recent years, significant advancements in technologies have enabled the study of TFs (Landt et al. 2012). Chromatin immunoprecipitation (ChIP) is a method to identify genomic *loci* that physically interact with a protein of interest. The protein may be a TF, a histone modifier or other chromatin-associated protein (Kaufmann et al. 2010). In a ChIP experiment, protein-DNA complexes are purified using a specific antibody against the DNA-interacting protein or against a tag fused protein. Specific antibodies for chromatin modifications (e.g. histone acetylation, histone or DNA methylation) can also be used to study the epigenetic status of a genome. A ChIP assay consists of five steps: (i) protein crosslink with the DNA, (ii) chromatin isolation, (iii) chromatin fragmentation, (iv) immunoprecipitation, (v) recovery and purification of the fragmented DNA and the identification of the DNA sequences (Figure I.9) (Yamaguchi et al. 2014). The outcome of a ChIP experiment consists of purified DNA fragments where the protein of interest was bound. These fragments can be analysed by quantitative real-time PCR (ChIP-qPCR), hybridisation to genome arrays (ChIP-CHIP) or more recently using Next Generation Sequencing (ChIP-Seq), which allows the identification of all direct targets at the genome level by a bioinformatics approach (Yamaguchi et al. 2014). The ChIP methodology in plant sciences is a laborious work, partly due to technical difficulties in ChIP sample preparation related to the complex composition of plant tissues, associated with the presence of components of the cell walls, such as polysaccharides and polyphenolic compounds, and other derived-components from the secondary metabolism only present in plant species. Recently, Li et al. (2014), developed a ChIP protocol for wood-forming tissues in the model woody plant *Populus trichocarpa* (Li et al. 2014). The authors identified all the steps that were ineffective in standard ChIP protocols and developed for the first time an optimized and robust protocol for woody and recalcitrant tissues to standard ChIP procedures.

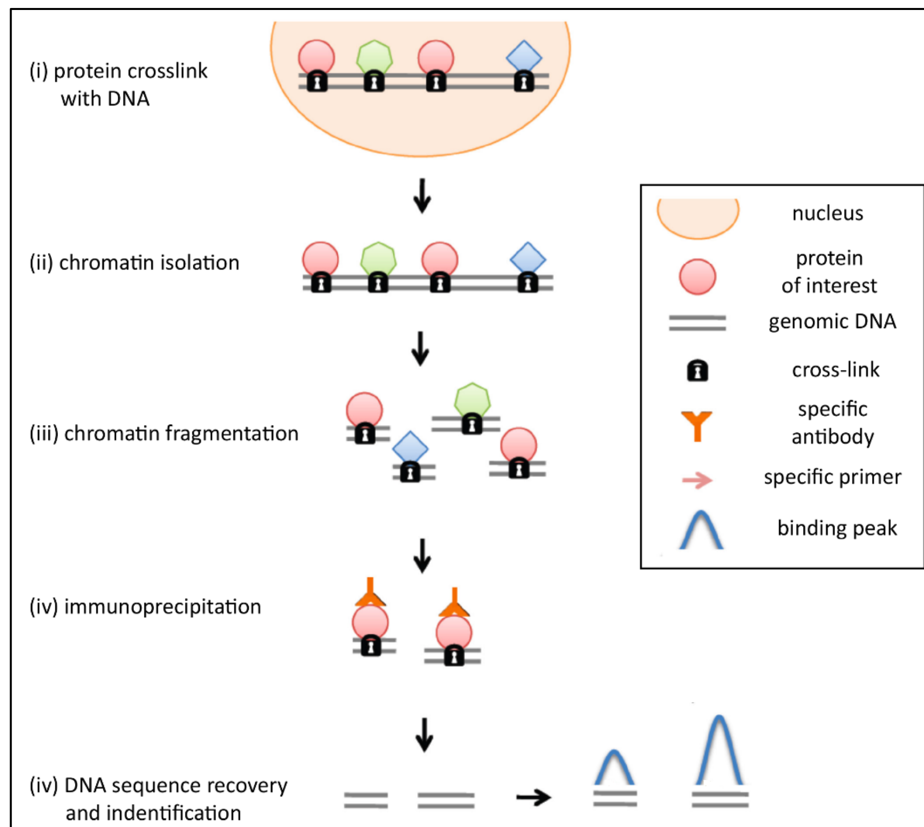


Figure I.9 | Generic ChIP procedure outline (*Adapted from Yamaguchi et al. 2014*)

8. CHIP-SEQ AS A ROBUST TECHNIQUE TO DISCOVER PROTEIN-DNA INTERACTIONS IN PLANTS

High-throughput low-cost sequencing was the main driver behind the development of the ChIP-Seq technique (Landt et al. 2012). ChIP-Seq uses massive parallel sequencing to sequence DNA fragments representatives of the target of a protein in the case of TFs. These fragments are then mapped against a reference genome allowing the identification of the TF on a genome-wide scope. The large genome coverage, the high sensitivity to the DNA targets and the site resolution are the main advantages of ChIP-Seq over similar techniques for generate data to map a TF binding site to a species genome (Arrigoni et al. 2015). The experimental design in ChIP-Seq involves antibody validation to ensure high specificity to the TF, the selection of a correct number of biological replicates and the appropriate sequencing technology. The number of reads to be generated is a crucial step since it ensures the generation of reproducible and meaningful data (Landt et al. 2012). The ENCODE consortium (Landt et al. 2012) recommends a minimum of 10 million reads for each of the two independent

replicates and two types of controls because of the non-random nature of fragmentation by sonication. The “input” control is cross-linked and fragmented but not immuno-precipitated (IP) while the “IgG” control uses an antibody, which binds irrelevant proteins to generate random IP DNA (Li et al. 2014). Both controls are important, although not essential to generate high quality data.

The generated reads are then mapped against the reference genome using a read aligner (Langmead and Salzberg 2012). After the reads are mapped to the genome, the TF binding sites can be identified by a “peak calling” software. There are more than 30 peak-calling software tools developed for ChIP-Seq data (Landt et al. 2012). Peak finders locate regions of enrichment based on the number or significance of enrichment and in most of them the TF binding sites DNA specific sequence are centred on the discovered peaks (Wilbanks and Facciotti 2010; Whitington et al. 2011).

ChIP-Seq is a widely accepted technique by researchers for genome-wide identification of TF binding sites and analysis of TF specificity. It has several advantages over ChIP-CHIP (Table I. 1), the preferred technique to study TF DNA binding before ChIP-Seq emerged. This was mainly because of the fast development of low-cost high throughput sequencing technologies. In addition, a ChIP-Seq experiment requires a small amount of sample, can map all TF binding sites in the genome and may be used on any sequenced genome version. However, as with any other technology, ChIP-Seq has several disadvantages (Table I. 1) (Park 2009). In the last years, the main disadvantage has been the cost, compared with ChIP-CHIP. However, with the decreasing cost of next generation sequencing this issue tends to be overcome. All the profiling technologies produce artefacts that may be considered in order to reduce errors. ChIP-Seq is not an exception and besides sequencing errors have been reduces substantially with the sequencing technology improving this kind of error is still present in the generated data, especially at the end of each read. There is also bias towards GC-rich content in fragment selection, both in library preparation and in amplification before and during sequencing, although notable improvements have been made recently (Li et al. 2015). In addition, when an insufficient number of reads is generated, there is loss of sensitivity or specificity in detection of enriched regions. There are also technical issues in performing the experiment, such as loading the correct amount of sample: not enough sample

will result in too few tags; too much sample will result in fluorescent labels that are too close to one another, and therefore lower quality data (Park 2009).

Table I. 1 | Comparsion between ChIP-Chip and ChIP-Seq (*Adaptated from Park 2009*)

	ChIP-Chip	**ChIP-Seq**
Resolution	Array-specific (30-100bp)	Single nucleotide
Coverage	Limited by the hybridizated sequences on the array; repetitive regions are usually masked out	Limited only by alignability of reads to the genome; increases with read length; many repetitive regions can be covered
Cost	lower	higher
Source platform noise	Cross-hybridization between probes and nonspecific targets	Some GC bias can be present
Required amount of ChIP DNA	High (micrograms)	Low (nanograms)
Dynamic range	Lower detection limit; saturation at high signal	Not limited
Amplification	Usually a requirement	Less required; single-molecule sequencing without amplification is a possibility
Multiplexing	Not applicable	Possible

9. MOTIF DISCOVERY FROM CHIP-SEQ DATA

The first step to determine TF binding specificity is the motif identification. A motif may be defined as a sequence pattern repeat in related proteins or DNA sequences with a particular function. A motif may be simply defined as a pattern of DNA sequences representing TF binding sites. In order to discover motifs in genomic data, there are exhaustive approaches, which searches for all words in a sequence, requiring large amounts of data processing (Pavesi and Pesole 2006). Alternatively, probabilistic techniques that use a model of the sequence data to search for a motif that are most likely to reproduce the observed results are commonly used. The probabilistic methods are based on the Expectation Maximization as in Multiple Expectation Maximization for Motif Elicitation (MEME) algorithm (Bailey et al. 2006), and stochastic approaches based on Gibbs sampling applied in motif discovery tools. Gibbs sampling generates a Markov chain Monte Carlo of samples, each of which is correlated with nearby samples

allowing Bayesian inference to obtain a sample from the posterior distribution. This allows the mapping of ChIP-Seq ambiguous tags and the determination of the probabilities used to infer the correct tag map positions (Wang et al. 2010). These techniques have been improved over the last decade to suit new data sets generated from new platforms and a major shift to the use of ChIP-based techniques led to the development of ChIP-MEME (Machanick and Bailey 2011). A number of consensus-based methods have also been designed for large-scale ChIP-Seq analysis (e.g. DREME (Bailey 2011), CisFinder (Sharov and Ko 2009), cERMIT (Georgiev et al. 2010), RSAT Peak-motifs (Medina-Rivera et al. 2015), and MCES (Yu et al. 2014)). When new motifs are identified, motif comparison tools such as TOMTOM (Gupta et al. 2007) can be used to compare with already identified and annotated motifs in databases such as JASPAR (Mathelier et al. 2015), TRANSFAC (Wingender et al. 2000), DAP motifs (O'Malley et al. 2017) or PBM motifs (Franco-Zorrilla et al. 2014). TOMTOM compares an input motif to those in motif databases and returns the most similar based on statistical motif-motif similarity, identifying a similarity with other already described motifs (Gupta et al. 2007).

10. SOMATIC EMBRYOGENESIS IN PLANTS

In flowering plants, the process of double fertilization involves a haploid sperm which fertilizes a haploid egg cell to form a diploid zygote. The zygote follows a series of morphological, biochemical, and molecular events in order to form an embryo. This stage of development is denominated as zygotic embryogenesis (Goldberg et al. 1994). Somatic embryogenesis (SE) occurs when somatic cells dedifferentiate to totipotent embryogenic stem cells (Verdeil et al. 2007; Ikeuchi et al. 2015). Since the first description of somatic embryogenesis in carrot (*Daucus carota*) cell cultures (Reinert 1958; Steward et al. 1958), this process has been reported in various woody plant genera as *Betula*, *Eucalyptus*, *Pinus*, *Populus*, *Pseudotsuga* and *Quercus*, among others. SE can be induced with the appropriate conditions recurring to hormone regulators. In contrast to primary SE induced from explant cells, secondary somatic embryogenesis is the phenomenon whereby new somatic embryos are induced through existing somatic embryos. In woody plants, secondary

embryogenesis is able to maintain the embryogenic competence for several years, providing a useful research material and a starting material for *in vitro* propagation of woody plants (Guan et al. 2016).

10.1 Regulatory genes in somatic embryogenesis

In *Arabidopsis*, four main types of TFs directly involved in SE positive regulation have been described. The first, include the *LEAFY COTYLEDON (LEC)* genes *LEC1* and *LEC2*, which were identified by loss of function mutants phenotypes in embryo identity and seed maturation (Lotan et al. 1998; Braybrook and Harada 2008). *AtLEC1* encodes the HEME-ACTIVATED PROTEINS (HAP3) subunit of the CCAAT box-binding TF, and ectopic *AtLEC1* expression is sufficient to induce SE from vegetative cells (Lotan et al. 1998; Gaj et al. 2005). *AtLEC1* acts as a regulator of cell fate determination and integrates hormone and light signalling pathways in somatic and zygotic embryogenesis (Braybrook and Harada 2008; Junker et al. 2012; Radoeva and Weijers 2014; Huang et al. 2015a; Huang et al. 2015b). *AtLEC2* functions as regulator of embryogenesis by repressing expression of the *GIBBERELLIN 3-BETA-DIOXYGENASE 2* (*AtGA3ox2*) gene and promoting the auxin pathway by up-regulating the auxin biosynthesis genes *INDOLE-3-PYRUVATE MONOOXYGENASE 2* (*AtYUCCA2*) and *AtYUCCA4*, and the auxin signalling gene *INDOLE-3-ACETIC ACID INDUCIBLE 30* (*AtIAA30)* (Stone et al. 2001; Curaba et al. 2004; Braybrook et al. 2006). In turn, *AtLEC2* together with *AtFUSCA3* when ectopically expressed resulted in the exhibition of embryonic traits in adult tissues (Luerssen et al. 1998; Stone et al. 2001). Moreover, *AtFUSCA3* negatively regulates gibberellic acid (GA) accumulation by repressing GA biosynthesis genes *AtGA3ox1* and *AtGA3ox2* (Curaba et al. 2004; Gazzarrini et al. 2004) which, in turn, promotes embryogenesis.

The second type of TFs involved in somatic embryogenesis includes *AGAMOUS-like 15* (*AtAGL15*), which encodes a MADS-box TF. The ectopic overexpression of *AtAGL15* enhances the somatic embryo initiation from shoot apical meristems and is directly regulated by the AtLEC2 protein (Heck et al. 1995; Harding et al. 2003). The expression of *AtAGL15* reduces GA levels by promoting *AtGA2ox6* expression which in turn inactivates GA (Wang et al. 2004).

In parallel, *AtAGL15* upregulates the expression of *AtIAA30*, which is also involved in the promotion of somatic embryo development (Zheng et al. 2009).

The third type of TFs that positively regulate somatic embryogenesis includes *BABY BOOM* (*AtBBM*) and *EMBRYOMAKER* (*AtEMK*), both APETALA2/Ethylene Responsive element transcription factors. The ectopic expression of *AtBBM* is able to induce the formation of somatic embryos from leafs (Boutilier et al. 2002). It was recently reported that BBM activates the LEC1-FUSCA3-LEC2 network to induce somatic embryogenesis (Horstman et al. 2017). *EMBRYOMAKER* (*AtEMK*), which is expressed in developing and mature embryos, promotes the formation of embryo-like structures from cotyledons when ectopically overexpressed (Tsuwamoto et al. 2010).

The last group includes genes from the *WUSCHELL-related homeobox* (*WOX*) gene family which is able to promote somatic embryogenesis by activating *AtLEC* genes (Wang et al. 2009a). In fact *AtWOX* has been described with specialized functions in various developmental processes, such as embryogenic patterning and stem cell maintenance (Gallois et al. 2004; Su et al. 2009).

Besides the well described positive regulatory role of the referred genes in embryogenesis, there are also other described genes involved in the process. It is the case of R2R3-type MYB transcription factors *AtMYB118* and *AtMYB115* which positively regulate *AtLEC1* expression inducing the formation of somatic embryos (Wang et al. 2009a). Also the *SOMATIC EMBRYOGENESIS RECEPTOR KINASE1* (*AtSERK1*) which encodes a leucine-rich repeat transmenbranar receptor-like kinase seems to be involved in the signalling pathway responsible for somatic cell initiation (Hecht et al. 2001). Furthermore AtSERK1 forms a protein complex with AtAGL15 (Karlova et al. 2006) and when overexpressed increases the efficiency of somatic cell initiation in Arabidopsis (Hecht et al. 2001).

Several repressors of SE have also been identified. In *Arabidopsis, PICKLE* (*AtPKL*) which encodes a chromodomain/helicase/DNA binding domain (CHD3) chromatin-remodelling factor act as a negative regulator of embryogenic identity by repressing the expression of *AtLEC* genes (Ogas et al. 1999; Rider et al. 2003). Moreover, *AtPKL* regulates epigenetically SE by modulating H3K37me3 levels (Zhang et al. 2012). Other epigenetic regulators as the polycomb repressive *complex (PRC) 1* and PRC2 are also involved in the

epigenetic regulation of SE in *Arabidopsis*. It is known that both PRC1 and PRC2 play important roles in the transition from the embryogenic to post-embryogenic stage by the suppression of the *AtLEC* genes expression (Chen et al. 2010; Bouyer et al. 2011). Recently, it was shown that AtPRC2 inhibits the dedifferentiation of mature somatic cells in *Arabidopsis* roots and that the differentiated cells are also able to dedifferentiate and produce somatic embryos once AtPRC2 epigenetic repression is removed (Ikeuchi et al. 2015).

After the identification of key genes involved in somatic embryogenesis in *Arabidopsis*, several studies have emerged in order to identify homologous genes in other plant species. In somatic embryos of European larch (*Larix decidua*), several homologous genes of *AtBBM*, *AtLEC1*, *AtWOX2* and *AtSERK* were identified. *LdLEC1* and *LdWOX2* are mainly expressed during early embryogenesis, while *LdBBM* and *LdSERK* are preferentially expressed later in development (Rupps et al. 2016). The overexpression of *LdLEC1* in *Arabidopsis* affects germination and cotyledon formation and produces a similar phenotype as in *AtLEC1* overexpression lines, leading to the formation of embryo-like structures (Lotan et al. 1998).

In *Theobroma cacao BBM*, *LEC1-like (L1L), LEC2 and SERK* were also characterized (De Oliveira Santos et al. 2005; Alemanno et al. 2008; Zhang et al. 2014; Florez et al. 2015). *TcSERK* has an increased expression in induced embryogenic *callus* (De Oliveira Santos et al. 2005), while *TcL1L* has an increased expression in young and immature zygotic or somatic embryos. Moreover, the ectopic expression of *TcL1L* in *Arabidopsis* is able to partially rescue the *Atlec1* mutant lines (Alemanno et al. 2008). Also the overexpression of *TcBBM* in *Arabidopsis* and in *T. cacao* led to the formation of somatic embryos without requiring the use of exogenous hormones (Florez et al. 2015). In addition, the overexpression of *TcLEC2* in cacao explants increases the frequency of regeneration of transformed somatic embryos (Zhang et al. 2014). In parallel, the overexpression of the *Citrus sinensis L1L* gene (*CsL1L*) led to similar results suggesting that these genes may have the capability to transit cells from vegetative to embryogenic phases (Zhu et al. 2014).

The epigenetic regulation is also an important aspect during the somatic embryogenesis in woody plants. Recently it was proposed that a crosstalk between DNA methylation and histone modifications may occur during the

earliest embryogenic stages in *Coffea canephora* by H3K27m3 action on *CcLEC1* and *CcBBM1* (Nic-Can et al. 2013). In *Quercus suber*, *QsPKL* and the *VP1/ABSCISIC ACID INSENSITIVE3-LIKE 1* (*VAL1*) may be necessary for the correct development of somatic embryogenesis (Pérez et al. 2015).

Although significant progress in characterizing the molecular mechanism involved in SE have been made, most of the studies of SE in woody plants mainly rely on physiological aspects, such as selection of culture media, explants and hormones (Isah 2016). Little is known about the molecular mechanisms behind the most critical aspect in SE, the cell embryogenic competence transition. Further research genomics, epigenomics and transcriptomics may address the answers on somatic embryogenesis in woody plants.

10.2 Somatic embryogenesis in *Quercus suber*

Somatic embryogenesis has the potential to generate large number of plants by *in vitro* propagation and has become achievable for many tree species. It may be defined as a non-sexual developmental process which produces a bipolar structure (embryo) from a somatic tissue (Namasivayam 2007). The somatic embryos present similar developmental stages, typically passing through globular, torpedo, and cotyledonary stages in dicots, or globular, scutellar, and coleoptillar stages for monocots (Mordhorst et al. 1997). In woody plants, SE constitutes a powerful tool for improvement of forest trees as it is considered one of the most effective process of *in vitro* regeneration (Vieitez et al. 2012). Also, large-scale propagation of selected biological material, genetic transformation, and cryopreservation are achievable by application of SE in several species of the *Quercus* family, namely in *Quercus robur* (Zegzouti et al. 2001; Sunderlikova and Wilhelm 2002; Concepción Sánchez et al. 2003; Toribio et al. 2004; Prewein et al. 2004; Martínez et al. 2008; Sanchez et al. 2008), *Quercus rubra* (Vengadesan and Pijut 2009)*, Quercus acutissima* (Yong-Wook 2000)*, Quercus ilex* (Mauri and Manzanera 2004) and *Quercus suber* (García-Martín 2001; Pinto et al. 2002b; Bueno et al. 2003; Valladares et al. 2004; García-Martín et al. 2005; Toribio et al. 2005; Álvarez and Ordás 2007; Pintos et al. 2010; Hernández et al. 2011).

In the last two decades, several studies have reported somatic embryo induction and development in *Quercus suber*. Hernandez et al. (2001), reported for the first time the induction of SE in mature oak trees, from leaves of epicormic shoots (Hernandéz et al. 2001). Pinto et al. (2002) obtained embryogenic *calli* in leaves from a 60 years old cork tree (Pinto et al., 2002). Further studies were carried out for a improved characterization of the induction of SE in cork oak, originating consistent embryogenic systems for the propagation of several trees with more than 100 years old (Hernández et al. 2003). It is however clear that induction of SE in *Quercus suber* is genotype dependent as in other woody plants (Isah 2016) and optimal induction treatments may differ from genotype to genotype (Hernandéz et al. 2003). Recently, Pérez *et al.* (2013) reported a temporary immersion system for large cork oak somatic embryogenic culture proliferation and somatic embryo production with increased efficiency (Pérez et al. 2013).

Once embryo initiation is achieved, a large number of somatic embryos can be obtained by secondary embryogenesis. After the isolation of a somatic embryo from the original explant, a recurrent embryogenesis mechanism can be triggered if the embryo are cultured in an appropriate proliferation medium, originating clonal embryogenic lines (Vieitez et al. 2012). Typically a semisolid proliferation medium supplemented with either a low concentration of cytokinins and auxins, or a medium without plant growth regulators are used for the multiplication of the embryogenic cell lines, however a genotypic effect was also documented in the proliferation of cork oak embryogenic lines (Hernandez et al., 2003; Pinto et al., 2002).

10.2.1 Embryos maturation, germination and plantlet conversion

The rate of somatic embryo conversion to plantlets is low in cork oak, which limits the application of embryogenesis at a large-scale. This is mainly because embryo maturation is hampered by repetitive embryogenesis, embryo dormancy, and precocious germination (Vieitez et al. 2012). Somatic embryos should accumulate specific storage metabolites during development, as occurs in zygotic embryos. Hormones such as abscisic acid (ABA) play a fundamental role in embryo maturation, regulating the synthesis of storage proteins and late

embryogenesis-abundant proteins as well as reducing the secondary embryogenesis rate and inhibiting the precocious germination (Von Arnold et al. 2002; Arc et al. 2013). Several satisfactory attempts were made to increase the rate of conversion in several *Quercus* spp., using different types and concentration of carbohydrates in the culture medium (Vieitez et al. 2012). Although somatic embryos of cork oak respond better to conversion with a presence of high concentration of sucrose in the medium as for *Q. robur* (Concepción Sánchez et al. 2003), the conversion rate is not altered, unless accompanied by a previous chilling treatment (Hernández et al. 2003). In order to increase embryo conversion, studies in *Quercus* have also focused on changing the concentration of agar, gibberellic acid (Concepción Sánchez et al. 2003), 6-benzilaminopurine (BAP) (Vieitez et al. 2009), indole-3-butyric acid (IBA), and thidiazuron (TDZ) in the medium (Concepción Sánchez et al. 2003). In some cases, root and shoot development in emblings are favoured, thus yielding a high frequency of conversion. However, some results are contradictory and in some cases the rates of conversion did not improve. Some authors defend the hypothesis that these differences maybe be a result of genotype (Hernández et al. 2003; Concepción Sánchez et al. 2003; Vieitez et al. 2012), while others defend the hypothesis of some genetic instability over cultures acquired from high rates of cell proliferation as is observed in secondary embryogenesis (Hornero et al. 2001; Concepción Sánchez et al. 2003; Lopes et al. 2006).

11. RNA-SEQUENCING, A WAY TO DISCOVER THE GENES INVOLVED IN SOMATIC EMBRYOGENESIS

In order to understand the mechanisms underlying somatic embryogenesis, embryo maturation as well as embryo germination in cork oak, it is crucial to proper understand which specific genes are being expressed and in which relative levels.

The evolution of NGS technologies allows in our days to profile the transcriptome at a tissue level or even at single-cell level (Saliba et al. 2014). Due to its accuracy and the ease of meaningful comparisons of samples not necessarily generated together, or even as part of the same experiment, RNA-sequencing (RNA-seq) is replacing other methods of quantifying transcript

expression, including complementary DNA (cDNA)-based microarray platforms and expressed sequenced tags (ESTs) (Wang et al. 2009b). A particular advantage of RNA-Seq is that it can capture transcripts dynamics across different tissues or conditions without sophisticated normalization of data sets (Cloonan and Grimmond 2008; Wilhelm et al. 2008). RNA-Seq has been used to accurately monitor gene expression during various development process in plants as in floral development (Zeng et al. 2017), petal development (Han et al. 2017), embryo developing (Chen et al. 2014; Elbl et al. 2015) or even somatic embryogenesis (Xu et al. 2012; Wickramasuriya and Dunwell 2015; Cao et al. 2017), between others. The technique is so robust that it allowed the decoding of the regulatory network landscape of somatic embryogenesis between *japonica* and *indica* rice subspecies (Indoliya et al. 2016) or even, the comparative transcriptome analysis between cotton somatic and zygotic embryos during development (Jin et al. 2014). In fact RNA-seq is considered a revolutionary tool for transcriptomics generating a digital measurement of gene expression in different tissues or stages of development (Wang et al. 2009b).

12. THE RNA-SEQUENCING METHODOLOGY

Currently, there are several technologies available to perform high-throughput sequencing of DNA molecules. The three principal NGS systems commercially available are from 454 Life Sciences, Illumina and Applied Biosciences (Pillai et al. 2017). Nevertheless, other technologies as from Pacific Biosciences, Oxford Nanopore and Helicos are establishing in the market with potentially even higher quality throughputs (Rhoads and Au 2015). Although, the essence of these systems relies in the same principle of miniaturizing individual sequencing reactions, the most commonly used is the Illumina platforms which will be described in further paragraphs (Illumina 2014) .

The typical RNA-Seq experiment starts with the purification of mRNAs from the bulk of the total RNA, selecting the ploy-(A) RNA molecules with poly-(T) olio-attached magnetic beads (Tariq et al. 2011). After purification the RNA is fragmented via hydrolysis or nebulisation to RNA fragments with 100-300 base-pair long or be performed on cDNA by DNase I treatment or sonication (Illumina 2011a). While the fragmentation of cDNA presents more bias towards the 3’ end

of the transcript, RNA fragmentation provides a better coverage along the gene body, however with an associated depletion in both 5' and 3' ends (Wilhelm and Landry 2009). RNA fragments are then converted into cDNA by a reverse transcriptase and then double-stranded using DNA polymerase I. These steps, initiate the conversion of total RNA into a library of template molecules representative of a transcriptome which is suitable high throughput DNA sequencing. It involves the addition of an adenine base to the double-stranded cDNA chain and the final product is generated when a thymine nucleotide from an adaptor hybridizes with the inserted adenine (Illumina 2011a). A desired range of DNA length is then purified by gel extraction selecting the size of the fragments and removing non-ligated adapters and adapter dimers. Finally, a set of primers are annealed to the adapters tail and the purified cDNA molecules are amplified by PCR. Prior to sequencing, single-stranded DNA are bridged amplified to form clonal clusters inside the sequencing flow cell. In order to achieve that, the double-stranded DNA molecules that resulted from the PCR amplification step are denatured into single stranded DNA (Illumina 2011b). First, the DNA templates are copied from the hybridized primers by 3' extension through a DNA polymerase. Then, the original templates are denatured which leaves the copied DNA immobilized on the flow cell surface which is further bridge-amplified, linearized and hybridized with sequencing primers. This procedure is repeated on each template by cycles of isothermal denaturation and amplification to create clonal clusters containing at least 1000 molecules per cluster. Four labelled reversible terminators and DNA polymerase are added to the reaction. Four fluorescently tagged nucleotides are added to the DNA template and the colour of its fluorescent label is registered by laser excitation (Illumina 2011b). The process follows during repeated sequencing cycles and capturing the signal of the next base to identity, generating sequence reads representative of each amplified cDNA template which are registered on an informatic file addressed with quality scores.

13. Bioinformatics analysis to unravel differential genes expression

After producing sequenced reads by any of the available NGS platforms the generated data needs to be analysed in order to extract biological information. A first step is to check the quality control of the sequence reads which usually also takes into account possible duplication rates. Then follows the mapping where the assembly of all the quality controlled reads, which are aligned to a genome reference or in the case of a non-available genome for the species a *de novo* transcriptome assembly is generated. In this phase, the transcriptome is reconstructed and transcripts can be analysed, discovered or even quantified by a variety of tools available to perform this data processing (Yang and Kim 2015). Once gene expression has been quantified and normalised, statistical testing between conditions is usually performed allowing to determine differential gene expression (Mutz et al. 2013). As a final step, the list of differential expressed genes may be grouped into gene expression clusters, which will represent groups of genes with similar expression levels among conditions, as well be grouped into common pathways. The pathway analysis identifies differentially active pathways and, ultimately, allows the correspondence between the transcriptome with a specific phenotype (active pathways) (Emmert-Streib and Glazko 2011).

14. Developing a method to identify the DNA-targets of QsMYB1 in cork oak

As mentioned before, ChIP-Seq is currently one of the most powerful techniques to identify protein DNA-targets. In order to perform a ChIP-Seq assay in a non-model species as *Quercus suber*, it is necessary to take into account several aspects of the technique and to perform intense optimization to the usual ChIP-Seq protocols (Kaufmann et al. 2010; Yamaguchi et al. 2014). Also, it is mandatory to evaluate how to perform the immunoprecipitation step of a protein without a specific antibody available. In order to circumvent these issues, the selected strategy may involve the production of transgenic plant expressing the

protein of interest fused with an epitope with high affinity for a specific antibody (Yamaguchi et al. 2014). One of the best epitope tags for ChIP-Seq assays is the triple FLAG epitope, commonly known by 3xFLAG (Ueda et al. 2011). The 3xFLAG structure is a synthetic peptide of 23 amino-acids within three tandem repeats of the Asp-Tyr-Lys-Xaa-Xaa-Asp motif developed for compatibility with the proteins where it is fused to, in order to improve the detection and recovery of recombinant proteins (Ueda et al. 2011). The 3xFLAG is more hydrophilic than other epitope tags commonly used, and therefore less likely to denature or inactive the proteins to which it is appended to. Theoretically, 3xFLAG is suitable to be fused with any protein. Using the nucleotide sequence of the protein of interest fused with the 3xFLAG nucleotide sequences, vectors for expression may be used to express a recombinant fused protein of interest.

The Gateway® technology and particularly the set of binary vector system available for recombinant protein expression in plants is suitable to plant transformation and recombinant protein expression *in vivo* (Krenek et al., 2015). However, a proper tissue to express the fused protein and the adjacent DNA is needed. In order to answer these requirements, the present PhD work started with the establishment of cork oak somatic embryogenic cell lines for recombinant protein production. Genetic modified embryos were then subject to a ChIP-Seq assay to identify putative protein DNA-targets in relative short time frame period. However, an important aspect regarding the expression of QsMYB1 expression native tissues have to be consider for future conclusions: *QsMYB1* is natively up-regulated in secondary growth associated tissues (e.g. phellogen and phellem). These secondary tissues are not developed in somatic embryos and therefore the cell molecular environment as well as the chromatin and DNA status are not similar, which may constrain the QsMYB1 ligation and access to some specific DNA-targets. Nevertheless, the use of an overexpression system will produce higher amounts of QsMYB1 protein which will have more coverage over these targets.

15. References

Alemanno L, Devic M, Niemenak N, Sanier C, Guilleminot J, Rio M, et al. (2008) Characterization of leafy cotyledon1-like during embryogenesis in Theobroma cacao L. Planta 227:853–866

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. J Plant Physiol 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 Physiol Biochem 73:274–281

Álvarez R, Ordás RJ (2007) Improved genetic transformation protocol for cork oak (Quercus suber L.). Plant Cell Tissue Organ Cult 91:45–52

Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes as regulators for plant responses: An overview. Physiol Mol Biol Plants 19:307–321

Arc E, Sechet J, Corbineau F, Rajjou L, Marion-Poll A (2013) ABA crosstalk with ethylene and nitric oxide in seed dormancy and germination. Front Plant Sci 4:63

Aronson J, Pereira JS, Pausas JG (2009) Cork Oak Woodlands on the Edge: Ecology, Adaptive Management, and Restoration. Island Press, Washington, DC

Arrigoni L, Richter AS, Betancourt E, Bruder K, Diehl S, Manke T, et al. (2015) Standardizing chromatin research: a simple and universal method for ChIP-seq. Nucleic Acids Res 44:67

Bailey TL (2011) DREME: Motif discovery in transcription factor ChIP-seq data. Bioinformatics 27:1653–1659

Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: Discovering and analyzing DNA and protein sequence motifs. Nucleic Acids Res 34:

Beisson F, Li Y, Bonaventure G, Pollard M, Ohlrogge JB (2007) The Acyltransferase GPAT5 Is Required for the Synthesis of Suberin in Seed Coat and Root of Arabidopsis. Plant Cell 19:351–368

Bernards MA (2002) Demystifying suberin. Can J Bot 80:227–240

Bernards MA, Lewis NG (1998) The macromolecular aromatic domain in suberized tissue: a changing paradigm. Phytochemistry 47:915–933

Bernards MA, Razem FA (2001) The poly(phenolic) domain of potato suberin: A non-lignin cell wall bio-polymer. Phytochemistry 57:1115–1122

Biłas R, Szafran K, Hnatuszko-konka K, Kononowicz AK (2016) Cis -regulatory elements used to control gene expression in plants. Plant Cell, Tissue Organ Cult 127:269–287

Bird D, Beisson F, Brigham A, Shin J, Greer S, Jetter R, et al. (2007) Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant J 52:485–498

Boavida LC, Varela MC, Feijó JA (1999) Sexual reproduction in the cork oak (Quercus suber L.). I. The progamic phase. Sex Plant Reprod 11:347–353

Bomal C, Bedon F, Caron S, Mansfield SD, Levasseur C, Cooke JEK, et al. (2008) Involvement of Pinus taeda MYB1 and MYB8 in phenylpropanoid metabolism and secondary cell wall biogenesis: a comparative in planta analysis. J Exp Bot 59:3925–3939

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, et al. (2002) Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth. Plant Cell 14:1737–1749

Bouyer D, Roudier F, Heese M, Andersen ED, Gey D, Nowack MK, et al. (2011) Polycomb repressive complex 2 controls the embryo-to-seedling phase transition. PLoS Genet 7:e1002014

Braybrook SA, Harada JJ (2008) LECs go crazy in embryo development. Trends Plant Sci. 13:624–630

Braybrook SA, Stone SL, Park S, Bui AQ, Le BH, Fischer RL, et al. (2006) Genes directly regulated by LEAFY COTYLEDON2 provide insight into the control of embryo maturation and somatic embryogenesis. Proc Natl Acad Sci 103:3468–3473

Bueno MA, Gomez A, Sepulveda F, Segui JM, Testillano PS, Manzanera JA, et al. (2003) Microspore-derived embryos from Quercus suber anthers mimic zygotic embryos and maintain haploidy in long-term anther culture. J Plant Physiol 160:953–960

Bugalho MN, Caldeira MC, Pereira JS, Aronson J, Pausas JG (2011) Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. Front Ecol Environ 9:278–286

Cao A, Zheng Y, Yu Y, Wang X, Shao D, Sun J, et al. (2017) Comparative Transcriptome Analysis of SE initial dedifferentiation in cotton of different SE capability. Sci Rep 7:8583

Chaffey N (2011) An introduction to plant structure and development. Plant anatomy for the Twenty-first Century. In: Annals of Botany. Oxford University Press, pp 271–280

Chen D, Molitor A, Liu C, Shen WH (2010) The arabidopsis PRC1-like ring-finger proteins are necessary for repression of embryonic traits during vegetative growth. Cell Res 20:1332–1344

Chen J, Zeng B, Zhang M, Xie S, Wang G, Hauck A, et al. (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. Plant Physiol 166:252–

264

Cloonan N, Grimmond SM (2008) Transcriptome content and dynamics at single-nucleotide resolution. Genome Biol 9:234

Compagnon V, Diehl P, Benveniste I, Meyer D, Schaller H, Schreiber L, et al. (2009) CYP86B1 Is Required for Very Long Chain ω-Hydroxyacid and α,ω-Dicarboxylic Acid Synthesis in Root and Seed Suberin Polyester. Plant Physiol 150:1831–1843

Concepción Sánchez M, Teresa Martínez M, Valladares S, Ferro E, Viéitez AM, Sanchez MC, et al. (2003) Maturation and germination of oak somatic embryos originated from leaf and stem explants: RAPD markers for genetic analysis of regenerants. J Plant Physiol 160:699–707

Curaba J, Moritz T, Blervaque R, Parcy F, Raz V, Herzog M, et al. (2004) AtGA3ox2, a key gene responsible for bioactive gibberellin biosynthesis, is regulated during embryogenesis by LEAFY COTYLEDON2 and FUSCA3 in Arabidopsis. Plant Physiol 136:3660–3669

Davuluri R V, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, et al. (2003) AGRIS: Arabidopsis Gene Regulatory Information Server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. BMC Bioinformatics 4:25

De Oliveira Santos M, Romano E, Yotoko KSC, Tinoco MLP, Dias BBA, Aragão FJL (2005) Characterisation of the cacao somatic embryogenesis receptor-like kinase (SERK) gene expressed during somatic embryogenesis. Plant Sci 168:723–729

Deeken R, Saupe S, Klinkenberg J, Riedel M, Leide J, Hedrich R, et al. (2016) The Nonspecific Lipid Transfer Protein AtLtpI-4 Is Involved in Suberin Formation of Arabidopsis thaliana Crown Galls. Plant Physiol 172:1911–1927

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

Ding K, Pei T, Bai Z, Jia Y, Ma P, Liang Z (2017) SmMYB36, a Novel R2R3-MYB Transcription Factor, Enhances Tanshinone Accumulation and Decreases Phenolic Acid Content in Salvia miltiorrhiza Hairy Roots. Sci Rep 7:5104

Domínguez E, Heredia-Guerrero JA, Heredia A (2015) Plant cutin genesis: Unanswered questions. Trends Plant Sci 20:551–558

Du H, Feng B-R, Yang S-S, Huang Y-B, Tang Y-X (2012) The R2R3-MYB Transcription Factor Gene Family in Maize. PLoS One 7:e37463

Dubos C, Le Gourrierec J, Baudry A, Huep G, Lanet E, Debeaujon I, et al. (2008) MYBL2 is a new regulator of flavonoid biosynthesis in Arabidopsis thaliana. Plant J 55:940–953

Dubos C, Stracke R, Grotewold E, Weisshaar B, Martin C, Lepiniec L (2010) MYB transcription factors in Arabidopsis. Trends Plant Sci. 15:573–581

Edstam MM, Blomqvist K, Eklöf A, Wennergren U, Edqvist J (2013) Coexpression patterns indicate that GPI-anchored non-specific lipid transfer proteins are involved in accumulation of cuticular wax, suberin and sporopollenin. Plant Mol Biol 83:625–649

Elbl P, Lira BS, Andrade SCS, Jo L, dos Santos ALW, Coutinho LL, et al. (2015) Comparative transcriptome analysis of early somatic embryo formation and seed development in Brazilian pine, Araucaria angustifolia (Bertol.) Kuntze. Plant Cell Tissue Organ Cult 120:903–915

Elena-Rossello JA, de Rio JM, Valdecantos JLG, Santamaria IG (1993) Ecological aspects of the floral phenology of the cork-oak (Q suber L): why do annual and biennial biotypes appear? Ann For Sci 50:114s–121s

Emmert-Streib F, Glazko G V. (2011) Pathway analysis of expression data: Deciphering functional building blocks of complex diseases. PLoS Comput Biol 7:e1002053

Evert RF (2006) Xylem: Secondary Xylem and Variations in Wood Structure. In: Esau's Plant Anatomy. John Wiley & Sons, Inc., pp 291–322

Feng C, Andreasson E, Maslak A, Mock HP, Mattsson O, Mundy J (2004) Arabidopsis MYB68 in development and responses to environmental cues. Plant Sci 167:1099–1107

Florez SL, Erwin RL, Maximova SN, Guiltinan MJ, Curtis WR (2015) Enhanced somatic embryogenesis in Theobroma cacao using the homologous BABY BOOM transcription factor. BMC Plant Biol 15:121

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc Natl Acad Sci U S A 111:2367–72

Franke R, Schreiber L (2007) Suberin-a biopolyester forming apoplastic plant interfaces. Curr Opin Plant Biol 10:252–259

Gaj MD, Zhang S, Harada JJ, Lemaux PG (2005) Leafy cotyledon genes are essential for induction of somatic embryogenesis of Arabidopsis. Planta 222:977–988

Gallois J-L, Nora FR, Mizukami Y, Sablowski R (2004) WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem service WUSCHEL induces shoot stem cell activity and developmental plasticity in the root meristem. Genes Dev 18:375–380

García-Martín G (2001) Quercus suber L. somatic embryo germination and plant conversion: pretreatments and germination conditions. Vitr Cell Dev Biol - Plant 37:190–198

García-Martín G, Manzanera JA, González-Benito ME (2005) Effect of exogenous ABA on embryo maturation and quantification of endogenous levels of ABA and IAA in Quercus suber somatic embryos. Plant Cell Tissue Organ Cult 80:171–177

Gazzarrini S, Tsuchiya Y, Lumba S, Okamoto M, McCourt P (2004) The transcription factor FUSCA3 controls developmental timing in Arabidopsis through the hormones Gibberellin and abscisic acid. Dev Cell 7:373–385

Georgiev S, Boyle AP, Jayasurya K, Ding X, Mukherjee S, Ohler U (2010) Evidence-ranked motif identification. Genome Biol 11:R19

Gil L (2009) Cork Composites: A Review. Materials (Basel) 2:776–789

Gil L, Varela MC (2008) Quercus suber - Technical Guidelines for genetic conservation and use for cork oak. EUFORGEN

Gilad Y, Rifkin SA, Pritchard JK (2008) Revealing the architecture of gene regulation: the promise of eQTL studies. Trends Genet. 24:408–415

Girard a.-L, Mounet F, Lemaire-Chamley M, Gaillard C, Elmorjani K, Vivancos J, et al. (2012) Tomato GDSL1 Is Required for Cutin Deposition in the Fruit Cuticle. Plant Cell 24:3119–3134

Goldberg RB, de Paiva G, Yadegari R (1994) Plant embryogenesis: zygote to seed. Science 266:605–614

Gou J-Y, Yu X-H, Liu C-J (2009) A hydroxycinnamoyltransferase responsible for synthesizing suberin aromatics in Arabidopsis. Proc Natl Acad Sci U S A 106:18855–60

Gou M, Hou G, Yang H, Zhang X, Cai Y, Kai G, et al. (2017) The MYB107 Transcription Factor Positively Regulates Suberin Biosynthesis. Plant Physiol 173:1045–1058

Graça J (2015) Suberin: the biopolyester at the frontier of plants. Front Chem 3:62

Graça J (2010) Hydroxycinnamates in suberin formation. Phytochem Rev 9:85–91

Graça J, Pereira H (1997) Cork Suberin: A Glyceryl Based Polyester. Holzforschung 51:225–234

Graça J, Pereira H (2000) Methanolysis of bark suberins: Analysis of glycerol and acid monomers. Phytochem Anal 11:45–51

Graça J, Santos S (2007) Suberin: a biopolyester of plants' skin. Macromol Biosci 7:128–35

Graça J, Santos S (2006) Glycerol-derived ester oligomers from cork suberin. Chem Phys Lipids 144:96–107

Guan Y, Li S-G, Fan X-F, Su Z-H (2016) Induction of somatic embryogenesis in explants of shoot cultures established from adult Eucalyptus globulus and E. saligna × E. maidenii trees. Front Plant Sci 7:938

Guo A, He K, Liu D, Bai S, Gu X, Wei L, et al. (2005) DATF: A database of Arabidopsis

transcription factors. Bioinformatics 21:2568–2569

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble W (2007) Quantifying similarity between motifs. Genome Biol 8:R24

Han Y, Wan H, Cheng T, Wang J, Yang W, Pan H, et al. (2017) Comparative RNA-seq analysis of transcriptome dynamics during petal development in Rosa chinensis. Sci Rep 7:43382

Harding EW, Tang W, Nichols KW, Fernandez DE, Perry SE (2003) Expression and maintenance of embryogenic potential is enhanced through constitutive expression of AGAMOUS-Like 15. Plant Physiol 133:653–663

Haslam TM, Haslam R, Thoraval D, Pascal S, Delude C, Domergue F, et al. (2015) ECERIFERUM2-LIKE proteins have unique biochemical and physiological functions in very-long-chain fatty acid elongation. Plant Physiol 167:682–92

He Q, Jones DC, Li W, Xie F, Ma J, Sun R, et al. (2016) Genome-Wide Identification of R2R3-MYB Genes and Expression Analyses During Abiotic Stress in Gossypium raimondii. Sci Rep 6:22980

Hecht V, Vielle-Calzada JP, Hartog M V, Schmidt ED, Boutilier K, Grossniklaus U, et al. (2001) The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol 127:803–16

Heck GR, Perry SE, Nichols KW, Fernandez DE (1995) AGL15, a MADS domain protein expressed in developing embryos. Plant Cell 7:1271–82

Heldt H-W, Piechulla B (2011) Phenylpropanoids comprise a multitude of plant secondary metabolites and cell wall components. In: Plant Biochemistry. pp 431–449

Hernández I, Celestino C, Alegre J, Toribio M (2003) Vegetative propagation of Quercus suber L. by somatic embryogenesis. II. Plant regeneration from selected cork oak trees. Plant Cell Rep 21:765–70

Hernandéz I, Celestino C, Martinez I, Manjon JL, Diez J, Fernandez-Guijarro B, et al. (2001) Cloning mature cork oak (Quercus suber L.) trees by somatic embryogenesis. In: Melhoramento (Portugal)

Hernandéz I, Celestino C, Toribio M (2003) Vegetative propagation of Quercus suber L. by somatic embryogenesis. I. Factors affecting the induction in leaves from mature cork oak trees. Plant Cell Rep 21:759–764

Hernández I, Cuenca B, Carneros E, Alonso-Blazquez N, Ruiz M, Celestino C, et al. (2011) Application of Plant Regeneration of Selected Cork Oak Trees by Somatic Embryogenesis to Implement Multivarietal Forestry for Cork Production. Tree For Sci Biotechnol 19–26

Höfer R, Briesen I, Beck M, Pinot F, Schreiber L, Franke R (2008) The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid ω-hydroxylase involved in suberin monomer biosynthesis. J Exp Bot 59:2347–2360

Hooke R (1665) Micrographia, or, Some physiological descriptions of minute bodies made by magnifying glasses: With observations and inquiries thereupon., Printed. Martyn, J and Allestry, J, London

Hornero J, Martínez I, Celestino C, Gallego FJ, Torres V, Toribio M (2001) Early Checking of Genetic Stability of Cork Oak Somatic Embryos By Aflp Analysis. Int J Plant Sci 162:827

Horstman A, Li M, Heidmann I, Weemen M, Chen B, Muiño JM, et al. (2017) The BABY BOOM transcription factor activates the LEC1-ABI3-FUS3-LEC2 network to induce somatic embryogenesis. Plant Physiol 175:848–587

Huang M, Hu Y, Liu X, Li Y, Hou X (2015a) Arabidopsis LEAFY COTYLEDON1 controls cell fate determination during post-embryonic development. Front Plant Sci 6:955

Huang M, Hu Y, Liu X, Li Y, Hou X (2015b) Arabidopsis LEAFY COTYLEDON1 Mediates Postembryonic Development via Interacting with PHYTOCHROME-INTERACTING FACTOR4. Plant Cell 27:3099–3111

Iida K, Seki M, Sakurai T, Satou M, Akiyama K, Toyoda T, et al. (2005) RARTF: Database and tools for complete sets of Arabidopsis transcription factors. DNA Res 12:247–256

Ikeuchi M, Iwase A, Rymen B, Harashima H, Shibata M, Ohnuma M, et al. (2015) PRC2 represses dedifferentiation of mature somatic cells in Arabidopsis. Nat Plants 1:15089

Illumina (2014) Sequencing Methods Review (A review of publications featuring Illumina Technology). Illumina Protoc 199:27

Illumina (2011a) TruSeq ™ RNA and DNA Sample Preparation Kits v2. Illumina Man. 2–5

Illumina (2011b) cBot Fully automated clonal cluster generation for Illumina sequencing. Illumina Man. 1–4

Indoliya Y, Tiwari P, Chauhan AS, Goel R, Shri M, Bag SK, et al. (2016) Decoding regulatory landscape of somatic embryogenesis reveals differential regulatory networks between japonica and indica rice subspecies. Sci Rep 6:23050

Isah T (2016) Induction of somatic embryogenesis in woody plants. Acta Physiol. Plant. 38:118

Ito M (2005) Conservation and diversification of three-repeat Myb transcription factors in plants. J. Plant Res. 118:61–69

Jin F, Hu L, Yuan D, Xu J, Gao W, He L, et al. (2014) Comparative transcriptome analysis

between somatic embryos (SEs) and zygotic embryos in cotton: Evidence for stress response functions in SE development. Plant Biotechnol J 12:161–173

Jothi R, Cuddapah S, Barski A, Cui K, Zhao K (2008) Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res 36:5221–5231

Junker A, Mönke G, Rutten T, Keilwagen J, Seifert M, Thi TMN, et al. (2012) Elongation-related functions of LEAFY COTYLEDON1 during the development of Arabidopsis thaliana. Plant J 71:427–442

Kang J, Park J, Choi H, Burla B, Kretzschmar T, Lee Y, et al. (2011) Plant ABC Transporters. Arabidopsis Book 9:e0153

Kang YH, Kirik V, Hulskamp M, Nam KH, Hagely K, Lee MM, et al. (2009) The MYB23 Gene Provides a Positive Feedback Loop for Cell Fate Specification in the Arabidopsis Root Epidermis. PLANT CELL ONLINE 21:1080–1094

Karlova R, Boeren S, Russinova E, Aker J, Vervoort J, de Vries S (2006) The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE1 protein complex includes BRASSINOSTEROID-INSENSITIVE1. Plant Cell 18:626–638

Kaufmann K, Muiño JM, Østerås M, Farinelli L, Krajewski P, Angenent GC (2010) Chromatin immunoprecipitation (ChIP) of plant transcription factors followed by sequencing (ChIP-SEQ) or hybridization to whole genome arrays (ChIP-CHIP). Nat Protoc 5:457–472

Keller T (2006) Arabidopsis REGULATOR OF AXILLARY MERISTEMS1 Controls a Leaf Axil Stem Cell Niche and Modulates Vegetative Development. PLANT CELL ONLINE 18:598–611

Kolattukudy PE (2001) Polyesters in higher plants. Adv Biochem Eng Biotechnol 71:1–49

Koo SC, Yoon HW, Kim CY, Moon BC, Cheong YH, Han HJ, et al. (2007) Alternative splicing of the OsBWMK1 gene generates three transcript variants showing differential subcellular localizations. Biochem Biophys Res Commun 360:188–193

Kosma DK, Murmu J, Razeq FM, Santos P, Bourgault R, Molina I, et al. (2014) AtMYB41 activates ectopic suberin synthesis and assembly in multiple plant species and cell types. Plant J 80:216–229

Kurdyukov S, Faust A, Nawrath C, Bär S, Voisin D, Efremova N, et al. (2006) The epidermis-specific extracellular BODYGUARD controls cuticle development and morphogenesis in *Arabidopsis*. Plant Cell 18:321–339

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22:1813–1831

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Lashbrooke J, Cohen H, Levy-Samocha D, Tzfadia O, Panizel I, Zeisler V, et al. (2016) MYB107 and MYB9 Homologs Regulate Suberin Deposition in Angiosperms. Plant Cell 28:2097–2116

Lauria M, Rossi V (2011) Epigenetic control of gene regulation in plants. Biochim Biophys Acta 1809:369–78

Lee S-B, Jung S-J, Go Y-S, Kim H-U, Kim J-K, Cho H-J, et al. (2009) Two Arabidopsis 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. Plant J 60:462–475

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, et al. (2013) Acyl-Lipid Metabolism. Arabidopsis Book 11:e0161

Li W, Lin Y-C, Li Q, Shi R, Lin C-Y, Chen H, et al. (2014) A robust chromatin immunoprecipitation protocol for studying transcription factor-DNA interactions and histone modifications in wood-forming tissue. Nat Protoc 9:2180–93

Li X, Nair A, Wang S, Wang L (2015) Quality control of RNA-seq experiments. In: RNA Bioinformatics. pp 137–146

Li Y, Beisson F, Koo AJK, Molina I, Pollard M, Ohlrogge J (2007) Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. Proc Natl Acad Sci U S A 104:18339–18344

Lopes T, Pinto G, Loureiro J, Costa A, Santos C (2006) Determination of genetic stability in long-term somatic embryogenic cultures and derived plantlets of cork oak using microsatellite markers. Tree Physiol 26:1145–1152

Lotan T, Ohto M, Yee KM, West MA., Lo R, Kwong RW, et al. (1998) Arabidopsis LEAFY COTYLEDON1 Is Sufficient to Induce Embryo Development in Vegetative Cells. Cell 93:1195–1205

Luerssen H, Kirik V, Herrmann P, Misera S (1998) FUSCA3 encodes a protein with a conserved VP1/AB13-like B3 domain which is of functional importance for the regulation of seed maturation in Arabidopsis thaliana. Plant J 15:755–764

Machanick P, Bailey TL (2011) MEME-ChIP: Motif analysis of large DNA datasets. Bioinformatics 27:1696–1697

Martínez T, Corredoira E, Valladares S, Jorquera L, Vieitez AM (2008) Germination and conversion of somatic embryos derived from mature Quercus robur trees: the effects of cold storage and thidiazuron. Plant Cell Tissue Organ Cult 95:341

Mathelier A, Fornes O, Arenillas DJ, Chen C, Denay G, Lee J, et al. (2015) JASPAR 2016: a major expansion and update of the open-access database of transcription

factor binding profiles. Nucleic Acids Res 44:D110-115

Mauri P V, Manzanera JA (2004) Effect of abscisic acid and stratification on somatic embryo maturation and germination of holm oak (Quercus ilex L.). Vitr Cell Dev Biol - Plant 40:495–498

Medina-Rivera A, Defrance M, Sand O, Herrmann C, Castro-Mondragon JA, Delerce J, et al. (2015) RSAT 2015: Regulatory sequence analysis tools. Nucleic Acids Res 43:W50–W56

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

Mordhorst AP, Toonen MAJ, de Vries SC, Meinke D (1997) Plant Embryogenesis. CRC Crit Rev Plant Sci 16:535–576

Muller D (2006) Blind Homologous R2R3 Myb Genes Control the Pattern of Lateral Meristem Initiation in Arabidopsis. PLANT CELL ONLINE 18:586–597

Mutz K-O, Heilkenbrinker A, Lönne M, Walter J-G, Stahl F (2013) Transcriptome analysis using next-generation sequencing. Curr Opin Biotechnol 24:22–30

Namasivayam P (2007) Acquisition of embryogenic competence during somatic embryogenesis. Plant Cell. Tissue Organ Cult. 90:1–8

Negrel J, Pollet B, Lapierre C (1996) Ether-linked ferulic acid amides in natural and wound periderms of potato tuber. Phytochemistry 43:1195–1199

Nic-Can GI, López-Torres A, Barredo-Pool F, Wrobel K, Loyola-Vargas VM, Rojas-Herrera R, et al. (2013) New insights into somatic embryogenesis: LEAFY COTYLEDON1, BABY BOOM1 and WUSCHEL-related homeobox4 are epigenetically regulated in Coffea canephora. PLoS One 8:e72160

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. (2017) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165:1280–1292

Ogas J, Kaufmann S, Henderson J, Somerville C (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in Arabidopsis. Proc Natl Acad Sci U S A 96:13839–13844

Panikashvili D, Savaldi-Goldstein S, Mandel T, Yifhar T, Franke RB, Hofer R, et al. (2007) The Arabidopsis DESPERADO/AtWBC11 transporter is required for cutin and wax secretion. Plant Physiol 145:1345–1360

Panikashvili D, Shi JX, Bocobza S, Franke RB, Schreiber L, Aharoni A (2010) The Arabidopsis DSO/ABCG11 transporter affects cutin metabolism in reproductive organs and suberin in roots. Mol Plant 3:563–575

Panikashvili D, Shi JX, Schreiber L, Aharoni A (2009) The Arabidopsis DCR Encoding a

Soluble BAHD Acyltransferase Is Required for Cutin Polyester Formation and Seed Hydration Properties. PLANT Physiol 151:1773–1789

Panikashvili D, Shi JX, Schreiber L, Aharoni A (2011) The Arabidopsis ABCG13 transporter is required for flower cuticle secretion and patterning of the petal epidermis. New Phytol 190:113–124

Parida SK, Dalal V, Singh AK, Singh NK, Mohapatra T (2009) Genic non-coding microsatellites in the rice genome: characterization, marker design and use in assessing genetic and evolutionary relationships among domesticated groups. BMC Genomics 10:140

Park PJ (2009) ChIP–seq: advantages and challenges of a maturing technology. Nat Rev Genet 10:669–680

Pascal S, Bernard A, Sorel M, Pervent M, Vile D, Haslam RP, et al. (2013) The Arabidopsis cer26 mutant, like the cer2 mutant, is specifically affected in the very long chain fatty acid elongation process. Plant J 73:733–746

Patzlaff A, McInnis S, Courtenay A, Surman C, Newman LJ, Smith C, et al. (2003) Characterisation of a pine MYB that regulates lignification. Plant J 36:743–754

Pavesi G, Pesole G (2006) Using Weeder for the discovery of conserved transcription factor binding sites. Curr Protoc Bioinformatics Chapter 2:Unit 2.11

Pereira H (2007a) The structure of cork. In: Cork: Biology, Production and Uses. Elsevier Science, pp 33–53

Pereira H (2007b) The formation and growth of cork. In: Cork: Biology, Production and Uses. Elsevier Science, pp 7–31

Pérez M, Bueno MA, Escalona M, Toorop P, Rodríguez R, Cañal MJ (2013) Temporary immersion systems (RITA®) for the improvement of cork oak somatic embryogenic culture proliferation and somatic embryo production. Trees - Struct Funct 27:1277–1284

Pérez M, Viejo M, LaCuesta M, Toorop P, Ca??al MJ (2015) Epigenetic and hormonal profile during maturation of Quercus Suber L. somatic embryos. J Plant Physiol 173:51–61

Pighin J a, Zheng H, Balakshin LJ, Goodman IP, Western TL, Jetter R, et al. (2004) Plant cuticular lipid export requires an ABC transporter. Science 306:702–4

Pillai S, Gopalan V, Lam AKY (2017) Review of sequencing platforms and their applications in phaeochromocytoma and paragangliomas. Crit. Rev. Oncol. Hematol. 116:58–67

Pinto G, Santos C, Neves L, Araujo C (2002a) Somatic embryogenesis and plant regeneration in Eucalyptus globulus Labill. Plant Cell Rep 21:

Pinto G, Valentim H, Costa A, Castro S, Santos C (2002b) Somatic embryogenesis in

leaf callus from a mature Quercus suber L. Tree. Vitr Cell Dev Biol - Plant 38:569–572

Pintos B, Manzanera J a., Bueno MA (2010) Oak somatic and gametic embryos maturation is affected by charcoal and specific aminoacids mixture. Ann For Sci 67:205

Pollard M, Beisson F, Li Y, Ohlrogge JB (2008) Building lipid barriers: biosynthesis of cutin and suberin. Trends Plant Sci 13:236–246

Prewein C, Vagner M, Wilhelm E (2004) Changes in water status and proline and abscisic acid concentrations in developing somatic embryos of pedunculate oak (Quercus robur) during maturation and germination. Tree Physiol 24:1251–1257

Radoeva T, Weijers D (2014) A roadmap to embryo identity in plants. Trends Plant Sci 19:709–716

Ranathunge K, Schreiber L, Franke R (2011) Suberin research in the genomics era-New interest for an old polymer. Plant Sci. 180:339–413

Reinert J (1958) Morphogenese und ihre Kontrolle an Gewebekulturen aus Carotten. Naturwissenschaften 45:344–345

Rhoads A, Au KF (2015) PacBio Sequencing and Its Applications. Genomics, Proteomics Bioinforma. 13:278–289

Riaño-Pachón DM, Ruzicic S, Dreyer I, Mueller-Roeber B (2007) PlnTFDB: an integrative plant transcription factor database. BMC Bioinformatics 8:42

Rider SD, Henderson JT, Jerome RE, Edenberg HJ, Romero-Severson J, Ogas J (2003) Coordinate repression of regulators of embryonic identity by PICKLE during germination in Arabidopsis. Plant J 35:33–43

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110

Rupps A, Raschke J, Rümmler M, Linke B, Zoglauer K (2016) Identification of putative homologs of Larix decidua to BABYBOOM (BBM), LEAFY COTYLEDON1 (LEC1), WUSCHEL-related HOMEOBOX2 (WOX2) and SOMATIC EMBRYOGENESIS RECEPTOR-like KINASE (SERK) during somatic embryogenesis. Planta 243:473–488

Saliba A-E, Westermann AJ, Gorski SA, Vogel J (2014) Single-cell RNA-seq: advances and future challenges. Nucleic Acids Res 42:8845–60

Salminen TA, Blomqvist K, Edqvist J (2016) Lipid transfer proteins: classification, nomenclature, structure, and function. Planta 244:971–997

Samuels L, Kunst L, Jetter R (2008) Sealing plant surfaces: cuticular wax formation by epidermal cells. Annu Rev Plant Biol 59:683–707

Sanchez C, Martinez MT, Vidal N, San-Jose MC, Valladare S, Vieitez AM (2008) Preservation of Quercus robur germplasm by cryostorage of embryogenic cultures derived from mature trees and RAPD analysis of genetic stability. Cryo Letters 29:493–504

Schreiber L, Hartmann K, Skrabs M, Zeier J (1999) Apoplastic barriers in roots: chemical composition of endodermal and hypodermal cell walls. J Exp Bot 50:1267–1280

Seo PJ, Xiang F, Qiao M, Park JY, Lee YN, Kim SG, et al. (2009) The MYB96 transcription factor mediates abscisic acid signaling during drought stress response in Arabidopsis. Plant Physiol 151:275–289

Serra O, Hohn C, Franke R, Prat S, Molinas M, Figueras M (2010) A feruloyl transferase involved in the biosynthesis of suberin and suberin-associated wax is required for maturation and sealing properties of potato periderm. Plant J 62:277–290

Serra O, Soler M, Hohn C, Franke R, Schreiber L, Prat S, et al. (2009a) Silencing of StKCS6 in potato periderm leads to reduced chain lengths of suberin and wax compounds and increased peridermal transpiration. J Exp Bot 60:697–707

Serra O, Soler M, Hohn C, Sauveplane V, Pinot F, Franke R, et al. (2009b) CYP86A33-Targeted Gene Silencing in Potato Tuber Alters Suberin Composition, Distorts Suberin Lamellae, and Impairs the Periderm's Water Barrier Function. Plant Physiol 149:1050–1060

Sharov AA, Ko MSH (2009) Exhaustive search for over-represented DNA sequence motifs with cisfinder. DNA Res 16:261–273

Shockey JM, Fulda MS, Browse JA (2002) Arabidopsis contains nine long-chain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerolipid metabolism. Plant Physiol 129:1710–1722

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

Steward FC, Mapes MO, Mears K (1958) Growth and organized development of cultured cells. II. Organization in cultures grown from freely suspended cells. Am J Bot 45:705–708

Stone SL, Kwong LW, Yee KM, Pelletier J, Lepiniec L, Fischer RL, et al. (2001) LEAFY COTYLEDON2 encodes a B3 domain transcription factor that induces embryo development. Proc Natl Acad Sci 98:11806–11811

Su YH, Zhao XY, Liu YB, Zhang CL, O'Neill SD, Zhang XS (2009) Auxin-induced WUS expression is essential for embryonic stem cell renewal during somatic embryogenesis in Arabidopsis. Plant J 59:448–460

Sunderlikova 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

Tariq MA, Kim HJ, Jejelowo O, Pourmand N (2011) Whole-transcriptome RNAseq analysis from minute amount of total RNA. Nucleic Acids Res 39:e120

Tominaga-Wada R, Nukumizu Y, Sato S, Kato T, Tabata S, Wada T (2012) Functional divergence of MYB-related genes, WEREWOLF and AtMYB23 in Arabidopsis. Biosci Biotechnol Biochem 76:883–887

Toribio M, Celestino C, Molinas M (2005) Cork Oak, Quercus suber L. In: Jain SM, Gupta PK (eds) Protocol for Somatic Embryogenesis in Woody Plants. Springer Netherlands, Dordrecht, pp 445–457

Toribio M, Fernández C, Celestino C, Martínez MT, San-José MC, Vieitez AM (2004) Somatic embryogenesis in mature Quercus robur trees. Plant Cell Tissue Organ Cult 76:283–287

Tsuwamoto R, Yokoi S, Takahata Y (2010) Arabidopsis EMBRYOMAKER encoding an AP2 domain transcription factor plays a key role in developmental change from vegetative to embryonic phase. Plant Mol Biol 73:481–492

Ueda M, Manabe Y, Mukai M (2011) The high performance of 3XFLAG for target purification of a bioactive metabolite: A tag combined with a highly effective linker structure. Bioorganic Med Chem Lett 21:1359–1362

Valladares S, Toribio M, Celestino C, Vieitez AM (2004) Cryopreservation of embryogenic cultures from mature Quercus suber trees using vitrification. Cryo-Letters 25:177–186

Vengadesan G, Pijut PM (2009) In vitro propagation of northern red oak (Quercus rubra L.). Vitr Cell Dev Biol - Plant 45:474–482

Verdeil JL, Alemanno L, Niemenak N, Tranbarger TJ (2007) Pluripotent versus totipotent plant stem cells: dependence versus autonomy? Trends Plant Sci 12:245–252

Vieitez AM, Corredoira E, Ballester A, Muñoz F, Durán J, Ibarra M (2009) In vitro regeneration of the important north american oak species quercus alba, quercus bicolor and quercus rubra. Plant Cell Tissue Organ Cult 98:135–145

Vieitez AM, Corredoira E, Martínez MT, San-José MC, Sánchez C, Valladares S, et al. (2012) Application of biotechnological tools to Quercus improvement. Eur J For Res 131:519–539

Vishwanath SJ, Delude C, Domergue F, Rowland O (2015) Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. Plant Cell Rep 34:573–586

Von Arnold S, Sabala I, Bozhkov P, Dyachok J, Filonova L (2002) Developmental pathways of somatic embryogenesis. Plant Cell Tissue Organ Cult 69:233–249

Wang H, Caruso L V, Downie a B, Perry SE (2004) The Embryo MADS Domain Protein

AGAMOUS-Like 15 Directly Regulates Expression of a Gene Encoding an Enzyme Involved in Gibberellin Metabolism. Plant Cell 16:1206–1219

Wang J, Huda A, Lunyak V V, Jordan IK (2010) A Gibbs sampling strategy applied to the mapping of ambiguous short sequence tags. Bioinformatics 26:2501–2508

Wang XC, Niu QW, Teng C, Li C, Mu J, Chua NH, et al. (2009a) Overexpression of PGA37/MYB118 and MYB115 promotes vegetative-to-embryonic transition in Arabidopsis. Cell Res 19:224–235

Wang Z, Gerstein M, Snyder M, Hayer K, Black MB, Dueck H, et al. (2009b) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57

Whitington T, Frith MC, Johnson J, Bailey TL (2011) Inferring transcription factor complexes from ChIP-seq data. Nucleic Acids Res 39:e98

Wickramasuriya AM, Dunwell JM (2015) Global scale transcriptome analysis of Arabidopsis embryogenesis in vitro. BMC Genomics 16:301

Wilbanks EG, Facciotti MT (2010) Evaluation of algorithm performance in ChIP-seq peak detection. PLoS One 5:e11471

Wilhelm BT, Landry JR (2009) RNA-Seq-quantitative measurement of expression through massively parallel RNA-sequencing. Methods 48:249–257

Wilhelm BT, Marguerat S, Watt S, Schubert F, Wood V, Goodhead I, et al. (2008) Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 453:1239–1243

Wingender E, Chen X, Hehl R, Karas H, Liebich I, Matys V, et al. (2000) TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res 28:316–319

Wittkopp PJ, Kalay G (2011) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet 13:59–69

Xu H, Gao Y, Wang J (2012) Transcriptomic analysis of rice (Oryza sativa) developing embryos using the RNA-Seq technique. PLoS One 7:e30646

Yadav V, Molina I, Ranathunge K, Castillo IQ, Rothstein SJ, Reed JW (2014) ABCG Transporters Are Required for Suberin and Pollen Wall Extracellular Barriers in Arabidopsis. Plant Cell 26:3569–3588

Yamaguchi N, Winter CM, Wu M-F, Kwon CS, William DA, Wagner D (2014) PROTOCOL: Chromatin Immunoprecipitation from Arabidopsis Tissues. Arab B 12:e0170

Yan B, Stark RE (2000) Biosynthesis, molecular structure, and domain architecture of potato suberin: A 13C NMR study using isotopically labeled precursors. J Agric Food Chem 48:3298–3304

Yang IS, Kim S (2015) Analysis of Whole Transcriptome Sequencing Data: Workflow and Software. Genomics Inform 13:119

Yang W, Simpson JP, Li-Beisson Y, Beisson F, Pollard M, Ohlrogge JB (2012) A Land-Plant-Specific Glycerol-3-Phosphate Acyltransferase Family in Arabidopsis: Substrate Specificity, sn-2 Preference, and Evolution. Plant Physiol 160:638–652

Yeats TH, Huang W, Chatterjee S, Viart HMF, Clausen MH, Stark RE, et al. (2014) Tomato Cutin Deficient 1 (CD1) and putative orthologs comprise an ancient family of cutin synthase-like (CUS) proteins that are conserved among land plants. Plant J 77:667–675

Yeats TH, Martin LBB, Viart HM-F, Isaacson T, He Y, Zhao L, et al. (2012) The identification of cutin synthase: formation of the plant polyester cutin. Nat Chem Biol 8:609–611

Yeats TH, Rose JKC (2013) The formation and function of plant cuticles. Plant Physiol 163:5–20

Yong-Wook K (2000) Somatic Embryogenesis in Quercus Acutissima. In: Jain SM, Gupta PK, Newton RJ (eds) Somatic Embryogenesis in Woody Plants: Volume 6. Springer Netherlands, Dordrecht, pp 671–685

Yu Q, Huo H, Chen X, Guo H, Vitter JS, Huan J (2014) An efficient motif finding algorithm for large DNA data sets. In: Bioinformatics and Biomedicine (BIBM), 2014 IEEE International Conference on. pp 397–402

Zegzouti R, Arnould M-F, Favre J-M (2001) Histological investigation of the multiplication step in secondary somatic embryogenesis of Quercus robur L. Ann For Sci 58:681–690

Zeng F, Biligetu B, Coulman B, Schellenberg MP, Fu Y-B (2017) RNA-Seq analysis of gene expression for floral development in crested wheatgrass (Agropyron cristatum L.). PLoS One 12:e0177417

Zhang H, Bishop B, Ringenberg W, Muir WM, Ogas J (2012) The CHD3 Remodeler PICKLE Associates with Genes Enriched for Trimethylation of Histone H3 Lysine 27. PLANT Physiol 159:418–432

Zhang X-N, Mount SM (2009) Two alternatively spliced isoforms of the Arabidopsis SR45 protein have distinct roles during normal plant development. Plant Physiol 150:1450–1458

Zhang Y, Clemens A, Maximova SN, Guiltinan MJ (2014) The Theobroma cacao B3 domain transcription factor TcLEC2 plays a duel role in control of embryo development and maturation. BMC Plant Biol 14:106

Zhao Q, Dixon RA (2011) Transcriptional networks for lignin biosynthesis: More complex than we thought? Trends Plant Sci. 16:227–233

Zheng Y, Ren N, Wang H, Stromberg AJ, Perry SE (2009) Global Identification of Targets of the Arabidopsis MADS Domain Protein AGAMOUS-Like15. PLANT CELL

ONLINE 21:2563–2577

Zhu S ping, Wang J, Ye J li, Zhu AD, Guo W wu, Deng X xin (2014) Isolation and characterization of LEAFY COTYLEDON 1-LIKE gene related to embryogenic competence in Citrus sinensis. Plant Cell Tissue Organ Cult 119:1–13

CHAPTER II

CORK OAK SOMATIC EMBRYOS, A MODEL FOR FUNCTIONAL STUDIES: INDUCTION, SELECTIVE AGENT CHARACTERIZATION, TRANSFORMATION, AND CRYOPRESERVATION

Chapter submitted as original article in a SCI journal

Capote T, Gonçalves S, Morais-Cecílio L. Cork Oak somatic embryos, a model for functional studies: induction, selective agent characterization, transformation and cryopreservation. *Plant Cell, Tissue and Organ Culture*. *Submitted*

TABLE OF CONTENTS – CHAPTER II

Cork Oak somatic embryos, a model for functional studies: induction, selective agent characterization, transformation, and cryopreservation 59

1. Abstract 62

2. Introduction.. 63

3. Materials and Methods 64

3.1. Explant source and preparation 64

3.2. Media composition 65

3.3. Procedures and culture conditions.. 65

3.4. Experimental design and data analysis of somatic embryogenesis induction 66

3.5. Determination of selective agent concentration 66

3.6. Bacterial strains and binary plasmids.. 67

3.7. Embryo transformation.. 67

3.8. RT-qPCR transcript detection of QsMYB1::3xFLAG expression 68

3.9. Embryos encapsulation and cryopreservation 68

4. Results and Discussion 69

4.1. Influence of treatment type, tree type and genotype in somatic embryogenesis 69

4.2. Kanamycin as selective agent of genetic modified cork oak somatic embryos.. 73

4.3. Effective transformation of cork oak embryogenic cell lines.................. 75

4.4. Storage of embryogenic cell lines for future use 76

5. Conclusion.. 77

6. References.. 78

7. Supplementary Material.. 81

7.1. Supplementary Material 1 81

1. ABSTRACT

Cork oak is a species with an important socio-economic role in the Mediterranean basin but has been in decline over the last decades. Future breeding and reforestation programs may have the capacity to invert this trend where integration of somatic embryogenesis is widely accepted as the tool of selection to produce embryos in large-scale.

Methods for inducing high-frequency somatic embryos in cork oak mature leaves from 11 individuals producing high or low cork quality were investigated by two induction treatments. After statistical analysis to study the genotypes, treatment and provenience influence two selected embryogenic cell lines were chosen to produce high quality somatic embryos in order to further produce genetic modified cork oak embryos. These embryos overexpressing QsMYB1 gene and producing the recombinant protein were successfully maintained through secondary embryogenesis. The maintenance of those lines was achieved by cryopreservation with an efficient optimized protocol.

Results demonstrate a genotype disposition to somatic embryogenesis induction in cork oak. Genotypes associated with high quality cork producing trees showed to have higher susceptibility to somatic embryogenesis induction. Nevertheless, to achieve the best yields of somatic embryogenesis induction it is convenient to determine the best treatment conditions for each genotype, since some variation among different genotypes is observed.

The developed methodology has demonstrated to be efficient to produce cork oak somatic embryos in large-scale. In addition, this work reports the generation of a valuable resource for the scientific community: somatic embryogenesis lines that are made available to researchers, allowing the realization of functional and biochemical studies in a non-model and unique species such as *Quercus suber*.

2. INTRODUCTION

Cork oak (*Quercus suber* L.) is one of the most important Mediterranean forest tree species. Ecologically, it plays a role as environmental protector populating dried and semi-arid regions around the Mediterranean basin. Socially, the species growth represents employment among the rural populations due to high economic value of cork (Aronson et al. 2009). Cork is the outer bark of this oak and is sustainably extracted by human hand usually once in a decade up to around 200 years causing no damage to the trees. Cork or phellem is part of the periderm that surrounds stems, branches and root (Pereira 2007a).The thickness of the tissue varies from individual to individual and its range of variation may be very large. Also, an important factor that can be highly variable in the same population is cork quality (Pereira 2007b). Cork industry and producers have several classifications from low to high quality cork, being its value completely dependent on quality. The understanding of the genetic variability associated to these variations is a challenge for plant science community and a valuable knowledge for cork producers and industry. A Portuguese national consortium (COEC - Cork Oak EST Consortium) was created in order to collect information of expressed sequence tags (ESTs). This project generated more than 6,000 ESTs (Pereira-Leal et al. 2014) for this species. The cork oak genome sequencing is also in course, which will reveal the parts of the genome not covered by transcriptome sequencing, including repeated sequences, pseudogenes and introns (Ramos et al. *submitted*)

The molecular genetics about cork formation have always interested researchers, but knowledge on this subject has progressed slowly due to the limitations associated with a long-life cycle species. The availability of tools to perform functional studies involving gene modifications has, more recently, allowed the study of genes and pathways involved in cork formation and development. Although cork oak shows a recalcitrant behaviour in vitro, there are tools available that enable the modification of the *Quercus suber* genome, even from selected adult trees (Álvarez et al. 2004; Pinto et al. 2002; Hernandez et al. 2003).

A practical and effective way to produce embryos as a starting material to obtain genetic modified plants for functional studies is somatic embryogenesis.

Although a genotype dependent response to somatic embryogenesis is observed in several species (Gaj 2004; Lopes et al. 2006; Vieitez et al. 2009; Vieitez et al. 2012; Corredoira et al. 2015; Florez et al. 2015; Correia et al. 2016; Martins et al. 2016), it is possible to develop customized protocols for certain genotypes of interest. In this context, preliminary studies to identify the best induction treatment for each selected genotype in order to develop a stable and characterized somatic cell line are required.

The present study reports the work involved in developing high somatic embryogenesis induction (SEI) from leaves of two selected groups of trees: high quality cork, and low quality cork producing trees. Antibiotic sensitivity assays were performed to obtain the essential information to transform somatic embryos with *Agrobacterium tumefaciens*. Cryopreservation procedures were also optimized and transgenic embryos were successfully stored and recovered.

3. Materials and Methods

3.1. Explant source and preparation

Branches up to 5 cm in diameter were collected from eleven adult cork oak trees growing in Cercal do Alentejo, Portugal, classified as low-quality cork producers (Ce 1, Ce 2, Ce 4, Ce 6) and in São Brás de Alportel, Portugal, classified as producers of high-quality cork (Al 1, Al 2, Al 3, Al 4, Al 5, Al 9, Al 10) in April, 2012. Shoot cuttings without lateral branches and leaves, 20 cm in length and between 2 and 4 cm in diameter were forced to sprout in a plant growth chamber, S600PLH (Aralab, Portugal) at 25±1ºC, 16 h photoperiod provided by OSRAM Lumilux cool-white fluorescent tubes (200 $\mu mol.m^{-2}.s^{-1}$) and 90-100% relative humidity. Plant material were sprayed weekly with a solution of 0.1g/L of proparmocarb hydrochloride, 0.75 g/L of Derosal (60% carbendazim) and 0.5 g/L of benomyl. Expanding leaves randomly collected from sprouts representing each genotype, 1-2 cm from the base to the apex were excised from the epicormic shoots and used as initial explants. Leaves were surface-sterilized in 70% ethanol for 30s, immersed in 0.3% sodium hypochlorite solution plus two drops of Tween 20 for 10 min, and finally rinsed four times in sterile distilled water.

3.2. Media composition

The induction of somatic embryogenesis was carried out following two different protocols, T1 and T2, described by Hernández et al. (2003) and Pinto et al. (2002), respectively. Three culture media were used in T1 protocol: (1) pre-conditioning medium (PCM) consisted of half-concentration Gamborg PRL-4-C (Gamborg 1966) macronutrients and micronutrients, vitamins and Fe-EDTA from MS medium (Murashige and Skoog 1962). Culture medium PCM was supplemented with 1% sucrose and 0.6% agar. (2) primary induction medium (MSSH1) consisted of SH macronutrients (Schenk and Hildebrandt 1972), 3% sucrose and supplementation with 10 μM 6-benzilaminopurine (BAP) plus 50 μM 1-Naphthaleneacetic acid (NAA). (3) secondary medium (MSSH2) had the same composition of the PM except the composition of plant growth regulators that were reduced both to 0.5 μM.

The T2 induction protocol was employed by using E3 medium which consisted in MS macro and micronutrients, vitamins and Fe-EDTA, 3% sucrose, 3% gelrite and supplemented with the growth regulators 4.5 μM 2,4-Dichlorophenoxyacetic acid (2,4-D) and 9 μM zeatin.

Proliferation media used after T1 (MSSH) and T2 (MSWH) had the same composition of MSSH2 and E3, respectively, lacking growth regulators.

The pH of all media used was adjusted to 5.6 (for T1) or 5.8 (for T2) and autoclaved at 121ºC for 20 min.

3.3. Procedures and culture conditions

For induction protocols, four surface-sterilized leaves were placed (abaxial side down) in 9 cm diameter sterilized Petri dishes with 20 mL of medium and sealed with Parafilm. In T1, aseptic leaves were cultured on PCM in dark at 25±1 ºC for 7 days. Leaves were then transferred to MSSH1 and incubated in dark at 25±1 ºC for 30 days with a following transfer to MSSH2 with a 16 h photoperiod driven by cool-white fluorescent tubes (100-150 $\mu mol.m^{-2}.s^{-1}$) 25±1 ºC for 30 days. At a final stage of induction, explants were transferred to MSSH under the same conditions. The somatic clusters were finally isolated from leaves and subcultured every 30 days. In T2, aseptic leaves were cultured on E3 medium for 30 days in dark at 25±1 ºC. Explants were then transferred to a photoperiod of

16 h provided by cool-white fluorescent tubes (100-150 $\mu mol.m^{-2}.s^{-1}$), 25±1 ºC for 60 days. When embryogenic clusters were detected, embryos were detached from the explant and transferred to MSWH. Both somatic embryos and embryogenic clusters from the two treatments were subcultured every 30 days. The effect of the antibiotic kanamycin on the embryogenic proliferation was assayed on MSSH medium supplemented with the appropriate antibiotic concentration in dark at 25±1 ºC for 120 days.

3.4. Experimental design and data analysis of somatic embryogenesis induction

The influence of the induction treatment, cork quality producer trees group genotype and their interaction on the percentage of leaves producing somatic embryogenesis (% SEI) was studied. Induction treatments were divided in two phases, an earlier phase (EP) and late phase (LP). EP refers to day 37 of T1 and day 30 of T2, while LP corresponds to day 67 and day 60, respectively. % SEI was evaluated in these time points since they are the crucial steps for somatic embryogenesis induction. The influence of the genotype in % SEI was carried out with eleven genotypes, four from low quality cork producing trees (Ce1 to Ce4) and seven from high quality cork producing (Al1 to Al5, Al9 and Al10). At least 23 uncontaminated leaves per genotype and per replica were assembled in a triplicate experimental design.

Percentage data were analyzed after an arcsine transformation by a two-way ANOVA followed by Sidak's multiple comparisons test using GraphPad Prism version 6.0b for Mac OS X, GraphPad Software, La Jolla, CA, USA, www.graphpad.com. Circular layouts were generated with Circos tool (Krzywinski et al. 2009).

3.5. Determination of selective agent concentration

The effect of kanamycin on somatic embryos proliferation was assayed in the presence of different antibiotic concentrations (12.5, 25, 37.5, 50, 75, 100, 150 $mg.L^{-1}$). Embryos weights were measured every 20 days and relative weight increment (ΔRWI) was calculated as ΔRWI= (embryos weight - embryos $weight_{t-}$

$_1$)/ embryos weight$_{t-1}$, where t is the day of each subculture. All the assays were performed with triplicates, according with a complete randomized scheme.

3.6. Bacterial strains and binary plasmids

Overexpression binary vectors were obtained from Gateway™ technology (Invitrogen, Carlsbad, CA, USA). QsMYB1 opening reading frame (ORF) was amplified with the specific primers QsMYB1_attB1F and QsMYB1_attB2R to generate a QsMYB1 ENTRY clone (pENTRY_QsMYB1) after a BP clonase reaction with the pDONR221 vector (see primer list in Table II. 1S, Supplementary Material 1). A 3xFLAG nucleotide sequence was created by annealing the 3xFLAG_F_Olig and the 3xFLAG_R_Olig oligonucleotides (Table II. 1S, Supplementary Material 1). The generated 3xFLAG sequence was amplified with the 3xFLAGattB1F and the 3xFLAGattB2R primers (Table II. 1S, Supplementary Material 1) to generate a 3xFLAG ENTRY clone (pENTRY_QsMYB1). A vector harbouring a triple FLAG epitope (3xFLAG) sequence fused with QsMYB1 ORF was generated by PCR-fusion/Gateway™ cloning procedure for gene fusion (Atanassov et al. 2009) and recombined with the pK7WG2,0 destination vector by LR clonase, generating the overexpression vector designated further by pK7MYB1::3xFLAG (Figure II. 1S and Table II. 2S, Supplementary Material 1).

The resulting binary vector was mobilized into *A. tumefaciens* AGL1 by triparental mating (Ditta et al. 1980) using *E. coli* MC1012 harbouring the mobilizing plasmid pRK2013.

3.7. Embryo transformation

Cork oak genetic transformation was performed by the method described by Álvarez and Ordás (2007). In brief, *A. tumefaciens* AGL1 carrying the binary vector pK7MYB1::3xFLAG was grown at 28ºC on a orbital shaker at 250 rpm for 24 h in liquid YEP medium containing 50 mg.L^{-1} spectinomycin. The bacterial suspension was collected by centrifugation at 4,000 rpm, washed with 10 mM $MgSO_4$ and resuspended in MSSH medium. Embryogenic *calli* of Ce1 cell line were collected 20 days after subculture and inoculated in an AGL1 pK7MYB1::3xFLAG suspension ($OD_{600\ nm}$ = 0.5) for 20 min at 25ºC and

cocultivated in a dark chamber for 2 days at 25 ºC. Embryos were then washed with MSSH medium containing 500 mg.L^{-1} cefotaxime (Duchefa). The putative genetic modified embryogenic *calli* were further selected in MSSH medium containing 500 mg.L^{-1} cefotaxime and 25 mg.L^{-1} kanamycin and analysed by PCR in order to check the presence of *nptII*, *QsMYB1::3xFLAG* and *VirG* genes with primers referred in Table II. 3S of the Supplementary Material 1.

3.8. RT-qPCR transcript detection of QsMYB1::3xFLAG expression

Aleatory independent transgenic embryo clusters were selected for quantitative real time PCR (RT-qPCR) analysis in order to screen the expression of the QsMYB1::3xFLAG transcript. Specific primers, qPCRQsMYB1F and SR(C) (Table II. 3S, Supplementary Material 1) were used in the RT-qPCR analysis. One microgram of total RNA was used for cDNA synthesis using the QuantiTect Reverse Transcription kit (Qiagen, Valencia, CA, USA), which includes an additional genomic DNA elimination step and uses a mix of oligo(dT) and random hexamer primers. RT-qPCR experiments were carried out in a iCycler iQ5 Instrument (Bio-Rad Laboratories, Hercules, CA, USA) using the Sso Advanced Universal SYBR Green Master mix (Bio-Rad Laboratories, Hercules, CA, USA) in 96-well plates. Three replicates were performed in reaction mixtures of 20 µL containing 10 µL of 2X Sso Advanced Universal SYBR Green Master mix, 400 nM of each specific primer pair (Forward/Reverse) and 1 µL from a 1:100 dilution of original cDNA as template. Normalization between samples was performed using two reference genes: ACTIN (ACT) and CLATHRIN ADAPTOR COMPLEXES (CACs). Normalized relative quantities (NRQ) were calculated by $\mathrm{NRQ} = \frac{\mathrm{E}_{\mathrm{goi}}^{\Delta \mathrm{Ct,goi}}}{\sqrt[\mathrm{f}]{\prod_{\mathrm{o}}^{\mathrm{f}} \mathrm{E}_{\mathrm{ref_o}}^{\Delta \mathrm{Ct,ref_o}}}}$ where E is the amplification efficiency for each primer pair, f the number of reference genes used to normalize the data, goi the gene of interest, ref the reference gene and $\Delta\mathrm{Ct}$ is the Ct of the sample with higher Ct across samples minus the Ct value of the sample in test (Hellemans et al. 2007).

3.9. Embryos encapsulation and cryopreservation

Embryo clusters were loaded in liquid MSSH supplemented with 3% (w/v) sodium alginate low viscosity and 0.5 M sucrose. The mixture was dripped with

a sterile pipette into 0.1 M calcium chloride solution composed of liquid MSSH with 0.1 calcium chloride and 0.5 M sucrose for 20 min, under constant swirling. This spherification process was repeated once again for each embryo cluster in order to achieve a double spherification. Beads were then incubated for three days in liquid MSSH medium sucrose-enriched (0.7M), followed by a desiccation process by blotting the embryo beads in a cellulose filter paper and placed in a horizontal laminar flow chamber for dehydration. Mass loss was monitored and the water content calculated according to Verleysen *et al*. (2005). Embryo beads were then transferred to 2 mL cryopreservation vials and directly stored in liquid nitrogen. For rehydration and regrowth of the cryopreserved embryos, the cryogenic vials were rapidly removed from liquid nitrogen and immediately rewarmed at 38ºC in a water bath for 3 min. Embryo beads were then rehydrated in liquid MSSH medium supplemented with 1M sucrose for 1h, and placed on MSSH medium for embryo proliferation.

4. Results and Discussion

4.1. Influence of treatment type, tree type and genotype in somatic embryogenesis

Treatments T1 and T2 were applied to cork oak leaves to induce somatic embryogenesis from adult cork oak trees. Two representative populations characterized by low and high quality cork producing trees were assessed. In general, treatment T1 was more efficient than T2 inducing SE [$F(1, 128)=13.77$, $p<0.003$, $\alpha=0,05$] (Figure II. 1 and Table II. 1). Regarding the two induction periods where the number of somatic masses were observed after 30 and 60 days (Figure II. 2) of treatment initiation, clearly there are differences in the time of induction of several genotypes [$F(1, 128)=23.08$, $p<0.0001$, $\alpha=0,05$] (Figure II. 1 and Table II. 1) which is consistent with the results reported by Hernandez *et al.* (2003a). These results indicate that some genotypes need more time to respond to SEI. This finding is an important information for future propagation of *Quercus suber* by somatic embryogenesis and constitute a crucial parameter for commercial and industry purposes since a shorter time of induction may reduce the coast of production.

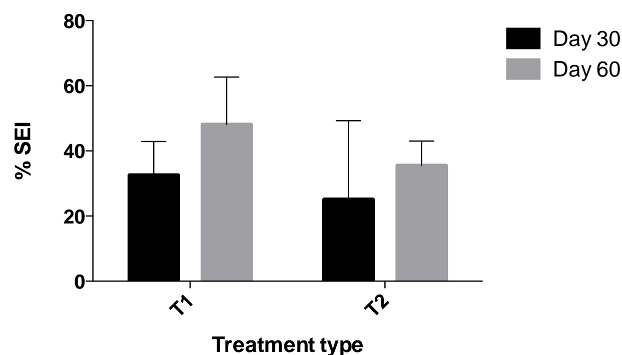


Figure II. 1 | Effect of treatment type and period of induction on somatic embryogenesis induction percentage (% SEI) from non-contaminated leaves after 30 days and 60 days of treatment initiation. Data from genotypes that do not show somatic embryos after 60 days were excluded. Analyzed data refers to % SEI ± SD of 3 independent experiments per genotype.

Table II. 1 | Analysis of variance of % SEI per two treatments for the two periods of induction for a significance level of 0.05.

Source	*df*	MS	*F ratio*	*P value*
Treatment type (*Tt*)	1	3.317	F (1, 128) = 13.77	P < 0.0003
Time (*Ti*)	1	5.560	F (1, 128) = 23.08	P < 0.0001
Tt × Ti	1	211.4	F (1, 128) = 0.8774	P < 0.3507
Residual	128	240.9		

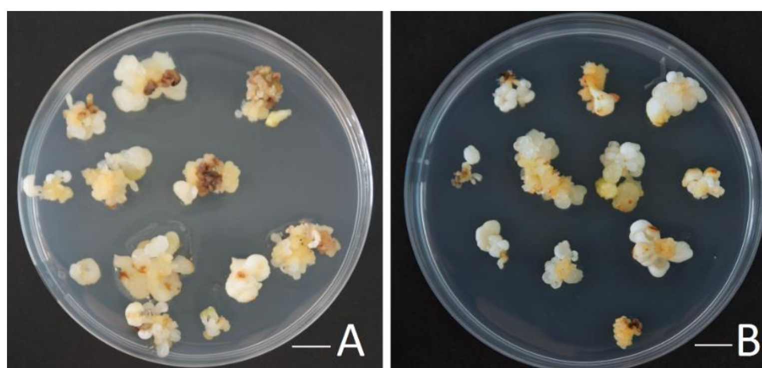


Figure II. 2 | Embryo clusters inducted after 30 day of treatment initiation for Ce1 cell line (A), and after 60 days of treatment initiation for Al3 cell line (B). bars = 1 cm.

Leaf explants from high quality cork trees have responded better to both treatments [F(1, 80)=43.09, p<0.0001, α=0,05] (Figure II. 3 and Table II. 2). The genotype Al3 showed the highest SEI value (82±6 %), although a great variance was observed between the different treatments and the genotypes studied, which justifies a more accurate analysis to understand the best treatment conditions for each genotype. In fact, SEI is also affected by the type of treatment [F(1,

80)=7.601, p=0.0072, α=0,05] (Figure II. 3 and Table II. 2) applied to different genotypes. Both factors, treatment and genotype affect significantly the % SEI but a trend for the high quality producing trees genotypes to respond better to SEI is observed. This is reasonable, and may be explained by the fact of that individuals belonging to the two quality cork groups are also from two distinct locations and cultivated in different agroforestry systems which may possibly contribute for the response variance observed.

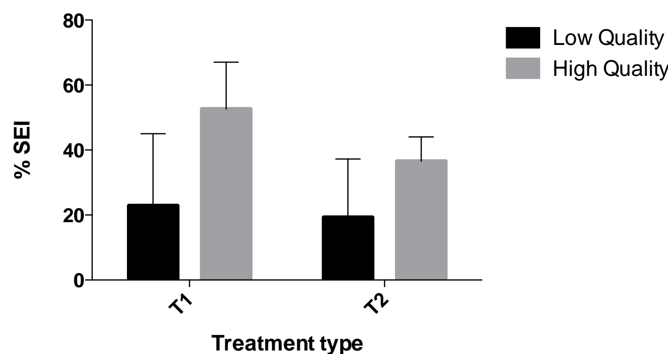


Figure II. 3 | Effect of treatment type and cork quality group on % SEI from non-contaminated leaves. Low quality group comprises all Ce genotypes. High quality group contain Al genotypes. Analysed data refers to % SEI ± SD of 3 independent experiments per genotype after 67 days of treatment.

Table II. 2 | Analysis of variance of SEI per two treatments by quality group, for a significance level of 0.05.

Source	*df*	MS	*F ratio*	*P value*
Treatment type (*Tt*)	1	2.044	F (1, 80) = 7.601	P = 0.0072
Cork Quality Group (Q)	1	11.586	F (1, 80) = 43.09	P < 0.0001
Tt × *Q*	1	812.6	F (1, 80) = 3.022	P = 0.0860
Residual	80	268.9		

Furthermore, we analysed if the treatment type affected differently the SEI for each individual tree genotype (Figure II. 4). T1 and the T2 treatment affect significantly the % SEI [F(1, 44)=195.0 p<0.0001, α=0,05] (Table II. 3) for each single genotype. Also, there are differences in SEI associated to the genotype [F(10, 44)=43.49 p<0.0001, α=0,05] (Table II.3), which may be explained as a specific genotype predisposition to SEI. With these findings we can conclude that the frequency of SEI is highly genotype dependent in cork oak as previously reported for other cork oak individuals (Pinto et al. 2002; Hernández et al. 2003;

Concepción Sánchez et al. 2003; Vieitez et al. 2012) In addition, the same treatment revealed to affect differently each genotype (Figure II. 4). In fact, there is an interaction between Genotype and Treatment factors demonstrating that these factors do not completely explain the source of variation individually. Therefore, we reinforce the idea that each genotype may be analysed independently to choose the best treatment for SEI.

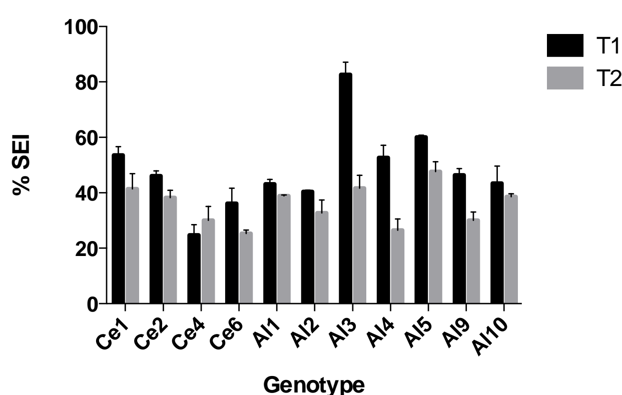


Figure II. 4 | Effect of treatment type per genotypes on % SEI from non-contaminated leaves. Data from genotypes that do not show somatic embryos after 60 days were excluded. Analysed data refers to % SEI ± SD of 3 independent experiments per genotype after 67 days of treatment.

Table II. 3 | – Analysis of variance of SEI per genotype by quality group, for a significance level of 0.05.

Source	*df*	MS	*F ratio*	*P value*
Genotype (*G*)	10	580.3	F (10, 44) = 43.49	P < 0.0001
Treatment type (*Tr*)	1	2601	F (1, 44) = 195.0	P < 0.0001
G × *Tr*	10	225.8	F (10, 44) = 16.92	P < 0.0001
Residual	80	268.9		

This finding is also in agreement with the work of Hernández *et al.*, (2003b), which observed that different plant growth regulators may be necessary to obtain high rates of SEI for some genotypes.

The relationship between the response of genotypes and the treatments applied is represented in Figure II. 5. While genotype Al3 responded better to T1 (67%), genotype Ce4 have a high % SEI with T2, however, there are genotypes that have a similar treatment response to SEI, as Ce1 or Al10. Globally these results show that for several genotypes there are induction treatments that increase the % SEI, however both T1 and T2 treatments are capable to induce somatic embryogenesis from adult material. To achieve the best yields of SEI it

is convenient to determine for each independent genotype the best treatment conditions since some variation among different genotypes is observed. Previous studies have already pointed to the possible influence of the genetic background of *Quercus suber* on its morphogenic capacity (Wilhelm 2000), and our results are in accordance with this.

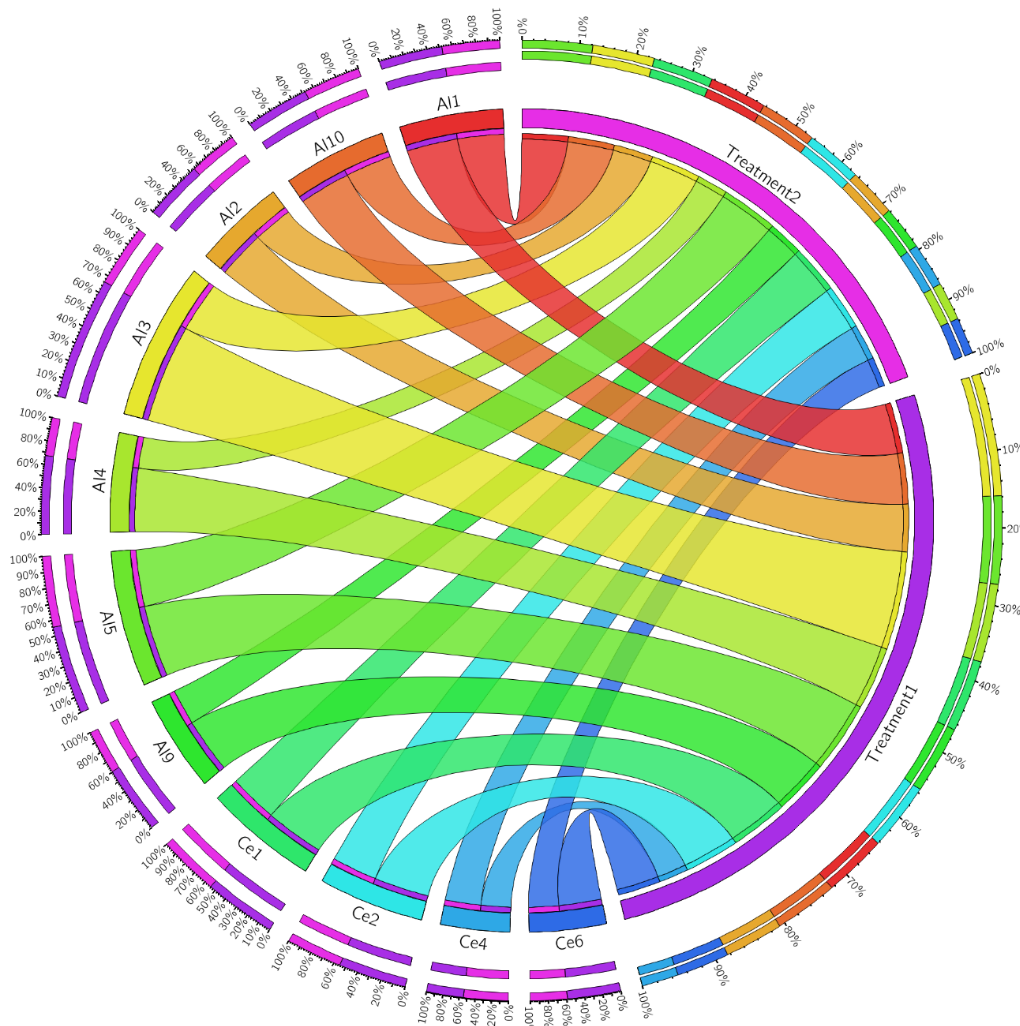


Figure II. 5 | Relation between genotype and treatment in % SEI. Each genotype is linked to the both treatments by data show on Figure II. 3 (mean). Relative quantification % SEI are showed in the outer semi-circle.

4.2. Kanamycin as selective agent of genetic modified cork oak somatic embryos

In order to determine the effect of kanamycin in somatic embryos induction and development for its future use as selective agent in transformation process, the Al1 and Ce1 somatic embryogenic lines were selected due to their higher rates of proliferation (data not shown). Embryo masses were cultured on MSSH

culture medium supplemented with different concentrations of kanamycin (Figure II. 6). Kanamycin inhibits the proliferation of somatic embryos in both Al1 and Ce1 cell lines. In the case of Al1, a relative higher concentration of kanamycin (>50 mg/L) is necessary to inhibit the rate of proliferation. In addition, this effect is just effective after 80 days of culture. For Ce1 line, the concentration of the antibiotic necessary to inhibit cell proliferation is lower (<25mg/L), and inhibition of cell proliferation is observed after the 20th day on culture which indicates that transformants could be selected after this period. Therefore, Ce1 line was selected for further transformation with the pK7MYB1::3xFLAG, which carries the *nptII* gene conferring resistance to kanamycin to the transformants.

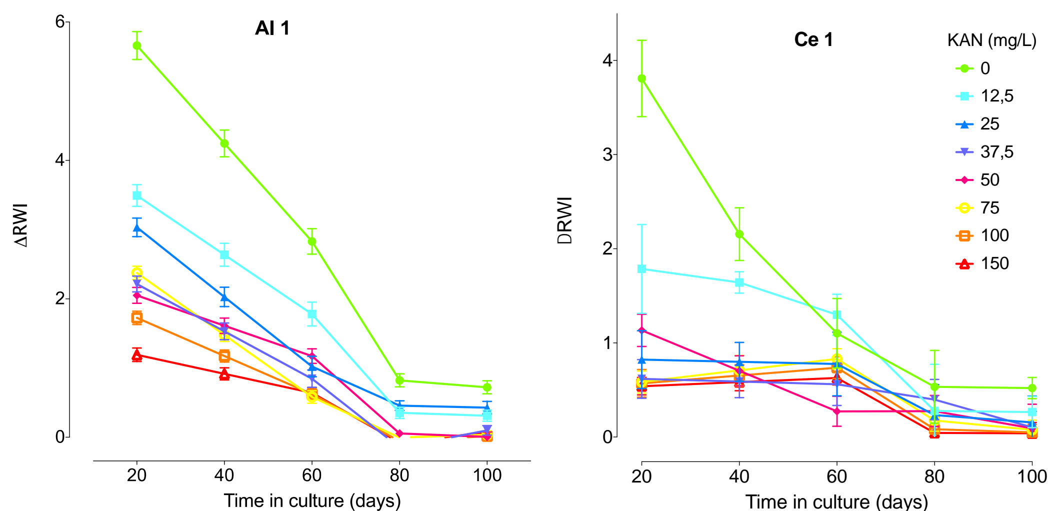


Figure II. 6 | Proliferation of cork oak somatic embryos in presence of 0, 12.5, 25, 37.5, 50, 75, 100, 150 mg.L-1 kanamycin. Marks represent the relative fresh weight increment [(ΔRWI= (embryos weight - embryos weightt-1)/ embryos weightt-1, where t is the day of each subcultivation)]. Mean and standard errors of three independent experiments are shown.

A concentration of 25 mg/mL of kanamycin was used for selection of genetic modified embryos. This work also confirms previous studies developed with three Spanish genetic provenances which had already pointed the need to determine sensitivity individually to the selective agent in each genotype. (Álvarez and Ordás 2008).

4.3. Effective transformation of cork oak embryogenic cell lines

The *Agrobacterium* strain AGL1 was used for embryos clusters transformation. Embryo clusters overexpressing the QsMYB1::3xFLAG were cultured in proliferation medium containing the selective agent which promoted recurrent embryogenesis. The genetic modified embryos clusters were selected during two years. The presence of the *nptII* gene and the QsMYB1::3xFLAG nucleotide sequence in the embryos DNA, assured in a first phase the incorporation of the transfer DNA (T-DNA) in the embryos genome (Figure II. 7B).

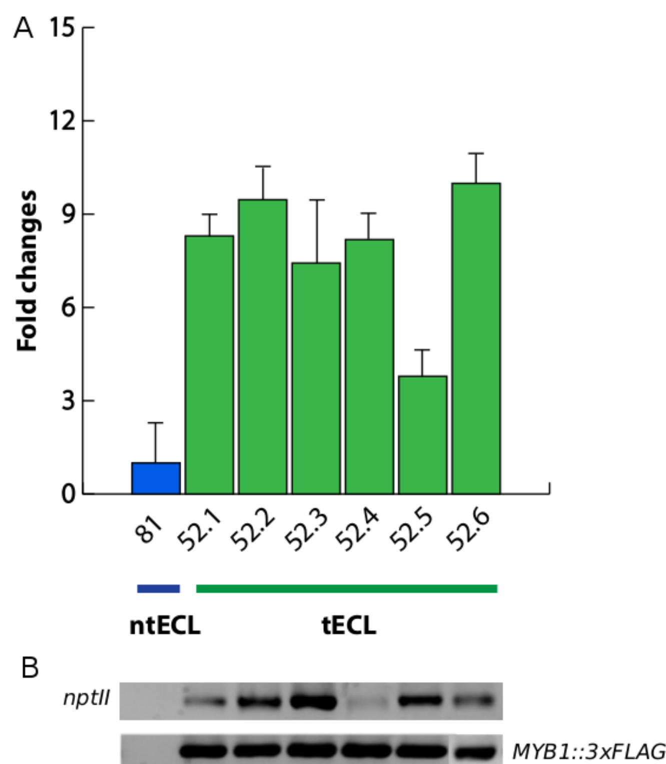


Figure II. 7 | (A) MYB1::3xFLAG transcripts levels from six transformed embryogenic cell lines (tECL) clusters (52.1-52.6) quantified by RT-qPCR and normalised against non-transformed embryogenic cell lines (ntECL). Mean and standard errors of three independent experiments are shown. **(B)** PCR products for ntpII and MYB1::3xFLAG.

The levels of expression of the QsMYB1::3xFLAG were then analysed by RT-qPCR in order to understand the differences among embryos in terms of QsMYB1::3xFLAG DNA integration and expression. The results identified a diversity of relative expression values lines (Figure II. 7A), which is typical in plant genetic modified tissue transformed by *Agrobacterium* (Álvarez et al. 2004). Several factors can explain this variation, the most obvious being the copy

number effect (Cervera et al. 2000; Álvarez et al. 2009), but other phenomena such as the position effect of the insertion of transgene into the host genome (Matzke and Matzke 1998) or complex configurations of the integrated T-DNA (Stam et al. 1997) should be considered. Together the results showed the establishment of cork stable genetic modified cell lines overexpressing the *QsMYB1::3xFLAG* construct for at least 2 years, through secondary embryogenesis, which allowed the use of different genetic modified embryo clusters in further work, according to the level of *QsMYB1::3xFLAG* expression (Chapter IV).

4.4. Storage of embryogenic cell lines for future use

One of the best methods to preserve embryogenic cell lines is by cryopreservation. *Quercus suber* embryos have been successfully cryopreserved after dehydration and encapsulation (Fernandes et al. 2008) or after vitrification (Valladares et al. 2004) with relative high efficiency. We optimized the method described by Verleysen et al. (2005) for cork oak embryos performing simple (Figure II. 8A) and double spherification (Figure II. 8B) with an alginate matrix.

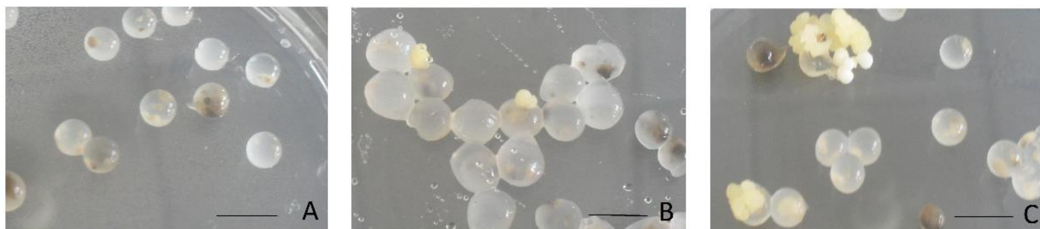


Figure II. 8 | Encapsulated embryos in alginate matrix with simple **(A)** and double spherification **(B)** before cryopreservation. Rehydrated embryos after 6 months stored in liquid nitrogen and 10 days growing in MSSH medium **(C)**. bars = 1 cm.

After cryopreservation for more than 6 months the cryopreserved embryos were collected from nitrogen, rewarmed and rehydrated (recovered). Successful proliferation was observed for 96.67% of the embryos with simple spherification, with no distinct morphology differences from the original Ce1 embryo line (Table II. 4, Figure II. 8C). In the case of embryos cryopreserved with double spherification the results are also very satisfactory with a rate of 85.56% of embryo proliferation (Table II. 4). The results of this optimized protocol show that

is possible to cryopreserve *Quercus suber* embryos with a simple and efficient method as in other *Quercus* family members as described by (Fernandes et al. 2008). Both, double and simple spherification are suitable processes to protect the embryos; simple spherification should be considered for industrial proposes since it involves less manipulation, less material and less time consumption.

Table II. 4 | Number of cryopreserved embryos; recovered; proliferating after 30 days in MSSH medium; and respective rate of recovery for the two spheferification methods used.

	Number of embryos			
	Cryopreserved	*Recovered*	*Proliferating*	***% of recovery***
Single spherification	90	88	87	96.67%
Double spherification	90	90	77	85.56%

5. Conclusion

Somatic embryogenesis constitutes an excellent choice to produce large-scale cork oak embryos. In addition, it allows the production of genetic modified embryos which may be used to perform functional studies in species with a long-life cycle such as cork oak. With this work we achieved an important goal for the study of *Quercus suber* biology, by describing an efficient method to generate and store genetic modified embryos. In addition, in the present work several somatic embryogenesis cell lines were produced which are available as a functional tool to other research groups, therefore shortening the required timeframe to achieve embryos. These lines therefore constitute a valuable resource for the scientific community allowing the realization of functional and biochemical studies in a non-model and unique species such as *Quercus suber*. The generation of transgenic lines, RNAi silenced lines as well as the development of ChIP-Seq assays for detection of transcription factors binding sites are now possible in cork oak by the model presented in our study.

6. REFERENCES

Álvarez R, Alonso P, Cortizo M, Celestino C, Hernández I, Toribio M, et al. (2004) Genetic transformation of selected mature cork oak (Quercus suber L.) trees. Plant Cell Rep 23:218–223

Álvarez R, Álvarez JM, Humara JM, Revilla A, Ordás RJ (2009) Genetic transformation of cork oak (Quercus suber L.) for herbicide resistance. Biotechnol Lett 31:1477–83

Álvarez R, Ordás RJ (2007) Improved genetic transformation protocol for cork oak (Quercus suber L.). Plant Cell Tissue Organ Cult 91:45–52

Aronson J, Pereira JS, Pausas JG (2009) Cork Oak Woodlands on the Edge: Ecology, Adaptive Management, and Restoration. Island Press, Washington, DC

Atanassov II, Atanassov II, Etchells JP, Turner SR (2009) A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. Plant Methods 5:14

Cervera M, Pina J a., Juárez J, Navarro L, Peña L (2000) A broad exploration of a transgenic population of citrus: stability of gene expression and phenotype. TAG Theor Appl Genet 100:670–677

Concepción Sánchez M, Teresa Martínez M, Valladares S, Ferro E, Viéitez AM, Sanchez MC, et al. (2003) Maturation and germination of oak somatic embryos originated from leaf and stem explants: RAPD markers for genetic analysis of regenerants. J Plant Physiol 160:699–707

Corredoira E, Ballester A, Ibarra M, Vieitez AM (2015) Induction of somatic embryogenesis in explants of shoot cultures established from adult Eucalyptus globulus and E. saligna × E. maidenii trees. Tree Physiol 35:678–690

Correia SI, Alves AC, Veríssimo P, Canhoto JM (2016) Somatic embryogenesis in broad-leaf woody plants: What we can learn from proteomics. In: Methods in Molecular Biology. pp 117–129

Fernandes P, Rodriguez E, Pinto G, Roldán-Ruiz I, De Loose M, Santos C (2008) Cryopreservation of Quercus suber somatic embryos by encapsulation- dehydration and evaluation of genetic stability. Tree Physiol 28:1841–1850

Florez SL, Erwin RL, Maximova SN, Guiltinan MJ, Curtis WR (2015) Enhanced somatic embryogenesis in Theobroma cacao using the homologous BABY BOOM transcription factor. BMC Plant Biol 15:121

Gaj MD (2004) Factors influencing somatic embryogenesis induction and plant regeneration with particular reference to Arabidopsis thaliana (L.) Heynh. Plant Growth Regul. 43:27–47

Gamborg OL (1966) Aromatic Metabolism in Plants: II. Enzymes of the Shikimate Pathway in Suspension Cultures of Plant Cells. Can J Biochem 44:791–799

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 Biol 8:R19

Hernández I, Celestino C, Alegre J, Toribio M (2003) Vegetative propagation of Quercus suber L. by somatic embryogenesis. II. Plant regeneration from selected cork oak trees. Plant Cell Rep 21:765–70

Hernandéz I, Celestino C, Toribio M (2003) Vegetative propagation of Quercus suber L. by somatic embryogenesis. I. Factors affecting the induction in leaves from mature cork oak trees. Plant Cell Rep 21:759–764

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. (2009) Circos: An information aesthetic for comparative genomics. Genome Res 19:1639–1645

Lopes T, Pinto G, Loureiro J, Costa A, Santos C (2006) Determination of genetic stability in long-term somatic embryogenic cultures and derived plantlets of cork oak using microsatellite markers. Tree Physiol 26:1145–1152

Martins JF, Correia SI, Canhoto JM (2016) Somatic embryogenesis induction and plant regeneration in strawberry tree (Arbutus unedo L.). In: Methods in Molecular Biology. pp 329–339

Matzke AJM, Matzke MA (1998) Position effects and epigenetic silencing of plant transgenes. Curr Opin Plant Biol 1:142–148

Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiol Plant 15:473–497

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

Pereira H (2007a) The formation and growth of cork. In: Cork: Biology, Production and Uses, 1st edn. Elsevier, Amsterdam, pp 7–31

Pereira H (2007b) The extraction of cork. In: Pereira H (ed) Cork: Biology, Production and Uses, 1st edn. Elsevier, Amesterdam, pp 127–144

Pinto G, Valentim H, Costa A, Castro S, Santos C (2002) Somatic embryogenesis in leaf callus from a mature Quercus suber L. Tree. Vitr Cell Dev Biol - Plant 38:569–572

Ramos AM, Usié A, Barbosa P, Barros P, Capote T, Chaves I, et al. The draft genome sequence of cork oak. Sci Data

Schenk RU, Hildebrandt AC (1972) Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. Can J Bot 50:199–204

Stam M, Mol JNM, Kooter JM (1997) The Silence of Genes in Transgenic Plants. Ann

Bot 79:3–12

Valladares S, Toribio M, Celestino C, Vieitez AM (2004) Cryopreservation of embryogenic cultures from mature Quercus suber trees using vitrification. Cryo-Letters 25:177–186

Verleysen H, Van Bockstaele E, Debergh P (2005) An encapsulation-dehydration protocol for cryopreservation of the azalea cultivar “Nordlicht” (Rhododendron simsii Planch.). Sci Hortic (Amsterdam) 106:402–414

Vieitez AM, Corredoira E, Ballester A, Muñoz F, Durán J, Ibarra M (2009) In vitro regeneration of the important north american oak species quercus alba, quercus bicolor and quercus rubra. Plant Cell Tissue Organ Cult 98:135–145

Vieitez AM, Corredoira E, Martínez MT, San-José MC, Sánchez C, Valladares S, et al. (2012) Application of biotechnological tools to Quercus improvement. Eur J For Res 131:519–539

Wilhelm E (2000) Somatic embryogenesis in oak (Quercus spp.). Vitr Cell Dev Biol - Plant 36:349–357

7. Supplementary Material

7.1. Supplementary Material 1

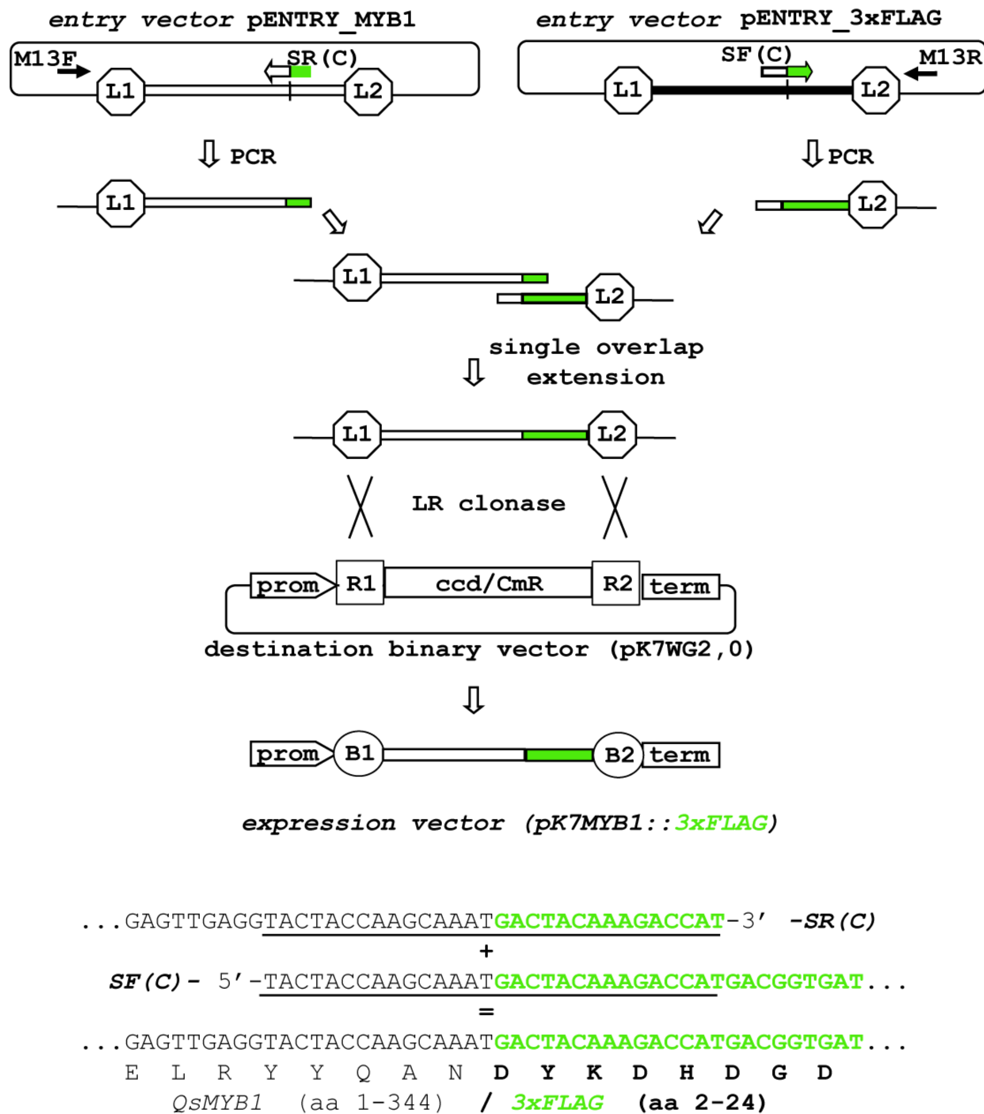


Figure II. 1S | Gene construction strategy used for QsMYB::triple FLAG epitope fusion protein production.

Table II. 1S | Primers used to generate the pENTRY_MYB1 vector and the pENTRY_3xFLAG vector pK7MYB1::3xFLAG

Primer	Sequence
QsMYB1attB1F	5′-G GGG ACA AGT TTG TAC AAA AAA GA GGC TTA GAA GGA GAT AGA ACC ATG GGG AGA GCT CCA TGT TGT GAC AAA G-3’
QsMYB1attB2R	5´-GGG GAC CAC TTT GTA CAA GAA AGC TGG GTC TTA ATT TGC TTG GTA GTA CCT CAA CTC T -3’
3xFLAG_F_Olig	5’-AGC TTA GAC TAC AAA GAC CAT GAC GGT GAT TAT AAA GAT CAT GAC ATC GAT TAC AAG GAT GAC GAT GAC AAG TGA TAT CG-3’
3xFLAG_R_Olig	5’-GAT CCG ATA TCA CTT GTC ATC GTC ATC CTT GTA ATC GAT GTC ATG ATC TTT ATA ATC ACC GTC ATG GTC TTT GTA GTC TA-3’
3xFLAGattB1F	5’- GGG GAC AAG TTT GTA CAA AAA AGC AGG CTT CGA AGG AGA TAG AAC CAT GGA CTA CAA AGA CCA TGA CGG TGA-3’
3xFLAG1attB2R	5' – GGG G ACC ACT TTG TAC AAG AAA GCT GGG TCT CAC TTG TCA TCG TCA TCC TTG TAG-3’

Table II. 2S | Primers used to generate the overexpression destination vector pK7MYB1::3xFLAG

Primer	Sequence
M13F	5’-GTA AAA CGA CGG CCA GT-3’
M13R	5’-CAG GAA ACA GCT ATG AC-3’
SF(C)	5’-TAC TAC CAA GCA AAT GAC TAC AAA GAC CAT-3’
SR(C)	5’-ATG GTC TTT GTA GTC ATT TGC TTG GTA GTA-3’
MYB1ns_attB1F	5’- AAA AAG CAG GCT TAG AAG GAG ATA GAA CCA TGG GGA GAG CTC CAT GTT GTG ACA AAG -3’
MYB1ns_attB2R	5′- AGAAAGCTGGGTCATTTGCTTGGTAGTACCTCAACTCT -3’
attB1 adapter	5′- G GGG ACA AGT TTG TAC AAA AAA GCA GGC T -3′
attB2 adapter	5′- G GGG ACC ACT TTG TAC AAG AAA GCT GGG T -3′

Table II. 3S | Primers used to confirm the integration of the foreign DNA delivered by the destination vector plasmid.

Primer	Sequence
qPCRQsMYB1F	5'-AGCCTAAAGCAAGAGATGAAGAGAG-3'
SR(C)	5'-ATG GTC TTT GTA GTC ATT TGC TTG GTA GTA-3'
nptIIF	5'-GAGGCTATTCGGCTATGACTG-3'
nptIIR	5'-ATCGGGAGCGGCGATACCGTA-3'
virGF	5'-AAGGTGAGCCGTTGAAACAC-3'
virGR	5'-ATCTCAAGCCCATCTTCACG-3'

CHAPTER III

TRANSCRIPTOME DYNAMICS OF CORK OAK (*QUERCUS SUBER*) SOMATIC EMBRYOGENESIS REVEALS ACTIVE GENE PLAYERS IN TRANSCRIPTION REGULATION AND HORMONE HOMEOSTASIS OF EMBRYO DEVELOPMENT

Chapter submitted as original article in a SCI journal

Capote T, Usié A, Barbosa P, Ramos AM, Morais-Cecílio L, Gonçalves S. Transcriptome dynamics of cork oak (Quercus suber) somatic embryogenesis reveals active gene players in transcription regulation and hormone homeostasis of embryo development. Tree Genetics & Genomes. *Submitted*

TABLE OF CONTENTS – CHAPTER III

Transcriptome dynamics of cork oak (*Quercus suber*) somatic embryogenesis reveals active gene players in transcription regulation and hormone homeostasis of embryo development 85

1. Abstract 89

2. Introduction 90

3. Materials and Methods 92

 3.1. Plant material and RNA isolation 92

 3.2. Sequencing sample preparation 92

 3.3. Sequence data processing, assembly and protein prediction 93

 3.4. Read mapping and differential expression analysis 93

 3.5. Transcriptome Annotation 94

 3.6. GO enrichment 94

 3.7. Gene Clustering Analysis 94

 3.8. Quantitative RT-PCR analysis and RNA-Seq data correlation 94

 3.9. Identification of transcription mediators, hormone and embryogenesis-related genes 95

4. Results 96

 4.1. RNA-Seq analysis and global functional annotation 96

 4.2. Gene expression and differential gene expression analysis 100

 4.3. Differential gene expression validation by RT-qPCR 102

 4.4. GO analysis of differentially expressed genes 102

 4.5. Transcription factors, transcription regulators and chromatin regulator genes involved in cork oak somatic embryogenesis 105

 4.6. Expression profile of phytohormones homeostasis-related genes 108

 4.7. Expression profile of genes involved in germination 109

 4.8. Expression profiles of other embryogenesis related-genes 109

5. Discussion 111
5.1. Transcription regulators involved in cork oak somatic embryogenesis and somatic embryo development 112
5.2. Hormones-related genes involved in cork oak somatic embryogenesis and somatic embryo development 114
5.3. Genes involved in germination... 118
5.4. Genes involved in root development... 121
5.5. Other relevant genes in cork oak somatic embryogenesis.................. 123
6. Data Archiving 125
7. References 125
8. Supplementary Material.. 137
8.1. Supplementary Material 1 – Smear plots and Volcano plots................ 137
8.2. Supplementary Material 2 – GO slim annotations............................... 138
8.3. Supplementary Material 3 – Differential expression of transcription factors, transcription regulators and chromatin regulators 138
8.4. Supplementary Material 4 – Hormone-related genes 138
8.5. Supplementary Material 5 – Embryogenesis related genes................. 138
8.6. Supplementary Material 6 – Primers List and Details 138

1. ABSTRACT

Cork oak (*Quercus suber L.*) is one of the most important Mediterranean forest tree species. The last decades have been marked by a decline in this species. The implementation of breeding programs is fundamental to revert this trend. Somatic embryogenesis is the system of election for clonal propagation, constituting a valuable tool for embryo production and improved genotypes testing. In this study, the transcriptome of the cork oak somatic embryogenesis was characterized in four stages of development to identify relevant genes in the process and to understand the molecular and biochemical events occurring in each specific stage.

A total of 66,693 candidate coding regions were predicted from the generated *de novo* transcriptome assembly. Differential gene expression analysis identified 11,507 genes distributed in 30 clusters with distinct gene expression patterns and enriched in various biological process GO-terms. Results show 1,159 differential expressed genes coding for transcription regulators, namely transcription factors (76%) with important roles in embryogenesis, like orthologous of AINTEGUMENTA-like, PLETHORA, CYTOKIN RESPONSE FACTOR, GATA transcription factors and AUXIN RESPONSE FACTORs genes. Results also show 250 differential expressed phytohormone-related genes involved in important aspects of embryogenesis as tissue specification, differentiation and embryogenesis competence. Finally, it was identified a group of genes with functions in cellular protection and abiotic stress tolerance coding for LATE EMBRYOGENEIS ABUNDANT proteins.

The cork oak embryogenesis transcriptome characterization represents a tool for future biotechnological applications. Results provide a molecular insight into embryo development, establishing the basis for further research towards improvement of somatic embryogenesis in cork oak.

2. INTRODUCTION

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

In the last two decades, the increase in climate changes, exotic pathogens and pests translocation from ecosystems separated in the past have caused extensive decline and mortality of several trees species worldwide (Allen et al. 2010). One of the main tree genera affected by this event is *Quercus* (Braisier 1996). In the Iberian Peninsula, a severe decline affecting evergreen oak forests has been well reported since the 1980s (Brasier 1992; Braisier 1996), in which cork oak is largely representative. A breeding program for cork oak, where elite selected trees with selected phenotypes will constitute the basis of a reproductive population for reforestation, may constitute the solution to stop the extensive declined observed and invert this trend. However, future cork oak breeding programs are compromised by limited natural regeneration and recalcitrant seeds (Valladares et al. 2004). Moreover, the use of seeds for propagation does not allow propagation of the valuable phenotypes (Valladares et al. 2004), identified in adult trees.

Since the capacity for vegetative propagation of the *Quercus* genera through traditional methods is very low (Vieitez et al. 2012), somatic embryogenesis emerges as a valuable tool for integration in breeding programs. The induction of *Q. suber* somatic embryogenesis has been successfully achieved by distinct research groups (Bueno et al. 2000; Pinto et al. 2002; Pintos et al. 2008; Vieitez et al. 2012). In addition, the possibility to modulate somatic embryogenesis and embryo development allowed the development of functional genomic studies such as genetic transformation for herbicide resistance (Álvarez et al. 2009), gene silencing (Zhang et al. 2012) and gene overexpression (Mallón et al. 2014) or even precise genome editing by new emerged techniques like CRISPR/CRISPR-associated protein 9 (Cas9) system (Fan et al. 2015).

Somatic embryos follow similar developmental patterns as their zygotic counterparts. Embryo development and maturation requires the concerted action

of several signaling pathways integrating genetic, epigenetic and hormonal regulation (Gutierrez et al. 2007; Feng et al. 2010). Embryo development proceeds through a series of spatially and temporally regulated gene expression network and many genes have already been characterized (Gutierrez et al. 2007). Also, an effort to explain the key regulatory process involved in distinct developmental stages of *Q. suber* somatic embryogenesis has been made by a proteomic approach (Gomez-Garay et al. 2013). However, the precise molecular mechanisms involved in somatic embryo induction, maturation and germination in this species are still poorly known.

An assessment of the transcriptome of embryo development will contribute to the comprehension of genes involved in this process and may help to modulate the process of *in vitro* embryo production. The identification of gene expression patterns associated with specific stages of embryo development is critical to understand the molecular and biochemical basis characteristic of each specific stage of embryo development. The knowledge generated by such studies will contribute to enhance the quality of mature embryos, for use in forestation processes.

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

3. MATERIALS AND METHODS

3.1. Plant material and RNA isolation

Branches up to 5 cm in diameter were collected from one adult cork oak tree growing in Cercal do Alentejo, Portugal in April 2011. Segmented branches without lateral branches and leaves, 20 cm in length and between 2 and 4 cm in diameter were forced to sprout in a fitoclima S600PLH (Aralab) at 25±1ºC, 16 h photoperiod provided by cool-white fluorescent tubes (200 μmol.m^{-2}.s^{-1}) and 90-100% relative humidity. Expanding leaves with 1-2 cm from the base to the apex were excised from the epicormic shoots and used as initial explants for the induction of somatic embryogenesis, according to Álvarez *et al.,* 2004 (Álvarez et al. 2004). *Quercus suber* embryogenic clusters were grown in MSSH medium, containing MS micronutrients (Murashige and Skoog 1962), SH macronutrients (Schenk and Hildebrandt 1972) and supplemented with MS vitamins (Murashige and Skoog 1962), on a climatic chamber at 25ºC under a light/dark cycle of 16/8h. Somatic embryogenic clusters were subculture every 30 days. After 4 years in culture, embryos were selected and pooled according to the developmental stage: globular, heart/torpedo, cotyledonar and mature cotyledonar (Von Arnold et al. 2002). Total RNA was extracted from the 4 embryo stages described above using the AMBION RNAqueous® Total RNA Isolation kit (Thermo Fisher Scientific, Waltham, MA, USA) including the Plant RNA Isolation Aid solution (Thermo Fisher Scientific, Waltham, MA, USA), according to the manufacturer's instructions. RNA samples were treated with AMBION TURBO DNA-*free*™ Kit (Thermo Fisher Scientific, Waltham, MA, USA) for DNA removal. RNA samples were analysed by electrophoretic separation of nucleic acids in a Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA) and used for cDNA synthesis and library construction for sequencing when RIN>8.

3.2. Sequencing sample preparation

RNA-seq library preparation and sequencing was performed by Beijing Genomics Institute BGI (Hong-Kong, China). cDNA synthesis and paired-end library was prepared following the protocol of Illumina® TruSeq RNA Library Preparation kit and sequenced in one lane using the Illumina HiSeq™ 2000

platform, with a read length of 100nt. Raw reads were cleaned by removing adaptor sequences, empty reads and low quality sequences.

3.3. Sequence data processing, assembly and protein prediction

Illumina paired-end reads were pre-processed according to quality (minimum of 20) and length (minimum of 80) using a custom Perl script combined with Sickle (Joshi and Fass 2011) in order to remove or trim from the dataset reads with low average quality and reads containing undetermined nucleotides (N's). Pre-processed reads were assembled with the Trinity platform (Grabherr et al. 2011). Assembly statistics were generated by quast 2.3 (Gurevich et al. 2013). Protein sequences prediction was performed with Transdecoder (http://transdecoder.github.io) from the transcriptome assembly, which extracts the longest open reading frames (ORFs), and identifies ORFs with homology to known proteins using Blastp and hmmer against the Uniprot and Pfam databases producing a set of candidate coding regions.

3.4. Read mapping and differential expression analysis

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

3.5. Transcriptome Annotation

Transcriptome assembly and the whole set of DEGs were functionally annotated using Blastx against the non-redundant (nr) protein plant sequence database from NCBI. Results were loaded into Blast2GO (Conesa and Gotz 2008) to perform a protein domain and Gene Ontology (GO) annotation using the InterProScan (Zdobnov and Apweiler 2001; Jones et al. 2014) function. ANNEX Annotation (Annotation Argumentation) function was used to refine annotations (Conesa and Gotz 2008) and GOSlim (plants) was used to generate plant specific GO terms. All these results were exported and processed with custom Python scripts in order to be analysed with Cytoscape (Shannon et al. 2003) and graphically represented.

3.6. GO enrichment

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

3.7. Gene Clustering Analysis

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

3.8. Quantitative RT-PCR analysis and RNA-Seq data correlation

A set of 19 DEGs was analysed by Reverse transcription quantitative real-time PCR (RT-qPCR) in order to validate the expression profiles obtained with the bioinformatics analyses (Supplementary Material 6). Two micrograms of total RNA were synthesized to cDNA using the QuantiTect Reverse Transcription kit (Qiagen, Valencia, CA, USA), which includes an additional genomic DNA

elimination step and uses a mix of oligo(dT) and random hexamer primers. Specific primers were designed by Primer3Plus (Untergasser et al. 2007). RT-qPCR experiments were carried out in a iCycler iQ5 Instrument (Bio-Rad Laboratories, Hercules, CA, USA) using the Sso Advanced Universal SYBR Green Master mix (Bio-Rad Laboratories, Hercules, CA, USA) in 96-well plates. Three replicates were performed in reaction mixtures of 20 µL containing 10 µL of 2X Sso Advanced Universal SYBR Green Master mix, 400 nM of each specific primer pair (Forward/Reverse) and 1 µL of cDNA with a dilution of 1:100 as template. All the selected genes were amplified with the following PCR program with a single fluorescent reading taken at the end of each cycle: 95 ºC at 10 min, 45 cycles of 10 s at 95 ºC, 15 s at 60 ºC, 20 s at 60 ºC and 15 s at 72 s, except for *AINTEGUMENTA (ANT)*, *NUCLEAR CAP-BINDING PROTEIN SUBUNIT 1 (ABH1)* and *TRANSLATIONALLY-CONTROLLED TUMOR PROTEIN 1 (TCPC1)* for which the annealing temperature was 61 ºC. To distinguish specific from nonspecific products and primer dimers, a melting curve was obtained immediately after amplification. Normalization was performed using two reference genes: *ACTIN (ACT)* and *CLATHRIN ADAPTOR COMPLEXES (CACs)*. Normalized relative quantities (NRQ) were calculated by $NRQ = \frac{E_{goi}^{\Delta Ct,goi}}{\sqrt[f]{\prod_{o}^{f} E_{ref_o}^{\Delta Ct,ref_o}}}$ where E is the amplification efficiency for each primer pair, f the number of reference genes used to normalize the data, goi the gene of interest, ref the reference gene and ΔCt is the Ct of the sample with higher Ct across samples minus the Ct value of the sample in test (Hellemans et al. 2007). Expression values from RNA-Seq experiments were calculated by the edgeR package. Both RNA-Seq and RT-qPCR gene expression values were submitted to log2 transformation and compared by Pearson correlation.

3.9. Identification of transcription mediators, hormone and embryogenesis-related genes

To identify genes encoding for transcription factors (TFs), transcription regulators (TRs) and chromatin regulators (CRs), all DEGs were analysed with the PlantTFcat tool (Dai et al. 2013). For analysis and classification of hormone-related genes, all DEGs were compared with the protein sequences from the

Arabidopsis Hormone Database (Jiang et al. 2011) using Blastx (E-value ≤ 10^{-10}, % identity > 60, query coverage > 70). Other embryogenesis-related genes were identified manually searching for genes annotated with “embryo” and “embryogenesis” terms.

4. RESULTS

Somatic embryogenesis is a regeneration process that starts with the formation of proembryogenic masses (PEMs) followed by somatic embryo formation, maturation, desiccation and plant regeneration (Von Arnold et al. 2002). In this study we selected four distinct *Q. suber* embryo developmental stages (Figure III. 1A): 1) the globular stage (ST1), in which embryos start to emerge from the PEMs acquiring a radial symmetrical morphology; 2) the heart/torpedo stage (ST2), where emergence of the first discernible organs occurs, the two cotyledons, and a change in embryo morphology from radial to bilateral symmetry is visible; 3) the immature cotyledonar stage (ST3) where the cotyledons are well defined and most embryogenic tissue and organs are formed; 4) and the mature cotyledonar stage (ST4), in which embryos pass through a maturation phase where cellular processes that are associated with embryo germination occur and cellular expansion and embryo growth is observed.

4.1. RNA-Seq analysis and global functional annotation

A *de novo* transcriptome sequence assembly was generated to study *Q. suber* somatic embryo development and maturation. Total RNA was isolated from four different tissue stages of embryo development previously described and used to prepare four cDNA libraries, one for each embryo developmental stage. A total of 160,366,471 paired-end reads were generated, of which 156,088,387 were kept after pre-processing. The transcriptome assembly resulted in 143,960 contigs ranging from 500 bp to 16,793 bp and a N50 length of 1,734 bases.

The total number of contigs was analysed with TransDecoder in order to identify candidate coding regions. A total of 66,693 candidate coding regions within the assembled transcriptome were identified, from which 22,109 were co-expressed among the four embryo developmental stages (Figure III. 2). In addition, 671, 811, 1,014 and 1,480 genes were found exclusively present in ST1,

ST2, ST3, ST4, respectively. Also, there were 373 genes uniquely present in ST1 and ST2, 786 genes that only appeared in ST2 and ST3 and 1088 genes that were expressed only in ST3 and ST4.

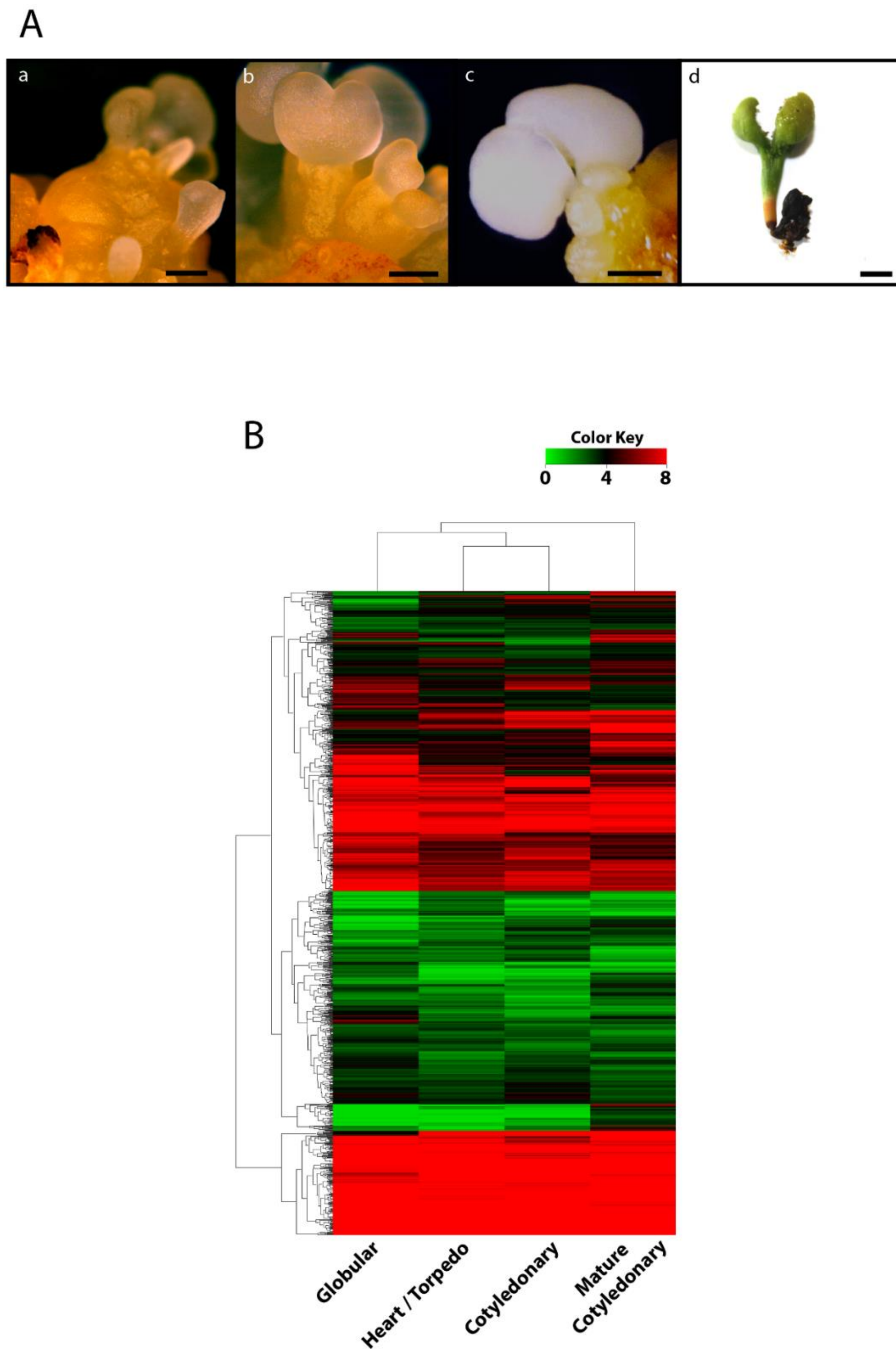


Figure III. 1 | (A) Morphology of cork oak somatic embryos in the different developmental stages used for RNA-Seq analysis. Somatic embryos at the globular (ST1) (a), heart/torpedo-shaped (ST2) (b), early cotyledonar (ST3) (c) and mature cotyledonar (ST4) (d) stages. **(B)** Hierarchical clustering of DE genes among libraries. Green indicates downregulated and red indicates upregulated genes. Bars = 0.5 mm in (a,b), 0.5 cm in (c,d).

After normalization and Log_2 transformation, a heatmap was generated to show a global overview of gene expression, during somatic embryo development (Figure III. 1B). Hierarchical clustering showed that ST2 and ST3 grouped closer to ST1 while ST4 clustered in a different group (Figure III. 1B). Moreover, multiple gene groups were clustered together showing similar gene expression patterns across developmental stages.

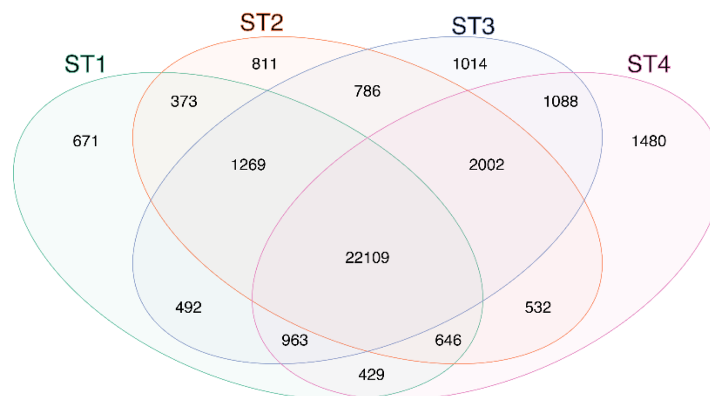


Figure III. 2 | Venn diagram of all expressed genes between the four developmental stages.

Functional categories were assigned to all genes in terms of gene GO. A total of 29,725 genes (44.57 %) were assigned to one or more plant GO-slim categories. From those, 18,360 (61.77%) were involved in biological processes, 24,712 (83.14%) with molecular functions, and 14,088 (47.40%) with cellular components. Also, a total of 8,558 (28.80%) genes were annotated to the three main GO categories. Regarding biological processes, the main represented GO categories at level 2 are cellular process (31.73%) and metabolic process (28.28%). Other well represented biological process such as single-organism process (8.77%), response to stimulus (5.68%), cellular component organization or biogenesis (4.87%), localization (4.20%), multicellular organismal process (3.63%) and developmental process (3.49%) were also identified (Figure III. 3). In terms of the molecular function categories, the GO terms at level 3 associated with the annotated genes reveals that heterocyclic compound binding (23.08%), organic cyclic compound binding (23.08%), transferase activity (17.11%), small molecule binding (16.34%), hydrolase activity (13.46%), protein binding (3.01%) and carbohydrate binding (0.94%) are the most represented GO terms (Figure III. 3). The cellular components associated-GOs at level 4, reveals that most of the genes belong to intracellular membrane-bounded organelle (28.75%), cytoplasm (28.28%), cytoplasmic part (23.73%), intracellular non-membrane-

bounded organelle (4.25%), thylakoid (2.69%), cell wall (2.46%) and Golgi apparatus (2.32%) components (Figure III. 3).

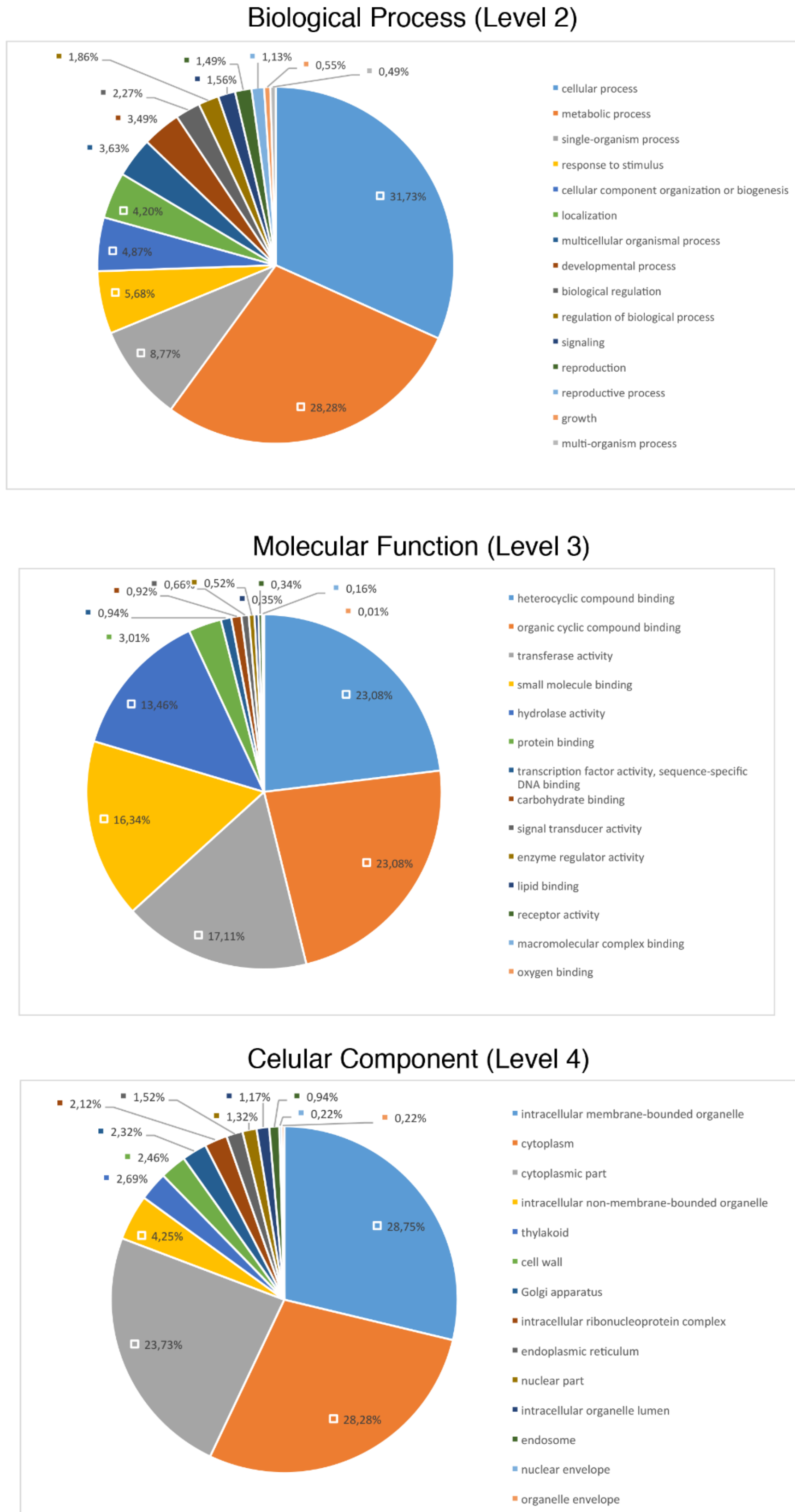


Figure III. 3 | - Gene Ontology functional classification of all expressed genes on biological process at level 2, molecular function at level 3 and cellular localization at level 4.

4.2. Gene expression and differential gene expression analysis

The unique mapped read reads of each cDNA library were used for gene expression analysis by the featureCounts R function. The analysis identified 11,507 DEGs (FDR<0.01) between all developmental stages. The list includes all genes that were differentially expressed when compared ST1 vs ST2 (ST12), ST2 vs ST3 (ST23), ST3 vs ST4 (ST34) and non-sequential development stages (Figure III. 4). Smear plots of the Log(Fold Change) vs Average Log(Count per million) and the respective volcano plots shows the distribution of the DGEs amongst the sequential developmental stages (Supplementary Material 1). In ST12 the analysis identified 2,995 up and 2,473 down-regulated genes (Figure III. 4). In ST23, 2,265 and 2,420 up and down-regulated genes were found, respectively (Figure III. 4). Finally, in ST34 the total number of DGEs is higher when compared with ST12 and ST23, accounting 3,565 genes up-regulated and 3,164 down-regulated (Figure III. 4).

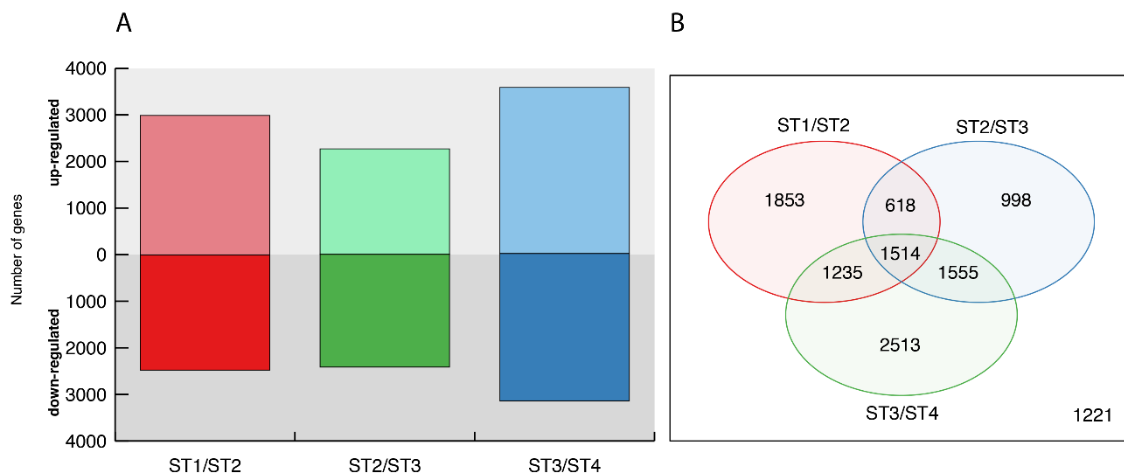


Figure III. 4 | Down-regulated and Up-regulated genes **(A)** and Venn diagram showing the overlaps of differentially expressed genes among three comparisons of developmental stages pairs **(B)**.

In order to understand the patterns of gene expression during somatic embryo development a hierarchical clustering analysis with K-means method using Euclidean Distance after Log_2 expression values transformation was performed. K-means clustering revealed 30 clusters with different expression patterns containing between 121 and 743 significantly regulated genes for the four developmental stages (Figure III. 5).

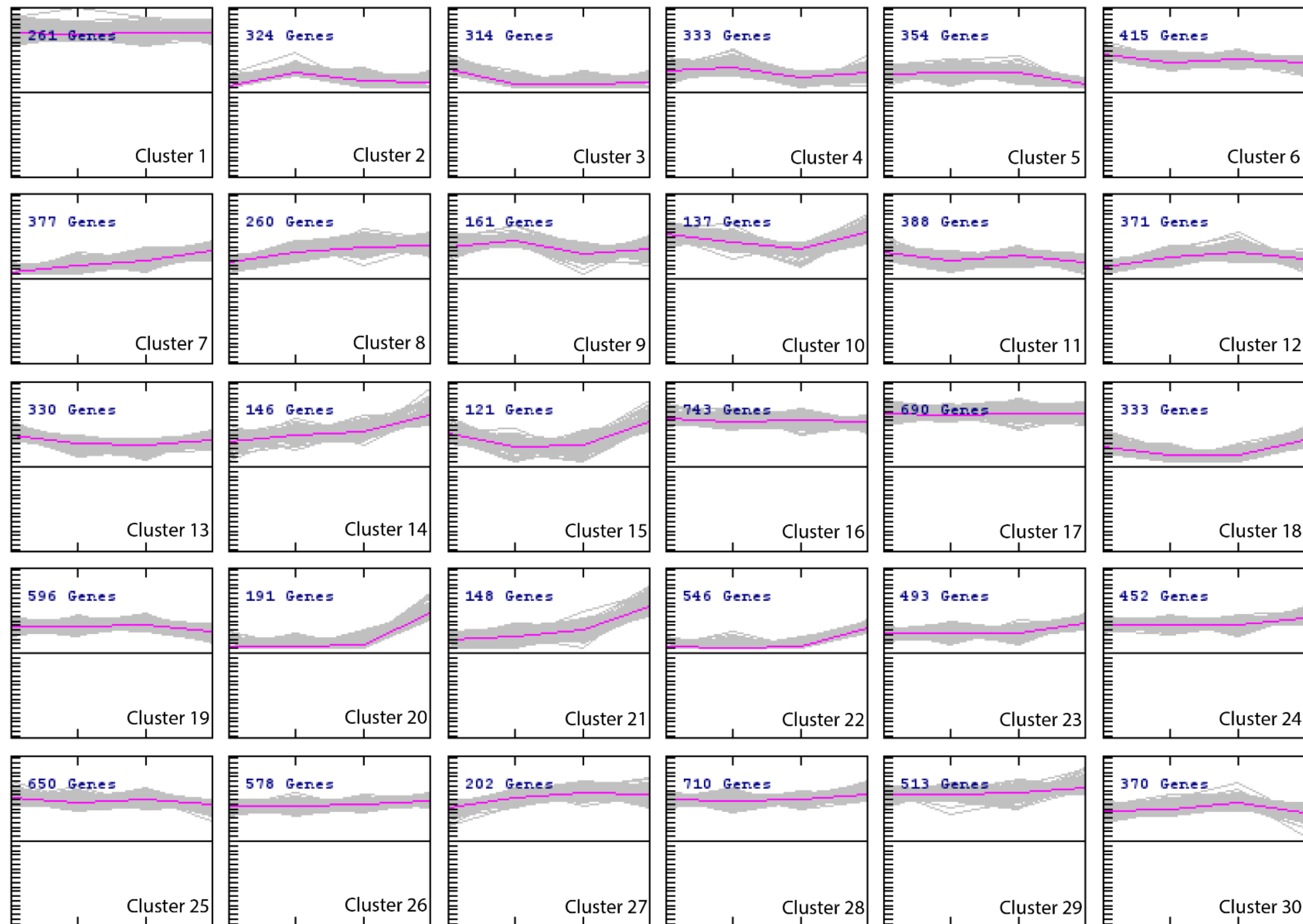


Figure III. 5 | Cluster analysis of differential expressed genes by the K-means method from the gene expression profile.

4.3. Differential gene expression validation by RT-qPCR

In order to confirm the expression levels obtained by bioinformatic analyses, nineteen genes related with embryogenesis, and with distinct profiles across the somatic embryos developmental stages, were chosen for the validation of the expression profiles by RT-qPCR (Figure III. 6). The correlation of gene expression levels obtained by RT-qPCR and Illumina sequencing was demonstrated by Pearson correlation in which the majority of the genes showed a high correlation level: $r>0.9$. A moderate correlation was also observed ($0.6> r > 0.9$) for six of the selected genes.

4.4. GO analysis of differentially expressed genes

The GO annotation for the 11,507 DEGs revealed that cellular and metabolic processes are the main represented categories in biological processes, each accounting with 31.51% and 28.31%, respectively. Genes involved in other important biological processes such as single-organism processes (8.95%), response to stimulus (6.60%), cellular component organization (4.57%), localization (3.99%), multicellular organism processes (3.38%) and developmental processes (3.33%) were also identified (Figure III. 7). Regarding the molecular function categories, GO terms for the DEG are mainly related with binding and catalytic activities. In the catalytic subset, the main groups represented were: heterocyclic compound binding (21.71%), organic cyclic compound binding (21.71%), small molecule binding (15.05%), protein binding (4.09%) and carbohydrate binding (1.34%). In the subset of catalytic activities two main groups were represented: transferase activity (18.63%) and hydrolase activity (13.67%). Transcription factor activity (1.30%) and signal transducer activity (0.80%) were also represented (Figure III. 7). In terms of cellular components, the identified genes are mainly related to intracellular membrane-bounded organelle (29.07%), cytoplasm (27.41%), cytoplasmic part (23.27%), intracellular non-membrane-bounded organelle (4.12%), thylakoid (3.76%) and cell wall (2.53%) (Figure III. 7). Data from all GO-slim annotated DE genes is resumed in Supplementary Material 2. In order to understand the enrichment occurrence of the overrepresented biological process

GO categories in the DEGs clusters defined in K-means hierarchical analysis, we compared gene clusters proportion with its transcriptome assembly occurrence.

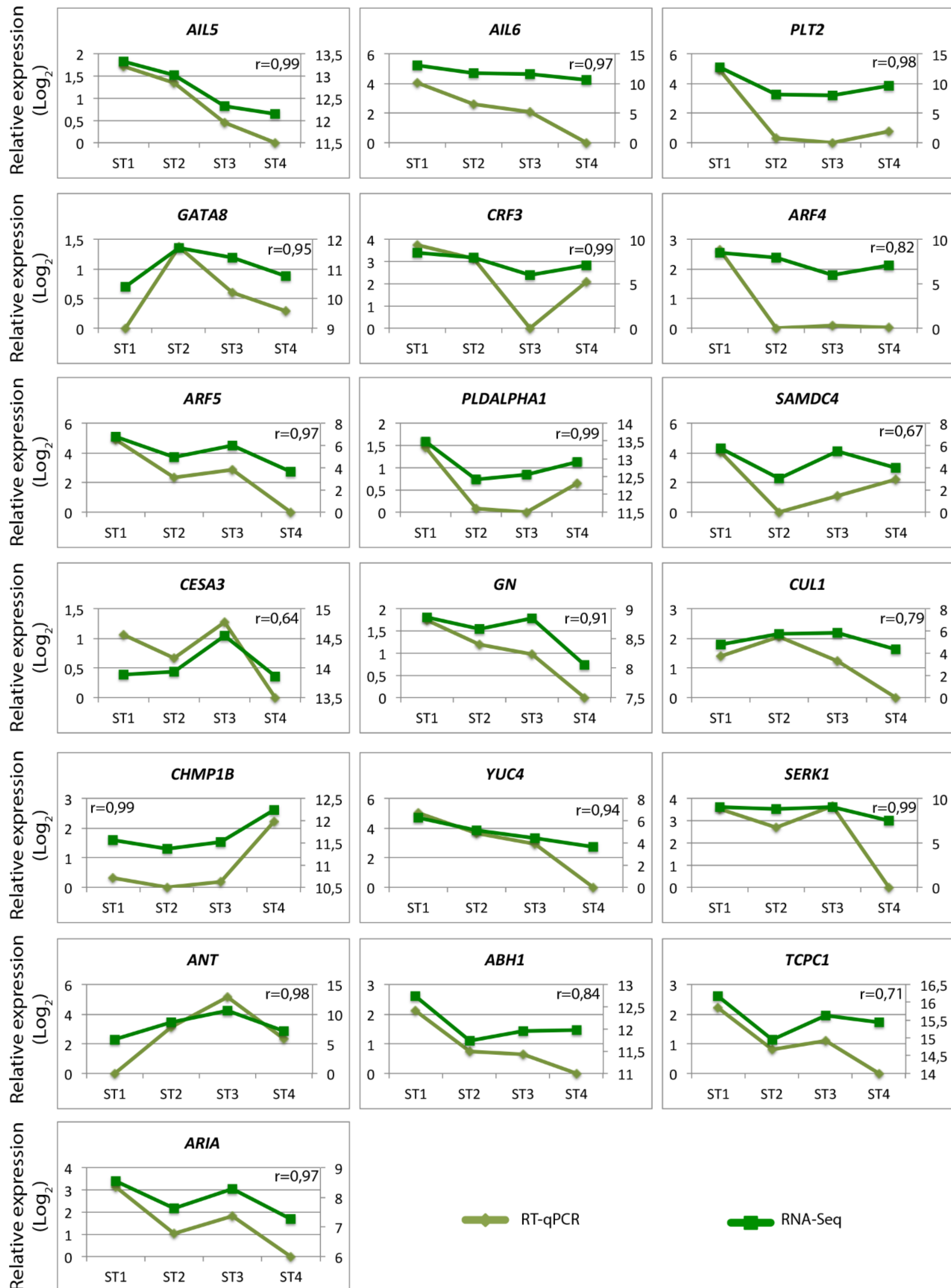


Figure III. 6 | Validation of RNA-Seq transcripts expression profiles. The comparison of the transcripts expression profile from RNA-Seq and from RT-qPCR assays was calculated with the Pearson correlation and expressed by the r-value. In the y-axis is represented the Log2 of the

relative expression level values from RNA-Seq data (right y-axis scale) and RT-qPCR (left y-axis scale) in each developmental stage.

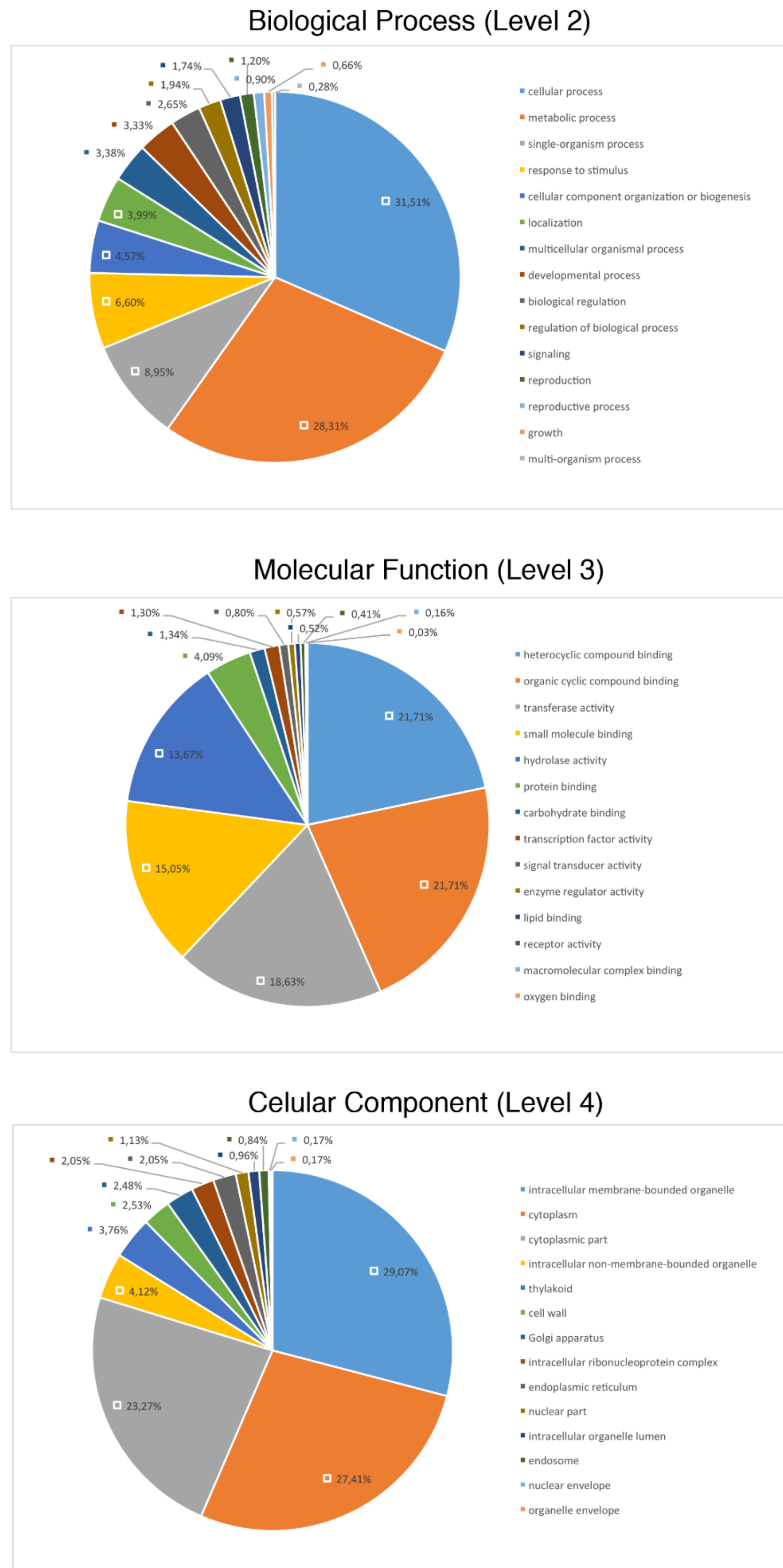


Figure III. 7 | GO functional classification of differential expressed genes on biological process at level 2, molecular function at level 3 and cellular localization at level 4.

This analysis showed that the significantly enriched GO ontologies were: generation of precursor metabolites and energy (clusters 5, 8 and 25), secondary metabolic processes (cluster 8 and 11), cell cycle (cluster 15 and 22), photosynthesis (cluster 5 and 8), DNA metabolic processes (cluster 15), translation (cluster 16), cellular component organization, DNA metabolic processes, embryo development, and epigenetic regulation of gene expression (all at cluster 22), cellular protein modification process and carbohydrate metabolic process (cluster 28) and response to biotic and external stimulus (cluster 29) (Figure III. 8). These results give a broad view of the overrepresented biological processes involved in some gene clusters, allowing monitoring the gene expression patterns with a complementary association to biological processes, molecular function and cellular components.

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

In order to identify and categorize the TFs, TRs and CRs genes related with somatic embryo developmental stages, a list of plant TFs, TRs and CRs genes was generated from libraries ST1, ST2, ST3 and ST4. The analysis identified 2,674 TFs, TRs and CRs genes from which 1,159 were differentially expressed among somatic embryos developmental stages (Supplementary Material 3). The most predominant family type of regulators is TFs (76%), followed by transcription factor iterators (8%) and CRs (4%). Among TFs, the most abundant belong to the C2H2, MYB-HB-like, WD40-like, AP2-EREBP, bHLH, WRKY, NAM and bZIP families of TFs, to the CCHC(Zn) family of TRs and to the PHD family of CRs. GO annotation enrichment analysis was performed to identify the overrepresented biological processes related with the identified differential expressed regulators. The most relevant GO terms were associated with response to endogenous stimulus, post-embryonic development, signal transduction, response to abiotic stimulus, cell communication, flower development, cell differentiation and anatomical structure morphogenesis (Figure III. 9).

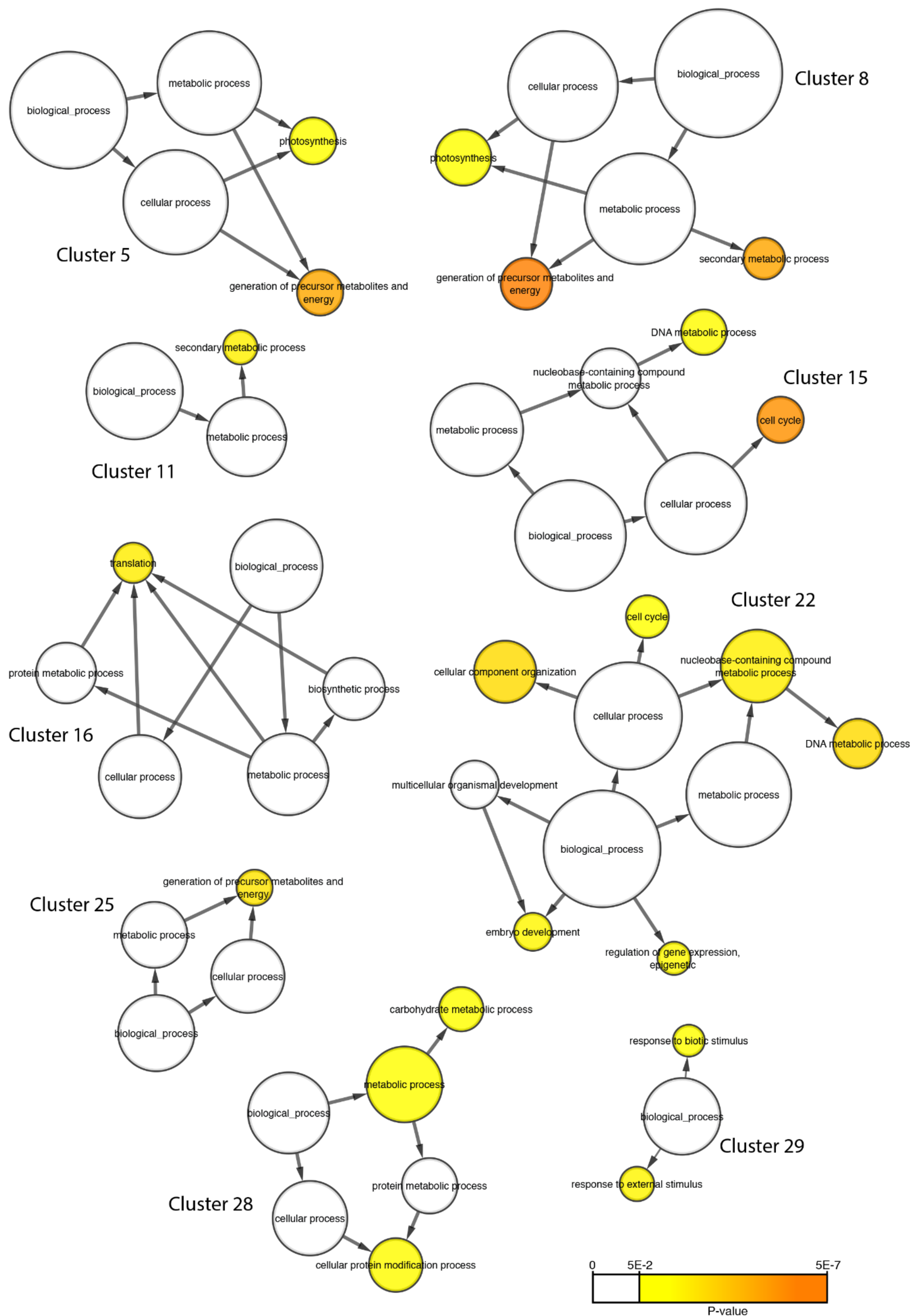


Figure III. 8 | Biological process network of GO term enrichment for DE genes in K-means clusters with over-represented GO terms. The node size represents the number of genes associated to a given GO term and node fill colour reflects the adjusted P-value.

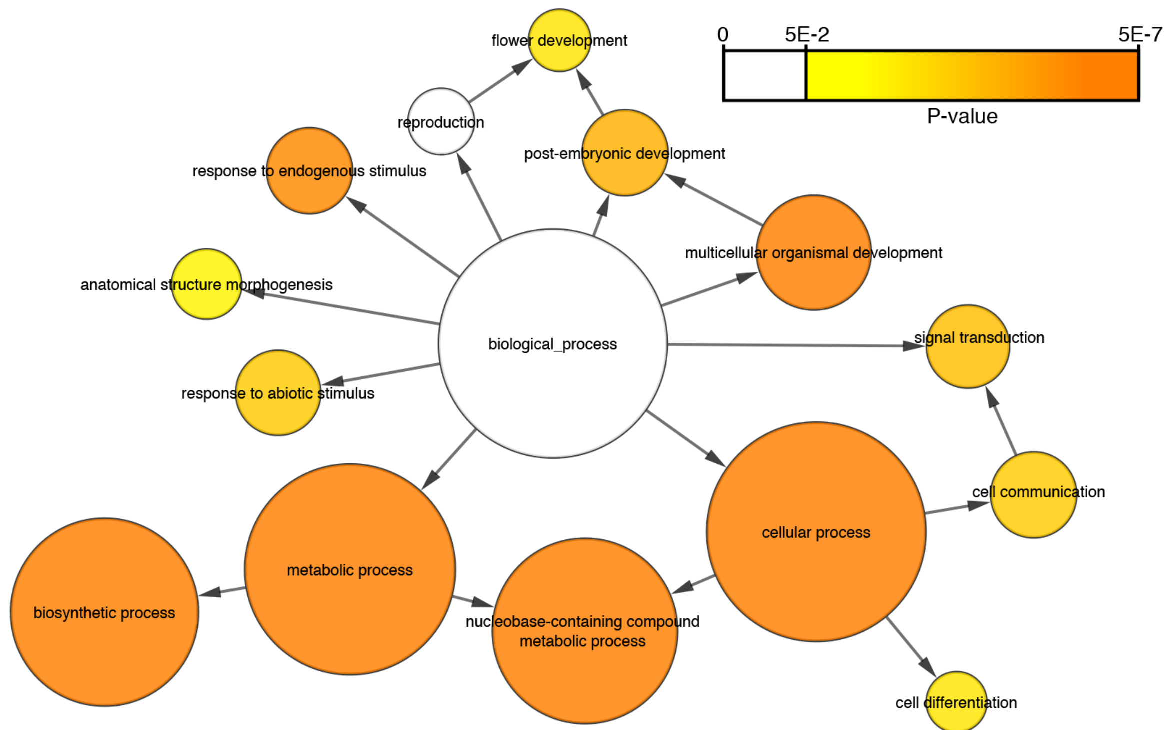


Figure III. 9 | Biological process network of GO term enrichment of DE transcription factors. The node size represents the number of genes associated to a given GO term and node fill colour reflects the adjusted P-value.

A search for genes involved in embryo developmental and maturation was made taking into account blast similarity and gene function in *Arabidopsis thaliana*. AP2-EREBP TFs with high homology for the *AINTEGUMENTA*-like (*AIL*) were identified, namely homologs of *AINTEGUMENTA-LIKE 5 (AIL5)* and *AINTEGUMENTA-LIKE 6 (AIL6)* genes. Interestingly, a homolog of *PLETHORA 2 (PLT2)* which is also an AP2-EREBP transcription factor is up-regulated in ST1, being 25-27 times more expressed than in ST2 and ST3, and a homolog of the *AINTEGUMENTA (ANT)* gene was found to be up-regulated in ST3. With an opposite expression pattern, a *CYTOKININ RESPONSE FACTOR 3 (CRF3)* homolog was identified as down-regulated in ST3 stage. We also found a homolog of the *GATA transcription factor 8 (GATA8)* expressed in all developmental stages, with a particularly higher expression in ST2 and ST3.

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

4.6. Expression profile of phytohormones homeostasis-related genes

In order to explore the hormone-mediated transcriptional regulation in gene expression during embryo development, we mapped the DEGs to eight categorical hormone related pathways including abscisic acid (ABA), auxin (AUX), brassinosteroid (BR), cytokinin (CK), ethylene (ET), gibberellin (GA), jasmonic acid (JA) and salicylic acid (SA). The analysis revealed that 250 genes associated with hormone biosynthesis, response, signaling, receptors and metabolism showed significant differential expression DE during somatic embryo development (Supplementary Material 4). The main group that showed significant DE is related with ABA (28%), followed by AUX (23%), ET (20%), JA and SA (7.6%), GA (5.6%), BR (5.2%) and CK (3.2%). In terms of biological functions, we found relevant *Arabidopsis thaliana* homolog genes within the ABA group, namely the *PHOSPHOLIPASE D ALPHA 1 (PLDALPHA1)* and the *NUCLEAR CAP-BINDING PROTEIN SUBUNIT 1 (ABH1),* which were highly expressed in all developmental stages. Within the AUX group we found homologs of *AUXIN EFFLUX CARRIER COMPONENT 2 (PIN2)*, which was highly up-regulated in ST1; homologs of *CULLIN-1 (CUL1)* and *TOPLESS (TPL)* were found to be up-regulated in both ST2 and ST3. A homolog of *S-ADENOSYLMETHIONINE DECARBOXYLASE PROENZYME 4 (SAMDC4)*, was also identified as up-regulated in ST1 and ST3. Related with AUX biosynthesis, a homolog of *INDOLE-3-PYRUVATE MONOOXYGENASE YUCCA4 (YUC4)* was identified as up-regulated in ST1 and the *GUANINE-NUCLEOTIDE EXCHANGE FACTOR GNOM (GN)* was down-regulated at ST4. Also related with AUX homeostasis we identified a putative gene encoding *ESCRT-RELATED PROTEIN CHMP1B (CHMP1B)*, which was highly expressed in all embryo developmental stages, especially in ST4. Down-regulated in ST4, was a putative coding sequence related with the BR signaling pathway, a homolog of the *SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 (SERK1)* gene. The approach used to identify the hormone-mediated transcriptional regulated genes, also allowed the identification of an ET and JA responsive gene highly expressed in all developmental stages, especially in ST3, the *CELLULOSE SYNTHASE A CATALYTIC SUBUNIT 3 (CESA3)* homolog. A group of hormone-related genes

not expressed in at least one developmental stage were also identified, namely *THERMOSPERMINE SYNTHASE ACAULIS5 (ACL5) XYLOGLUCAN ENDOTRANSGLUCOSYLASE/ HYDROLASE PROTEIN 24 (XTH24)* and three *HISTIDINE KINASE* isoforms *(HK1, HK3 and HK4)*.

4.7. Expression profile of genes involved in germination

Plant hormones can affect several biological process including seed dormancy and germination (Graeber et al. 2012). Although seeds do not occur in the *in vitro* somatic embryo culture system used in this study, genes involved in seed germination may have analog function in the embryo maturation process. In fact, genes described as involved in seed germination and related with hormone homeostasis were found to be differentially expressed, namely transcripts with homology for ABA and GA-related genes, which are the major class of hormones responsible for seed germination (Miransari and Smith 2014). Within these, we highlight ABA-related genes like homologs of the *ABC TRANSPORTER C FAMILY MEMBER5 (ABCC5)*, *ARM REPEAT PROTEIN INTERACTING WITH ABF2 (ARIA)* and *BETA-D-XYLOSE 6 (BXL6)* which were up-regulated in ST3. With high expression levels in all development stages was found a homolog of *MAGNESIUM-CHELATASE SUBUNIT CHLH (CHLH)* up-regulated in ST4. Related with CK homeostasis, two *HISTIDINE KINASE (HK)* were identified, an *HK3* homolog up-regulated in ST4 and an *HK4* homolog down-regulated in ST2. Related with the GA signaling pathway was found one homolog of the *DELLA PROTEIN (RGL1)* up-regulated in ST3. Related with JA biosynthesis, a homolog of *3-KETOACYL-COA THIOLASE 2 (PED1)* which showed an increasing expression profile throw ST1 to ST4 and a homolog of the *TWO PORE CALCIUM CHANNEL PROTEIN 1 (TPC1)* down-regulated at ST4 were reported.

4.8. Expression profiles of other embryogenesis related-genes

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

genes involved in embryo development than genes which occur in the described categories. In order to bypass this issue, we manually searched the DE genes dataset for blast descriptions and GO terms related with embryo and embryogenesis. Furthermore we excluded genes previously described in this study as TFs, TRs, CRs and phytohormones homeostasis-related genes. A list of 67 genes was compiled and an enrichment map using as reference set the DEGs was generated (Figure III. 10).

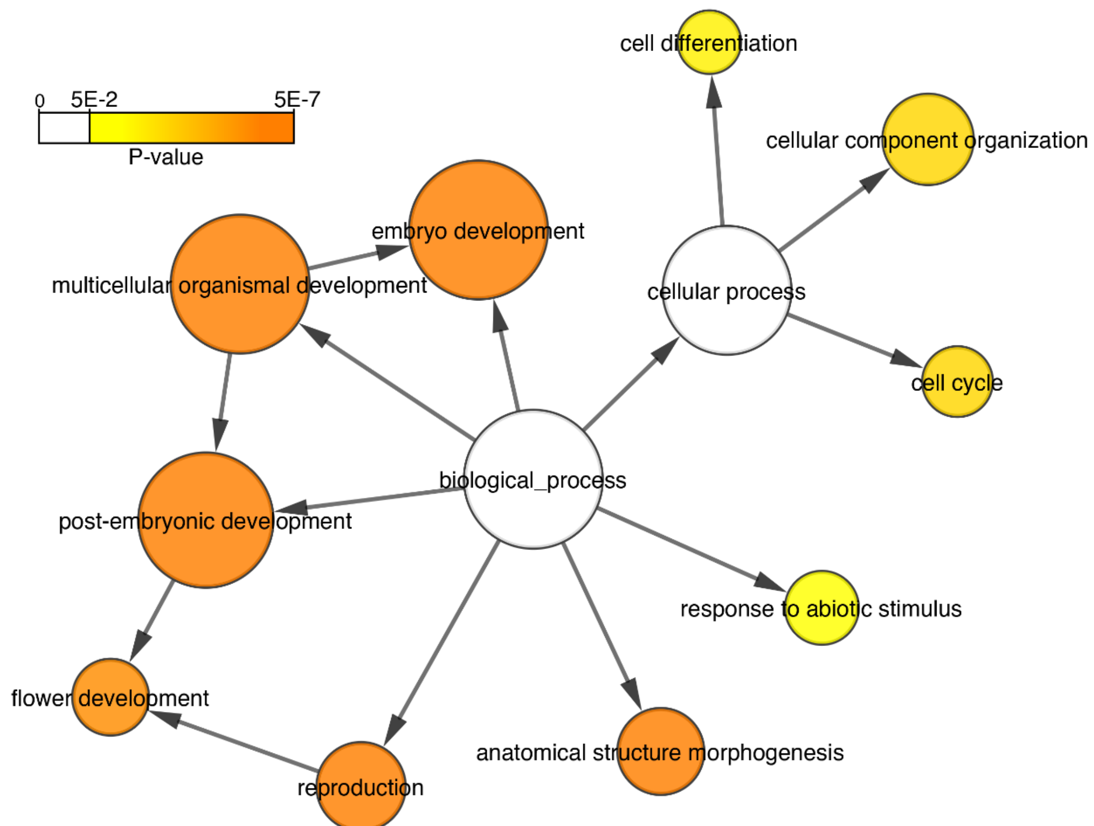


Figure III. 10 | Biological process network of GO term enrichment of DE genes related with embryogenesis GO terms not included in the transcription factors and hormone related-genes sets. The node size represents the number of genes associated to a given GO term and node fill colour reflects the adjusted P-value.

GO related with embryo development and post-embryonic development were the most significantly enriched categories in this subset of genes. Besides of the relevant role in embryogenesis, these genes are also involved in anatomical structure morphogenesis, response to reproduction and flower development, abiotic stimulus and cellular processes as cell cycle, cellular component organization and cell differentiation. A list of these genes, as well as

its expression levels is summarized in Supplementary Material 5. It is relevant to highlight gene groups with a representative occurrence in the list, namely embryo defective, embryo abundant and late embryogenesis abundant protein (LEA) homologs with different expression profiles during the four developmental stages, three pentatricopeptide repeat protein (PPR) homologs down-regulated in ST2, namely the *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN At3g06430*, *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN At3g49240* and *PENTATRICOPEPTIDE REPEAT-CONTAINING PROTEIN At3g53700*. A homolog of *SUMO-ACTIVATING ENZYME SUBUNIT 2 (SAE2)* and a homolog of *SERINE/THREONINE-PROTEIN KINASE-LIKE PROTEIN ACR4 (ACR4)* which were both up-regulated in ST1 and a homolog of the ABC transporter I family member 6 (ABCI6) up-regulated in ST4. Interestingly, a homolog of *TRANSLATIONALLY-CONTROLLED TUMOR PROTEIN 1 (TCPC1)* with high expression values during all developmental stages, particularly in ST1, was identified.

5. Discussion

Embryogenesis is traditionally referred as the developmental period where zygote passes through a series of differentiation events, which will result in a mature embryo with cotyledons, shoot apical meristem, hypocotyl, root, and root apical meristem. Higher plant embryogenesis requires two fertilization events for reproduction, producing the zygote and the endosperm. Endosperm is an extraembrionic tissue system that is essential for zygotic embryo development, functioning as a nutrient provider for embryos, however endosperm does not occur in somatic embryogenesis (Von Arnold et al. 2002; Jin et al. 2014). Plants are able to produce embryos by different ways: zygotic, from egg cells as a result of fertilization and somatic, which is defined as a biological process where a bipolar structure similar to a zygotic embryo develops from a non-zygotic cell (Von Arnold et al. 2002). Somatic embryogenesis occurs through a series of stages identical to zygotic embryogenesis (Jin et al. 2014) and by this reason somatic embryos are used as a model system in plant embryological studies. Somatic embryos also constitute a high valuable resource in large-scale vegetative

propagation and embryogenic cultures open the possibility to establish gene banks and genetic modifications (Von Arnold et al. 2002).

Quercus suber somatic embryos are important biological resource for future breeding programs in this species. We provide here the first overview of the transcriptome dynamics of *Q. suber* somatic embryos development which will help to understand the complex mechanism underlying this process.

A total of 66,693 genes were identified with distinct roles in various biological processes, 1,1507 of them with DE between embryo stages. Several showed to be specifically expressed in the studied stages of somatic embryo development. We centred the analyses on transcripts with potential functions in embryogenesis, namely TFs, CRs and transcripts related with phytohormones homeostasis during embryogenesis and germination.

5.1. Transcription regulators involved in cork oak somatic embryogenesis and somatic embryo development

In the last years, several genetic studies have identified genes involved in embryogenesis initiation and progression as well as embryo development (Tzafrir et al. 2004; Abid et al. 2010; Radoeva and Weijers 2014). Many of these genes encode TRs, particularly TFs which are able to control various plant biological processes regulating gene expression such as embryonic and post-embryonic development or embryo germination (Radoeva and Weijers 2014). In this study, 1159 differential expressed TFs with gene ontologies deeply related with response to stimulus, signal transduction and structures development were identified (Figure III. 9). Particularly, we found a group of AP2-EREBP TFs, including *AIL5*, *AIL6*, *PLT2* and *ANT2* differentially expressed across the four stages of cork somatic embryo development. AP2-EREBP transcription factor family play an essential role in cell proliferation and embryogenesis (Ouakfaoui et al. 2010). *BABY BOOM,* a member of the family, is preferentially expressed in developing embryos and when overexpressed it induces the formation of somatic embryos from leafs, cotyledons and shoot apical meristem (Boutilier et al. 2002). A close homolog of *BBM* is *AIL5*, also known as *EMBRYOMAKER (EMK)* or *PLT5.* Our data revealed an orthologue of *AIL5* in cork oak. Its expression pattern shows that the gene is highly expressed in the four developmental stages,

however it is relative highly expressed in stages ST1 and ST2. This evidence suggests that the *QsAIL5* has an active function during cork oak somatic embryogenesis and similarly to its orthologue in *Arabidopsis* may have an important role in the embryonic identity maintenance. The fact that the relative expression values decreases along the embryo maturation, indicates an expression profile similar to the observed in *Arabidopsis*. In fact, the rapid disappearance of *AtAIL5* expression is essential for the developmental transition between the embryonic and vegetative phases in *Arabidposis* (Tsuwamoto et al. 2010). Previous work has shown that *ANT* and *AIL6,* which is also known as *PLT3,* have partially redundant roles in shoot and flower development, specifically in floral meristem initiation and floral patterning (Krizek 2009; Krizek 2011). In addition, *AIL6* together with *PLT1*, *PLT2* and *BBM* have redundant functions in root meristem and embryo differentiation and in combination with *AIL5* controls shoot phyllotaxis (Galinha et al. 2007; Prasad et al. 2011). *PLT2* is also described as responsible by controlling the patterning within the root stem cell niche and is required for root stem cell activity during embryogenesis (Aida et al. 2004). Therefore, we hypothesize that *QsAIL6* and *QsPLT2* may also play a role in cork oak hypocotyl and root development. Transcription of the PLT genes is stimulated by AUX and is dependent of ARF TFs (Benjamins and Scheres 2008; Mahonen et al. 2014). In this context is particular interesting to observe that both genes show identical decreasing expression from ST1 to ST4, which may explain the difficulty of cork oak somatic embryo root formation without the use of plant growth regulators. AUX signalling is vital for plant growth and development, and plays an important role in embryo patterning and shoot elongation (Smit and Weijers 2015). The transcriptional activator *MONOPTEROS* (*MP*) also known as *ARF5* is crucial to embryogenesis as demonstrated by defects in most AUX-dependent embryo patterning in *arf5* mutant (Benjamins and Scheres 2008). Specifically, *ARF5* is critical for embryonic root formation (Schlereth et al. 2010), embryo axis formation and vascular development (Hardtke and Berleth 1998). In this study we found an orthologue of *ARF5*, which is up-regulated in ST1. The relative high expression of *QsARF5*, *QsPLT2, QsAIL6* in ST1 suggests that these TFs may act together in early stages of cork oak somatic somatic embryogenesis to control the cell identity and differentiation of the embryo like is observed in *Glycine max* (Ouakfaoui et al. 2010). However, the

detection of other members of the ARF family in the same stage points to a redundant function of genes of this family. In fact phylogenetic analysis in *Arabidopsis* consider several ARF sister pairs (Chandler 2016), namely *ARF3* and *ARF4*. We found an orthologue of *ARF4* also up-regulated in ST1. *ARF4* and *ARF3* seem to be redundant in the establishment of tissue patterning and floral meristem determinacy (Liu et al. 2014). This points to the importance and occurrence of tissue organization and patterning in early stages of cork oak development, however further studies have been reported in order to understand the function of the referred transcription factor genes in cork oak embryogenesis. CRFs genes also belong to the AP2-EREBP TFs family. Analysis of loss-of-function mutations revealed that CRFs function redundantly in order to regulate embryos, cotyledons and leaves development (Rashotte et al. 2006). In this study we found an orthologue of *CRF3* with DE in ST1. *AtCRF3* is a component of the CK signalling pathway and seems to be involved in embryo development (Rashotte et al. 2006). As a transcription factor, *QsCRF3* may have a key role in cork oak embryo development linking the complex transcriptional regulatory network involving TFs and phytohormones action. Within the differential expressed TFs identified we found also a member of the GATA TF family with DE during embryogenesis. Its orthology with *AtGATA8*, which is a positive regulator of seed germination and is expressed in embryo axis (Liu et al. 2005), lead us to hypothesise a putative role of *QsGATA8* in somatic embryo germination since we observed a reduction of the gene expression levels at ST3 and ST4, which may also be related with the difficulty for cork oak somatic embryos to convert into emblings, documented in several studies (Pinto et al. 2002)(García-Martín 2001; Hernández et al. 2003). Moreover, the lack of *GATA8* in *Arabidopsis* led to a reduced rate of germination even after a stratification treatment (Liu et al. 2005).

5.2. Hormones-related genes involved in cork oak somatic embryogenesis and somatic embryo development

Phytohormones are chemical messengers that regulate cellular activities, pattern formation, plant growth, plant development and stress responses but how these compounds act at the molecular level is still not fully understood.

Hormones are involved in modulating gene expression by controlling the abundance of two types of gene regulatory proteins, TFs and transcriptional repressors, through regulated protein degradation or by modifying their activities through post-translational modifications. In this context it is particularly important to understand how hormones regulate the biosynthesis and transport of other hormones and how hormone interactions converge on regulation of TFs, which integrate and coordinate the plant developmental response. In this study we found 250 DEGs associated with hormone biosynthesis, response, signalling, receptors and metabolism and focused on genes involved in embryogenesis, embryo development and germination. ABA is responsible to promote the initiation of embryogenesis and inhibit seed germination (Miransari and Smith 2014). Related with ABA, an orthologue of *PLDALPHA1* and an orthologue of *ABH1* highly expressed in all developmental stages and up-regulated in ST1, was identified. *PLDALPHA1* plays an important role in various cellular processes, namely in phytohormone action and response to stress, acting as positive regulators of ABA-activated signalling pathway (Guo et al. 2012). This is particular important in the mechanism of water loss mediated by ABA, regulating the opening and closing of stomata (Uraji et al. 2012). It is involved in ABA sensitivity of seed germination and also in ABA-induced stomatal closing (Hugouvieux et al. 2001). *ABH1* is also a component of the cap-binding complex (CBC), which is involved in miRNA-mediated RNA interference and is required for primary miRNA processing (Gregory et al. 2008). The high level of both transcripts in all developmental stages, particularly in ST1, evidences the importance of these two genes in ABA signaling of early cork oak somatic embryo development. AUX is a phytohormone with a profound effect on plant growth and development. Related with AUX, we identified at least 3 genes deeply involved in embryogenesis and post-embryonic development, namely *CUL1*, *SAMDC4 and YUC4*. CUL1, together with SKP1 and an F-box protein form a proteic complex named SCF ubiquitin ligase which targets various proteins involved in hormone response and signalling, morphogenesis and control of circadian clock (Shen et al. 2002; Willems et al. 2004; Harmon et al. 2008; Mockaitis and Estelle 2008). Several studies showed a reduced AUX response in *cul1* loss-of-function mutants development (Shen et al. 2002; Hellmann et al. 2003; Moon et al. 2007) and an arrestment in early embryogenesis stages in *Arabidopsis* (Shen et al. 2002). In

this study, *CUL1* is expressed in all developmental stages and may play a key role in AUX homeostasis in cork oak embryogenesis. An orthologue of *SAMDC4,* an important decarboxylase in plant polyamine biosynthetic pathway, was also identified. In plants, polyamines are related with membrane fluidity and biotic and abiotic stress response (Watson and Malmberg 1995; Cowley and Walters 2002; Perez-Amador 2002; Capell et al. 2004; Tonon et al. 2004). Several studies also indicate that polyamines also affects the formation of plant architecture, such as internode elongation (Hanzawa et al. 1997; Alcázar et al. 2005), root branching (Ben-Hayyim et al. 1994) and shoot apical dominance (Geuns et al. 2001). In *Arabidopsis SAMDC*4 is essential for plant embryogenesis, normal growth and development and may be involved in the response to AUX and/or CK (Ge et al. 2006). In this study, the *QsSAMDC4* is expressed during all developmental stages and is down-regulated in ST2 and ST4. *YUC4* is a AUX responsive gene and is related with AUX biosynthesis and plant development. YUC genes are mainly expressed in meristems, young primordia, vascular tissues, and reproductive organs. In *Arabidopsis*, the overexpression of four important YUC genes (*YUC1*, *YUC2*, *YUC4* and *YUC6*) leads to AUX overproduction, whereas the lack of expression of each gene individually does not result in developmental defects. This reflects the complex regulation and mode of action of AUX homeostasis. Also in *Arabidopsis,* Cheng and co-author (Cheng et al. 2007) showed that the AUX synthesized by YUC genes is essential for the establishment of the basal body region during embryogenesis and the formation of embryonic and postembryonic organs. More specifically, the *YUC1* and *YUC4* were expressed during embryogenesis and their expression overlaps with other YUC genes, namely *YUC10* and *YUC11*. In fact the quadruple mutants *yuc1 yuc4 yuc10 yuc11* failed to develop a hypocotyl and root meristem, a phenotype similar to the *arf5* mutant discussed previously. Our data identified an orthologue of YUC4 in cork oak, which is up-regulated in ST1. These results, may suggest a potential important involvement of *QsYUC4* in cork oak embryogenesis initiation and maintenance. Related with multivesicular body formation and sorting of the *PIN* AUX carriers and essential for embryo development the *CHMP1B* gene was also identified (Spitzer et al. 2009). In *Arabidopsis CHMP1B* together with *CHIMP1A* are involved in cellular differentiation and embryo symmetry establishment. In fact, homozygous *chmp1a chmp1b* mutant embryos showed

limited polar differentiation and failed to establish bilateral symmetry as well as an incapacity to establish normal AUX gradients (Spitzer et al. 2009). In our data a *QsCHMP1B* is expressed in all developmental stages, with high relative expression values, especially in ST4. We hypothesize that *QsCHMP1B* may be relevant in the embryo symmetry establishment and cellular differentiation in cork oak embryo development, particularly in ST4 were *QsCHMP1B* is up-regulated. BR are plant hormones which have similar function to steroid hormones in other organisms (Arora et al. 2008; Miransari and Smith 2014). BR are able to promote germination by controlling the inhibitory effect of ABA on seed germination (Finkelstein et al. 2008). We identified an orthologue of *SERK1,* a BR receptor, as down-regulated in ST4. *SERK* genes have been reported to play specialized functions in embryogenesis (Hecht et al. 2001; Hu et al. 2005). In addition, *SERK1* is described as a key factor in embryogenic competence regulation, as it regulates cell death and increases stress resistance (Hu et al. 2005). Moreover, *Arabidopsis* seedlings overexpressing *AtSERK1* exhibited an increased efficiency for initiation of somatic embryogenesis, which led to consider that an increase in *AtSERK1* expression of 3-to 4-fold is sufficient to confer embryogenic competence in culture (Hecht et al. 2001). Secondary embryogenesis was observed in embryo clusters in ST1 to ST3 stage, but this capacity was lost in ST4 stage, where embryos follow a maturation pathway. The expression pattern of *SERK1* orthologue correlates with these observations. *QsSERK1* is expressed from ST1 to ST3 with similar mRNA expression levels, however in ST4 *QsSERK1* decreases about 2.5-to 3 fold. Therefore, *QsSERK1* may have the ability to confer embryogenic competence like its orthologue in *Arabidopsis*. A *CESA3* orthologue was also identified, which is related with ET and JA signalling (Ellis and Turner 2001; Ellis et al. 2002). It is demonstrated that *CESA3* is involved in morphogenesis during embryo development and primary cell wall formation, especially in roots (Beeckman et al. 2002; Li et al. 2014). Our data indicates that this orthologue of *CESA3* is up-regulated in ST3 and down-regulated in ST4. Our data also revealed several differential expressed genes not expressed in at least one developmental stage. *ACL5* is not expressed in ST1 but its expression increase from ST2 to ST4. In *Arabidopsis*, *ACL5* is required for xylem specification by regulating and preventing the premature death of the xylem vessel elements (Hanzawa et al. 2000; Kakehi et al. 2008; Muñiz et al. 2008). It

is demonstrated that thermospermine produced by *ACL5* is one of the factors downstream to AUX contributing to the regulation of vascular differentiation. In agreement, *ACL5* is regulated positively by AUX (Hanzawa et al. 2000; Cui et al. 2010) and negatively by exogenous thermospermine probably by a feedback mechanism involving the translational regulation of the SUPRESSOR OF ACAULIS 51 (SAC51) b-HLH TF (Imai et al. 2006; Kakehi et al. 2010; Yoshimoto et al. 2012). Exogenously added thermospermine supresses AUX-inducible xylem differentiation and may fine-tune the timing and spatial pattern of xylem differentiation (Yoshimoto et al. 2012). Recently, Milhinhos and co-workers (Milhinhos et al. 2013) described a feedback regulatory path in poplar secondary xylem in which thermospermine levels are controlled by an AUX-dependent feedback-loop mechanism involving the poplar orthologue of *ARABIDOPSIS THALIANA HOMEOBOX 8 (ATHB8). ATHB8* belong to the homoeodomain-leucine zipper (HD-ZIP) III family of TF which are genes actively expressed in vasculature (Miyashima et al. 2013). Specifically, *ATHB8* is expressed in procambial cells and is directly regulated by *ARF5* (Baima et al. 2014). Our data reveals the presence of two *ATHB8* orthologues, expressed in all developmental stages and with similar expression patterns. Our report describes the presence of well-known key players in AUX-mediated process including hypocotyl elongation and differentiation of xylem precursor cells. Taking into account the recalcitrance of cork oak somatic embryos to elongate and form leaves in vitro, the identification of these players may constitute important findings to design modulation strategies for cork oak plantlets conversion in a near future.

5.3. Genes involved in germination

Somatic embryos are formed by somatic cells differentiation and morphologically resemble zygotic embryos. After maturation, functional root and shoot apices are established, originating an embling. Conversion in cork oak somatic embryos can be obtained by different methods however a low proportion of embryo germination is usually reported (García-Martín 2001; Pinto et al. 2002; Hernández et al. 2003; Hernández et al. 2011). In our study, several expressed genes are related with germination, namely genes involved in ABA regulation. *ABCC5*, up-regulated in stage ST3 and ST4, codes for a pump for glutathione S-

conjugates involved in the root AUX content regulation that controls primary root elongation, lateral root formation and also plays a role in ABA-mediated germination inhibition (Gaedeke et al. 2001; Martinoia et al. 2002; Klein et al. 2003; Lee et al. 2004). Recently, *AtABCG25,* a member of the ABC transporter family was referred as the first identified ABA transporter (Kuromori et al. 2010). At the same time, another ABC transporter, *AtABCG40*, was reported to function as ABA importer in plant cells (Kang et al. 2010). The identification of ABA transporters it is particularly important for the production of plants with increase drought tolerance and faster response to stress (Boursiac et al. 2013). In this study, we identified several transcripts with homology to members of the ABC transporter family, namely 5 transcripts with homology to *AtABCC5* which in turn is related with ABA germination (Klein et al. 2003). These transcripts may have important roles in germination, which opens the gate for future studies about cork oak somatic embryos germination. In addition, our results revealed the presence of *ARIA* orthologues*,* which acts as positive regulator of ABA response by modulating the transcriptional activity of *ABF*2, a TF that controls ABA-dependent gene expression via the G-box-type ABA-responsive elements. *ABF*2 negatively regulates seed germination and young seedling growth in *Arabidopsis* (Kim et al. 2004). In our data, the putative *QsARIA* is down-regulated in ST4 which may reflect a decreasing effect of ABA germination inhibition, however the gene is still expressed which may be insufficient for embryo germination to take place in same cases. Also involved in the regulation of germination is *TPC1* which inhibits ABA-dependent germination (Peiter et al. 2005) and is involved in JA homeostasis (Bonaventure et al. 2007). In this study, we identified a differential expressed orthologue of *TPC1.* Taking together the expression of *ARIA* and *TPC1*, the results led us to hypothesis the presence of a genetic driving force inhibiting cork oak embryos germination, specifically at ST4. One particular aspect in hormone action is signal transduction, in part by the presence of specific hormone receptors at tissue and cellular level. *CHLH* is a multifunctional protein which have functions related with ABA signalling perception (Tsuzuki et al. 2011). It has a role during seed germination and binds ABA acting as a positive regulator of ABA signalling (Du et al. 2012). In our data, a *CHLH* orthologue is highly up-regulated in ST4 which may be related with a strong ABA-mediated germination inhibition. On the contrary with ABA regulation of germination, GA promotes

seeds germination in many plant species (Seo et al. 2009). In fact GA is required to break seed dormancy leading to its germination (Debeaujon and Koornneef 2000). The regulation of GA homeostasis is regulated in part by *DELLA* proteins (Cao et al. 2005). In our data, we found an orthologue of *RGL1*, a member of the *DELLA* family, which is a well-known negative regulator of GA responses (Wen and Chang 2002), up-regulated in ST3 and ST4. *RGL1* function is related with negative regulation of the GA-signalling pathway, therefore the high expression levels observed particularly in ST3 and ST4 may explain in part the low cork oak embryo germination rates by impairing the GA action in germination. Another plant hormone that plays positive and negative regulatory roles in many aspects of plant growth and development is CK (Riefler et al. 2006). Studies in CK-deficient plants has shown that CK plays opposite roles in shoot and root meristems growth (Werner et al. 2003; Riefler et al. 2006). In *Arabidopsis*, the CK signal transduction is perceived by at least three sensor histidine kinases, *AtHK2*, *AtHK3* and *AtHK4* (Inoue et al. 2001; Ueguchi et al. 2001; Yamada et al. 2001; Riefler et al. 2006). These three receptors show a high degree of sequence homology, however their function seems to be different although partially redundant. *AtHK3* and *AtHK4* play key roles in early vascular development in early stages of embryogenesis (Mahonen et al. 2000) and cell differentiation in root meristem (Dello Ioio et al. 2007). In addition, Riefler and co-workers (Riefler et al. 2006) showed that these different receptors were able to mediate CK control of seed germination, particularly *AtHK4* and *AtHK3*. In fact the study shows that CK is a negative regulator of seed germination and that distinct pathways are being controlled by different CK-hormone receptors. In agreement with this idea is the fact that GA and CK signalling pathway can interact, which reinforces the idea of a germination arrestment effect observed (Greenboim-Wainberg et al. 2005). In other hand *AtHK2* and *AtHK3* are involved in inhibition of the primary root development in *Arabidopsis* (Riefler et al. 2006), which is also observed in plants with a reduced endogenous CK content (Werner et al. 2003). We identified in our data orthologues of *HK3* and *HK4*. The expression levels of the two genes have a negative correlation from ST1 to ST4, meaning that when the expression of one gene decreases the expression of the another increases. This evidence may be related with redundant function of the HK receptors. Notably, *QsHK3* is highly expressed in all developmental stages, especially in ST4. These findings

may reveal the players acting in cork oak germination inhibition observed in somatic embryos however further studies are required to clarify the effect of these genes in cork oak somatic embryos conversion. It will be interesting to study how the genes responses to different hormone treatments, namely to ABA, GA and CK and how the repression and overexpression of these gene will affect conversion. These findings may have a deep impact in a future cork oak reforestation and breeding program since the low percentage of somatic embryo germination is a crucial factor in these processes.

5.4. Genes involved in root development

One of the most critical steps in producing large-scale *Q. suber* embryos by somatic embryogenesis is the low frequency of embryo conversion into plantlets. In *in vitro* cultures of *Q. suber* embryos the lack of cotyledons or root is observed in large percentage of embryo, even when proper treatments are applied. In the case of root formation and development the best options are to apply a unique cold treatment or to submit the embryos to a partial desiccation under conditions of high humidity, or controlled starvation, prior to a cold treatment. However, very few embryos still to develop spontaneously a well-formed root system. In our data we found DEGs directly related with root formation and development, which may explain the difficulty of *Q. suber* somatic embryo root development. *PIN2* which acts as a component of the AUX efflux carrier (Abas et al. 2006; Benjamins and Scheres 2008) is involved in the root-specific AUX transport, and mediates root gravitropism in *Arabidopsis* (Guyomarc'h et al. 2012). Its localization also suggests a role in translocation of AUX towards the elongation zone (Müller et al. 1998; Benjamins and Scheres 2008). We identified a cork oak orthologue of *PIN2* up-regulated in ST1, where embryos are characterized by a smaller number of differentiated cells. This suggests that *QsPIN2* is a important player in early embryo AUX efflux and may be involved on earlier events of root development and gravitropism as in *Arabidopsis* (Guyomarc'h et al. 2012). Moreover, the relative low expression of the gene in the other three developmental stages could be associated with the difficulty of cork oak somatic embryos to develop an *in vitro* root system. Related with AUX transport and AUX-dependent plant growth, an orthologue of *GN* was

also identified. *GN* is an ARF guanine-nucleotide exchange factor which activates ARF proteins (Richter et al. 2010). It regulates the vesicle trafficking required for the coordinated polar localization of AUX efflux carriers, which in turn determines the direction of AUX flow (Naramoto et al. 2010). GN also mediates the sorting and polarization of PIN proteins in hypocotyl endodermal cells (Rakusová et al. 2011) and is involved in the specification of the apical-basal pattern formation in the early embryo and during root formation (Richter et al. 2010). The *QsGN* transcripts are expressed evenly during ST1 to ST3, however in ST4, which is an embryo fully mature stage, it is down-regulated. The involvement of *GN* in root formation and the relation with PIN proteins also puts *QsGN* as a candidate gene related with the difficulty of root formation in cork oak somatic embryos. Another important gene in root formation is *TPL* which is a transcriptional co-repressor of AUX-regulated genes (Szemenyei et al. 2008) targeting directly the *PLT1* and *PLT2* genes (Smith and Long 2010). In *tpl* mutants, *PLT1* and *PLT2* are expressed in both apical and basal ends of the embryo leading to the formation of two roots, due to the lack of TPL binding to the *PLT* gene promoter and therefore not exerting its repressor effect in the basal pole of the embryo (Smith and Long 2010). We identified an orthologous gene of the *AtTPL* in cork oak embryogenesis. The *QsTPL* is highly expressed during all the developmental stages especially in ST2 and ST3 where is up-regulated, although a down-regulation of the gene is observed in ST4. ST4 is a developmental stage were mature embryos occur with fully developed cotyledons. It was expected that the root system was formed in this stage, however this was not observed. We expect that the down-regulation of *QsTPL*, together with other key players in root formation, like the transcription factor *QsPLT2* also referred in this study, may contribute to the understanding of the root formation in cork oak somatic embryos, however further studies are required. CK is well know by positive and negative roles in many aspects of plant development as stimulate the formation and activity of shoot meristems or inhibiting root growth and branching (Riefler et al. 2006). CK receptors like *HK3* and particularly *HK4* are related with primary root formation and root branching (Riefler et al. 2006). In our data we found *QsHK3* and *QsHK4* highly expressed in all developmental stages specially *QsHK3*. The high expression of these receptors will in turn increase CK cellular perception and the hormonal signal cascade probably increasing inhibitory effect of root

development. Together, these findings will open the possibility of further studies about cork oak root formation and development, which may have a great impact in the efficiency of embryo root formation and therefore cork oak somatic embryos conversion.

5.5. Other relevant genes in cork oak somatic embryogenesis

Higher plants have developed several genetic responses to biotic and abiotic stress during evolution. An important group is the LEA protein gene family, which have functions in cellular protection during abiotic stress tolerance (Hincha and Thalhammer 2012). LEA proteins are mainly expressed in seeds and are accumulated during late embryo developmental stages, however they have also been identified in seedlings, roots and other organs (Hong-Bo et al. 2005; Gao and Lan 2016). In this study, we identified several LEA transcripts, namely hydroxyproline-rich glycoproteins and pentatricopeptide repeat (PPR) proteins. PPR gene mutants are related with photosynthetic defects, aberrant leaf development, changes in leaf pigmentation, restricted growth, hypersensitivity to abiotic and ABA, restricted growth and defective seed and embryo development (Barkan and Small 2014). In our data, we found three PPR genes differential expressed and closed related with embryo development aspects. The first PPR gene identified was an orthologue of the *PPR At3g06430* also known by *AtPPR2*, which its mutation is related with delayed embryogenesis, leading to embryonic lethality (Lu et al. 2011). In addition, the mutation in *AtEMB2750,* which appears to be a weak mutant allele of the *AtPPR2* locus, also results in defective seed (Lu et al. 2011). In the same study, the majority of *emb2750* mutant seeds were able to germinate, but their cotyledons were albino and often deformed adding the fact that *emb2750* seedling growth were arrested after germination (Lu et al. 2011). The second *PPR* gene identified in cork oak somatic embryogenesis is an orthologue of *PPR At3g49240* also known as *AtEMB1796*. The *emb1796* mutants resulted in developmental failure beyond globular stage, exhibiting consistent and severe developmental arrest at the globular stage (Cushing et al. 2005). The third PPR gene identified is an orthologue of *PPR At3g53700* also know as *MATERNAL EFFECT EMBRYO ARREST 40 (MEE40)* which is involved in protein degradation, cell death, signal transduction and transcriptional regulation

required for early embryogenesis (Pagnussat et al. 2005). These three genes are important players in early embryogenesis in where mutations led to developing arrestment or embryonic lethality. Furthermore, we identified orthologues of *ABCI6* and *SAE2*, other two genes essential for embryogenesis. *AtABCI6* also known as *AtNAP7* plays an important role e plastid Fe-S cluster maintenance and repair during embryogenesis and as *AtSAE2* their mutants are embryonic lethal (Xu and Møller 2004; Saracco et al. 2007). We observe an active expression of these orthologues, which indicates the importance of these genes for embryo development cork oak somatic embryogenesis. The presence of these genes in our data constitutes evidences that they may be part of a group of essential genes for cork embryo development. Another important aspect of plant growth and development is a coordinated and regulated cell proliferation and growth. *TCPC1* is a positive regulator of mitotic growth controlling the duration of the cell cycle both in animals as in plants and therefore acting in the complex and poorly understood regulation network of these processes (Brioudes et al. 2010). *ACR4* is another important player in cell division, particularly in meristems, root tips and lateral initiation zones of the pericycle (Watanabe et al. 2004; De Smet et al. 2008; Stahl et al. 2009). In addition it is required during embryogenesis and embryo development, probably for the differentiation of the protoderm in epidermal cells (Tanaka et al. 2002). In cork oak somatic embryogenesis these two genes are actively expressed. The expression of *TCPC1* increases from ST1 to ST4, probably promoting the cell division and embryo growth observed through the developmental stages. In other hand *ACR4* is up-regulated in ST1 and ST3 which may indicate a high meristematic activity at these stages of development. The identification of these genes represents an important advance in the knowledge of cork oak somatic embryogenesis. Nevertheless, functional studies will be needed in a near future in order to determine the exact role of each referred gene.

This work allowed the characterization of cork oak somatic embryogenesis and the identification of several genes and its expression levels in four distinct stages of cork oak embryo development showing up and down-regulation expressed genes in specific stages of development. Genes related with transcriptional regulation were specially addressed in this study due to their importance in embryogenesis. TFs were the main class of regulators identified

throughout embryo developmental process and were involved in over-represented specific processes such as response to cell differentiation, endogenous stimulus, post-embryonic development and signal transduction. Hormone-related genes were also deeply investigated by their relevance in embryo development. Several phytohormone homeostasis related-genes involved in root formation and development as well as genes involved in germination showed active expression dynamics in cork oak somatic embryogenesis. Functional studies of the genes identified in this work will be of great importance to better understand their role in somatic embryo development. The knowledge gathered in this study and in upcoming functional studies will be valuable for the design of successful regeneration protocols that can be integrated in future breeding programs. Additionally, the dataset generated in this work contributes to the enlargement of cork oak genomics resources.

6. Data Archiving

Raw reads of transcriptomic data have been deposited to the NCBI Short Read Archive (SRA) under accessions numbers SRS1739225, SAMN05898490, SAMN05898491 and SAMN05898492. The four records correspond to *Quercus suber* somatic embryo development Biosamples stage 1 (SAMN05898489), stage 2 (SAMN05898490), stage 3 (SAMN05898491) and stage 4 (SAMN05898492).

7. References

Abas L, Benjamins R, Malenica N, Paciorek T, Wiśniewska J, Moulinier-Anzola JC, et al. (2006) Intracellular trafficking and proteolysis of the Arabidopsis auxin-efflux facilitator PIN2 are involved in root gravitropism. Nat Cell Biol 8:249–256

Abid G, Jacquemin J, Sassi K, Muhovski Y (2010) Gene expression and genetic analysis during higher plants embryogenesis. Biotechnol Agron Soc Environ 14:667–680

Aida M, Beis D, Heidstra R, Willemsen V, Blilou I, Galinha C, et al. (2004) The PLETHORA genes mediate patterning of the Arabidopsis root stem cell niche. Cell 119:119–120

Alcázar R, García-Martínez JL, Cuevas JC, Tiburcio AF, Altabella T (2005) Overexpression of ADC2 in Arabidopsis induces dwarfism and late-flowering

through GA deficiency. Plant J 43:425–436

Allen CD, Macalady AK, Chenchouni H, Bachelet D, McDowell N, Vennetier M, et al. (2010) A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. For Ecol Manage 259:660–684

Álvarez R, Alonso P, Cortizo M, Celestino C, Hernández I, Toribio M, et al. (2004) Genetic transformation of selected mature cork oak (Quercus suber L.) trees. Plant Cell Rep 23:218–223

Álvarez R, Álvarez JM, Humara JM, Revilla A, Ordás RJ (2009) Genetic transformation of cork oak (Quercus suber L.) for herbicide resistance. Biotechnol Lett 31:1477–83

Arora N, Bhardwaj R, Sharma P, Arora HK (2008) Effects of 28-homobrassinolide on growth, lipid peroxidation and antioxidative enzyme activities in seedlings of Zea mays L. under salinity stress. Acta Physiol Plant 30:833–839

Baima S, Forte V, Possenti M, Peñalosa A, Leoni G, Salvi S, et al. (2014) Negative feedback regulation of auxin signaling by ATHB8/ACL5-BUD2 transcription module. Mol Plant 7:1006–1025

Barkan A, Small I (2014) Pentatricopeptide repeat proteins in plants. Annu Rev Plant Biol 65:415–42

Beeckman T, Przemeck GKH, Stamatiou G, Lau R, Terryn N, De Rycke R, et al. (2002) Genetic complexity of cellulose synthase a gene function in Arabidopsis embryogenesis. Plant Physiol 130:1883–1893

Ben-Hayyim G, Damon JP, Martin-Tanguy J, Tepfer D (1994) Changing root system architecture through inhibition of putrescine and feruloyl putrescine accumulation. FEBS Lett 342:145–148

Benjamins R, Scheres B (2008) Auxin: the looping star in plant development. Annu Rev Plant Biol 59:443–465

Bonaventure G, Gfeller A, Proebsting WM, Hörtensteiner S, Chételat A, Martinoia E, et al. (2007) A gain-of-function allele of TPC1 activates oxylipin biogenesis after leaf wounding in Arabidopsis. Plant J 49:889–898

Boursiac Y, Léran S, Corratgé-Faillie C, Gojon A, Krouk G, Lacombe B (2013) ABA transport and transporters. Trends Plant Sci 18:325–333

Boutilier K, Offringa R, Sharma VK, Kieft H, Ouellet T, Zhang L, et al. (2002) Ectopic Expression of BABY BOOM Triggers a Conversion from Vegetative to Embryonic Growth. Plant Cell 14:1737–1749

Braisier CM (1996) Phytophthora cinnamomi and oak decline in southern Europe. Environmental constraints including climate change. Ann des Sci For 53:347–358

Brasier CM (1992) Oak tree mortality in Iberia. Nature 360:539

Brioudes F, Thierry A-M, Chambrier P, Mollereau B, Bendahmane M (2010) Translationally controlled tumor protein is a conserved mitotic growth integrator in animals and plants. Proc Natl Acad Sci U S A 107:16384–9

Bueno MA, Gómez A, Manzanera JA (2000) Somatic Embryogenesis in Woody Plants: Volume 6. In: Jain SM, Gupta PK, Newton RJ (eds). Springer Netherlands, Dordrecht, pp 479–508

Cao D, Hussain A, Cheng H, Peng J (2005) Loss of function of four DELLA genes leads to light- and gibberellin-independent seed germination in Arabidopsis. Planta 223:105–113

Capell T, Bassie L, Christou P (2004) Modulation of the polyamine biosynthetic pathway in transgenic rice confers tolerance to drought stress. Proc Natl Acad Sci U S A 101:9909–9914

Chandler JW (2016) Auxin response factors. Plant Cell Environ 39:1014–1028

Cheng Y, Dai X, Zhao Y (2007) Auxin synthesized by the YUCCA flavin monooxygenases is essential for embryogenesis and leaf formation in Arabidopsis. Plant Cell 19:2430–9

Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:619832

Cowley T, Walters DR (2002) Polyamine metabolism in an incompatible interaction between barley and the powdery mildew fungus, Blumeria graminis f. sp. hordei. J Phytopathol 150:581–586

Cui X, Ge C, Wang R, Wang H, Chen W, Fu Z, et al. (2010) The BUD2 mutation affects plant architecture through altering cytokinin and auxin responses in Arabidopsis. Cell Res 20:576–586

Cushing DA, Forsthoefel NR, Gestaut DR, Vernon DM (2005) Arabidopsis emb175 and other ppr knockout mutants reveal essential roles for pentatricopeptide repeat (PPR) proteins in plant embryogenesis. Planta 221:424–436

Dai X, Sinharoy S, Udvardi M, Zhao PX (2013) PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. BMC Bioinformatics 14:321

De Smet I, Vassileva V, De Rybel B, Levesque MP, Grunewald W, Van Damme D, et al. (2008) Receptor-like kinase ACR4 restricts formative cell divisions in the Arabidopsis root. Science 322:594–597

Debeaujon I, Koornneef M (2000) Gibberellin requirement for Arabidopsis seed germination is determined both by testa characteristics and embryonic abscisic acid. Plant Physiol 122:415–424

Dello Ioio R, Linhares FS, Scacchi E, Casamitjana-Martinez E, Heidstra R, Costantino

P, et al. (2007) Cytokinins Determine Arabidopsis Root-Meristem Size by Controlling Cell Differentiation. Curr Biol 17:678–682

Du S-Y, Zhang X-F, Lu Z, Xin Q, Wu Z, Jiang T, et al. (2012) Roles of the different components of magnesium chelatase in abscisic acid signal transduction. Plant Mol Biol 80:519–537

Ellis C, Karafyllidis I, Wasternack C, Turner JG (2002) The Arabidopsis mutant cev1 links cell wall signaling to jasmonate and ethylene responses. Plant Cell 14:1557–1566

Ellis C, Turner JG (2001) The Arabidopsis mutant cev1 has constitutively active jasmonate and ethylene signal pathways and enhanced resistance to pathogens. Plant Cell 13:1025–1033

Fan D, Liu T, Li C, Jiao B, Li S, Hou Y, et al. (2015) Efficient CRISPR/Cas9-mediated Targeted Mutagenesis in Populus in the First Generation. Sci Rep 5:12217

Feng S, Jacobsen SE, Reik W (2010) Epigenetic Reprogramming in Plant and Animal Development. Science (80-) 330:622–627

Finkelstein R, Reeves W, Ariizumi T, Steber C (2008) Molecular aspects of seed dormancy. Annu Rev Pant Bol 59:387–415

Gaedeke N, Klein M, Kolukisaoglu U, Forestier C, Muller A, Ansorge M, et al. (2001) The Arabidopsis thaliana ABC transporter AtMRP5 controls root development and stomata movement. EMBO J 20:1875–1887

Galinha C, Hofhuis H, Luijten M, Willemsen V, Blilou I, Heidstra R, et al. (2007) PLETHORA proteins as dose-dependent master regulators of Arabidopsis root development. Nature 449:1053–1057

Gao J, Lan T (2016) Functional characterization of the late embryogenesis abundant (LEA) protein gene family from Pinus tabuliformis (Pinaceae) in Escherichia coli. Sci Rep 6:19467

Gao M, Gropp G, Wei S (2012) Combinatorial networks regulating seed development and seed filling. Embryog …. doi: 10.5772/2143

García-Martín G (2001) Quercus suber L. somatic embryo germination and plant conversion: pretreatments and germination conditions. Vitr Cell Dev Biol - Plant 37:190–198

Ge C, Cui X, Wang Y, Hu Y, Fu Z, Zhang D, et al. (2006) BUD2, encoding an S-adenosylmethionine decarboxylase, is required for Arabidopsis growth and development. Cell Res 16:446–56

Geuns JMC, Smets R, Struyf T, Prinsen E, Valcke R, Van Onckelen H (2001) Apical dominance in Pssu-ipt-transformed tobacco. Phytochemistry 58:911–921

Gomez-Garay A, Lopez JA, Camafeita E, Bueno MA, Pintos B (2013) Proteomic perspective of Quercus suber somatic embryogenesis. J Proteomics 93:314–325

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotech 29:644–652

Graeber K, Nakabayashi K, Miatton E, Leubner-Metzger G, Soppe WJJ (2012) Molecular mechanisms of seed dormancy. Plant Cell Environ 35:1769–1786

Greenboim-Wainberg Y, Maymon I, Borochov R, Alvarez J, Olszewski N, Ori N, et al. (2005) Cross talk between gibberellin and cytokinin: the Arabidopsis GA response inhibitor SPINDLY plays a positive role in cytokinin signaling. Plant Cell 17:92–102

Gregory BD, O'Malley RC, Lister R, Urich MA, Tonti-Filippini J, Chen H, et al. (2008) A Link between RNA Metabolism and Silencing Affecting Arabidopsis Development. Dev Cell 14:854–866

Guo L, Mishra G, Markham JE, Li M, Tawfall A, Welti R, et al. (2012) Connections between sphingosine kinase and phospholipase D in the abscisic acid signaling pathway in Arabidopsis. J Biol Chem 287:8286–8296

Gurevich A, Saveliev V, Vyahhi N, Tesler G (2013) QUAST: Quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075

Gutierrez L, Van Wuytswinkel O, Castelain M, Bellini C (2007) Combined networks regulating seed maturation. Trends Plant Sci 12:294–300

Guyomarc'h S, Leran S, Auzon-Cape M, Perrine-Walker F, Lucas M, Laplaze L (2012) Early development and gravitropic response of lateral roots in Arabidopsis thaliana. Philos Trans R Soc B Biol Sci 367:1509–1516

Hanzawa Y, Takahashi T, Komeda Y (1997) ACL5: an Arabidopsis gene required for internodal elongation after flowering. Plant J 12:863–874

Hanzawa Y, Takahashi T, Michael AJ, Burtin D, Long D, Pineiro M, et al. (2000) ACAULIS5, an Arabidopsis gene required for stem elongation, encodes a spermine synthase. EMBO J 19:4248–4256

Hardtke CS, Berleth T (1998) The Arabidopsis gene MONOPTEROS encodes a transcription factor mediating embryo axis formation and vascular development. EMBO J 17:1405–1411

Harmon F, Imaizumi T, Gray WM (2008) CUL1 regulates TOC1 protein stability in the Arabidopsis circadian clock. Plant J 55:568–579

Hecht V, Vielle-Calzada JP, Hartog M V, Schmidt ED, Boutilier K, Grossniklaus U, et al. (2001) The Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1 gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. Plant Physiol 127:803–16

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 Biol 8:R19

Hellmann H, Hobbie L, Chapman A, Dharmasiri S, Dharmasiri N, Del Pozo C, et al. (2003) Arabidopsis AXR6 encodes CUL1 implicating SCF E3 ligases in auxin regulation of embryogenesis. EMBO J 22:3314–3325

Hernández I, Celestino C, Alegre J, Toribio M (2003) Vegetative propagation of Quercus suber L. by somatic embryogenesis. II. Plant regeneration from selected cork oak trees. Plant Cell Rep 21:765–70

Hernández I, Cuenca B, Carneros E, Alonso-Blazquez N, Ruiz M, Celestino C, et al. (2011) Application of Plant Regeneration of Selected Cork Oak Trees by Somatic Embryogenesis to Implement Multivarietal Forestry for Cork Production. Tree For Sci Biotechnol 19–26

Hincha DK, Thalhammer A (2012) LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. Biochem Soc Trans 40:1000–1003

Hong-Bo S, Zong-Suo L, Ming-An S (2005) LEA proteins in higher plants: Structure, function, gene expression and regulation. Colloids Surfaces B Biointerfaces 45:131–135

Hu H, Xiong L, Yang Y (2005) Rice SERK1 gene positively regulates somatic embryogenesis of cultured cell and host defense response against fungal infection. Planta 222:107–117

Hugouvieux V, Kwak JM, Schroeder JI (2001) An mRNA cap binding protein, ABH1, modulates early abscisic acid signal transduction in Arabidopsis. Cell 106:477–487

Imai A, Hanzawa Y, Komura M, Yamamoto KT, Komeda Y, Takahashi T (2006) The dwarf phenotype of the Arabidopsis acl5 mutant is suppressed by a mutation in an upstream ORF of a bHLH gene. Development 133:3575–3585

Inoue T, Higuchi M, Hashimoto Y, Seki M, Kobayashi M, Kato T, et al. (2001) Identification of CRE1 as a cytokinin receptor from Arabidopsis. Nature 409:1060–1063

Jiang Z, Liu X, Peng Z, Wan Y, Ji Y, He W, et al. (2011) AHD2.0: An update version of arabidopsis hormone database for plant systematic studies. Nucleic Acids Res 39:1–7

Jin F, Hu L, Yuan D, Xu J, Gao W, He L, et al. (2014) Comparative transcriptome analysis between somatic embryos (SEs) and zygotic embryos in cotton: Evidence for stress response functions in SE development. Plant Biotechnol J 12:161–173

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, et al. (2014) InterProScan 5: Genome-scale protein function classification. Bioinformatics 30:1236–1240

Joshi N, Fass J (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at

https://github.com/najoshi/sickle.

Kakehi JI, Kuwashiro Y, Motose H, Igarashi K, Takahashi T (2010) Norspermine substitutes for thermospermine in the control of stem elongation in Arabidopsis thaliana. FEBS Lett 584:3042–3046

Kakehi JI, Kuwashiro Y, Niitsu M, Takahashi T (2008) Thermospermine is Required for Stem Elongation in Arabidopsis thaliana. Plant Cell Physiol 49:1342–1349

Kang J, Hwang J-U, Lee M, Kim Y-Y, Assmann SM, Martinoia E, et al. (2010) PDR-type ABC transporter mediates cellular uptake of the phytohormone abscisic acid. Proc Natl Acad Sci U S A 107:2355–2360

Kim S, Choi H, Ryu H-J, Park JH, Kim MD, Kim SY (2004) ARIA, an Arabidopsis arm repeat protein interacting with a transcriptional regulator of abscisic acid-responsive gene expression, is a novel abscisic acid signaling component. Plant Physiol 136:3639–3648

Klein M, Perfus-Barbeoch L, Frelet A, Gaedeke N, Reinhardt D, Mueller-Roeber B, et al. (2003) The plant multidrug resistance ABC transporter AtMRP5 is involved in guard cell hormonal signalling and water use. Plant J 33:119–129

Krizek B (2009) AINTEGUMENTA and AINTEGUMENTA-LIKE6 act redundantly to regulate Arabidopsis floral growth and patterning. Plant Physiol 150:1916–1929

Krizek BA (2011) Aintegumenta and Aintegumenta-Like6 regulate auxin-mediated flower development in Arabidopsis. BMC Res Notes 4:176

Kuromori T, Miyaji T, Yabuuchi H, Shimizu H, Sugimoto E, Kamiya A, et al. (2010) ABC transporter AtABCG25 is involved in abscisic acid transport and responses. Proc Natl Acad Sci U S A 107:2361–2366

Lee EK, Kwon M, Ko J-H, Yi H, Hwang MG, Chang S, et al. (2004) Binding of sulfonylurea by AtMRP5, an Arabidopsis multidrug resistance-related protein that functions in salt tolerance. Plant Physiol 134:528–538

Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079

Li S, Bashline L, Lei L, Gu Y (2014) Cellulose synthesis and its regulation. Arabidopsis Book 12:e0169

Liu PP, Koizuka N, Martin RC, Nonogaki H (2005) The BME3 (Blue Micropylar End 3) GATA zinc finger transcription factor is a positive regulator of Arabidopsis seed germination. Plant J 44:960–971

Liu X, Dinh TT, Li D, Shi B, Li Y, Cao X, et al. (2014) AUXIN RESPONSE FACTOR 3 integrates the functions of AGAMOUS and APETALA2 in floral meristem

determinacy. Plant J 80:629–641

Lu Y, Li C, Wang H, Chen H, Berg H, Xia Y (2011) AtPPR2, an Arabidopsis pentatricopeptide repeat protein, binds to plastid 23S rRNA and plays an important role in the first mitotic division during gametogenesis and in cell proliferation during embryogenesis. Plant J 67:13–25

Maere S, Heymans K, Kuiper M (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. Bioinformatics 21:3448–3449

Mahonen AP, Bonke M, Kauppinen L, Riikonen M, Benfey PN, Helariutta Y (2000) A novel two-component hybrid molecule regulates vascular morphogenesis of the Arabidopsis root. Genes Dev 14:2938–2943

Mahonen AP, Tusscher K ten, Siligato R, Smetana O, Diaz-Trivino S, Salojarvi J, et al. (2014) PLETHORA gradient formation mechanism separates auxin responses. Nature 515:125–129

Mallón R, Valladares S, Corredoira E, Vieitez AM, Vidal N (2014) Overexpression of the chestnut CsTL1 gene coding for a thaumatin-like protein in somatic embryos of Quercus robur. Plant Cell, Tissue Organ Cult 116:141–151

Martinoia E, Klein M, Geisler M, Bovet L, Forestier C, Kolukisaoglu U, et al. (2002) Multifunctionality of plant ABC transporters--more than just detoxifiers. Planta 214:345–355

McCarthy DJ, Chen Y, Smyth GK (2012) Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res 40:4288–4297

Milhinhos A, Prestele J, Bollhöner B, Matos A, Vera-Sirera F, Rambla JL, et al. (2013) Thermospermine levels are controlled by an auxin-dependent feedback loop mechanism in Populus xylem. Plant J 75:685–698

Miransari M, Smith DL (2014) Plant hormones and seed germination. Environ Exp Bot 99:110–121

Miyashima S, Sebastian J, Lee J-Y, Helariutta Y (2013) Stem cell function during plant vascular development. EMBO J 32:178–193

Mockaitis K, Estelle M (2008) Auxin receptors and plant development: a new signaling paradigm. Annu Rev Cell Dev Biol 24:55–80

Moon J, Zhao Y, Dai X, Zhang W, Gray WM, Huq E, et al. (2007) A new CULLIN 1 mutant has altered responses to hormones and light in Arabidopsis. Plant Physiol 143:684–696

Müller A, Guan C, Gälweiler L, Tänzler P, Huijser P, Marchant A, et al. (1998) *AtPIN2* defines a locus of *Arabidopsis* for root gravitropism control. EMBO J 17:6903–6911

Muñiz L, Minguet EG, Singh SK, Pesquet E, Vera-Sirera F, Moreau-Courtois CL, et al. (2008) ACAULIS5 controls Arabidopsis xylem specification through the prevention of premature cell death. Development 135:2573–82

Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiol Plant 15:473–497

Naramoto S, Kleine-Vehn J, Robert S, Fujimoto M, Dainobu T, Paciorek T, et al. (2010) ADP-ribosylation factor machinery mediates endocytosis in plant cells. Proc Natl Acad Sci U S A 107:21890–21895

Ouakfaoui S El, Schnell J, Abdeen A, Colville A, Labbé H, Han S, et al. (2010) Control of somatic embryogenesis and embryo development by AP2 transcription factors. Plant Mol Biol 74:313–326

Pagnussat GC, Yu H-J, Ngo QA, Rajani S, Mayalagu S, Johnson CS, et al. (2005) Genetic and molecular identification of genes required for female gametophyte development and function in Arabidopsis. Development 132:603–614

Peiter E, Maathuis FJM, Mills LN, Knight H, Pelloux J, Hetherington AM, et al. (2005) The vacuolar Ca2+-activated channel TPC1 regulates germination and stomatal movement. Nature 434:404–408

Perez-Amador MA (2002) Induction of the Arginine Decarboxylase ADC2 Gene Provides Evidence for the Involvement of Polyamines in the Wound Response in Arabidopsis. Plant Physiol 130:1454–1463

Pinto G, Valentim H, Costa A, Castro S, Santos C (2002) Somatic embryogenesis in leaf callus from a mature Quercus suber L. Tree. Vitr Cell Dev Biol - Plant 38:569–572

Pintos B, Bueno MA, Cuenca B, Manzanera JA (2008) Synthetic seed production from encapsulated somatic embryos of cork oak (Quercus suber L.) and automated growth monitoring. Plant Cell Tissue Organ Cult 95:217–225

Prasad K, Grigg SP, Barkoulas M, Yadav RK, Sanchez-Perez GF, Pinon V, et al. (2011) Arabidopsis PLETHORA transcription factors control phyllotaxis. Curr Biol 21:1123–1128

Radoeva T, Weijers D (2014) A roadmap to embryo identity in plants. Trends Plant Sci 19:709–716

Rakusová H, Gallego-Bartolomé J, Vanstraelen M, Robert HS, Alabadí D, Blázquez MA, et al. (2011) Polarization of PIN3-dependent auxin transport for hypocotyl gravitropic response in Arabidopsis thaliana. Plant J 67:817–826

Rashotte AM, Mason MG, Hutchison CE, Ferreira FJ, Schaller GE, Kieber JJ (2006) A subset of Arabidopsis AP2 transcription factors mediates cytokinin responses in concert with a two-component pathway. Proc Natl Acad Sci U S A 103:11081–5

Richter S, Anders N, Wolters H, Beckmann H, Thomann A, Heinrich R, et al. (2010) Role

of the GNOM gene in Arabidopsis apical-basal patterning - From mutant phenotype to cellular mechanism of protein action. Eur J Cell Biol 89:138–144

Riefler M, Novak O, Strnad M, Schmu T (2006) Arabidopsis Cytokinin Receptor Mutants Reveal Functions in Shoot Growth , Leaf Senescence , Seed Size , Germination , Root Development , and Cytokinin Metabolism. 18:40–54

Robinson MD, McCarthy DJ, Smyth GK (2009) edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140

Saeed AI, Bhagabati NK, Braisted JC, Liang W, Sharov V, Howe EA, et al. (2006) TM4 microarray software suite. Methods Enzymol 411:134–193

Saracco S a, Miller MJ, Kurepa J, Vierstra RD (2007) Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. Plant Physiol 145:119–134

Schenk RU, Hildebrandt AC (1972) Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. Can J Bot 50:199–204

Schlereth A, Möller B, Liu W, Kientz M, Flipse J, Rademacher EH, et al. (2010) MONOPTEROS controls embryonic root initiation by regulating a mobile transcription factor. Nature 464:913–6

Seo M, Nambara E, Choi G, Yamaguchi S (2009) Interaction of light and hormone signals in germinating seeds. Plant Mol Biol 69:463–472

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. (2003) Cytoscape: A software Environment for integrated models of biomolecular interaction networks. Genome Res 13:2498–2504

Shen W-H, Parmentier Y, Hellmann H, Lechner E, Dong A, Masson J, et al. (2002) Null mutation of AtCUL1 causes arrest in early embryogenesis in Arabidopsis. Mol Biol Cell 13:1916–1928

Smit ME, Weijers D (2015) The role of auxin signaling in early embryo pattern formation. Curr Opin Plant Biol 28:99–105

Smith ZR, Long J a (2010) Control of Arabidopsis apical-basal embryo polarity by antagonistic transcription factors. Nature 464:423–426

Spitzer C, Reyes FC, Buono R, Sliwinski MK, Haas TJ, Otegui MS (2009) The ESCRT-related CHMP1A and B proteins mediate multivesicular body sorting of auxin carriers in Arabidopsis and are required for plant development. Plant Cell 21:749–766

Stahl Y, Wink RH, Ingram GC, Simon R (2009) A signaling module controlling the stem cell niche in Arabidopsis root meristems. Curr Biol 19:909–914

Szemenyei H, Hannon M, Long JA (2008) TOPLESS mediates auxin-dependent

transcriptional repression during Arabidopsis embryogenesis. Science 319:1384–6

Tanaka H, Watanabe M, Watanabe D, Tanaka T, Machida C, Machida Y (2002) ACR4, a putative receptor kinase gene of Arabidopsis thaliana, that is expressed in the outer cell layers of embryos and plants, is involved in proper embryogenesis. Plant Cell Physiol 43:419–428

Tonon G, Kevers C, Faivre-Rampant O, Grazianil M, Gaspar T (2004) Effect of NaCl and mannitol iso-osmotic stresses on proline and free polyamine levels in embryogenic Fraxinus angustifolia callus. J Plant Physiol 161:701–8

Tsuwamoto R, Yokoi S, Takahata Y (2010) Arabidopsis EMBRYOMAKER encoding an AP2 domain transcription factor plays a key role in developmental change from vegetative to embryonic phase. Plant Mol Biol 73:481–492

Tsuzuki T, Takahashi K, Inoue S, Okigaki Y, Tomiyama M, Hossain MA, et al. (2011) Mg-chelatase H subunit affects ABA signaling in stomatal guard cells, but is not an ABA receptor in Arabidopsis thaliana. J Plant Res 124:527–538

Tzafrir I, Pena-muralla R, Dickerman A, Berg M, Rogers R, Hutchens S, et al. (2004) Identification of Genes Required for Embryo Development in Arabidopsis. Plant Physiol 135:1206–1220

Ueguchi C, Sato S, Kato T, Tabata S (2001) The AHK4 Gene Involved in the Cytokinin-Signaling Pathway as a Direct Receptor Molecule in Arabidopsis thaliana. Plant Cell Physiol 42:751–755

Untergasser A, Nijveen H, Rao X, Bisseling T, Geurts R, Leunissen JAM (2007) Primer3Plus, an enhanced web interface to Primer3. Nucleic Acids Res 35:W71-4

Uraji M, Katagiri T, Okuma E, Ye W, Hossain MA, Masuda C, et al. (2012) Cooperative function of PLDdelta and PLDalpha1 in abscisic acid-induced stomatal closure in Arabidopsis. Plant Physiol 159:450–460

Valladares S, Toribio M, Celestino C, Vieitez AM (2004) Cryopreservation of embryogenic cultures from mature Quercus suber trees using vitrification. Cryo-Letters 25:177–186

Vieitez AM, Corredoira E, Martínez MT, San-José MC, Sánchez C, Valladares S, et al. (2012) Application of biotechnological tools to Quercus improvement. Eur J For Res 131:519–539

Von Arnold S, Sabala I, Bozhkov P, Dyachok J, Filonova L (2002) Developmental pathways of somatic embryogenesis. Plant Cell Tissue Organ Cult 69:233–249

Watanabe M, Tanaka H, Watanabe D, Machida C, Machida Y (2004) The ACR4 receptor-like kinase is required for surface formation of epidermis-related tissues in Arabidopsis thaliana. Plant J 39:298–308

Watson MB, Malmberg RL (1995) Regulation of Arabidopsis thaliana (L .) Heynh

Arginine Decarboxylase by Potassium Deficiency Stress. 1077–1083

Wen C-K, Chang C (2002) Arabidopsis RGL1 encodes a negative regulator of gibberellin responses. Plant Cell 14:87–100

Werner T, Motyka V, Laucou V, Smets R, Van Onckelen H, Schmulling T (2003) Cytokinin-deficient transgenic Arabidopsis plants show multiple developmental alterations indicating opposite functions of cytokinins in the regulation of shoot and root meristem activity. Plant Cell 15:2532–2550

Willems AR, Schwab M, Tyers M (2004) A hitchhiker's guide to the cullin ubiquitin ligases: SCF and its kin. Biochim Biophys Acta - Mol Cell Res 1695:133–170

Xu XM, Møller SG (2004) AtNAP7 is a plastidic SufC-like ATP-binding cassette/ATPase essential for Arabidopsis embryogenesis. Proc Natl Acad Sci U S A 101:9143–8

Yamada H, Suzuki T, Terada K, Takei K, Ishikawa K, Miwa K, et al. (2001) The Arabidopsis AHK4 Histidine Kinase is a Cytokinin-Binding Receptor that Transduces Cytokinin Signals Across the Membrane. Plant Cell Physiol 42:1017–1023

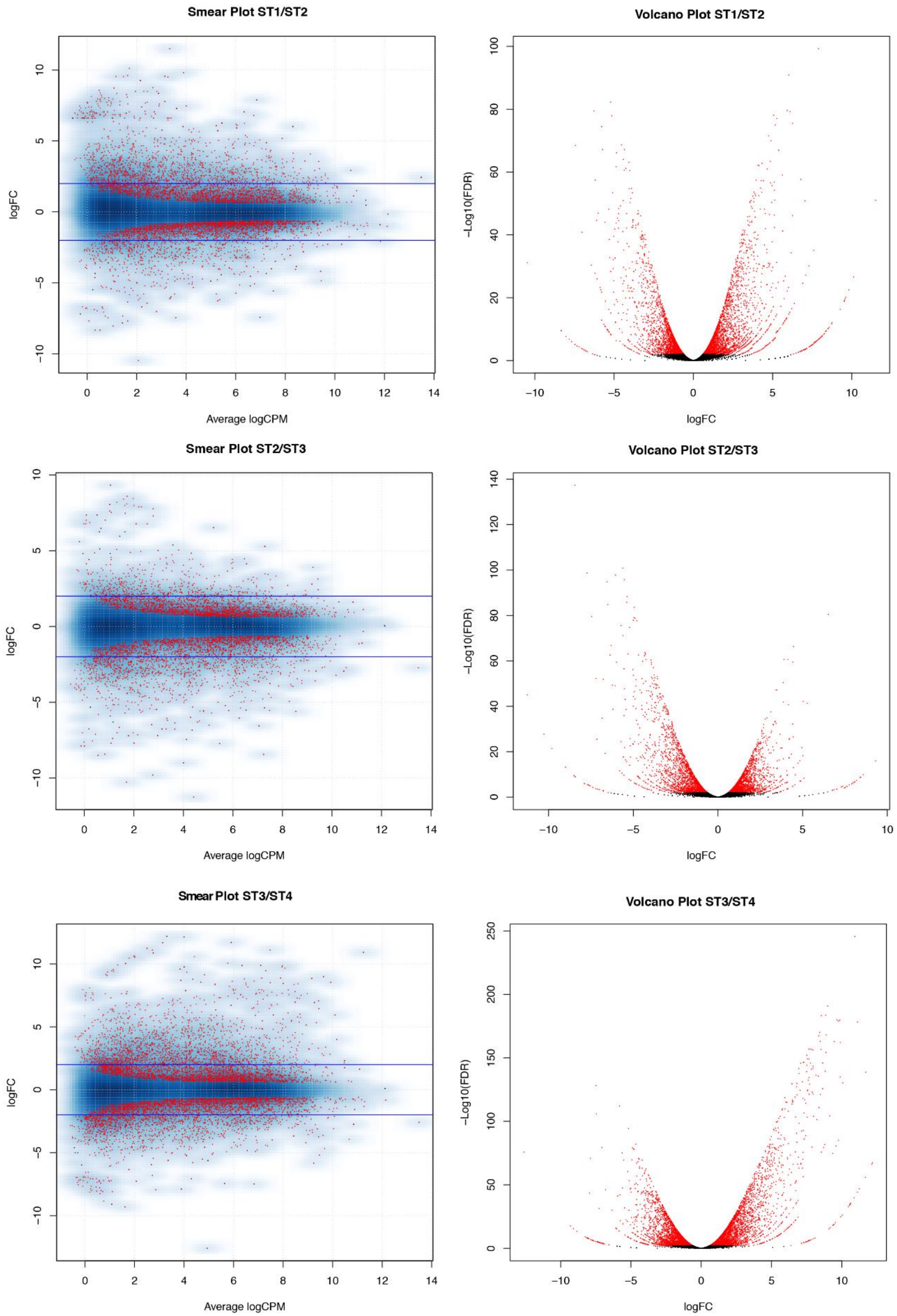
Yoshimoto K, Noutoshi Y, Hayashi KI, Shirasu K, Takahashi T, Motose H (2012) A chemical biology approach reveals an opposite action between thermospermine and auxin in xylem development in Arabidopsis thaliana. Plant Cell Physiol 53:635–645

Zdobnov EM, Apweiler R (2001) InterProScan--an integration platform for the signature-recognition methods in InterPro. Bioinformatics 17:847–8

Zhang Y, Wang Y, Wang C (2012) Gene overexpression and gene silencing in Birch using an Agrobacterium-mediated transient expression system. Mol Biol Rep 39:5537–5541

8. Supplementary Material

8.1. Supplementary Material 1 – Smear plots and Volcano plots



8.2. Supplementary Material 2 – GO slim annotations

Data in file ESM_2.xlsx

8.3. Supplementary Material 3 – Differential expression of transcription factors, transcription regulators and chromatin regulators

Data in file ESM_3.xlsx

8.4. Supplementary Material 4 – Hormone-related genes

Data in file ESM_4.xlsx

8.5. Supplementary Material 5 – Embryogenesis related genes

Data in file ESM_5.xlsx

8.6. Supplementary Material 6 – Primers List and Details

Data in file ESM_6.xlsx

CHAPTER IV

CHIP-SEQ REVEALS THAT QSMYB1 DIRECTLY TARGETS GENES INVOLVED IN LIGNIN AND SUBERIN BIOSYNTHESIS PATHWAYS IN CORK OAK (*QUERCUS SUBER*)

Chapter submitted as original article in a SCI journal

Capote T, Barbosa P, Usié A, Ramos AM, Inácio V, Ordás R, Gonçalves S, Morais-Cecílio L. ChIP-Seq reveals that QsMYB1 directly targets genes involved in lignin and suberin biosynthesis pathways in cork oak (Quercus suber). *BMC Plant Biology. Submitted*

TABLE OF CONTENTS – CHAPTER IV

ChIP-Seq reveals that QsMYB1 directly targets genes involved in lignin and suberin biosynthesis pathways in cork oak (*Quercus suber*).......................... 139

1. Abstract 143

2. Introduction.. 143

3. Methods... 145

3.1. Plant material and growth conditions 145

3.2. *Quercus suber* genetic modified cell lines.. 146

3.3. RT-qPCR validation of QsMYB1::3xFLAG expression......................... 146

3.4. Detection of QsMYB1::3xFLAG fused protein in *Q. suber* embryos 147

3.5. Immunodetection 148

3.6. ChIP-Seq assay 148

3.7. Library preparation and Illumina sequencing 149

3.8. Sequencing data pre-processing and read mapping 149

3.9. Peak calling and downstream processing.. 150

3.10. Analysis of targets related with transcriptional elements.................... 150

3.11. Motif analysis 150

3.12. Metabolic pathway analysis 151

4. Results 151

4.1. QsMYB1::3xFLAG somatic embryo selection for ChIP-Seq 151

4.2. QsMYB1target genes identified by ChIP-Seq 153

4.3. Binding Motif Analysis Reveals QsMYB1 *cis*-regulatory Elements 154

4.4. QsMYB1 directly targets other transcriptional elements 155

4.5. Enzyme encoding genes targeted by QsMYB1 156

4.6. MYB1 target genes related with lipid transporters 162

5. Discussion.. 162

5.1. QsMYB1 targets phenylpropanoid pathway genes regulating the biosynthesis of essential lignin and suberin precursors 163

5.2. QsMYB1 is targeting genes encoding enzymes responsible for suberin biosynthesis 164

5.3. QsMYB1 targets putative transporters of suberin-lipid components 166

5.4. Putative regulation of glycerolipids metabolism by QsMYB1 167

5.5. QsMyb1 may be involved in abiotic stress responses through the regulation of the glycerophospholid pathway 168

5.6. QsMYB1 targets other transcriptional elements expanding its regulatory action 169

6. Conclusions 171

7. References 171

8. Supplementary Material 180

8.1. Supplementary Material 1 180

8.2. Supplementary Material 2 - Generation of cork oak genome draft....... 182

8.2.1. DNA extraction and sequencing 182

8.2.2. Pre-processing and evaluation of high-throughput sequence data 182

8.2.3. Genome assembly and scaffolding 182

8.2.4. Assembly validation and annotation 184

8.2.5. References 184

8.3. Supplementary Material 3 185

8.4. Supplementary Material 4 186

8.5. Supplementary Material 5 186

8.6. Supplementary Material 6 186

8.7. Supplementary Material 7 186

1. Abstract

Gene activity is largely controlled by transcriptional regulation through the action of transcription factors and other regulators. QsMYB1 is a member of the R2R3-MYB transcription factor family related to secondary growth, and in particular, with the cork development process. In order to identify the putative gene targets of QsMYB1 across the cork oak genome we developed a ChIP-Seq strategy.

Results provide direct evidence that QsMY1B targets genes encoding for enzymes involved in the lignin and suberin pathways as well as genes encoding for ABCG transporters and LTPs implicated in the transport of monomeric suberin units across the cellular membrane. These results highlight the role of QsMYB1 as a regulator of lignin and suberin biosynthesis, transport and assembly.

To our knowledge, this work constitutes the first ChIP-Seq experiment performed in cork oak, a non-model plant species with a long-life cycle, and these results will contribute to deepen the knowledge about the molecular mechanisms of cork formation and differentiation.

2. Introduction

Regulation of gene activity at the transcriptional level is the most common form of gene control. Regulation of transcription generally occurs via changes in the amounts and activities of transcription factors (TFs), which modulate the transcription of specific genes either activating or repressing the rate of transcription (Hong 2016). TFs function in networks in which a protein may regulate the expression of another to control directly or indirectly the expression of a particular gene or group of genes in a temporal and spatial fashion manner, allowing the unique expression of each gene in different cell types and during development (Hong 2016).

A distinct characteristic of TFs is that they present a specific DNA-binding domain, which confers the ability to bind to specific DNA regions, then controlling target genes. Chromatin immunoprecipitation (ChIP) followed by high-throughput DNA sequencing (ChIP-Seq) is a method widely used to

identify the binding sites of a target protein across a genome. In a ChIP assay, a TF, cofactor, or other chromatin protein is enriched by immunoprecipitation from cross-linked cells along with its associated DNA (Landt et al. 2012). The resulted DNA is then sequenced and mapped against the species genome in order to identify the binding sites of protein of interest. This allows the subsequent identification of their gene targets, unravelling potential regulatory networks (Landt et al. 2012).

Plant TFs are characterized by a larger number of genes and by the diversity of families. In *Arabidopsis* there are around 2000 TFs genes belonging to approximately 30 different TF families (Riechmann et al. 2000). Most plant TFs as APETALA2/-Ethylene Responsive element binding factor (AP2/ERF), NAC, MADS box, basic helix-loop-helix (bHLH), basic leucine zipper (bZIP) and myeloblastosis (MYB) form large domain families playing important roles in the control of plant growth and development (Hong 2016). The MYB family constitute the most abundant group of TFs in plants. In *Arabidopsis,* 198 MYB TFs were identified, of which 126 belong to the R2R3-MYB subfamily (Yanhui et al. 2006). R2R3-MYB TFs are characterized by their role in a variety of plant-specific processes, such as cell shape and morphogenesis, cellular proliferation and differentiation, hormone response, abiotic and biotic stress, and regulation of primary and secondary metabolism such as phenylpropanoid, lignin and suberin metabolism (Ambawat et al. 2013; Gou et al. 2017). AtMYB41, for example, it is known to be involved although indirectly in the regulation of suberin biosynthesis, export, assembly and deposition in plants under stress conditions. It was recently reported a relation between AtMYB107 and AtMYB9 which synchronize the transcriptional induction of aliphatic and aromatic monomer biosynthesis as well as suberin transport and polymerization in seed outer integument layer (Lashbrooke et al. 2016). Gou et al., (2017) reported AtMYB107 as a positive regulator for seed coat suberin synthesis (Gou et al. 2017), highlighting the important role of MYB TFs in suberin synthesis regulation.

Cork oak (*Quercus suber L.)* is an evergreen broadleaved tree species native to the Mediterranean basin. It is a valuable economic resource due to the sustainable exploitation of its thick bark, the cork (Aronson et al. 2009; Bugalho et al. 2011). Cork or phellem is a tissue derived from the meristematic

activity of the phellogen characterized by a layered deposition of suberized death cells (Graça and Pereira 2004). The high content in suberin provides cork with unique insulator and elastic properties that translates into a large variety of industry applications. Despite the importance of cork, the knowledge on the molecular mechanisms on cork formation and development, are still poorly understood.

Almeida and co-workers (2013) have characterized a R2R3-MYB gene previously identified as related with cork formation and differentiation (Soler et al. 2007), which was named QsMYB1. Authors showed that QsMYB1 is mainly active in organs and tissues with secondary growth resulting from the activity of phellogen (Almeida et al. 2013a). Moreover, *QsMYB1* transcripts are more abundant in cork, a highly suberized tissue, than in wood, a lignified but non-suberized tissue. The authors have also found several *cis*-acting regulatory elements related to phenylpropanoid pathway in the *QsMYB1* promotor region. These findings led the authors to hypothesize that QsMYB1 may be regulating one or more metabolic pathways involved in cork formation.

The present work aimed to validate these results, by developing a Chip-Seq strategy to identify QsMYB1 target genes. Results showed that QsMYB1 acts directly genes of the lignin and suberin metabolic pathways, confirming the important role of QsMYB1 in regulating cork formation and development.

3. METHODS

3.1. Plant material and growth conditions

A stable *Q. suber* somatic embryogenesis line, Ce1, was obtained as described elsewhere (Capote et al., in preparation). Embryo clusters were grown under long day conditions (16h light per 8h dark cycle) at 25 ºC on MSSH medium, containing Murashige and Skoog (MS) micronutrients (Murashige and Skoog 1962), Schenk and Hildebrandt (SH) macronutrients (Schenk and Hildebrandt 1972) and supplemented with MS vitamins (Murashige and Skoog 1962), 3% (w/v) sucrose and 1% (w/v) agar.

3.2. *Quercus suber* genetic modified cell lines

A binary vector for overexpression of QsMYB1::triple FLAG epitope fusion protein was constructed using Gateway™ technology (Gateway® Recombination Cloning, Invitrogen, Lifetechnologies Corporation, CA, USA). QsMYB1 coding sequence (accession number JF970262) was amplified with primers MYB1ns_attB1F and MYB1ns_attB2R (Supplementary File 1) using attB1 and attB2 adapters and recombined with the pDONR221 vector to generate a QsMYB1 ENTRY vector (pENTRY_QsMYB1). Triple FLAG epitope (3xFLAG) nucleotide sequence was generated by annealing the 3xFLAG_F_Olig and the 3xFLAG_R_Olig oligonucleotides (Supplementary File 1) and recombined with the pDONR221 vector in order to produce a 3xFLAG ENTRY vector (pENTRY_3xFLAG). A PCR-fusion/Gateway procedure for gene fusion (Atanassov et al. 2009) was applied to generate an overexpression pK7WG2,0 vector carrying the QsMYB1::3xFLAG fused coding sequence (pK7MYB1::3xFLAG). The resulting binary vector was mobilized into *A. tumefaciens* AGL1 by triparental mating (Ditta et al. 1980) using *E. coli* MC1012 harbouring the mobilizing plasmid pRK2013. After a 10 days sub-culture period, embryos were transformed as described by Álvarez and Ordás (2007) (Álvarez and Ordás 2007).

3.3. RT-qPCR validation of QsMYB1::3xFLAG expression

Genetic transformed embryo clusters were grown by repetitive embryogenesis in kanamycin selective medium during 24 months with regular sub-culturing every 30 days, to allow for transformed embryos selection. After 24 months of embryo transformation, aleatory independent embryo clusters were selected for RT-qPCR analysis in order to screen the expression of the QsMYB1::3xFLAG transcript. Specific primers qPCRQsMYB1F and SR(C) (Supplementary File 1) were used in the RT-qPCR experiments. One microgram of total RNA was used for cDNA synthesis using the QuantiTect Reverse Transcription kit (Qiagen, Valencia, CA, USA), which includes an additional genomic DNA elimination step and uses a mix of oligo(dT) and random hexamer primers. RT-qPCR experiments were carried out in a iCycler iQ5 Instrument (Bio-Rad Laboratories, Hercules, CA, USA) using the Sso

Advanced Universal SYBR Green Master mix (Bio-Rad Laboratories, Hercules, CA, USA) in 96-well plates. Three replicates were performed in reaction mixtures of 20 µL containing 10 µL of 2X Sso Advanced Universal SYBR Green Master mix, 400 nM of each specific primer pair (Forward/Reverse) and 1 µL of cDNA with a dilution of 1:100 as template. Normalization between samples was performed using two reference genes: ACTIN (ACT) and CLATHRIN ADAPTOR COMPLEXES (CACs) shown to be constitutively expressed (data not shown). Normalized relative quantities (NRQ) were calculated by $\mathrm{NRQ} = \frac{\mathrm{E}_{\mathrm{goi}}^{\Delta\mathrm{Ct,goi}}}{\sqrt[\mathrm{f}]{\prod_{\mathrm{o}}^{\mathrm{f}} \mathrm{E}_{\mathrm{ref_o}}^{\Delta\mathrm{Ct,ref_o}}}}$ where E is the amplification efficiency for each primer pair, f the number of reference genes used to normalize the data, goi the gene of interest, ref the reference gene and $\Delta\mathrm{Ct}$ is the Ct of the sample with higher Ct across samples minus the Ct value of the sample in test (Hellemans et al. 2007).

3.4. Detection of QsMYB1::3xFLAG fused protein in *Q. suber* embryos

Expression of full-length QsMYB1 protein fused with 3xFLAG tag was confirmed by Western Blot. Total protein content of genetic modified and non-genetic modified embryos was extracted using TCA-acetone method (Méchin et al. 2007). Supernatants were collected, and total protein concentrations quantified according to Lowry method (LOWRY et al. 1951), using BSA as protein standard. Protein components of cell lysates (25-40 µg protein) were separated on sodium dodecyl sulphate 10% polyacrylamide gel, and then transferred onto polyvinylidene difluoride (PVDF) membranes (Amersham Biosciences, Buckinghamshire, UK). PVDF membranes were blocked with 5% (w/v) of non-fat dry milk at room temperature for 1 h, and incubated overnight at 4°C with a mouse monoclonal ANTI-FLAG® M2 (Sigma-Aldrich, St Louis, MO) primary antibody (1:2000). Bands were visualized by chemiluminescence using anti-mouse horseradish peroxidase-conjugated secondary antibodies, and developed with ECL reagents (Amersham Biosciences, Buckinghamshire, UK), according to manufacturer's instructions.

3.5. Immunodetection

Quercus suber embryos were fixed in a freshly prepared solution of 4 % paraformaldehyde in PBS (phosphate buffered saline: 137 mM NaCl; 0.27 mM KCl; 1 mM phosphate buffer, pH 7.4). The fixed embryos were sectioned (10- to 20-µm-thick) under water using vibratome. The sections were kept at 4 ºC in a 0.1 % solution of formaldehyde in PBS until further use. Subsequently, sections were incubated with blocking solution (5% (w/v) Bovine Serum Albumin (BSA) in 1x PBST – PBS supplemented with 0.5% (v/v) Tween 20). After washing in 1xPBST, sections were incubated in a mouse monoclonal ANTI-FLAG® M2 (Sigma-Aldrich, St Louis, MO) primary antibody (1:200) supplemented with 1% BSA for 16h at 4ºC. The primary antibody was detected with a goat polyclonal secondary antibody to mouse - Alexa Fluor® 488 IgG - H&L (ab150113, Abcam Cambridge, UK) (1:200 dilution in PBST supplemented with 1% BSA) for 1h at 37ºC. Sections were mounted with Vectashield mounting medium with DAPI (Vector Laboratories, USA). Images were acquired on an epifluorescence microscope Axio Imager.Z1.

3.6. ChIP-Seq assay

ChIP-Seq assay of *Q. suber* embryos expressing QsMYB1::3xFLAG was performed by cross-linking and isolating chromatin embryos using the Abcam Plant Chromatin Extraction Kit (ab156906, Abcam, Cambridge, UK) protocol, with slight modifications to the cross-linking procedure: starting with 3 g of fresh embryos, 1 volume of Plant RNA Isolation Aid (Thermo Fisher Scientific, Waltham, MA, USA) per unit mass of embryo tissue (mL/mg) was added to 1% formaldehyde solution and vacuum infiltration step was performed during 20 minutes at 4 ºC. Chromatin was sheared with a Bioruptor® Plus (Diagenode): high intensity pulses, 5 rounds of 10 cycles, 30 seconds ON / 30 seconds OFF, with a sample volume of 150 µL to generate ~300-bp average fragment sizes, as determined by agarose gel and capillary electrophoresis (Agilent 2100 Bioanalyzer) using the High Sensitivity DNA ChIP kit (Agilent, 5067–4626). ChIP was performed following the protocol developed by Li and collaborators (Li et al. 2014b) using mouse monoclonal ANTI-FLAG® M2 antibody (Sigma-Aldrich, St Louis, MO) and mouse IgG serum (Sigma-Aldrich,

St Louis, MO), generating ChIP and Mock samples, respectively. Mock samples were generated in the same experimental conditions, except the immunoprecipitation step that was performed with IgG serum instead of the ANTI-FLAG® M2 specific antibody. After reverse cross-link and DNA purification, the precipitated DNA was quantified and analysed by capillary electrophoresis (Agilent 2100 Bioanalyzer) using the High Sensitivity DNA ChIP kit (Agilent, 5067–4626).

3.7. Library preparation and Illumina sequencing

DNA from ChIP of 3xFLAG (ChIP) and from ChIP of IgG serum (Mock) was used to produce sequencing libraries using the KAPA Hyper Prep Kit (KAPA Biosystems, Wilmington, MA). A total of four single-end libraries were prepared, one for each of the two replicates of the ChIP experiment and two of the control (Mock), and sequenced using the HiSeq 4000 system (Illumina, San Diego, CA), in two independent sequencing lanes (L1 and L5), with a read length of 50 bp.

3.8. Sequencing data pre-processing and read mapping

FastQC (Andrews 2010) was used as the first data control tool to check the quality of the sequencing reads. Raw reads were pre-processed with Sickle (Joshi and Fass 2011) using 20 for the quality threshold and 40 for the read length. Mapping of the reads against a draft version of the cork oak genome (Supplementary File 2) was performed with Bowtie2 (Langmead and Salzberg 2012) and unique mapped reads were extracted using the mapping quality value of 255. Library complexity was measured by the non-redundant fraction (NRF) and PCR Bottlenecking Coefficients (PBC). NRF was calculated as described by Nakato and Shirahige (Nakato and Shirahige 2017) by setting the threshold for non-redundant reads to 2. Non-redundant reads were extracted using the output of MACS2 (Zhang et al. 2008) filterdup utility. PBC metrics were obtained using BEDtools (Quinlan and Hall 2010) genomecov. Phantompeakqualtools (Landt et al. 2012) was applied for the cross-correlation analysis of the ChIP and Mock libraries.

3.9. Peak calling and downstream processing

An irreproducible discovery rate (IDR) analysis was employed to find the most consistent set of peaks across replicates. Self-pseudoreplicates and pooled-pseudoreplicates were also generated and peaks were called with Model Based Analysis for ChIP-Seq data (MACS2) (Zhang et al. 2008), using a q-value < 0.05, based on a Poisson distribution comparing the ChIP and Mock sequenced samples. A maximum number of 8 duplicate tags were defined, as it seemed the appropriate value for which the number of peaks reached a plateau phase using the fragment length estimated by Phantompeakqualtools.. IDR v2.0.2 (Li et al. 2011) was used to measure the reproducibility of replicates ranking the peaks by their p-value with an IDR threshold of 0.05. Peaks passing the IDR threshold by comparing true replicates were selected for downstream analysis according to the ENCODE (Encyclopedia of DNA elements) ChIP-Seq guidelines (Landt et al. 2012).

Based on the structural annotation generated for the genome draft (Supplementary File 2), custom python scripts were created to assign a gene to each IDR peak and its respective location regarding gene boundaries. Promotor regions were defined to be up to 2000 bp upstream of the beginning of each gene and terminator regions were set to be up to 1000 bp downstream of the end of each gene.

3.10. Analysis of targets related with transcriptional elements

Identification of target genes encoding for TFs, transcription regulators and chromatin regulators were analyzed with the PlantTFcat tool (Dai et al. 2013) submitting the gene sequences correspondent of each peak to the web platform analysis.

3.11. Motif analysis

A motif discovery analysis was performed using MEME-ChIP software (Machanick and Bailey 2011). MEME-ChIP analysis was done with the gene models available for the cork oak draft genome (Supplementary File 2) and selected based on the location of detected peaks. Sequences with 250 bp from both sides of peaks summits were retrieved and these 501 bp sequences were

given as input in MEME-ChIP software to identify common motifs. Once motifs were identified, the motifs comparison tool TOMTOM (Gupta et al. 2007) was used to compare with already identified and annotated motifs as the Arabidopsis motifs detected in protein microarray (PBM) (Franco-Zorrilla et al. 2014) or Arabidopsis motifs identified by DNA affinity purification sequencing (DAP-Seq) (O'Malley et al. 2017).

3.12. Metabolic pathway analysis

Genomic sequences with 500 bp from both sides of peak summits were checked against the NCBI non-redundant protein (Nr) database, annotated with gene ontology plant-terms, enriched and refined using ANNEX function and used for metabolic pathway analysis using through the Blast2GO software (Conesa and Gotz 2008). Metabolic pathway analysis was then performed using the Blast2GO interface to access Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto 2000). Pathway maps were downloaded and inspected manually focusing in the identified enzymes of pathways related with suberin and lignin metabolism, namely phenylpropanoid biosynthesis, fatty acid biosynthesis and degradation, glycerolipid metabolism and glycerophospholipid metabolism.

4. Results

4.1. QsMYB1::3xFLAG somatic embryo selection for ChIP-Seq

Stable modified cork oak somatic embryo cell lines overexpressing QsMYB1::triple FLAG fused (QsMYB1::3xFLAG) protein were produced using Ce1 line, previously generated (unpublished observation). Ce1 line, stable and with a high rate of secondary embryogenesis, was previously characterized for kanamycin resistance, allowing the determination of the somatic embryo natural resistance to grow under the presence of the antibiotic. Concentrations over 25 mg/mL of kanamycin showed to inhibit embryo proliferation setting kanamycin at this concentration as selective agent of putative transformants. After 30 days subcultures during 24 months, MYB1::3xFLAG transcripts were

detected in several transformed embryo clusters by quantitative real time PCR (RT-qPCR). When comparing with the expression of *QsMYB1* in non-transformed embryos, MYB1::3xFLAG transcripts presented a 4-12 fold change in the level of expression (Figure IV. 1A). Furthermore, the specificity of the anti-FLAG antibody was validated by western blot (Figure IV. 1B) and the MYB1::3xFLAG protein was detected in the nucleus by fluorescence microscopy (Figure IV. 1C), confirming the nuclear location of the protein. Due to the lack of known target DNA of QsMYB1 and consequent impossibility to control the enrichment quality of ChIPed DNA, we selected embryo clusters with fold change expression of *QsMYB1::3xFLAG* higher than 6 (Figure IV. 1A) in order to have enough protein suitable to perform the ChIP-Seq assay.

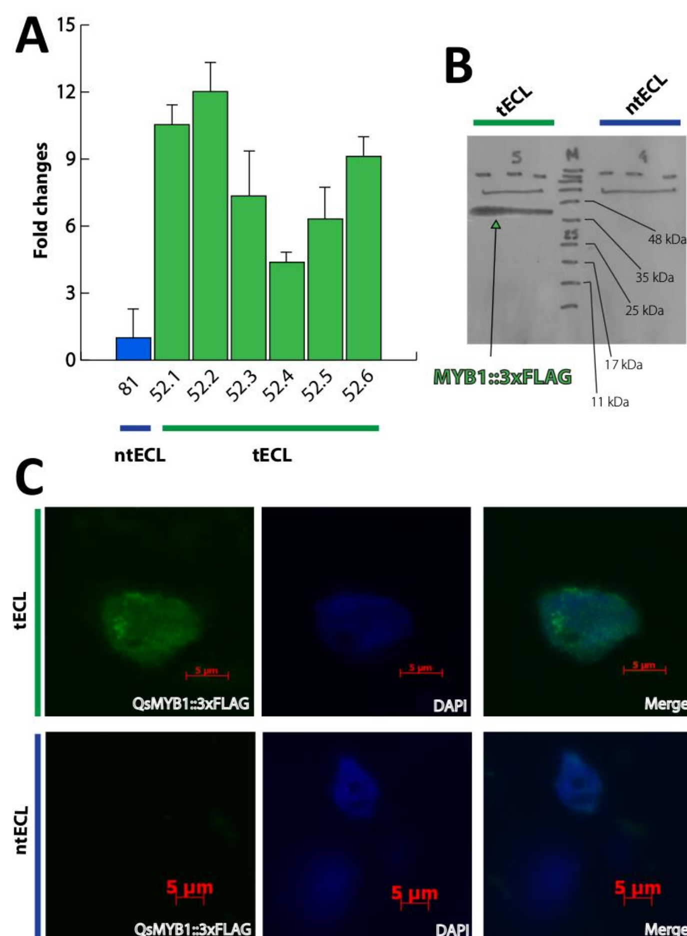


Figure IV. 1 | MYB1::3xFLAG transcripts and protein detection. **(A)** MYB1::3xFLAG transcripts levels from six transformed embryogenic cell lines (tECL) clusters (52.1-52.6) quantified by RT-qPCR and normalised against non-transformed embryogenic cell lines (ntECL). Mean and standard errors of three independent experiments are shown. **(B)** Western Blot of MYB1::3xFLAG proteins in tECL. Total proteins from non-transformed ECL (ntECL) were used as negative control. **(C)** MYB1::3xFLAG protein nuclear detection by fluorescence microscopy in tECL. Nucleus magnification is shown for tECL and ntECL.

4.2. QsMYB1target genes identified by ChIP-Seq

In order to identify the DNA targets of QsMYB1, ChIP-Seq was performed by immunoprecipitating QsMYB1 protein and the cross-linked DNA after chromatin fragmentation of ~300 bp (Supplementary File 3). DNA was purified and further sequenced with the Illumina HiSeq 4000 system producing ~70-90 million reads per sample of which ~ 123 million reads uniquely mapped to the cork oak draft genome (Table IV. 1). One ChIP (anti-FLAG) and one mock (IgG) library were sequenced with one technical duplicate each. A total of 18,165 putative binding sites were identified. It was firstly analysed whether the peaks in the ChIP experiment were distributed by genic regions. Based on these analyses, MYB1 binding sites are located in genic regions of 14,290 genes: 13.4% in promotor regions, 8.1% in 5' untranslated regions (UTR), 19.2% in intron regions, 25.5% in exon regions, 6.5% in 3' UTR and 6.0% in the terminator regions (Figure IV. 2A). Intergenic regions represent 18.4% of the total binding sites and 2.9% of the binding sites were not annotated (unknown) (Figure IV. 2A). The Peak-calling analysis identified several target genes reflecting the binding of QsMYB1 to specific DNA loci. Between these genes, QsMYB1 targets other TFs, genes encoding for enzymes related with the lipid metabolism and transport, as well as enzymes from the phenylpropanoid pathway. Amongst these genes, several are related with various aspects of cork formation, as suberin and lignin biosynthesis or transport and deposition of suberin monomeric units. In order to explore the regulatory functions of QsMYB1 in cork formation and differentiation we focused the analysis in these genes, although QsMYB1 potentially targets a vaster number of genes.

Table IV. 1 | Sequencing throughput and mapping results obtained for ChIP and mock samples.

Sample	Raw reads	Pre-processed	UMR reads	% UMR
ChIP_L1	75,987,297	72,855,787	28,558,358	37.58%
ChIP_L5	66,520,992	63,594,856	24,887,132	37.41%
Mock_L1	93,072,373	89,005,882	37,335,088	40.11%
Mock_L5	80,189,676	76,512,503	32,053,662	39.97%

4.3. Binding Motif Analysis Reveals QsMYB1 *cis*-regulatory Elements

To explore the MYB1 biding motifs environment, 1-kb flanking sequences around all peaks summit were analysed by the motif discovery tool MEME-ChIP (Figure IV. 2B). Motifs CWHCAA (E-value=3.5e-50), CYTCBTC (E-value=8.8e-38) and BKTGG (E-value=2.9e-30) were the most statistically relevant. The BKTGG motif is presented in 12,308 peaks (67.9%), while the CWHCAA and the CYTCBTC motifs were identified in 7,926 (43.7%) and 2,016 (11.1%) peak sequences, respectively. When the occurrence of motifs by binding locations is individually analysed, it is observed that some genic regions present more than one identity motif. Nevertheless, each binding feature putative motif has a similarity with the 3 global motifs identified (Figure IV. 2A).

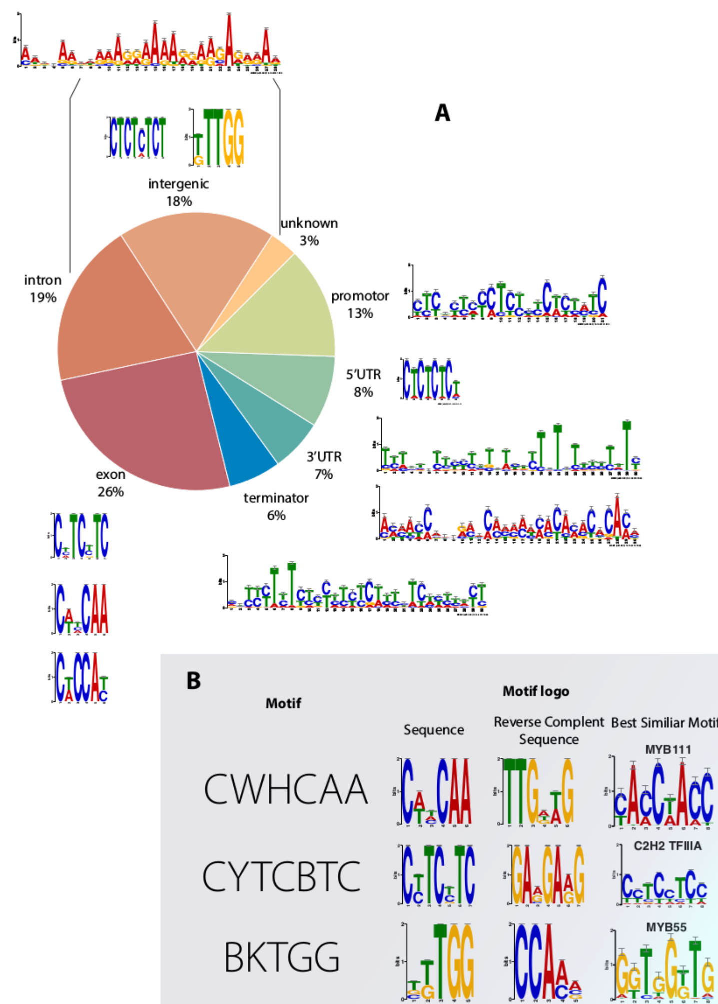


Figure IV. 2 | (A) Distribution of peaks by genic and intergenic regions and correspondent motifs of QsMYB1 preferentially binding. (B) Summary of motifs analysis of all peaks. The three more representative motifs of all peaks detected and the best similar motif known are represented.

4.4. QsMYB1 directly targets other transcriptional elements

To detect genes regulated by QsMYB1 gene, the QsMYB1 DNA-targets putatively encoding for TFs, transcription regulators and chromatin regulators were analysed with the PlantTFcat tool. Results showed that 414 regulatory genes are targeted by QsMYB1: TF (42%), transcription regulators (50.5%) and chromatin regulators (7.5%) comprising different types of regulators families (Figure IV. 3). From these families the most representative putatively encodes for CCHC zinc finger proteins (CCHC) (192), C2H2-type zinc fingers TFs (C2H2) (43), WD40-like TFs (19), BED-type zinc finger proteins (BED-type(Zn)) (13), AP2/ERF TFs (10) and bHLH TFs (10) (Supplementary File 4).

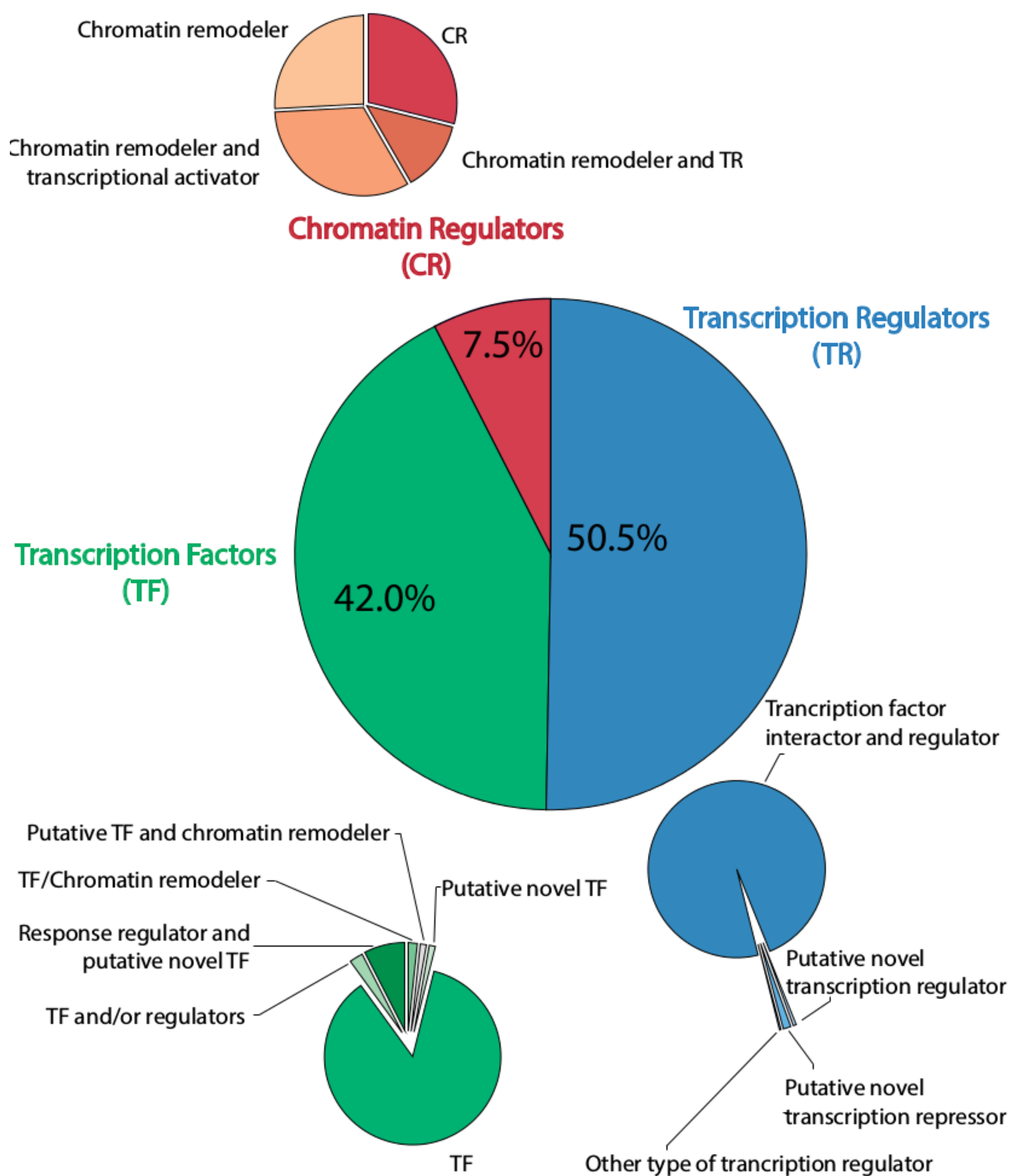


Figure IV. 3 | Types of transcription factors, transcription regulators and chromatin regulators targeted by QsMYB1.

4.5. Enzyme encoding genes targeted by QsMYB1

To identify the targets of QsMYB1 directly related with secondary growth and cork formation, all QsMYB1 putative target genes were submitted to the KEGG database giving focus to the ones that present a fold enrichment higher or equal to 3 (Supplementary File 5). With this approach, 16 genes encoding three distinct enzymes essential to the phenylpropanoid pathway (Figure IV. 4) were identified, namely 4-coumarate: CoA ligase (4CL - 6.2.1.12), cinnamyl alcohol dehydrogenase (CAD - 1.1.1.195) and class III plant peroxidase (PPO - 1.11.1.7). Also related with the phenylpropanoid metabolism, the results show that QsMYB1 putatively targets three distinct genes encoding for β-galoctosidase (β-GAL - 3.2.1.21).

Diverse QsMYB1 target genes encoding for key enzymes on the metabolism of lipids were identified, namely enzymes related with fatty acid biosynthesis (Figure IV 5) as acetyl-CoA carboxylase (ACC – 6.4.1.2), long-chain-fatty-acid-CoA ligase (LACS – 6.2.1.3) and members of fatty acid synthase complex: enoyl-[acyl-carrier-protein] reductase (NADH) (FABI – 1.3.1.9) and acyl-[acyl-carrier-protein] desaturase (ACP - 1.14.19.2). Moreover, we found two Arabidopsis homolog genes of β-ketoacyl-CoA synthase in our data. Related with fatty acid oxidation (Figure IV. 6), the results revealed two genes encoding for a putative acyl-CoA oxidase (ACX – 1.3.3.6) and two genes encoding for a aldehyde dehydrogenase (NAD+) (ALDH - 1.2.1.3). Also, two distinct target genes related the ω-hydroxylation of various fatty acids and with high homology for the Arabidopsis Cytochrome P450 (CYP) 86A8 (*AtCYP86A8*) and for the Arabidopsis CYP 96A1 (*AtCYP96A1*) were found.

The results additional show several genes encoding for enzymes involved in the glycerol metabolism (Figure IV. 7) as aldehyde dehydrogenase (NAD+) (ALDH - 1.2.1.3), NADP-dependent alcohol dehydrogenase (ADH – 1.2.1.1), NADPH-dependent aldehyde reductase (ADR – 1.1.1.21), glycerol 3-phosphate acyltransferase (GPAT – 2.3.1.15), 1,2-diacyl-3-β-D-galactosyl-*sn*-glycerol acylhydrolase (DGL – 3.1.1.26), triacylglycerol acylhydrolase (LIP - 3.1.1.3) and diacylglycerol *O*-acyltransferase (DGAT - 2.3.1.20).

Figure IV. 4 | Enzymes targeted by QsMYB1 in the phenylpropanoid biosynthesis pathway (coloured boxes). Numbers correspond to enzymes E.C. numbers.

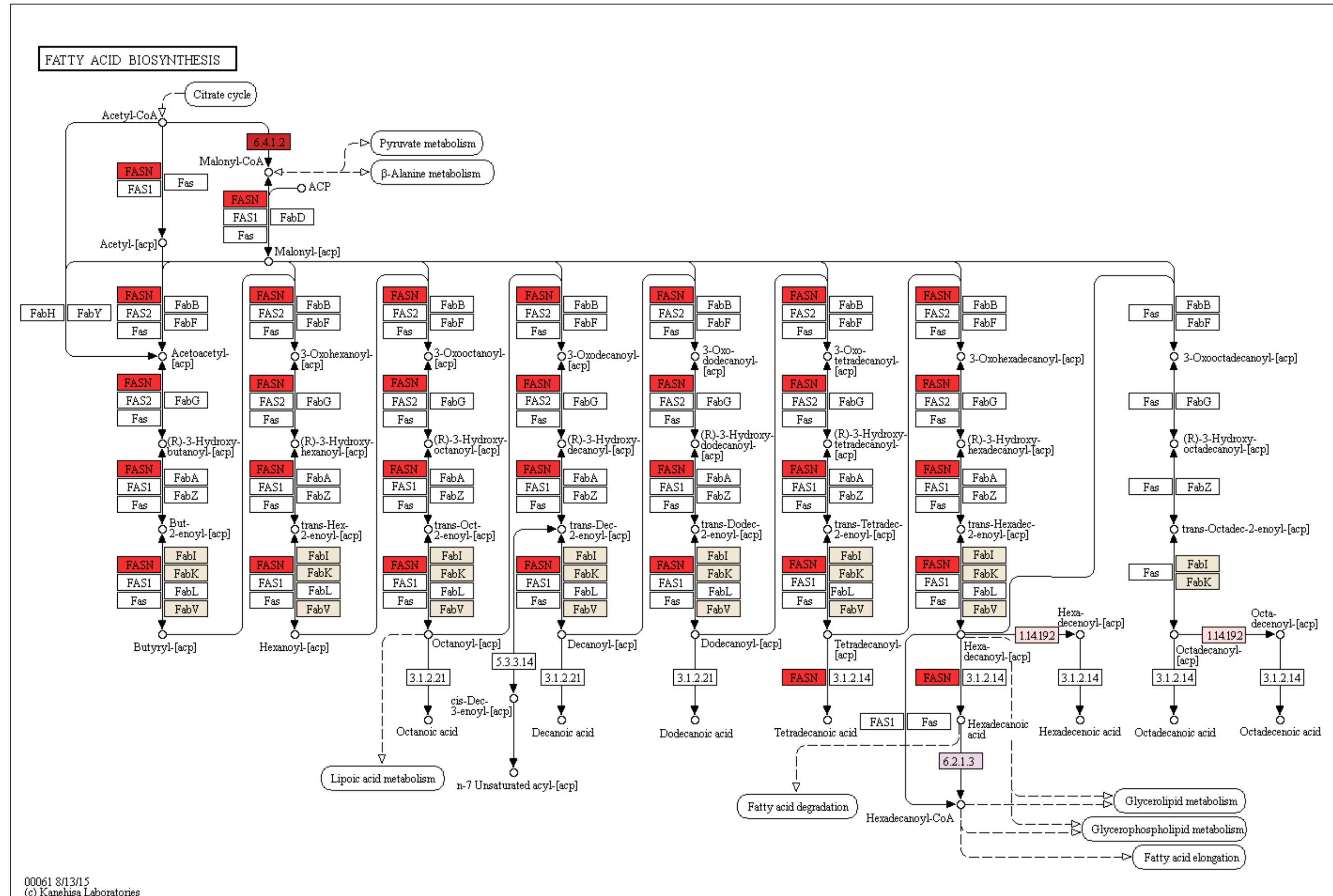


Figure IV. 5 | Enzymes targeted by QsMYB1 in the fatty acid biosynthesis pathway (coloured boxes). Numbers correspond to enzymes E.C. numbers.

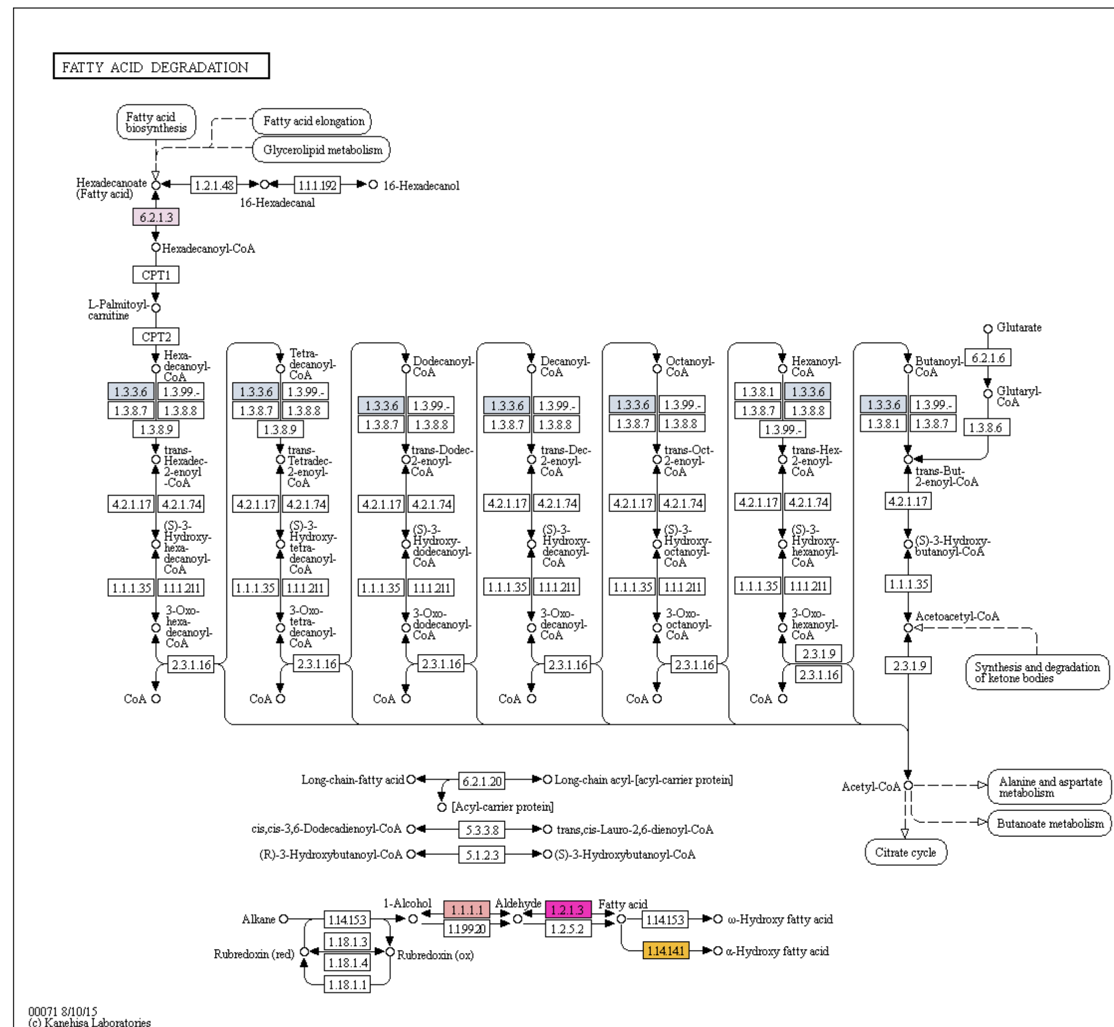


Figure IV. 6 | Enzymes targeted by QsMYB1 in the fatty acid degradation biosynthesis pathway (coloured boxes). Numbers correspond to enzymes E.C. numbers.

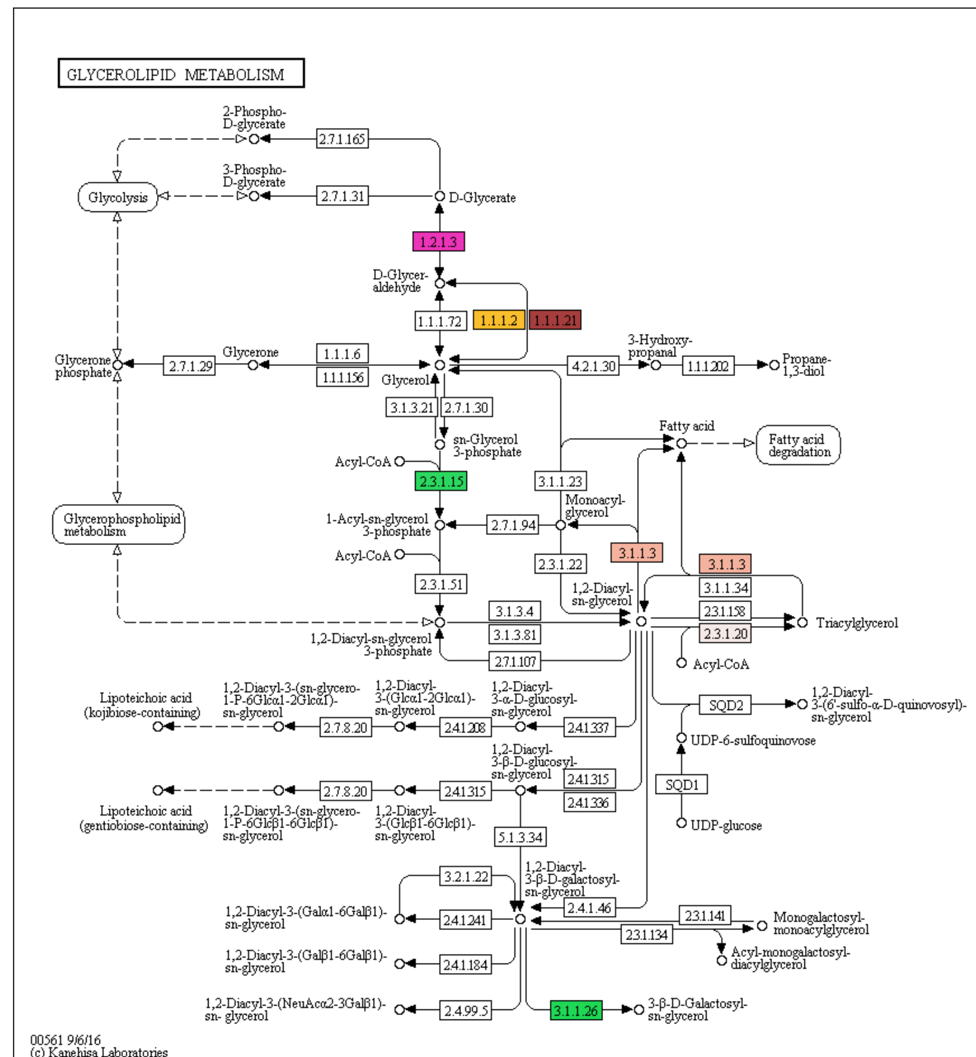


Figure IV. 7 | Enzymes targeted by QsMYB1 in the glycerolipid metabolism (coloured boxes). Numbers correspond to enzymes E.C. numbers.

Figure IV. 8 | Enzymes targeted by QsMYB1 in the glycerophospholipid metabolism (coloured boxes). Numbers correspond to enzymes E.C. number

Several genes encoding for enzymes related with the glycerophopspholipid metabolism (Figure IV. 8) were also identified: glycerol-3-phosphate dehydrogenase (NAD+) (GPDH - 1.1.1.8), glycerol-3-phosphate dehydrogenase [NAD(P)+] (GPDH - 1.1.1.94), phospholipase A1 (PLA1 - 3.1.1.32) and the phosphoethanolamine N-methyltransferase (NMT – 2.1.1.103).

4.6. MYB1 target genes related with lipid transporters

Our ChIP-Seq data reveals that QsMYB1 is targeting several ATP-binding cassette proteins (ABC) or ABC-like transporters. Of these, 26 present high homology for the G family of ABC (ABCG) genes (Supplementary File 6). Five of these genes have high similarity with *AtABCG11*, six of them with *AtABCG37* and six with the *AtABCG40*. In addition, we identified five genes encoding for lipid-transfer proteins (LTPs) (Supplementary File 7), three of them with similarity to protease inhibitor/seed storage/LTPs and two of them putative LTPs genes in *Q. suber*.

5. Discussion

Quercus suber MYB1 (QsMYB1) is a TF that is mainly expressed in organs and tissues with secondary growth resulting from the activity of the phellogen (Almeida et al. 2013a). This points to a putative regulation of QsMYB1 in the lignin and suberization processes and therefore possibly in cork formation and differentiation. In fact, *Arabidopsis* mutants generated by gene trap insertion in *AtMYB68,* the *QsMYB1* orthologue, produce increased biomass and lignin levels relatively to wild type. In addition its closest homolog, *AtMYB84,* exhibit an overlapping expression suggesting a partly redundant function (Feng et al. 2004).

In order to identify the genes regulated by QsMYB1 in a genome-wide scale, a cork oak embryo model was established, allowing protein overexpression, isolation and easy access to genomic material. In this context, genetic modified cork oak embryos overexpressing QsMYB1 were produced and used in a ChIP-Seq experiment allowing for the identification of the QsMYB1 putative binding sites. Although this is a proxy model, as embryos do not present neither the

developmental stage nor the secondary growth tissues where *QsMYB1* is naturally up-regulated, it allowed for the identification of several QsMYB1 target genes with specific and defined DNA binding domains related with several aspects of tissue lignification, suberization and consequently with cork formation, which are further discussed. Furthermore QsMYB1 may not target all the binding sites as at the secondary tissues where it has been found to be expressed (Almeida et al. 2013a). However, the overexpression system used in this study, which forces the production of QsMYB1, is an attempt to mimic the up-regulation and DNA-binding status observed in other tissues or organs (Almeida et al. 2013a).

5.1. QsMYB1 targets phenylpropanoid pathway genes regulating the biosynthesis of essential lignin and suberin precursors

The results of our study clearly show that QsMYB1, belonging to R2R3-MYB family, is targeting genes coding for key enzymes responsible for the biosynthesis of phenylpropanoids, which constitute phenolic components of suberin and lignin polymers. Members of the R2R3-MYB TF family are well known for their regulatory function in diverse metabolisms as the phenylpropanoid (Liu et al. 2015), lignin (Tian et al. 2013; Li et al. 2014a) and lipid (Liu et al. 2014) metabolism among others. Phenylpropanoids are phenolic derivatives which contain a phenyl ring and a C_3 side chain that comprise a multitude of plant secondary metabolites and cell wall components as lignin and suberin (Graça and Santos 2007; Vanholme et al. 2010). The *Arabidopsis MYB41,* phylogenetically close to *AtMYB68* (the Arabidopsis homolog of *QsMYB1*), when overexpressed has an up-regulatory effect on phenylpropanoid, lignin and suberin synthesis and related-genes expression (Kosma et al. 2014). Like AtMYB41, our results indicate an involvement of QsMYB1 in the phenylpropanoid metabolism. QsMYB1 targets 4-coumarate:CoA ligase which is involved in the conversion of ferrulic, caffeic and coumaric acids to ω-OH-alkyl-ferulates/n-alkyl ferulates, n-alkyl caffeates and n-alkyl coumarates, that in turn are linked to fatty alcohols to produce alkyl hydroxycinnamates, an essential constituent in suberin-associated waxes (Vishwanath et al. 2015). In addition, QsMYB1 targets cinnamyl alcohol

dehydrogenase and class III plant peroxidase, which together with 4-coumarate:CoA are involved in the biosynthesis of the monomeric precursor of lignin *p*-hidroxy-cinnamyl alcohols, the so-called monolignols. Furthermore class III plant peroxidases are also involved in polymerization of suberin monomers in the apoplast (Arrieta-Baez and Stark 2006). Together, the results indicate

an involvement of QsMYB1 in the synthesis of phenolic compounds from the phenylpropanoid pathway evidencing an earlier regulatory action in the synthesis of essential lignin and suberin precursors.

5.2. QsMYB1 is targeting genes encoding enzymes responsible for suberin biosynthesis

Suberin may be defined as a complex glycerol-based polymer consisting of a polyaliphatic polyester linked with phenolic components and embedded waxes (Bernards 2002; Franke and Schreiber 2007; Pollard et al. 2008). Besides suberin composition vary between developmental stages, tissue and plant species, in *Q. suber* depolymerisation of cork suberin include monoacylglycerols of ω-hydroxy acids and α,ω-dicarboxilic acids, ferrulic acid linked to ω-hydroxi acids, trimeric diesters of glycerol linked to α,ω-dicarboxilic acids and ω-hydroxi acids linked to ferrulic acid and glycerol (Graça and Pereira 2000; Graça and Santos 2006). Our results reveal that QsMYB1 is targeting genes that encode enzymes involved in synthesis of the suberin polyester building blocks; one *QsLACS* and two *QsKCS* predicted genes, constituting evidence of genes associated with suberin monomers synthesis in *Q. suber*. LACS catalyses the acyl activation reaction of free fatty acids to fatty thiosteres (Shockey et al. 2002). In *Arabidopsis LACS1* and *LACS2* have overlapping roles activating fatty acids in cutin and wax synthetic pathways. In addition, *LACS2* is involved in suberin formation besides the capability of LACS enzymes may also act on modified fatty acids as ω-hydroxy acids and α,ω-dicarboxilic acids before esterification to glycerol (Li-Beisson et al. 2013). In turn, β-ketoacyl-CoA synthase (KCS) is responsible for controlling the extension of elongation of the long-chain fatty acyl-CoAs (Millar and Kunst 1997). In *Arabidopsis*, two β-ketoacyl-CoA synthases, *AtKCS2* and *AtKCS20* are

involved in elongation of C_{20} acyl chain suberin precursors (Franke et al. 2009; Lee et al. 2009). In potato, silencing of *StKCS6* gene resulted in reduction of suberin monomers with chain lengths of C_{28} and higher in the tuber periderm indicating that StKCS6 is involved in elongation of suberin precursors to C_{28} or higher chain lengths (Serra et al. 2009a).

Several CYP86 family members were identified in our data, namely homologues of *AtCYP86A1* and of *AtCYP86A8.* CYP86A8 monooxigenase are capable of ω-hydroxylating saturated and unsaturated fatty acids with chain lengths from C_{12} to C_{18} and is likely to be involved in cutin biosynthesis (Wellesen et al. 2001). Two major monomers of the suberin polyester are ω-hydroxy acids and α,ω-dicarboxylic acids, which in plants are formed by the hydroxiliation of the terminal methyl group (ω-position) catalysed by enzymes of the CYP86 subfamily of cytochrome P450 monooxygenases (Höfer et al. 2008). The *Arabidopsis* CYP86A1 and CYP86B1 are involved in root suberin synthesis (Li et al. 2007; Höfer et al. 2008; Compagnon et al. 2009). Similarly, in potato, the StCYP86A33 orthologue of AtCYP86A1 is important for the production of ω-functionalized suberin monomers of the tuber periderm resulting in significant reduction in 18:1 ω-hydroxy acids and α,ω-dicarboxilic acids when the gene is silenced (Serra et al. 2009b).

Our results reveal a *QsGPAT* homolog target by QsMYB1 indicating a regulatory role of this TF in the production of monoacylglycerols, which may be considered as the initial step in the aliphatic suberin biosynthesis. In *Arabidopsis*, GPAT4, 6 and 8 are bifunctional enzymes with *sn*-2 acyltransferase and phosphatase activity, catalysing the formation of *sn*-2 monoacylglycerol products. GPAT5 in turn is involved in the transfer of very long-chain aliphatics to a glycerol based acceptor in *Arabidopsis* root and seed coat suberin (Beisson et al. 2007). *Arabidopsis* mutants of *GPAT5* show a large reduction in long-chain unsubstituted fatty acids, ω-hydroxy acids and α,ω-dicarboxilic acids in root and seed coat suberin with a total reduction of 50% in total suberin content (Beisson et al. 2007). In the case of GPAT7, the *GPAT7* gene is induced by wounding producing soluble suberin-like precursors in overexpressing lines (Yang et al. 2012). These results clearly indicate that QsMYB1 is targeting genes encoding for enzymes that are responsible for the production of aliphatic domain of suberin.

5.3. QsMYB1 targets putative transporters of suberin-lipid components

By directly targeting genes coding for 4-coumarate:CoA, cinnamyl alcohol dehydrogenase, LACSs, KCSs, GPATs amongst others, QsMYB1 is likely to regulate the phenylpropanoid and the suberin biosynthesis pathways. In addition, QsMYB1 may have a regulatory effect in the expression of suberin transporters.

The various suberin lipidic precursors are generically transported from the endoplasmic reticulum where they are synthesized to and across the plasma membrane, and then polymerized in the apoplasm to form the suberin barrier (Li et al. 2016). We identified several members of the ABCG gene family targeted by QsMYB1, namely homologs of *AtABCG11* and *AtABCG6,* which implicates directly QsMYB1 in regulation of lipid transport and suberin formation. Transport of aliphatic lipids across the plasma membrane involves plasma membrane-localized transporters of the ABCG family which consists of half-transporters that oligomerise to form the functional transporter (McFarlane et al. 2010). In *Arabidopsis*, the ABCG transporter ABCG11 is implicated with cutin and suberin biosynthesis (Bird et al. 2007; Panikashvili et al. 2007; Panikashvili et al. 2010) while ABCG12 and ABCG13 are involved in export of cuticular wax monomers (Panikashvili and Aharoni 2008; Panikashvili et al. 2010). Because homo- and heterodimers of ABCG11 and ABCG12 are reported as important for wax export and ABCG13 seems to be required for cutin deposition in flowers (Panikashvili et al. 2011), it is widely accepted that tissue-dependent combinations of different half-size ABCG proteins involved in the transport of wax, cutin and suberin components are possible to occur (Li et al. 2016). Furthermore, ABCG11 can form dimers with ABCG9 and ABCG14 which links it to a function in lipid homeostasis regulation in vascular organs as phloem (Le Hir et al. 2013). Additionally, a recent study reported that half-size ABCG2, ABCG6 and ABCG20 are involved in the formation of suberin layers in seed coats and roots where *abcg2/6/20* triple mutants had increased permeability, altered suberin structure and reduce aliphatic components (Yadav et al. 2014). Another type of proteins involved in lipid excretion into the extracellular space are LTPs. Evidence of LTPs involvement in cuticular wax,

suberin and sporopollenin assembly is increasing in recent literature (Salminen et al. 2016; Deeken et al. 2016). Our data also identified five genes coding for LTPs reinforcing the involvement of QsMYB1 in the regulation of genes involved in lipid traffic across the cellular membrane and possibly in suberin assembly and deposition.

5.4. Putative regulation of glycerolipids metabolism by QsMYB1

Glycerolipids are fatty acid esters of glycerol. Fatty acids are constituents of membrane lipids of every cell. The first step in *the novo* synthesis of fatty acids is the carboxylation of acetyl-CoA to malonyl-CoA by acetyl-CoA carboxylase (Heldt and Piechulla 2011). In *Arabidopsis* acetyl-CoA carboxylase is required for very long chain fatty acid elongation (Baud et al. 2003). Our data suggests that QsMYB1 is regulating a homolog gene of acetyl-CoA carboxylase, which indicates direct regulation on a fundamental key step of fatty acid synthesis and elongation pathways. The second enzyme responsible for fatty acids biosynthesis is the fatty acid synthase (FAS) which in plants is a type II FAS consisting of a multienzyme complex of acyl carrier proteins (Heldt and Piechulla 2011). The results also revealed that at least 2 genes with homology for *Arabidopsis* enoyl-[acyl-carrier-protein] reductases (NADH) are QsMYB1 gene targets. One acyl-[acyl-carrier-protein] desaturase was also identified in our data. This type of desaturases are responsible for fatty acid modifications introducing *cis*-double ligation in the acyl chain acyl-[acyl-carrier-protein] and ultimately in the correspondent fatty acid (Heldt and Piechulla 2011). Therefore, QsMYB1 may regulate the production of unsaturated fatty acids, which in turn is used as precursors for the suberin monomeric units synthesis.

Suberin, in cork oak, presents monoacylglycerols, diglycerol diesteres and diacylglycerols in is chemical composition (Graça and Santos 2006; Graça and Santos 2007). The availability of glycerol-derived molecules may constitute a critical limitation step in glycerol-derived monomers synthesis of suberin (Beisson et al. 2007). In this study, were identified two genes coding for aldehyde dehydrogenases which are responsible for the production of D-glyceraldehyde, which in turn is converted to glycerol by aldehyde reductase

and alcohol dehydrogenase (both also identified in our data) to form glycerol in plants (Heldt and Piechulla 2011). These results indicate that QsMYB1 is acting on transcription regulation of genes involved in the synthesis of glycerol. In addition, the results revealed that QsMYB1 is exerting its action on a gene encoding for diacylglycerol O-acyltransferase which is responsible for the conversion of diacylglycerols to triacylglycerols (Li et al. 2010). Interestingly, QsMYB1 also seems to act on the opposite conversion reaction of triacylglycerols to the correspondent diacylglycerols by targeting the homolog gene of a triacylglycerol acylhydrolase. Triacylglycerol acylhydrolases are also described to have the ability for catalyzing the reaction of diacylglycerols to monoacylglycerols (El-Kouhen et al. 2005; Seo et al. 2009). This finding may constitute an evidence for the transcription regulation of diacylglycerols and monoacylglycerols synthesis in *Q. suber*, however further studies have to be performed in order to confirm this hypothesis.

5.5. QsMyb1 may be involved in abiotic stress responses through the regulation of the glycerophospholid pathway

Glycerophospholipids are major components of cellular membranes which are synthesized from glycerol-3-phosphate and have functions as second messengers in plant growth regulation and cellular response to environmental change or stress (Heldt and Piechulla 2011). Lipid membrane composition, remodelling and activation of a variety of phospholipid-based pathways is a strategy developed by plants to survive and adapt to osmotic stress (Munnik and Meijer 2001).

Almeida *et al.* reported a putative function of QsMYB1 in the regulatory network of cork oak response to heat and drought stresses (Almeida et al. 2013b). Accordingly, we found that QsMYB1 is targeting a gene coding for phospholipase A1, a *sn*-1-specific phospholipase that releases free fatty acids from phospholipids and can act on the cellular membrane galactolipids and triacylglycerols (Wang et al. 2012). In *Arabidopsis,* the *DEFECTIVE IN ATHER DEHISCENCE1 (DAD1)* gene has been shown to code phospholipase A1 catalyzing the initial step of jasmonic acid biosynthesis (Ellinger et al. 2010). The *Arabidopsis* phospholipase coded by the At2g42690 gene, which is UV-B

inducible, was also reported as having properties that suggest a possible involvement in the biosynthesis of jasmonic acid (Lo et al. 2004). These studies point to involvement of phospholipases A in hormonal response to stress and may indicate evidence of a relation between QsMYB1 and response to stress mediated by hormones. Furthermore, our results revealed that QsMYB1 is targeting a *phosphoethanolamine N-methyltransferase* homolog gene of *Arabidopsis* which catalyzes key steps in choline biosynthesis, namely the N-methylation of phosphoethanolamine (Cruz-Ramirez et al. 2004). Choline is well reported as an important player in plant grow and development (Lin et al. 2015). Choline is also oxidized to glycinebetaine which is a strong osmoprotectant that confers tolerance to salinity, drought and other stress (Waditee et al. 2005; Shirasawa et al. 2006). In *Arabidopsis*, the silencing *of phosphoethanolamine N-methyltransferase* results in temperature-sensitive male sterility and salt hypersensitivity, which demonstrates that choline biosynthesis also contributes to tolerance to environmental stress (Mou et al. 2002). Our results suggest an involvement of QsMYB1 response to drought stress specifically involving a key choline biosynthesis enzyme supporting the hypothesis by Almeida *et al.* that QsMYB1 expression is modulated in response to heat and drought stresses (Almeida et al. 2013b).

5.6. QsMYB1 targets other transcriptional elements expanding its regulatory action

Three motifs ([CWHCAA], [CYTCBTC] and [BKTGG] were identified as the principal motifs characteristic of the QsMYB1-DNA binding. The [CWHCAA] and [BKTGG] motifs present high homology with already described R2R3-MYB binding site motifs (Franco-Zorrilla et al. 2014), while [CYTCBTC] has a similarity with zinc finger C2H2-type motif of TF IIIA, which in *Arabidopsis* encodes a protein required for transcription of 5S ribosomal RNA (Layat et al. 2012). However, the presence of secondary motifs signatures is a constant in the dataset. Secondary motifs results from two events: two TFs binding to neighbouring sites, or one TF binding to another TF that in turn binds to DNA. Taking into account that members of the MYB family are well known by acting as homo- or heterodimers (Pireyre and Burow 2015), the data suggest that

QsMYB1 is either binding to DNA or to other TFs resulting in an extensive and complex transcriptional regulation of genes. Therefore, QsMYB1 may be involved in a complex regulatory network targeting both transcriptional elements and a group of specific genes with direct function in several biological processes. We found that QsMYB1 is targeting other transcription elements, which supports the idea that QsMYB1 directly regulates the expression of genes and also exerts its effect through other TFs, transcription regulators and chromatin regulators. Amongst these, the more abundant gene targets of QsMYB1 are involved: in fine tuning transcription, translation and post-transcriptional modifications as CCHC zinc finger proteins (CCHC) family, which specifically interacts with single-stranded DNA or RNA oligonucleotides (Brown 2005) or C2H2-type zinc fingers TFs (C2H2) which besides DNA binding also provides protein-protein and RNA-protein interactions (Razin et al. 2012) or BED-type zinc finger proteins, which function as either transcription activators or repressors by modifying local chromatin structure on binding to GC-rich sequences (Saghizadeh et al. 2009; Aravind 2017); in regulation of cell division, cell-fate determination, transmenbranar signalling and vesicle fusion as the WD40-like TFs; and in integrating gene regulatory networks which controls the metabolic, hormonal and environmental signals in plant growth, development and response to plant abiotic and biotic stresses as the AP2/ERF TF (Dietz et al. 2010; Chen et al. 2016) and the bHLH TF families (Pireyre and Burow 2015).

Together with MYB, bHLH represents a large proportion of the TFs in *Arabidopsis* both regulating multiple biological process (Feller et al. 2011). Members of the two families are described by acting as homodimers or heterodimers consisting of proteins from the same family or in complexes with TFs from other families including the WD40-like TFs previously referred (Feller et al. 2011). In many cases, MYB/bHLH complexes have been described and associated with the developmental and metabolic plasticity present in plants (Feller et al. 2011). The regulation of pathways controlled by these two types of TFs as well as the mechanism regulating their activity is therefore linked, which is in line with our results. Thus, by forming heterodimers with other transcriptional elements, QsMYB1 may trigger additional regulatory mechanisms at chromatin, transcription and post-transcriptional level.

6. Conclusions

This study has contributed to increase the knowledge about the molecular mechanisms behind cork formation and differentiation. To address the role of QsMYB1 we performed a genome-wide analysis of the DNA targets of this TF using ChIP-Seq. QsMYB1 was shown to be involved in lignin and suberin biosynthesis and assembly, acting through a complex regulatory network directly by targeting enzyme genes, and indirectly by targeting genes encoding for other transcriptional regulatory elements. The complex QsMYB1 regulatory mechanisms that we disclosed may represent an example of how *Q. suber* has optimised its developmental regulatory systems during evolution. Furthermore, we found QsMYB1 as a master regulatory factor of cork formation and differentiation since it acts on genes belonging to the lignin and suberin biosynthetic pathways, the two main components of cork. In addition, QsMYB1 has the capability to modulate the process of lignin and suberin synthesis through the regulation of specific genes of the phenylpropanoid pathway. Further studies are required to confirm and reveal the specific role of the reported genes and to clarify the exact function of QsMYB1 in the process. Even so, the generated data is a starting point to an in-depth understanding of suberin biosynthesis and deposition. By starting to unravel the biosynthetic pathways involved in cork formation we provide important knowledge to support the development of breeding programs based e.g, on the design of cork metabolic engineering strategies.

7. 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. J Plant Physiol 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 Physiol Biochem 73:274–281

Álvarez R, Ordás RJ (2007) Improved genetic transformation protocol for cork oak (Quercus suber L.). Plant Cell Tissue Organ Cult 91:45–52

Ambawat S, Sharma P, Yadav NR, Yadav RC (2013) MYB transcription factor genes

as regulators for plant responses: An overview. Physiol Mol Biol Plants 19:307–321

Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. Babraham Bioinforma http://www.bioinformatics.babraham.ac.uk/projects/

Aravind L (2017) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. Trends Biochem Sci 25:421–423

Aronson J, Pereira JS, Pausas JG (2009) Cork Oak Woodlands on the Edge: Ecology, Adaptive Management, and Restoration. Island Press, Washington, DC

Arrieta-Baez D, Stark RE (2006) Modeling suberization with peroxidase-catalyzed polymerization of hydroxycinnamic acids: cross-coupling and dimerization reactions. Phytochemistry 67:743–753

Atanassov II, Atanassov II, Etchells JP, Turner SR (2009) A simple, flexible and efficient PCR-fusion/Gateway cloning procedure for gene fusion, site-directed mutagenesis, short sequence insertion and domain deletions and swaps. Plant Methods 5:14

Baud S, Guyon V, Kronenberger J, Wuillème S, Miquel M, Caboche M, et al. (2003) Multifunctional acetyl-CoA carboxylase 1 is essential for very long chain fatty acid elongation and embryo development in Arabidopsis. Plant J 33:75–86

Beisson F, Li Y, Bonaventure G, Pollard M, Ohlrogge JB (2007) The Acyltransferase GPAT5 Is Required for the Synthesis of Suberin in Seed Coat and Root of Arabidopsis. Plant Cell 19:351–368

Bernards MA (2002) Demystifying suberin. Can J Bot 80:227–240

Bird D, Beisson F, Brigham A, Shin J, Greer S, Jetter R, et al. (2007) Characterization of Arabidopsis ABCG11/WBC11, an ATP binding cassette (ABC) transporter that is required for cuticular lipid secretion. Plant J 52:485–498

Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol 17:1519–33

Brown RS (2005) Zinc finger proteins: Getting a grip on RNA. Curr. Opin. Struct. Biol. 15:94–98

Bugalho MN, Caldeira MC, Pereira JS, Aronson J, Pausas JG (2011) Mediterranean cork oak savannas require human use to sustain biodiversity and ecosystem services. Front Ecol Environ 9:278–286

Chen L, Han J, Deng X, Tan S, Li L, Li L, et al. (2016) Expansion and stress responses of AP2/EREBP superfamily in Brachypodium distachyon. Sci Rep 6:21623

Compagnon V, Diehl P, Benveniste I, Meyer D, Schaller H, Schreiber L, et al. (2009) CYP86B1 Is Required for Very Long Chain ω-Hydroxyacid and α,ω-Dicarboxylic Acid Synthesis in Root and Seed Suberin Polyester. Plant Physiol 150:1831–

1843

Conesa A, Gotz S (2008) Blast2GO: A comprehensive suite for functional analysis in plant genomics. Int J Plant Genomics 2008:619832

Cruz-Ramirez A, Lopez-Bucio J, Ramirez-Pimentel G, Zurita-Silva A, Sanchez-Calderon L, Ramirez-Chavez E, et al. (2004) The xipotl mutant of Arabidopsis reveals a critical role for phospholipid metabolism in root system development and epidermal cell integrity. Plant Cell 16:2020–2034

Dai X, Sinharoy S, Udvardi M, Zhao PX (2013) PlantTFcat: an online plant transcription factor and transcriptional regulator categorization and analysis tool. BMC Bioinformatics 14:321

Deeken R, Saupe S, Klinkenberg J, Riedel M, Leide J, Hedrich R, et al. (2016) The Nonspecific Lipid Transfer Protein AtLtpI-4 Is Involved in Suberin Formation of Arabidopsis thaliana Crown Galls. Plant Physiol 172:1911–1927

Dietz K-J, Vogel MO, Viehhauser A (2010) AP2/EREBP transcription factors are part of gene regulatory networks and integrate metabolic, hormonal and environmental signals in stress acclimation and retrograde signalling. Protoplasma 245:3–14

Ditta G, Stanfield S, Corbin D, Helinski DR (1980) Broad host range DNA cloning system for gram-negative bacteria: construction of a gene bank of Rhizobium meliloti. Proc Natl Acad Sci U S A 77:7347–7351

El-Kouhen K, Blangy S, Ortiz E, Gardies A-M, Ferte N, Arondel V (2005) Identification and characterization of a triacylglycerol lipase in Arabidopsis homologous to mammalian acid lipases. FEBS Lett 579:6067–6073

Ellinger D, Stingl N, Kubigsteltig II, Bals T, Juenger M, Pollmann S, et al. (2010) DONGLE and DEFECTIVE IN ANTHER DEHISCENCE1 lipases are not essential for wound- and pathogen-induced jasmonate biosynthesis: redundant lipases contribute to jasmonate formation. Plant Physiol 153:114–127

Feller A, Machemer K, Braun EL, Grotewold E (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. Plant J 66:94–116

Feng C, Andreasson E, Maslak A, Mock HP, Mattsson O, Mundy J (2004) Arabidopsis MYB68 in development and responses to environmental cues. Plant Sci 167:1099–1107

Franco-Zorrilla JM, López-Vidriero I, Carrasco JL, Godoy M, Vera P, Solano R (2014) DNA-binding specificities of plant transcription factors and their potential to define target genes. Proc Natl Acad Sci U S A 111:2367–72

Franke R, Hofer R, Briesen I, Emsermann M, Efremova N, Yephremov A, et al. (2009) The DAISY gene from Arabidopsis encodes a fatty acid elongase condensing

enzyme involved in the biosynthesis of aliphatic suberin in roots and the chalaza-micropyle region of seeds. Plant J 57:80–95

Franke R, Schreiber L (2007) Suberin-a biopolyester forming apoplastic plant interfaces. Curr Opin Plant Biol 10:252–259

Gou M, Hou G, Yang H, Zhang X, Cai Y, Kai G, et al. (2017) The MYB107 Transcription Factor Positively Regulates Suberin Biosynthesis. Plant Physiol 173:1045–1058

Graça J, Pereira H (2004) The periderm development in Quercus suber. IAWA J 25:325–335

Graça J, Pereira H (2000) Methanolysis of bark suberins: Analysis of glycerol and acid monomers. Phytochem Anal 11:45–51

Graça J, Santos S (2007) Suberin: a biopolyester of plants' skin. Macromol Biosci 7:128–35

Graça J, Santos S (2006) Glycerol-derived ester oligomers from cork suberin. Chem Phys Lipids 144:96–107

Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble W (2007) Quantifying similarity between motifs. Genome Biol 8:R24

Heldt H-W, Piechulla B (2011) Lipids are membrane constituents and function as carbon stores. In: Plant Biochemistry. Academic Press, San Diego, pp 359–398

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 Biol 8:R19

Höfer R, Briesen I, Beck M, Pinot F, Schreiber L, Franke R (2008) The Arabidopsis cytochrome P450 CYP86A1 encodes a fatty acid ω-hydroxylase involved in suberin monomer biosynthesis. J Exp Bot 59:2347–2360

Hong JC (2016) Chapter 3 - General Aspects of Plant Transcription Factor Families A2 - Gonzalez, Daniel H. BT - Plant Transcription Factors. Academic Press, Boston, pp 35–56

Joshi N, Fass J (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Joshi NA FJN (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files

Kanehisa M, Goto S (2000) Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 28:27–30

Kosma DK, Murmu J, Razeq FM, Santos P, Bourgault R, Molina I, et al. (2014) AtMYB41 activates ectopic suberin synthesis and assembly in multiple plant species and cell types. Plant J 80:216–229

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, et al. (2012) ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res 22:1813–1831

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Lashbrooke J, Cohen H, Levy-Samocha D, Tzfadia O, Panizel I, Zeisler V, et al. (2016) MYB107 and MYB9 Homologs Regulate Suberin Deposition in Angiosperms. Plant Cell 28:2097–2116

Layat E, Cotterell S, Vaillant I, Yukawa Y, Tutois S, Tourmente S (2012) Transcript levels, alternative splicing and proteolytic cleavage of TFIIIA control 5S rRNA accumulation during Arabidopsis thaliana development. Plant J 71:35–44

Le Hir R, Sorin C, Chakraborti D, Moritz T, Schaller H, Tellier F, et al. (2013) ABCG9, ABCG11 and ABCG14 ABC transporters are required for vascular development in Arabidopsis. Plant J 76:811–824

Lee S-B, Jung S-J, Go Y-S, Kim H-U, Kim J-K, Cho H-J, et al. (2009) Two Arabidopsis 3-ketoacyl CoA synthase genes, KCS20 and KCS2/DAISY, are functionally redundant in cuticular wax and root suberin biosynthesis, but differentially controlled by osmotic stress. Plant J 60:462–475

Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, et al. (2013) Acyl-Lipid Metabolism. Arabidopsis Book 11:e0161

Li C, Wang X, Lu W, Liu R, Tian Q, Sun Y, et al. (2014a) A poplar R2R3-MYB transcription factor, PtrMYB152, is involved in regulation of lignin biosynthesis during secondary cell wall formation. Plant Cell, Tissue Organ Cult 119:553–563

Li N, Xu C, Li-Beisson Y, Philippar K (2016) Fatty Acid and Lipid Transport in Plant Cells. Trends Plant Sci 21:145–158

Li Q, Brown JB, Huang H, Bickel PJ (2011) Measuring reproducibility of high-throughput experiments. Ann Appl Stat 5:1752–1779

Li R, Yu K, Hildebrand DF (2010) DGAT1, DGAT2 and PDAT expression in seeds and other tissues of epoxy and hydroxy fatty acid accumulating plants. Lipids 45:145–157

Li W, Lin Y-C, Li Q, Shi R, Lin C-Y, Chen H, et al. (2014b) A robust chromatin immunoprecipitation protocol for studying transcription factor-DNA interactions and histone modifications in wood-forming tissue. Nat Protoc 9:2180–93

Li Y, Beisson F, Koo AJK, Molina I, Pollard M, Ohlrogge J (2007) Identification of acyltransferases required for cutin biosynthesis and production of cutin with suberin-like monomers. Proc Natl Acad Sci U S A 104:18339–18344

Lin Y-C, Liu Y, Nakamura Y (2015) The Choline/Ethanolamine Kinase Family in

Arabidopsis: Essential Role of CEK4 in Phospholipid Biosynthesis and Embryo Development. Plant Cell 27:1497–511

Liu J, Osbourn A, Ma P (2015) MYB transcription factors as regulators of phenylpropanoid metabolism in plants. Mol. Plant 8:689–708

Liu Y-F, Li Q-T, Lu X, Song Q-X, Lam S-M, Zhang W-K, et al. (2014) Soybean GmMYB73 promotes lipid accumulation in transgenic plants. BMC Plant Biol 14:73

Lo M, Taylor C, Wang L, Nowack L, Wang T-W, Thompson J (2004) Characterization of an ultraviolet B-induced lipase in Arabidopsis. Plant Physiol 135:947–958

LOWRY OH, ROSEBROUGH NJ, FARR AL, RANDALL RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193:265–275

Machanick P, Bailey TL (2011) MEME-ChIP: Motif analysis of large DNA datasets. Bioinformatics 27:1696–1697

McFarlane HE, Shin JJH, Bird DA, Samuels AL (2010) Arabidopsis ABCG transporters, which are required for export of diverse cuticular lipids, dimerize in different combinations. Plant Cell 22:3066–75

Méchin V, Damerval C, Zivy M (2007) Total protein extraction with TCA-acetone. Methods Mol Biol 355:1–8

Millar a a, Kunst L (1997) Very-long-chain fatty acid biosynthesis is controlled through the expression and specificity of the condensing enzyme. Plant J 12:121–131

Mou Z, Wang X, Fu Z, Dai Y, Han C, Ouyang J, et al. (2002) Silencing of Phosphoethanolamine N-Methyltransferase Results in Temperature-Sensitive Male Sterility and Salt Hypersensitivity in Arabidopsis. Plant Cell Online 14:2031–2043

Munnik T, Meijer HJG (2001) Osmotic stress activates distinct lipid and MAPK signalling pathways in plants. FEBS Lett 498:172–178

Murashige T, Skoog F (1962) A Revised Medium for Rapid Growth and Bio Assays with Tobacco Tissue Cultures. Physiol Plant 15:473–497

Nakato R, Shirahige K (2017) Recent advances in ChIP-seq analysis: From quality management to whole-genome annotation. Brief Bioinform 18:279–290

O'Malley RC, Huang SC, Song L, Lewsey MG, Bartlett A, Nery JR, et al. (2017) Cistrome and Epicistrome Features Shape the Regulatory DNA Landscape. Cell 165:1280–1292

Panikashvili D, Aharoni A (2008) ABC-type transporters and cuticle assembly: Linking function to polarity in epidermis cells. Plant Signal Behav 3:806–809

Panikashvili D, Savaldi-Goldstein S, Mandel T, Yifhar T, Franke RB, Hofer R, et al. (2007) The Arabidopsis DESPERADO/AtWBC11 transporter is required for cutin

and wax secretion. Plant Physiol 145:1345–1360

Panikashvili D, Shi JX, Bocobza S, Franke RB, Schreiber L, Aharoni A (2010) The Arabidopsis DSO/ABCG11 transporter affects cutin metabolism in reproductive organs and suberin in roots. Mol Plant 3:563–575

Panikashvili D, Shi JX, Schreiber L, Aharoni A (2011) The Arabidopsis ABCG13 transporter is required for flower cuticle secretion and patterning of the petal epidermis. New Phytol 190:113–124

Pireyre M, Burow M (2015) Regulation of MYB and bHLH transcription factors: A glance at the protein level. Mol Plant 8:378–388

Pollard M, Beisson F, Li Y, Ohlrogge JB (2008) Building lipid barriers: biosynthesis of cutin and suberin. Trends Plant Sci 13:236–246

Quinlan AR, Hall IM (2010) BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26:841–842

Razin S V, Borunova V V, Maksimenko OG, Kantidze OL (2012) Cys2His2 zinc finger protein family: classification, functions, and major members. Biochemistry (Mosc) 77:217–226

Riechmann JL, Heard J, Martin G, Reuber L, Jiang C, Keddie J, et al. (2000) Arabidopsis transcription factors: genome-wide comparative analysis among eukaryotes. Science 290:2105–2110

Saghizadeh M, Akhmedov NB, Yamashita CK, Gribanova Y, Theendakara V, Mendoza E, et al. (2009) ZBED4, a BED-Type Zinc-Finger Protein in the Cones of the Human Retina. Invest Ophthalmol Vis Sci 50:3580–3588

Sahlin K, Vezzi F, Nystedt B, Lundeberg J, Arvestad L (2014) BESST--efficient scaffolding of large fragmented assemblies. BMC Bioinformatics 15:281

Salminen TA, Blomqvist K, Edqvist J (2016) Lipid transfer proteins: classification, nomenclature, structure, and function. Planta 244:971–997

Schenk RU, Hildebrandt AC (1972) Medium and techniques for induction and growth of monocotyledonous and dicotyledonous plant cell cultures. Can J Bot 50:199–204

Seo YS, Kim EY, Kim JH, Kim WT (2009) Enzymatic characterization of class I DAD1-like acylhydrolase members targeted to chloroplast in Arabidopsis. FEBS Lett 583:2301–2307

Serra O, Soler M, Hohn C, Franke R, Schreiber L, Prat S, et al. (2009a) Silencing of StKCS6 in potato periderm leads to reduced chain lengths of suberin and wax compounds and increased peridermal transpiration. J Exp Bot 60:697–707

Serra O, Soler M, Hohn C, Sauveplane V, Pinot F, Franke R, et al. (2009b) CYP86A33-Targeted Gene Silencing in Potato Tuber Alters Suberin Composition, Distorts

Suberin Lamellae, and Impairs the Periderm's Water Barrier Function. Plant Physiol 149:1050–1060

Shirasawa K, Takabe T, Takabe T, Kishitani S (2006) Accumulation of glycinebetaine in rice plants that overexpress choline monooxygenase from spinach and evaluation of their tolerance to abiotic stress. Ann Bot 98:565–571

Shockey JM, Fulda MS, Browse JA (2002) Arabidopsis contains nine long-chain acyl-coenzyme a synthetase genes that participate in fatty acid and glycerolipid metabolism. Plant Physiol 129:1710–1722

Simpson JT (2014) Exploring genome characteristics and sequence quality without a reference. Bioinformatics 30:1228–1235

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

Tian Q, Wang X, Li C, Lu W, Yang L, Jiang Y, et al. (2013) Functional Characterization of the Poplar R2R3-MYB Transcription Factor PtoMYB216 Involved in the Regulation of Lignin Biosynthesis during Wood Formation. PLoS One 8:76369

Vanholme R, Demedts B, Morreel K, Ralph J, Boerjan W (2010) Lignin Biosynthesis and Structure. PLANT Physiol 153:895–905

Vishwanath SJ, Delude C, Domergue F, Rowland O (2015) Suberin: biosynthesis, regulation, and polymer assembly of a protective extracellular barrier. Plant Cell Rep 34:573–586

Waditee R, Bhuiyan MNH, Rai V, Aoki K, Tanaka Y, Hibino T, et al. (2005) Genes for direct methylation of glycine provide high levels of glycinebetaine and abiotic-stress tolerance in Synechococcus and Arabidopsis. Proc Natl Acad Sci U S A 102:1318–1323

Wang G, Ryu S, Wang X (2012) Plant phospholipases: an overview. Methods Mol Biol 861:123–137

Wellesen K, Durst F, Pinot F, Benveniste I, Nettesheim K, Wisman E, et al. (2001) Functional analysis of the LACERATA gene of Arabidopsis provides evidence for different roles of fatty acid omega -hydroxylation in development. Proc Natl Acad Sci U S A 98:9694—9699

Yadav V, Molina I, Ranathunge K, Castillo IQ, Rothstein SJ, Reed JW (2014) ABCG Transporters Are Required for Suberin and Pollen Wall Extracellular Barriers in Arabidopsis. Plant Cell 26:3569–3588

Yang W, Simpson JP, Li-Beisson Y, Beisson F, Pollard M, Ohlrogge JB (2012) A Land-Plant-Specific Glycerol-3-Phosphate Acyltransferase Family in Arabidopsis: Substrate Specificity, sn-2 Preference, and Evolution. Plant Physiol 160:638–652

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, et al. (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol 60:107–124

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. (2008) Model-based analysis of ChIP-Seq (MACS). Genome Biol 9:R137

Zoldos V, Papes D, Brown SC, Panaud O, Siljak-Yakovlev S (1998) Genome size and base composition of seven *Quercus* species: inter- and intra-population variation. Genome 41:162–168

8. Supplementary Material

8.1. Supplementary Material 1

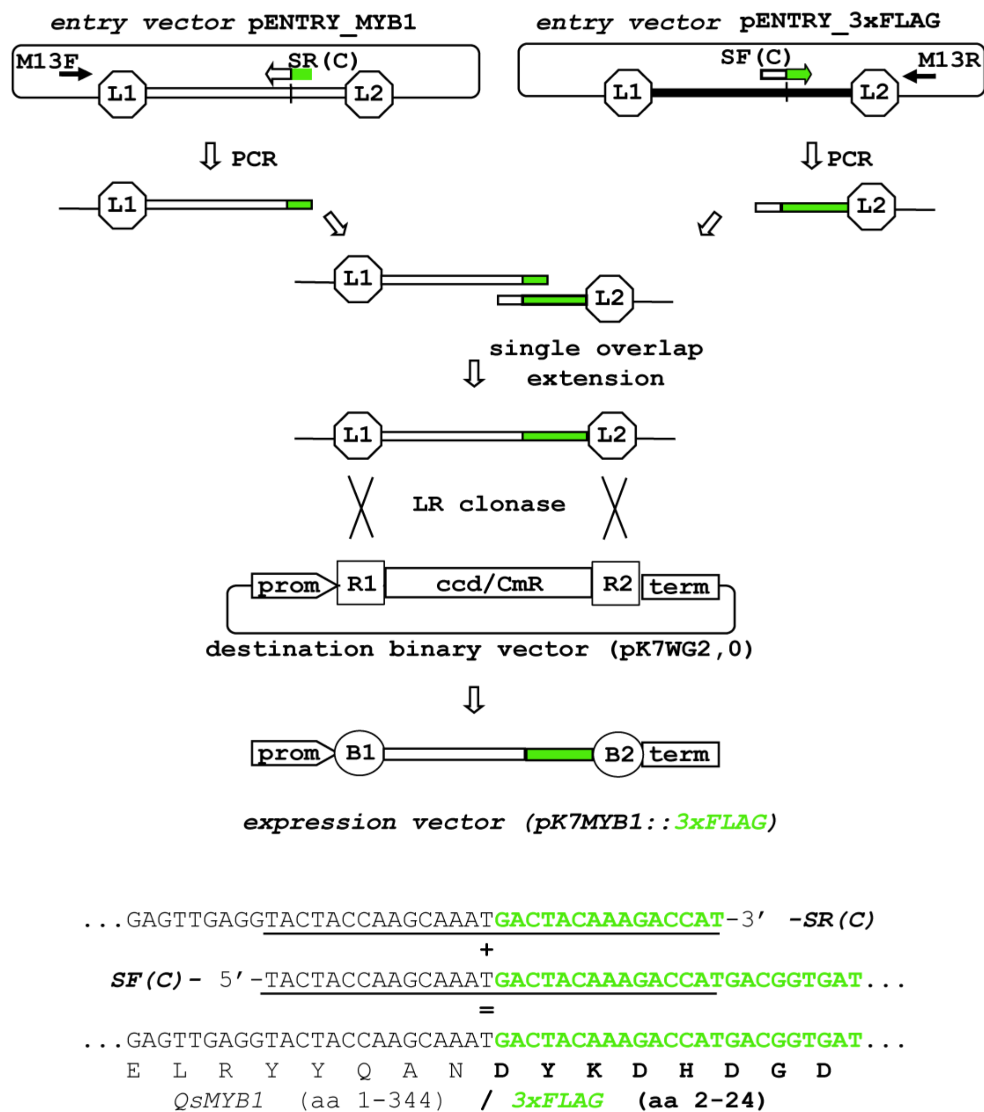


Figure IV. SM 1 | Gene construction strategy used for QsMYB::triple FLAG epitope fusion protein production.

Table IV. SM 1 | Primers used to generate the overexpression destination vector pK7MYB1::3xFLAG.

Primer	Sequence
M13F	5'-GTA AAA CGA CGG CCA GT-3'
M13R	5'-CAG GAA ACA GCT ATG AC-3'
SF(C)	5'-TAC TAC CAA GCA AAT GAC TAC AAA GAC CAT-3'
SR(C)	5'-ATG GTC TTT GTA GTC ATT TGC TTG GTA GTA-3'
3xFLAG_F_Olig	5'-AGC TTA GAC TAC AAA GAC CAT GAC GGT GAT TAT AAA GAT CAT GAC ATC GAT TAC AAG GAT GAC GAT GAC AAG TGA TAT CG -3'
3xFLAG_R_Olig	5'-GAT CCG ATA TCA CTT GTC ATC GTC ATC CTT GTA ATC GAT GTC ATG ATC TTT ATA ATC ACC GTC ATG GTC TTT GTA GTC TA -3'
MYB1ns_attB1F	5'- AAA AAG CAG GCT TAG AAG GAG ATA GAA CCA TGG GGA GAG CTC CAT GTT GTG ACA AAG -3'
MYB1ns_attB2R	5′- AGA AAG CTG GGT CAT TTG CTT GGT AGT ACC TCA ACT CT -3'
attB1 adapter	5′- GGG GAC AAG TTT GTA CAA AAA AGC AGG CT -3′
attB2 adapter	5′- GGG GAC CAC TTT GTA CAA GAA AGC TGG GT -3′

Table IV. SM 2 | Primers used to confirm the integration of the foreign DNA delivered by the destination vector plasmid.

Primer	Sequence
qPCRQsMYB1F	5'-AGC CTA AAG CAA GAG ATG AAG AGA G-3'
SR(C)	5'-ATG GTC TTT GTA GTC ATT TGC TTG GTA GTA-3'
nptIIF	5'-GAG GCT ATT CGG CTA TGA CTG-3'
nptIIR	5'-ATC GGG AGC GGC GAT ACC GTA-3'
virGF	5'-AAG GTG AGC CGT TGA AAC AC-3'
virGR	5'-ATC TCA AGC CCA TCT TCA CG-3'

8.2. Supplementary Material 2 - Generation of cork oak genome draft

8.2.1. DNA extraction and sequencing

Nuclear DNA extractions were performed from leaf material using the innuPREP Plant DNA Kit (Analytik Jena). A total of nine paired-end (PE) libraries and nine mate-pair (MP) libraries, of different insert sizes, were prepared and sequenced using the Illumina Hiseq2000 platform and a read length of 100 bp (PE) or 50 bp (MP), which yielded a total of 4,451,211,066 reads (Table IV. SM2).

8.2.2. Pre-processing and evaluation of high-throughput sequence data

Sickle (Joshi and 2011) was used to filter out the reads, using as criteria minimum quality (20) and read length (80 bp and 40 bp for the PE and MP libraries, respectively). Furthermore, any read with an ambiguous base (N) was also discarded (Table IV. SM2.1). After removal of the low quality reads a total of 83.1% (PE) and 85.3% (MP) from the initial number of reads remained in the dataset. Genome size size estimations were done with the PE reads using SGA preqc module (Simpson 2014), which extends the k-mer count based analysis by including methods to explicitly distinguish true genomic k-mers from artificial k-mers containing sequencing errors. The predicted genome size was 715.9 Mb, which represented a smaller size when compared with previous flow cytometry estimations (Zoldos et al. 1998).

8.2.3. Genome assembly and scaffolding

Paired-end reads were assembled with Ray v2.3.1 (Boisvert et al. 2010) with a k-mer size of 81. Mate-pair libraries were mapped onto the Ray assembly

with bowtie2 v2.2.7 (Langmead and Salzberg 2012) and those alignments were used to scaffold the genome using BESST v1.3 (Sahlin et al. 2014). The cork oak genome draft assembly contained 50,595 scaffolds, with a minimum size of 1 kb. The total genome length contained in the draft assembly was 655.5 Mb, and the percentage of Ns was 5.6%. The assembly N50 was 87,392 bp, while the longest scaffold contained 665,812 bp.

Table IV. SM 3 | Genome Draft Reads characteristics.

Sample Name	Number of raw reads	Insert size (bp)	Cleaned reads	% Kept
WHABPI001738-44	352,588,096	170	315,869,872	89.59%
HL8.2	321,172,452	170	290,834,920	90.55%
HL8.4	309,545,950	170	280,005,244	90.46%
WHAIPI001737-45	212,784,794	500	169,685,260	79.75%
HL8.5	334,949,460	500	276,589,646	82.58%
HL8.6	376,329,674	500	332,800,750	88.43%
HL8.7	218,435,846	800	160,517,058	73.48%
HL8.8	228,831,248	800	165,292,408	72.23%
HL8.9	175,494,582	800	128,782,240	73.38%
Total paired-end	2,530,132,102	-	2,120,377,398	83.81%
HL8.10	217,902,622	2,000	189,416,746	86.93%
HL8.14	228,392,402	2,000	196,420,348	86.00%
HL8.16	224,235,198	2,000	195,654,282	87.25%
Sob19	277,442,886	2,000	236,518,096	85.25%
Sob20	189,991,228	2,000	161,768,988	85.15%
Sob21	281,147,166	2,000	237,842,630	84.60%
Sob15	197,695,978	5,000	168,796,986	85.38%
Sob17	158,274,708	5,000	129,562,092	81.86%
Sob18	145,996,776	5,000	122,714,074	84.05%
Total mate-pair	1,921,078,964	-	1,638,694,242	85.30%

8.2.4. Assembly validation and annotation

The 50,592 scaffolds were used in a CEGMA analysis, which showed that the majority of core genes were already present in the draft (81.9% in full length and 96.4% partially represented). Augustus was used for gene prediction using *Arabidopsis thaliana* as the training model. A total of 109,847 genes were detected, of which 90,978 (82.8%) represented full length features (with a start and stop codon) and 18,869 (17.2%) were partial. Rapsearch2 annotations against NCBI-nr and NCBI plants and an Interproscan run displayed 73,630 (67.0%), 70,792 (64.5%) and 71,913 (65.5%) genes with a match, respectively.

8.2.5. References

Boisvert S, Laviolette F, Corbeil J (2010) Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies. J Comput Biol 17:1519–33

Joshi N, Fass J (2011) Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle

Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359

Sahlin K, Vezzi F, Nystedt B, Lundeberg J, Arvestad L (2014) BESST--efficient scaffolding of large fragmented assemblies. BMC Bioinformatics 15:281

Simpson JT (2014) Exploring genome characteristics and sequence quality without a reference. Bioinformatics 30:1228–1235

Zoldos V, Papes D, Brown SC, Panaud O, Siljak-Yakovlev S (1998) Genome size and base composition of seven Quercus species: inter- and intra-population variation. Genome 41:162–168

8.3. Supplementary Material 3

A

Fluorescence arbitrary units

1500 bp
1000 bp
500 bp
100 bp

1500 1000 700 600 500 400 300 200 100

DNA lenght in base pairs (bp)

Figure IV. SM 4 | Quantification of ChIP-DNA fragments distribution by agarose gel 2.5% (w/v).

B

Fluorescence arbitrary units

DNA lenght in base pairs (bp)

Figure IV. SM 5 | ChIP-DNA fragments distribution quantified by Bioanalyzer.

8.4. Supplementary Material 4

Data in file Supplementary File 4.xlsx

8.5. Supplementary Material 5

Data in file Supplementary File 5.xlsx

8.6. Supplementary Material 6

Data in file Supplementary File 6.xlsx

8.7. Supplementary Material 7

Data in file Supplementary File 7.xlsx

CHAPTER V

CONCLUSIONS AND FINAL REMARKS

CONCLUSIONS AND FINAL REMARKS

QsMYB1 is a R2R3-Myb transcription factor up-regulated in organs and tissues derived from the activity of the cork cambium, mainly in the newly formed cork (Almeida et al. 2013). Besides the potential importance of this gene to better understand the molecular mechanism of cork formation, the mechanism behind the regulatory functions of QsMYB1 was not fully understood, so further studies were needed in order to unravel the regulatory network between QsMYB1 and the cork biosynthesis pathways(s).

This thesis contributes to unravel the QsMYB1 mode of action at genome-wide level in a first attempt to identify QsMYB1 targets involved in cork biosynthesis pathways. The main results of this work are summarized in Figure V. 1.

In order to discover the putative DNA targets of QsMYB1 on a genome-wide scale, it was necessary to develop an expression system where QsMYB1 transcription factor was able to be pull-downed together with its DNA binding sites. To identify these regions, genetic modified somatic embryogenic lines overexpressing QsMYB1 fused to 3xFLAG were successfully generated (Chapter II). With this work, stable and characterized cork oak embryos were produced. The use of somatic embryogenesis is an efficient approach for the purpose and is proved with this work that the system may be used as model to perform functional and biochemistry studies in a non-model plant species with a long-life cycle. The embryos and original stable cell lines produced in this study were successfully cryopreserved providing important material for further investigation, namely as a functional tool to other research groups. Since the genomic resources of cork oak are increasing quickly, including the knowledge of the genome sequence, this cell lines represent a starting point for a broad of research projects about the cork oak genes functions.

Although somatic embryos were used to study QsMYB1-DNA interactions, the availability of cork somatic embryos in different development phases led to design a transcriptomic study during somatic embryo development (Chapter III). The identification of a group of genes with relevance in various aspects of embryogenesis and embryo development, as transcription regulators, hormone-

related genes, genes involved in germination and root development, constitute important knowledge to fill a gap in the field, the lack of information of the cork oak specific genes involved in embryogenesis. These results provide knowledge not only for the biological investigation of the species but also constitute a very useful information for future breeding and reforestation programs.

A chromatin immunoprecipitation protocol was then optimized for cork oak somatic embryos, in order to perform a ChIP-Seq, which revealed the putative QsMYB1 DNA binding sites (Chapter IV). The optimization of ChIP protocols is recursively demanded and time spending. Probably by this reason, did not exist until date a definitive protocol for ChIP in *Quercus suber.* This thesis provides for the first a time, an efficient straight-performing ChIP-Seq protocol for cork oak embryos which will enable to perform important chromatin based studies, as epigenetics, in the species.

The ChIP-Seq data analysis evidences that QsMYB1 has multiple putative targets over the cork oak genome, binding preferentially to [CWHCAA], [CYTCBTC] and [BKTGG] DNA sequence, which constitutes the three principal characteristic motifs of the QsMYB1-DNA binding. The DNA binding activity of QsMYB1 occurs in diverse regions of the genome modulating gene expression by binding to gene promotor regions or other distal regions as enhancers or silencers regions, between others. The putative binding occurs preferentially on exons (26%), introns (19%), intergenic regions (18%), and promotors (13%), besides 5' and 3' untranslated regions or terminator regions (Figure V. 1). However, is proper to consider that results reflect the occurrence of fortuitous binding sites in the genome by an overexpressing system and biologically not all the DNA binding sites are relevant. Nevertheless, the results of the present thesis constitute an important information about the mode of action of QsMYB1 evidencing it as a master regulator with putative affinity for multi-binding regions across the cork oak genome.

Taking into account these considerations the generated bioinformatic data was extensively analysed and annotated in order to retrieve the putative role of QsMYB1 on: (1) metabolic pathways; (2) processes involved in suberin and lignin biosynthesis, transport and assembly; and (3) gene expression regulation (Figure V. 1). Based on the *Quercus suber* genome draft annotation and by pathway analysis, putative genes coding for key enzymes in phenylpropanoid pathway,

suberin biosynthesis, and fatty acid and lipid metabolism were identified as susceptible of regulation by QsMYB1 (Chapter IV) (Figure V. 1). Furthermore, QsMYB1 is a putative regulator of lipid transmembrane transporters genes as *ABCGs* and *LTPs* which are involved in suberin transport and assembly (Figure V. 1). Finally, this work revealed that QsMYB1 may target other transcriptional regulators as CCHC ZF, C2H2-type ZF and BED-type ZF proteins as well as WD40-like, AP2/ERF and bHLH TFs, revealing a complex regulatory mechanism of genes transcription.

As final conclusion, the results evidence QsMYB1 as a potential master regulator of cork formation and differentiation. However further studies are required to confirm and reveal the specific role of the reported genes and to clarify the exact function of QsMYB1 in these cellular and metabolic processes. Even so, the generated data represent important information about the pathways involved in the differentiation of this unique tissue, and is a starting point to an in-depth understanding of suberin biosynthesis and deposition in cork oak. By continuing to unravel the regulation of biosynthetic pathways involved in cork formation, initiated with this thesis, important knowledge will be gathered to support the development of breeding programs based e.g, on the design of cork metabolic engineering strategies.

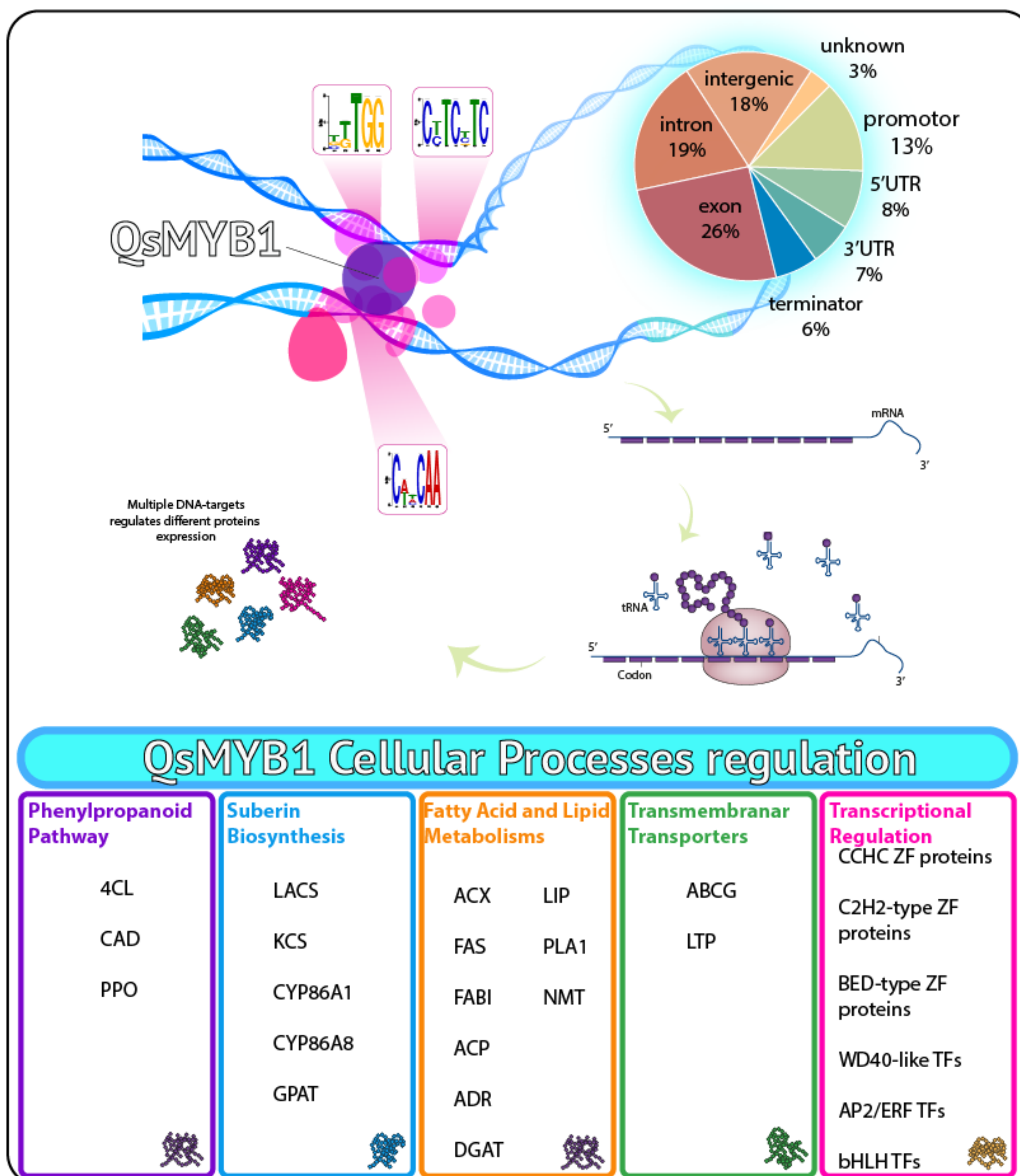


Figure V. 1 | Schematic representation of the QsMYB1 DNA-binding and the putative cellular processes regulated. *4CL*, 4-Coumarate:CoA ligase; *CAD*, Cinnamyl alcohol dehydrogenase; *PPO*, class III plant peroxidase; *LACS*, Long-chain acyl-CoA synthetase; *KCS*, β-ketoacyl-CoA synthase; *CYP86A1*, Cytochrome P450 enzymes subfamily 86A1; *CYP86A1*, Cytochrome P450 enzymes subfamily 86A8; *GPAT*, Glycerol 3-phosphate acyltransferase; *ACX*, Acyl-CoA oxidase; *FA*, Fatty acid synthase; *FABI*, enoyl-[acyl-carrier-protein] reductase (NADH); *ACP*, Acyl-carrier protein; *ADR*, NADPH-dependent aldehyde reductase; *DGAT*, diacylglycerol *O*-acyltransferase; *LIP*, triacylglycerol acylhydrolase; *PLA1*, phospholipase A1; *NMT*, phosphoethanolamine N-methyltransferase; *ABCG*, Plasma membrane-localized ATP-binding cassette transporter of the G-subfamily; *LTP*, Lipid transfer protein; *ZF*, zinc finger; *bHLH*, basic helix-turn-helix; *UTR*, Untranslated region, *mRNA*, messeger RNA; *tRNA*, transfer RNA.

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. J Plant Physiol 170:172–178

CHAPTER VI

ACKNOWLEDGEMENTS

ACKNOWLEDGEMENTS

Porque esta Tese de Doutoramento também é o trabalho de outras pessoas, não posso deixar de agradecer a todos os que fizeram parte dela e a tornaram concretizável.

Em primeiro lugar agradeço às minhas grandes orientadoras, Professora Maria Leonor Mota Morais Cecílio e Doutora Sónia Cláudia Morgado Gonçalves; por toda a dedicação, empenho, paciência, sabedoria e principalmente por toda a amizade ao longo destes anos.

Ao Professor Ricardo Ordás, que disponibilizou sempre o seu laboratório e os seus conhecimentos, mesmo a altas horas da noite! Sem ele este trabalho não seria um êxito.

A todos os colegas e amigos do CEBAL, em especial à Tânia e à Ana que desde os primeiros dias me acompanharam neste desafio.

Aos colegas do ISA, Vera, Teresa, Miguel, Edna e Alex, pelos belos momentos de discussões e argumentações científicas.

A todos os meus amigos que sempre se interessaram por este trabalho, apesar da minha dificuldade em por vezes o explicar.

Aos meus pais, por tudo o que me transmitiram nesta Vida. Como eu sou vossa obra, sem dúvida que este trabalho é de alguma forma vosso.

Aos meus avós, tios e primos. Um muito obrigado por estarem sempre presentes!

Aos meus sogros, duas pessoas com um coração de ouro. Um muito obrigado por toda a ajuda e paciência. Sim, já acabei!

À Ana, que sempre me acompanhou nesta jornada, me apoiou, incentivou, me ergueu nos momentos mais complicados e acima de tudo comemorou junto a mim todas as vitórias. Obrigado minha lindona! Fhfhfhfhhhh!!!

Por fim agradeço à Fundação para a Ciência e Tecnologia, pela bolsa de Doutoramento (SFRH/BD/69785/2010) que me foi atribuída e me permitiu desenvolver este trabalho.