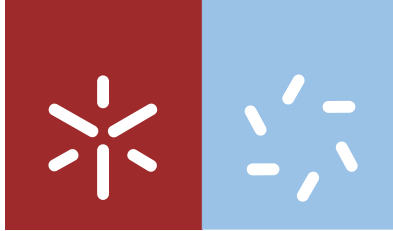


Universidade do Minho
Escola de Ciências

Alexandre Miguel Papadopoulos Magalhães

**RNA-seq analysis of the *Quercus suber*
root response to drought**



Universidade do Minho

Escola de Ciências

Alexandre Miguel Papadopoulos Magalhães

RNA-seq analysis of the *Quercus suber* root response to drought

Msc thesis in
Plant Molecular Biology, Biotechnology and
Bioentrepreneurship

Work supervised by
Doctor Rui Tavares
Doctor Herlânder Azevedo
Doctor Pedro Humberto Castro

April 2015

Declaração

Nome: Alexandre Miguel Papadopoulos Magalhães

Endereço electrónico: xandinho2468@gmail.com

Telefone: +351 917001036

Número do Cartão de Cidadão: 13739531

Título tese de Mestrado:

RNAseq analysis of the Quercus suber root response to drought

Orientadores:

Doutor Rui Manuel Peixoto Tavares

(Departamento de Biologia, Escola de Ciências, Universidade do Minho, Portugal)

Doutor Herlânder Anselmo Queirós Pereira Azevedo

(CIBIO, InBIO - Research Network in Biodiversity and Evolutionary Biology, Universidade do Porto, Portugal)

Doutor Pedro Humberto Castro

(Departamento de Biologia, Escola de Ciências, Universidade do Minho, Portugal)

Ano de conclusão: 2015

Designação do Mestrado:

Mestrado em Biologia Molecular, Biotecnologia e Bioempreendedorismo em Plantas

É AUTORIZADA A REPRODUÇÃO INTEGRAL DESTA TESE APENAS PARA EFEITOS DE INVESTIGAÇÃO, MEDIANTE DECLARAÇÃO ESCRITA DO INTERESSADO, QUE A TAL SE COMPROMETE.

Universidade do Minho, ____/____/____

Assinatura: _____

“When you eliminate the impossible, whatever remains, however improbable, must be the truth”

Leonard Nimoy – Star Trek VI: The Undiscovered Country, 1991

“Research is what I’m doing when I don’t know what I’m doing”

Wernher von
Aos Doutores Rui Tavares,
Herlander Azevedo e
Pedro Humberto Castro

Acknowledgments

It is finally over!!! I could not emphasize this more. I learned a lot, met some new people, expanded my horizon, and thanks to a few I am now a better student and researcher than last year. Now it is time for thanking those people.

First and foremost my supervisors. To Professor Rui Tavares for helping me a lot throughout the thesis and for always being straight and honest with me. To Herlander Azevedo for giving me the opportunity of entering a field, that despite being straight up to difficult and horrendous to some, has me fascinated and fully committed to the future. Your patience and dedication helped me evolve and master this field of bioinformatics. I cannot thank you enough for this. And to Humberto that with a strict supervision taught precision and carefulness is essential for excellence in the lab. To Teresa Lino Neto for being always present throughout the project and for having always an optimistic view even when things do not go as planned. Finally to Manuela costa even if not being my supervisor was always helpful and provided excellent advice when needed.

Many thanks to all my colleagues at the lab for supporting me. To Joana, Helena, João, Telma, Francisca for the help and company. Special note has to be made to Romulo for always having the time to advice and answer even the smallest of doubts, and Sara Freitas for the laughs, those little breaks for coffee to help with the morning and for the little tutoring always helpful. To Nuno and Miguel my war buddies many thanks for the friendship and company... we entered together and we conquered. To all above I am certain that you will have great success in your scientific career and even greater accomplishments at the personal level. To you Tiago our time was little but I hope you have a great future.

For all the company, an especial thanks Luis, Vasco and Vera. Even not being members of the lab you were always present, especially during the breaks and the necessary coffee, helping to take my mind off science.

Must acknowledge all my colleagues from the first year of the masters. A special thanks from the heart to Helena and Raquel for this last two years. You were always my company throughout the first year and the thesis. We enduring great difficulties, work and disappointments is always better when in the company of great friends. We shared allot, we laugh, we fought and sometimes you cried but in the end the friendship was always

present. To you Angela also an especial thank you, you were not with us throughout the thesis but when present your smile gave us joy to continue.

To all my friends thank you for the support. Special remark to those that accompanied me. To Luis you were always there from the beginning and you will always be like a brother to me. To Sandra for those talks and support when needed. To Pedro Calheiros for everything you have done for me. And at last a great appreciation to Viriato, Avelino and Manel (Jorge) for the latte drinks and friendship. You all help me endure those last academic years and for that I will always be grateful. I hope you can share more moments together and you will be always present in my life.

All of this was not possible without my parents and grandparents. You paid for my education, you supported me, and you carried me until the end. I will always be grateful and hope I can be all that you want me to be.

Last but not least to Caroline. You were always my company, always the smile that brought me back up. Thank you for always being there and always putting up with me. You will always be in my heart.

Abstract

Cork oak (*Quercus suber* L.) is a typical species from the Portuguese Mediterranean forest, usually growing in dry environments. Mainly due to its high economic value, cork oak has been considered a protected plant species and a national species of interest. Factors resulting from climate changes, e.g. increased temperature, high light intensity, drought and air pollution are directly affecting the development of plants and reducing their productivity. Next generation sequencing, in particular RNA-seq, has proven to be a powerful tool with a diverse range of applications, and provides fundamental insight into the structure and regulatory aspects of gene networks, particularly when organisms without available OMICs resources, such as a sequence genome, are studied.

To address the cork oak's response and regulation to water availability, differential transcriptomic analysis was performed in roots of young plantlets subjected to moderate and high drought conditions, comparing to well-irrigated plants. Photosynthetic activity measured by PAM fluorometry and photosynthetic pigment content were used as physiological indicators of plant fitness. cDNA was obtained from previously drought-stressed roots RNA, and sequencing was carried out by Roche's 454 technology resulting in three libraries, of a 1.8 million reads in total, later assembled into 21012 unigenes, that would be used for mapping steps, in order to identify differential expression. This resulted in the identification of 353 up-regulated and 193 down-regulated genes. Primary *in silico* analysis identified a significant number of effector genes traditionally associated with drought responses, such as dehydrins, LEA proteins, cell wall remodeling, transcription factors and ubiquitinating-associated genes, suggesting a tight control of drought responses at both the transcriptional and protein turnover levels.

Further *in silico* analysis was accomplished by annotating the genes against the genome of the model plant *Arabidopsis thaliana*, allowing for ortholog identification. This allowed us to establish gene networks based on functional characteristics, such as gene co-expression, protein-protein interactions and co-localization. Coupled with cis-element enrichment and transcription factor identification, this analysis enabled the establishment of functional and regulatory relationships between differentially expressed genes. Ultimately, we demonstrated the induction, in drought stressed *Quercus suber* roots, of a complete, ABA-dependent signaling cascade. Elements of this signaling

cascade range from ABA-sensing components to transcription factors and then to effector genes involved in the drought response.

Resumo

O sobreiro (*Quercus suber*) é uma espécie típica da floresta Mediterrânea Portuguesa e cresce usualmente em ambientes secos. Devido ao seu elevado valor económico, o sobreiro é considerado uma espécie vegetal protegida e de interesse nacional. Fatores decorrentes das alterações climáticas como, temperatura e intensidade luminosa elevadas, seca e bem como poluentes afetam o desenvolvimento vegetal, reduzindo a produtividade de cultivares de interesse. A sequenciação de nova geração, em particular RNA-seq, tem se destacando como uma poderosa ferramenta de aplicações diversas, providenciando conhecimentos fundamentais sobre aspetos estruturais e reguladores de redes genéticas, particularmente em estudos em organismos sem recursos das OMICas, como um genoma sequenciado.

Para estudar a resposta do sobreiro a condições que comprometem a disponibilidade de água, foi efetuada análise diferencial dos transcriptomas de raízes de plantas jovens sujeitas a condições de secura moderada e severa, e os resultados foram comparados com os obtidos a partir de raízes de plantas bem irrigadas. A actividade fotossintética foi medida por fluimetria PAM e o conteúdo de prolina e pigmentos fotossintéticos foram considerados como indicadores fisiológicos do *fitness* das plantas. cDNA, obtido a partir de RNA extraído destas raízes, foi sequenciado por tecnologia 454 da Roche, resultando em 3 bibliotecas de 1.8 milhões de *reads* total, mais tarde *assembladas* em 21012 unigenes para serem utilizadas nos passos de mapeamento utilizado em expressão diferencial. Esta análise resultou em 353 genes positivamente induzidos e 193 genes negativamente induzidos. Análise primária *in silico* identificou um número significativo de genes efetores tradicionalmente associados à resposta à secura, como desidrinas, proteínas LEA, genes de remodelação da parede celular, fatores de transcrição e genes associados a ubiquitina, sugerindo um controlo apertado da resposta à secura dos níveis transcricionais, bem como de abundância proteica.

Nova análise *in silico* foi realizada através da identificação de genes ortólogos, anotando-se os genes de sobreiro contra o genoma de *Arabidopsis thaliana*. Este estudo possibilitou o estabelecimento de redes genéticas baseadas em características funcionais como co-expressão, interação proteína-proteína e co-localização. Conjuntamente com análise de enriquecimento de elementos *cis* e identificação de fatores de transcrição, foi possível estabelecer relações funcionais e de regulação entre os genes. Os resultados permitem sugerir que ocorre indução de uma cascata de sinalização dependente do ABA

em raízes de *Q. suber* em condições de secura, onde estão envolvidos sensores do ABA, fatores de transcrição e genes efectores envolvidos na resposta à secura.

Table of Contents

Acknowledgments	v
Abstract	vii
Resumo	ix
Table of contents	xi
Abbreviations and symbols	xiii
List of figures	xiv
List of tables	xv
Chapter 1 – Introduction	1
1.1 <i>Quercus suber</i> (cork oak) biology	3
1.2 <i>Quercus suber</i> (cork oak) biology	3
1.3 Molecular tools in <i>Quercus suber</i> research	7
1.3.1 DNA extraction	8
1.3.2 RNA extraction	8
1.3.3 Real-Time Quantitative PCR (qPCR)	9
1.3.4 Molecular markers and population genomics	9
1.3.5 OMICS and NGS	10
1.4 Next-generation sequencing in non-model plants	11
1.5 Objectives and outline of the thesis	14
Chapter 2 – Materials and Methods	15
2.1 Plant material and physiological assessments	17
2.2 454 sequencing	17
2.3 Sequence assembly and differential expression analysis	17
2.4 Protein prediction, functional annotation and ortholog assessment	19
2.5 Functional data mining in <i>Arabidopsis</i>	19
2.5.1 Gene Network analysis	19
2.5.2 Transcription factor identification	19
2.5.3 Cis-element enrichment	19
2.5.4 Phylogenetic analysis	20
2.6 Quantitative Real-Time PCR	20
Chapter 3 – Results and Discussion	23
3.1. Physiological response to long-term drought	25
3.2. Assembly of the <i>Quercus suber</i> root transcriptome	26

3.3. Differential expression analysis of the drought root transcriptome	30
3.4. Annotation of differentially expressed genes	32
3.5. Functional data mining of DEGs	36
3.6. Identification of an ABA-dependent signaling network	46
Chapter 4 – Concluding remarks and future perspectives	53
Chapter 5 – Bibliography	57
Appendixes	67
Appendix I	

Abbreviations and Symbols

°C	degrees Celsius	NGS	next-generation sequencing
aa	amino acids	ORF	open reading frame
ABA	abscisic acid	PAM	pulse amplitude modulated
BLAST	basic local alignment search tool	PCR	polymerase chain reaction
bp	base pair	qPCR	real-time quantitative PCR
cDNA	complementary DNA	RNA	ribonucleic acid
CDS	coding sequence	Rpm	rotations per minute
DEG	differentially expressed gene	ROS	reactive oxygen species
DNA	deoxyribonucleic acid	s	second
EDTA	ethylenediaminetetraacetic acid	SA	salicylic acid
ERF	ethylene response factor	SD	standard deviation
EST	expressed sequence tag	SEM	standard error of the means
g	gram	SOTA	self-organizing tree algorithm
<i>g</i>	relative centrifuge force	TAIR	The Arabidopsis information
GO	gene ontology	resource	
h	hour	<i>Taq</i>	<i>Thermus aquaticus</i> polymerase
HQ	high quality	TF	transcription factor
L	liter	Tm	melting temperature
M	molar	Ub	ubiquitin
m	meter	v/v	volume per volume
MAPK	Mitogen-activated protein kinases	w/v	weight per volume
min	minute		
mol	mole		

List of Figures

Figure 1.1 Adult cork oak specimen in a montado.	4
Figure 1.2 The cork oak natural distribution area.	5
Figure 1.3 <i>Q. suber</i> cork extraction.	6
Figure 1.4 Cork oak industry products.	6
Figure 1.5 Overview of the workflow for analysis of RNA-Seq data.	12
Figure 3.1 Characterization of morphological and physiological traits of <i>Q. suber</i> plantlets subjected to long-term drought stress.	26
Figure 3.2 Outline of the RNA-Seq strategy employed in the identification of the <i>Q. suber</i> differential root transcriptome in response to long-term drought.	27
Figure 3.3 Characterization of sequence data regarding processing and assembly of reads.	29
Figure 3.4 Venn diagram representation of the number of homology hits ($e < 10^{-6}$) between <i>Q. suber</i> unigenes and Arabidopsis thaliana TAIR 10 gene models, <i>Q. rubra</i> putative transcripts and the <i>Castanea mollissima</i> predicted transcriptome.	29
Figure 3.5 Characterization of differential gene expression in drought-stressed Quercus suber roots.	30
Figure 3.6 Characterization of differential gene expression in D100/90 vs D50, D100/90 vs D25/10 and D50 vs D25/10 transcriptional comparisons.	31
Figure 3.7 Annotation of differentially expressed genes by BLAST homology search.	33
Figure 3.8 Gene network analysis of Arabidopsis DEG orthologs regarding functional features.	37
Figure 3.9 Additional gene network analysis, performed in the Genemania App of Cytoscape, of Arabidopsis DEG orthologs.	40
Figure 3.10 Phylogenetic reconstruction of NAC TF class, containing <i>Q. suber</i> differentially expressed unigenes, and the complete annotated set of Arabidopsis thaliana NAC gene family members.	43
Figure 3.11 Phylogenetic reconstruction of AP2-ERF TF class, containing <i>Q. suber</i> differentially expressed unigenes, and the complete annotated set of Arabidopsis thaliana AP2-ERF superfamily members.	44
Figure 3.12 Phylogenetic reconstruction of MYB TF class, containing <i>Q. suber</i> differentially expressed unigenes, and the complete annotated set of Arabidopsis thaliana MYB gene family members.	45
Figure 3.13 Phylogenetic reconstruction of WRKY TF class, containing <i>Q. suber</i> differentially expressed unigenes, and the complete annotated set of Arabidopsis thaliana WRKY gene family members.	46
Figure 3.14 Model for the regulation of long-term drought responses in <i>Q. suber</i> roots: ABF-dependent signalling pathway components are up-regulated in <i>Q. suber</i> roots in response to drought.	48
Figure 3.15 Phylogenetic reconstruction of PP2C identified in the <i>Q. suber</i> differentially expressed unigenes and the complete annotated set of Arabidopsis thaliana genes from the same class.	50

Figure 3.16 Real-time RT-PCR analysis of selected *Q. suber* drought stress responsive genes.

52

List of Tables

Table 2.1 List of quantitative qPCR primers.	20
Table 3.1 Statistics of the sequencing and assembly steps.	28
Table 3.2 Selection and categorization of significant <i>Quercus suber</i> unigenes differentially expressed in response to drought.	34
Table 3.3 Cis-elements over-represented in the promoter region of DEG orthologs.	38
Table 3.4 List of transcription factors (TFs) present within DEG orthologs.	42

Chapter 1

Introduction

CONTENTS

1.1 Plant response to abiotic stress	3
1.2 <i>Quercus suber</i> (cork oak) biology	3
1.3 Molecular tools in <i>Quercus suber</i> research	7
1.3.1 DNA extraction	8
1.3.2 RNA extraction	8
1.3.3 Real-Time Quantitative PCR (qPCR)	9
1.3.4 Molecular markers and population genomics	9
1.3.5 OMICs and NGS	9
1.4 Next-generation sequencing in non-model plants	11
1.5 Objectives and outline of the thesis	14

1.1. Plant response to abiotic stress

Living organisms are continuously affected by a variety of environmental factors. Plants, in particular, have evolved a variety of advance response mechanisms to cope with the challenges facing a sessile organism. Those adverse environmental factors limit survival, biomass production, development and yield. Abiotic stress, that can consist of drought, salinity, extreme temperatures, heavy metal oxidative stress, high UV irradiation, or often a combination of these stresses, is a particularly significant threat to crop productivity and is the major cause for worldwide yield losses in crop plants (Wang et al., 2003; Ahmad and Prasad, 2011; Qin et al., 2011; Koyro et al., 2012). To cope with these daily and seasonal environmental changes, plants response mechanisms often involve gene regulation, engaging signal transduction to trigger changes in metabolic processes, management of resources, and organ morphology (Wang et al., 2003; Danquah et al., 2014). Currently, many aspects of the regulatory pathways involved in these responses have been described. They include environmental sensing mechanisms, membrane-localized elements, signalling transduction components such as MAP kinase cascades, hormone-dependent signalling modules, and induction of several classes of transcription factors, such as HSF, AP2/ERF, NAC, MYB, ABF, DREB/CBF (Wang et al., 2003; Wang et al., 2004; Yoshida et al., 2010; Lata and Prasad, 2011; Atkinson and Urwin, 2012; Chen et al., 2012; Mizoi et al., 2012; Nakashima et al., 2012; Osakabe et al., 2013; Danquah et al., 2014; Nakashima et al., 2014). These regulatory events manage the expression of functional downstream response genes, involved in growth regulation, water management, ROS scavenging and induction secondary metabolism (Wang et al., 2003; Atkinson and Urwin, 2012). Highlighting the level of complexity associated with abiotic stress signalling, many novel regulatory mechanisms have been uncovered in recent years, including microRNA regulation (Lu and Huang, 2008; Khraiweh et al., 2012), chromatin remodelling (Golldack et al., 2011; Luo et al., 2012) and examples of post-translational modification of proteins (Castro et al., 2012). Uncovering of new regulatory mechanisms may, one day, ensure plant productivity for future generations facing ever-changing climate conditions.

1.2. *Quercus suber* (cork oak) biology

Cork oak (*Quercus suber*) is an evergreen species of the Fagaceae family, being one of the most significant forest species in the Mediterranean region. Their ecological dominance and economic value define this species as one of the main oak species in the western part of the Mediterranean Basin (Nixon, 1993; Belahbib et al., 2001). *Q. suber* (Figure 1.1) is a slow

Chapter 1. INTRODUCTION

growing, extremely long-lived tree, reaching a height of up to 20 m, with massive branches forming a round crown (Faria et al., 1996). It is a diploid ($2n=24$), monoecious species, with a protandrous to ensure cross-pollination. Natural propagation in wild populations occurs by natural regeneration, through seed (acorn) dispersal and subsequent germination. Cork oak natural regeneration is mostly assured by wind and animals, which is consistent with the strategy of most evergreen species (Magri et al., 2007).



Figure 1.1. Adult cork oak specimen in a *montado*. (www.panoramio.com)

Cork oak are distributed around overlapping habitats. In the Western Mediterranean Basin, *Q. ilex* and *Q. suber* are the dominant broadleaved oak species. *Q. suber* has quite a narrow geographical range when compared to that of other main Mediterranean oak species, mainly due to its ecological restrictions (Toumi and Lumaret, 1998; Toumi and Lumaret, 2001). The modern distribution of cork oak is rather discontinuous, ranging from the Atlantic coasts of North Africa and Iberian Peninsula to the south-eastern regions of Italy, Sicily and Sardinia, as well as the coastal belts of Algeria and Tunisia, France and Spain (Figure 1.2) (Lumaret et al., 2005; Magri et al., 2007).

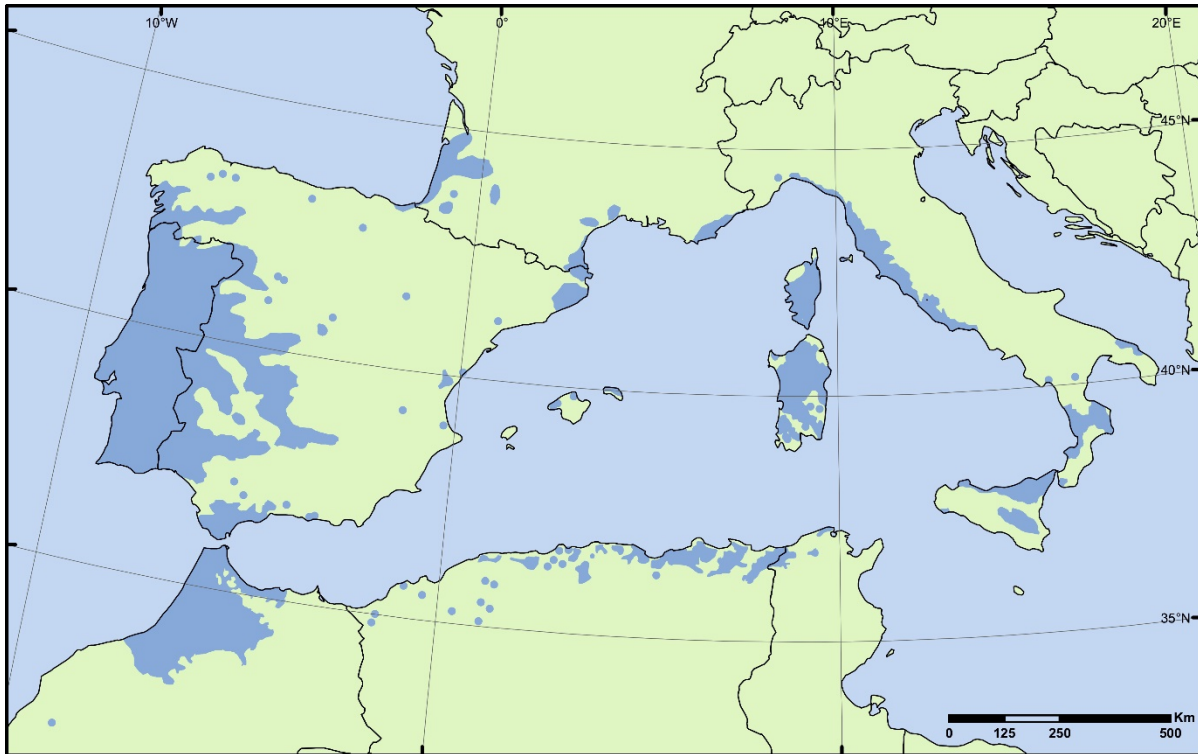


Figure 1.2. The cork oak natural distribution area. Adapted from EUFORGEN2009. (www.euforgen.org)

As a species that is endemic to the Western Mediterranean region, *Q. suber* is mostly present in semi-natural stands known as *Montados*, which are open woods with a delicate and particular ecosystem, created and maintained by man (Sánchez-González et al., 2008). *Montados* and Cork oak management represent an important economical resource, as they are associated not only with the harvesting of acorns, but also with the use of bark as the source of cork (Lopes et al., 2001). Trees are stripped of cork from the trunk (lower portion) (Figure 1.3), when they are approximately 18-25 years of age, and subsequently every 9-12 years (first and second strips are usually discarded for quality reasons). This process can be sustained for 100 to 500 years without any evident consequence on the tree's physiology (Lopes et al., 2001; Oliveira and Costa, 2012). Presently, *Q. suber* forests cover 2.2 million hectares worldwide, providing 340,000 tons/year of cork. The largest stands, covering about 737,000 ha, are in Portugal, and correspond to 23% of the forest area in Portugal and to 34% of the world's cork producing area. Currently, the cork industry represents 3% of all Portuguese exports. Cork stoppers (Figure 1.4, A) used in the wine industry are the most representative product of this industry, responsible for 70% of exports (ICNF, 2013).

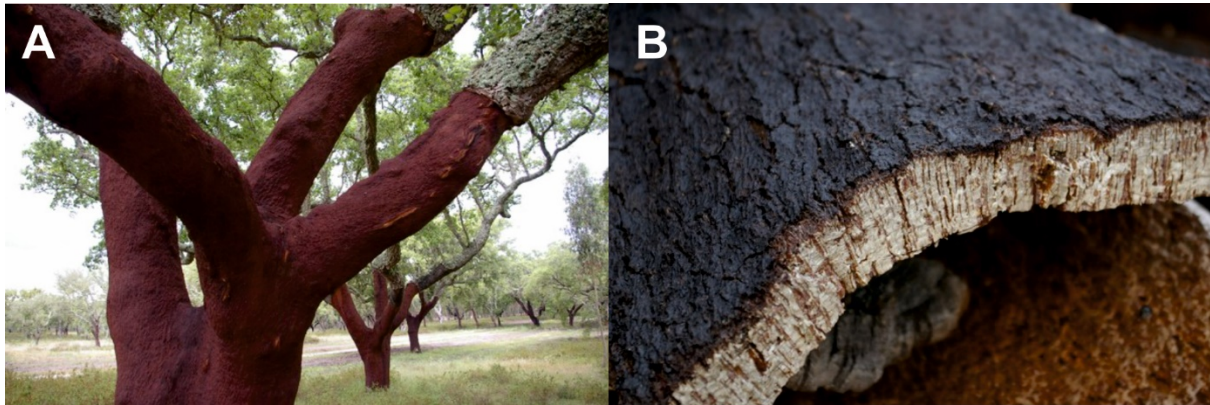


Figure 1.3. *Q. suber* cork extraction.

(A) Adult tree main trunks after cork extraction. (www.wineanorak.com)

(B) Close-up of a high-quality piece of bark used in wine stoppers. (www.wineanorak.com)

Other applications look at the remarkable qualities of cork as a light-weight material, compressible, elastic and flexible. It is also impermeable to most liquid and gaseous substances and an extremely good insulator of heat, sound, shock and electricity, making it a target for industrial utilization, with applications ranging from natural materials used in everyday construction (Figure 1.4, B) to high performance insulation used in the aerospace industry including the Space Shuttle, Delta and Atlas rockets (www.corkfacts.com).



Figure 1.4. Cork oak industry products.

(A) Cork stoppers for wine bottles. (www.corkfacts.com)

(B) Cork tiles used for flooring. (www.corkfacts.com)

The Cork oak's economic and environmental importance has to be taken into consideration as the unparalleled decline in cork oak plantations that has been occurring in the Iberian Peninsula and in Morocco is threatening the entire ecosystem (Coelho et al., 2006). Although the marginal and natural populations of cork oak are possibly the most endangered, Iberian Montados are also currently threatened and in decline due to multiple factors. The main factor contributing to this decline is the occurrence of very severe drought periods over several consecutive years (Toumi and Lumaret, 1998). As opposed to other oaks which have a

great ecological amplitude, this species is a sclerophyllous tree that is adapted to a 4-month-long hot-dry summer period (with at least 450 mm mean annual rainfall), maximum temperatures approaching 35 to 40°C, mid-day irradiance exceeding 2,000 $\mu\text{mol m}^{-2} \text{s}^{-1}$ photosynthetically active radiation (PAR) (Faria et al., 1996; Toumi and Lumaret, 1998; Toumi and Lumaret, 2001; Catry et al., 2012). The Mediterranean-climate regions are characterized by a cycle of temperatures out of phase with the rainfall, producing mild to cool rainy winters and dry summers. The hydrological variability of Mediterranean climate regions is due to a combination of rainfall (irregularly distributed in time and space) and heterogeneous land topography. The existing variety of climatic conditions in these areas is due to the specific configuration of land surface and distance from sea. The climate of Mediterranean areas induces a temporal variability in the soil moisture dynamics that is strongly affected by seasonal fluctuations between humid and dry conditions (Toumi and Lumaret, 1998; Toumi and Lumaret, 2001; Sofo et al., 2008; Catry et al., 2012).

Mediterranean vegetation dealing with this peculiar soil moisture dynamics has developed a number of physiological mechanisms to tolerate drought stress and growth under adverse climatic conditions, which include early responses involving stomatal closure, to prevent or delay tissue dehydration, and antioxidant biosynthesis as a photoprotection mechanism (Flexas et al., 2014). Other long term acclimation responses include decreased growth, to reduce water and nutrient demands (Nuche et al., 2014), changes in allocation of resources, from support tissues to assimilating organs (Nuche et al., 2014) and development of strategies to prevent xylem cavitation, like refilling mechanisms, regulation of hydraulic conductance and xylem margin reinforcement and repair (Nardini et al., 2014). Moreover, studies at population level suggest that Mediterranean species have overall phenotypic plasticity to global changes, and many species like *Q. suber* have significant differences in adaptive evolutionary potential, with important implications for species distribution (Matesanz and Valladares, 2014).

1.3. Molecular tools in *Quercus suber* research

As environmental changes in the Mediterranean basin keep casting a shadow of doubt on the resilience of this species, and as numbers of natural population of cork oak keep declining, more and more efforts are driven toward the understanding of the biology and physiology of such an important economic species. Over the last years, several tools have been developed to study several aspects of the species at the genetic and molecular levels, such as gene/protein characterization, population genomics and dEST transcriptomics, aiding,

as well, the development of OMICs-based approaches that are essential for characterisation of cork oak at a systems level.

1.3.1. DNA extraction

Nucleic acid extraction has been accomplished with a number of approaches ranging from commercial kits to more complex strategies. DNA extraction has been amply extracted from different tissues, mainly following the protocol developed by Doyle (1990) with uses Cetyl Trimethyl Ammonium Bromide (CTAB) and is optimized for fresh tissue. A number of authors have adopted this method for extraction (Soto et al., 2003; Coelho et al., 2006; Soto et al., 2007; Rocheta et al., 2012; Almeida et al., 2013), reporting high yields. Kit-based approaches have relied mostly on Qiagen's Plant DNeasy Kit, reported to successfully extract genomic DNA from a number of different tissues (Alvarez et al., 2004; Rocheta et al., 2012; Modesto et al., 2014).

1.3.2. RNA extraction

RNA extraction methods have been more diverse. Given the recalcitrant nature of tissues in woody species (Kozłowski and Pallardy, 2002), methods have to account for various pigments, polysaccharides and secondary metabolites, including phenolic compounds that affect not only downstream applications (Loomis, 1974; Richards et al., 1994; Wilkins and Smart, 1996) as well as RNA integrity (Azevedo et al., 2003) and quantification of a given sample (Wilkins and Smart, 1996).

The most recent implementation of Next Generation Sequencing (NGS) technology to study the cork oak's transcriptome employed a number of methodologies that could yield high-quality RNA suitable for such an application. Methods range from commercial kits, like RNAqueous Kit (Ambion) (Rocheta et al., 2014) or RNeasy Plant Mini Kit (QIAGEN) (Pereira-Leal et al., 2014) to more complex protocols like the Hot Borate method (Wan and Wilkins, 1994; Sebastiana et al., 2014). Also within NGS applications a number of "hybrid" approaches have been used, in which commercial kits have been employed together with extra reagents, to enhance extraction quality and yield (Pereira-Leal et al., 2014; Teixeira et al., 2014) or other cases where an additional cleaning step was introduced to eliminate any leftover contaminants (Pereira-Leal et al., 2014). Other studies have also used commercial kits (Chaves et al., 2011; Rocheta et al., 2012; Almeida et al., 2013) or kit-less approaches as described above (Soler et al., 2007; Soler et al., 2008; Rocheta et al., 2012; Almeida et al., 2013; Almeida et al., 2013), the exception being the method implemented by Marum et al.

(2012) and Pereira-Leal et al. (2014) which is a CTAB-based method with selective precipitation.

1.3.3. Real-Time Quantitative PCR (qPCR)

One of the most useful downstream applications of the above mentioned RNA extraction methods (1.2.2) was the establishment of a set of genes that display constitutive expression in different tissues, developmental stages and stress-related responses, to be implemented as internal controls for qPCR applications. Marum et al. (2012) reported ten candidate reference genes based on gene sets reported in other plants like *Arabidopsis thaliana* (Czechowski et al., 2005; Remans et al., 2008; Hong et al., 2010) but focusing on reports in woody trees like poplar, spruce and logan (Brunner et al., 2004; Gutierrez et al., 2008; Phillips et al., 2009; Lin and Lai, 2010; Xu et al., 2011). Using a number of methodologies the list was validated and the candidate genes were separated into groups, depending on the stability of expression given the stages/conditions tested and concluding that Actin (*Act*) and Clathrin adaptor complexes medium subunit family protein (CACs) were in fact the most stable of the set.

1.3.4. Molecular markers and population genomics

The selection of high quality trees for propagation can improve productivity and quality of cork and other byproducts, and give some insights into the *Quercus* genetic inheritance. Present strategies have been focused on molecular markers, like microsatellites, to establish the population genomics of *Q. suber*. Genome fingerprinting tools using Inter Simple Sequence Repeat (ISSR) and Single Sequence Repeat (SSR) microsatellites, reported as conserved in related species (Bornet and Branchard, 2001) and implemented by Bornet and Branchard (2001) and López-Aljorna et al. (2007), have helped expand the knowledge of the genetic diversity of the species. Special focus was made, in the later report, to the development of a selection procedure of individuals that produced high quality cork, for future propagation efforts. In addition, the implementation of such tools had the objective of identifying hybrid individuals of *Q. suber* and *Q. ilex* using SSR markers (Soto et al., 2003). Expanding on this idea Burgarella et al. (2009) documented the frequency of ongoing interspecific gene exchange between two Mediterranean oaks in F1 hybrids using a population genomic strategy, managing to address not only rate of hybridization but also hybrid identity.

1.3.5. OMICs and Next Generation Sequencing

Before the introduction of NGS, earlier OMICs-based studies were almost non-existent, and information about the genetics, biochemistry and physiology of *Q. suber* was scarce. Before the first NGS effort there were only about 7,000 ESTs available for *Q. suber* in NCBI's dbEST.

Proteomic analysis studies in embryos were carried to identify expression patterns between haploid and diploid stages (Gomez et al., 2009) and in somatic embryogenesis (Gomez-Garay et al., 2013), to aid the difficulties of breeding the species. Also important are findings of cork synthesis-related proteins detected in early stages of development that could lead to the development of markers for cork quality, aiding in selective breeding programmes for cork production and industry.

Possibly the most important development to come out in *Q. suber* research was the application of high throughput analysis to uncover the species' transcriptome. A Portuguese nation-wide initiative, the Cork oak ESTs Consortium (COEC), performed the de novo sequencing of the *Q. suber* transcriptome using 454 sequencing technology. cDNA samples derived from 21 normalized libraries extending multiple tissues, organs developmental stages and physiological/environmental conditions were processed and assemble, identifying ~159,000 unigenes (Pereira-Leal et al., 2014). All sequences, annotation and functional characterization were made available (www.corkoakdb.org). Within this framework, a number of side projects were carried out with the objective of analysing differential expression of the *Q. suber* transcriptome on non-normalized libraries related to the ones mentioned above. Sebastiana et al. (2014) conducted RNA-seq analysis in roots, comparing tissue inoculated with ectomycorrhizal fungi, *Pisolithus tinctorius*, and control samples without mycorrhizae. Results suggested that roots undergo extensive remodelling of cell wall, activation of biosynthesis pathway and induction of gene expression related to fungal recognition. Rocheta et al. (2014) conducted similar analysis with the objective of identifying the molecular mechanisms that are involved in flower identity and development. Samples comprised of male and female cDNA libraries in early and late developmental stages. Results highlighted some genes related to pollen development and ovule formation and a few uncharacterised or not previously reported and determinant genes in flowering. One other important study was carried out by Teixeira et al. (2014) to understand differences in transcription prolife between distinct phellogenic tissues, comprising both good and bad

quality cork. RNA-seq analysis suggested that a number of genes related to abiotic stress, especially those related to heat-shock, are responsible for greater quality of cork in *Q. suber*.

In recent years the GenoSuber project was initiated, aiming to completely sequence the 772 Mbp genome of *Q. suber*. Illumina sequencing technology is to be used for that purpose and is expected to reach a depth of 63X (www.genosuber.com). This advancement in genomic resources in *Q. suber* is of great importance, implying major research possibilities, like high depth and resolution RNA-seq experiments, not relying on de novo sequencing, and expanding the range of assays that could be carried out, like epigenetic analysis through bisulphite sequencing or population genomics and phenotype profiling, that really on genomic sequences.

1.4 Next-generation sequencing in non-model plants

Next generation sequencing (NGS) technology is revolutionizing research in distinct areas of biology. The ability to generate massive sequencing libraries with a lower cost and high degree of performance has profoundly affected our research methodologies, moving towards the “OMICS” age. The number of plant genomes being sequenced has been increasing in recent years and has been driving the entering of a genomics age. Accompanying this trend, RNA-sequencing (RNA-seq) has also provided an explosion in the number of information generated by NGS. This method has been replacing other transcriptomics methodologies as the main research tool for RNA expression profiling, surpassing other approaches with high resolution transcriptomes capable of quantifying more accurately and at higher detection limits (Martin et al., 2013). Coupling this analysis with reference genomic data enables rapid acquisition of expression profiles, isoform characterization, annotation improvement, novel gene discovery and polymorphism, using mapping methodologies with reference genome sequences (Mochida et al., 2011). When genome sequences are not available, RNA-seq analysis still enables acquisition of transcriptomic expression profiles and preliminary sequence data, enabling for downstream application and providing important characterization data in non-model plant species (Mochida et al., 2011).

Effective RNA-seq analysis in non-model plant species requires new methods of analysis, as strategies of expression profiling through differential expression analysis require a mapping step based on genome sequences, not available in this group of plant species. Apart from this procedure the RNA-seq workflow is generally the same for transcriptome or

genome mapping, but important considerations have to be considered. The general outline of the workflow is illustrated in Figure 1.5. and addressed bellow.

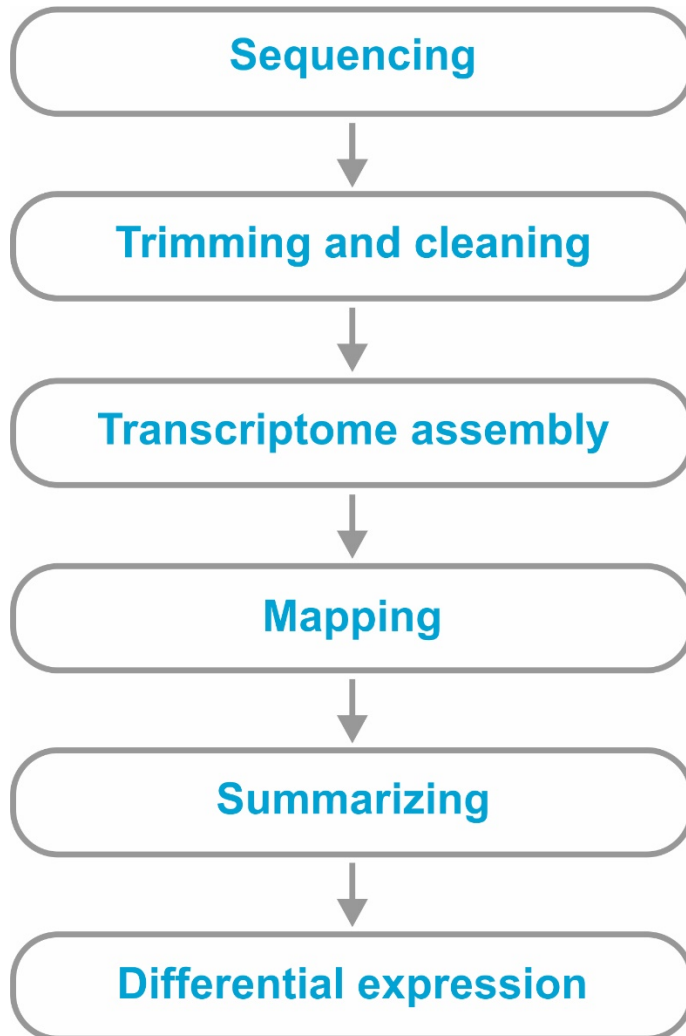


Figure 1.5. Overview of the workflow for analysis of RNA-Seq data.

A typical RNA-seq experiment starts with library preparation, that includes steps for cDNA synthesis, fragmentation and sequencing platform specific procedures like adapter ligation. Then, the sequencing step will be performed. At the present state of sequencing technology, Illumina's platforms are the most commonly applied for RNA-seq. HiSeq platforms allow for a range of sequencing length reads and read sequencing strategies, achieving high sequencing depth and allowing generation of reads suitable for mapping and assembly of highly repetitive genomes/transcriptomes (Van Verk et al., 2013). In the past, the most common platform used in de novo sequencing was Roche's 454 sequencing, providing the larger read advantage, but nowadays Illumina's MiSeq platform is catching Roche's 454 performance, but even at those levels of performance, current HiSeq sequencing depth and 100 to 150bp read length, coupled with pair-end sequencing and improvements in de novo

assembly tools, provide a better alternative to single end long-range strategies adopted in Roche's 454, due to the higher costs of the platform (Garg and Jain, 2013).

The next step, after receiving the sequences is trimming and cleaning. Adapters, poly(A) tails and low quality reads have to be trimmed and contaminants, such as PCR amplification artifacts, have to be filtered. After quality of reads is assured and artifacts that can conflict with downstream de novo assembly can be carried out. Optimized assembly strategies have been designed for short-reads coming from the Illumina HiSeq platform (Velvet, Oases, AbySS, SOAPdenovo and Trinity), based on kmer assembly (Li et al., 2014; Zhao et al., 2011), and long-read for Roche 454 and Illumina MiSeq platforms (MIRA and Newbler) based on overlap-mapping (Li et al., 2014), but former short-read tools can be applied (Garg and Jain, 2013). Hybrid approaches have also emerged, combining both data types (Martin et al., 2013), that are based on clustering strategies after independent assembly (Garg and Jain, 2013). In either strategies, the assembled transcriptome needs to be able to serve as reliable mapping reference and help with downstream applications like gene categorization or proteome sequencing (Mochida et al., 2011).

Addressing further downstream analysis, a mapping step has to be performed to align reads to reference transcriptomes. For that purpose two software tools exist for long and short read libraries and the more popular are BWA and Bowtie (Van Verk et al., 2013), respectively. The quantification step requires a step of summarization for counting the number of reads aligned to each gene/sequence. After that, statistical models can be applied to normalize count data and estimate expression. The most commonly used tool is Cufflinks, which uses RPKM/FPKM normalization (Van Verk et al., 2013) that accounts for differences between sample depth and within-sample differences. This normalization methods have been shown to introduce a bias to differential expression in favor of longer transcripts (Oshlack and Wakefield, 2009). In recent years more suitable methods have been developed accounting for read length bias, as well as the other two sequencing bias mentioned above. Those include median count ratio (DESeq) or TMM (edgeR) (Van Verk et al., 2013).

Further downstream methods include annotation and data-mining using model plant databases and tools, or closely related plant species resources available for such purposes. This can include annotation data, co-expression network assessment, subcellular localization, interaction maps, metabolic pathway mapping and gene family characterization (Mochida et al., 2011). With single-molecule sequencing emerging, the potential to sequence complete transcripts at once without the need of an amplification step or even cDNA synthesis could drastically reduce computation requirements and reduce bias introduced by those methods.

This opens possibilities for whole-genome expression profiling at an unprecedented level of detail.

1.5. Objectives and outline of the thesis

Changes in Mediterranean climate and recent declines to native populations of *Quercus suber* are of major concern due to the economic and ecological value of the species. Recent advancements in sequencing technology and transcriptomics analysis could give an insight into the species response mechanisms to drought, given it is one of the emerging adverse conditions to affect the Mediterranean basin. The main objective of the present work is to uncover drought response mechanisms through RNA-seq analysis at root level, focusing on the regulation, signaling, and downstream mechanisms. Since this species is a non-model organism it lacks many important tools to aid in the subsequent analysis and downstream data mining. To address this issue a de novo assembly approach was adopted to generate a transcriptome, to be used as reference to read alignment steps and ORF discovery. After the determination of differentially expressed genes (DEGs) and general characterization of such genes, further analysis was carried out on the model plant *Arabidopsis thaliana*, using the DEGs' homologous genes, thus providing access to the extensive knowledge and tools that are available in this species. To elucidate the mechanisms underlying drought stress responses, *in silico* analysis was performed through network analysis, *cis*-element enrichment and transcription factor identification. To confirm RNA-seq results and major contributors to drought response qPCR was performed and key determinants were identified for future studies.

Chapter 2

MATERIALS AND METHODS

CONTENTS

2.1 Plant material and physiological assessments	18
2.2 454 sequencing	18
2.3 Sequence assembly and differential expression analysis	18
2.4 Protein prediction, functional annotation and ortholog assessment	19
2.5 Functional data mining in Arabidopsis	20
2.5.1 Gene Network analysis	20
2.5.2 Transcription factor identification	20
2.5.3 Cis-element enrichment	20
2.5.4 Phylogenetic analysis	20
2.6 Quantitative Real-Time PCR	21

2.1. Plant material and physiological assessments

Quercus suber acorns were supplied by Instituto Superior de Agronomia (ISA, Lisbon). Acorns were stratified at 4°C for two weeks and subsequently sowed in a soil mixture consisting of a 2:1:1 volume mix of turf, vermiculite and silicate sand. A total of 123 vases were sowed and grown in a 16 h light/8 h dark photoperiod regime (100 $\mu\text{mol Photon m}^{-2} \text{ s}^{-1}$). After 80 days, drought treatments were initiated. Treatments used 90 plants with the best fitness, divided into eighteen groups, with five plants each. Plants from each group were subjected to five different drought treatments, corresponding to 100%, 90%, 50%, 25% and 10% of the field capacity of the soil mixture, respectively identified as D100, D90, D50, D25 and D10. Watering was performed three times a week. In each group, plants were watered with 100%, 90%, 50%, 25% or 10% of the pot weight loss measured in D100 samples.

Quantum yield of photosystem II (F_v/F_m) was assessed by pulse amplitude modulated (PAM) fluorometry, after adapting plants to the dark for 15 min, using a portable Pulse Amplitude Modulated fluorometer (Junior-PAM, Gademann Instruments GmbH, Germany). Blue light was used as the light source. For each condition, three healthy leaves were selected in three independent plants, and F_v/F_m was evaluated throughout the drought assay. Chlorophyll and carotenoid pigments were quantified spectrophotometrically as previously reported (Andralojc 1987). Quantification of L-proline levels in roots was performed using the ninhydrin method, as previously reported (Bates et al. 1973). Commercially available L-Proline was used for a standard curve.

2.2. 454 sequencing

Total RNA was isolated from secondary roots. These were snap-frozen and grinded in liquid nitrogen. Total RNA was extracted following the Hot Borate method (Wan and Wilkins, 1994). Total RNA was further purified using RNeasy plant mini kit (Qiagen). High quality RNA was sequenced in a Genome Sequencer GS FLX Titanium (Roche-454 Life Sciences, Brandford, CT, USA), according to the standard manufacturer's instructions at Biocant (Cantanhede, Portugal). Each library (D10/25, D50 and D90/100) occupied one half of a picotiter plate. Raw reads were subsequently subjected to trimming, removal of adapters/primers and low quality reads using the service provider (Biocant) in-house pipeline.

2.3. Sequence assembly and differential expression analysis

De novo assembly was carried out using Newbler (GS De Novo Assembler V2.9,) with default parameters and the resulting assembly file was submitted to the CD-HIT-EST web server (Li and Godzik, 2006; Huang et al., 2010; Fu et al., 2012) to eliminate

Chapter 2. MATERIALS AND METHODS

redundancies originated from the *de novo* assembly. Sequences with less than 100bp were removed and the remaining isotigs and contigs were named as *Unigenes*. Unigenes were given an accession number using an in-house Pearl script. Sequences were deposited in the National Center for Biotechnology Information (NCBI), under the Short Read Archive (SRA) accession number SRP055382. Coverage for the assembled transcriptome was assessed with BLASTx analysis against the *Castanea molissima* predicted transcriptome, *Quercus rubra* Unigene v3 transcriptome from Fagaceae Genomics Web (<http://www.fagaceae.org/>) and *Arabidopsis thaliana* predicted transcriptome (<http://www.arabidopsis.org/>) with an e-value threshold of 10^{-6} .

To determine the set of differentially expressed genes (DEGs) between the three experimental conditions, HQ reads from each sequencing effort were mapped against the assembled transcriptome using the Burrows-Wheeler Aligner (BWA) (bio-bwa.sourceforge.net). The algorithm chosen was BWA-MEM as it features long-read support and split alignment that is suitable for 454 reads, and is faster and more accurate when compared to BWA-SW. Parameters were set to promote single alignments for each read. Summarization of results was performed using Tablet (Milne et al., 2012). Venn diagrams were generated using Venny (bioinfogp.cnb.csic.es/tools/venny/index.html). Identification of differentially expressed genes was performed in DESeq (Anders and Huber, 2010), a software package for R (R Development Core Team, 2008). Using DESeq we modelled count data using a negative binomial distribution, paring all conditions (D100/D90 vs D50; D25/D10 vs D50; D100/D90 vs D25/D10). We retained unigenes that displayed differential expression (significance value <0.05) in at least one comparison. CDBFasta (<http://cdbfasta.sourceforge.net/>) was used to retrieve specific fasta sequences from the assembled unigenes file.

DEG unigene lists from all three pairs were subjected to comparison via an online Venn generator (<http://bioinformatics.psb.ugent.be/webtools/Venn/>). For Hierarchical clustering (HCL; (Eisen et al., 1998)), normalized read counts for each DEG were retrieved from DESeq; HCL was performed by Euclidean distance and average linkage clustering on Multiple Experiment Viewer (MeV v4.0; <http://www.tm4.org/mev.htm>). Gene clusters were identified using the MeV feature Self Organizing Tree Algorithm (SOTA) (Herrero et al., 2001) by Euclidean distance. Expression plots were rendered considering SOTA clustering and DEGs were separated into UP-regulated and Down-regulated based on Expression plot profile.

2.4. Protein prediction, functional annotation and ortholog assessment

Differentially expressed unigenes were subjected to homology search using BLASTX against the non-redundant protein database (nr) with an e-value threshold of 10^{-6} (Supplementary File S1). The ORF Predictor (<http://proteomics.ysu.edu/tools/OrfPredictor.html>) tool was used to identify open reading frames (ORFs). For this purpose, BLASTX results and nucleotide sequences were uploaded to the ORF Predictor server, so frame shift errors and incomplete sequences could be accounted for. Corrected CDS and protein sequences were retrieved for later analysis. Nucleotide sequences were also analyzed using the Blast2GO pipeline (Conesa et al., 2005) using standard settings and the local BLASTX previously generated. To establish *Arabidopsis thaliana* orthologs, DEG protein sequences were subjected to BLASTp analysis (e-value of 10^{-6}) against the reference *Arabidopsis thaliana* TAIR10 annotation (www.arabidopsis.org). Only the gene model was considered for later analysis.

2.5 Functional data mining in Arabidopsis

2.5.1 Gene Network analysis

Gene networks were predicted in Cytoscape, using the Genemania plugin. Network layouts were as follows: functional features, Prefuse Force-Directed Layout; sub-cellular localization, Edge-weighted Spring Embedded Layout; co-expression, Prefuse Force-Directed Layout. Default parameters were used but changes were made so that only the DEG list was considered for analysis. For functional network prediction Gene Ontology co-annotation was used to highlight clusters or genes of interest. For co-localization network, GO ontology annotation of DEG orthologs was carried by AgBase database web server (<http://agbase.msstate.edu/>). Genes were mapped for the second GO slim level and given distinct colours within the predicted network.

2.5.2 Transcription factor identification

TFs were identified from Arabidopsis DEG orthologs using the AGRIS (Davuluri et al., 2003) Arabidopsis transcription factor database (AtTFDB, <http://arabidopsis.med.ohio-state.edu/AtTFDB/>), MapMan (transcription pathway bin; (Thimm et al., 2004; Usadel et al., 2005; Usadel et al., 2009)), and manual curation

2.5.3 Cis-element enrichment

Cis-element enrichment analysis was carried using AtCOECis (Vandepoele et al., 2009) and Athena (O'Connor et al., 2005), with default parameters.

2.5.4 Phylogenetic analysis

For each class of transcription factors (TFs), we used automated gene family annotation to retrieve the a.a. sequences of all family members present in the *Arabidopsis thaliana* genome. For that purpose the Plaza web-based platform (version 2.5; (Proost et al., 2009; Van Bel et al., 2011)) was used. Phylogenetic analysis was performed in SeaView v4.5. (Gouy et al., 2010) Sequences of *Q. suber* differentially expressed family members and *Arabidopsis* family members were aligned with the MUSCLE algorithm (Edgar, 2004). The PhyML v3.1 (Guindon et al., 2010) feature of Seaview was used to compute maximum likelihood trees using the LG substitution model (Le and Gascuel, 2008), with branch support using the aLRT SH-like method (Anisimova and Gascuel, 2006).

2.6 Quantitative Real-Time PCR

For quantitative Real-Time PCR (qPCR) analysis, a second independent drought induction assay was performed, and total RNA was isolated from secondary roots by a CTAB-based method adapted from Azevedo et al. (2003). RNA quantity and quality were assessed using a Nanodrop ND-1000 spectrophotometer and electrophoretic analysis. RNA samples were treated with Recombinant DNase I (Takara Biotechnology), and cDNA was subsequently generated using a Superscript III Reverse Transcriptase Kit (Invitrogen). For qPCR reactions, Ssofast Evagreen Supermix (Bio-Rad) was used according to the manufacturer's indications. Reactions used 1:20 cDNA dilutions. Analysis was run in a CFX96 Touch Real-Time PCR Detection System (BIORAD), as per the manufacturer's instructions. Relative expression was determined by the Pfaffl's equation (Pfaffl, 2001), normalizing with PP2A gene. Values were calculated between severe drought conditions (D25/10) and well watered conditions (D100/90).

Primers for qPCR were designed using NCBI Primer-BLAST (Ye et al., 2012) and design for product length amplification of 100 to 200 bp, ~60°C on T_m and 50 to 60% of GC content.

Primers are listed below:

Table 2.1 – List of quantitative qPCR primers. Primers were designed to ensure specific amplification, 100-200 bp amplification products, and ~60°C T_m.

Gene Name (Unigene ID)	Primer Sequence (5' to 3')	Length (nt)	T _m (°C)	Product size (bp)
QsHAI2	CAGGTGGTCGGGTCATCTAC	20	60.4	181
(QSDrought_01105)	ATCACGTCCCAAAGCCCATC	20	59.5	

Table 2.1 (Continued)

QsHAI3	GTTCTGGGAATTTCTGACTGGG	21	60.1	176
(QSDrought_03606)	CCACTCTGAACCTCGTCATCC	21	58.9	
QsSNRK2.6	AGCGGAATGAAGAGCCAGAG	20	60.6	170
(QSDrought_03551)	GAGCCTGTCCACCACACTTC	20	59.8	
QsABF2	TCGGATCCAAGAAGCTGGTG	20	60.9	165
(QSDrought_02295)	CGCTCGCCAGGATTAGACAC	20	59.8	
QsRD29B	GCCAAGAGGATCACACTCCC	20	59.8	167
(QSDrought_03061)	TCTCCTGGCATCTTTGCCTC	20	60.1	
QsRD26	ACAGCCACTCTTCCCCAAAC	20	59.8	175
(QSDrought_03374)	CCTACCCTGCCAATTCCTCC	20	60.2	
QsCOR47	GTAGATGAACATGGGCCCCG	20	60.3	145
(QSDrought_05686)	CACCGCCATGAGTAGGTACG	20	60.5	
QsNCED3	TGCATGACTCCAGCTGACTC	20	60.3	176
(QSDrought_05518)	TTCTTCCAAGCCTGTTCCG	20	59.8	
QsXTH15	CCCTTCCCTAAGAGCCAACC	22	59.9	179
(QSDrought_06025)	TGCCATGTAGAATCGGAGGAAG	20	59.7	
QsNAC100	AAGGGTAGAGCCCCAAAAGG	22	58.1	111
(QSDrought_06406)	CAGTCTCCAGTCTCACTATCCC	20	60.1	
QsPP2A-3	GGGTTCCCAACATCAAGTTC	20	56.6	174
(QSDrought_03365)	TGACCTGATCACTTGACTGC	20	57.5	

Chapter 3

Results and discussion

CONTENTS

3.1 Physiological response to long-term drought	25
3.2 Assembly of the <i>Quercus suber</i> root transcriptome	26
3.3 Differential expression analysis of the drought root transcriptome	30
3.4 Annotation of differentially expressed genes	32
3.5 Functional data mining of DEGs	36
3.6 Identification of an ABA-dependent signalling network	46

3.1. Physiological response to long-term drought

In the present work, the cork oak root-level response to drought was investigated at the transcriptional level by an RNA-Seq approach. Prior to transcriptional analysis, *Q. suber* plantlets were subjected to prolonged drought stress and characterized in their morphological and physiological responses. More specifically, 80-day-old plantlets were subjected for one month to five different watering regimes (herein designated D100, D90, D50, D25 and D10), that restored 100%, 90%, 50%, 25% or 10% of the water lost between watering periods. This imposed a progressive loss in water availability (Figure 3.1A), resulting in moderate to extreme drought stress over an extended period of time. The consequences of drought stress imposition were visible in leaf morphology, with leaves displaying increasing stress-related symptoms such as leaf area drop, leaf rolling, edge and tip necrosis, and changes in colouration observable as reddening and yellowing (Figure 3.1B-D). In order to support morphological data, we monitored alterations to plant fitness throughout the treatment period using Pulse-amplitude-modulation (PAM) fluorometry, since this technique has been successfully used to study alterations in photosynthetic electron transport *in vivo* (Sixto et al., 2006; Papageorgiou, 2007), measuring photosynthetic performance determined by the F_v/F_m parameter (maximum quantum efficiency of PSII photochemistry). Here, we could observe that a progressive loss in efficiency of PSII was visible for all major drought treatments, especially for D25 and D10 (Figure 3.1E). This may relate to loss in stability of the PSII and degradation in photochemical activity of this system under water stress (Havaux, 1992). Our results are also consistent with other studies in oak species undergoing water deprivation, in which a decrease in F_v/F_m was observed (Epron et al., 1993; Méthy et al., 1996). In addition, we estimated chlorophyll and carotenoid contents in leaves at the end of the experiment, observing a reduction in both pigments that was proportional to the severity of the drought treatment (Figure 3.1F, G). Overall results are consistent with previous reports that associate molecular and physiological responses to water deficit to chlorophyll bleaching, decline in photosynthesis, accumulation of carotenoid-like molecules, remobilization of nutrients, dismantling of cellular organelles, and programmed cell death events (Shao et al., 2008). Finally, to confirm drought stress imposition we quantified root proline content. Proline is a standard indicator of low Ψ_w caused by environmental stresses, acting as a compatible solute that accumulates to high concentrations following decreases in water availability (Singh et al., 1972; Ibarra-Caballero et al., 1988; Lambers et al., 2008). In the present report we were able to observe a significant and dose dependent increase in the proline content of drought-stressed *Q. suber* roots (Figure 3.1H), thus confirming the imposition of drought stress. Overall, we

Chapter 3. RESULTS AND DISCUSSION

clearly established that in our physiological model, different watering regimes matched different intensities of long-term drought stress imposition.

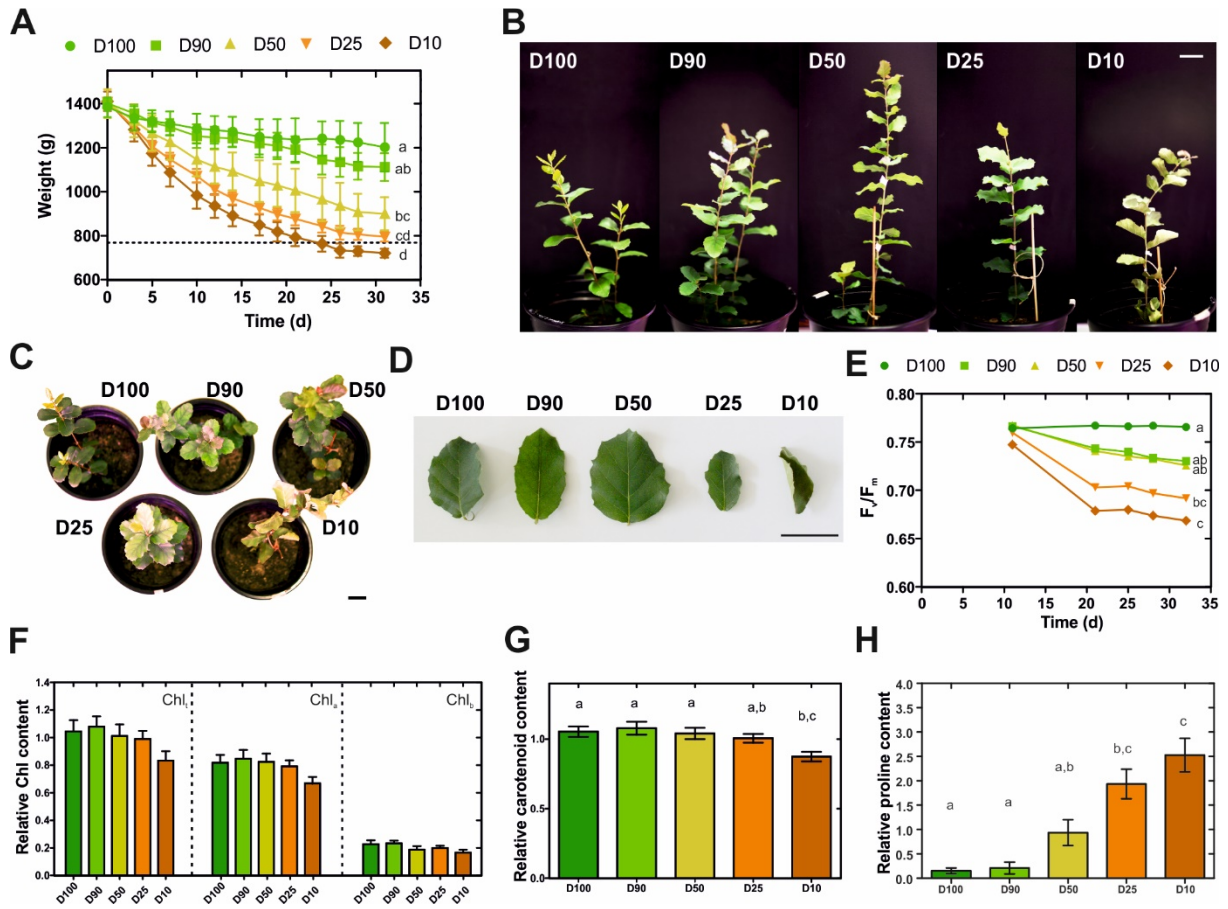


Figure 3.1. Characterization of morphological and physiological traits of *Q. suber* plantlets subjected to long-term drought stress.

(A) Eighty-day-old plantlets were subjected to five different watering regimens (D100, D90, D50, D25, D10) for one month; pot weight was determined before each watering period, (n=6).

(B-C) Plant (B, C) and leaf (D) morphology at the end of the drought stress period.

(E) Effect of each watering regimen in the quantum yield of photosystem II (F_v/F_m), (n=9).

(F) Total chlorophyll (Chl), chlorophyll a (Chla) and chlorophyll b (Chlb) contents at the end of the drought stress period, (n=15).

(G) Carotenoid content at the end of the drought stress period, (n=15).

(H) Quantification of L-proline in *Q. suber* roots at the end of the drought stress period, (n=15)

Letters represent statistically similar results (one-way ANOVA with post-hoc Tukey test to compare all columns; $p > 0.05$). Scale bars represent 3 cm.

3.2. Assembly of the *Quercus suber* root transcriptome

We were involved in a nation-wide initiative, the *Cork oak ESTs Consortium* (COEC) that performed *de novo* sequencing of the *Q. suber* transcriptome by 454 NGS (www.corkoakdb.org/, (Pereira-Leal et al., 2014)). Within this initiative, we performed an RNA-Seq analysis of the *Q. suber* root response to long term drought. RNA was extracted

from roots of plantlets subjected to different water usage conditions (D100+D90, D50 and D25+D10), and used to generate non-normalized libraries that were then sequenced by 454 GS FLX Titanium technology, using half-plate runs for each experimental condition (Figure 3.2).

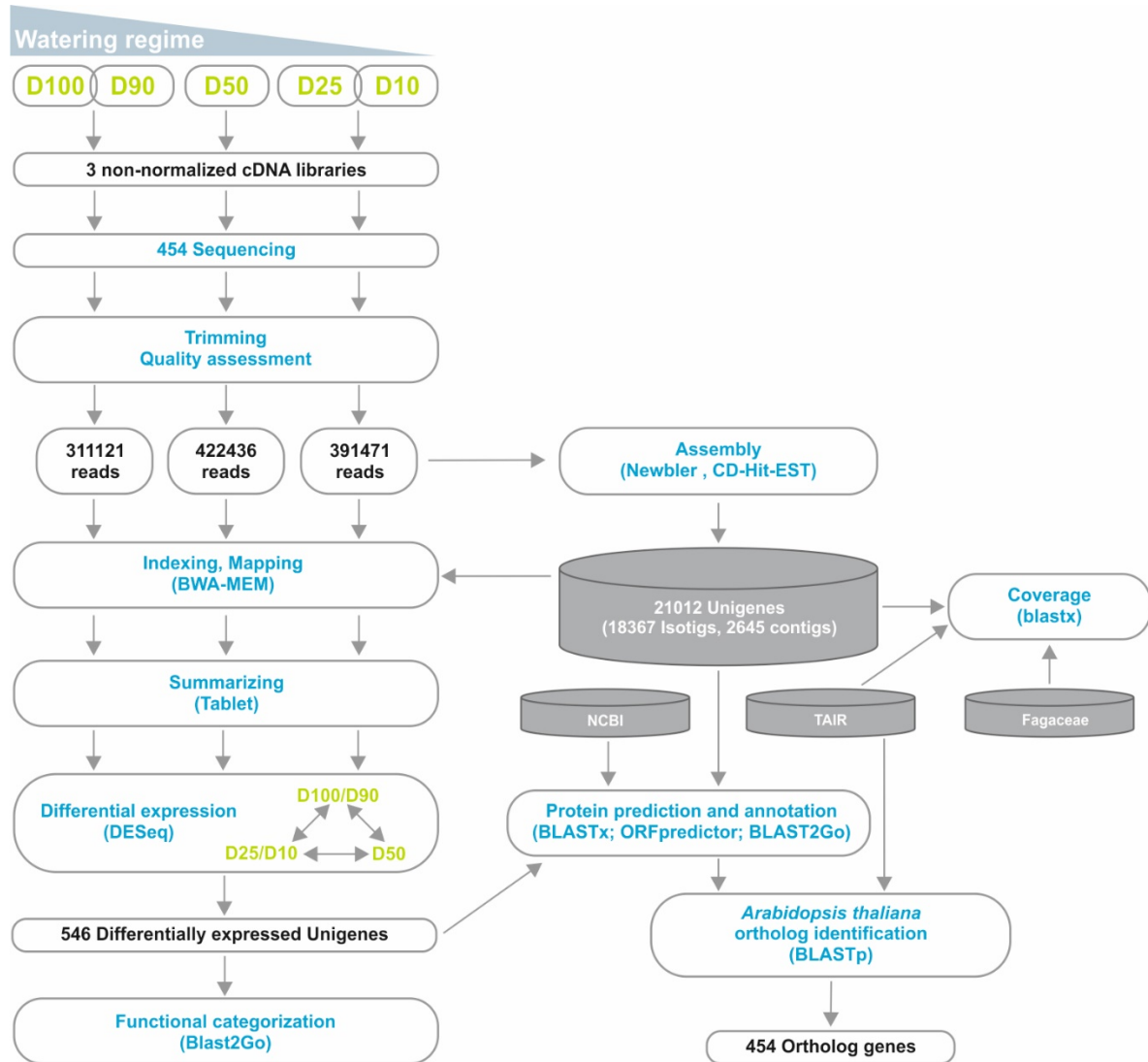


Figure 3.2. Outline of the RNA-Seq strategy employed in the identification of the *Q. suber* differential root transcriptome in response to long-term drought. Functional characterization was extended by identification of *Arabidopsis thaliana* orthologs for the differentially expressed genes.

Results from the sequencing effort are presented in Table 3.1. Raw read number (1.8 million) and length (~400 bp) followed expected standards for this technology. Raw reads were subjected to a pre-processing/trimming step to remove short or low quality sequences and adaptor/primer sequences. High quality (HQ) reads amounted to 88.1% of original reads, and totalled 470.8 Mbp (Table 3.1). HQ reads displayed no significant differences in number

Chapter 3. RESULTS AND DISCUSSION

or length frequency, between different drought stress regimes (Table 3.1; Figure 3.3A), corroborating the quality of the cDNA library and sequencing processes. Subsequently, we assembled a reference transcriptome using HQ reads, resulting in a total of 22,455 raw sequences; these were uploaded to the CD-HIT-EST web server (Li and Godzik, 2006; Huang et al., 2010; Fu et al., 2012) for clustering and redundancy elimination, resulting in 21,012 unigenes of which 18,367 were isotigs and 2,645 were contigs (Figure 3.2, Table 3.1). Frequency analysis demonstrated that the assembled unigenes were as long as 2,600 bp, averaged at 759 bp and peaked in the 400-500 bp range (Figure 3.3B).

Table 3.1. - Statistics of the sequencing and assembly steps.

	D100/D90	D50	D25/D10	Total
Raw reads				
Number	686,891.00	620,615.00	501,296.00	1,808,826.00
Total Number of Bases	270,949,869.00	250,845,326.00	198,528,142.00	720,323,337.00
Average length (bp)	394.60	404.30	396.00	398.30
Processed reads				
Number	548,456.00	601,692.00	442,674.00	1,592,822.00
Total Number of Bases	163,984,625.00	172,383,797.00	134,446,167.00	470,814,589.00
Average length (bp)	299.00	286.50	303.70	296.40
Assembly				
Number of Contigs	-	-	-	2,645.00
Number of Isotigs	-	-	-	18,367.00
Number of Isogroups	-	-	-	19,579.00
Number of Unigenes	-	-	-	21,012.00

For quality validation, assembled transcriptome coverage was established against the *Quercus rubra* Unigene V2 transcriptome, *Castanea molissima* predicted transcriptome and *Arabidopsis thaliana* transcriptome (Figure 3.4). Results demonstrated that the assembled *Q. suber* transcriptome had a very high coverage, with matches representing between 50-67% of previously established transcriptomes, and 86.5% of the *Q. suber* assembled transcriptome. Together with other recent *Q. suber* 454 RNA-Seq efforts, results also suggest that Newbler-based assembly (present report; (Rocheta et al., 2014; Teixeira et al., 2014) is more suited for 454 libraries than MIRA-based assembly (Pereira-Leal et al., 2014; Sebastiana et al., 2014),

since the latter studies presented poorer contig metrics. This supports the previous claim that Newbler outperforms other assemblers with regards to 454 pyrosequencing data (Kumar and Blaxter, 2010).

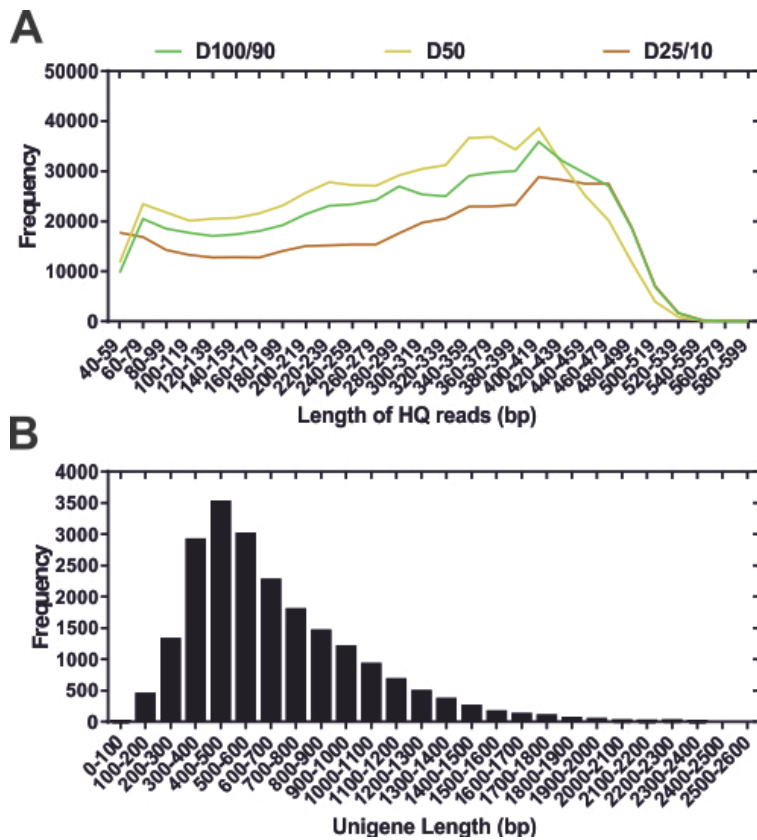


Figure 3.3. Characterization of sequence data regarding processing and assembly of reads.
(A) Distribution of the length of HQ reads obtained for each watering regime.
(B) Distribution of the length of unigenes obtained after assembly.

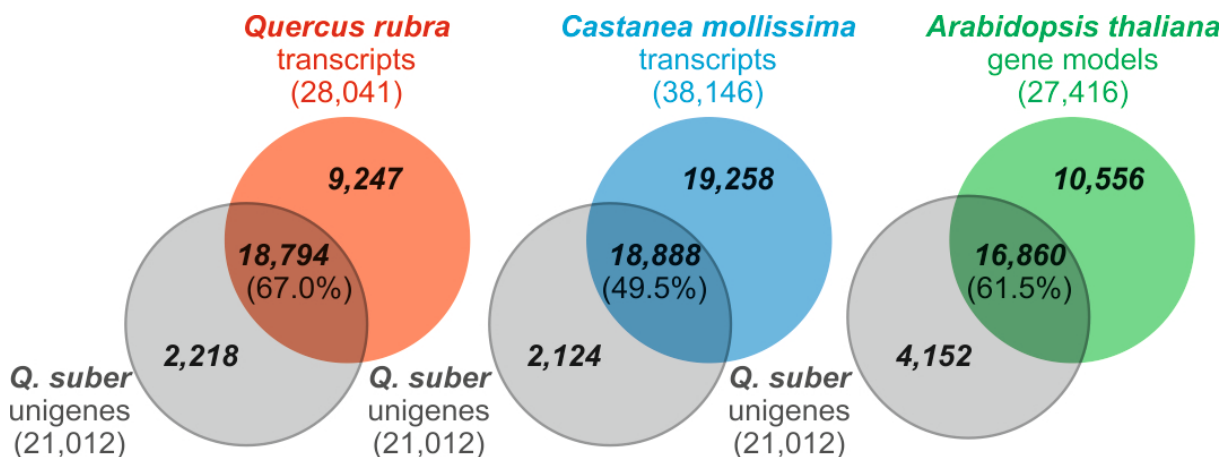


Figure 3.4. Venn diagram representation of the number of homology hits ($e < 10^{-6}$) between *Q. suber* unigenes and *Arabidopsis thaliana* TAIR 10 gene models, *Q. rubra* putative transcripts and the *Castanea mollissima* predicted transcriptome. Percentages indicate the proportion of reference transcriptomes that match *Q. suber* unigene models.

3.3 Differential expression analysis of the drought root transcriptome

For differential expression analysis, HQ reads from each sample were mapped to the reference transcriptome (Figure 3.2). Distribution analysis for the number of reads mapped to each unigene demonstrated that approximately 98% of unigenes displayed a fairly low number of reads (<100) (Figure 3.5A). Consistently, the different watering regimes showed similar distribution profiles (Figure 3.5A).

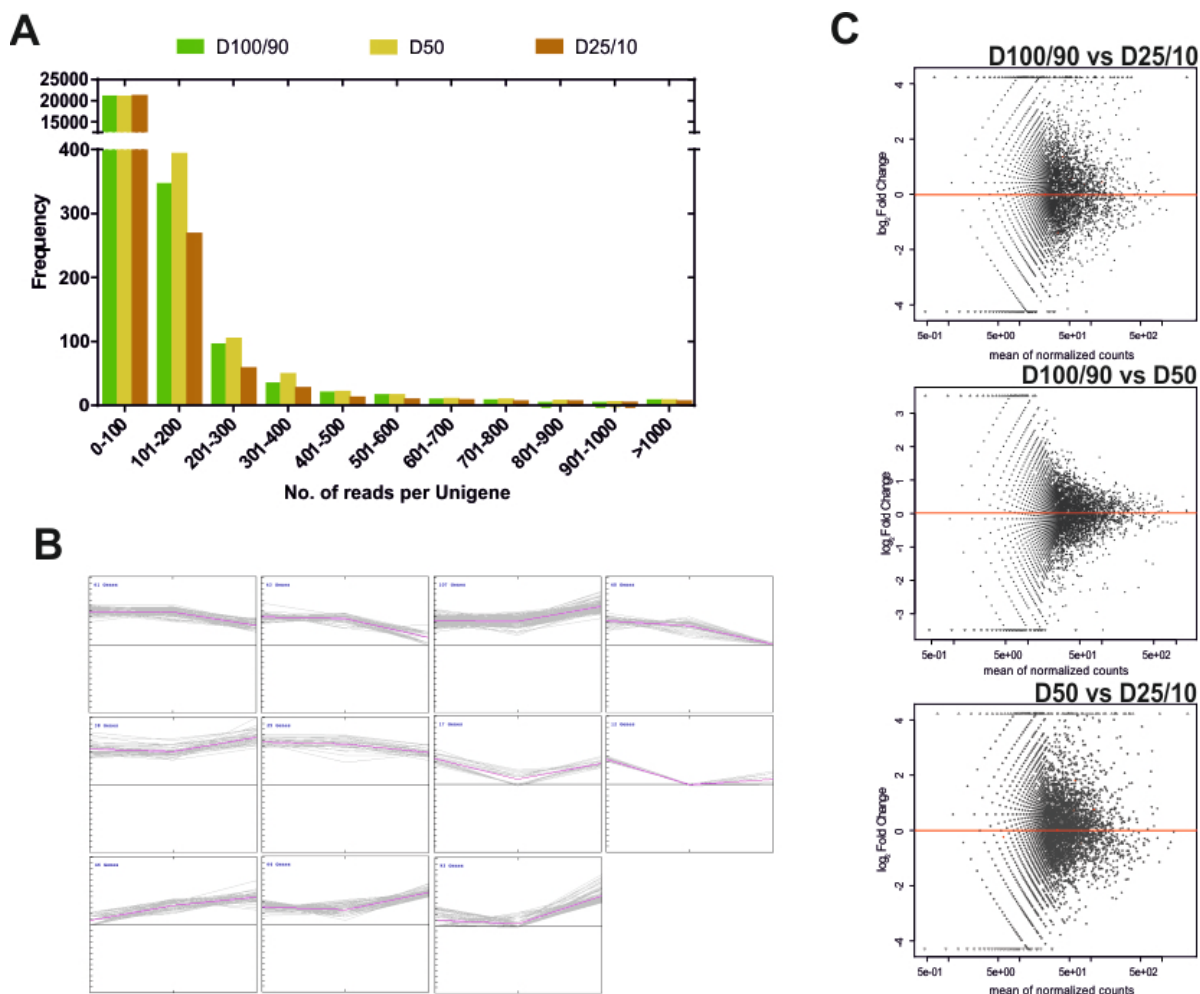


Figure 3.5. Characterization of differential gene expression in drought-stressed *Quercus suber* roots. (A) Distribution, for each watering regime, of the number of reads mapped to each unigene (unigene coverage). (B) Expression plot of the 11 clusters identified by SOTA analysis. (C) MA plots of the paired differential expressions for D100/90 vs D50, D100/90 vs D25/10 and D50 vs D25/10. Graphs represent scatter plots of each individual unigene's $\log_2$ fold change in expression, as a function of the mean of normalized counts. Note that most of the genes are represented as having basal expression values (between zero and one fold change) and few have large changes in expression, thereby validation DESeq parameters used for the differential expression analysis.

To establish differential expression, unigenes were then subjected to read count differential analysis by Negative Binomial Distribution (Huber and Anders). Briefly, all three experimental situations were paired for differential expression and genes that displayed p value <0.05 in at least one comparison were selected, resulting in a total of 546 differentially expressed genes (DEGs) (Figure 3.2; Figure 3.6).

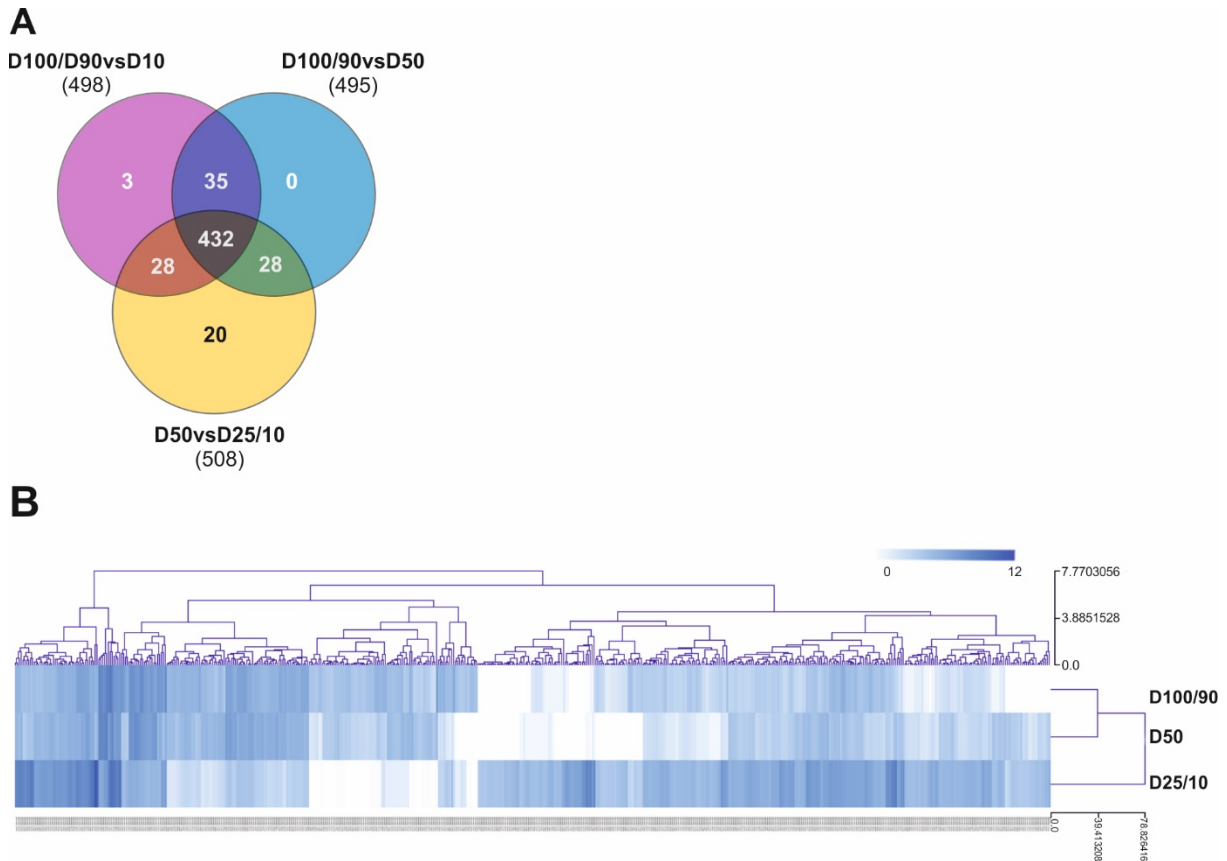


Figure 3.6. Characterization of differential gene expression in D100/90 vs D50, D100/90 vs D25/10 and D50 vs D25/10 transcriptional comparisons.

(A) Venn diagram representing unigenes in each experimental comparison.

(B) Hierarchical clustering analysis of DEG gene expression values.

Interestingly, comparative analysis showed that most DEGs (79%) were expressed differentially throughout the three transcriptional pools (Figure 3.6A), suggesting the existence of dramatic/extensive transcriptional rearrangement between control, moderately- and severely-stressed plants. Hierarchical clustering analysis was subsequently performed on DEGs (Figure 3.6B). Based on expression patterns, we differentiated 11 gene clusters containing 353 up-regulated and 193 down-regulated unigenes (Figure 3.6B; Figure 3.5B). In response to long term drought, up-regulated genes surpassed down-regulated genes by over two-fold. Results support that the present transcriptomics strategy was efficient in resolving transcriptional changes. Data specifically suggests that the switch from moderate to severe

drought (*D50* vs *D25/D10*) was transcriptionally more dramatic, considering that it evidenced (1) the highest number of unique DEGs (Figure 3.6A), (2) the most scattered pattern in MA plot projections (Figure 3.5C), (3) the most significant differences in fold change values (Figure 4B; Figure 3.5B).

3.4 Annotation of differentially expressed genes

For functional annotation, DEGs were subjected to (1) BLASTx homology search against the Non-redundant GenBank Protein Database, (2) Blast2Go analysis (Conesa et al., 2005), and the complete set of annotated genes is compiled in Appendix 1. Quality of the sequencing/assembly strategy was validated by various forms. First, Figure 3.7A depicts that the majority of unigene homology hits had extremely low E-values. Also, only 5.7% of DEGs failed to annotate with known proteins, even when threshold E-values were raised up to 10^{-1} (data not shown).

This is a substantial improvement over the rate of 45% to 35% unsuccessful annotation obtained previously for the *Q. suber* transcriptome (Pereira-Leal et al., 2014; Sebastiana et al., 2014). This is less prominent when compared with other Newbler-based assembly efforts that range between 70% and 80% (Rocheta et al., 2014; Teixeira et al., 2014).

Taking into account BLAST E-value distribution, peaking at extremely low E-values and annotation coverage observed between assembly methodologies suggests that the raise in of overlap identity, 95% compared to 90% adopted by other *Q. suber* transcriptomes assembled with Newbler was greatly beneficial, successfully generated less assembly errors and more complete sequences.

Finally, species distribution for blast hits placed unigene homologues within plant species (Figure 3.7B), supporting the initial trimming step that eliminated foreign sequences and contaminants.

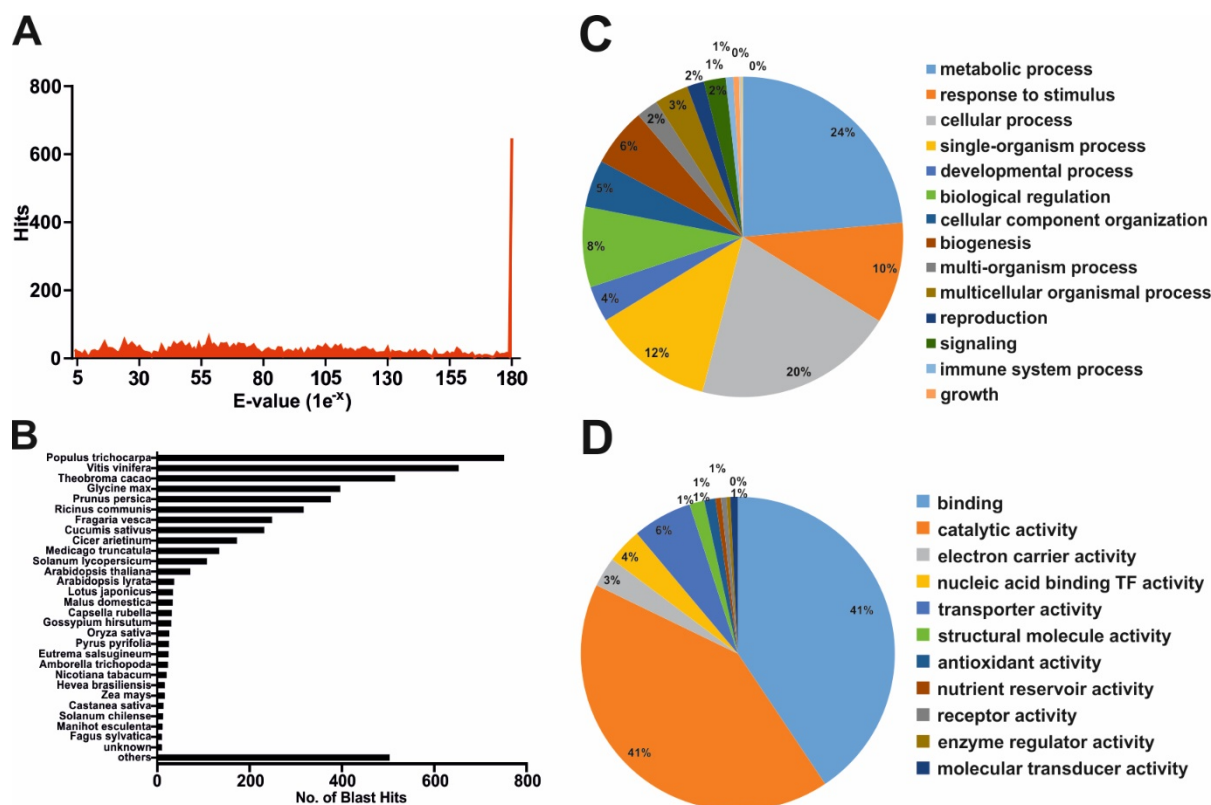


Figure 3.7. Annotation of differentially expressed genes by BLAST homology search.
(A) E-value distribution of the first 50 BLAST hits per each unigene.
(B) Species distribution of the first 50 BLAST hits per each unigene.
(C) Functional categorization of annotated genes by BLAST2GO (GO *Biological Process*) analysis.
(D) Functional categorization of annotated genes by BLAST2GO (GO *Molecular Function*) analysis.

Table 3.2 highlights the most significant DEGs based on functional categories that are of interest to our study. DEGs included an important number of effector proteins traditionally associated with drought responses, such as proteins with chaperone activity (dehydrins and LEA proteins) (Umezawa et al., 2006; Shao et al., 2008), solute transporters, and biosynthetic genes of compatible osmolites (Wang et al., 2003) and proteins associated with cell wall remodelling (Tenhaken, 2014). DEGs also included regulatory components of known drought-related signalling pathways, such as Protein Phosphatase 2C (PP2Cs) (Schweighofer et al., 2004) and serine/threonine kinases (SNRK2s) (Yoshida et al., 2002), as well as genes associated with ubiquitin-mediated protein degradation (Lee and Kim, 2011) (Table 3.2), suggesting a tight control of drought responses at both the transcriptional and protein turnover levels. Several transcription factors (TFs) were also identified (Appendix 1) that were later the focus of more in-depth analysis (see below).

Chapter 3. RESULTS AND DISCUSSION

Table 3.2. - Selection and categorization of significant *Quercus suber* unigenes differentially expressed in response to drought.

Unigene ID	DEG	<i>Arabidopsis thaliana</i> AGI	Gene Name	Annotation
Signalling				
QSDrought_00285	Up	AT4G30960	CIPK6	ATCIPK6 CIPK6 SIP3 SNRK3.14
QSDrought_03501	Up	AT5G55090	MAPKKK15	MAPKKK15 mitogen-activated protein kinase 15
QSDrought_03551	Up	AT4G33950	ATOST1	ATOST1 SNRK2.6 SRK2E Protein kinase superfamily protein
QSDrought_04452	Up	AT5G10930	SnRK3.24	CIPK5 SnRK3.24 CBL-interacting protein kinase 5
QSDrought_06388	Down	AT5G50860	AT5G50860	Protein kinase superfamily protein
QSDrought_06651	Up	AT2G48010	RKF3	RKF3 receptor-like kinase in in flowers 3
QSDrought_06886	Up	AT3G25250	AGC2	AGC2 AtOXI1 AGC (cAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein
QSDrought_08928	Up	AT5G47070	AT5G47070	Protein kinase superfamily protein
QSDrought_09509	Up	AT3G45640	ATMPK3	ATMAPK3 ATMPK3 MPK3 mitogen-activated protein kinase 3
QSDrought_09888	Down	AT1G11330	AT1G11330	S-locus lectin protein kinase family protein
QSDrought_11117	Down	AT2G21480	AT2G21480	Malectin/receptor-like protein kinase family protein
QSDrought_14166	Up	AT2G33580	AT2G33580	Protein kinase superfamily protein
QSDrought_20555	Up	AT5G60550	GRIK2	GRIK2 geminivirus rep interacting kinase 2
LEA				
QSDrought_09652	Up	AT5G06760	LEA4-5	Late Embryogenesis Abundant 4-5
QSDrought_06522	Up	AT2G35980	NHL10	LEA hydroxyproline-rich glycoprotein 10
QSDrought_07106	Up	AT2G35981	NHL10	LEA hydroxyproline-rich glycoprotein 10
QSDrought_08053	Up	AT1G52690	LEA7	Late Embryogenesis Abundant 7
Chaperone activity				
QSDrought_16809	Up	AT3G03773	HSP20-like	HSP20-like chaperones superfamily protein
QSDrought_12322	Up	AT1G20450	EDR10	Dehydrin family protein EDR10
QSDrought_07218	Up	AT1G76180	ERD14	Dehydrin family protein ERD14
QSDrought_13469	Up	AT4G10710	SPT16	global transcription factor C
Protein serine/threonine phosphate activity				
QSDrought_01105	Up	AT1G07430	HAI2	HAI2 highly ABA-induced PP2C gene 2
QSDrought_03606	Up	AT2G29380	HAI3	HAI3 highly ABA-induced PP2C gene 2
QSDrought_06016	Up	AT3G62260	AT3G62260	Protein phosphatase 2C family protein

Table 3.2 (Continued)

QSDrought_07848	Up	AT1G72770	HAB1	HAB1 homology to ABI1
QSDrought_09366	Up	AT2G30020	AP2C1	AP2C1 clade B of the PP2C-superfamily
QSDrought_11908	Up	AT1G34750	AT1G34750	Protein phosphatase 2C family protein
QSDrought_12112	Up	AT5G59220	HAI1	HAI1 SAG113 highly ABA-induced PP2C gene 1
Cell wall remodelling				
QSDrought_00248	Up	AT3G45970	ATEXLA1	Expansin-like A1
QSDrought_01805	Up	AT4G17030	ATEXLB1	Expansin-like B1
QSDrought_05775	Up	AT4G38400	ATEXLA2	Expansin-like A2
QSDrought_11848	Up	AT4G17030	ATEXLB1	Expansin-like B1
QSDrought_06025	Down	AT4G14130	XTH15	Xyloglucan endotransglucosylase 15
QSDrought_05100	Up	AT4G25810	XTH23	Xyloglucan endotransglycosylase 6
QSDrought_07044	Up	AT5G57550	XTH25	Xyloglucan endotransglucosylase/hydrolase 25
Oxygen binding				
QSDrought_00336	Down	AT5G63450	CYP94B1	CYP94B1 cytochrome P450, family 94, subfamily B, polypeptide 1
QSDrought_00338	Up	AT3G01900	CYP94B2	CYP94B2 cytochrome P450, family 94, subfamily B, polypeptide 2
QSDrought_02537	Up	AT2G29090	CYP707A2	CYP707A2 cytochrome P450, family 707, subfamily A, polypeptide 2
QSDrought_02886	Down	AT5G36110	CYP716A1	CYP716A1 cytochrome P450, family 716, subfamily A, polypeptide 1
QSDrought_06155	Down	AT5G07990	CYP75B1	CYP75B1 D501 TT7 Cytochrome P450 superfamily protein
QSDrought_09220	Up	AT5G42650	AOS	AOS CYP74A DDE2 allene oxide synthase
QSDrought_09914	Down	AT3G26330	CYP71B37	CYP71B37 cytochrome P450, family 71, subfamily B, polypeptide 37
QSDrought_10631	Down	AT3G14610	CYP72A7	CYP72A7 cytochrome P450, family 72, subfamily A, polypeptide 7
QSDrought_16485	Up	AT4G37360	CYP81D2	CYP81D2 cytochrome P450, family 81, subfamily D, polypeptide 2
QSDrought_19848	Down	AT5G06905	CYP712A2	CYP712A2 cytochrome P450, family 712, subfamily A, polypeptide 2
QSDrought_19940	Down	AT5G06900	CYP93D1	CYP93D1 cytochrome P450, family 93, subfamily D, polypeptide 1
Sugar:hydrogen symporter activity				
QSDrought_00469	Up	AT2G43330	ATINT1	ATINT1 INT1 inositol transporter 1
QSDrought_00694	Up	AT3G18830	ATPLT5	ATPLT5 ATPMT5 PMT5 polyol/monosaccharide transporter 5
QSDrought_01264	Down	AT1G11260	STP1	ATSTP1 STP1 sugar transporter 1
QSDrought_02773	Up	AT1G30220	ATINT2	ATINT2 INT2 inositol transporter 2
QSDrought_04922	Up	AT5G26340	MSS1	ATSTP13 MSS1 STP13 Major facilitator superfamily protein
QSDrought_07023	Down	AT1G22710	SUC2	ATSUC2 zSUT1 sucrose-proton symporter 2

Subsequently, Blast2Go was used to classify genes according to Gene Ontology (GO) terms (Figure 3.7C, D). In support of our differential expression analysis, foremost *GO Biological Process* terms included metabolic processes and response to stimulus, while significant *GO Molecular Function* terms included nucleic acid binding TF activity and transporter activity. Interestingly, a series of genes normally considered to have housekeeping functions could be observed, including ribosomal-, actin- and tubulin-gene family members (Appendix 1). Since genes belonging these categories have already been used as constitutive expressors in *Q. suber* (Marum et al., 2012; Rocheta et al., 2014; Sebastiana et al., 2014; Teixeira et al., 2014), the present and ongoing NGS-based transcriptional studies will be important for a future revision of constitutive expressor genes.

3.5. Functional data mining of DEGs

To obtain functional data from *Q. suber* DEGs, we identified their orthologs in the model plant *Arabidopsis thaliana* (herein designated DEG orthologs), using BLASTp homology search against the Arabidopsis TAIR 10 genome annotation (Figure 3.2). Of the initial 546 *Q. suber* unigenes, 41 failed to annotate to any known sequence, and 15 did not annotate to Arabidopsis protein models, resulting in 490 blast hits. We also observed redundancy within these hits, leading to the identification of 454 unique DEG orthologs (Figure 3.2, Appendix 1). Functional data mining was subsequently performed by means of gene network analysis, using the Genemania plugin at Cytoscape (Cline et al., 2007; Warde-Farley et al., 2010; Saito et al., 2012). First, a functional network was created to represent known genetic interactions, protein interactions and protein structural features (Figure 3.8; Figure 3.9A). In the analysis we uncovered a central gene cluster that was enriched in water responsive genes (the most over-represented GO category in the Genemania analysis), including various ABA-signalling-related genes such as phosphatases of the PP2C family (highlighted) found to regulate components within the ABA-signalling pathway (Saez et al., 2004; Zhang et al., 2013) but also regulating other genes associated with response to low Ψ_w (Bhaskara et al., 2012).

A

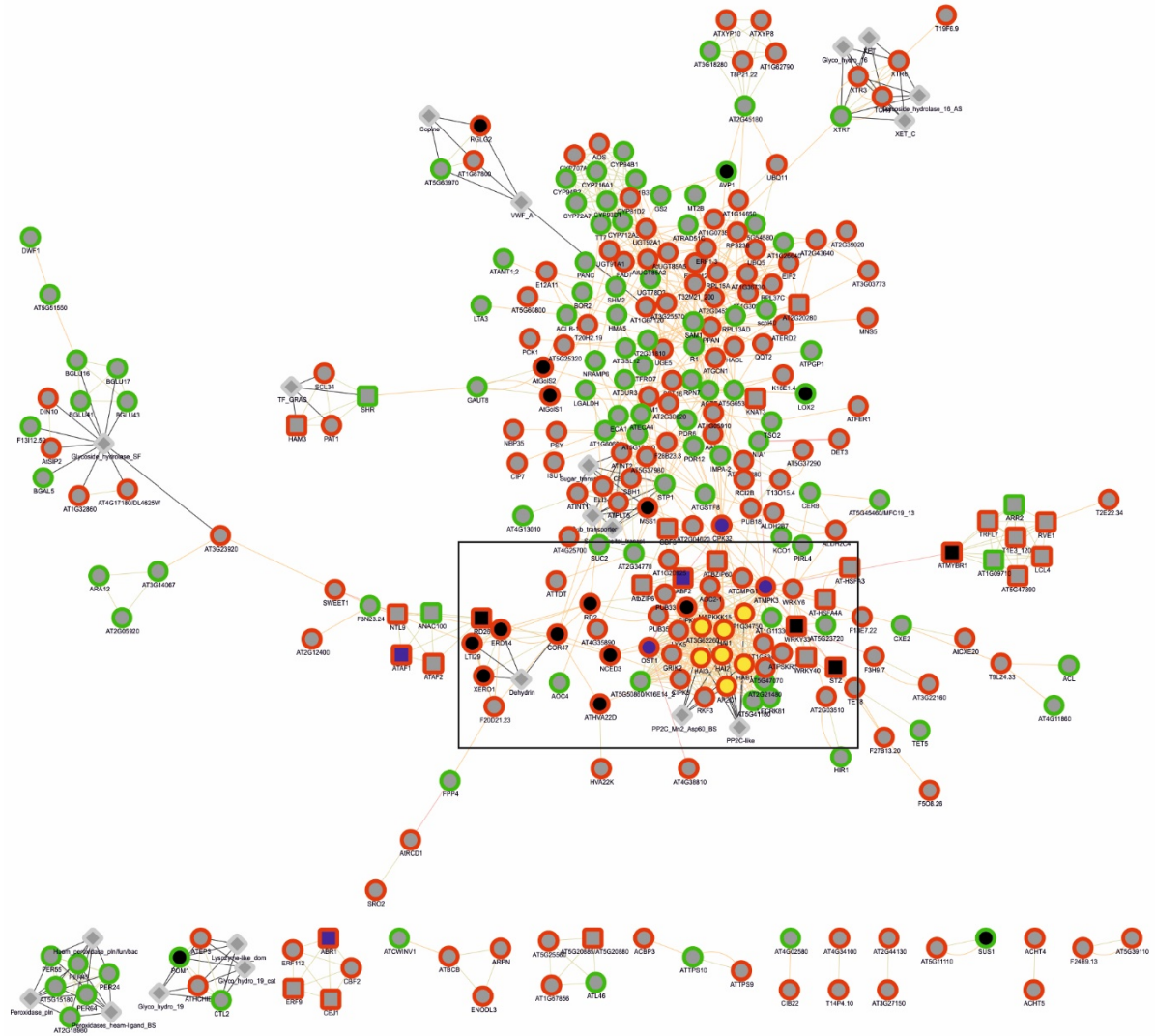
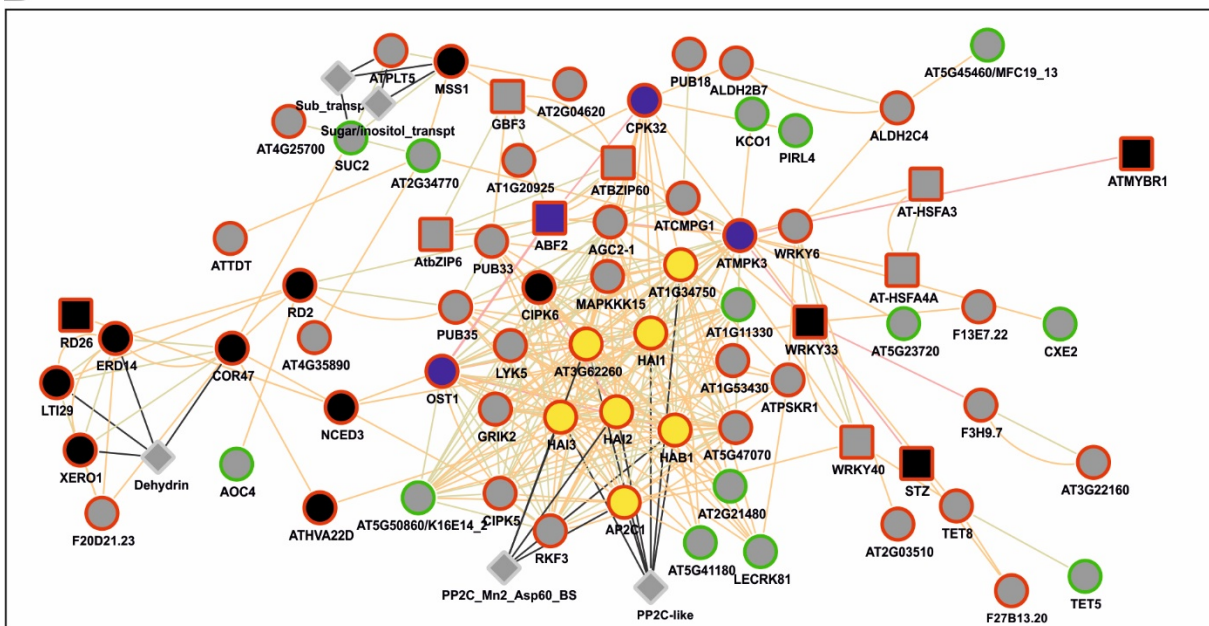
**B**

Figure 3.8. Gene network analysis of Arabidopsis DEG orthologs regarding functional features.

(A) Functional network, performed in the Genemania App of Cytoscape, selecting attributes, genetic interactions, physical interactions, predicted protein interactions and shared protein domains.

(B) Highlight of the main cluster of the previous functional feature analysis. Highlighted core gene cluster refers to three most over-represent Gene Ontology co-annotation categories (Called “Gene Ontology-based weighting”).

Genes are generally represented as grey circles; attributes are represented as grey diamonds; genes automatically annotated as belonging to the GO category Response to water are represented as black circles, and within these, those belonging to ABA-related subcategories are represented by blue circles; genes automatically annotated as belonging to the GO category Protein serine/threonine phosphatase activity are represented as yellow circles. Borders indicate up-regulated (red) and down-regulated (green) genes. Interactions are represented as edges (edge thickness represents normalized link weights) and edge colour indicates physical interaction (magenta), predicted interaction (orange), shared protein domain (light green) and InterPro protein classification (grey).

A second network was generated to represent cellular co-localization (Figure 3.9B). Analysis revealed a major gene cluster associated with protein targeting to the nucleus, suggesting once more the existence of extensive transcriptional regulatory events. Based on this evidence, and considering that genes with identical expression patterns will be controlled by the same TFs and share common *cis*-elements in their promoters, we identified statistically over-represented *cis*-elements in the promoters of DEG orthologues (Table 3.3). We could observe that a significant number of over-represented TF-binding site motifs belonged to the conserved ABA-responsive *cis* element (ABRE) and ABRE-like motifs. These elements are the binding site of *ABRE-binding factor* (ABF) TFs that have been previously associated to promoter regions of ABA-regulated, drought-responsive genes (Yoshida et al., 2010). All DEG orthologs that displayed ABRE and/or ABRE-like motifs were considered for co-expression network prediction using Cytoscape, and as expected, a complex gene co-expression network could be observed, indicating the existence of an ABA-mediated ABF-dependent regulon (Figure 3.9B).

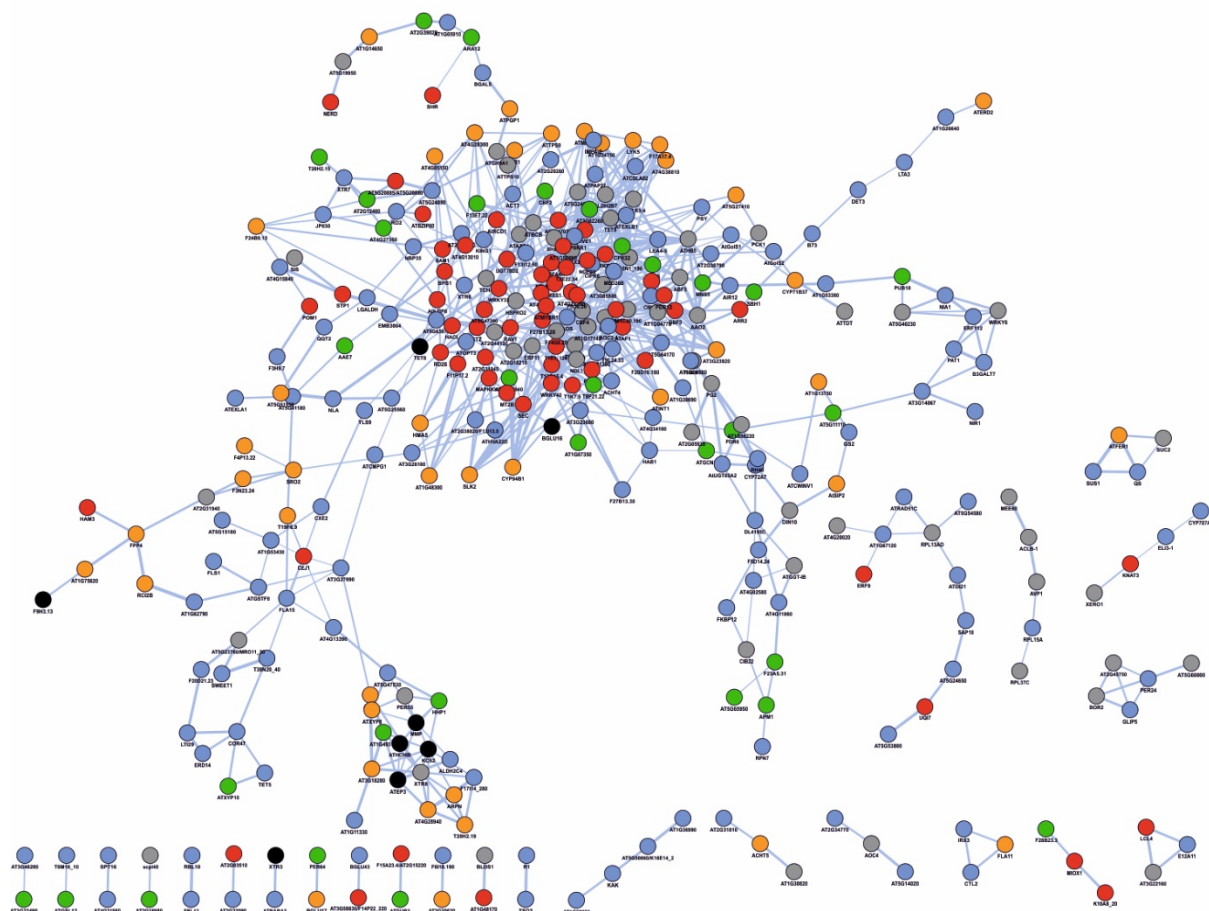
Table 3.3. - *Cis*-elements over-represented in the promoter region of DEG orthologs. The subset of up- and down-regulated genes were submitted to AtCOECis (Vandepoele et al., 2009) and Athena (O'Connor et al., 2005).

Cis Element	Motif	p-value (AtCOECis)	p-value (Athena)
<i>Up-regulated genes</i>			
ABFs binding site motif	CACGTGGC	2.17E-03	< 10E-4
ABRE binding site motif	BACGTGKM	2.57E-04	< 10E-4
ABRE-like binding site motif	YACGTGGC	9.75E-05	< 10E-10
ABREATRD22	RYACGTGGYR	3.22E-04	< 10E-3
ACGTABREMOTIFA2OSEM	ACGTGKC		< 10E-10

Table 3.3 (Continued)

CACGTGMOTIF	CACGTG		< 10E-7
DRE core motif	GACCGACTA	1.80E-03	< 10E-4
DREB1A/CBF3	RCCGACNT		< 10E-4
G-box plus G	CACGTGG	1.30E-05	
GADOWNAT	ACGTGTC		< 10E-8
GAREAT	TAACAAR		< 10E-3
GBOXLERBCS	MCACGTGGC		< 10E-3
LRE	ACGTGGCA	7.00E-04	
NCS-motif	Various	1.52E-07	
TATA-box Motif	TATAAA		< 10E-4
TGA1 binding site motif	TGACGTGG		< 10E-3
UPRMOTIFIAT	CCACGTCA		< 10E-3
<i>Down-regulated genes</i>			
MYB binding site promoter	CACCTAAC	1.09E-03	
NCS-motif	Various	<2,64E-06	

A



B

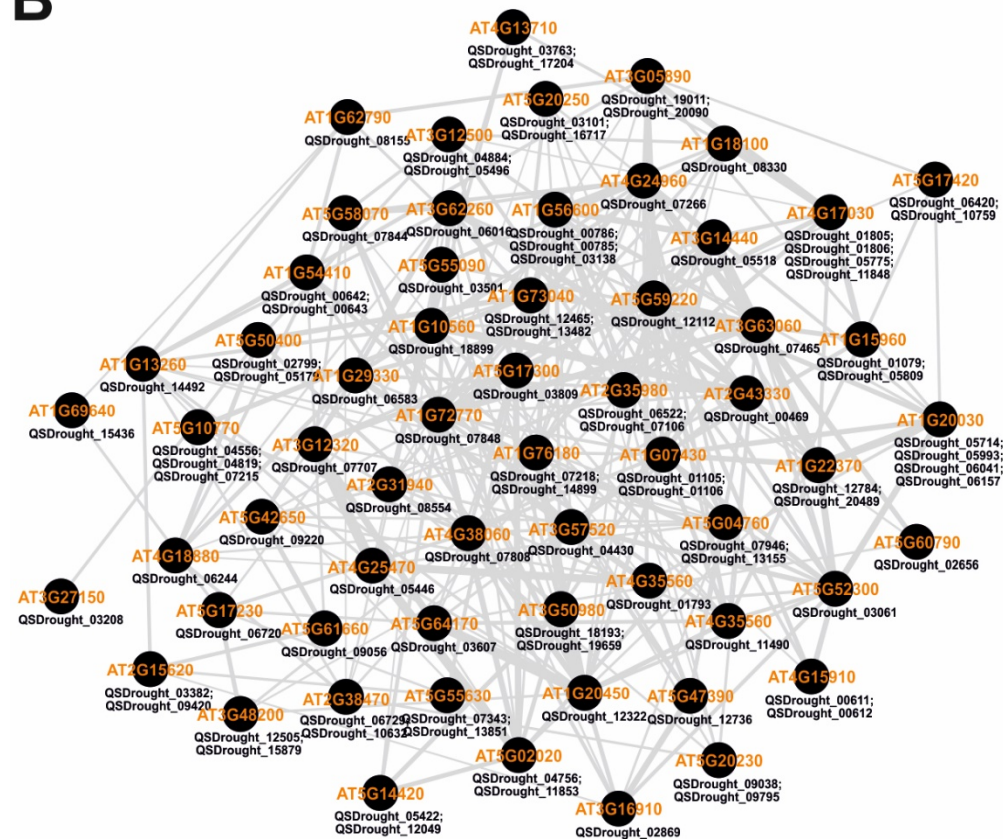


Figure 3.9. Additional gene network analysis, performed in the Genemania App of Cytoscape, of Arabidopsis DEG orthologs.

(B) Predicted protein co-localization analysis. Edge thickness represents normalized co-localization strength of the gene pair. Genes are represented by colour code, reflecting predicted sub-cellular targeting of their coding proteins, based on GO *Cellular Component* categorization towards the nucleus (red), cytoplasm (blue), plasma membrane (orange), organelle (green), extracellular region (black) and unknown (grey).

(C) Co-expression network of DEGs containing ABRE and ABRE-like motifs. Genes are represented as black circles. Edge thickness represents normalized gene co-expression strength. Arabidopsis AGI gene codes are represented (orange), as well as their orthologs from the *Q. suber* DEG unigene set.

Next we performed an automated search for TFs within DEG orthologs, and observed that drought imposition transcriptionally regulated a huge variety of TFs (Table 3.4). The most over-represented classes included NAC, AP2/ERF, WRKY and MYB TFs. The existence of various TF classes in the differential transcriptome of *Q. suber* roots responding to drought is extremely noteworthy. NAC, AP2/ERF, WRKY and MYB TFs have all been functionally associated with drought stress response mechanisms (Wang et al., 2003; Mahajan and Tuteja, 2005; Shao et al., 2008). For instance, in *Arabidopsis thaliana*, loss- or gain-of-function mutants for *ATAF1* (Lu et al., 2007), *MYB44* (Persak and Pitzschke, 2014), *ABR1* (Pandey et al., 2005) and *WRKY40* (Chen et al., 2010; Wang et al., 2013) have been established as drought/osmotic stress determinants, regulating signalling pathways, ROS scavenging, cell wall remodelling and stomatal aperture. To validate blast homology searches and the existence of drought-related *Q. suber* TFs, these were phylogenetically resolved against the totality of annotated Arabidopsis TFs, in each of the four most represented TF classes (Figs 3.10, 3.11, 3.12 and 3.13). Results allowed us to identify the closest Arabidopsis orthologs to our *Q. suber* genes-of-interest, which is important for two reasons. On the one hand, identification of Arabidopsis orthologs with known functional association with drought resistance can be used to functionally validate *Q. suber* genes as novel cork oak determinants of root level drought stress resistance, namely by complementation of Arabidopsis functional mutants with *Q. suber* orthologs. Secondly, phylogenetic mapping of *Q. suber* drought-responsive genes can be used in exploratory analysis to identify novel drought stress determinants in Arabidopsis, which can then be validated by subsequent functional mutant approaches in this model species. For instance, QSDrought_01793, a MYB TF, is the ortholog of AT5G67300 and is likely to be its functional equivalent. On the other hand, within the same MYB TF family QSDrought_16911 suggests that the functionally unresolved AT1G26580 is possibly a drought stress determinant.

Chapter 3. RESULTS AND DISCUSSION

Table 3.4. - List of transcription factors (TFs) present within DEG orthologs. For identification purposes, genes were submitted to AtTFDB search at the AGRIS platform (Davuluri et al., 2003) and MapMan (transcription pathway bin) (Thimm et al., 2004; Usadel et al., 2005; Usadel et al., 2009).

TF Family Name	AGI Code	Gene Name, Synonym	QS ID	Differential expression
Alfin-Like AP2- EREBP	AT5G20510	AL5	QSDrought_11056	Down-regulated
	AT2G33710		QSDrought_04959	Up-regulated
	AT3G50260	ATERF11, CEJ1, RAP2.1	QSDrought_10869	Up-regulated
	AT4G25470	CBF2, DREB1C, FTQ4	QSDrought_05446	Up-regulated
	AT5G44210	ERF9	QSDrought_09244	Up-regulated
	AT5G64750	ABR1	QSDrought_13237	Up-regulated
ARR-B	AT4G16110	ARP5, ARR2	QSDrought_08994	Down-regulated
bHLH	AT1G35460		QSDrought_18928	Down-regulated
bZIP	AT1G42990	bZIP60	QSDrought_12067	Up-regulated
	AT1G45249	ABF2, AREB1	QSDrought_02295	Up-regulated
	AT2G22850	bZIP6	QSDrought_06726	Up-regulated
	AT2G46270	GBF3	QSDrought_20020	Up-regulated
C2H2	AT1G10480	ZFP5	QSDrought_06423	Up-regulated
	AT1G27730	STZ, ZAT10	QSDrought_01447	Up-regulated
	AT2G27100	SE	QSDrought_07992	Up-regulated
C3H	AT5G20885		QSDrought_17648	Up-regulated
GRAS	AT4G00150	SCL6	QSDrought_09538	Up-regulated
	AT5G48150	PAT1	QSDrought_16917	Up-regulated
	AT4G37650	SGR7, SHR	QSDrought_03719	Down-regulated
Homeobox	AT5G25220	KNAT3	QSDrought_04401	Up-regulated
HSF	AT4G18880	HSFA4A, HSF21	QSDrought_06244	Up-regulated
	AT5G03720	HSFA3	QSDrought_07727	Up-regulated
MYB	AT5G67300	MYB44, MYBR1	QSDrought_01793	Up-regulated
	AT4G12350	MYB42	QSDrought_06966	Up-regulated
	AT1G26580		QSDrought_16911	Up-regulated
	AT3G60460	DUO1	QSDrought_17304	Down-regulated
MYB-Like	AT4G01280		QSDrought_13589	Up-regulated
	AT5G04760		QSDrought_07946	Up-regulated
	AT5G17300		QSDrought_03809	Up-regulated
	AT5G47390		QSDrought_12736	Up-regulated
NAC	AT1G01720	ANAC002, ATAF1	QSDrought_04760	Up-regulated
	AT4G27410	ANAC072, RD26	QSDrought_03374	Up-regulated
	AT4G35580	NTL9	QSDrought_00747	Up-regulated
	AT5G08790	ANAC081, ATAF2	QSDrought_06066	Up-regulated
	AT5G61430	ANAC100, ATNAC5	QSDrought_06406	Down-regulated

Table 3.4 (Continued)

RAV	AT1G13260	RAV1	QSDrought_14492	Up-regulated
WRKY	AT1G62300	WRKY6	QSDrought_05433	Up-regulated
	AT1G80840	WRKY40	QSDrought_03742	Up-regulated
	AT2G38470	WRKY33	QSDrought_06729	Up-regulated
ZF(CCCH-type)	AT2G20280		QSDrought_12281	Up-regulated

NAC

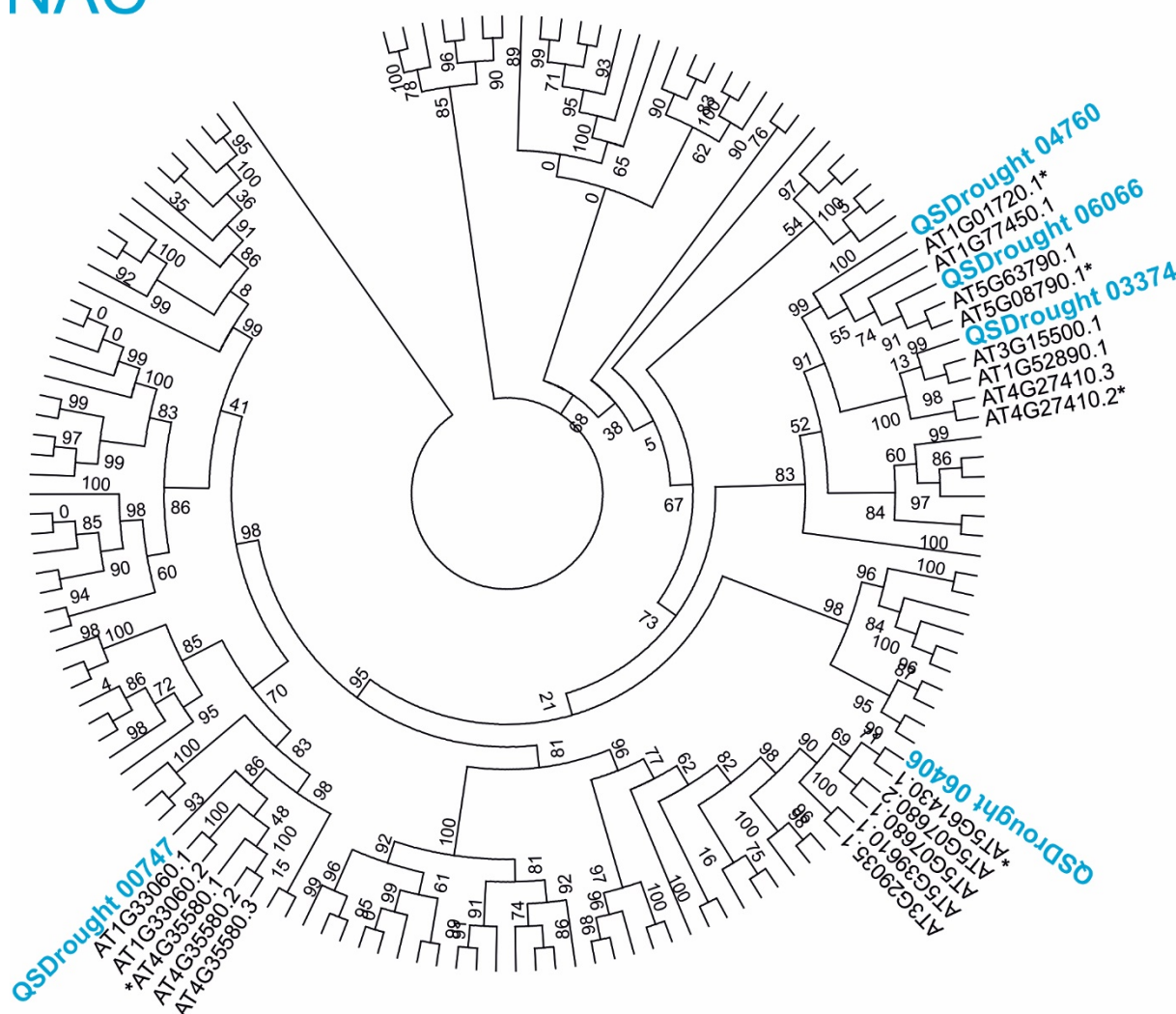


Figure 3.10. Phylogenetic reconstruction of NAC TF class, containing *Q. suber* differentially expressed unigenes, and the complete annotated set of *Arabidopsis thaliana* NAC gene family members. The first BLAST hit from the ortholog identification step is marked with an asterisk.

AP2-ERF

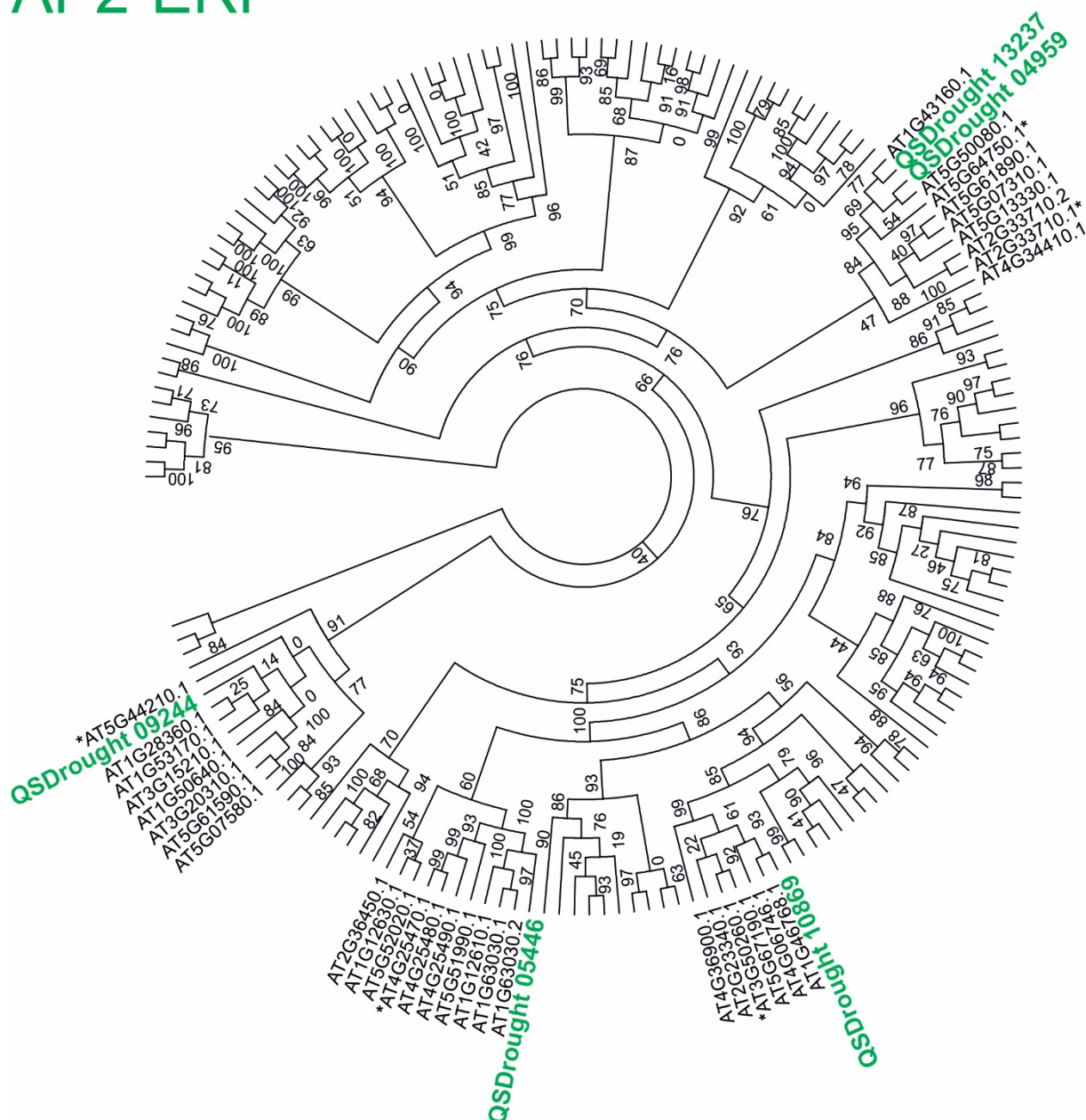


Figure 3.11. Phylogenetic reconstruction of AP2-ERF TF class, containing *Q. suber* differentially expressed unigenes, and the complete annotated set of Arabidopsis thaliana AP2-ERF superfamily members. The first BLAST hit from the ortholog identification step is marked with an asterisk.

MYB

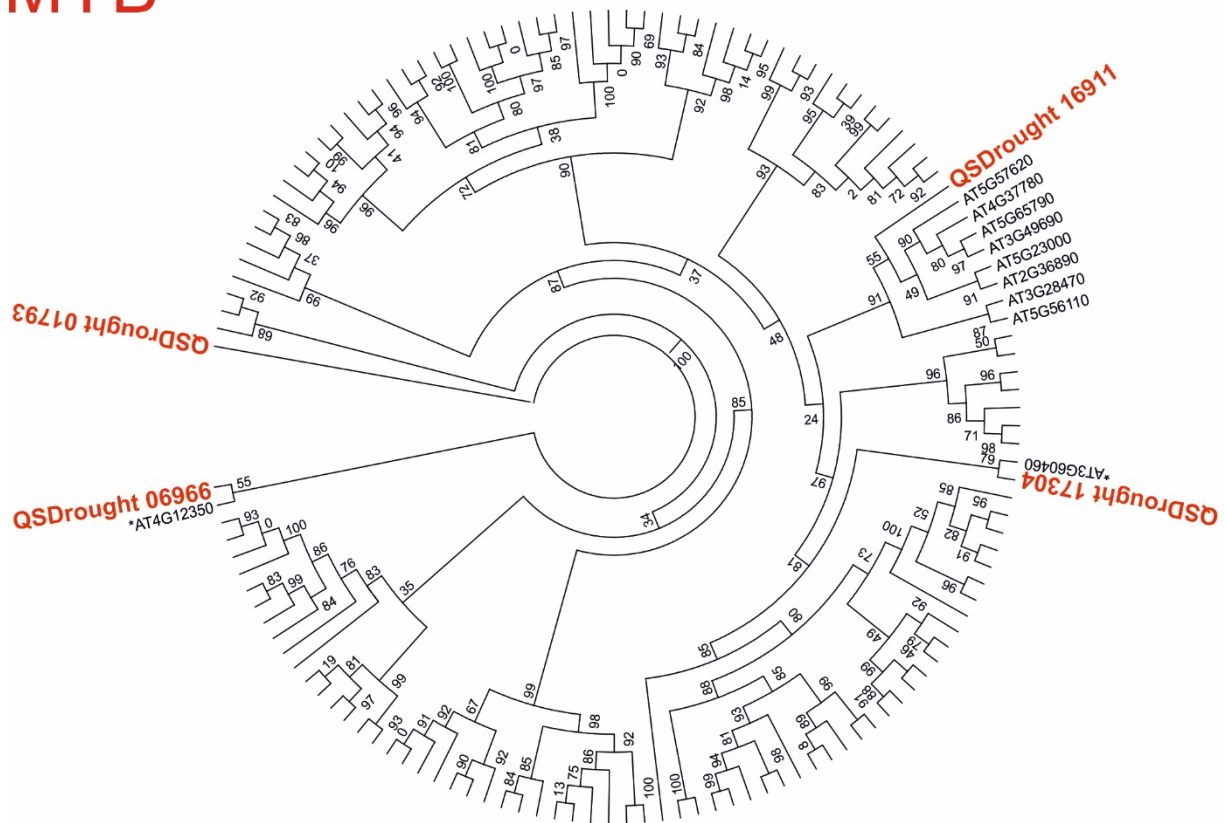
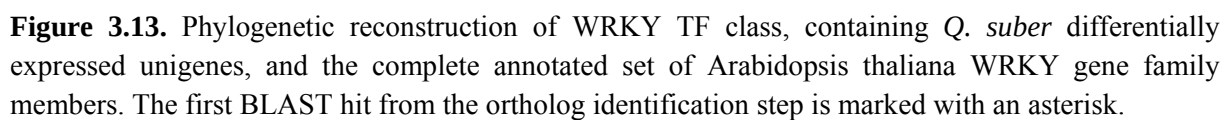


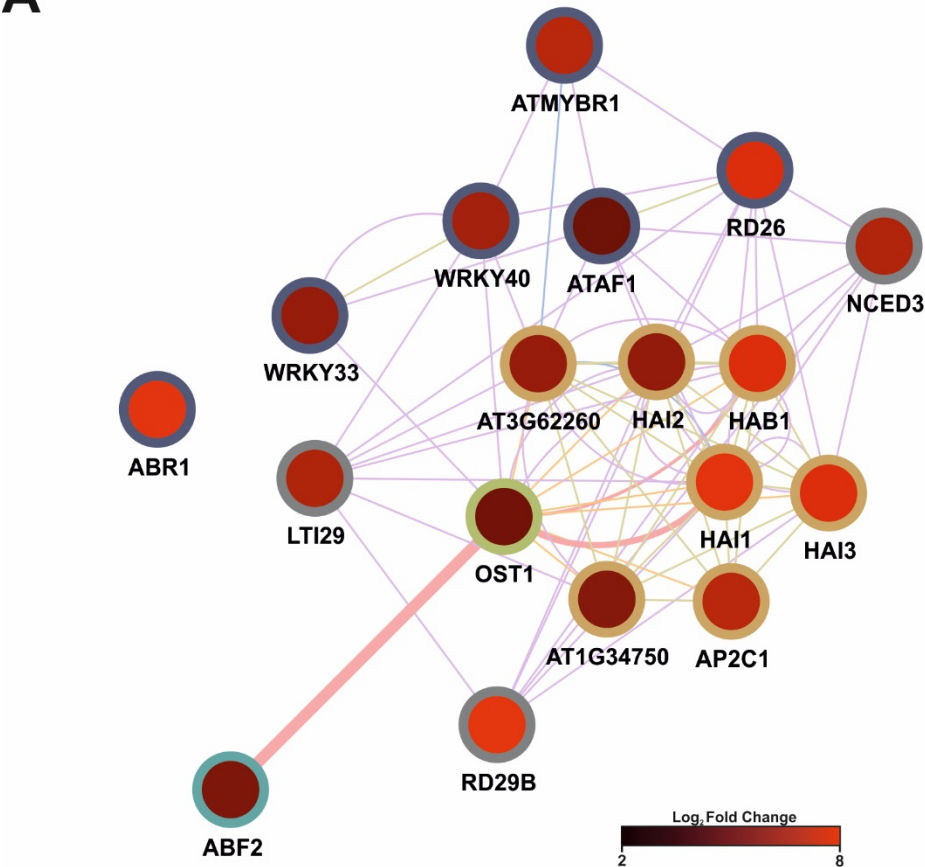
Figure 3.12. Phylogenetic reconstruction of MYB TF class, containing *Q. suber* differentially expressed unigenes, and the complete annotated set of *Arabidopsis thaliana* MYB gene family members. The first BLAST hit from the ortholog identification step is marked with an asterisk.



In plants, drought stress responses have been functionally mapped to four major signalling pathways that involve both ABA-dependent and ABA-independent pathways (Shinozaki and Yamaguchi-Shinozaki, 2007; Kuromori et al., 2014). Present results strongly supported the involvement of an ABA-dependent signalling network in the *Q. suber* root-level response to drought, given that (1) analysis of GO category enrichment highlighted the up-regulation of known ABA-signalling components, (2) *cis*-element analysis proposed an enrichment in *ABRE* and *ABRE-like* motifs, (3) several ABA-related TFs and response marker genes were identified. Present evidence specifically suggested the existence of *ABF*-related

signalling events, therefore we performed an intersection network, merging the DEG functional network and components of the *ABF*-dependent signal transduction pathway (Figure 3.14A). Surprisingly, we uncovered evidence for the activation of the complete canonical pathway in long-term drought-stressed *Q. suber* roots. ABA levels can be sensed by a series of different cellular receptors (Golldack et al., 2014). In the most well established case, PYR/PYL/RCAR nucleocytoplasmic receptors bind ABA and inhibit PP2Cs, preventing them from negatively regulating SnRK2 kinases. SnRK2s are then free to phosphorylate and activate key components that include ABF TFs, thus modulating the activity of a series of ABA-responsive genes (Figure 3.14) (Christmann et al., 2006; Hubbard et al., 2010; Raghavendra et al., 2010; Nakashima and Yamaguchi-Shinozaki, 2013; Golldack et al., 2014).

A



B

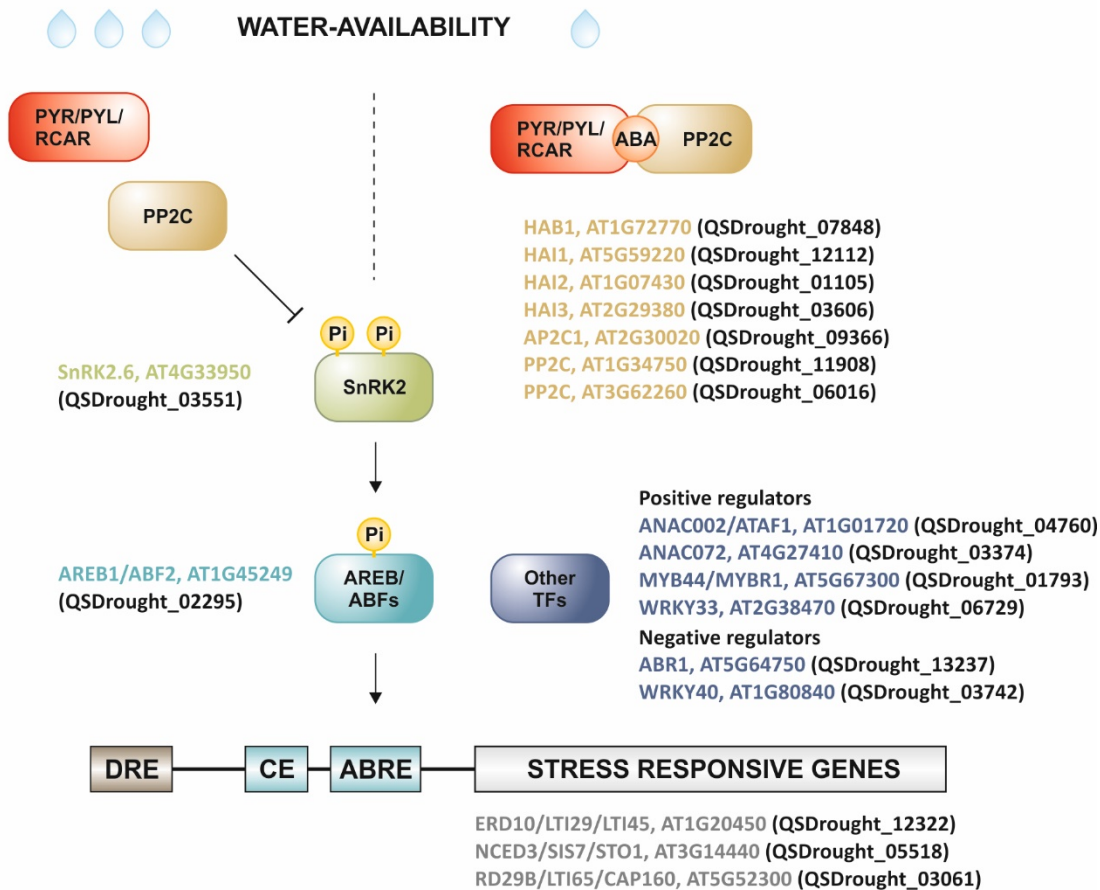


Figure 3.14. Model for the regulation of long-term drought responses in *Q. suber* roots: *ABF*-dependent signalling pathway components are up-regulated in *Q. suber* roots in response to drought.

(A) Functional network based on Arabidopsis ABA-dependent signalling components (Miyakawa et al., 2013; Nakashima and Yamaguchi-Shinozaki, 2013) merged with the complete DEG functional network, comprising different functional features. Edge thickness represents normalized link weights and edge colour indicates physical interaction (magenta), predicted interaction (orange), shared protein domain (light green) and co-expression (purple). Node colour represents expression in Log₂ of fold change, while border colour represents different classes of signalling components.

(B) Schematic model of the proposed ABF-dependent signalling cascade: colours refer to different Arabidopsis signalling cascade components (gene name plus AGI code), while brackets highlight matching Quercus ID unigenes found in the present analysis to be differentially expressed. Low water availability induces ABA production which is recognized by the receptors PYR/PYL/RCAR. PYR/PYL/RCAR structural conformation changes after binding to ABA, thus recruiting Protein Phosphatase 2C (PP2C) phosphatases. PP2Cs are unavailable to dephosphorylate SnRK2 kinases. Phosphorylated (active) SnRK2s activates AREB/ABF transcription factors (TFs), which recognize ABRE cis-element motifs, mediating induction of stress-related gene expression (e.g. RD29b and NCED3). Other TF families are likely to play a role in ABA-signalling such as ANAC, AP2/EREB, MYB and WRKYs

In the present report we observed that, with the exception of the PYR/PYL/RCAR sensor, all components could be found within *Q. suber* DEGs (and their Arabidopsis orthologs), and all appeared as up-regulated upon stress imposition. One *Q. suber* PYR/PYL/RCAR gene was found in the assembled *Q. suber* transcriptome, suggesting that the sensor is constitutively expressed. The most observed enrichment related to the PP2C gene family, with 7 up-regulated genes (Table 3.2, Figure 3.8). *Q. suber* PP2C were phylogenetically resolved against the Arabidopsis PP2C gene family (Fig 3.15). Components included orthologs of Arabidopsis HAB1, a negative regulator of ABA signalling pathway (Saez et al., 2004), HAI1, also associated with negative regulation of ABA signalling pathway (Zhang et al., 2013), HAI2 and HAI3, that together with HAI1 have recently been associated with negative regulation of osmotic adjustment-related genes (Bhaskara et al., 2012). Ortholog PP2Cs also include AP2C1, reported as being both a negative regulator of MAPK signalling proteins, and a regulator of ABA-related genes and signalling components (Schweighofer et al., 2007; Brock et al., 2010). Positive components of the pathway (*SnRK2* and *ABF* elements) were also observed and found to be up-regulated, possibly contributing for the transcriptional accumulation of drought stress marker genes such as *RD29b* and *NCED3*, both of which respond to ABF by containing *ABRE* cis-elements (Figure 3.14) (Nakashima et al., 2006; Behnam et al., 2013). NCED3 is a component of the ABA biosynthetic pathway (Iuchi et al., 2001), and may thus contribute to signal amplification via a positive feedback loop.

PP2C

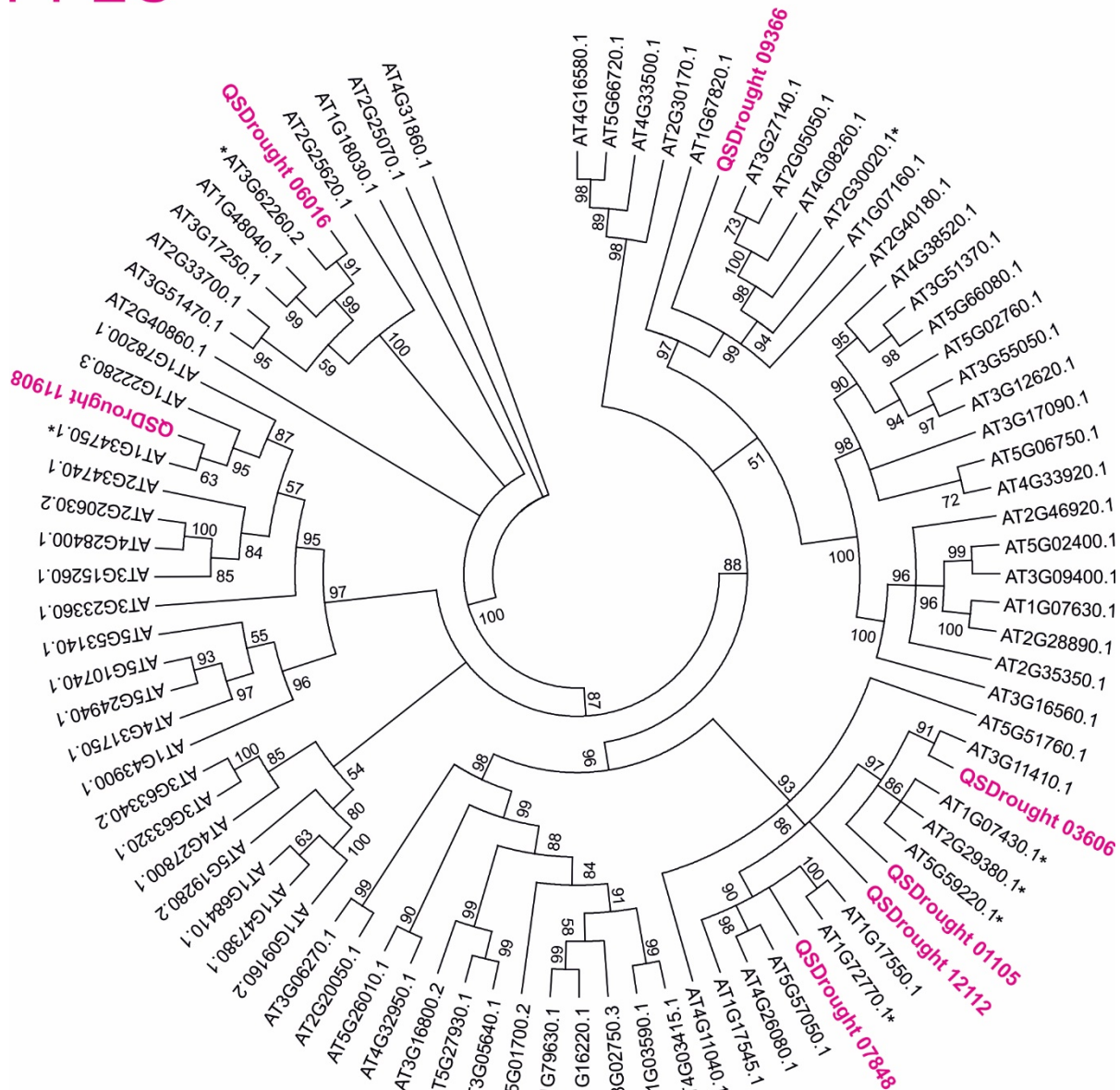


Figure 3.15. Phylogenetic reconstruction of PP2C identified in the *Q. suber* differentially expressed unigenes and the complete annotated set of *Arabidopsis thaliana* genes from the same class. Asterisks represent first hit in the ortholog identification step.

Additional signalling pathways may be involved in the *Q. suber* root response to long-term drought, given the extension of TFs and genes involved in the overall transcriptional response to drought and the capacity of this organism to sustain long periods of water deficit. Figure 3.8 acknowledges the existence of additional regulatory elements of the *Q. suber* differential transcriptome, namely TF that have been previously highlighted (Table 3.4), and that act as both positive and negative determinants of drought-related effector genes (Fujita et

al., 2004; Pandey et al., 2005; Lu et al., 2007; Yoshida et al., 2010; Nakashima and Yamaguchi-Shinozaki, 2013; Wang et al., 2013; Persak and Pitzschke, 2014).

Quantitative Real-Time PCR (qPCR) was used to validate the current RNA-Seq data, particularly the induction of the above mentioned ABA-dependent ABF signalling pathway, and therefore selected genes included genes for all components of the pathway (Figure 3.16). All ten genes closely matched the relative expression values calculated via RNA-Seq analysis, thus supporting the current strategy and the existence of an ABA-dependent signalling network that could have a major role in the *Q. suber* response to drought.

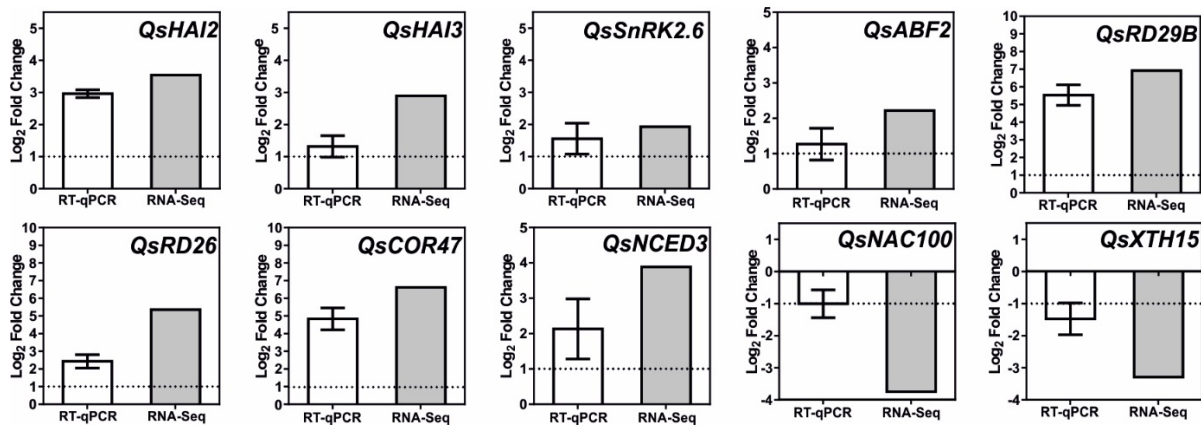


Figure 3.16. Real-time RT-PCR analysis of selected *Q. suber* drought stress responsive genes. White bar represents changes in mRNA levels of root tissues under severe drought stress (D25/D10), in comparison to control samples (D100/90). Expression was normalized using *QsPP2A-3* as a reference gene. Results represent mean values $\pm$ SD ($n = 3$) and are expressed as Log₂ values of the fold change. Grey bar represents matching relative expression values, derived from RNA-Seq analysis of the D100/90 vs D25/10 condition.

Chapter 4

CONCLUDING REMARKS AND FUTURE PERSPECTIVES

The Mediterranean basin is a unique and rich environment facing rapid climate changes due to extreme weather and ever changing conditions. The cork oak forest is an excellent example of a multi-functional forest, which provides many ecological functions and important economic resources, facing such environmental changes. Understanding how cork oak responds to environmental stress is critical from a conservation step and could possibly expand the knowledge on the subject, revealing novel mechanisms of response. Being a relatively new target for research, *Quercus suber* has the implication of not having a lot of resources for studying the species at the molecular level.

Next generation sequencing (NGS) is now an established platform for research in non-model organisms like cork oak, bringing the possibility of high throughput data generation and integration of such data with orthogonal data sets, further accelerating analysis and comprehension of the underlying question that is being asked. RNA-seq in particular has been widely implemented as a method to drastically reduce costs of research and enable rapid and accurate transcriptional analysis without the need to have genome resources available for the organism.

RNA-seq was the experimental approach implemented in this work to uncover the transcriptional response to drought in *Q. suber* roots. First, plantlets were subjected to different drought conditions and characterised based on their physiological responses. Having established that plantlets displayed physiological signs associated with drought stress, RNA was extracted and sequenced by Roche's 454 technology. After receiving sequencing libraries, a standard pipeline was implemented adding a de novo assembly step to generate a reference transcriptome. Assembly metrics, coverage analysis, differential analysis characterization and functional annotation of DEGs suggest that the methods applied to the sequencing libraries were suited for the objectives of this study. Further characterization of the DEG library was conducted through ortholog discovery for the *Arabidopsis thaliana* plant model, enabling in-depth analysis and integration with major resources and data bases, suitable for data mining. Results highlighted that common transcription factors and effector genes, which are involved in drought response, are regulating the response of cork oak during drought. Moreover the identification of the ABA signalling pathway revealed the underlying regulation pathway in roots for that species and established a regulation network governing responses to drought at the root level. Although all the components were identified and qPCR results confirmed this major result, other response pathways could be involved and there could even be crosstalk events between pathways.

Chapter 4. CONCLUDING REMARKS AND FUTURE PERSPECTIVES

To expand knowledge on the subject, future analysis is critical. Exploratory analysis of novel drought stress determinants in *Arabidopsis thaliana* with loss- and gain-of-function mutants could reveal new insights into previously uncharacterized genes. Moreover the functional characterization of *Q. suber* genes in *A. thaliana* could help define new functions to characterized genes and uncover novel characteristics to the *Q. suber* genes, and establish the molecular and evolutionary path of this species.

From a productivity point of view, research into the molecular response in leaf is critical. Stress determinants and pathway components were identified throughout this study, and could be screened at the transcriptional level. The qPCR protocol could be implemented in leaf, to determine if a similar pathway is regulating the response in that tissue. For a more in-depth study, Illumina-based RNA sequencing should be implemented. Having already established a reference transcriptome, de novo assembly efforts will be computationally less demanding and enable the use of hybrid approaches to assembly and mapping (Ward et al., 2012) with Illumina HiSeq sequencing data, without the need to compromise in throughput and resolution in favour of read length.

The prospect of a sequenced genome for this species brings new possibilities for future work. Integrating the present data with genomics annotation can lead to better annotation resources for the project and mapping unknown sequences or unmapped reads could identify novel genes in *Q. suber*. It also opens the way for RNA-seq analysis based in genome mapping. This enables more in-depth studies like SNP-profiling and splice variance abundance analysis. More tantalising even is the prospect that genome data is used to facilitate the study of other levels of regulation. Epigenetic analysis, for example, is highly dependent on genome data for mapping steps. Integration on multiple layers of OMICs data can give rise to a deeper understanding of the species and the major regulation of stress responses, which are critical for the conservation of this economically significant species.

Chapter 5

BIBLIOGRAPHY

- Ahmad P, Prasad MNV** (2011) Abiotic stress responses in plants: metabolism, productivity and sustainability. Springer Science & Business Media
- Almeida T, Menendez E, Capote T, Ribeiro T, Santos C, Goncalves S** (2013) 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** (2013) QsMYB1 expression is modulated in response to heat and drought stresses and during plant recovery in *Quercus suber*. *Plant Physiol Biochem* **73**: 274-281
- Alvarez R, Alonso P, Cortizo M, Celestino C, Hernandez I, Toribio M, Ordas RJ** (2004) Genetic transformation of selected mature cork oak (*Quercus suber* L.) trees. *Plant Cell Rep* **23**: 218-223
- Anders S, Huber W** (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**: R106
- Anisimova M, Gascuel O** (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol* **55**: 539-552
- Atkinson NJ, Urwin PE** (2012) The interaction of plant biotic and abiotic stresses: from genes to the field. *J Exp Bot* **63**: 3523-3543
- Azevedo H, Lino-Neto T, Tavares RM** (2003) An improved method for high-quality RNA isolation from needles of adult maritime pine trees. *Plant Molecular Biology Reporter* **21**: 333-338
- Behnam B, Iuchi S, Fujita M, Fujita Y, Takasaki H, Osakabe Y, Yamaguchi-Shinozaki K, Kobayashi M, Shinozaki K** (2013) Characterization of the promoter region of an Arabidopsis gene for 9-cis-epoxycarotenoid dioxygenase involved in dehydration-inducible transcription. *DNA Res* **20**: 315-324
- Belahbib N, Pemonge MH, Ouassou A, Sbay H, Kremer A, Petit RJ** (2001) Frequent cytoplasmic exchanges between oak species that are not closely related: *Quercus suber* and *Q. ilex* in Morocco. *Mol Ecol* **10**: 2003-2012
- Bhaskara GB, Nguyen TT, Verslues PE** (2012) Unique drought resistance functions of the highly ABA-induced clade A protein phosphatase 2Cs. *Plant Physiol* **160**: 379-395
- Bornet B, Branchard M** (2001) Nonanchored Inter Simple Sequence Repeat (ISSR) markers: Reproducible and specific tools for genome fingerprinting. *Plant Mol Biol Report* **19**: 209-215
- Brock AK, Willmann R, Kolb D, Grefen L, Lajunen HM, Bethke G, Lee J, Nürnberger T, Gust AA** (2010) The Arabidopsis Mitogen-Activated Protein Kinase Phosphatase PP2C5 Affects Seed Germination, Stomatal Aperture, and Abscissic Acid-Inducible Gene Expression. *Plant Physiol* **153**: 1098-1111
- Brunner AM, Yakovlev IA, Strauss SH** (2004) Validating internal controls for quantitative plant gene expression studies. *BMC Plant Biol* **4**: 14
- Burgarella C, Lorenzo Z, Jabbour-Zahab R, Lumaret R, Guichoux E, Petit RJ, Soto A, Gil L** (2009) Detection of hybrids in nature: application to oaks (*Quercus suber* and *Q. ilex*). *Heredity* **102**: 442-452
- Castro PH, Tavares RM, Bejarano ER, Azevedo H** (2012) SUMO, a heavyweight player in plant abiotic stress responses. *Cel Mol Life Sci* **69**: 3269-3283
- Catry FX, Moreira F, Pausas JG, Fernandes PM, Rego F, Cardillo E, Curt T** (2012) Cork oak vulnerability to fire: the role of bark harvesting, tree characteristics and abiotic factors. *PLoS One* **7**: e39810
- Chaves I, Passarinho JA, Capitao C, Chaves MM, Fevereiro P, Ricardo CP** (2011) Temperature stress effects in *Quercus suber* leaf metabolism. *Jf Plant Physiol* **168**: 1729-1734

- Chen H, Lai Z, Shi J, Xiao Y, Chen Z, Xu X** (2010) Roles of Arabidopsis WRKY18, WRKY40 and WRKY60 transcription factors in plant responses to abscisic acid and abiotic stress. *BMC Plant Biol* **10**: 281
- Chen L, Song Y, Li S, Zhang L, Zou C, Yu D** (2012) The role of WRKY transcription factors in plant abiotic stresses. *Biochim Biophys Acta* **1819**: 120-128
- Christmann A, Moes D, Himmelbach A, Yang Y, Tang Y, Grill E** (2006) Integration of Absciscic Acid Signalling into Plant Responses. *Plant Biol* **8**: 314-325
- Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang P-L, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD** (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat. Protocols* **2**: 2366-2382
- Coelho AC, Lima MB, Neves D, Cravador A** (2006) Genetic diversity of two evergreen oaks [*Quercus suber* (L.) and *Quercus ilex* subsp. *rotundifolia* (Lam.)] in Portugal using AFLP markers. *Silvae Genet* **55**: 105-118
- Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M** (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674-3676
- Czechowski T, Stitt M, Altmann T, Udvardi MK, Scheible W-R** (2005) Genome-wide identification and testing of superior reference genes for transcript normalization in Arabidopsis. *Plant Physiol* **139**: 5-17
- Danquah A, de Zelicourt A, Colcombet J, Hirt H** (2014) The role of ABA and MAPK signaling pathways in plant abiotic stress responses. *Biotechnology Adv* **32**: 40-52
- Davuluri RV, Sun H, Palaniswamy SK, Matthews N, Molina C, Kurtz M, Grotewold E** (2003) AGRIS: Arabidopsis gene regulatory information server, an information resource of Arabidopsis cis-regulatory elements and transcription factors. *BMC Bioinformatics* **4**: 25
- Doyle JJ** (1990) Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-15
- Edgar RC** (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**: 1792-1797
- Eisen MB, Spellman PT, Brown PO, Botstein D** (1998) Cluster analysis and display of genome-wide expression patterns. *PNAS* **95**: 14863-14868
- Epron D, Dreyer E, Aussenac G** (1993) A comparison of photosynthetic responses to water stress in seedlings from 3 oak species: *Quercus petraea* (Matt) Liebl, *Q. rubra* L and *Q. cerris* L. *Ann For Sci* **50**: 48s-60s
- Faria T, Garcia-Plazaola JI, Abadia A, Cerasoli S, Pereira JS, Chaves MM** (1996) Diurnal changes in photoprotective mechanisms in leaves of cork oak (*Quercus suber*) during summer. *Tree Physiol* **16**: 115-123
- Flexas J, Diaz-Espejo A, Gago J, Gallé A, Galmés J, Gulías J, Medrano H** (2014) Photosynthetic limitations in Mediterranean plants: A review. *Env Expl Bot* **103**: 12-23
- Fu L, Niu B, Zhu Z, Wu S, Li W** (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**: 3150-3152
- Fujita M, Fujita Y, Maruyama K, Seki M, Hiratsu K, Ohme-Takagi M, Tran LSP, Yamaguchi-Shinozaki K, Shinozaki K** (2004) A dehydration-induced NAC protein, RD26, is involved in a novel ABA-dependent stress-signaling pathway. *Plant J* **39**: 863-876
- Garg, R., & Jain, M.** (2013) RNA-Seq for transcriptome analysis in non-model plants. In *Legume Genomics* (pp. 43-58). Humana Press.

- Golldack D, Lüking I, Yang O** (2011) Plant tolerance to drought and salinity: stress regulating transcription factors and their functional significance in the cellular transcriptional network. *Plant Cell Rep* **30**: 1383-1391
- Golldack D, Li C, Mohan H, Probst N** (2014) Tolerance to drought and salt stress in plants: Unraveling the signaling networks. *Front Plant Sci* **5**: 151
- Gomez-Garay A, Lopez JA, Camafeita E, Bueno MA, Pintos B** (2013) Proteomic perspective of *Quercus suber* somatic embryogenesis. *J Proteomics* **93**: 314-325
- Gomez A, Lopez JA, Pintos B, Camafeita E, Bueno MA** (2009) Proteomic analysis from haploid and diploid embryos of *Quercus suber* L. identifies qualitative and quantitative differential expression patterns. *Proteomics* **9**: 4355-4367
- Gouy M, Guindon S, Gascuel O** (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biol Evol* **27**: 221-224
- Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O** (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307-321
- Gutierrez L, Mauriat M, Guénin S, Pelloux J, Lefebvre JF, Louvet R, Rusterucci C, Moritz T, Guerineau F, Bellini C** (2008) The lack of a systematic validation of reference genes: a serious pitfall undervalued in reverse transcription-polymerase chain reaction (RT-PCR) analysis in plants. *Plant Biotech J* **6**: 609-618
- Havaux M** (1992) Stress tolerance of photosystem II in vivo antagonistic effects of water, heat, and photoinhibition stresses. *Plant Physiol* **100**: 424-432
- Herrero J, Valencia A, Dopazo J** (2001) A hierarchical unsupervised growing neural network for clustering gene expression patterns. *Bioinformatics* **17**: 126-136
- Hong SM, Bahn SC, Lyu A, Jung HS, Ahn JH** (2010) Identification and testing of superior reference genes for a starting pool of transcript normalization in Arabidopsis. *Plant Cell Physiol* **51**: 1694-1706
- Huang Y, Niu B, Gao Y, Fu L, Li W** (2010) CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**: 680-682
- Hubbard KE, Nishimura N, Hitomi K, Getzoff ED, Schroeder JI** (2010) Early abscisic acid signal transduction mechanisms: newly discovered components and newly emerging questions. *Genes & Dev* **24**: 1695-1708
- Huber S, Anders S** Differential expression analysis for sequence count data. *Genome Biol* **11**: R106
- Ibarra-Caballero J, Villanueva-Verduzco C, Molina-Galán J, Sánchez-De-Jiménez E** (1988) Proline Accumulation as a Symptom of Drought Stress in Maize: A Tissue Differentiation Requirement. *J Exp Bot* **39**: 889-897
- ICNF** (2013) IFN6 – Áreas dos usos do solo e das espécies florestais de Portugal continental. Resultados preliminares. [pdf], 34 pp, Instituto da Conservação da Natureza e das Florestas. Lisboa.
- Iuchi S, Kobayashi M, Taji T, Naramoto M, Seki M, Kato T, Tabata S, Kakubari Y, Yamaguchi-Shinozaki K, Shinozaki K** (2001) Regulation of drought tolerance by gene manipulation of 9-cis-epoxycarotenoid dioxygenase, a key enzyme in abscisic acid biosynthesis in Arabidopsis. *Plant J* **27**: 325-333
- Khraiwesh B, Zhu J-K, Zhu J** (2012) Role of miRNAs and siRNAs in biotic and abiotic stress responses of plants. *Biochimica et Biophysica Acta (BBA) - Gene Reg Mech* **1819**: 137-148
- Koyro H-W, Ahmad P, Geissler N** (2012) Abiotic stress responses in plants: an overview. *In* Environmental Adaptations and Stress Tolerance of Plants in the Era of Climate Change. Springer, pp 1-28

- Kozlowski TT, Pallardy SG** (2002) Acclimation and adaptive responses of woody plants to environmental stresses. *Bot Rev* **68**: 270-334
- Kumar S, Blaxter ML** (2010) Comparing de novo assemblers for 454 transcriptome data. *BMC genomics* **11**: 571
- Kuromori T, Mizoi J, Umezawa T, Yamaguchi-Shinozaki K, Shinozaki K** (2014) Drought Stress Signaling Network. In SH Howell, ed, *Molecular Biology*, Vol 2. Springer New York, pp 383-409
- Lambers H, Chapin FS, III, Pons T** (2008) Plant Water Relations. In *Plant Physiological Ecology*. Springer New York, pp 163-223
- Lata C, Prasad M** (2011) Role of DREBs in regulation of abiotic stress responses in plants. *J Exp Bot* **62**: 4731-4748
- Le SQ, Gascuel O** (2008) An improved general amino acid replacement matrix. *Mol Biol Evol* **25**: 1307-1320
- Lee J-H, Kim W** (2011) Regulation of abiotic stress signal transduction by E3 ubiquitin ligases in Arabidopsis. *Mol Cells* **31**: 201-208
- Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J. A., Stewart, R., & Dewey, C. N.** (2014) Evaluation of de novo transcriptome assemblies from RNA-Seq data. *Genome Biol*, 15: 553.
- Li W, Godzik A** (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**: 1658-1659
- Lin Y, Lai Z** (2010) Reference gene selection for qPCR analysis during somatic embryogenesis in longan tree. *Plant Sci* **178**: 359-365
- Loomis W** (1974) Overcoming problems of phenolics and quinones in the isolation of plant enzymes and organelles. *Meth Enzymol* **31**: 528-544
- Lopes MH, Barros AS, Pascoal Neto C, Rutledge D, Delgadillo I, Gil AM** (2001) Variability of cork from Portuguese *Quercus suber* studied by solid-state (13)C-NMR and FTIR spectroscopies. *Biopolymers* **62**: 268-277
- López-Aljorna A, Bueno M, Aguinagalde I, Martín J** (2007) Fingerprinting and genetic variability in cork oak (*Quercus suber* L.) elite trees using ISSR and SSR markers. *Ann For Sci* **64**: 773-779
- Lu P-L, Chen N-Z, An R, Su Z, Qi B-S, Ren F, Chen J, Wang X-C** (2007) A novel drought-inducible gene, ATAF1, encodes a NAC family protein that negatively regulates the expression of stress-responsive genes in Arabidopsis. *Plant Mol Biol* **63**: 289-305
- Lu X-Y, Huang X-L** (2008) Plant miRNAs and abiotic stress responses. *Biochem Biophys Res Commun* **368**: 458-462
- Lumaret R, Tryphon-Dionnet M, Michaud H, Sanuy A, Ipotesi E, Born C, Mir C** (2005) Phylogeographical variation of chloroplast DNA in cork oak (*Quercus suber*). *Ann Bot* **96**: 853-861
- Luo M, Liu X, Singh P, Cui Y, Zimmerli L, Wu K** (2012) Chromatin modifications and remodeling in plant abiotic stress responses. *Biochim Biophys Acta* **1819**: 129-136
- Magri D, Fineschi S, Bellarosa R, Buonamici A, Sebastiani F, Schirone B, Simeone MC, Vendramin GG** (2007) The distribution of *Quercus suber* chloroplast haplotypes matches the palaeogeographical history of the western Mediterranean. *Mol Ecol* **16**: 5259-5266
- Mahajan S, Tuteja N** (2005) Cold, salinity and drought stresses: An overview. *Arch Biochem Biophys* **444**: 139-158
- Marum L, Miguel A, Ricardo CP, Miguel C** (2012) Reference gene selection for quantitative real-time PCR normalization in *Quercus suber*. *PLoS ONE* **7**: e35113

- Martin, L. B., Fei, Z., Giovannoni, J. J., & Rose, J. K.** (2013) Catalyzing plant science research with RNA-seq. *Front Plant Sci* **4**: 66
- Matesanz S, Valladares F** (2014) Ecological and evolutionary responses of Mediterranean plants to global change. *Env Exp Bot* **103**: 53-67
- Méthy M, Damesin C, Rambal S** (1996) Drought and photosystem II activity in two Mediterranean oaks. *Ann For Sci* **53**: 255-262
- Milne I, Stephen G, Bayer M, Cock PJ, Pritchard L, Cardle L, Shaw PD, Marshall D** (2012) Using Tablet for visual exploration of second-generation sequencing data. *Brief Bioinform* **14**: 193-202
- Miyakawa T, Fujita Y, Yamaguchi-Shinozaki K, Tanokura M** (2013) Structure and function of abscisic acid receptors. *Trends Plant Sci* **18**: 259-266
- Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K** (2012) AP2/ERF family transcription factors in plant abiotic stress responses. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mech* **1819**: 86-96
- Modesto IS, Miguel C, Pina-Martins F, Glushkova M, Veloso M, Paulo OS, Batista D** (2014) Identifying signatures of natural selection in cork oak (*Quercus suber* L.) genes through SNP analysis. *Tree Gen Gen* **10**: 1645-1660
- Mochida, K., & Shinozaki, K.** (2011) Advances in omics and bioinformatics tools for systems analyses of plant functions. *Plant Cell Physiol* **52**: 2017-2038.
- Nakashima K, Fujita Y, Katsura K, Maruyama K, Narusaka Y, Seki M, Shinozaki K, Yamaguchi-Shinozaki K** (2006) Transcriptional Regulation of ABI3- and ABA-responsive Genes Including RD29B and RD29A in Seeds, Germinating Embryos, and Seedlings of Arabidopsis. *Plant Mol Biol* **60**: 51-68
- Nakashima K, Takasaki H, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K** (2012) NAC transcription factors in plant abiotic stress responses. *Biochim Biophys Acta* **1819**: 97-103
- Nakashima K, Yamaguchi-Shinozaki K** (2013) ABA signaling in stress-response and seed development. *Plant Cell Rep* **32**: 959-970
- Nakashima K, Yamaguchi-Shinozaki K, Shinozaki K** (2014) The transcriptional regulatory network in the drought response and its crosstalk in abiotic stress responses including drought, cold, and heat. *Front Plant Sci* **5**: 170
- Nardini A, Lo Gullo MA, Trifilò P, Salleo S** (2014) The challenge of the Mediterranean climate to plant hydraulics: Responses and adaptations. *Env Exp Bot* **103**: 68-79
- Nixon K** (1993) Infrageneric classification of *Quercus* (Fagaceae) and typification of sectional names. *Ann For Sci* **50**: 25s-34s
- Nuche P, Komac B, Camarero JJ, Alados CL** (2014) Developmental instability as an index of adaptation to drought stress in a Mediterranean oak. *Ecol Indic* **40**: 68-75
- O'Connor TR, Dyreson C, Wyrick JJ** (2005) Athena: a resource for rapid visualization and systematic analysis of Arabidopsis promoter sequences. *Bioinformatics* **21**: 4411-4413
- Oliveira G, Costa A** (2012) How resilient is *Quercus suber* L. to cork harvesting? A review and identification of knowledge gaps. *For Ecol Manage* **270**: 257-272
- Osakabe Y, Yamaguchi-Shinozaki K, Shinozaki K, Tran L-SP** (2013) Sensing the environment: key roles of membrane-localized kinases in plant perception and response to abiotic stress. *J Exp Bot* **64**: 445-458
- Oshlack, A. and Wakefield, M.** (2009) Transcript length bias in RNAseq data confounds systems biology. *Biol. Direct.* **4**: 14
- Pandey GK, Grant JJ, Cheong YH, Kim BG, Li L, Luan S** (2005) ABR1, an APETALA2-domain transcription factor that functions as a repressor of ABA response in Arabidopsis. *Plant Physiol* **139**: 1185-1193

- Papageorgiou GC** (2007) Chlorophyll a Fluorescence: A Signature of Photosynthesis. Kluwer Academic, NY
- Pereira-Leal JB, Abreu IA, Alabaça CS, Almeida MH, Almeida P, Almeida T, Amorim MI, Araújo S, Azevedo H, Badia A** (2014) A comprehensive assessment of the transcriptome of cork oak (*Quercus suber*) through EST sequencing. *BMC Genomics* **15**: 371
- Persak H, Pitzschke A** (2014) Dominant repression by Arabidopsis transcription Factor MYB44 causes oxidative damage and hypersensitivity to abiotic stress. *Int J Mol Sci* **15**: 2517-2537
- Pfaffl MW** (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**: e45-e45
- Phillips MA, D'Auria JC, Luck K, Gershenzon J** (2009) Evaluation of candidate reference genes for real-time quantitative PCR of plant samples using purified cDNA as template. *Plant Mol Biol Rep* **27**: 407-416
- Proost S, Van Bel M, Sterck L, Billiau K, Van Parys T, Van de Peer Y, Vandepoele K** (2009) PLAZA: a comparative genomics resource to study gene and genome evolution in plants. *Plant Cell* **21**: 3718-3731
- Qin F, Shinozaki K, Yamaguchi-Shinozaki K** (2011) Achievements and challenges in understanding plant abiotic stress responses and tolerance. *Plant Cell Physiol* **52**: 1569-1582
- Raghavendra AS, Gonugunta VK, Christmann A, Grill E** (2010) ABA perception and signalling. *Trends Plant Sci* **15**: 395-401
- Remans T, Smeets K, Opdenakker K, Mathijsen D, Vangronsveld J, Cuypers A** (2008) Normalisation of real-time RT-PCR gene expression measurements in *Arabidopsis thaliana* exposed to increased metal concentrations. *Planta* **227**: 1343-1349
- Richards E, Reichardt M, Rogers S** (1994) Preparation of genomic DNA from plant tissue. *In* *Cur Protocols Mol Biol*, pp 2.3. 1-2.3. 7
- Rocheta M, Carvalho L, Viegas W, Morais-Cecilio L** (2012) Corky, a gypsy-like retrotransposon is differentially transcribed in *Quercus suber* tissues. *BMC Res Notes* **5**: 432
- Rocheta M, Sobral R, Magalhães J, Amorim MI, Ribeiro T, Pinheiro M, Egas C, Morais-Cecilio L, Costa MM** (2014) Comparative transcriptomic analysis of male and female flowers of monoecious *Quercus suber*. *Front Plant Sci* **5**: 599
- Saez A, Apostolova N, Gonzalez-Guzman M, Gonzalez-Garcia MP, Nicolas C, Lorenzo O, Rodriguez PL** (2004) Gain-of-function and loss-of-function phenotypes of the protein phosphatase 2C HAB1 reveal its role as a negative regulator of abscisic acid signalling. *Plant J* **37**: 354-369
- Saito R, Smoot ME, Ono K, Ruschinski J, Wang P-L, Lotia S, Pico AR, Bader GD, Ideker T** (2012) A travel guide to Cytoscape plugins. *Nat Methods* **9**: 1069-1076
- Sánchez-González M, Cañellas I, Montero G** (2008) Base-age invariant cork growth model for Spanish cork oak (*Quercus suber* L.) forests. *Eur J For Res* **127**: 173-182
- Schweighofer A, Hirt H, Meskiene I** (2004) Plant PP2C phosphatases: emerging functions in stress signaling. *Trends Plant Sci* **9**: 236-243
- Schweighofer A, Kazanaviciute V, Scheikl E, Teige M, Doczi R, Hirt H, Schwanninger M, Kant M, Schuurink R, Mauch F** (2007) The PP2C-type phosphatase AP2C1, which negatively regulates MPK4 and MPK6, modulates innate immunity, jasmonic acid, and ethylene levels in Arabidopsis. *Plant Cell* **19**: 2213-2224
- Sebastiana M, Vieira B, Lino-Neto T, Monteiro F, Figueiredo A, Sousa L, Pais MS, Tavares R, Paulo OS** (2014) Oak root response to ectomycorrhizal symbiosis

- establishment: RNA-Seq derived transcript identification and expression profiling. *PLoS One* **9**: e98376
- Shao H-B, Chu L-Y, Jaleel CA, Zhao C-X** (2008) Water-deficit stress-induced anatomical changes in higher plants. *C R Biol* **331**: 215-225
- Shinozaki K, Yamaguchi-Shinozaki K** (2007) Gene networks involved in drought stress response and tolerance. *J Exp Bot* **58**: 221-227
- Singh T, Aspinall D, Paleg L** (1972) Proline accumulation and varietal adaptability to drought in barley: a potential metabolic measure of drought resistance. *Nature* **236**: 188-190
- Sixto H, Aranda I, Grau J** (2006) Assessment of salt tolerance in *Populus alba* clones using chlorophyll fluorescence. *Photosynthetica* **44**: 169-173
- Sofo A, Manfreda S, Fiorentino M, Dichio B, Xiloyannis C** (2008) The olive tree: a paradigm for drought tolerance in Mediterranean climates. *Hydrol Earth Syst Sci Disc* **12**: 293-301
- Soler M, Serra O, Molinas M, Garcia-Berthou E, Caritat A, Figueras M** (2008) Seasonal variation in transcript abundance in cork tissue analyzed by real time RT-PCR. *Tree Physiol* **28**: 743-751
- Soler M, Serra O, Molinas M, Huguet G, Fluch S, Figueras M** (2007) A genomic approach to suberin biosynthesis and cork differentiation. *Plant Physiol* **144**: 419-431
- Soto A, Lorenzo Z, Gil L** (2003) Nuclear microsatellite markers for the identification of *Quercus ilex* L. and *Q. suber* L. hybrids. *Silvae Genet* **52**: 63-66
- Soto A, Lorenzo Z, Gil L** (2007) Differences in fine-scale genetic structure and dispersal in *Quercus ilex* L. and *Q. suber* L.: consequences for regeneration of mediterranean open woods. *Heredity* **99**: 601-607
- Team RDC** (2008) R: A Language and Environment for Statistical Computing.
- Teixeira RT, Fortes AM, Pinheiro C, Pereira H** (2014) Comparison of good-and bad-quality cork: application of high-throughput sequencing of phellogenetic tissue. *J Exp Bot* **65**: 4887-490
- Tenhaken R** (2014) Cell wall remodeling under abiotic stress. *Front Plant Sci* **5**: 771
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Kruger P, Selbig J, Muller LA, Rhee SY, Stitt M** (2004) MAPMAN: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* **37**: 914-939
- Thimm O, Blasing O, Gibon Y, Nagel A, Meyer S, Krüger P, Selbig J, Müller LA, Rhee SY, Stitt M** (2004) mapman: a user-driven tool to display genomics data sets onto diagrams of metabolic pathways and other biological processes. *Plant J* **37**: 914-939
- Toumi L, Lumaret R** (1998) Allozyme variation in cork oak (*Quercus suber* L.): the role of phylogeography and genetic introgression by other Mediterranean oak species and human activities. *Theor Appl Genet* **97**: 647-656
- Toumi L, Lumaret R** (2001) Allozyme characterisation of four Mediterranean evergreen oak species. *Biochem Syst Ecol* **29**: 799-817
- Umezawa T, Fujita M, Fujita Y, Yamaguchi-Shinozaki K, Shinozaki K** (2006) Engineering drought tolerance in plants: discovering and tailoring genes to unlock the future. *Current Op. Biotech* **17**: 113-122
- Usadel B, Nagel A, Thimm O, Redestig H, Blaesing OE, Palacios-Rojas N, Selbig J, Hannemann J, Piques MC, Steinhauser D** (2005) Extension of the visualization tool MapMan to allow statistical analysis of arrays, display of coresponding genes, and comparison with known responses. *Plant Physiol.* **138**: 1195-1204
- Usadel B, Poree F, Nagel A, Lohse M, CZEDIK-EYSENBERG A, Stitt M** (2009) A guide to using MapMan to visualize and compare Omics data in plants: a case study in the crop species, Maize. *Plant, Cell & Environ* **32**: 1211-1229

- Ward JA, Ponnala L, Weber CA** (2012) Strategies for transcriptome analysis in nonmodel plants. *Amer J Bot* **99**: 267-276
- Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K** (2012) Dissecting plant genomes with the PLAZA comparative genomics platform. *Plant Physiol* **158**: 590-600.
- Van Verk, M. C., Hickman, R., Pieterse, C. M., & Van Wees, S. C.** (2013) RNA-Seq: revelation of the messengers. *Trends Plant Sci* **18**: 175-179.
- Vandepoele K, Quimbaya M, Casneuf T, De Veylder L, Van de Peer Y** (2009) Unraveling transcriptional control in Arabidopsis using cis-regulatory elements and coexpression networks. *Plant Physiol* **150**: 535-546
- Wan C-Y, Wilkins TA** (1994) A modified hot borate method significantly enhances the yield of high-quality RNA from cotton (*Gossypium hirsutum* L.). *Anal Biochem* **223**: 7-12
- Wang W, Vinocur B, Altman A** (2003) Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance. *Planta* **218**: 1-14
- Wang W, Vinocur B, Shoseyov O, Altman A** (2004) Role of plant heat-shock proteins and molecular chaperones in the abiotic stress response. *Trends Plant Sci* **9**: 244-252
- Wang X, Du B, Liu M, Sun N, Qi X** (2013) Arabidopsis Transcription Factor WRKY33 Is Involved in Drought by Directly Regulating the Expression of Cesa8. *Amer J Plant Sci* **4**: 21-27
- Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q** (2010) The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res* **38**: W214-W220
- Wilkins TA, Smart LB** (1996) Isolation of RNA from plant tissue. A laboratory guide to RNA: isolation, analysis, and synthesis: 21-42
- Xu M, Zhang B, Su X, Zhang S, Huang M** (2011) Reference gene selection for quantitative real-time polymerase chain reaction in *Populus*. *Anal Biochem* **408**: 337-339
- Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL** (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134
- Yoshida R, Hobo T, Ichimura K, Mizoguchi T, Takahashi F, Aronso J, Ecker JR, Shinozaki K** (2002) ABA-activated SnRK2 protein kinase is required for dehydration stress signaling in Arabidopsis. *Plant Cell Physiol* **43**: 1473-1483
- Yoshida T, Fujita Y, Sayama H, Kidokoro S, Maruyama K, Mizoi J, Shinozaki K, Yamaguchi-Shinozaki K** (2010) AREB1, AREB2, and ABF3 are master transcription factors that cooperatively regulate ABRE-dependent ABA signaling involved in drought stress tolerance and require ABA for full activation. *Plant J* **61**: 672-685
- Zhang J, Li X, He Z, Zhao X, Wang Q, Zhou B, Yu D, Huang X, Tang D, Guo X, Liu X** (2013) Molecular character of a phosphatase 2C (PP2C) gene relation to stress tolerance in Arabidopsis thaliana. *Mol Biol Rep* **40**: 2633-2644
- Zhao, Q. Y., Wang, Y., Kong, Y. M., Luo, D., Li, X., & Hao, P.** (2011) Optimizing de novo transcriptome assembly from short-read RNA-Seq data: a comparative study. *BMC Bioinformatics* **12**: Suppl 14, S2.

Appendixes

APPENDIX I

ANNOTATION TABLE OF THE DIFFERENTIAL EXPRESSED GENES, OBTAINED WITH BLAST2GO ANNOTATION AGAINST NON-REDUNDANT PROTEIN DATABASE (NR) AND BLASTx ANNOTATION AGAINST TAIR10 PROTEIN DATABASE.

<i>Quercus suber</i>			<i>Arabidopsis thaliana</i>			<i>Quercus suber</i>	
Unigene ID	Differential expression	Gene cluster	Annotation (BLAST2GO)	AGI	Annotation (TAIR 10)	At Prot Length	Qs Prot Length
QSDrought_00149	Up	5	rna-binding family protein isoform 2	AT1G07350	SR45a RNA-binding (RRM/RBD/RNP motifs) family protein	382	134
QSDrought_00248	Up	11	beta-expansin 1a	AT3G45970	ATEXLA1_ATEXPL1_ATHEXP BETA 2.1_EXLA1_EXPL1 expansin-like A1	265	270
QSDrought_00266	Up	10	—NA—				109
QSDrought_00278	Down	2	galacturonosyltransferase 8-like	AT3G25140	GAUT8_QUA1 Nucleotide-diphospho-sugar transferases superfamily protein	559	231
QSDrought_00285	Up	5	cbl-interacting serine threonine-protein	AT4G30960	ATCIPK6_CIPK6_SIP3_SNRK3.14 SOS3-interacting protein 3	441	434
QSDrought_00293	Down	1	protein aspartic protease in guard cell 1-like	AT3G18490	Eukaryotic aspartyl protease family protein	500	254
QSDrought_00333	Down	2	auxin-induced protein 5ng4-like	AT5G40230	nodulin MtN21 /EamA-like transporter family protein	370	220
QSDrought_00336	Down	4	cytochrome p450 cyp94a1 family protein	AT5G63450	CYP94B1 cytochrome P450, family 94, subfamily B, polypeptide 1	510	407

QSDrought_00338	Down	1	cytochrome p450 94a2-like	AT3G01900	CYP94B2 cytochrome P450, family 94, subfamily B, polypeptide 2	496	103
QSDrought_00354	Up	11	glucan endo- -beta-glucosidase	AT1G32860	Glycosyl hydrolase superfamily protein	426	376
QSDrought_00370	Up	10	unnamed protein product	AT2G04620	Cation efflux family protein	798	244
QSDrought_00371	Up	10	unnamed protein product				153
QSDrought_00437	Up	10	u-box domain-containing protein 21-like	AT1G66160	ATCMPG1_CMPG1 CYS, MET, PRO, and GLY protein 1	431	337
QSDrought_00469	Up	3	inositol transporter 1-like	AT2G43330	ATINT1_INT1 inositol transporter 1	509	261
QSDrought_00603	Up	10	polyubiquitin	AT4G05050	UBQ11 ubiquitin 11 UBQ11 ubiquitin 11 UBQ11 ubiquitin 11	229	226
QSDrought_00611	Up	3	late embryogenesis-abundant protein	AT4G15910	ATDI21_DI21 drought-induced 21	104	109
QSDrought_00612	Up	3	late embryogenesis-abundant protein	AT4G15910	ATDI21_DI21 drought-induced 21	104	106
QSDrought_00632	Up	10	proline-rich protein PRP1				102
QSDrought_00633	Up	11	proline-rich protein PRP1				88
QSDrought_00642	Up	11	iron transport protein 2	AT1G54410	dehydrin family protein dehydrin family protein	98	92
QSDrought_00643	Up	11	iron transport protein 2	AT1G54410	dehydrin family protein dehydrin family protein	98	87
QSDrought_00647	Down	1	nitrate reductase	AT1G77760	GNR1_NIA1_NR1 nitrate reductase 1	917	723
QSDrought_00668	Up	5	phosphoenolpyruvate carboxykinase	AT4G37870	PCK1_PEPCK phosphoenolpyruvate carboxykinase 1	671	492
QSDrought_00694	Up	11	sorbitol transporter	AT3G18830	ATPLT5_ATPMT5_PMT5 polyol/monosaccharide transporter 5	539	426
QSDrought_00717	Down	2	E3 ubiquitin-protein ligase	AT3G26100	Regulator of chromosome condensation (RCC1) family protein	532	432
QSDrought_00747	Up	3	nac transcription factor 055	AT4G35580	NTL9 NAC transcription factor-like 9	534	501

QSDrought_00785	Up	11	galactinol synthase family protein	AT1G56600	AtGolS2_GolS2 galactinol synthase 2 AtGolS2_GolS2 galactinol synthase 2 AtGolS2_GolS2 galactinol synthase 2	335	335
QSDrought_00786	Up	10	galactinol synthase family protein	AT1G56600	AtGolS2_GolS2 galactinol synthase 2 AtGolS2_GolS2 galactinol synthase 2 AtGolS2_GolS2 galactinol synthase 2	335	328
QSDrought_00834	Up	11	plastid-lipid-associated chloroplastic-like	AT4G22240	Plastid-lipid associated protein PAP / fibrillin family protein	310	235
QSDrought_00975	Up	10	omega-3 fatty acid desaturase	AT3G11170	AtFAD7_FAD7_FADD fatty acid desaturase 7	446	270
QSDrought_01049	Up	9	gpn-loop gtpase 1 homolog	AT4G21800	QQT2 P-loop containing nucleoside triphosphate hydrolases superfamily protein	379	354
QSDrought_01079	Down	1	nramp transporter	AT1G15960	ATNRAMP6_NRAM6 NRAMP metal ion transporter 6 ATNRAMP6_NRAM6 NRAMP metal ion transporter 6	527	273
QSDrought_01105	Up	5	probable protein phosphatase 2c 24-like	AT1G07430	HAI2 highly ABA-induced PP2C gene 2 HAI2 highly ABA-induced PP2C gene 2	442	264
QSDrought_01106	Up	9	probable protein phosphatase 2c 24-like	AT1G07430	HAI2 highly ABA-induced PP2C gene 2 HAI2 highly ABA-induced PP2C gene 2	442	214
QSDrought_01150	Down	7	dnase i-like superfamily protein isoform 3	AT4G30900	DNAse I-like superfamily protein	316	242
QSDrought_01191	Down	8	abc1 kinase	AT3G07700	Protein kinase superfamily protein	724	254
QSDrought_01227	Up	11	chitinase 1				306
QSDrought_01264	Down	1	hexose transporter 1	AT1G11260	ATSTP1_STP1 sugar transporter 1	522	286
QSDrought_01287	Up	5	acyl- binding protein 3b	AT4G24230	ACBP3 acyl-CoA-binding domain 3	366	355

QSDrought_01408	Down	1	endo- -beta-d-glucanase-like	AT3G23600	alpha/beta-Hydrolases superfamily protein	239	200
QSDrought_01437	Down	5	60s ribosomal protein l13a-4-like	AT5G48760	Ribosomal protein L13 family protein	206	206
QSDrought_01447	Up	5	zinc finger protein zat10-like	AT1G27730	STZ_ZAT10 salt tolerance zinc finger	227	248
QSDrought_01637	Up	5	histone h1	AT2G30620	winged-helix DNA-binding transcription factor family protein	273	212
QSDrought_01646	Up	7	60s ribosomal protein l15	AT4G16720	Ribosomal protein L23/L15e family protein	204	204
QSDrought_01718	Down	2	germin-like protein 5	AT5G39110	RmlC-like cupins superfamily protein	222	217
QSDrought_01793	Up	11	sucrose responsive element binding protein	AT5G67300	RmlC-like cupins superfamily protein ATMYB44_ATMYBR1_MYB44_MYB	305	165
QSDrought_01805	Up	5	expansin-like b1-like	AT4G17030	R1 myb domain protein r1 AT-EXPR_ATEXLB1_ATEXPR1_ATHEXP	250	147
QSDrought_01806	Up	5	major pollen allergen ory s 1	AT4G17030	BETA 3.1_EXLB1_EXPR expansin-like B1 AT-EXPR_ATEXLB1_ATEXPR1_ATHEXP	250	174
QSDrought_01863	Up	9	probable mediator of rna polymerase ii transcription subunit 26b- partial	AT5G05140	BETA 3.1_EXLB1_EXPR expansin-like B1 Transcription elongation factor (TFIIS) family protein	436	167
QSDrought_01903	Up	9	nadp-dependent alkenal double bond reductase p2-like	AT5G37980	Zinc-binding dehydrogenase family protein	353	105
QSDrought_01949	Up	5	11s globulin seed storage protein 2	AT1G07750	RmlC-like cupins superfamily protein	356	221
QSDrought_02089	Up	10	bromodomain protein	AT1G05910	cell division cycle protein 48-related / CDC48-related	1210	64

QSDrought_02243	Up	5	af334835_1proline-rich protein prp1				118
QSDrought_02263	Up	11	Pc13g14020				211
QSDrought_02266	Down	8	achain structure of a molybdopterin-bound cnx1g domain links molybdenum and copper metabolism	AT5G20990	B73_CHL6_CNX_CNX1_SIR4 molybdopterin biosynthesis CNX1 protein / molybdenum cofactor biosynthesis enzyme CNX1 (CNX1)	670	162
QSDrought_02287	Up	5	inactive poly	AT1G32230	ATP8_AtRCD1_CEO_CEO1_RCD1 WWE protein-protein interaction domain protein family	589	192
QSDrought_02295	Up	5	abscisic acid-insensitive 5-like protein	AT1G45249	ABF2_AREB1_AtABF2_ATAREB1 abscisic acid responsive elements-binding factor 2	416	96
QSDrought_02345	Down	1	metallothionein-like protein	AT5G02380	MT2B metallothionein 2B	77	77
QSDrought_02471	Up	11	—NA—				74
QSDrought_02537	Up	5	aba 8 -hydroxylase	AT2G29090	CYP707A2 cytochrome P450, family 707, subfamily A, polypeptide 2	482	144
QSDrought_02597	Down	6	duf246 domain-containing protein at1g04910-like	AT1G22460	O-fucosyltransferase family protein	565	83
QSDrought_02649	Up	3	RING/U-box domain-containing protein isoform 2				101
QSDrought_02656	Up	3	abc transporter f family member 1-like	AT5G60790	ABCF1_ATGCN1_GCN1 ABC transporter family protein	595	599
QSDrought_02686	Down	4	oligopeptide transporter 3-like	AT4G16370	ATOPT3_OPT3_OPT3 oligopeptide transporter	737	529
QSDrought_02700	Up	3	hydroxyproline-rich glycoprotein	AT4G28300	Protein of unknown function (DUF1421)	496	530
QSDrought_02760	Down	6	acetolactate synthase small subunit chloroplastic-like	AT2G31810	ACT domain-containing small subunit of acetolactate synthase protein	491	481

QSDrought_02773	Up	3	probable inositol transporter 2-like	AT1G30220	ATINT2_INT2 inositol transporter 2	580	335
QSDrought_02776	Up	5	quinolinate synthase	AT5G50210	OLD5_QS_SUFE3 quinolinate synthase	718	228
QSDrought_02799	Down	4	nucleotide pyrophosphatase phosphodiesterase	AT5G50400	ATPAP27_PAP27 purple acid phosphatase 27 ATPAP27_PAP27 purple acid phosphatase 27	611	431
QSDrought_02834	Down	1	glutamine synthetase	AT5G35630	ATGSL1_GLN2_GS2 glutamine synthetase 2	430	432
QSDrought_02866	Up	5	mediator of rna polymerase ii transcription subunit 33b-like	AT2G48110	MED33B_REF4 reduced epidermal fluorescence 4	1275	427
QSDrought_02869	Up	3	acetate butyrate- ligase peroxisomal-like	AT3G16910	AAE7_ACN1 acyl-activating enzyme 7	569	471
QSDrought_02886	Down	4	cytochrome p450	AT5G36110	CYP716A1 cytochrome P450, family 716, subfamily A, polypeptide 1	477	481
QSDrought_02899	Up	3	auxin-induced in root cultures protein 12	AT5G47530	Auxin-responsive family protein	395	399
QSDrought_02969	Up	9	nuclease harbi1-like	AT4G29780	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G12010.1); Has 945 Blast hits to 944 proteins in 87 species: Archae - 0; Bacteria - 0; Metazoa - 519; Fungi - 43; Plants - 365; Viruses - 0; Other Eukaryotes - 18 (source: NCBI BLINK).	540	453
QSDrought_02995	Down	1	low quality protein: laccase-14-like	AT5G09360	LAC14 laccase 14	569	490
QSDrought_03019	Up	5	beta-amylase chloroplastic-like	AT3G23920	BAM1_BMY7_TR-BAMY beta-amylase 1	575	401
QSDrought_03061	Up	11	stress-responsive family	AT5G52300	LTI65_RD29B CAP160 protein	619	321

QSDrought_03076	Down	4	protein calcium-transporting atpase endoplasmic reticulum-type- like	AT1G07670	ATECA4_ECA4 endomembrane-type CA-ATPase 4	1061	410
QSDrought_03101	Up	3	hydrolyzing o-glycosyl	AT5G20250	DIN10_RS6 Raffinose synthase family protein DIN10_RS6 Raffinose synthase family protein	844	329
QSDrought_03109	Up	5	thioredoxin-like 1- chloroplastic-like	AT1G08570	ACHT4 atypical CYS HIS rich thioredoxin 4	275	296
QSDrought_03138	Up	5	galactinol synthase family protein	AT1G56600	AtGolS2_GolS2 galactinol synthase 2 AtGolS2_GolS2 galactinol synthase 2 AtGolS2_GolS2 galactinol synthase 2	335	315
QSDrought_03140	Down	4	s-adenosylmethionine synthetase	AT1G02500	AtSAM1_MAT1_METK1_SAM- 1_SAM1 S-adenosylmethionine synthetase 1	393	393
QSDrought_03145	Down	1	linoleate 13s-lipoxygenase 2- chloroplastic-like	AT3G45140	ATLOX2_LOX2 lipoxygenase 2	896	447
QSDrought_03208	Up	5	f-box kelch-repeat protein at3g27150-like	AT3G27150	Galactose oxidase/kelch repeat superfamily protein	422	400
QSDrought_03225	Down	2	copine family isoform 1	AT5G63970	Copine (Calcium-dependent phospholipid-binding protein) family	367	389
QSDrought_03335	Down	2	l-galactose dehydrogenase	AT4G33670	NAD(P)-linked oxidoreductase superfamily protein	319	106
QSDrought_03363	Up	3	PREDICTED: uncharacterized protein LOC101297283	AT2G12400	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 13 growth	541	278

					stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G25270.1); Has 177 Blast hits to 172 proteins in 23 species: Archae - 0; Bacteria - 2; Metazoa - 3; Fungi - 0; Plants - 164; Viruses - 0; Other Eukaryotes - 8 (source: NCBI BLink).		
QSDrought_03374	Up	5	nac domain transcriptional regulator superfamily protein isoform 2	AT4G27410	ANAC072_RD26 NAC (No Apical Meristem) domain transcriptional regulator superfamily protein	314	328
QSDrought_03381	Down	2	cell division protein homolog 2- chloroplastic-like	AT3G52750	FTSZ2-2 Tubulin/FtsZ family protein	473	343
QSDrought_03382	Down	4	nitrite reductase	AT2G15620	ATHNIR_NIR_NIR1 nitrite reductase 1 ATHNIR_NIR_NIR1 nitrite reductase 1	586	198
QSDrought_03390	Down	4	cell elongation protein	AT3G19820	CBB1_DIM_DIM1_DWF1_EVE1 cell elongation protein / DWARF1 / DIMINUTO (DIM)	561	333
QSDrought_03433	Down	1	subtilisin-like protease-like	AT5G67360	ARA12 Subtilase family protein	757	379
QSDrought_03436	Down	6	bro1 domain-containing protein brox	AT5G14020	Endosomal targeting BRO1-like domain-containing protein	401	390
QSDrought_03461	Down	1	pleiotropic drug resistance protein 1-like	AT1G15520	ABCG40_ATABCG40_ATPDR12_PD R12 pleiotropic drug resistance 12	1423	228
QSDrought_03501	Up	10	mitogen-activated protein kinase kinase kinase a-like	AT5G55090	MAPKKK15 mitogen-activated protein kinase kinase kinase 15	448	410
QSDrought_03503	Down	1	galactose oxidase	AT1G75620	glyoxal oxidase-related protein	547	424
QSDrought_03528	Up	3	PREDICTED: uncharacterized protein LOC101314743	AT1G13480	Protein of unknown function (DUF1262)	387	391
QSDrought_03551	Up	5	serine threonine-protein kinase sapk3-like	AT4G33950	ATOST1_OST1_P44_SNRK2-6_SNRK2.6_SRK2E Protein kinase	362	339

QSDrought_03563	Up	5	ctp synthase family protein	AT1G30820	superfamily protein CTP synthase family protein	600	302
QSDrought_03606	Up	5	protein phosphatase 2c	AT2G29380	HAI3 highly ABA-induced PP2C gene 3	362	251
QSDrought_03607	Up	11	dentin sialophospho isoform 3	AT5G64170	dentin sialophosphoprotein-related	616	374
QSDrought_03640	Up	3	snf1-related protein kinase regulatory subunit gamma-1- like	AT3G48530	KING1 SNF1-related protein kinase regulatory subunit gamma 1	424	412
QSDrought_03665	Up	5	u-box domain-containing protein 4-like	AT3G01400	ARM repeat superfamily protein	355	357
QSDrought_03705	Down	1	polygalacturonase-1 non- catalytic subunit beta	AT1G70370	PG2 polygalacturonase 2	626	341
QSDrought_03719	Down	2	gras family transcription factor	AT4G37650	SGR7_SHR GRAS family transcription factor	531	202
QSDrought_03742	Up	9	probable wrky transcription factor 40-like	AT1G80840	ATWRKY40_WRKY40 WRKY DNA- binding protein 40	302	310
QSDrought_03743	Up	5	bromo-adjacent homology domain-containing family protein	AT4G11560	bromo-adjacent homology (BAH) domain-containing protein	587	256
QSDrought_03754	Up	3	thioredoxin-like 1- chloroplastic-like	AT5G61440	ACHT5 atypical CYS HIS rich thioredoxin 5	245	273
QSDrought_03763	Down	4	probable pectate lyase 15-like	AT4G13710	Pectin lyase-like superfamily protein	470	274
QSDrought_03769	Down	1	mta sah	AT4G28940	Phosphorylase superfamily protein	348	355
QSDrought_03777	Down	1	peroxidase 27	AT5G15180	Peroxidase superfamily protein	329	329
QSDrought_03809	Up	10	dna binding	AT5G17300	RVE1 Homeodomain-like superfamily protein	387	254
QSDrought_03820	Down	1	flavonol synthase	AT5G08640	ATFLS1_FLS_FLS1 flavonol synthase 1	336	331
QSDrought_03835	Up	5	udp-glucose 4-epimerase gepi48-like	AT4G10960	UGE5 UDP-D-glucose/UDP-D- galactose 4-epimerase 5	351	349
QSDrought_03883	Down	1	gds1 esterase lipase 5-like	AT1G53920	GLIP5 GDSL-motif lipase 5	385	385

QSDrought_03929	Up	5	modifier of isoform 1	AT4G24680	MOS1 modifier of snc1	1427	274
QSDrought_03961	Up	11	u-box domain-containing protein 21-like	AT3G02840	ARM repeat superfamily protein	379	351
QSDrought_03976	Down	1	peroxidase 16 precursor family protein	AT4G30170	Peroxidase family protein	325	323
QSDrought_03994	Up	3	e3 ubiquitin-protein ligase upl3-like	AT4G38600	KAK_UPL3 HEAT repeat ;HECT-domain (ubiquitin-transferase)	1888	291
QSDrought_03995	Up	10	carboxylesterase 1-like	AT5G62180	AtCXE20_CXE20 carboxyesterase 20	327	220
QSDrought_04022	Down	6	rhomboid protease glup	AT1G25290	ATRBL10_RBL10 RHOMBOID-like protein 10	343	343
QSDrought_04063	Up	5	spfh domain-containing protein 2	AT2G03510	SPFH/Band 7/PHB domain-containing membrane-associated protein family	356	363
QSDrought_04101	Down	1	amt1 2	AT1G64780	AMT1;2_ATAMT1;2 ammonium transporter 1;2	514	99
QSDrought_04107	Down	1	ankyrin repeat-containing protein at3g12360	AT4G03500	Ankyrin repeat family protein	652	318
QSDrought_04114	Up	5	cytosolic fe-s cluster assembly factor nubp1 homolog	AT5G50960	ATNBP35_NBP35 nucleotide binding protein 35	350	355
QSDrought_04181	Down	1	peroxidase 55-like	AT5G14130	Peroxidase superfamily protein	330	323
QSDrought_04184	Down	4	pyrophosphate-energized vacuolar membrane proton pump 1-like	AT1G15690	ATAVP3_AtVHP1;1_AVP-3_AVP1_FUGU5 Inorganic H pyrophosphatase family protein	770	288
QSDrought_04196	Up	3	ring finger and chy zinc finger domain-containing protein 1-like	AT5G25560	CHY-type/CTCHY-type/RING-type Zinc finger protein	328	250
QSDrought_04215	Up	5	calcium ion binding	AT4G38810	Calcium-binding EF-hand family protein	375	364
QSDrought_04231	Down	4	sieve element occlusion a	AT3G01680	SEOb_SEOR1 CONTAINS InterPro DOMAIN/s: Mediator complex	740	285

					subunit Med28 (InterPro:IPR021640); BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G01670.1); Has 122 Blast hits to 112 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 122; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLINK).		
QSDrought_04306	Up	5	enth vhs family protein isoform 1	AT3G26990	ENTH/VHS family protein	513	369
QSDrought_04311	Down	2	callose synthase 3-like	AT5G13000	ATGSL12_CALS3_gsl12_GSL12 glucan synthase-like 12	1955	319
QSDrought_04367	Up	5	conserved hypothetical protein	AT4G27350	Protein of unknown function (DUF1223)	278	278
QSDrought_04385	Down	1	nad -linked oxidoreductase superfamily protein	AT1G60690	NAD(P)-linked oxidoreductase superfamily protein	345	348
QSDrought_04386	Down	1	chitinase-like protein 2	AT1G05850	ATCTL1_CTL1_ELP_ELP1_ERH2_ HOT2_POM1 Chitinase family protein	321	313
QSDrought_04401	Up	5	homeobox protein knotted-1- like 3-like	AT5G25220	KNAT3 KNOTTED1-like homeobox gene 3	431	295
QSDrought_04409	Down	1	probable boron transporter 2- like	AT3G62270	HCO3- transporter family	703	316
QSDrought_04428	Down	4	unknown	AT1G64140	BEST Arabidopsis thaliana protein match is: loricrin-related (TAIR:AT5G64550.1); Has 3596 Blast hits to 2004 proteins in 150 species: Archae - 0; Bacteria - 54; Metazoa - 1597; Fungi - 38; Plants -	646	228

QSDrought_04430	Up	5	probable galactinol-sucrose galactosyltransferase 2-like	AT3G57520	513; Viruses - 9; Other Eukaryotes - 1385 (source: NCBI BLink). AtSIP2_RS2_SIP2 seed imbibition 2	773	306
QSDrought_04452	Up	5	cbl-interacting protein kinase 25	AT5G10930	CIPK5_SnRK3.24 CBL-interacting protein kinase 5	445	348
QSDrought_04517	Up	3	r3h domain containing	AT3G56680	Single-stranded nucleic acid binding R3H protein	353	353
QSDrought_04556	Down	1	aspartyl protease family protein	AT5G10770	Eukaryotic aspartyl protease family protein	474	146
QSDrought_04567	Up	5	probable e3 ubiquitin-protein ligase bah1-like 1-like	AT1G02860	BAH1_NLA SPX (SYG1/Pho81/XPR1) domain-containing protein	335	256
QSDrought_04568	Up	10	glucan endo- -beta-glucosidase 5-like	AT4G17180	O-Glycosyl hydrolases family 17 protein	475	344
QSDrought_04602	Up	5	probable udp-n-acetylglucosamine-peptide n-acetylglucosaminyltransferase sec-like	AT3G04240	SEC Tetratricopeptide repeat (TPR)-like superfamily protein	977	144
QSDrought_04621	Up	5	translation initiation factor if2 if5 isoform 1	AT1G36730	Translation initiation factor IF2/IF5	439	289
QSDrought_04638	Up	5	lipase 1-like	AT1G18460	alpha/beta-Hydrolases superfamily protein	701	184
QSDrought_04726	Up	10	upf0496 protein 4-like	AT1G01550	BPS1 Protein of unknown function (DUF793)	349	327
QSDrought_04756	Up	11	predicted protein	AT5G02020	SIS Encodes a protein involved in salt tolerance, names SIS (Salt Induced Serine rich).	149	150
QSDrought_04760	Up	3	nam-like protein	AT1G01720	ANAC002_ATAF1 NAC (No Apical Meristem) domain transcriptional regulator superfamily protein	289	267

QSDrought_04780	Down	4	tonoplast intrinsic	AT4G01470	ATTIP1.3_GAMMA-TIP3_TIP1;3 tonoplast intrinsic protein 1;3	252	252
QSDrought_04783	Down	1	exordium like 3	AT5G51550	EXL3 EXORDIUM like 3	337	340
QSDrought_04815	Down	4	peroxidase superfamily protein	AT2G39040	Peroxidase superfamily protein	350	324
QSDrought_04819	Down	4	eukaryotic aspartyl protease family isoform partial	AT5G10770	Eukaryotic aspartyl protease family protein Eukaryotic aspartyl protease family protein	474	325
QSDrought_04851	Up	11	probable inactive poly	AT1G23550	SRO2 similar to RCD one 2	323	323
QSDrought_04879	Down	1	hypersensitive-induced response protein	AT5G62740	ATHIR1_AtHIR4_HIR1_HIR4 SPFH/Band 7/PHB domain- containing membrane-associated protein family	286	286
QSDrought_04884	Up	3	class i chitinase	AT3G12500	ATHCHIB_B-CHI_CHI- B_HCHIB_PR-3_PR3 basic chitinase	335	321
QSDrought_04922	Up	5	at5g26340 f9d12_17	AT5G26340	ATSTP13_MSS1_STP13 Major facilitator superfamily protein	526	294
QSDrought_04944	Down	1	beta-galactosidase 3	AT1G45130	BGAL5 beta-galactosidase 5	732	97
QSDrought_04959	Up	5	ap2 erf domain-containing transcription factor	AT2G33710	Integrase-type DNA-binding superfamily protein	231	335
QSDrought_04987	Up	5	unknown protein	AT2G01100	unknown protein; Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	247	258
QSDrought_05007	Down	8	serine arginine repetitive matrix protein 2 isoform 2	AT4G29790	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G19390.1); Has 538 Blast hits to 357 proteins in	1211	245

124 species: Archae - 0; Bacteria - 74; Metazoa - 109; Fungi - 58; Plants - 105; Viruses - 2; Other Eukaryotes - 190 (source: NCBI BLink).

QSDrought_05018	Up	11	nadp-dependent alkenal double bond reductase p2	AT1G26320	Zinc-binding dehydrogenase family protein	351	311
QSDrought_05027	Down	6	leucine-rich repeat-containing	AT4G35470	PIRL4 plant intracellular ras group-related LRR 4	549	288
QSDrought_05040	Down	4	glutathione s-transferase	AT2G47730	ATGSTF5_ATGSTF8_GST6_GSTF8 glutathione S-transferase phi 8	263	215
QSDrought_05065	Down	1	cell wall invertase	AT3G13790	ATBFRUCT1_ATCWINV1 Glycosyl hydrolases family 32 protein	584	312
QSDrought_05100	Up	10	probable xyloglucan endotransglucosylase hydrolase protein 23-like	AT4G25810	XTH23_XTR6 xyloglucan endotransglycosylase 6	286	298
QSDrought_05104	Up	3	uncharacterized loc101210213	AT1G61260	Protein of unknown function (DUF761)	344	323
QSDrought_05163	Down	1	quinone-oxidoreductase chloroplastic-like	AT4G13010	Oxidoreductase, zinc-binding dehydrogenase family protein	329	299
QSDrought_05179	Down	6	nucleotide pyrophosphatase phosphodiesterase	AT5G50400	ATPAP27_PAP27 purple acid phosphatase 27	611	306
QSDrought_05204	Up	10	monoglyceride lipase-like	AT1G73480	alpha/beta-Hydrolases superfamily protein	463	321
QSDrought_05268	Up	5	probable -trehalose-phosphate synthase	AT1G23870	ATTPS9_TPS9_TPS9 trehalose-phosphatase/synthase 9	867	274
QSDrought_05293	Up	10	f-box family protein	AT4G15563	unknown protein; Has 36 Blast hits to 36 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 36; Viruses - 0; Other Eukaryotes - 0 (source: NCBI	167	253

QSDrought_05306	Down	1	glucosyltransferase-like protein	AT5G22740	BLink). ATCSLA02_ATCSLA2_CSLA02_CSLA2 cellulose synthase-like A02	534	234
QSDrought_05422	Up	5	e3 ubiquitin-protein ligase rglg2	AT5G14420	RGLG2 RING domain ligase2 RGLG2 RING domain ligase2	468	260
QSDrought_05433	Up	5	wrky transcription	AT1G62300	ATWRKY6_WRKY6 WRKY family transcription factor	553	293
QSDrought_05446	Up	11	cbf4 transcription factor	AT4G25470	ATCBF2_CBF2_DREB1C_FTQ4 C-repeat/DRE binding factor 2	216	177
QSDrought_05470	Up	3	uncharacterized loc101218277	AT5G24890	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 13 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G24550.1); Has 1807 Blast hits to 1807 proteins in 277 species: Archae - 0; Bacteria - 0; Metazoa - 736; Fungi - 347; Plants - 385; Viruses - 0; Other Eukaryotes - 339 (source: NCBI BLink).	240	215
QSDrought_05496	Up	5	class i chitinase isoform partial	AT3G12500	ATHCHIB_B-CHI_CHI-B_HCHIB_PR-3_PR3 basic chitinase ATHCHIB_B-CHI_CHI-B_HCHIB_PR-3_PR3 basic chitinase	335	263
QSDrought_05518	Up	11	9-cis-epoxycarotenoid dioxygenase	AT3G14440	ATNCED3_NCED3_SIS7_STO1 nine-cis-epoxycarotenoid	599	256

QSDrought_05528	Up	10	sucrose phosphate synthase	AT5G11110	dioxygenase 3 ATSPS2F_KNS2_SPS1_SPS2F_SP SA2 sucrose phosphate synthase 2F	1047	316
QSDrought_05540	Up	3	senescence-associated protein	AT2G23810	TET8 tetraspanin8	273	269
QSDrought_05572	Up	10	atp binding	AT5G58040	ATFIP1[V]_ATFIPS5_FIP1[V]_FIPS5 homolog of yeast FIP1 [V]	1196	147
QSDrought_05577	Up	3	s-adenosylmethionine decarboxylase	AT3G25570	Adenosylmethionine decarboxylase family protein	349	206
QSDrought_05583	Down	2	uncharacterized loc101220556	AT1G75060	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G19330.2); Has 104 Blast hits to 104 proteins in 22 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 104; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	242	223
QSDrought_05651	Up	10	probable xyloglucan endotransglucosylase hydrolase protein 23-like	AT5G57560	TCH4_XTH22 Xyloglucan endotransglucosylase/hydrolase family protein	284	295
QSDrought_05659	Up	5	copine family isoform 1	AT1G67800	Copine (Calcium-dependent phospholipid-binding protein) family	453	162
QSDrought_05678	Up	5	low quality protein: midasin- like	AT1G67120	ATPases;nucleotide binding;ATP binding;nucleoside- triphosphatases;transcription factor binding	5393	257
QSDrought_05686	Up	3	dehydrin partial	AT1G20440	AtCOR47_COR47_RD17 cold 47	265	224

QSDrought_05711	Down	1	multidrug resistance protein	AT2G36910	ABCB1_ATPGP1_PGP1 ATP binding cassette subfamily B1	1286	296
QSDrought_05714	Up	3	thaumatin-like protein	AT1G20030	Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein	316	243
QSDrought_05727	Down	2	pantoate-beta-alanine ligase-like	AT5G48840	ATPTS_PANC_PTS homolog of bacterial PANC	310	297
QSDrought_05735	Up	11	atp binding	AT5G63950	CHR24 chromatin remodeling 24	1090	258
QSDrought_05757	Down	6	probable methyltransferase pmt15-like	AT2G45750	S-adenosyl-L-methionine-dependent methyltransferases superfamily protein	631	259
QSDrought_05775	Up	10	major pollen allergen ory s 1	AT4G17030	AT-EXPR_ATEXLB1_ATEXPR1_ATHEXP BETA 3.1_EXLB1_EXPR expansin-like B1	250	257
QSDrought_05809	Down	1	metal transporter nramp5-like	AT1G15960	ATNRAMP6_NRAMP6 NRAMP metal ion transporter 6 ATNRAMP6_NRAMP6 NRAMP metal ion transporter 6	527	137
QSDrought_05835	Down	6	predicted protein	AT4G23410	TET5 tetraspanin5	281	205
QSDrought_05836	Up	5	probable lrr receptor-like serine threonine-protein kinase at1g53430-like	AT1G53430	Leucine-rich repeat transmembrane protein kinase	1030	142
QSDrought_05868	Down	8	protein cmss1-like	AT2G43110	unknown protein; Has 212 Blast hits to 211 proteins in 96 species: Archae - 0; Bacteria - 6; Metazoa - 62; Fungi - 91; Plants - 31; Viruses -	288	285

					0; Other Eukaryotes - 22 (source: NCBI BLINK).		
QSDrought_05926	Up	9	pirin-like	AT1G50590	RmlC-like cupins superfamily protein	310	226
QSDrought_05934	Up	5	PREDICTED: uncharacterized protein LOC100249003	AT1G20070	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; EXPRESSED IN: 20 plant structures; EXPRESSED DURING: 11 growth stages; Has 26 Blast hits to 26 proteins in 9 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 2; Plants - 24; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLINK).	193	230
QSDrought_05993	Up	11	thaumatin-like protein	AT1G20030	Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein	316	242
QSDrought_06016	Up	5	probable protein phosphatase 2c 49-like	AT3G62260	Protein phosphatase 2C family protein	384	196
QSDrought_06025	Down	1	xyloglucan endotransglycosylase hydrolase xth-21	AT4G14130	XTH15_XTR7 xyloglucan endotransglucosylase/hydrolase 15	289	280
QSDrought_06041	Up	3	thaumatin-like protein	AT1G20030	Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein Pathogenesis-related	316	247

QSDrought_06066	Up	5	nac domain ipr003441	AT5G08790	thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein anac081_ATAF2 NAC (No Apical Meristem) domain transcriptional regulator superfamily protein	283	175
QSDrought_06091	Down	2	dehydroquinate dehydratase shikimate dehydrogenase isoform 1	AT3G06350	EMB3004_MEE32 dehydroquinate dehydratase, putative / shikimate dehydrogenase, putative	603	129
QSDrought_06092	Down	4	PREDICTED: uncharacterized protein LOC101501710 isoform X2				202
QSDrought_06132	Up	10	tyramine n-feruloyltransferase 4 11-like	AT2G39020	Acyl-CoA N-acyltransferases (NAT) superfamily protein	236	238
QSDrought_06155	Down	1	cytochrome p450 71a1-like	AT5G07990	CYP75B1_D501_TT7 Cytochrome P450 superfamily protein	513	276
QSDrought_06157	Up	11	thaumatin-like protein	AT1G20030	Pathogenesis-related thaumatin superfamily protein Pathogenesis- related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein Pathogenesis-related thaumatin superfamily protein	316	268
QSDrought_06199	Down	3	probable -trehalose-phosphate synthase	AT1G60140	ATTPS10_TPS10_TPS10 trehalose phosphate synthase	861	173
QSDrought_06244	Up	9	heat shock factor	AT4G18880	AT-HSFA4A_HSF A4A heat shock transcription factor A4A	401	270
QSDrought_06310	Down	1	peroxidase 45-like	AT2G18980	Peroxidase superfamily protein	323	230
QSDrought_06336	Up	5	udp-glycosyltransferase 85a2	AT1G22360	AtUGT85A2_UGT85A2 UDP- glucosyl transferase 85A2	481	186
QSDrought_06348	Down	1	copper-transporting atpase 3-	AT1G63440	HMA5 heavy metal atpase 5	995	278

			like				
QSDrought_06356	Down	2	beta-glucosidase 24-like	AT2G44480	BGLU17 beta glucosidase 17	517	199
QSDrought_06375	Down	2	beta-glucosidase 41-like	AT5G54570	BGLU41 beta glucosidase 41	535	285
QSDrought_06386	Up	11	hypothetical protein PRUPE_ppa020923mg	AT2G38790	unknown protein; Has 21 Blast hits to 21 proteins in 5 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 19; Viruses - 0; Other Eukaryotes - 2 (source: NCBI BLink).	204	205
QSDrought_06388	Down	6	probable serine threonine- protein kinase at1g54610-like isoform x1	AT5G50860	Protein kinase superfamily protein	580	265
QSDrought_06406	Down	2	nac domain ipr003441	AT5G61430	ANAC100_ATNAC5_NAC100 NAC domain containing protein 100	336	355
QSDrought_06420	Down	4	cellulose synthase	AT5G17420	ATCESA7_CESA7_IRX3_MUR10 Cellulose synthase family protein	1026	171
QSDrought_06423	Up	10	zinc finger protein 5-like	AT1G10480	ZFP5 zinc finger protein 5	211	253
QSDrought_06426	Down	6	chitinase-like protein 2	AT3G16920	ATCTL2_CTL2 chitinase-like protein 2	348	247
QSDrought_06480	Up	9	b-cell receptor-associated 31- like protein	AT5G42570	B-cell receptor-associated 31-like	218	221
QSDrought_06522	Up	5	late embryogenesis abundant hydroxyproline-rich glycoprotein	AT2G35980	ATNHL10_NHL10_YLS9 Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	227	230
QSDrought_06538	Up	5	uncharacterized loc101205361				158
QSDrought_06564	Up	10	vq motif-containing family protein	AT3G22160	VQ motif-containing protein	192	237
QSDrought_06575	Up	9	biogenesis of lysosome-related organelles complex 1 subunit	AT2G30330	BLOS1 GCN5L1 family protein	152	154

QSDrought_06583	Up	5	1-like er lumen protein retaining receptor-like	AT1G29330	AERD2_ATERD2_ERD2 ER lumen protein retaining receptor family protein	215	215
QSDrought_06589	Down	2	udp-glucose:flavonoid 3-o- glucosyltransferase	AT5G17050	UGT78D2 UDP-glucosyl transferase 78D2	460	236
QSDrought_06632	Up	5	inositol oxygenase 1-like	AT1G14520	MIOX1 myo-inositol oxygenase 1	311	254
QSDrought_06636	Up	11	zinc finger family protein	AT1G67856	RING/U-box superfamily protein	133	198
QSDrought_06651	Up	11	probable lrr receptor-like serine threonine-protein kinase rkf3-like	AT2G48010	RKF3 receptor-like kinase in in flowers 3	617	258
QSDrought_06653	Up	5	atp adp transporter	AT1G80300	ATNTT1_NTT1 nucleotide transporter 1	624	250
QSDrought_06654	Up	11	late embryogenesis abundant protein d-34-like	AT3G22490	Seed maturation protein	262	178
QSDrought_06655	Up	10	60s ribosomal protein	AT3G01170	Ribosomal protein L34e superfamily protein	215	231
QSDrought_06705	Up	5	adenine guanine permease azg1-like	AT3G10960	ATAZG1_AZG1 AZA-guanine resistant1	579	267
QSDrought_06716	Down	2	cxe carboxylesterase	AT1G47480	alpha/beta-Hydrolases superfamily protein	314	280
QSDrought_06720	Up	9	phytoene synthase	AT5G17230	PSY PHYTOENE SYNTHASE	437	220
QSDrought_06726	Up	11	ocs element-binding	AT2G22850	AtbZIP6_bZIP6 basic leucine-zipper 6	227	185
QSDrought_06729	Up	10	wrky transcription factor 17	AT2G38470	ATWRKY33_WRKY33 WRKY DNA- binding protein 33	519	596
QSDrought_06749	Up	11	class iv chitinase	AT3G54420	ATCHITIV_ATEP3_CHIV_EP3 homolog of carrot EP3-3 chitinase	273	264
QSDrought_06756	Down	2	anthranilate n- benzoyltransferase	AT1G24430	HXXXD-type acyl-transferase family protein	435	240
QSDrought_06822	Down	2	mitochondrial import inner	AT5G24650	Mitochondrial import inner	259	246

			membrane translocase subunit tim17 tim22 tim23 family protein		membrane translocase subunit Tim17/Tim22/Tim23 family protein		
QSDrought_06845	Up	9	polygalacturonase-1 non-catalytic subunit beta	AT1G23760	JP630_PG3 BURP domain-containing protein	622	210
QSDrought_06886	Up	11	serine threonine-protein kinase oxi1-like	AT3G25250	AGC2_AGC2-1_AtOXI1_OXI1 AGC (cAMP-dependent, cGMP-dependent and protein kinase C) kinase family protein	421	261
QSDrought_06895	Up	9	predicted protein	AT2G46940	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G62070.1); Has 143 Blast hits to 141 proteins in 14 species: Archae - 0; Bacteria - 0; Metazoa - 4; Fungi - 0; Plants - 139; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLINK).	252	238
QSDrought_06913	Down	8	Uncharacterized protein isoform 1	AT2G38820	Protein of unknown function (DUF506)	310	264
QSDrought_06966	Up	9	telomeric repeat binding	AT1G06910	TRFL7 TRF-like 7	390	189
QSDrought_06972	Up	10	PREDICTED: uncharacterized protein LOC100255921	AT1G13360	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G25870.1).	194	217
QSDrought_07023	Down	2	sucrose transport	AT1G22710	ATSUC2_SUC2_SUT1 sucrose-proton symporter 2	512	139
QSDrought_07044	Up	5	probable xyloglucan	AT5G57550	XTH25_XTR3 xyloglucan	284	236

			endotransglucosylase hydrolase protein 23-like		endotransglucosylase/hydrolase 25		
QSDrought_07065	Down	1	calcium-transporting atpase endoplasmic reticulum-type- like	AT1G07810	ACA3_ATECA1_ECA1 ER-type Ca2+-ATPase 1	1061	145
QSDrought_07074	Up	7	dead-box atp-dependent rna helicase 40-like	AT3G06480	DEAD box RNA helicase family protein	1088	257
QSDrought_07106	Up	10	syntaxin-24-like	AT2G35980	ATNHL10_NHL10_YLS9 Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family ATNHL10_NHL10_YLS9 Late embryogenesis abundant (LEA) hydroxyproline-rich glycoprotein family	227	222
QSDrought_07123	Down	5	tetraketide alpha-pyrone reductase 1-like	AT5G19440	NAD(P)-binding Rossmann-fold superfamily protein	326	265
QSDrought_07134	Up	5	p27-like protein	AT2G15220	Plant basic secretory protein (BSP) family protein	225	227
QSDrought_07212	Up	5	transcriptional corepressor	AT5G62090	SLK2 SEUSS-like 2	816	230
QSDrought_07215	Down	2	aspartyl protease family protein	AT5G10770	Eukaryotic aspartyl protease family protein Eukaryotic aspartyl protease family protein	474	221
QSDrought_07218	Up	3	phosphoprotein ecpp44-like	AT1G76180	ERD14 Dehydrin family protein ERD14 Dehydrin family protein	185	227
QSDrought_07266	Up	9	protein hva22-like	AT4G24960	ATHVA22D_HVA22D HVA22 homologue D	135	139
QSDrought_07328	Up	7	u-box domain-containing protein 35-like	AT4G25160	U-box domain-containing protein kinase family protein	835	261
QSDrought_07343	Down	6	calcium-activated outward- rectifying potassium channel 1-like	AT5G55630	ATKCO1_ATTPK1_KCO1_KCO1_TP K1 Outward rectifying potassium channel protein	363	186

					ATKCO1_ATTPK1_KCO1_KCO1_TP K1 Outward rectifying potassium channel protein		
QSDrought_07352	Up	10	scarecrow-like protein 1	AT2G29065	GRAS family transcription factor	630	235
QSDrought_07380	Down	6	auxin-induced in root cultures protein 12	AT3G07390	AIR12 auxin-responsive family protein	273	244
QSDrought_07393	Down	2	ribonucleoside-diphosphate reductase large subunit	AT2G21790	ATRNR1_CLS8_R1_RNR1 ribonucleotide reductase 1	816	204
QSDrought_07432	Down	1	endo- -beta-d-glucanase	AT5G49720	ATGH9A1_DEC_GH9A1_IRX2_KOR _KOR1_RSW2_TSD1 glycosyl hydrolase 9A1	621	220
QSDrought_07438	Down	4	fasciclin-like arabinogalactan protein 12	AT5G03170	ATFLA11_FLA11 FASCICLIN-like arabinogalactan-protein 11	246	247
QSDrought_07460	Down	1	allene oxide cyclase chloroplastic-like	AT1G13280	AOC4 allene oxide cyclase 4	254	183
QSDrought_07465	Up	11	eid1-like f-box protein 3-like	AT3G63060	EDL3 EID1-like 3	272	247
QSDrought_07472	Up	11	bidirectional sugar transporter sweet1-like	AT1G21460	AtSWEET1_SWEET1 Nodulin MtN3 family protein	247	221
QSDrought_07517	Down	1	beta-glucosidase 44-like	AT3G18070	BGLU43 beta glucosidase 43	501	216
QSDrought_07528	Down	6	sun domain-containing protein 1-like	AT5G04990	ATSUN1_SUN1 SAD1/UNC-84 domain protein 1	471	141
QSDrought_07590	Up	5	Proline-rich protein expressed in expanding leaves, stems, flowers, and siliques.	AT2G21140	ATPRP2_PRP2 proline-rich protein 2	321	176
QSDrought_07594	Up	9	probable rhamnogalacturonate lyase b-like	AT1G09890	Rhamnogalacturonate lyase family protein	617	78
QSDrought_07602	Down	6	nucleolysin tiar-like	AT5G54580	RNA-binding (RRM/RBD/RNP motifs) family protein	156	135
QSDrought_07617	Down	6	dihydrolipoyllysine-residue acetyltransferase component 1 of pyruvate dehydrogenase	AT3G52200	LTA3 Dihydrolipoamide acetyltransferase, long form protein	713	110

QSDrought_07640	Up	5	mitochondrial-like hypothetical protein POPTR_0009s01300g	AT5G60800	Heavy metal transport/detoxification superfamily protein	302	151
QSDrought_07707	Up	11	math and lrr domain- containing protein isoform 2	AT3G12320	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G06980.4); Has 102 Blast hits to 102 proteins in 16 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 98; Viruses - 0; Other Eukaryotes - 4 (source: NCBI BLink).	269	217
QSDrought_07727	Up	11	dna binding	AT5G03720	AT-HSFA3_HSFA3 heat shock transcription factor A3	412	169
QSDrought_07808	Up	10	PREDICTED: uncharacterized protein LOC101213956	AT4G38060	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G65480.1); Has 63 Blast hits to 63 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 63; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	136	171
QSDrought_07844	Up	5	temperature-induced lipocalin	AT5G58070	ATTIL_TIL temperature-induced lipocalin	186	185
QSDrought_07848	Up	11	protein phosphatase 2c 16-like	AT1G72770	HAB1 homology to ABI1	511	191
QSDrought_07868	Down	1	fatty acid 2-hydroxylase-like	AT2G34770	ATFAH1_FAH1 fatty acid hydroxylase 1	237	162
QSDrought_07892	Down	4	fasciclin-like arabinogalactan protein 17-like	AT3G52370	FLA15 FASCICLIN-like arabinogalactan protein 15 precursor	436	196
QSDrought_07904	Up	9	lipid binding	AT2G44300	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin	204	199

QSDrought_07906	Up	11	alcohol dehydrogenase	AT4G37980	superfamily protein ATCAD7_CAD7_ELI3_ELI3-1 elicitor-activated gene 3-1	357	208
QSDrought_07946	Up	11	transcription factor divaricata-like	AT5G04760	Duplicated homeodomain-like superfamily protein Duplicated homeodomain-like superfamily protein	215	223
QSDrought_07992	Up	7	serrate rna effector molecule-like	AT2G27100	SE C2H2 zinc-finger protein SERRATE (SE)	720	210
QSDrought_08021	Down	1	serine mitochondrial-like	AT5G26780	SHM2 serine hydroxymethyltransferase 2	533	158
QSDrought_08053	Up	11	late embryogenesis abundant protein family protein	AT1G52690	LEA7 Late embryogenesis abundant protein (LEA) family protein	169	166
QSDrought_08111	Up	9	amino acid binding	AT5G25320	ACT-like superfamily protein	500	122
QSDrought_08125	Down	8	cc-nbs-lrr class disease resistance	AT5G43470	HRT_RCY1_RPP8 Disease resistance protein (CC-NBS-LRR class) family	908	125
QSDrought_08155	Up	5	lipid transfer-like protein vas-like	AT1G62790	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	150	148
QSDrought_08207	Up	10	universal stress family protein	AT2G21620	RD2 Adenine nucleotide alpha hydrolases-like superfamily protein	193	176
QSDrought_08216	Down	6	amino-acid permease inda1				207
QSDrought_08217	Down	1	peroxidase 64-like	AT5G42180	PER64 Peroxidase superfamily protein	317	211
QSDrought_08235	Up	3	phase-change related protein				175
QSDrought_08330	Up	11	mft-like protein	AT1G18100	E12A11_MFT PEBP (phosphatidylethanolamine-binding protein) family protein	173	175
QSDrought_08409	Up	5	est gb	AT1G23710	Protein of unknown function (DUF1645)	295	239

QSDrought_08412	Up	10	f-box protein skip27-like	AT4G21510	F-box family protein	225	191
QSDrought_08531	Up	10	e3 ubiquitin-protein ligase march3	AT1G02610	RING/FYVE/PHD zinc finger superfamily protein	221	216
QSDrought_08554	Up	11	uncharacterized loc101219783	AT2G31940	unknown protein; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 16 plant structures; EXPRESSED DURING: 6 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G19875.1); Has 227 Blast hits to 227 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 227; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	120	129
QSDrought_08572	Up	3	phase-change related protein				207
QSDrought_08584	Up	5	—NA—	AT5G09530	PELPK1_PRP10 hydroxyproline-rich glycoprotein family protein	370	141
QSDrought_08624	Up	9	hypothetical protein POPTR_0017s03260g				142
QSDrought_08638	Down	6	adipor-like receptor cg5315-like	AT5G20270	HHP1 heptahelical transmembrane protein1	332	229
QSDrought_08665	Up	9	non-specific lipid-transfer protein at2g13820-like	AT3G22600	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	170	167
QSDrought_08800	Down	2	actin 1	AT5G09810	ACT7 actin 7	377	186
QSDrought_08821	Up	11	aldehyde dehydrogenase	AT1G23800	ALDH2B_ALDH2B7 aldehyde dehydrogenase 2B7	534	143
QSDrought_08831	Up	11	hva22-like protein k-like	AT4G36720	HVA22K HVA22-like protein K	200	185
QSDrought_08832	Down	1	ubiquitin-like protein	AT4G05050	UBQ11 ubiquitin 11 UBQ11	229	153

QSDrought_08888	Up	11	chloroplastic nifs-like cysteine desulfurase isoform 1	AT4G05050	ubiquitin 11 UBQ11 ubiquitin 11 UBQ11 ubiquitin 11 UBQ11	229	158
QSDrought_08928	Up	11	probable receptor-like protein kinase at5g47070-like	AT5G47070	ubiquitin 11 UBQ11 ubiquitin 11 Protein kinase superfamily protein	410	146
QSDrought_08994	Down	2	two-component response regulator arr2-like	AT4G16110	ARR2_RR2 response regulator 2	664	166
QSDrought_09000	Up	11	late embryogenesis abundant	AT3G51810	AT3_ATEM1_EM1_GEA1 Stress induced protein	152	132
QSDrought_09038	Up	11	blue copper	AT5G20230	ATBCB_BCB_BCB_SAG14 blue-copper-binding protein ATBCB_BCB_BCB_SAG14 blue-copper-binding protein	196	185
QSDrought_09056	Up	5	glycine-rich protein	AT5G61660	glycine-rich protein	134	151
QSDrought_09107	Down	1	—NA—				196
QSDrought_09220	Up	5	allene oxide chloroplastic	AT5G42650	AOS_CYP74A_DDE2 allene oxide synthase	518	148
QSDrought_09244	Up	5	ethylene-responsive transcription factor	AT5G44210	ATERF-9_ATERF9_ERF9 erf domain protein 9	200	201
QSDrought_09252	Up	5	lupus la	AT4G35890	winged-helix DNA-binding transcription factor family protein	523	198
QSDrought_09265	Up	10	histone deacetylase complex subunit sap18	AT2G45640	ATSAP18_SAP18 SIN3 associated polypeptide P18	152	151
QSDrought_09283	Up	5	—NA—				149
QSDrought_09329	Down	2	3-ketoacyl- synthase 11-like	AT1G04220	KCS2 3-ketoacyl-CoA synthase 2	528	197
QSDrought_09343	Down	2	PREDICTED: uncharacterized protein LOC100829005	AT5G45460	unknown protein; BEST Arabidopsis thaliana protein match is: Protein of unknown function (DUF594) (TAIR:AT5G45470.1); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria -	703	157

					1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).		
QSDrought_09366	Up	10	phosphatase 2c family protein	AT2G30020	Protein phosphatase 2C family protein	396	194
QSDrought_09420	Down	1	ferredoxin-nitrite chloroplastic-like	AT2G15620	ATHNIR_NIR_NIR1 nitrite reductase 1 ATHNIR_NIR_NIR1 nitrite reductase 1	586	166
QSDrought_09509	Up	5	big map kinase	AT3G45640	ATMAPK3_ATMPK3_MPK3 mitogen-activated protein kinase 3	370	141
QSDrought_09538	Up	10	gras family transcription factor	AT4G00150	ATHAM3_HAM3_LOM3 GRAS family transcription factor	558	140
QSDrought_09583	Down	6	methyl esterase	AT2G23600	ACL_ATME8_ATMES2_ME8_MES2 acetone-cyanohydrin lyase	263	149
QSDrought_09617	Up	5	eukaryotic translation initiation factor 3 subunit f-like	AT2G39990	AtelF3f_EIF2_eIF3F eukaryotic translation initiation factor 2	293	207
QSDrought_09652	Up	11	seed maturation protein lea 4	AT5G06760	AtLEA4-5_LEA4-5 Late Embryogenesis Abundant 4-5	158	171
QSDrought_09706	Down	1	drm3-like protein isoform 1	AT1G56220	Dormancy/auxin associated family protein	140	126
QSDrought_09728	Up	11	—NA—				96
QSDrought_09795	Up	3	blue copper	AT5G20230	ATBCB_BCB_BCB_SAG14 blue-copper-binding protein	196	182
QSDrought_09799	Up	9	f-box lrr-repeat protein 12	AT3G07550	RNI-like superfamily protein	395	214
QSDrought_09812	Up	5	hypothetical protein PRUPE_ppa009094mg	AT4G02040	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: chloroplast; Has 30 Blast hits to 30 proteins in 6 species: Archae - 0; Bacteria - 0;	152	142

QSDrought_09848	Up	9	conserved hypothetical protein	AT1G53380	Metazoa - 0; Fungi - 0; Plants - 30; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink). Plant protein of unknown function (DUF641)	453	145
QSDrought_09888	Down	2	g-type lectin s-receptor-like serine threonine-protein kinase at1g11300-like	AT1G11330	S-locus lectin protein kinase family protein	842	133
QSDrought_09914	Down	6	flavonoid 3 -monooxygenase-like	AT3G26330	CYP71B37 cytochrome P450, family 71, subfamily B, polypeptide 37	500	147
QSDrought_09952	Up	5	unknown protein	AT2G32210	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G32190.1); Has 214 Blast hits to 214 proteins in 21 species: Archae - 0; Bacteria - 0; Metazoa - 1; Fungi - 10; Plants - 203; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	71	71
QSDrought_09954	Up	5	lipid binding	AT2G37870	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	115	120
QSDrought_10028	Down	1	atp-citrate synthase beta chain protein 1-like	AT3G06650	ACLB-1 ATP-citrate lyase B-1	608	209
QSDrought_10044	Up	10	unnamed protein product	AT3G52240	unknown protein; Has 220 Blast hits to 193 proteins in 66 species: Archae - 0; Bacteria - 15; Metazoa - 53; Fungi - 33; Plants - 66; Viruses - 0; Other Eukaryotes - 53 (source: NCBI BLink).	680	147
QSDrought_10158	Up	11	subtilisin-like protease	AT5G09225	unknown protein; FUNCTIONS IN:	68	60

					molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLINK).		
QSDrought_10205	Down	2	bet v 1 allergen	AT1G24020	MLP423 MLP-like protein 423	155	160
QSDrought_10262	Up	7	protein maternal effect embryo arrest 60	AT5G05950	MEE60 maternal effect embryo arrest 60	175	173
QSDrought_10284	Down	8	dna repair protein rad51-like protein	AT2G45280	ATRAD51C_RAD51C RAS associated with diabetes protein 51C	387	172
QSDrought_10346	Up	11	nadph:quinone oxidoreductase	AT3G27890	NQR NADPH:quinone oxidoreductase	196	185
QSDrought_10586	Up	10	udp- rhamnose:rhamnosyltransfera se 1-like	AT2G22590	UDP-Glycosyltransferase superfamily protein	470	202
QSDrought_10631	Down	1	secologanin synthase-like	AT3G14610	CYP72A7 cytochrome P450, family 72, subfamily A, polypeptide 7	512	139
QSDrought_10632	Up	5	wrky 10	AT2G38470	ATWRKY33_WRKY33 WRKY DNA- binding protein 33	519	201
QSDrought_10669	Down	2	l-type lectin-domain containing receptor kinase -like	AT3G53380	Concanavalin A-like lectin protein kinase family protein	715	166
QSDrought_10681	Down	6	protein fam63b-like	AT4G11860	Protein of unknown function (DUF544)	682	99
QSDrought_10705	Down	6	long chain acyl- synthetase 1- like	AT2G47240	CER8_LACS1 AMP-dependent synthetase and ligase family protein	660	148

QSDrought_10723	Up	11	iron-binding protein	AT5G01600	ATFER1_FER1 ferretin 1	255	147
QSDrought_10759	Down	1	cellulose synthase	AT5G17420	ATCESA7_CESA7_IRX3_MUR10	1026	175
QSDrought_10770	Down	2	14 kda proline-rich protein	AT2G45180	Cellulose synthase family protein Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	134	139
QSDrought_10846	Up	7	germin-like protein 5	AT5G39110	RmlC-like cupins superfamily protein	222	179
QSDrought_10847	Down	6	low quality protein: trafficking protein particle complex subunit 11-like	AT5G65950	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 25 plant structures; EXPRESSED DURING: 15 growth stages; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF1683, C-terminal (InterPro:IPR012880), Foie gras liver health family 1 (InterPro:IPR021773); Has 30201 Blast hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).	1190	169
QSDrought_10866	Up	3	af334835_1proline-rich protein prp1				122
QSDrought_10869	Up	5	dehydration-responsive element-binding protein	AT3G50260	ATERF#011_CEJ1_DEAR1 cooperatively regulated by ethylene and jasmonate 1	153	148
QSDrought_10884	Down	3	—NA—				115

QSDrought_10997	Down	2	hypothetical protein PRUPE_ppa000819mg	AT1G19835	Plant protein of unknown function (DUF869)	982	103
QSDrought_11032	Down	6	atp binding	AT4G05150	Octicosapeptide/Phox/Bem1p family protein	477	188
QSDrought_11045	Down	1	patellin-4 isoform 1	AT1G30690	Sec14p-like phosphatidylinositol transfer family protein	540	169
QSDrought_11056	Down	7	phd finger family protein	AT5G20510	AL5 alfin-like 5	260	128
QSDrought_11089	Down	6	eg5651	AT1G32460	unknown protein; Has 19 Blast hits to 19 proteins in 8 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 19; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	92	86
QSDrought_11110	Up	10	mki67 fha domain-interacting nucleolar phospho	AT5G04600	RNA-binding (RRM/RBD/RNP motifs) family protein	222	143
QSDrought_11117	Down	6	probable receptor-like protein kinase at5g24010	AT2G21480	Malectin/receptor-like protein kinase family protein	871	185
QSDrought_11122	Up	10	asr5 protein				108
QSDrought_11123	Up	7	vq motif-containing protein	AT1G28280	VQ motif-containing protein	247	148
QSDrought_11137	Down	1	probable non-specific lipid- transfer protein akcs9-like	AT3G18280	Bifunctional inhibitor/lipid-transfer protein/seed storage 2S albumin superfamily protein	96	93
QSDrought_11209	Up	5	40s ribosomal protein s23-like	AT5G02960	Ribosomal protein S12/S23 family protein	142	142
QSDrought_11212	Down	2	glycosyl hydrolase family protein isoform 4	AT3G47000	Glycosyl hydrolase family protein	608	192
QSDrought_11265	Up	10	protein phosphatase regulatory	AT3G09880	ATB' BETA Protein phosphatase 2A regulatory B subunit family protein	499	79
QSDrought_11279	Down	2	—NA—				183
QSDrought_11432	Down	6	amino acid kinase family protein	AT1G26640	Amino acid kinase family protein	332	178

QSDrought_11490	Up	10	transducin wd40 repeat-like superfamily protein isoform 3	AT4G35560	DAW1 Transducin/WD40 repeat-like superfamily protein	1050	190
QSDrought_11591	Up	10	—NA—				22
QSDrought_11691	Up	3	Uncharacterized protein isoform 3, partial				93
QSDrought_11848	Up	11	major pollen allergen ory s 1	AT4G38400	ATEXLA2, ATEXPL2, ATHEXP BETA 2.2, EXLA2, EXPANSIN L2, EXPANSIN-LIKE A2, EXPL2	250	99
QSDrought_11853	Up	10	proteasome inhibitor-related	AT5G02020	SIS Encodes a protein involved in salt tolerance, names SIS (Salt Induced Serine rich).	149	135
QSDrought_11898	Up	5	eukaryotic peptide chain release factor subunit 1-3-like	AT3G26618	ERF1-3 eukaryotic release factor 1-3	435	123
QSDrought_11908	Up	5	protein phosphatase	AT1G34750	Protein phosphatase 2C family protein	282	149
QSDrought_11925	Down	2	ferric reduction oxidase chloroplastic-like	AT5G49740	ATFR07_FRO7 ferric reduction oxidase 7	747	134
QSDrought_11932	Up	9	PREDICTED: uncharacterized protein LOC100245830	AT2G25670	BEST Arabidopsis thaliana protein match is: copper ion binding (TAIR:AT4G32610.1); Has 43784 Blast hits to 26928 proteins in 1799 species: Archae - 86; Bacteria - 6347; Metazoa - 15971; Fungi - 5398; Plants - 1931; Viruses - 259; Other Eukaryotes - 13792 (source: NCBI BLink).	318	183
QSDrought_12049	Up	5	ring domain ligase2 isoform 1	AT5G14420	RGLG2 RING domain ligase2 RGLG2	468	158
QSDrought_12067	Up	5	bzip transcription factor 60-like	AT1G42990	ATBZIP60_BZIP60_BZIP60 basic region/leucine zipper motif 60	295	144
QSDrought_12112	Up	11	protein phosphatase	AT5G59220	HAI1_SAG113 highly ABA-induced	413	62

PP2C gene 1

QSDrought_12121	Up	9	hypothetical protein VITISV_037878				60
QSDrought_12254	Up	5	basic blue protein	AT2G02850	ARPN plantacyanin	129	125
QSDrought_12281	Up	11	zinc finger ccch domain- containing protein 11-like	AT2G20280	Zinc finger C-x8-C-x5-C-x3-H type family protein Zinc finger C-x8-C-x5- C-x3-H type family protein	371	103
QSDrought_12322	Up	5	early responsive to dehydration protein	AT1G20450	ERD10_LTI29_LTI45 Dehydrin family protein	260	132
QSDrought_12402	Up	10	malate transporter	AT5G47560	ATSDAT_ATTDT_TDT tonoplast dicarboxylate transporter	540	146
QSDrought_12465	Down	6	myrosinase-binding protein	AT1G73040	Mannose-binding lectin superfamily protein Mannose-binding lectin superfamily protein	176	156
QSDrought_12505	Down	6	uncharacterized loc101203522	AT3G48200	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; Has 210 Blast hits to 148 proteins in 42 species: Archae - 0; Bacteria - 118; Metazoa - 0; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 44 (source: NCBI BLink). unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane	1088	140

					system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; Has 210 Blast hits to 148 proteins in 42 species: Archae - 0; Bacteria - 118; Metazoa - 0; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 44 (source: NCBI BLink).		
QSDrought_12549	Up	5	eukaryotic translation initiation factor 1a-like	AT2G04520	Nucleic acid-binding, OB-fold-like protein	145	68
QSDrought_12566	Down	6	histone h3	AT5G65360	Histone superfamily protein	136	136
QSDrought_12698	Up	11	uncharacterized membrane protein at3g27390-like	AT4G37030	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G12680.1); Has 101 Blast hits to 99 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 101; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	569	143
QSDrought_12736	Up	9	dna binding	AT5G47390	myb-like transcription factor family protein	365	68
QSDrought_12784	Up	11	hypothetical protein POPTR_0098s003001g, partial	AT1G22370	AtUGT85A5_UGT85A5 UDP-glucosyl transferase 85A5	479	161
QSDrought_12799	Up	11	fkbp-type peptidyl-prolyl cis-trans isomerase	AT5G64350	ATFKBP12_FKBP12_FKP12 FK506-binding protein 12	112	118

QSDrought_12848	Down	7	hypothetical protein PRUPE_ppa005811mg	AT5G19950	Domain of unknown function (DUF1767)	443	173
QSDrought_12890	Up	11	udp- isoform 2	AT5G12890	UDP-Glycosyltransferase superfamily protein	488	123
QSDrought_12911	Up	11	beta- -galactosyltransferase 7-like	AT1G77810	Galactosyltransferase family protein	393	137
QSDrought_12956	Up	11	—NA—				70
QSDrought_12966	Down	2	soluble diacylglycerol acyltransferase	AT1G48300	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; Has 96 Blast hits to 87 proteins in 37 species: Archae - 0; Bacteria - 2; Metazoa - 12; Fungi - 2; Plants - 65; Viruses - 0; Other Eukaryotes - 15 (source: NCBI BLink).	285	97
QSDrought_12976	Down	1	gtp-binding family protein	AT1G01200	ATRA3-A3_ATRABA3_RABA3 RAB GTPase homolog A3	237	171
QSDrought_13155	Up	9	transcription factor divaricata-like	AT5G04760	Duplicated homeodomain-like superfamily protein Duplicated homeodomain-like superfamily protein	215	140
QSDrought_13178	Up	5	plastocyanin-like domain-containing family protein	AT4G28365	AtENODL3_ENODL3 early nodulin-like protein 3	199	169
QSDrought_13237	Up	11	dna binding	AT5G64750	ABR1 Integrase-type DNA-binding superfamily protein	391	129
QSDrought_13240	Up	11	brix domain-containing protein	AT5G61770	PPAN PETER PAN-like protein	346	153

QSDrought_13253	Up	5	iron-sulfur cluster assembly enzyme mitochondrial-like	AT4G22220	ATISU1_ISU1 SufE/NifU family protein	167	151
QSDrought_13320	Up	10	nadh dehydrogenase	AT4G34700	AtCIB22_CIB22 LYR family of Fe/S cluster biogenesis protein	117	119
QSDrought_13343	Up	10	f-box kelch-repeat protein at2g44130-like	AT2G44130	Galactose oxidase/kelch repeat superfamily protein	409	163
QSDrought_13344	Up	5	2-hydroxyacyl- lyase-like	AT5G17380	Thiamine pyrophosphate dependent pyruvate decarboxylase family protein	572	141
QSDrought_13380	Down	7	unnamed protein product	AT4G35510	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT2G17540.3); Has 182 Blast hits to 179 proteins in 73 species: Archae - 0; Bacteria - 87; Metazoa - 17; Fungi - 9; Plants - 50; Viruses - 0; Other Eukaryotes - 19 (source: NCBI BLink).	333	157
QSDrought_13395	Down	4	major latex family protein	AT1G70870	Polyketide cyclase/dehydrase and lipid transport superfamily protein	139	91
QSDrought_13469	Up	10	global transcription factor c isoform 1	AT4G10710	SPT16 global transcription factor C	1074	95
QSDrought_13482	Down	2	mannose-binding lectin superfamily protein	AT1G73040	Mannose-binding lectin superfamily protein Mannose-binding lectin superfamily protein	176	135
QSDrought_13535	Up	10	—NA—				54
QSDrought_13589	Up	9	transcription factor asg4-like	AT4G01280	Homeodomain-like superfamily protein	303	138
QSDrought_13706	Down	6	class iv family protein	AT5G27410	D-aminoacid aminotransferase-like PLP-dependent enzymes superfamily protein	559	136
QSDrought_13845	Up	10	nucleic acid binding	AT5G53800	unknown protein; Has 30201 Blast	351	151

hits to 17322 proteins in 780 species: Archae - 12; Bacteria - 1396; Metazoa - 17338; Fungi - 3422; Plants - 5037; Viruses - 0; Other Eukaryotes - 2996 (source: NCBI BLink).

QSDrought_13851	Down	6	outward rectifying potassium channel	AT5G55630	ATKCO1_ATTPK1_KCO1_KCO1_TP K1 Outward rectifying potassium channel protein ATKCO1_ATTPK1_KCO1_KCO1_TP K1 Outward rectifying potassium channel protein	363	124
QSDrought_13979	Up	11	hydroxyproline-rich glycoprotein family isoform 3	AT2G32840	proline-rich family protein	337	125
QSDrought_14000	Down	2	benzyl alcohol o-benzoyltransferase-like	AT5G17540	HXXXD-type acyl-transferase family protein	461	159
QSDrought_14019	Up	3	Dehydrin ERD10				138
QSDrought_14029	Down	6	urea-proton symporter dur3-like	AT5G45380	ATDUR3_DUR3 solute:sodium symporters;urea transmembrane transporters	704	110
QSDrought_14166	Up	11	brassinosteroid insensitive 1-associated receptor kinase 1	AT2G33580	Protein kinase superfamily protein	664	156
QSDrought_14181	Down	6	dual specificity protein phosphatase phs1-like	AT5G23720	PHS1 dual specificity protein phosphatase family protein	929	125
QSDrought_14394	Down	6	ribonucleoside-diphosphate reductase small chain-like	AT3G27060	ATTSO2_TSO2 Ferritin/ribonucleotide reductase-like family protein	332	80
QSDrought_14407	Up	10	rubber elongation factor	AT3G05500	Rubber elongation factor protein (REF)	246	123
QSDrought_14458	Up	9	phospholipid glycerol acyltransferase family protein	AT1G80950	Phospholipid/glycerol acyltransferase family protein	398	112

QSDrought_14492	Up	10	dna-binding protein	AT1G13260	EDF4_RAV1 related to ABI3/VP1 1	344	154
QSDrought_14595	Up	10	unknown mRNA				153
QSDrought_14606	Down	4	abc transporter g family member 39-like	AT2G36380	ABCG34_ATPDR6_PDR6 pleiotropic drug resistance 6	1453	108
QSDrought_14660	Down	1	nadh dehydrogenase	AT4G02580	NADH-ubiquinone oxidoreductase 24 kDa subunit, putative	255	59
QSDrought_14793	Down	2	subtilisin-like serine protease	AT2G05920	Subtilase family protein	754	103
QSDrought_14899	Up	3	early responsive to dehydration protein	AT1G76180	ERD14 Dehydrin family protein	185	139
QSDrought_14965	Down	6	beta-glucosidase 24-like	AT3G60130	BGLU16 beta glucosidase 16	514	138
QSDrought_15066	Up	11	iq-domain 3 isoform 3	AT5G03040	iqd2 IQ-domain 2	461	107
QSDrought_15166	Up	3	hypothetical protein				147
QSDrought_15198	Up	11	auxin efflux carrier family	AT1G20925	Auxin efflux carrier family protein	472	146
QSDrought_15206	Up	5	60s ribosomal protein l37	AT3G16080	Zinc-binding ribosomal protein family protein	95	95
QSDrought_15239	Up	5	rna-binding (rrm rbd rnp motifs) family protein isoform 1	AT5G42905	Polynucleotidyl transferase, ribonuclease H-like superfamily protein	258	48
QSDrought_15246	Up	9	beta-hydroxylase isoform partial	AT4G25700	B1_BCH1_BETA-OHASE 1_chy1 beta-hydroxylase 1	310	104
QSDrought_15252	Down	8	sister chromatid cohesion 2	AT5G15540	ATSCC2_EMB2773_SCC2 PHD finger family protein	1844	48
QSDrought_15263	Up	9	sequence-specific dna binding transcription factors	AT3G58630	sequence-specific DNA binding transcription factors	321	131
QSDrought_15304	Down	5	—NA—				45
QSDrought_15428	Up	11	cop1-interacting protein 7	AT4G27430	CIP7 COP1-interacting protein 7	1058	71
QSDrought_15436	Up	5	sphingoid base hydroxylase 2	AT1G69640	SBH1 sphingoid base hydroxylase 1	260	75
QSDrought_15572	Down	7	importin subunit alpha-1-like	AT4G16143	IMPA-2 importin alpha isoform 2	535	61
QSDrought_15599	Up	10	—NA—				140
QSDrought_15629	Up	9	at1g06530-like protein	AT1G06530	PMD2 Tropomyosin-related	323	99
QSDrought_15750	Up	11	metalloendoproteinase 1-like	AT1G70170	MMP matrix metalloproteinase	378	142

QSDrought_15791	Down	1	metallophosphatase family protein	AT1G13750	Purple acid phosphatases superfamily protein	613	141
QSDrought_15811	Up	7	hypothetical protein RCOM_1959270	AT4G13390	EXT12 Proline-rich extensin-like family protein	429	127
QSDrought_15879	Down	6	uncharacterized loc101203522	AT3G48200	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; Has 210 Blast hits to 148 proteins in 42 species: Archae - 0; Bacteria - 118; Metazoa - 0; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 44 (source: NCBI BLink). unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 21 plant structures; EXPRESSED DURING: 13 growth stages; Has 210 Blast hits to 148 proteins in 42 species: Archae - 0; Bacteria - 118; Metazoa - 0; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 44 (source: NCBI BLink).	1088	92
QSDrought_15889	Up	11	signal recognition particle 14 kda	AT2G43640	Signal recognition particle, SRP9/SRP14 subunit	121	113

QSDrought_15924	Up	11	zinc finger ccch domain-containing protein 11-like	AT2G20280	Zinc finger C-x8-C-x5-C-x3-H type family protein Zinc finger C-x8-C-x5-C-x3-H type family protein	371	126
QSDrought_15987	Up	7	armadillo repeat-containing protein 7-like	AT5G37290	ARM repeat superfamily protein	180	91
QSDrought_16000	Down	8	chloroplast ribosomal protein l22				106
QSDrought_16021	Up	11	metal ion binding	AT5G23760	Copper transport protein family	103	77
QSDrought_16132	Down	6	ion channel pollux-like 2-like	AT5G43745	Protein of unknown function (DUF1012)	817	111
QSDrought_16207	Down	1	subtilase family protein	AT3G14067	Subtilase family protein	777	114
QSDrought_16270	Up	5	—NA—				101
QSDrought_16303	Up	11	hypothetical protein PRUPE_ppa013922mg	AT2G31945	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 8 plant structures; EXPRESSED DURING: 4 anthesis, petal differentiation and expansion stage; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G05575.1); Has 61 Blast hits to 61 proteins in 10 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 61; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLINK).	95	92
QSDrought_16321	Up	11	aldehyde dehydrogenase	AT3G24503	ALDH1A_ALDH2C4_REF1 aldehyde dehydrogenase 2C4	501	113
QSDrought_16378	Up	11	—NA—				109

QSDrought_16382	Down	8	hypothetical protein SORBIDRAFT_10g005070	AT1G51172	unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; CONTAINS InterPro DOMAIN/s: Protein of unknown function DUF601 (InterPro:IPR006736); BEST Arabidopsis thaliana protein match is: Protein of unknown function, DUF601 (TAIR:AT1G46696.1); Has 1849 Blast hits to 1554 proteins in 272 species: Archae - 15; Bacteria - 116; Metazoa - 677; Fungi - 267; Plants - 154; Viruses - 11; Other Eukaryotes - 609 (source: NCBI BLink).	681	135
QSDrought_16485	Up	11	cytochrome family subfamily polypeptide	AT4G37360	CYP81D2 cytochrome P450, family 81, subfamily D, polypeptide 2	499	79
QSDrought_16513	Down	4	organ-specific protein s2				134
QSDrought_16548	Down	6	—NA—				61
QSDrought_16625	Up	9	zinc finger ccch domain- containing protein 19-like	AT2G16485	nucleic acid binding;zinc ion binding;DNA binding	1773	91
QSDrought_16653	Up	9	—NA—				97
QSDrought_16717	Up	5	probable galactinol-sucrose galactosyltransferase 6-like	AT5G20250	DIN10_RS6 Raffinose synthase family protein DIN10_RS6 Raffinose synthase family protein	844	91
QSDrought_16809	Up	11	p23 co-chaperone protein	AT3G03773	HSP20-like chaperones superfamily protein	204	74
QSDrought_16818	Up	5	annexin-like protein rj4-like	AT5G12380	ANNAT8 annexin 8	316	87

QSDrought_16841	Up	11	isoform 1 nadp-specific glutamate	AT4G24110	unknown protein; Has 76 Blast hits to 76 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 1; Fungi - 0; Plants - 75; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	250	95
QSDrought_16911	Up	9	atp binding	AT1G26580	FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: cellular_component unknown; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 15 growth stages; BEST Arabidopsis thaliana protein match is: ELM2 domain-containing protein (TAIR:AT2G03470.1); Has 161 Blast hits to 161 proteins in 17 species: Archae - 0; Bacteria - 0; Metazoa - 1; Fungi - 4; Plants - 156; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	493	68
QSDrought_16917	Up	9	chitin-inducible gibberellin-responsive protein 1-like	AT5G48150	PAT1 GRAS family transcription factor	490	126
QSDrought_16983	Up	11	calcium-dependent protein kinase	AT3G57530	ATCPK32_CDPK32_CPK32 calcium-dependent protein kinase 32	538	71
QSDrought_17204	Down	1	probable pectate lyase 15-like	AT4G13710	Pectin lyase-like superfamily protein	451	125
QSDrought_17304	Down	2	homeodomain-like superfamily protein isoform 2	AT1G09710	Homeodomain-like superfamily protein	610	65
QSDrought_17373	Up	9	mannosyl-oligosaccharide -	AT1G27520	Glycosyl hydrolase family 47 protein	574	59

QSDrought_17384	Up	11	alpha-mannosidase phytosulfokin receptor 1	AT2G02220	ATPSKR1_PSKR1 phytosulfokin receptor 1	1008	48
QSDrought_17452	Up	5	—NA—				90
QSDrought_17648	Up	10	ring-h2 finger protein	AT5G20885	RING/U-box superfamily protein	176	105
QSDrought_17808	Down	8	ring finger	AT5G40250	RING/U-box superfamily protein	376	65
QSDrought_17826	Down	2	probable lrr receptor-like serine threonine-protein kinase at1g63430-like	AT5G41180	leucine-rich repeat transmembrane protein kinase family protein	664	121
QSDrought_17873	Down	7	probable 26s proteasome regulatory subunit	AT4G24820	26S proteasome, regulatory subunit Rpn7;Proteasome component (PCI) domain	387	63
QSDrought_17955	Up	11	galactinol synthase	AT2G47180	AtGolS1_GolS1 galactinol synthase 1	344	83
QSDrought_17986	Up	9	2-oxoglutarate and fe - dependent oxygenase superfamily protein isoform 1	AT4G16330	2-oxoglutarate (2OG) and Fe(II)- dependent oxygenase superfamily protein	364	91
QSDrought_18075	Up	11	dna binding	AT1G30320	Remorin family protein	509	115
QSDrought_18101	Down	4	sucrose synthase 2	AT5G20830	ASUS1_atsus1_SUS1 sucrose synthase 1	808	63
QSDrought_18104	Down	8	unknown	AT1G48170	unknown protein; Has 78 Blast hits to 78 proteins in 34 species: Archae - 0; Bacteria - 0; Metazoa - 30; Fungi - 0; Plants - 48; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	153	57
QSDrought_18123	Up	10	eukaryotic peptide chain release factor	AT3G26618	ERF1-3 eukaryotic release factor 1-3	435	50
QSDrought_18193	Up	11	dehydrin 3	AT3G50980	XERO1 dehydrin xero 1 XERO1 dehydrin xero 1	128	73
QSDrought_18216	Down	6	geranylgeranyl transferase	AT2G39550	ATGGT-IB_GGB_PGGT-I	375	87

QSDrought_18406	Down	1	type-1 subunit beta-like organ-specific protein s2-like	AT1G49310	Prenyltransferase family protein unknown protein; FUNCTIONS IN: molecular_function unknown; INVOLVED IN: biological_process unknown; LOCATED IN: endomembrane system; EXPRESSED IN: 9 plant structures; EXPRESSED DURING: LP.04 four leaves visible, petal differentiation and expansion stage; Has 11 Blast hits to 11 proteins in 6 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 11; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	82	102
QSDrought_18459	Up	9	—NA—				96
QSDrought_18532	Down	2	malectin receptor protein kinase family protein				83
QSDrought_18573	Up	5	ubiquitin-like protein	AT3G62250	UBQ5 ubiquitin 5	157	78
QSDrought_18810	Up	11	—NA—				106
QSDrought_18838	Up	10	dag protein	AT4G20020	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT5G44780.1); Has 35333 Blast hits to 34131 proteins in 2444 species: Archae - 798; Bacteria - 22429; Metazoa - 974; Fungi - 991; Plants - 531; Viruses - 0; Other Eukaryotes - 9610 (source: NCBI BLink).	406	107
QSDrought_18899	Up	11	u-box domain-containing protein 19-like	AT1G10560	ATPUB18_PUB18 plant U-box 18	697	51

QSDrought_18928	Down	7	dna binding	AT1G35460	basic helix-loop-helix (bHLH) DNA-binding superfamily protein	259	106
QSDrought_18948	Up	11	PREDICTED: uncharacterized protein LOC100242324	AT2G01300	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT1G15010.1); Has 73 Blast hits to 73 proteins in 12 species: Archae - 0; Bacteria - 0; Metazoa - 0; Fungi - 0; Plants - 73; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	156	83
QSDrought_18969	Down	1	golgin candidate 6-like	AT3G27530	GC6_MAG4 golgin candidate 6	914	105
QSDrought_19011	Up	10	stress-induced hydrophobic peptide	AT3G05890	RCI2B Low temperature and salt responsive protein family	54	60
QSDrought_19058	Down	2	serine carboxypeptidase-like 40-like	AT3G63470	scpl40 serine carboxypeptidase-like 40	502	100
QSDrought_19071	Up	9	serine-rich adhesin for platelets-like isoform x2	AT1G36990	unknown protein; LOCATED IN: chloroplast; EXPRESSED IN: 24 plant structures; EXPRESSED DURING: 14 growth stages; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT4G08510.1); Has 5029 Blast hits to 1779 proteins in 339 species: Archae - 2; Bacteria - 1372; Metazoa - 990; Fungi - 933; Plants - 111; Viruses - 28; Other Eukaryotes - 1593 (source: NCBI BLink).	581	61
QSDrought_19073	Up	10	—NA—				69
QSDrought_19187	Up	11	—NA—				71
QSDrought_19201	Up	5	predicted protein	AT2G23090	Uncharacterised protein family	78	57

QSDrought_19227	Up	10	v-type proton atpase subunit c-like	AT1G12840	SERF ATVHA-C_DET3 vacuolar ATP synthase subunit C (VATC) / V-ATPase C subunit / vacuolar proton pump C subunit (DET3)	375	102
QSDrought_19366	Up	10	asr5 protein				90
QSDrought_19441	Up	11	—NA—				54
QSDrought_19659	Up	11	dehydrin 1	AT3G50980	XERO1 dehydrin xero 1 XERO1 dehydrin xero 1	128	84
QSDrought_19709	Up	7	—NA—				51
QSDrought_19775	Up	11	btb poz domain-containing protein isoform 2	AT4G15840	BTB/POZ domain-containing protein	836	66
QSDrought_19810	Up	10	—NA—				78
QSDrought_19848	Down	2	-n-methylcoclaurine 3 -hydroxylase isozyme 2-like	AT5G06905	CYP712A2 cytochrome P450, family 712, subfamily A, polypeptide 2	521	54
QSDrought_19895	Up	9	puromycin-sensitive aminopeptidase-like	AT4G33090	APM1_ATAPM1 aminopeptidase M1	879	46
QSDrought_19940	Down	6	flavonoid 3 -monooxygenase-like	AT5G06900	CYP93D1 cytochrome P450, family 93, subfamily D, polypeptide 1	507	63
QSDrought_20020	Up	9	g-box binding factor	AT2G46270	GBF3 G-box binding factor 3	382	74
QSDrought_20090	Up	10	low temperature and salt responsive protein lti6b	AT3G05890	RCI2B Low temperature and salt responsive protein family	54	58
QSDrought_20140	Up	9	u-box domain-containing protein 33-like	AT2G45910	U-box domain-containing protein kinase family protein	834	73
QSDrought_20184	Up	9	probable splicing factor 3a subunit 1-like	AT1G14650	SWAP (Suppressor-of-White-APricot)/surp domain-containing protein / ubiquitin family protein	785	45
QSDrought_20295	Up	11	e3 ubiquitin-protein ligase march6-like	AT4G34100	RING/U-box superfamily protein	1107	65
QSDrought_20376	Up	5	—NA—				52

QSDrought_20392	Up	5	—NA—				72
QSDrought_20436	Up	11	Uncharacterized protein TCM_014589	AT5G63350	unknown protein; BEST Arabidopsis thaliana protein match is: unknown protein (TAIR:AT3G48510.1); Has 103 Blast hits to 102 proteins in 13 species: Archae - 0; Bacteria - 0; Metazoa - 2; Fungi - 0; Plants - 101; Viruses - 0; Other Eukaryotes - 0 (source: NCBI BLink).	302	69
QSDrought_20484	Up	10	—NA—				70
QSDrought_20489	Up	11	low quality protein: udp- glycosyltransferase 85a7-like	AT1G22370	AtUGT85A5_UGT85A5 UDP- glucosyl transferase 85A5	479	69
QSDrought_20555	Up	9	probable serine threonine- protein kinase ddb_g0279405-like	AT5G60550	ATSNAK1_GRIK2 geminivirus rep interacting kinase 2	407	67
QSDrought_20788	Up	9	—NA—				16
QSDrought_20882	Up	9	—NA—				52
QSDrought_20885	Up	5	PEB3				34
QSDrought_20977	Up	11	—NA—				43
QSDrought_20990	Down	4	PEA34				40